I.11 Norms of Vectors ~~and Functions~~ and Matrices

A <u>norm</u> is a function $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}$ that satisfies:

(1) $\|v\| > 0$ if $v \neq 0$,

(2) $\|cv\| = |c| \cdot \|v\|$ for $v \in \mathbb{R}^n$ and $c \in \mathbb{R}$,

(3) $\|v+w\| \leq \|v\| + \|w\|$ (Triangle inequality).

Note that (2) implies that $\|0\| = 0$.

<u>Examples:</u>

2-norm (aka Euclidean norm):

$$\|v\|_2 = \sqrt{|v_1|^2 + \cdots + |v_n|^2} = \sqrt{v^* v}.$$

1-norm:

$$\|v\|_1 = |v_1| + \cdots + |v_n|.$$

p-norm:

$$\|v\|_p = \left(|v_1|^p + \cdots + |v_n|^p\right)^{1/p}.$$

Here $1 \leq p < \infty$. If $0 < p < 1$, then $\|\cdot\|_p$ is not a norm (triangle inequality fails).

If we let $p \to \infty$, we have

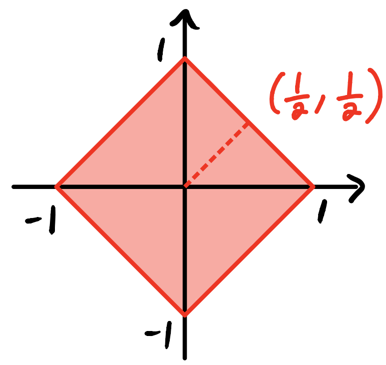
$$\|v\|_\infty = \max_{i=1,\dots,n} |v_i| .$$

∞-norm:

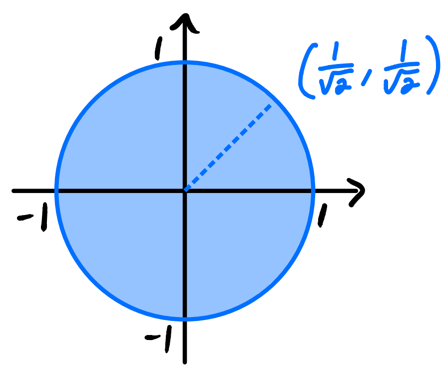
$$\|v\|_\infty = \max_{i=1,\dots,n} |v_i| .$$

The set of vectors with $\|v\|_p \le 1$ is different for different p.

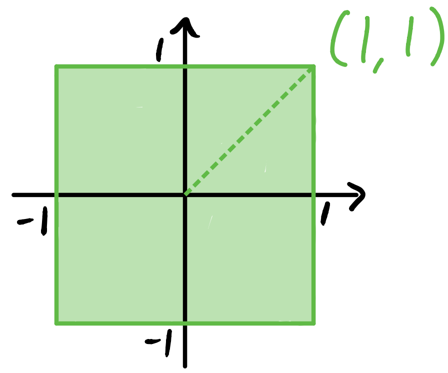
1-norm: $|v_1| + |v_2| \le 1$



2-norm: $v_1^2 + v_2^2 \le 1$



∞-norm: $|v_1| \le 1, \ |v_2| \le 1$

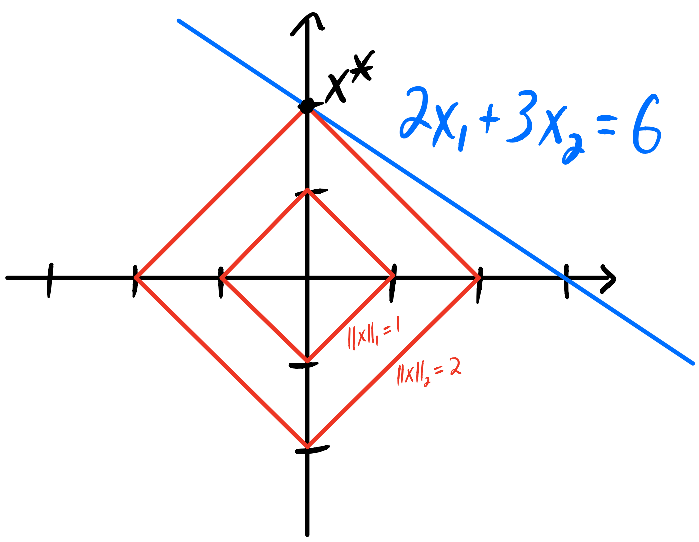


The 1-norm has a very important property that the optimization problem

$$\boxed{\min \ \|x\|_1 \quad \text{s.t.} \quad Ax = b}$$

always has an optimal sol'n x^* having at least one entry 0 (ie $x_i^* = 0$ for some i), and often x^* has many zero entries (ie x^* is sparse).

We can easily see why this is in $\mathbb{R}^2$. Suppose $A = [2 \ \ 3]$ and $b = [6]$.



$$\boxed{\begin{array}{ll} \min & \|x\|_1 \\ \text{s.t.} & 2x_1 + 3x_2 = 6 \end{array}}$$

$$x^* = (0, 2)$$

$$\|x^*\|_1 = 2$$

Grow the 1-norm ball until it touches the line.

Clearly the 1-norm ball will always touch any line at some point x^* where $x^* = (a, 0)$ or $x^* = (0, b)$.

Inner Products and Norms

The 2-norm is defined by the <u>inner product</u> $\langle v, w \rangle = v^T w = v_1 w_1 + \cdots + v_n w_n$:

$$\|v\| = \sqrt{v^T v}.$$

(no subscript means 2-norm)

We can prove the <u>triangle inequality</u>

$$\|v + w\| \leq \|v\| + \|w\|$$

using the <u>Cauchy-Schwarz inequality</u>

$$|v^T w| \leq \|v\| \|w\|$$

[Recall that $v^T w = \|v\| \|w\| \cos\theta$, where $0 \leq \theta \leq \pi$ is the angle between v and w.]

Pf of the triangle inequality:

$$\begin{aligned}\|v+w\|^2 &= (v+w)^T(v+w)\\ &= v^Tv + v^Tw + w^Tv + w^Tw\\ &= \|v\|^2 + 2v^Tw + \|w\|^2\\ &\le \|v\|^2 + 2\|v\|\|w\| + \|w\|^2\\ &= (\|v\| + \|w\|)^2\end{aligned}$$

Another inner product is the <u>S inner product</u>,

$$\langle v, w\rangle_S = v^T S w,$$

where S is a symmetric positive definite matrix. We then get the <u>S-norm</u>,

$$\|v\|_S = \sqrt{\langle v, v\rangle_S} = \sqrt{v^T S v}.$$

We can use S-norms to do <u>weighted least-squares</u>.

Norms of Matrices

A <u>matrix norm</u> must satisfy

(1) $\|A\| > 0$ if $A \neq 0$

(2) $\|cA\| = |c|\,\|A\|$ for scalars c

(3) $\|A+B\| \leq \|A\| + \|B\|$

(4) $\|AB\| \leq \|A\|\,\|B\|$ ← new property

The <u>Frobenius norm</u> of $A \in \mathbb{R}^{m \times n}$ is

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2 \right)^{1/2}$$

$$= \sqrt{\operatorname{tr}(A^T A)}$$

$$= \sqrt{\sigma_1^2 + \cdots + \sigma_r^2}.$$

The corresponding <u>inner product</u> is

$$\langle A, B \rangle = \operatorname{tr}(A^T B) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}.$$

Then $\|A\|_F = \sqrt{\langle A, A \rangle}$.

Induced Matrix Norms

If $\|\cdot\|$ is a vector norm, the corresponding <u>induced matrix norm</u> is

$$\|A\| = \max_{v \neq 0} \frac{\|Av\|}{\|v\|}.$$

Then $\|Av\| \leq \|A\|\|v\|$.

<u>Examples:</u>

2-norm: $\|A\|_2 = \sigma_1$

1-norm: $\|A\|_1 =$ largest 1-norm of the <u>columns</u> of A

∞-norm: $\|A\|_\infty =$ largest 1-norm of the <u>rows</u> of A

Matrix Norms and Singular Values

Define $\sigma(A) = \begin{bmatrix} \sigma_1 \\ \vdots \\ \sigma_r \end{bmatrix}$ where $\sigma_1 \geq \cdots \geq \sigma_r > 0$

are the singular values of A.

Then $\|A\|_F = \|\sigma(A)\|_2$

and $\|A\|_2 = \|\sigma(A)\|_\infty$.

What about $\|\sigma(A)\|_1$? This is the <u>nuclear norm</u> of A:

$$\|A\|_N = \sigma_1 + \cdots + \sigma_r .$$

Just as minimizing the 1-norm of a vector gives us a sparse vector, <u>minimizing the nuclear norm</u> gives us a <u>low rank matrix</u>.

<u>Example:</u> The Netflix Problem

$$X = \begin{array}{c|ccc} & \text{Movie } 1 & \cdots & \text{Movie } n \\ \hline \text{Person } 1 & 4 & \cdots & 2 \\ \text{Person } 2 & 1 & \cdots & 5 \\ \vdots & & & \\ \text{Person } m & 3 & \cdots & 1 \end{array}$$

We do not have ratings from all people for all movies. We want to fill in the missing ratings so we can give movie recommendations.

Since there are only a small number of different types of people and movies, we expect the <u>rating matrix</u> X to have low rank.

$$\boxed{\min \ \|X\|_N \quad \text{s.t.} \quad X_{ij} = r_{ij}, \quad (i,j) \in I}$$

Here I is the set of (i,j) where person i has rated movie j, and that rating is denoted r_{ij}.