Wednesday, March 1, 2017 3:48 PM

BASIC LINEAR ALGEBRA

$A \in \mathbb{R}^{m \times n}$ or $A \in \mathbb{C}^{m \times n}$ → Matrix.

It is a linear map, $A: \mathbb{C}^n \to \mathbb{C}^m$, $x \mapsto Ax$, $\begin{cases} A(\alpha x) = \alpha A x \\ A(x+y) = Ax + Ay \end{cases}$

- $\boxed{b = Ax}$, $b_i = \sum_{j=1}^{n} a_{ij} x_j$ (standard)

Can also be seen as

$$\underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}}_{\text{numbers}} = \underbrace{\left(a_1 \middle| a_2 \middle| \cdots \middle| a_n \right)}_{\text{vectors}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\text{numbers}} = \sum_{i=1}^{n} a_i x_i = b$$

linear combination of the columns of A

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = x_1 \underbrace{\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}}_{a_1} + x_2 \underbrace{\begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}}_{a_2} + \cdots + x_n \underbrace{\begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}}_{a_n}$$

2

- $B = AC$. $b_{ij} = \sum_{k=1}^{n} a_{ik} c_{kj}$ (standard)

$$\left(b_1 \middle| b_2 \middle| \cdots \middle| b_k\right) = A\left(c_1 \middle| c_2 \middle| \cdots \middle| c_k\right)$$

$$b_i = A c_i = \sum_{j=1}^{n} a_j c_{ji}$$

ith column of B is a linear combination of of the columns of A, with coefficients given by the ith column of C.

$$\begin{pmatrix} \boxed{b_{11}} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mk} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1k} \\ c_{21} & c_{22} & \cdots & c_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nk} \end{pmatrix}$$

$$\left(b_1 \middle| \cdots \middle| b_k\right) = \left(A c_1 \middle| \cdots \middle| A c_n\right)$$

Ex.: $u v^T = \begin{pmatrix} u \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \left(u v_1 \middle| u v_2 \middle| \cdots \middle| u v_n\right)$

outer prod.

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_m v_1 & u_m v_2 & \cdots & u_m v_n \end{pmatrix}$$

Ex.: Let R be $r_{ij} = 1$ if $j \geq i$, $r_{ij} = 0$ if $j < i$.

$$R = \begin{pmatrix} 1 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 1 & \cdots & 1 \\ 0 & 0 & 0 & 0 & \cdots & \\ & & \vdots & & & \end{pmatrix}$$

$$Z = AR$$

$$z_j = \sum_{k=1}^{m} a_k r_{kj}$$

$$= \sum_{k=1}^{j} a_k$$ Sum of the first j collumns of A.

- $\mathrm{Range}(A) = \{ y \in \mathbb{C}^m \mid y = Ax \text{ for some } x \in \mathbb{C}^n \}$
- $\mathrm{Null}(A) = \{ x \in \mathbb{C}^n \mid Ax = 0 \}$
- $\mathrm{Rank}(A) =$ # indep. cols of A.
  $=$ # indep. rows of A } Theorem (SVD)

A has <u>full rank</u> if $\mathrm{rank}(A) = \min(m,n)$

<u>Theo.</u> $\mathrm{Range}(A)$ is the space spanned by the cols. of A

- Inverse. $e_1 = (1, 0, 0, \ldots, 0)^T$
  $e_2 = (0, 2, 0, \ldots, 0)^T$
  $\vdots$

If A has full rank the $e_i = \sum_{j=1}^{m} a_j z_{ji}$, or $I = AZ$, so there is Z such that $AZ = I$. Z is the inverse of A, $Z = A^{-1}$.

Equivalent statements:

(1) A has an inverse A^{-1}
(2) A has full rank
(3) $\det A \neq 0$
(4) $\text{range}(A) = \mathbb{C}^m$
(5) $\text{null}(A) = \{0\}$
(6) $\lambda(A) \neq 0$
(7) $\sigma(A) \neq 0$

$\boxed{x = A^{-1}b}$ think about $Ax = b$, i.e.

$A^{-1}b$ is the vector of coefficients of the expansion of b in the basis columns of A.

Let $A \in \mathbb{C}^{m \times n}$. The Hermitian conjugate or adjoint of A is denoted by $A^{\dagger}$ and defined by

$$A^{\dagger} = \begin{pmatrix} a_{11}^{*} & a_{12}^{*} & \cdots & a_{1n}^{*} \\ \vdots & \vdots & & \vdots \\ a_{m1}^{*} & a_{m2}^{*} & \cdots & a_{mn}^{*} \end{pmatrix}^{T}$$ transpose conjugate

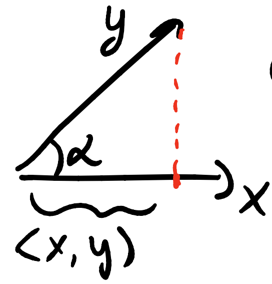
A is Hermitian if $A = A^{\dagger}$. If $A \in \mathbb{R}^{m \times n}$ then A is said to be symmetric, $A = A^{T}$.

Inner product:

Let $x, y \in \mathbb{C}^m$. We define
$$\langle x, y\rangle = x^\dagger y = \sum_{i=1}^{m} x_i^* y_i$$

The norm of x is $\|x\| = \sqrt{\langle x, x\rangle}$.

Geometric interpretation:



$$\cos\alpha = \frac{\langle x, y\rangle}{\|x\|\|y\|}$$

Bilinearity:
- $\langle x_1 + x_2, y\rangle = (x_1 + x_2)^\dagger y = x_1^\dagger y + x_2^\dagger y = \langle x_1, y\rangle + \langle x_2, y\rangle$
- $\langle x, y_1 + y_2\rangle = x^\dagger (y_1 + y_2) = x^\dagger y_1 + x^\dagger y_2 = \langle x, y_1\rangle + \langle x, y_2\rangle$
- $\langle x, \alpha y\rangle = x^\dagger \alpha y = \alpha x^\dagger y = \alpha \langle x, y\rangle$
- $\langle \alpha x, y\rangle = (\alpha x)^\dagger y = \alpha^* x^\dagger y = \alpha^* \langle x, y\rangle$

Orthogonal vectors: $\langle x, y\rangle = 0$

$S = \{x_1, \dots, x_n\}$ is an orthogonal set if $\langle x_i, x_j\rangle = 0$ when $i \neq j$. In addition, S is orthonormal if $\|x_i\| = 1 \;\forall i$, or, $\langle x_i, x_j\rangle = \delta_{ij}$.

Notice that $\|x\|^2 = \langle x, x\rangle = \sum_{i=1}^{m} x_i^* x_i = \sum_{i=1}^{m} |x_i|^2 \geq 0$,

Thus $\|x\| = 0 \iff x = 0$.

<u>Theorem</u>. The vectors in an orthogonal set are linear independent.

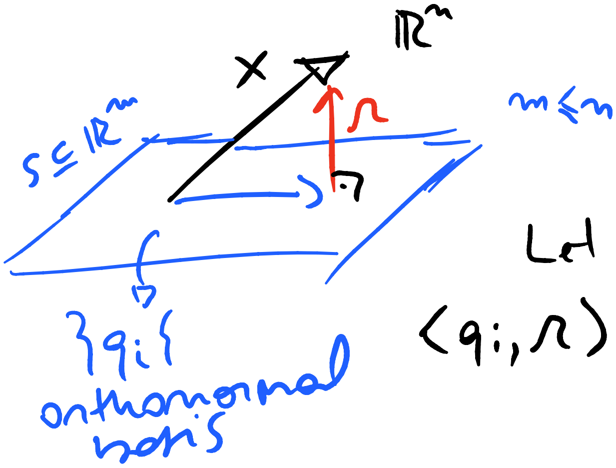
proof. Suppose they are not independent, i.e

$$x_k = \sum_{\substack{j=1 \\ j\neq k}}^{m}{}' \, a_j x_j .$$

Then $\langle x_k, x_k \rangle = \sum_{\substack{j=1 \\ j\neq k}}^{m}{}' \, a_j \langle x_k, x_j \rangle = 0$ since they are orthogonal by assumption. But this implies that $x_k = 0$ which is a contradiction. $\square$

<u>Corollary</u>. If $S \subseteq \mathbb{R}^n$ and S has n orthogonal vectors, then S is a basis of $\mathbb{R}^n$.

Inner products can be used to decompose a vector into orthogonal projections.



$$\bar{x} = \sum_{i=1}^{m} \langle q_i, x \rangle q_i$$

Let $r = x - \bar{x}$. Then,

$$\begin{aligned}
\langle q_i, r \rangle &= \langle q_i, x \rangle - \sum_{j=1}^{m} \langle q_j, x \rangle \underbrace{\langle q_i, q_j \rangle}_{\delta_{ij}} \\
&= \langle q_i, x \rangle - \langle q_i, x \rangle \\
&= 0
\end{aligned}$$

Thus $x = r + \sum_{i=1}^{m} \langle q_i, x \rangle q_i$

$$= r + \sum_{i=1}^{m} (q_i^+ x) q_i = r + \sum_{i=1}^{m} q_i (q_i^+ x)$$

$$= r + \left(\sum_{i=1}^{m} q_i q_i^+ \right) x$$

The operator $P_i = q_i q_i^+$ projects x into the ray of q_i. Notice that $\text{Rank}(P_i) = 1$. The operator $P = \sum_{i=1}^{m} q_i q_i^+$ projects x into $S \subseteq \mathbb{R}^m$ subspace.

$\text{rank}(P) = m$. If $m = n$ then $\{q_i\}$ is a basis for $\mathbb{R}^n$ and $x = \sum_{i=1}^{m} c_i q_i$, thus

$$\langle q_j, x \rangle = \sum_{i=1}^{m} c_i \underbrace{\langle q_j, q_i \rangle}_{\delta_{ij}} = c_j \quad \therefore \quad x = \sum_{i=1}^{m} \langle q_i, x \rangle q_i$$

or $x = \sum_{i=1}^{m} q_i^+ x \, q_i = \left(\sum_{i=1}^{m} q_i q_i^+ \right) x \quad \therefore \quad P = \sum_{i=1}^{m} q_i q_i^+ = I.$

This is known as completeness relation.

Unitary Matrices

$U U^+ = U^+ U = I$ (unitary, $U \in \mathbb{C}^{m \times m}$)

$Q Q^T = Q^T Q = I$ (orthogonal, $Q \in \mathbb{R}^{m \times m}$)

Viewing as before:

$$\begin{pmatrix} u_1^\dagger \\ \hline u_2^\dagger \\ \hline \vdots \\ \hline u_m^\dagger \end{pmatrix} \left(u_1 \middle| u_2 \middle| \cdots \middle| u_m \right) = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{pmatrix}$$

$$u_i^\dagger u_j = \delta_{ij}$$

So the columns of U are orthogonal vectors.

Ux is a linear combination of the columns of U. $U^\dagger b$ is the vectors whose components are the expansion of b in the basis of U.

$$\begin{tikzcd} b \; \{e_i\} \arrow[r, bend left, "U^\dagger"] & x \; \{u_i\} \arrow[l, bend left, "U"] \end{tikzcd}$$

Coeff. of b in $\{e_i\}$ basis

Coeff. of b in the $\{u_i\}$ basis

Now let $\tilde{x} = Ux$, $\tilde{y} = Uy$, thus

$$\langle \tilde{x}, \tilde{y} \rangle = x^\dagger U^\dagger U y = x^\dagger y = \langle x, y \rangle$$

Thus unitary transformations preserve the inner product, or norm of vectors.

vector Norms

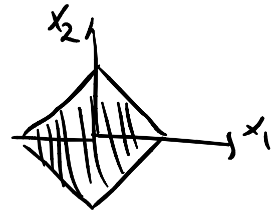
$\|\cdot\| : \mathbb{C}^m \to \mathbb{R}$ such that

1. $\|x\| \geq 0$ and $\|x\| = 0$ iff $x = 0$.
2. $\|x+y\| \leq \|x\| + \|y\|$ (triangle inequality)
3. $\|\alpha x\| = |\alpha| \|x\|$

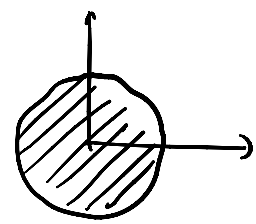
Unit ball: $B = \{ x \in \mathbb{C}^m \mid \|x\| \leq 1 \}$.

p-norms $\|x\|_p$ | B

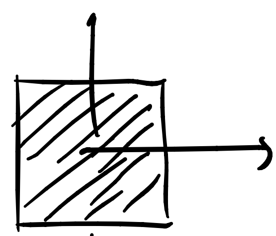
$$\|x\|_1 = \sum_{i=1}^{m} |x_i|$$



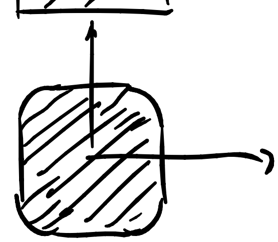
$$\|x\|_2 = \left(\sum_{i=1}^{m} |x_i|^2 \right)^{1/2}$$



$$\|x\|_\infty = \max \{ |x_i| \}$$



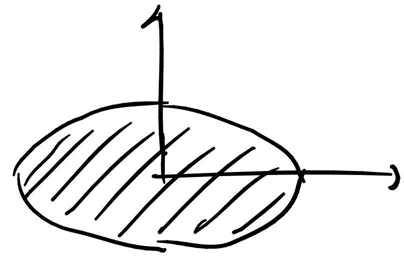
$$\|x\|_p = \left(\sum_{i=1}^{m} |x_i|^p \right)^{1/p}$$



weighted norm: $\|x\|_W = \|Wx\|$ where $W = \mathrm{diag}(w_1, \ldots, w_m)$

For example

$$\|x\|_w = \left(\sum_{i=1}^{m} |w_i x_i|^2 \right)^{1/2}$$



Induced Matrix Norms

Let $A \in \mathbb{C}^{m \times n}$, $\|x\|_{(n)}$ a norm of $x \in \mathbb{C}^n$, and $\|y\|_{(m)}$ a norm of $y \in \mathbb{C}^m$. These two vector norms induces a matrix norm $\|A\|_{(m,n)}$ as follows:

$$\|A\|_{(m,n)} = \sup_x \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \sup_{\|x\|_{(n)}=1} \|Ax\|_{(m)}$$

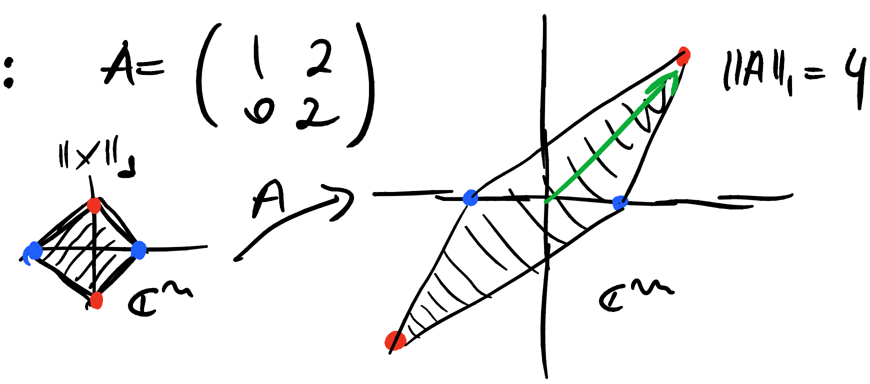
In other words $\|A\|_{(m,n)}$ is the smallest C such that

$$\|Ax\|_{(m)} \leq C \|x\|_{n}$$

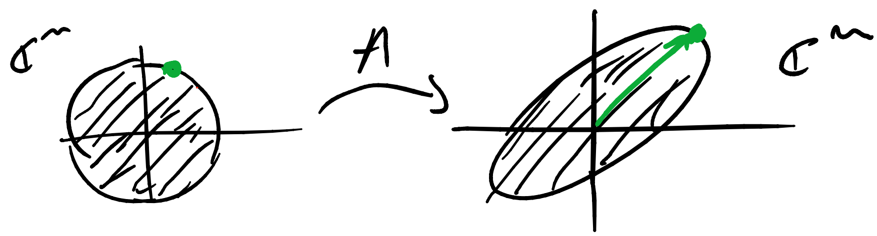
holds for any $x \in \mathbb{C}^n$. This is the maximum factor that A can stretch a vector.

Example in 2D: $A = \begin{pmatrix} 1 & 2 \\ 0 & 2 \end{pmatrix}$

1-Norm $\|A\|_1$

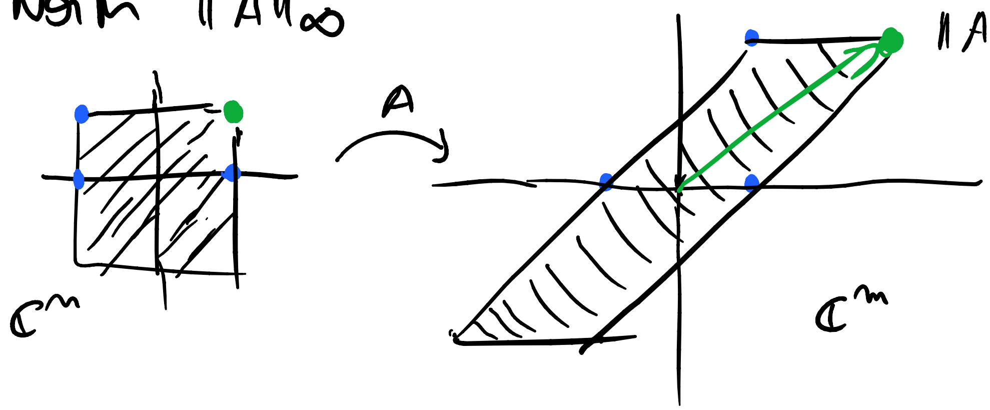


2-Norm $\|A\|_2$



$\|A\|_2 \simeq 2.9$

∞-Norm $\|A\|_\infty$



$\|A\|_\infty = 3$

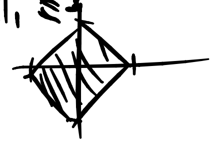
$$\|A\|_2 \leq \|A\|_\infty \leq \|A\|_1$$

<u>Example</u>. $D = \operatorname{diag}(d_1, \dots, d_m)$. The unit circle $\|x\|$ is mapped into an hyperellipse whose semi-axis have size $|d_i|$. Thus $\|D\|_2 = \max |d_i|$. The same holds for $\|D\|_p = \max |d_i|$.

Now consider the 1-norm of $A = \left(a_1 \middle| \cdots \middle| a_m \right)$.
We have
$$\|A\|_1 = \max_{\|x\|=1} \|Ax\|_1$$
$$= \max_{\|x\|=1} \Big\| \sum_{j=1}^{m} x_j a_j \Big\| \leq \max_{\|x\|=1} \sum_{j=1}^{m} |x_j| \, \|a_j\|_1$$

Consider the unit ball, i.e

$\|x\|_1 \leq 1$



$$\max_{\|x\|=1} \sum_{j=1}^{m} |x_j| = \max_{\|x\|=1} \|x\|_1 = 1$$

and replace $\|a_j\|_1 \to \max_j \|a_j\|_1$. Then

$$\|A\|_1 \leq 1 \cdot \max_j \|a_j\|_1$$

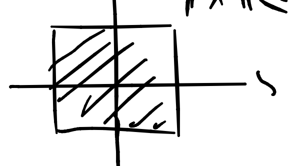
Now choose x such that $Ax = a_j$ where $a_j = \max_k \|a_k\|$. Thus the above bound is attainable and

$$\|A\|_1 = \max_j \|a_j\|_1$$

$\|A\|_1$ is the maximum "column sum" of A.

Consider $\|A\|_\infty$. $\|x\|_\infty = \max |x_i|$ and the unit ball is

$\|x\| \leq 1$



$$\|A\|_\infty = \max_{\|x\|=1} \|Ax\|_\infty = \max_{\|x\|=1} \{|(Ax)_i|\}$$

$$|(Ax)_i| = \left|\sum_{j=1}^{m} a_{ij} x_j\right| \leq \sum_{j=1}^{m} |a_{ij}||x_j| \leq \sum_{j=1}^{m} |a_{ij}| = \|\tilde{a}_i\|_1$$

where $\tilde{a}_i$ is the ith row of A.

where $\tilde{a}_i$ is the i-th row of A.

Thus $\|A\|_\infty \leqslant \max_i \|\tilde{a}_i\|_1$. But it is possible to choose a unit vector x such that the bound is attainable, thus

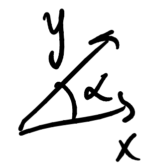
$$\|A\|_\infty = \max_i \|\tilde{a}_i\|$$

is the maximum "row sum" of A.

Cauchy-Schwarz and Hölder

We have Cauchy-Schwarz inequality

$$|x^t y| \leqslant \|x\|_2 \|y\|_2 .$$



$\langle x, y\rangle = \|x\| \|y\| \cos \alpha$

This can be generalized to

$$|x^t y| \leqslant \|x\|_p \|x\|_q$$

provided $\frac{1}{p} + \frac{1}{q} = 1$ for $p, q \geqslant 1$. This is Hölder's inequality.

Ex.: Outer product $A = uv^t = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix} (v_1^* \dots v_n^*)$

$$\|A\|_2 = \max \frac{\|Ax\|_2}{\|x\|_2} = \max \frac{\|uv^t x\|_2}{\|x\|_2} = \max \frac{\|u\|_2 |v^t x|}{\|x\|_2}$$

but $|v^t x| \leqslant \|v\|_2 \|x\|_2 \therefore \|A\|_2 \leqslant \max \|u\|_2 \|v\|_2$.

This is attainable, just make $x = v$. Therefore,

$\ldots | \leq \|v\|_2 \|x\|_2 \;\therefore\; \|A\|_2 \leqslant \max \|u\|_2 \|v\|_2.$

This is attainable, just make $x = v$. Therefore,

$$\|uv^*\| = \|u\|_2 \|v\|_2.$$

Notice that $\|Ax\| \leqslant C\|x\|$ where $C = \|A\|$. (14)

This is the maximum stretch. Thus $\|Ax\| \leqslant \|A\| \|x\|$ for any x. Now let $A \in \mathbb{C}^{\ell \times m}$, $B \in \mathbb{C}^{m \times n}$, $x \in \mathbb{C}^n$:

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

$$\|AB\| = \max_x \frac{\|ABx\|}{\|x\|} \;\therefore\; \|AB\| \leqslant \|A\| \|B\|$$

In general the equality is not attainable.

General Matrix Norms

Matrices $A \in \mathbb{C}^{m \times n}$ form a vector space, thus one can define an arbitrary norm $\|A\|$, which does not need to be induced by a vector norm, provided

1. $\|A\| \geqslant 0$, $\|A\| = 0 \Leftrightarrow A = 0$
2. $\|A + B\| \leqslant \|A\| + \|B\|$
3. $\|\alpha A\| = |\alpha| \|A\|$

hold. An important example is the Frobenius norm:

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

$$= \left(\sum_{i=1}^{m} \|a_i\|_2^2 \right)^{1/2}$$

Notice also that $(A^+A)_{kl} = \sum_{j=1}^{m} a_{jk}^* a_{jl}$.

Thus $(A^+A)_{kk} = \sum_{j=1}^{m} a_{jk}^* a_{jk} = \sum_{j=1}^{m} |a_{jk}|^2$. Summing over k given (15)

$$\|A\|_F = \mathrm{Tr}(A^+A) = \mathrm{Tr}(AA^+)$$

Let $C = AB$. $c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} = a_i^+ b_j$. ($b_j$ column, a_i^+ row)

$$\|AB\|_F^2 = \sum_{i=1}^{m} \sum_{j=1}^{l} |c_{ij}|^2 = \sum_{i=1}^{m} \sum_{j=1}^{l} |a_i^+ b_j|^2$$

$$\leq \sum_{i=1}^{m} \sum_{j=1}^{l} \|a_i^+\|_2^2 \|b_j\|_2^2$$

$$= \sum_{i=1}^{m} \|a_i^+\|_2^2 \sum_{j=1}^{l} \|b_j\|_2^2$$

$$= \|A\|_F^2 \|B\|_F^2$$

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

Theorem. Let U be unitary, $UU^+ = I = U^+U$. Then $\|UA\|_2 = \|A\|_2$ and $\|UA\|_F = \|A_F\|$.

proof. $\|UA\|_2^2 = \max_{\|x\|=1} \|UAx\|_2^2 = \max_{\|x\|=1} (UAx)^+(UAx)$

proof. $\|UA\|_2^2 = \max_{\|x\|=1} \|UAx\|_2^2 = \max_{\|x\|=1} (UAx)^\dagger (UAx)$

$= \max_{\|x\|=1} x^\dagger A^\dagger U^\dagger U A x = \max_{\|x\|=1} \|Ax\|_2^2 = \|A\|_2^2$

$\|UA\|_F = \mathrm{Tr}\left((UA)^\dagger UA\right) = \mathrm{Tr}\left(A^\dagger U^\dagger U A\right) = \mathrm{Tr}(A^\dagger A) = \|A\|_F$ $\square$