EL3370 Mathematical Methods in Signals, Systems and Control

Topic 3: Normed Spaces

Cristian R. Rojas

Division of Decision and Control Systems
KTH Royal Institute of Technology

Outline

Outline

Motivation and Definitions

Inner product spaces are useful (and easy to handle), but are not the only spaces of interest in system theory (*e.g.*, $C[a,b]$, H_∞, ... more on this last space later in the course). The metrics of these other spaces cannot be induced by inner products, but it is still possible to define a *norm* on them.

Definition
Let V be a real (complex) vector space. A *norm* on V is a mapping $\|\cdot\|: V \to \mathbb{R}_0^+$ s.t., for all $x, y \in V$ and $\lambda \in \mathbb{R}$ ($\mathbb{C}$),

(i) $\|x\| > 0$ if $x \neq 0$, and $\|0\| = 0$,

(ii) $\|\lambda x\| = |\lambda| \|x\|$,

(iii) $\|x + y\| \leqslant \|x\| + \|y\|$. (*triangle inequality*)

$(V, \|\cdot\|)$ is a *normed space*.

Example
Let X be a topological space, and $C(X) := \{f : X \to \mathbb{C} : f \text{ is continuous and bounded}\}$. $C(X)$ is then a vector space, and we can define the *supremum norm* $\|f\|_\infty := \sup_{x \in X} |f(x)|$.

Example

Consider the vector space $\mathbb{R}^n$. For $x \in \mathbb{R}^n$, its *p-norms* are defined as

$$\begin{aligned} \|x\|_1 &:= |x_1| + \cdots + |x_n|, \\ \|x\|_p &:= \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}, \qquad p \in [1,\infty) \\ \|x\|_\infty &:= \max\{|x_1|, \ldots, |x_n|\}. \end{aligned}$$

Difficulty: How to prove the triangle inequality for $1 \leqslant p < \infty$?

For $1 \leqslant p \leqslant \infty$, let $1 \leqslant q \leqslant \infty$ be s.t. $1/p + 1/q = 1$, and take $x, y \in \mathbb{R}^n$

Hölder inequality: $\displaystyle\sum_{k=1}^n |x_k y_k| \leqslant \|x\|_p \|y\|_q$ (with equality iff $\dfrac{|x_k|^p}{\|x\|_p^p} = \dfrac{|y_k|^q}{\|y\|_q^q}$ for all k)

Minkowski inequality: $\|x + y\|_p \leqslant \|x\|_p + \|y\|_p$

(*See bonus slides for proofs of these inequalities*.)

Example

Let $1 \leqslant p < \infty$. ℓ_p is the normed space of all sequences $x = (x_k)$ s.t. $\sum_{k=1}^{\infty} |x_k|^p < \infty$, together with the norm

$$\|x\|_p := \left(\sum_{k=1}^{\infty} |x_k|^p \right)^{1/p}, \qquad x = (x_k) \in \ell_p.$$

For $p = \infty$ we define ℓ_∞ as the normed space of all *bounded* sequences $x = (x_k)$ (*i.e.*, there is an $M > 0$ s.t. $|x_k| \leqslant M$ for all k), with the norm

$$\|x\|_\infty := \sup_{k \in \mathbb{N}} |x_k|, \qquad x = (x_k) \in \ell_\infty.$$

Observation

For $p < \infty$, the properties of the ℓ_p norm rely on an extension of Minkowski's inequality.

Exercise: Prove that $\|\cdot\|_\infty$ satisfies the triangle inequality.

Theorem

In a normed space $(V, \|\cdot\|)$, the function $d: V \times V \to \mathbb{R}_0^+$ given by $d(x,y) := \|x-y\|$ is a *translation-invariant* metric (*i.e.*, for every $x,y,z \in V$, $d(x+z, y+z) = d(x,y)$).

Proof. For every $x,y,z \in V$,

(1) $d(x,y) \geqslant 0$, and $d(x,y) = 0$ iff $\|x-y\| = 0$, or equivalently, iff $x = y$.

(2) $d(x,y) = \|x-y\| = |-1|\|x-y\| = \|y-x\| = d(y,x)$.

(3) $d(x,z) = \|x-z\| = \|x-y+y-z\| \leqslant \|x-y\| + \|y-z\| = d(x,y) + d(y,z)$.

Thus, d is a metric, and $d(x+z, y+z) = \|x+z-y-z\| = \|x-y\| = d(x,y)$, so it is translation-invariant. $\square$

Consequence

A normed space is a metric space, and inherits its *topological*/*convergence* properties.

Exercise: Prove that in a normed space V, the norm $\|\cdot\|: V \to \mathbb{R}_0^+$ is continuous.

Theorem
In a real normed space V, the addition $+\colon V \times V \to V$ and scalar multiplication $\cdot\colon \mathbb{R} \times V \to V$ are continuous operations (with respect to the product topologies of $V \times V$ and $\mathbb{R} \times V$, respectively).

Proof (for scalar multiplication; for addition the proof is similar)
Let $\varepsilon > 0$, and fix $\lambda \in \mathbb{R}$ and $x \in V$. For every $\mu \in \mathbb{R}$ and $y \in V$,

$$\|\lambda x - \mu y\| = \|\lambda x - \mu x + \mu x - \mu y\| \leqslant |\lambda - \mu| \|x\| + |\mu| \|x - y\|.$$

Then, if we define the open sets

$$U_\lambda := \left\{\mu \in \mathbb{R} \colon |\mu - \lambda| < \min\left(1, \frac{\varepsilon}{2(1 + \|x\|)}\right)\right\}, \quad V_x := \left\{y \in V \colon \|y - x\| < \frac{\varepsilon}{2(1 + |\lambda|)}\right\},$$

we have that for $\mu \in U_\lambda$ and $y \in V_x$: $|\mu| < |\lambda| + 1$ and

$$\|\lambda x - \mu y\| < \frac{\varepsilon}{2} \frac{\|x\|}{1 + \|x\|} + \mu \frac{\varepsilon}{2(1 + |\lambda|)} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

$\square$

Outline

Closed Linear Subspaces

Some properties that hold for finite dimensional normed spaces are not always valid in infinite dimensions. *E.g.*, in $\mathbb{C}^n$, linear subspaces are always closed (*i.e.*, if (x_n) is a sequence in a linear subspace, and $x_n \to x$, then x belongs to that subspace).

Example
Let ℓ_0 be the set of sequences (x_n) in ℓ_2 which have only a finite number of nonzero terms. Then ℓ_0 is a linear subspace of ℓ_2, but it is not closed: Take $x^k := (1, 1/2, 1/3, \dots, 1/k, 0, 0, \dots) \in \ell_0$. Then, $x^k \to x := (1, 1/2, 1/3, \dots) \in \ell_2$, because

$$\|x^k - x\| = \sqrt{\sum_{n=k+1}^{\infty} 1/n^2} \longrightarrow 0.$$

However, all the terms in x are nonzero, so $x \notin \ell_0$.

Theorem

The closure of a linear subspace (of a normed space) is also a linear subspace.

Proof. Let F be a linear subspace of V, and take $x, y \in \overline{F}$. Every nbd of x has an element of F, so there is a sequence (x_n) in F s.t. $x_n \to x$ (similarly, there is a $y_n \to y$). Since addition and scalar multiplication are continuous, $x_n + y_n \to x + y$ and $\lambda x_n \to \lambda x$ for every λ, and these limits belong to $\overline{F}$ (because it is closed). Hence, $\overline{F}$ is a linear subspace. □

Reminder

Let V be a normed space, and let $A \subseteq V$. The *linear span* of A, lin A, is the set of all (finite) linear combinations of points in A, *i.e.*,

$$\text{lin } A = \left\{ \sum_{n=1}^{m} \lambda_n a_n : m \in \mathbb{N}; \lambda_1, \dots, \lambda_m \in \mathbb{R}; a_1, \dots, a_m \in A \right\},$$

or, equivalently, the intersection of all linear subspaces that contain A.

Definition. The *closed linear span* of A, clin A, is the intersection of all closed linear subspaces containing A.

Theorem
Let V be a normed space. For every $A \subseteq V$, clin A is the closure of lin A.

Proof. Since the closure of lin A is a closed linear subspace that contains A, it has to contain clin A. Conversely, clin A is closed and contains lin A, thus clin $A \supseteq \overline{\text{lin } A}$; to see this, note that if $x \in [\text{clin } A]^c$, then there is a nbd of x completely contained in $[\text{clin } A]^c$ (because this set is open), so $x \notin \overline{\text{lin } A}$. This shows that clin $A = \overline{\text{lin } A}$. □

In finite dimensional vector spaces, topological issues are exactly the same as in $\mathbb{R}^n$:

Theorem. Every two norms in a real or complex finite dimensional space V generate the same topology.

Proof (for $\mathbb{R}$). Let $\{e_1, \ldots, e_n\}$ be a basis of V, and define the norm $\rho\left(\sum_{k=1}^n \lambda_k e_k\right) := \sqrt{\sum_{k=1}^n \lambda_k^2}$ for all $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. We will show that for every norm $\|\cdot\|$ on V there exist $K_1, K_2 > 0$ s.t.

$$K_1 \rho(x) \leqslant \|x\| \leqslant K_2 \rho(x)$$

for every $x \in V$ (*i.e.*, every open set in (V, ρ) is open in $(V, \|\cdot\|)$, and vice versa).

Proof (cont.)

1. $(\|x\| \leqslant K_2 \rho(x))$
   If $x = \sum_{k=1}^n \lambda_k e_k$, then

$$\|x\| = \left\| \sum_{k=1}^n \lambda_k e_k \right\| \leqslant \sum_{k=1}^n \|\lambda_k e_k\| = \sum_{k=1}^n |\lambda_k| \, \|e_k\| \leqslant \sqrt{\sum_{k=1}^n \lambda_k^2} \sqrt{\sum_{k=1}^n \|e_k\|^2} = K_2 \rho(x),$$

   where we can take $K_2 := \sqrt{\sum_{k=1}^n \|e_k\|^2}$.

2. $(K_1 \rho(x) \leqslant \|x\|)$
   We will prove that $\inf_{x \neq 0} \|x\|/\rho(x) > 0$. Since both norms scale with $x = \sum_{k=1}^n \lambda_k e_k$, we can restrict ourselves to the compact set $K = \left\{(\lambda_1, \dots, \lambda_n) \colon \sum_{k=1}^n \lambda_k^2 = 1\right\}$ (where $\rho(x) = 1$). On K, $\inf_{x \in K} \|x\| = \min_{x \in K} \|x\| > 0$, since otherwise there is an $x_0 = \sum_{k=1}^n \lambda_k^0 e_k$ s.t. $\|x_0\| = 0$, *i.e.*, $x_0 = 0$, which means that $\{e_1, \dots, e_n\}$ are l.d., which is a contradiction. Therefore, we can take $K_1 = \inf_{x \in K} \|x\| > 0$. $\square$

Corollary (Heine-Borel theorem for normed spaces)
In a finite-dimensional normed space $(V, \|\cdot\|)$, a set A is compact iff it is closed and bounded.

Proof
The norm ρ (from the proof of the previous theorem) makes (V, ρ), and thus $(V, \|\cdot\|)$, homeomorphic to $\mathbb{R}^n$, so A is closed in $(V, \|\cdot\|)$ iff it is closed in (V, ρ). Also, since $K_1 \rho(x) \leqslant \|x\| \leqslant K_2 \rho(x)$ for every $x \in V$, A is bounded in $(V, \|\cdot\|)$ iff it is bounded in (V, ρ). Heine-Borel can then be applied. $\square$

Theorem. The closed unit ball $\overline{B(0,1)} := \{x\colon \|x\| \leqslant 1\}$ in an infinite-dimensional normed space X is *not* compact.

Proof. Assume that $\overline{B(0,1)}$ is compact; we will show that $\dim X < \infty$. As $(B(x,2^{-1}))_{x\in X}$ is an open cover of $\overline{B(0,1)}$, there are $x_1,\dots,x_N \in B(0,1)$ s.t. $\overline{B(0,1)} \subseteq \bigcup_{k=1}^{n} B(x_k,2^{-1})$. Since $B(x_k,2^{-1}) = x_k + 2^{-1}B(0,1)$, it follows that $B(0,1) \subseteq \overline{B(0,1)} \subseteq Y + 2^{-1}B(0,1)$, where $Y = \operatorname{lin}\{x_1,\dots,x_N\}$. Thus,

$$B(0,1) \subseteq Y + 2^{-1}[Y + 2^{-1}B(0,1)] = Y + 2^{-2}B(0,1) \subseteq \cdots \subseteq Y + 2^{-3}B(0,1).$$

In general, $B(0,1) \subseteq Y + 2^{-n}B(0,1)$ for every $n \in \mathbb{N}$, so each $x \in B(0,1)$ can be written as $x = y_n + x_n$, where $y_n \in Y$ and $x_n \in B(0,2^{-n})$. As $x_n \to 0$, we have that $y_n \to x$, and $x \in \overline{Y} = Y$ (by Homework 2, every finite-dimensional subspace of a normed space is closed). This shows that $B(0,1) \subseteq Y$, which implies that $X \subseteq Y$, thus $X = Y$, so $\dim X = \dim Y = N < \infty$. □

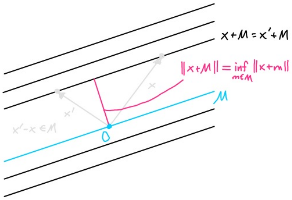
Notation. Let V be a vector space, $A,B \subseteq V$ and $x \in V$. Then, $x + A := \{x + y\colon y \in A\}$ and $A + B := \{x + y\colon x \in A,\ y \in B\}$.

Quotient Spaces

Let V be a vector space over $\mathbb{F}$, and $M \subseteq V$ a subspace. One can define an equivalence relation on V by $x \sim y$ iff $x - y \in M$ ("x and y are equivalent modulo M"). This relation partitions V into equivalence classes / cosets, corresponding to the *translates* $[x] := x + M$.

Definition. The *quotient space* V/M is the vector space consisting of the cosets $[x]$ of M in V, with the operations $[x] + [y] = [x + y]$ and $\alpha[x] = [\alpha x]$ for all $x, y \in V$ and $\alpha \in \mathbb{F}$.
The *co-dimension* of M in V is the dimension of V/M.

Exercise: Prove that these operations are well-defined, that V/M is a vector space, and that, if $\dim V < \infty$, then $\dim V/M = \dim V - \dim M$.



If V is a normed space, and $M \subseteq V$ is a closed subspace, V/M can be turned into a normed space with

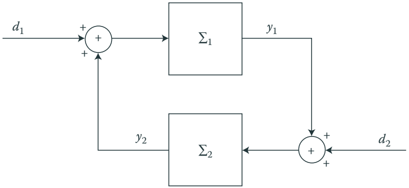
$$\|[x]\| := \inf_{m \in M} \|x + m\|, \quad x \in V.$$

The assumption that M is closed is needed to ensure that $\|[x]\| > 0$ if $[x] \neq [0]$.

Exercise: Prove that this is a norm on V/M.

Outline

Consider the following feedback interconnection:



d_1, d_2, y_1 and y_2 are signals, while Σ_1 and Σ_2 are systems, *i.e.*, mappings between signal spaces. A *signal* f is a real sequence, *i.e.*, $f: \mathbb{N} \to \mathbb{R}$, and its *truncation* $f_\tau: \mathbb{N} \to \mathbb{R}$ ($\tau \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$) is

$$f_\tau(n) := \begin{cases} f(n), & n \leqslant \tau, \\ 0, & n > \tau. \end{cases}$$

(Note that truncated sequences lie in ℓ_0.)

We want conditions on Σ_1 and Σ_2 to ensure that the feedback interconnection is stable. To this end, we first need to define stability...

Definition. Let Σ be a system with input u and output y, *i.e.*, $y = \Sigma(u)$. Σ is *stable* (with respect to the norm $\|\cdot\|$ in ℓ_0) if there is a *gain function* $\gamma\colon \mathbb{R}_0^+ \to \mathbb{R}_0^+$ which is continuous, non-decreasing and s.t. $\gamma(0) = 0$ and

$$\|y_\tau\| \leqslant \gamma(\|u_\tau\|), \text{ for all } \tau \in \mathbb{N}_0.$$

To study the stability of a feedback interconnection, we need some more definitions:

Definitions

(a) The *graph* of Σ is $G_\Sigma := \{(u, y)\colon\ y = \Sigma(u)\}$.

(b) The *inverse graph* of Σ is $G_\Sigma^I := \{(y, u)\colon\ y = \Sigma(u)\}$.

(c) G_Σ and G_Σ^I are subsets of an underlying normed space X, called the *ambient space*, where a norm can be defined as: $\|(u, y)_\tau\| := \|u_\tau\| + \|y_\tau\|$ (here, $(u, y)_\tau = (u_\tau, y_\tau)$).

(d) A *feedback interconnection* (Σ_1, Σ_2) is *well-defined* if, for all pairs of signals (d_1, d_2), there exist signals y_1, y_2 s.t. $y_1 = \Sigma_1(d_1 + y_2)$ and $y_2 = \Sigma_2(d_2 + y_1)$.

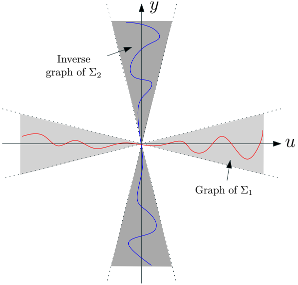
Theorem (Separation of graphs)
A well-defined interconnection (Σ_1, Σ_2) is stable iff there is a gain function γ s.t.

$$x \in G^I_{\Sigma_2} \quad \Longrightarrow \quad \|x_\tau\| \leqslant \gamma(d_\tau(x, G_{\Sigma_1})), \quad \text{for all } \tau \in \mathbb{N}_0, \qquad (*)$$

where $d_\tau(x, G_\Sigma) := \inf_{z \in G_\Sigma} \|(x-z)_\tau\|$.

To understand this fundamental theorem, consider the following figure:

The graph of Σ_1 is the subset of the Cartesian space of input-output pairs (u, y), where $y = \Sigma_1(u)$.
Similarly, the inverse graph of Σ_2 consists of those pairs (u, y) where $u = \Sigma_2(y)$.
The graph separation theorem says that (Σ_1, Σ_2) is stable if these two graphs do not intersect each other, except at the origin, and that the separation between these graphs should increase as one goes further away from the origin.



If the systems are not known exactly, stability can be guaranteed by imposing disjoint regions where the graphs of and are known to lie, as shown, *e.g.*, by the shaded cones.

Proof
Since the interconnection is well-defined, given (d_1, d_2) there are signals y_1, y_2 s.t. $y_1 = \Sigma_1(d_1 + y_2)$ and $y_2 = \Sigma_2(d_2 + y_1)$. Then, let

$$x = (y_2, y_1 + d_2) \in G^I_{\Sigma_2},$$

$$z = (y_2 + d_1, y_1) \in G_{\Sigma_1},$$

so $\|(x - z)_\tau\| = \|(-d_1, d_2)_\tau\| = \|(d_1, d_2)_\tau\|$, and $(*)$ becomes equivalent to

$$\left\|(y_2, y_1 + d_2)_\tau\right\| \leqslant \gamma\left(\left\|(d_1, d_2)_\tau\right\|\right), \quad \text{for all } \tau \in \mathbb{N}_0,$$

or, alternatively, to

$$\left\|(y_2)_\tau\right\| + \left\|(y_1 + d_2)_\tau\right\| \leqslant \gamma\left(\left\|(d_1)_\tau\right\| + \left\|(d_2)_\tau\right\|\right), \quad \text{for all } \tau \in \mathbb{N}_0.$$

If (Σ_1, Σ_2) is stable, then there is a gain function $\tilde{\gamma}$ s.t. $\left\|(y_1)_\tau\right\| + \left\|(y_2)_\tau\right\| \leqslant \tilde{\gamma}\left(\left\|(d_1)_\tau\right\| + \left\|(d_2)_\tau\right\|\right)$ for all $\tau \in \mathbb{N}_0$.
Therefore,

$$\left\|(y_2)_\tau\right\| + \left\|(y_1 + d_2)_\tau\right\| \leqslant \left\|(y_2)_\tau\right\| + \left\|(y_1)_\tau\right\| + \left\|(d_2)_\tau\right\| \leqslant \tilde{\gamma}\left(\left\|(d_1)_\tau\right\| + \left\|(d_2)_\tau\right\|\right) + \left\|(d_1)_\tau\right\| + \left\|(d_2)_\tau\right\|,$$

and the right-hand side becomes $\gamma\left(\left\|(d_1)_\tau\right\| + \left\|(d_2)_\tau\right\|\right)$ if we take $\gamma(x) := \tilde{\gamma}(x) + x$. Hence $(*)$ holds.

Proof (cont.)

Conversely, if $(*)$ holds, we have that, by the triangle inequality for the norm,

$$\begin{aligned}
\|(y_1)_\tau\| + \|(y_2)_\tau\| &\leqslant \|(y_1 + d_2)_\tau\| + \|(-d_2)_\tau\| + \|(y_2)_\tau\| \\
&= \|(y_1 + d_2)_\tau\| + \|(d_2)_\tau\| + \|(y_2)_\tau\| \\
&\leqslant \gamma\left(\|(d_1)_\tau\| + \|(d_2)_\tau\|\right) + \|(d_2)_\tau\| \\
&\leqslant \gamma\left(\|(d_1)_\tau\| + \|(d_2)_\tau\|\right) + \|(d_1)_\tau\| + \|(d_2)_\tau\| \\
&\leqslant \tilde{\gamma}\left(\|(d_1)_\tau\| + \|(d_2)_\tau\|\right),
\end{aligned}$$

where we have defined $\tilde{\gamma}(x) := \gamma(x) + x$. This shows that (Σ_1, Σ_2) is stable. $\square$

Application: Input-Output Stability (cont.)

Remarks

(a) This simple result pioneered the use of functional analysis in control theory. See, *e.g.*,

W.S. Levine. The Control Handbook, 2nd Ed., CRC Press, 2011,

M.G. Safonov. Stability and Robustness of Multivariable Feedback Systems, MIT Press, 1980.

(b) In spite of its simplicity, the graph separation theorem contains as special cases most sufficient conditions for stability, such as the small gain theorem, passivity theory, the Nyquist criterion, the Popov circle criterion, Lyapunov stability and integral quadratic constraints!

(c) The robustness of stability in feedback connections was intensively studied in the 1980's. Based on the graph separation theorem, a natural approach was developed by A.K. El-Sakkary in 1985 based on the so-called "gap metric" on the ambient space where the graphs of Σ_1 and Σ_2 lie. However, this metric is not easy to compute, so later G. Vinnicombe developed a new metric, the "ν-gap" in 1993, which induces the same topology as the gap metric but is computationally more tractable.

Next Topic

Hilbert and Banach Spaces

Outline

Motivation and Definitions

Closed Linear Subspaces

Application: Input-Output Stability

Bonus Slides

Theorem (Young's inequality) For all $a, b \geqslant 0$, where $1/p + 1/q = 1$,

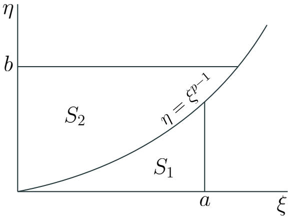
$$ab \leqslant \frac{a^p}{p} + \frac{b^q}{q},$$

with equality iff $a^p = b^q$.

Proof. Consider the figure to the right. The curve satisfies $\eta = \xi^{p-1}$ or $\xi = \eta^{1/(p-1)} = \eta^{q-1}$. The areas S_1 and S_2 are given by

$$S_1 = \int_0^a \xi^{p-1} d\xi = \frac{a^p}{p}, \quad S_2 = \int_0^b \eta^{q-1} d\eta = \frac{b^q}{q}.$$

It is clear from the figure that $ab \leqslant S_1 + S_2$, which implies the inequality. Equality holds iff $b = a^{p-1} = a^{p/q}$, or equivalently, iff $a^p = b^q$ (since $p/q = p(1-1/p) = p-1$). $\square$



Bonus: Proof of Hölder and Minkowski's Inequalities (cont.)

Proof of Hölder's inequality: $\displaystyle\sum_{k=1}^{n} |x_k y_k| \leqslant \|x\|_p \|y\|_q$

The inequality is trivial if $\|x\|_p = 0$ or $\|y\|_q = 0$. Otherwise, let us divide the inequality by the right-hand side, giving $\sum_{k=1}^{n} |\tilde{x}_k \tilde{y}_k| \leqslant 1$, with $\tilde{x}_k = x_k/\|x\|_p$ and $\tilde{y}_k = y_k/\|y\|_q$. This expression follows from Young's inequality, since

$$\sum_{k=1}^{n} |\tilde{x}_k \tilde{y}_k| \leqslant \sum_{k=1}^{n} \left(\frac{|\tilde{x}_k|^p}{p} + \frac{|\tilde{y}_k|^q}{q} \right) = \frac{1}{p} \|\tilde{x}\|_p^p + \frac{1}{q} \|\tilde{y}\|_q^q = \frac{1}{p} + \frac{1}{q} = 1,$$

with equality iff $\dfrac{|x_k|^p}{\|x\|_p^p} = \dfrac{|y_k|^q}{\|y\|_q^q}$ for all k. $\square$

Bonus: Proof of Hölder and Minkowski's Inequalities (cont.)

Proof of Minkowski's inequality: $\|x+y\|_p \leqslant \|x\|_p + \|y\|_p$

$$
\begin{aligned}
&\sum_{k=1}^{n} |x_k + y_k|^p \\
&\leqslant \sum_{k=1}^{n} |x_k||x_k + y_k|^{p-1} + \sum_{k=1}^{n} |y_k||x_k + y_k|^{p-1} \\
&\leqslant \left(\sum_{k=1}^{n} |x_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |x_k + y_k|^p\right)^{1/q} + \left(\sum_{k=1}^{n} |y_k|^p\right)^{1/p} \left(\sum_{k=1}^{n} |x_k + y_k|^p\right)^{1/q} \quad \text{(Hölder's ineq., with } (p-1)q = p\text{)} \\
&= \left(\|x\|_p + \|y\|_p\right) \|x+y\|_p^{p/q}.
\end{aligned}
$$

Since the left side is $\|x+y\|_p^p$, diving both sides by $\|x+y\|_p^{p/q}$ and noting that $p - p/q = p(1-1/q) = p/p = 1$ gives Minkowski's inequality. $\square$

Bonus: Hierarchy of ℓ_p Spaces

Theorem

If $1 \leqslant p_1 < p_2 \leqslant \infty$, then $\ell_{p_1} \subseteq \ell_{p_2}$.

Proof

Take $x \in \ell_{p_1}$. Then, $\|x\|_{p_1}^{p_1} = \sum_{k=1}^{\infty} |x_k|^{p_1} < \infty$, so there exists an $N \in \mathbb{N}$ s.t. $|x_k|^{p_1} < 1$ for all $k \geqslant N$, or, equivalently, $|x_k| < 1$. Therefore, $|x_k|^{p_2} = |x_k|^{p_1} |x_k|^{p_2 - p_1} < |x_k|^{p_1}$ for all $k \geqslant N$, so

$$\sum_{k=1}^{\infty} |x_k|^{p_2} = \sum_{k=1}^{N-1} |x_k|^{p_2} + \sum_{k=N}^{\infty} |x_k|^{p_2} < \sum_{k=1}^{N-1} |x_k|^{p_2} + \sum_{k=N}^{\infty} |x_k|^{p_1} < \infty.$$

This means that $x \in \ell_{p_2}$, so in general we have that $\ell_{p_1} \subseteq \ell_{p_2}$. $\square$