Lecture 3

Norms

(P-1)

Vector Norms

Norms encapsulate the notion of size and distances in a vector space.

e.g. errors, similarity of graphs or images, etc., are measured by norms.

Def^n: A vector norm is a function $\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ that assigns a real-valued length of a vector and satisfies the following conditions:

a) $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$
b) $\|x+y\| \leq \|x\| + \|y\|$ (Triangle inequality)
c) $\|\alpha x\| = |\alpha| \, \|x\|$, $\forall \alpha \in \mathbb{C}$

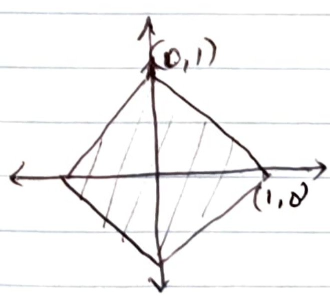
P-norms

Let $x \in \mathbb{C}^n$

$$\|x\|_p = \left(\sum |x_i|^p\right)^{1/p}$$

1-norm

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|$$



$\|x\|_1 \leq 1$

$|x_1| + |x_2| \leq 1$

2-norm

$$\|x\|_2 = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$

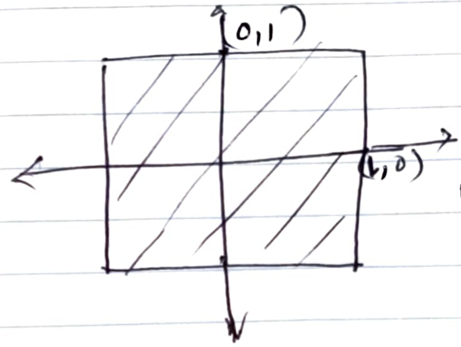
Circle of radius 1

$\|x\|_2 \leq 1$

$\left(|x_1|^2 + |x_2|^2\right)^{1/2} \leq 1$

∞ - norm

$$\|x\|_\infty = \max_{1 \le i \le n} |x_i|$$



$\|x\|_\infty \le 1$

$\max(|x_1|, |x_2|) \le 1$

Relevance

Linear Solvers : $Ax = b \rightarrow$ Find an approximate $\tilde{x}$ such that $\|A\tilde{x} - b\| < \varepsilon$

linear legression : $(x_I, y_I) \rightarrow$ Traning samples

Learn $y(x)$. Approximate $y = Wx + b$

$$\min_{W,b} \sum_I \| y(x_I) - y_I \|_p^p = \min_{W,b} \sum_I \| W x_I + b - y_I \|_p^p$$

Matrix Norms

$A \in \mathbb{C}^{m \times n}$ can be viewed as an mn-dimensional vector, and hence, any mn-dimensional norm can be used to measure the size of ~~of~~ A.

However, in the context of matrices there are more useful and interesting norms that we can define.

Induced Matrix Norms

Induced matrix norms are defined in terms of the effect of the matrix on a vector.

Let $A \in \mathbb{C}^{m \times n}$, then the norm on the domain of A can be denoted as $\|\cdot\|^{(n)}$ (because A acts on $x \in \mathbb{C}^n$). Similarly, the norm on the range of A can be represented as $\|\cdot\|^{(m)}$ (because $Ax \in \mathbb{C}^m$).

The induced matrix norm of A, denoted as $\|A\|^{(m,n)}$, is the smallest number C which satisfies :

$$\|Ax\|^{(m)} \leq C \|x\|^{(n)} \quad \forall x \in \mathbb{C}^n \qquad -\ (1)$$

In other words, $\|A\|^{(m,n)}$ represents the maximum factor by which A can stretch a vector x. We say that $\|\cdot\|^{(m,n)}$ is the matrix norm induced by $\|\cdot\|^{m}$ and $\|\cdot\|^{(n)}$.

From (1), note that $\|A\|^{(m,n)} = C$ can be taken as the maximum of $\dfrac{\|Ax\|^{(m)}}{\|x\|^{n}}$, i.e

$$\|A\|^{(m,n)} = \max_{\substack{x \in \mathbb{C}^n \\ x \neq 0}} \frac{\|Ax\|^{(m)}}{\|x\|^{(n)}} = \max_{\substack{x \in \mathbb{C}^n \\ \|x\|^{(n)} = 1}} \|Ax\|^{(m)}$$

Why ? (Define $x = \|x\|^{(n)} \hat{x}$, where $\hat{x}$ is a unit vector and substitute above)

Examples

$$A = \begin{bmatrix} 1 & k \\ 0 & k \end{bmatrix}$$

1 - Norm :

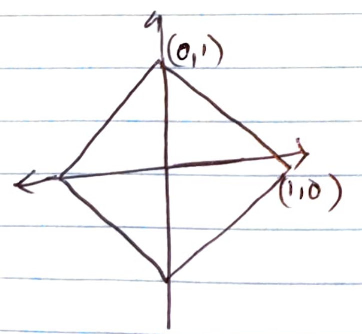
Let $x = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$ s.t. $\|x\|_1 = 1 \Rightarrow |\alpha| + |\beta| = 1$

$$\|A\|_1^{(2,2)} = \max_{\substack{\alpha,\beta \\ |\alpha|+|\beta|=1}} \|Ax\|_1^{(2)} = \max_{\substack{\alpha,\beta \\ |\alpha|+|\beta|=1}} |\alpha + k\beta| + |k\beta|$$

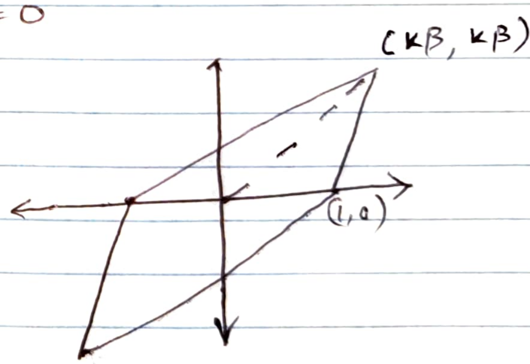
$$\leq \max_{\substack{\alpha,\beta \\ |\alpha|+|\beta|=1}} |\alpha| + 2|k||\beta| \quad \text{(Triangle inequality)}$$

$$= \max_{\beta} 1 + (2|k| - 1)|\beta|$$

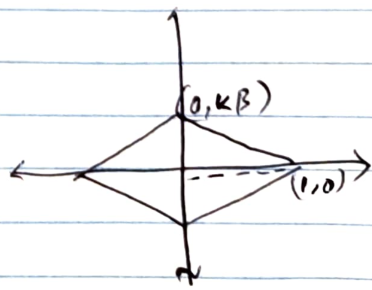
$|k| > 1/2$: $|\beta| = 1$ $\qquad$ $|k| < 1/2$: $\beta = 0$



$|k| > 1/2$



$|k| < 1/2$



$$\|A\|_1^{(2,2)} = \begin{cases} 2|k| & , \text{if } |k| > 1/2 \\ 1 & , \text{otherwise} \end{cases}$$

2-norm

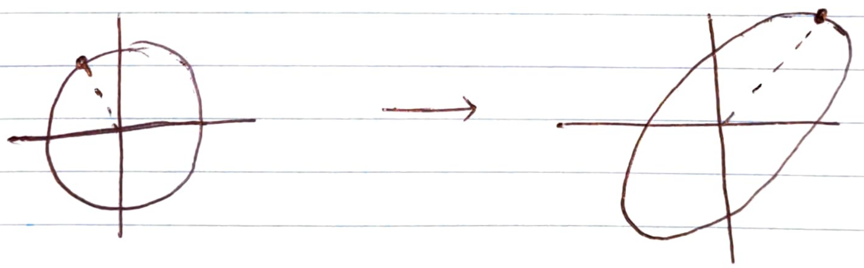
Let $x = \begin{bmatrix} \sin\theta \\ \cos\theta \end{bmatrix}$ (why ?) $\|x\|_2 = \sin^2\theta + \cos^2\theta = 1$

$$\|A\|_2^{(2,2)} = \max_{\theta} \underbrace{\left[(\sin\theta + k\cos\theta)^2 + (k\cos\theta)^2 \right]}_{f(\theta)}^{1/2}$$

Finding θ that maximizes $f(\theta)^{1/2}$ is same as finding θ that maximizes $f(\theta)$.

So find θ that maximizes $(\sin\theta + k\cos\theta)^2 + (k\cos\theta)^2$

$$\Rightarrow \quad \theta^* = \frac{1}{2}\tan^{-1}\left(\frac{2k}{2k^2-1}\right)$$



$$\|A\|_2^{(2,2)} = f(\theta^*)^{1/2}$$

∞-norm

$$\|A\|_\infty^{(2,2)} = 1 + |k| \quad \text{(Left as an exercise)}$$

Diagonal Matrix P-norms

$$D = \begin{bmatrix} d_1 & & & 0 \\ & d_2 & & \\ & & \ddots & \\ 0 & & & d_m \end{bmatrix} , \qquad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

$$\|D\|_p^{m,m} = \max_{\substack{x \in \mathbb{C}^m \\ \|x\|_p = 1}} \|Dx\|_p^{(m)} = \max_{\substack{x \in \mathbb{C}^m \\ \|x\|_p = 1}} \left[\sum_{i=1}^{m} (d_i x_i)^p \right]^{1/p}$$

$$\leq \max_{1 \leq i \leq m} |d_i| \left[\max_{\substack{x \in \mathbb{C}^m \\ \|x\|_p = 1}} \sum_i \left(\sum |x_i|^p \right)^{1/p} \right]$$

$$\leq \max_{1 \leq i \leq m} |d_i|$$

Note : The equality holds for $x = e_j$ (i.e. $x_{i \neq j} = 0$ & $x_j = 1$), for j such that $|d_j| = \max_{1 \leq i \leq m} |d_i|$

Result 1

The 1-norm of a matrix is its maximum column sum.

Proof : Let $A \in \mathbb{C}^{m \times n}$ and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ s.t. $\|x\|_1 = 1$

$$A = \left[\begin{array}{c|c|c|c} a_1 & a_2 & \cdots & a_n \end{array} \right]$$

$$\|Ax\|_1 = \left\| \sum_{i=1}^{n} x_i a_i \right\|_1$$

$$\leq \sum_{i=1}^{n} \|x_i a_i\|_1 \quad \text{(Triangle inequality)}$$

$$= \sum_{i=1}^{n} |x_i| \, \|a_i\|_1$$

$$\leq \max_{1 \leq i \leq n} \|a_i\|_1 \left(\sum_{i=1}^{n} |x_i| \right) \longrightarrow 1$$

$$= \max_{1 \leq i \leq n} \|a_i\|_1$$

Note : The equality holds for $x = e_j$, where j maximizes $\|a_i\|_1$

Result 2 (Exercise)

The ∞ - norm of a matrix is its maximum row sum

Holder - Inequality

Let p and q satisfy $\frac{1}{p} + \frac{1}{q} = 1$ with $1 \leq p, q < \infty$,

Holder inequality states that for any vectors x any y

$$|x^* y| \leq \|x\|_p \|y\|_q$$

If $p = q = 2$, we get the Cauchy - Schwartz inequality :

$$|x^* y| \leq \|x\|_2 \|y\|_2$$

Examples

Let $u \in \mathbb{C}^m$, $v \in \mathbb{C}^n$, and $A = uv^*$. Find $\|A\|_2^{(m,n)}$

$$\|A\|_2 = \max_{\substack{x \\ \|x\|_2 = 1}} \|Ax\|_2 = \max_{\substack{x \\ \|x\|_2 = 1}} \|uv^*x\|_2$$

$$= \max_{\substack{x \\ \|x\|_2 = 1}} |v^*x| \, \|u\|_2 \leq \max_{\substack{x \\ \|x\|_2 = 1}} \|v\|_2 \underbrace{\|x\|_2}_{=1} \|u\|_2$$

$$= \|u\|_2 \|v\|_2$$

$$\Rightarrow \|Ax\|_2 \leq \|u\|_2 \|v\|_2$$

The above is a tight bound (i.e. $\exists$ x for which the equality holds). How?

Take $x = \dfrac{v}{\|v\|_2}$

Hence $\|A\|^{(m,n)} = \|u\|_2 \|v\|_2$

Induced Norms of Product of Matrices

Let $A \in \mathbb{C}^{\ell \times m}$, $B \in \mathbb{C}^{m \times n}$. And let $\|\cdot\|^{(\ell)}$, $\|\cdot\|^{(m)}$, $\|\cdot\|^{(n)}$ denote the norms on $\mathbb{C}^\ell$, $\mathbb{C}^m$, $\mathbb{C}^n$, respectively.
Then for any $x \in \mathbb{C}^n$,

$$\|ABx\|^{(\ell)} \leq \|A\|^{(\ell,m)} \|Bx\|^{(m)} \leq \|A\|^{(\ell,m)} \|B\|^{(m,n)} \|x\|^{(n)}$$

Therefore, $\|AB\|^{(\ell,n)} \leq \|A\|^{(\ell,m)} \|B\|^{(m,n)}$

In other words, the induced norm of AB is less than or equal to the product of individual induced norms.

Note : for $\|AB\|^{(\ell,n)} \leq \|A\|^{(\ell,m)} \|B\|^{(m,n)}$, the equality is not always guaranteed. Why? ~~Although $\|A^2\|$~~
Take $B = A$, then although $\|A^2\| \leq \|A\|^2$,
$\|A^2\| = \|A\|^2$ does not hold in general.

General Matrix Norms

Matrix norms need not be induced by vector norms. The Matrix norm just needs to satisfy the three vector norm properties when applied in the mn-dimensional vector space of matrices :

a). $\|A\| \geq 0$ and $\|A\| = 0$ if only if $A = 0$
b). $\|A+B\| \leq \|A\| + \|B\|$
c). $\|\alpha A\| = |\alpha| \|A\|$

Is $\|A\| = \det(A)$ a valid matrix norm ?

Frobenius Norm

For $A \in \mathbb{C}^{m,n}$, the Frobenius (aka Hilbert-Schimdt) norm is defined as,

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

$\|A\|_F$ can also be ~~written~~ viewed as the 2-norm of the matrix, if one takes the matrix as an mn-dimensional vector.

If a_j denotes the j^{th} column of A,

$$\|A\|_F = \left(\sum_{j=1}^{n} \|a_j\|_2^2 \right)^{1/2}$$

Alternatively, $\|A\|_F = \sqrt{tr(A^*A)} = \sqrt{tr(AA^*)}$

where $tr(B)$ = sum of diagonal entries of B.

~~What~~ Frobenius norm of Matrix product:

Let $A \in \mathbb{C}^{l \times m}$, $B \in \mathbb{C}^{m \times n}$, and $C = AB$.

$c_{ij} = a_i^* b_j$ (i.e dot product of i^{th} row of A and j^{th} column of B)

$$\|AB\|_F^2 = \|C\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |c_{ij}|^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} |a_i^* b_j|^2$$

$$\leq \sum_{i=1}^{m} \sum_{j=1}^{n} \left(\|a_i\|_2 \|b_j\|_2 \right)^2 \quad \text{(Cauchy-Schwartz inequality)}$$

$$\leq \left(\sum_{i=1}^{m} \|a_i\|_2^2 \right) \left(\sum_{j=1}^{n} \|b_j\|_2^2 \right) = \|A\|_F^2 \|B\|_F^2$$

$$\Rightarrow \boxed{\|AB\|_F^2 \leq \|A_F\|^2 \|B_F\|^2}$$

Invariance under Unitary Multiplication

For any $A \in \mathbb{C}^{m \times n}$ and orthogonal matrix $Q \in \mathbb{C}^{m,n}$,

a) $\|QA\|_2 = \|A\|_2$
b). $\|QA\|_F = \|A\|_F$