# Basic Linear Algebra and Matrix Operations

## Notations

Scalars (numbers): normal font lower case letters

e.g. $a, b, x, y$ ...

$\mathbb{R}$ —— the set of all real numbers

$\mathbb{C}$ —— the set of all complex numbers

$a \in \mathbb{R}$ —— $a$ is a real number

$b \in \mathbb{C}$ —— $b$ is a complex number

Unless specified, a scalar is assumed real by default.

Vectors: bold lower case letters in printing.

(standard lower case letters in hand-writing)

e.g.,

$$a = [a_i]_{n\times 1} = \begin{bmatrix} a_1 \\ \vdots \\ a_i \\ \vdots \\ a_n \end{bmatrix}$$

— the i-th entry of $a$.

or dropped for simplicity

, where $a_i$ are scalars.

Unless specified, we refer vectors as column vectors.

The transpose of vector $a$ is denoted by $a^T$ and given by

$$a^T = [a_1 \cdots a_i \cdots a_n]$$

Obviously, $a^T$ is a length-n row vector.

Vector spaces: $\mathbb{R}^n$, $\mathbb{C}^n$

$a \in \mathbb{R}^n$ —— $a$ is a length-n column real vector.

$a \in \mathbb{C}^n$ —— $a$ is a length-n column complex vector.

"Real" spaces will be assumed if not specified

Special vectors:

length-n basis vector: $e_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$ ← the i-th component.

length-n zero vector: $0 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

length-n vector of all ones.

$$\mathbb{1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \sum_{i=1}^{n} e_i$$

Matrices: bold upper case letters in printing
standard upper case letters in hand-writing

An $n \times m$ matrix $A \in \mathbb{R}^{n \times m}$ (or $A \in \mathbb{C}^{n \times m}$)

$$A = [a_{ij}]_{n \times m} \text{ or } A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

called element or ntry

or dropped for simplicity

where $a_{ij} \in \mathbb{R}$ (or $\mathbb{C}$) for $i = 1, \cdots, n$, $j = 1, \cdots, m$.

Special matrices:

Identity matrix $I$

$$I = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & 1 \end{bmatrix}_{n\times n}$$

Matrix spaces: $\mathbb{R}^{m\times n}$, $\mathbb{C}^{m\times n}$

$A \in \mathbb{R}^{m\times n}$ —— $A$ is an $m\times n$ real matrix.

$A \in \mathbb{R}^{m\times n}$ —— $A$ is an $m\times n$ complex matrix.

Since a vector $a \in \mathbb{R}^n$ can also be viewed as an $n\times 1$ matrix, i.e., $a \in \mathbb{R}^{n\times 1}$,

$$\mathbb{R}^{n\times 1} = \mathbb{R}^n$$

and
$$\mathbb{C}^{n\times 1} = \mathbb{C}^n$$

Similarly, the space of all length-$n$ row vectors is

$$\mathbb{R}^{1\times n} \quad \text{or} \quad \mathbb{C}^{1\times n}.$$

Any matrix $A \in \mathbb{R}^{n\times m}$ (or $\mathbb{C}^{n\times m}$) can be written in a column vector form as

$$A = [a_{(1)}, a_{(2)}, \cdots a_{(m)}],$$

where $a_{(j)} \in \mathbb{R}^n$ (or $\mathbb{C}^n$), $j = 1, \cdots, m$, are columns of $A$.

Similarly, we can also write $A \in \mathbb{R}^{n\times m}$ (or $\mathbb{C}^{n\times m}$) in a row column vector form as

$$A = \begin{bmatrix} (a^{(1)})^T \\ (a^{(2)})^T \\ \vdots \\ (a^{(n)})^T \end{bmatrix}$$

where $a^{(i)} \in \mathbb{R}^m$ (or $\mathbb{C}^m$), $i = 1, \cdots, n$, are rows of $A$.

Matrix transpose:

Given $A = [a_{ij}]_{n\times m}$, its transpose $A^T$ is defined by

$$A^T = [a_{ji}]_{m\times n}$$

Example:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \Longleftrightarrow A^T = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \end{bmatrix}$$

Obviously $(A^T)^T = A$

Hemitian (Also called Adjoint and conjugate transpose)

$$A = [a_{ij}]_{n\times m} \Longleftrightarrow A^* = [\overline{a_{ji}}]_{m\times n}$$

Example:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \Longleftrightarrow A^* = \begin{bmatrix} \overline{a_{11}} & \overline{a_{21}} & \overline{a_{31}} \\ \overline{a_{12}} & \overline{a_{22}} & \overline{a_{32}} \end{bmatrix}$$

If $A \in \mathbb{R}^{n\times m}$, then $A^* = A^T$

If $A \in \mathbb{C}^{n\times m}$, then $A^* = \overline{A^T}$

Symmetry and Hermitian

$A \in \mathbb{R}^{n\times n}$ is symmetric if $A = A^T$

$A \in \mathbb{C}^{n\times n}$ is Hermitian if $A = A^*$

Example: $A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$ is symmetric

$A = \begin{bmatrix} 1 & i \\ -i & -1 \end{bmatrix}$ is Hermitian.

## Some Basic Operations of Matrices and Vectors

### ✩ Addition, Subtraction

Given two matrices $A=[a_{ij}]_{n\times m}$ and $B=[b_{ij}]_{n\times m}$, their summation/difference, denoted by $C$, is in $\mathbb{R}^{n\times m}$ and given by

$$C = A \pm B = \begin{bmatrix} a_{11}\pm b_{11} & a_{12}\pm b_{12} & \cdots & a_{1m}\pm b_{1m} \\ a_{21}\pm b_{21} & a_{22}\pm b_{22} & \cdots & a_{2m}\pm b_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\pm b_{n1} & a_{n2}\pm b_{n2} & \cdots & a_{nm}\pm b_{nm} \end{bmatrix}$$

### ✩ Scalar product

Let $c \in \mathbb{R}$ be a scalar and $A \in \mathbb{R}^{n\times m}$ be a matrix.
The scalar product of $c$ and $A$, denoted by $B$, is in $\mathbb{R}^{n\times m}$ and given by

$$B = cA = \begin{bmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{bmatrix}$$

pseudo code:

```
for i=1:n
    for j=1:m
        b_ij = c a_ij
    end
end
```

ij-loop

$B$ is obtained row-by-row

```
for j=1:m
    for i=1:n
        b_ij = c a_ij
    end
end
```

ji-loop

$B$ is obtained column-by-column

The two implementations may have different performances, depending on how the matrices are stored. If the matrices is stored row-by-row, then ij-loop may be better, because it can save communication cost of Cache and CPU.

## Inner product:

- Given two vectors $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$ and $b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$ in $\mathbb{R}^n$

  their inner product is

$$\langle a, b \rangle = \sum_{i=1}^{n} a_i b_i$$

- pseudo code:

```
inner_prod = 0;
for i = 1:n
    inner_prod = inner_prod + a_i b_i ;
end
```

- If $\langle a, b \rangle = 0$, then we say $a, b$ are perpendicular or othogonal.

## Matrix-Vector product:

- Let $A = [a_{ij}]_{m \times n} \in \mathbb{R}^{m \times n}$ and $x = [x_i]_{n \times 1} \in \mathbb{R}^n$.

  Then the matrix-vector product of $A$ and $x$, denoted $b$, is a vector in $\mathbb{R}^m$, given by

$$b = Ax = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}$$

- It is easy to see that

$$\langle a, b \rangle = [a_1\ a_2 \cdots a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = [b_1\ b_2 \cdots b_n] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

$$= a^T b \qquad\qquad = b^T a$$

  i.e., inner product is a special case of matrix-vector product.

- If we write $A$ in row vector form

$$(a^{(1)})^T$$
$$(a^{(2)})^T$$

$(a^{(m)})^T$

then

$$b = Ax = \begin{bmatrix} (a^{(1)})^T \\ (a^{(2)})^T \\ \vdots \\ (a^{(m)})^T \end{bmatrix} x = \begin{bmatrix} (a^{(1)})^T x \\ (a^{(2)})^T x \\ \vdots \\ (a^{(m)})^T x \end{bmatrix} = \begin{bmatrix} \langle a^{(1)}, x \rangle \\ \langle a^{(2)}, x \rangle \\ \vdots \\ \langle a^{(m)}, x \rangle \end{bmatrix}$$

i.e, entries of $b$ are inner products of row vectors $a^{(i)}$ and $x$.

This gives an implementation of $b = Ax$.

pseudo code 1:

ij-loop

```
for i = 1:m
    b_i = 0;
    for j = 1:n
        b_i = b_i + a_ij x_j;
    end
end
```

inner product $b_i = \langle a^{(i)}, x \rangle$

- If we write $A$ in column vector form

$A = [a_{(1)}, a_{(2)}, \cdots, a_{(n)}]$, where $a_{(j)} \in \mathbb{R}^m, j = 1, \cdots, n$ are columns of $A$,

then

$$b = Ax = [a_{(1)} \; a_{(2)} \; \cdots \; a_{(n)}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 a_{(1)} + x_2 a_{(2)} + \cdots \; x_n a_{(n)}$$

This reveals that $b$ is a linear combination of columns $a_{(j)}$ of $A$.

This gives another implementation of $b = Ax$:

pseudo code 2:

ji-loop

```
set vector b = 0
for j = 1:n
    for i = 1:m
        b_i = b_i + x_j a_ij
    end
end
```

add $b$ by $x_j a_{(j)}$.

- Comparison of two implementations.

| ij - loop | ji - loop |
|---|---|
| Outer loop on rows<br>Inner loop on columns | Outer loop on columns<br>Inner loop on rows |

Their efficiency depends on how the matrix $A$ stored.

If $A$ is stored column-wise, ji-loop is better

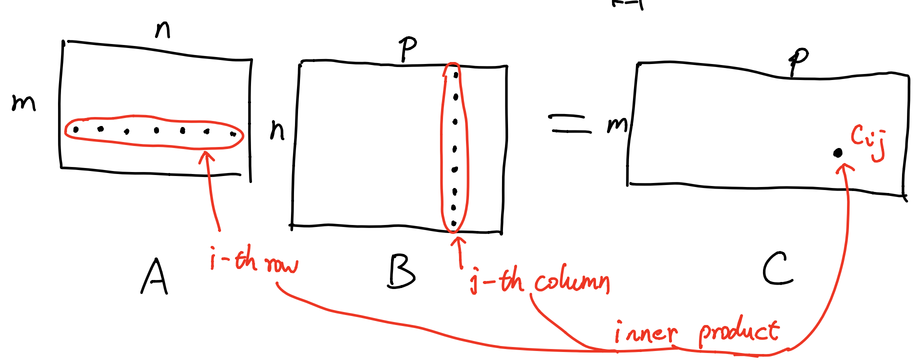
row-wise, ij-loop

## ☆ Matrix - Matrix product

- Let $A=[a_{ij}] \in \mathbb{R}^{m\times n}$ and $B=[b_{ij}] \in \mathbb{R}^{n\times p}$.

  Then the product of $A$ and $B$, denoted by $C$, is an $m\times q$ matrix, given by

$$C = [C_{ij}] \in \mathbb{R}^{m\times p}, \quad \text{where} \quad C_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj}$$




- Obviously, inner product and matrix-vector product are special cases of matrix-matrix product.
- We have different implementations.

  ① By the definition above.

```
for i = 1:m
    for j = 1:p
        Cij = 0
```

```
for j = 1:p
    for i = 1:m
        Cij = 0
```

```
for k=1:n
    c_ij = c_ij + a_ik b_kj
  end
 end
end
```

$ijk$ – loop

The product is computed row-by-row

```
for k=1:n
    c_ij = c_ij + a_ik b_kj
  end
 end
end
```

$jik$ – loop

The product is computed column-by-column

② Write $A$ in column form and $B$ in row form, i.e.,

$$A = [a_{(1)}\ a_{(2)}\ \cdots\ a_{(n)}], \qquad B = \begin{bmatrix} (b^{(1)})^T \\ (b^{(2)})^T \\ \vdots \\ (b^{(n)})^T \end{bmatrix}$$

Then,

$$C = AB = [a_{(1)}\ a_{(2)}\ \cdots\ a_{(n)}] \begin{bmatrix} (b^{(1)})^T \\ (b^{(2)})^T \\ \vdots \\ (b^{(n)})^T \end{bmatrix}$$

$$= \sum_{k=1}^{n} a_{(k)} (b^{(k)})^T.$$

We get another implementation

```
Initialize matrix C=0
for k=1:n
  for i=1:m
    for j=1:p
      c_ij = c_ij + a_ik b_kj
    end
  end
end
```

Add $C$ by $a_{(k)}(b^{(k)})^T$

$kij$ – loop.

Any permutation of the sequence $(i,j,k)$ is a possible implementation. So we have $3! = 6$ different implementations of $C=AB$.

☆ Computational complexity analysis

- How many scalar operations are involved in the computation of inner product of $a, b \in \mathbb{R}^n$ ?

  for an $i$, one scalar multiplication and one scalar addition

  $$\sum_{i=1}^{n} (1+1) = 2n$$

  We say the computational complexity is $O(n)$, meaning a constant multiple of $n$.

- Matrix-vector product of $A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$

  $$\sum_{i=1}^{n} \sum_{j=1}^{m} (1+1)$$

  for an $i, j$, one multiplication, one addition

  $$= 2nm \equiv O(nm)$$

- Matrix-Matrix product of $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times p}$

  $$\sum_{i=1}^{m} \sum_{j=1}^{p} \sum_{k=1}^{n} (1+1) = 2nmp \equiv O(nmp)$$

☆ All these basic operations are implemented efficiently in BLAS (Basic Linear Algebra Subroutine) library, which comes with CPU. or cuBLAS, which comes with NVIDIA GPU.

The real performance depends on many factors, e.g., computational complexity, CPU Cache, ...

## More concepts of matrices

Given a matrix $A = [a_{(1)}\ a_{(2)} \cdots a_{(n)}] \in \mathbb{R}^{m\times n}$ with $a_{(j)} \in \mathbb{R}^m$ columns of $A$,

- the range of $A$ is

$$Ran(A) = \{Ax \mid x \in \mathbb{R}^n\}$$
$$= \{x_1 a_{(1)} + x_2 a_{(2)} + \cdots + x_n a_{(n)} \mid x \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

Ran(A) is also called column space of A, because all vectors in Ran(A) is a linear combination of column vectors of A.

- The null space of A is
$$Null(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$$
- The rank of A is
$$Rank(A) = dim(Ran(A))$$
We must have $Rank(A) \le \min\{m, n\}$.

## Vector and Matrix Norms

To measure how accurate our algorithm is, we need to define metric, i.e., distance between two scalars, two vectors, or two matrices.

☆ Metric for scalars

For two scalars $x, y \in \mathbb{R}$, their distance is $|x-y|$.

☆ Metric for vectors

For two vectors $a, b \in \mathbb{R}^n$, we need to generalize the absolute function for scalars to vectors.

- For a scalar, the absolution function satisfies:

① $|x| \ge 0, \ \forall x \in \mathbb{R}$ and $|x| = 0 \Longleftrightarrow x = 0$.

② $|xy| = |x||y| \quad \forall x, y \in \mathbb{R}$

③ $|x+y| \le |x| + |y| \quad \forall x, y \in \mathbb{R}$

- Norm —— generalization of absolute value to vectors.

Definition: A function $\mathbb{R}^n \to \mathbb{R}$, denoted by $\|x\|$ for any $x \in \mathbb{R}^n$, is called a norm, if it satisfies:

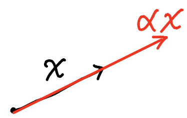
① $\|x\| \geq 0 \quad \forall\, x \in \mathbb{R}^n$ and $\|x\| = 0 \Leftrightarrow x = 0$.

② $\|\alpha x\| = |\alpha|\, \|x\| \quad \forall\, \alpha \in \mathbb{R},\ x \in \mathbb{R}^n$

③ $\|x+y\| \leq \|x\| + \|y\| \quad \forall\, x, y \in \mathbb{R}^n$

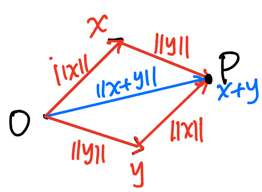
Remarks:

- Condition ① is to make a norm a magnitude.
- Condition ② says that the norm of a scaling of vectors is the scaling of the norm.



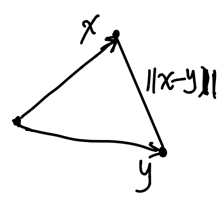

$\|\alpha x\| = |\alpha|\, \|x\|$

- Condition ③ is also known as the triangle inequality.



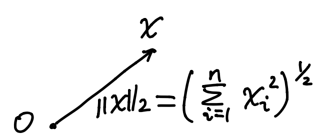

The length of indirect path (the red) from O to P is longer than that of the direct path (the blue)

- Given two vectors $x, y \in \mathbb{R}^n$, their distance is $\|x-y\|$




Example 1: Euclidean norm (2-norm)

$$\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \quad \forall\, x \in \mathbb{R}^n$$




$\|\cdot\|_2$ is indeed a norm, because

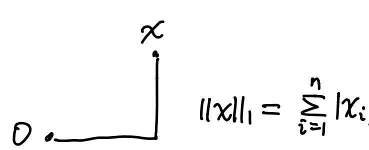
① $\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2} \geq 0$ and $\|x\|_2 = 0 \Leftrightarrow \sum_{i=1}^{n} x_i^2 = 0 \Leftrightarrow x_i = 0\ \forall i \Leftrightarrow x = 0$.

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

③ Need to use Cauchy-Schwartz to prove $\|x+y\|_2 \leq \|x\|_2 + \|y\|_2$.
(We omit it)

Example 2: Manhattan norm (1-norm)

$$\|x\|_1 = \sum_{i=1}^{n} |x_i| \quad \forall\, x \in \mathbb{R}^n$$



$\|\cdot\|_1$ is indeed a norm, because

① $\|x\|_1 = \sum_{i=1}^{n} |x_i| \geq 0$ and $\|x\|_1 = 0 \Leftrightarrow \sum_{i=1}^{n} |x_i| = 0 \Leftrightarrow |x_i| = 0 \ \forall i \Leftrightarrow x = 0.$

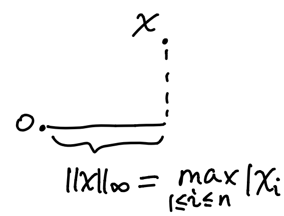
② $\|\alpha x\|_1 = \sum_{i=1}^{n} |\alpha x_i| = \sum_{i=1}^{n} |\alpha| |x_i| = |\alpha| \sum_{i=1}^{n} |x_i| = |\alpha| \, \|x\|_1 \qquad \forall \alpha \in \mathbb{R}, x \in \mathbb{R}^n$

③ $\|x+y\|_1 = \sum_{i=1}^{n} |x_i + y_i| \leq \sum_{i=1}^{n} (|x_i| + |y_i|) = \sum_{i=1}^{n} |x_i| + \sum_{i=1}^{n} |y_i| = \|x\|_1 + \|y\|_1$

$\forall x, y \in \mathbb{R}^n$

Example 3: Max norm ($\infty$-norm)

$$\|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \qquad \forall x \in \mathbb{R}^n$$




Question: Check $\|\cdot\|_\infty$ is a norm.

Example 4: $p$-norm ($p \geq 1$)

Let $p$ be a positive integer. Define

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

Then it can check $\|x\|_p$ is a norm for any $p \geq 1$.

$p$-norm indeed defines a metric for vectors.

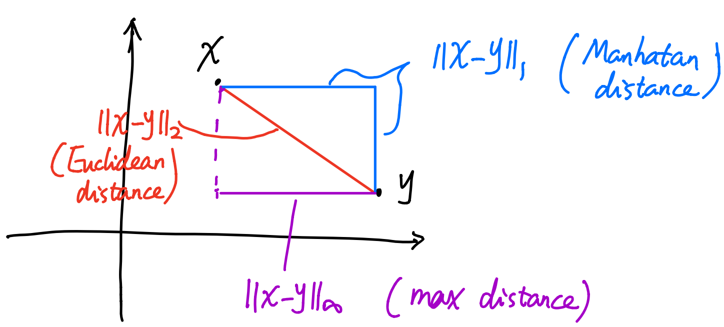
$p=1$, $\quad \|x\|_1 = \sum_{i=1}^{n} |x_i| \qquad$ (Manhatan norm)

$p=2$, $\quad \|x\|_2 = \left( \sum_{i=1}^{n} |x_i|^2 \right)^{1/2} \qquad$ (Euclidean norm)

$p \to \infty$, $\quad \lim_{p \to \infty} \|x\|_p = \max_{1 \leq i \leq n} |x_i| \qquad$ (Max norm)

Can you prove it? → P.T.O

(For this reason, we denoted $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$)




## ☆ Metric for Matrices

Similar to vector norms, we can define matrix norm.

$p \to \infty$ of the $L^p$ norm:

**Proof.** Let $\|x\|_\infty = \max_i |x_i|$ and $\|x\|_p = \left(\sum_i |x_i|^p\right)^{\frac{1}{p}}$. We wish to show that $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty$.

We have that

$$
\begin{aligned}
\|x\|_p &= \|x\|_\infty \frac{\left(\sum_i |x_i|^p\right)^{\frac{1}{p}}}{\|x\|_\infty} \\
&= \|x\|_\infty \left(\sum_i \frac{|x_i|^p}{\|x\|_\infty^p}\right)^{\frac{1}{p}} \\
&= \|x\|_\infty \left(\sum_i \left(\frac{|x_i|}{\|x\|_\infty}\right)^p\right)^{\frac{1}{p}} \\
&\le \|x\|_\infty n^{\frac{1}{p}}
\end{aligned}
$$

$\max_i |x_i| \geq |x_i|$

We arrive at the last item because $\left(\frac{|x_i|}{\|x\|_\infty}\right)^p \le 1$ for every $i$ (because $\|x\|_\infty \ge |x_i|$ for each $i$). Thus we have

$$\|x\|_\infty \le \|x\|_p \le \|x\|_\infty n^{\frac{1}{p}}$$

so, taking a limit as $p \to \infty$, we have

$$\|x\|_\infty \le \lim_{p\to\infty} \|x\|_p \le \|x\|_\infty \cdot \lim_{p\to\infty} n^{\frac{1}{p}} = \|x\|_\infty$$

In other words, we have $\lim_{p\to\infty} \|x\|_p$ sandwiched between $\|x\|_\infty$ and $\|x\|_\infty$, implying equality. Therefore $\lim_{p\to\infty} \|x\|_p = \|x\|_\infty = \max_i |x_i|$.

---

Consider also this proof:

**Definition:** A function $\mathbb{R}^{m\times n} \to \mathbb{R}$, denoted by $\|A\|$ for any $A \in \mathbb{R}^{m\times n}$, is called a norm, if it satisfies:

① $\|A\| \geq 0 \quad \forall A \in \mathbb{R}^{m\times n}$ and $\|A\| = 0 \Leftrightarrow A = 0$.

② $\|\alpha A\| = |\alpha|\,\|A\| \quad \forall \alpha \in \mathbb{R},\ A \in \mathbb{R}^{m\times n}$

③ $\|A+B\| \leq \|A\| + \|B\| \quad \forall A, B \in \mathbb{R}^{m\times n}$

- Since $m\times n$ matrices can be viewed as long vectors of length $mn$, i.e.,

$$\mathbb{R}^{m\times n} \longleftrightarrow \mathbb{R}^{mn}$$

we can apply norms for vectors to obtain norms for matrices.

Example 1: Frobenius norm (vector 2-norm)

$$\|A\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2\right)^{1/2} \qquad \forall A \in \mathbb{R}^{m\times n}$$

Example 2: (vector p-norm)

$$\|A\|_{vec,p} = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^p\right)^{1/p} \qquad \forall A \in \mathbb{R}^{m\times n}$$

Example 3: Consider the identity matrix

$$I = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix} \in \mathbb{R}^{n\times n}$$

unitary 包括 旋轉 / 反射, 不改變數值

Then $\|I\|_F = \sqrt{n}$ and $\|I\|_p = n^{\frac{1}{p}}$

Example 4: Consider a unitary matrix $U \in \mathbb{R}^{n\times n}$ (i.e., $UU^T = U^TU = I$)

$$\begin{bmatrix} U_{11} & U_{12} & U_{13} \\ U_{21} & U_{22} & U_{23} \\ U_{31} & U_{32} & U_{33} \end{bmatrix}\begin{bmatrix} U_{11} & U_{21} & U_{31} \\ U_{12} & U_{22} & U_{32} \\ U_{13} & U_{23} & U_{33} \end{bmatrix}$$

$$= U_{11}^2 + U_{12}^2 + U_{13}^2 = 1$$

Then $\|U\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij}^2\right)^{1/2} = \left(\sum_{i=1}^{n} 1\right)^{1/2} = \sqrt{n}$.

each row of $U$ are unit vectors in 2-norm

From Example 3 and 4 we see that: The Frobenius norm for identity and unitary matrices grows with respect to $n$. However, they are "unit" linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$, and their magnitude should not grow w.r.t. $n$. Therefore, these matrix norms generalized from vector norms are not suitable for linear transformations.

- We view matrices as linear transformations.

$$A: \mathbb{R}^{m\times n} \quad : \quad \underset{\substack{\uparrow \\ p\text{-norm}}}{\mathbb{R}^n} \to \underset{\substack{\uparrow \\ p\text{-norm}}}{\mathbb{R}^m}$$

So, we can define matrix operator p-norm (or simply matrix p-norm) as

$$\|A\|_p = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_p}{\|x\|_p}$$

→ 例如，如果 $(Ax)\in\mathbb{R}^m$ 係將一條好細的 vector $\vec{x}\in\mathbb{R}^n$ map 去一條好大的 $(Ax)\in\mathbb{R}^m$，咁呢個 ratio 就反映咗 A 呢個 linear transformation 的 amplitude 有幾勁.

In other words, matrix p-norm of $A$ is the largest magnifying factor of $A$ acting from $\mathbb{R}^n$ with p-norm to $\mathbb{R}^m$ with p-norm.

而要 take sup 的原因是：$x$ 是虛構的，我並不關心某一特定的 $x$ 於 $A$ 之影響，只是取其最大衡量 $A$ 的 amplitude.

Sup 的定義基本嘅

**Theorem:** $\|A\|_p = \max_{\|x\|_p=1} \|Ax\|_p$

proof. By definition,

$$\|A\|_p = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\substack{x\in\mathbb{R}^n \\ x\neq 0}} \left\|A\left(\frac{x}{\|x\|_p}\right)\right\|_p = \sup_{\|x\|_p\neq 0} \left\|A\frac{x}{\|x\|_p}\right\|_p$$

$$= \sup_{\left\|\frac{x}{\|x\|_p}\right\|_p=1} \left\|A\frac{x}{\|x\|_p}\right\|_p \underset{\substack{\uparrow \\ \text{Let } \tilde{x}=\frac{x}{\|x\|_p}}}{=} \sup_{\|\tilde{x}\|_p=1} \|A\tilde{x}\|_p \underset{\substack{\uparrow \\ \text{use } x \\ \text{to replace } \tilde{x}}}{=} \sup_{\|x\|_p=1} \|Ax\|_p$$

It remains to prove $\sup_{\|x\|_p=1} \|Ax\|_p = \max_{\|x\|_p=1} \|Ax\|_p$. To this end, we define $S = \{x\in\mathbb{R}^n \mid \|x\|_p = 1\}$.

- The p-norm $\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$, viewed as a function of $x$, is continuous, because all $|\cdot|^p$, $+$, $(\cdot)^{1/p}$ functions are continuous.

  Let $\{x^{(k)}\}_{k\in\mathbb{N}} \subset S$ and $\lim_{k\to\infty} x^{(k)} = x\in\mathbb{R}^n$. Then,

  $$1 = \lim_{k\to+\infty} \|x^{(k)}\|_p = \|\lim_{k\to+\infty} x^{(k)}\|_p = \|x\|_p \implies x\in S.$$

  So, $S$ is closed.
- For any $x\in S$, we have $|x_i|\leq 1 \ \forall i$. Thus, $S$ is bounded.
- Moreover, $\|Ax\|_p = \left(\sum_{i=1}^m \left|\sum_{j=1}^n a_{ij}x_j\right|^p\right)^{1/p}$ is continuous in $x$, because it is a composition of functions such as multiplication, addition, p-th power, p-th root, absolute value, which are all continuous.

Altogether, by using Weierstrass extreme value theorem (a continuous

function on a bouned and closed set obtains its exetreme values),
$\|Ax\|_p$ can attain its supremum on $S$, i.e.,

$$\sup_{x\in S}\|Ax\|_p = \max_{x\in S}\|Ax\|_p$$

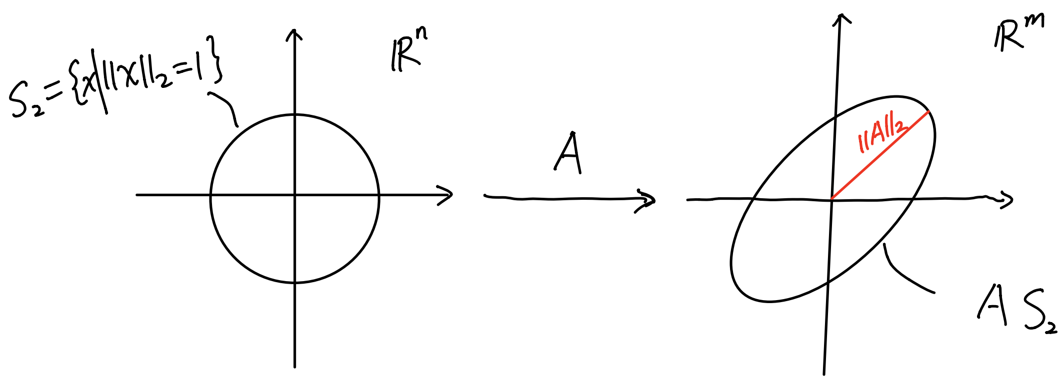
Example 1: 2-norm

$$\|A\|_2 = \max_{\|x\|_2=1}\|Ax\|_2 \iff \|A\|_2^2 = \max_{\|x\|_2=1}\|Ax\|_2^2 \overset{\text{because } \|y\|_2^2 = y^Ty \ \forall y.}{=} \max_{\|x\|_2=1} x^TA^TAx$$

$$\underset{\text{will be proved later in eigenvalue decomposition}}{=} \text{maximum eigenvalue of } A^TA$$

Therefore, $\|A\|_2 = (\text{max eigenvalue of } A^TA)^{1/2}$

Geometry:

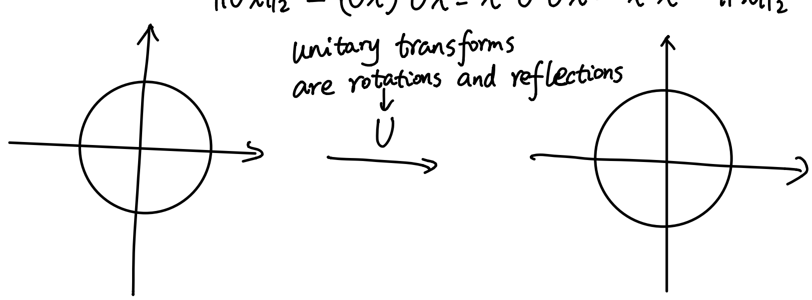



So,

$$\|I\|_2 = \max_{\|x\|_2=1}\|Ix\|_2 = \max_{\|x\|_2=1}\|x\|_2 = \max_{\|x\|_2=1} 1 = 1$$

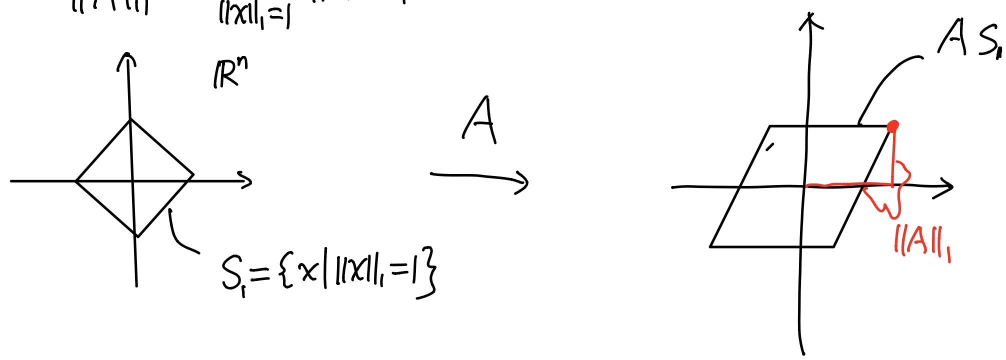
Also, for unitary $U\in\mathbb{R}^{n\times n}$ (i.e, $UU^T=U^TU=I$)

$$\|U\|_2 = \max_{\|x\|_2=1}\|Ux\|_2 \underset{\|Ux\|_2^2 = (Ux)^TUx = x^TU^TUx = x^Tx = \|x\|_2^2}{=} \max_{\|x\|_2=1}\|x\|_2 = 1$$



Example 2: 1-norm

$$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1$$




Theorem: $\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{1\le j\le n} \|a_{(j)}\|_1$,

where $A = [a_{(1)}\ a_{(2)}\ \cdots\ a_{(n)}] \in \mathbb{R}^{m\times n}$

$\boxed{\|A\|_1 \text{ is the max 1-norm of columns.}}$

proof. we prove the theorem by showing

$$\max_{\|x\|_1=1} \|Ax\|_1 \le \max_{1\le j\le n} \|a_{(j)}\|_1 \quad \text{and} \quad \max_{\|x\|_1=1} \|Ax\|_1 \ge \max_{1\le j\le n} \|a_{(j)}\|_1$$

① For "$\le$":

For an arbitrary $x$ satisfying $\|x\|_1 = 1$,

$$\|Ax\|_1 = \left\| [a_{(1)}\ a_{(2)}\ \cdots\ a_{(n)}] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right\|_1 = \left\| \sum_{\ell=1}^n x_\ell a_{(\ell)} \right\|_1$$

triangle inequality

because $\|a_{(\ell)}\|_1 \le \max_{1\le j\le 1} \|a_{(j)}\|_1$

$$\le \sum_{\ell=1}^n \|x_\ell a_{(\ell)}\|_1 = \sum_{\ell=1}^n |x_\ell| \|a_{(\ell)}\|_1 \le \left(\sum_{\ell=1}^n |x_\ell|\right) \cdot \left(\max_{1\le j\le n} \|a_{(j)}\|_1\right)$$

$$= \|x\|_1 \cdot \max_{1\le j\le n} \|a_{(j)}\|_1 = \max_{1\le j\le n} \|a_{(j)}\|_1$$

So, $\|Ax\|_1 \le \max_{1\le j\le n} \|a_{(j)}\|_1 \quad \forall x: \|x\|_1 = 1$

Taking max over all $x$ satisfying $\|x\|_1 = 1$

$$\max_{\|x\|_1=1} \|Ax\|_1 \le \max_{1\le j\le n} \|a_{(j)}\|_1$$

② For "$\ge$":

Let $j_0 = \arg\max_{1\le j\le n} \|a_{(j)}\|_1$, i.e., $\|a_{(j_0)}\|_1 = \max_{1\le j\le n} \|a_{(j_0)}\|_1$

Then $\|Ae_{j_0}\|_1 = \|a_{(j_0)}\|_1 = \max_{1\le j\le n} \|a_{(j)}\|_1$

Because $e_{j_0}$ satisfies $\|e_{j_0}\|_1 = 1$,

$\max_{\|x\|_1=1} \|Ax\|_1 \ge \|Ae_{j_0}\|_1$

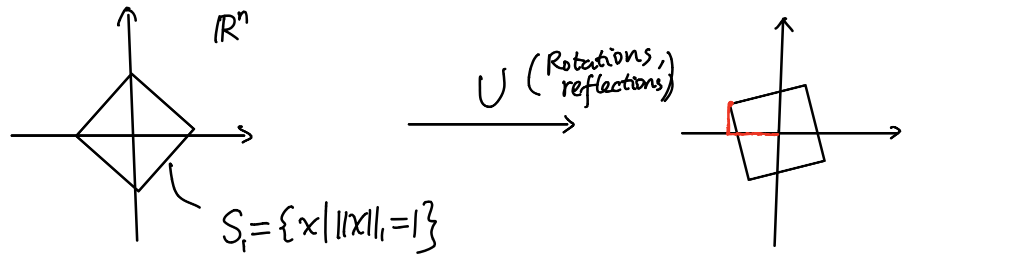
$$\Longrightarrow \max_{\|x\|_1=1} \|Ax\|_1 \ge \max_{1\le j\le n} \|a_{(j)}\|_1$$

⊠

For the identity matrix

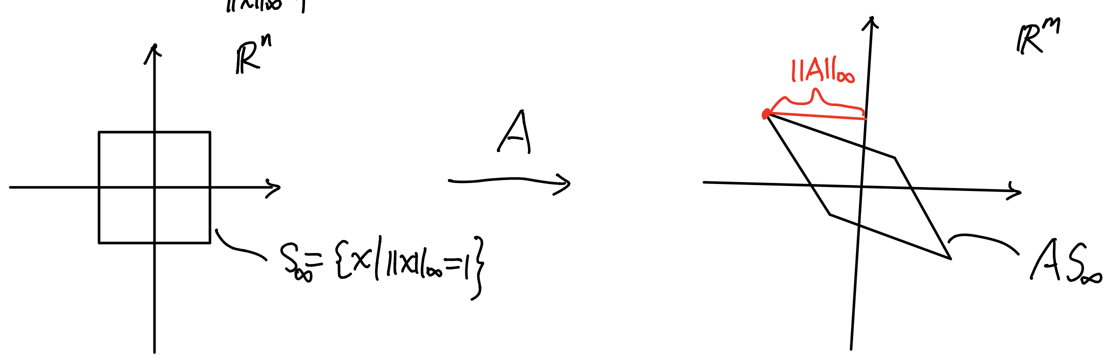
$$\|I\|_1 = \max_{\|x\|_1=1} \|Ix\|_1 = \max_{\|x\|_1=1} \|x\|_1 = 1$$

For a unitary matrix $U$, $\|U\|_1 \neq 1$ in general. (see the picture below)




Example 3: $\infty$-norm

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




Theorem: $\|A\|_\infty = \max_{1\le i\le m} \|a^{(i)}\|_1$,

where $A = \begin{bmatrix} (a^{(1)})^T \\ (a^{(2)})^T \\ \vdots \\ (a^{(m)})^T \end{bmatrix} \in \mathbb{R}^{m\times n}$ ☒

In other words, $\|A\|_\infty$ is the maximum 1-norm of row vectors.

$\boxed{\text{Question: Prove the theorem ?}}$

For matrix p-norms, besides ①②③ in the definition of matrix norm, the matrix p-norm also satisfies:

④ $\|Ax\|_p \le \|A\|_p \|x\|_p \quad \forall A \in \mathbb{R}^{m\times n}$ and $x \in \mathbb{R}^n$

⑤ $\|AB\|_p \le \|A\|_p \|B\|_p \quad \forall A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times q}$