

SI231b: Matrix Computations

Lecture 2: Basic Concepts (Part 1)

Yue Qiu(邱越)

qiuyue@shanghaitech.edu.cn

School of Information Science and Technology
ShanghaiTech University

Sept. 10, 2020

Basic Concepts: Part 1

- ▶ notation and conventions
- ▶ vector spaces
- ▶ vector norms and matrix norms
- ▶ subspaces

Notation and Conventions

\mathbb{R}	the set of real numbers, or real space
\mathbb{C}	the set of complex numbers, or complex space
\mathbb{R}^n	n -dimensional real space
\mathbb{C}^n	n -dimensional complex space
$\mathbb{R}^{m \times n}$	set of all $m \times n$ real-valued matrices
$\mathbb{C}^{m \times n}$	set of all $m \times n$ complex-valued matrices
\mathbf{x}	column vector
$x_i, [\mathbf{x}]_i$	i th entry of \mathbf{x}
\mathbf{A}	matrix
$a_{ij}, [\mathbf{A}]_{ij}$	(i, j) th entry of \mathbf{A}

- **vector:** $\mathbf{x} \in \mathbb{R}^n$ means that \mathbf{x} is a real-valued n -dimensional column vector, i.e.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R} \text{ for all } i.$$

Similarly, $\mathbf{x} \in \mathbb{C}^n$ means that \mathbf{x} is a complex-valued n -dimensional column vector.

- **transpose:** let $\mathbf{x} \in \mathbb{R}^n$. The notation \mathbf{x}^T means that

$$\mathbf{x}^T = [x_1, \quad x_2, \quad \dots, \quad x_n].$$

- **Hermitian transpose:** let $\mathbf{x} \in \mathbb{C}^n$. The notation \mathbf{x}^H means that

$$\mathbf{x}^H = [x_1^*, \quad x_2^*, \quad \dots, \quad x_n^*],$$

where the superscript $*$ denotes the complex conjugate.

- **matrix:** $\mathbf{A} \in \mathbb{R}^{m \times n}$ means that \mathbf{A} is real-valued $m \times n$ matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R} \text{ for all } i, j.$$

Similarly, $\mathbf{A} \in \mathbb{C}^{m \times n}$ means that \mathbf{A} is a complex-valued $m \times n$ matrix.

- unless specified, we denote the i th column of a matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ as $\mathbf{a}_i \in \mathbb{R}^m$, i.e.,

$$\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n].$$

The same notation applies to $\mathbf{A} \in \mathbb{C}^{m \times n}$.

- **transpose:** let $\mathbf{A} \in \mathbb{R}^{m \times n}$. The notation \mathbf{A}^T means that

$$\mathbf{A}^T = \begin{bmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \vdots & & & \vdots \\ a_{1n} & a_{m2} & \dots & a_{mn} \end{bmatrix} \in \mathbb{R}^{n \times m}.$$

- or, we have $\mathbf{B} = \mathbf{A}^T \iff b_{ij} = a_{ji}$ for all i, j .
- properties:
 - $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$
 - $(\mathbf{A}^T)^T = \mathbf{A}$
 - $(\mathbf{A} + \mathbf{B})^T = \mathbf{A}^T + \mathbf{B}^T$

- **Hermitian transpose:** let $\mathbf{A} \in \mathbb{C}^{m \times n}$. The notation \mathbf{A}^H means that

$$\mathbf{A}^H = \begin{bmatrix} a_{11}^* & a_{21}^* & \cdots & a_{m1}^* \\ a_{12}^* & a_{22}^* & \cdots & a_{m2}^* \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n}^* & a_{m2}^* & \cdots & a_{mn}^* \end{bmatrix} \in \mathbb{C}^{n \times m}.$$

- or, we have $\mathbf{B} = \mathbf{A}^T \iff b_{ij} = a_{ji}^*$ for all i, j .
- properties (the same as transpose):
 - $(\mathbf{AB})^H = \mathbf{B}^H \mathbf{A}^H$
 - $(\mathbf{A}^H)^H = \mathbf{A}$
 - $(\mathbf{A} + \mathbf{B})^H = \mathbf{A}^H + \mathbf{B}^H$

- **trace:** let $\mathbf{A} \in \mathbb{R}^{n \times n}$, the trace of \mathbf{A} is defined by

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}.$$

- properties:

► $\text{tr}(\mathbf{A}^T) = \text{tr}(\mathbf{A})$

► $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$

► $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$ for \mathbf{A}, \mathbf{B} of appropriate sizes

- **matrix power:** let $\mathbf{A} \in \mathbb{R}^{n \times n}$, then the notation \mathbf{A}^2 means $\mathbf{A}^2 = \mathbf{AA}$, and \mathbf{A}^k means

$$\mathbf{A}^k = \underbrace{\mathbf{AA} \cdots \mathbf{A}}_{k \text{ A's}}.$$

- ▶ **all-ones vector:** we use the notation

$$\mathbf{1} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

to denote a vector of all 1's.

- ▶ **zero vectors or matrices:** we use the notation $\mathbf{0}$ to denote either a vector of all zeros, or a matrix of all zeros.
- ▶ **unit vectors:** unit vectors are vectors that have only one nonzero element and the nonzero element is 1. We use the notation

$$\mathbf{e}_i = [0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0]^T$$

to denote a unit vector with the nonzero element at the i th entry.

- identity matrix:

$$\mathbf{I} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix},$$

where, as a convention, the empty entries are assumed to be zero.

- diagonal matrices: we use the notation

$$\text{diag}(a_1, \dots, a_n) = \begin{bmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{bmatrix}$$

to denote a diagonal matrix with diagonals a_1, \dots, a_n . We also use the shorthand notation $\text{diag}(\mathbf{a}) = \text{diag}(a_1, \dots, a_n)$.

- ▶ The matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ is said to be
 - square if $m = n$;
 - 'tall' if $m > n$;
 - 'fat' if $m < n$.
- ▶ The matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be
 - upper triangular if $a_{ij} = 0$ for all $i > j$;
 - strictly upper triangular if $a_{ij} = 0$ for all $i \geq j$;
 - lower triangular if $a_{ij} = 0$ for all $i < j$;
 - strictly lower triangular if $a_{ij} = 0$ for all $i \leq j$.

Examples:

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 2 & 0 \\ \frac{1}{8} & 3 & 2 \end{bmatrix}, \quad \mathbf{U} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}.$$

A vector space involves four things

- ▶ two sets \mathcal{V} and \mathcal{F}
- ▶ two algebraic operations, i.e., vector addition and scalar multiplication;

where

- ▶ \mathcal{V} is a nonempty set of objects called **vectors**;
- ▶ \mathcal{F} is a scalar field, for this course, either \mathbb{R} or \mathbb{C} ;
- ▶ vector addition, denoted by $\mathbf{x} + \mathbf{y}$ for $\mathbf{x}, \mathbf{y} \in \mathcal{V}$;
- ▶ scalar multiplication, denoted by $\alpha \mathbf{x}$ for $\alpha \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{V}$.

Axiomatic Definition of a Vector Space

The set \mathcal{V} is called a vector space over \mathcal{F} when the vector addition and scalar multiplication operations satisfy the following properties

- (A1) $\mathbf{x} + \mathbf{y} \in \mathcal{V}$ for all $\mathbf{x}, \mathbf{y} \in \mathcal{V}$, closure for vector addition
- (A2) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$ for every $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{V}$
- (A3) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$ for every $\mathbf{x}, \mathbf{y} \in \mathcal{V}$
- (A4) There is an element $\mathbf{0} \in \mathcal{V}$ such that $\mathbf{x} + \mathbf{0} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$
- (A5) $\forall \mathbf{x} \in \mathcal{V}, \exists (-\mathbf{x}) \in \mathcal{V}$ such that $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$
- (M1) $\alpha \mathbf{x} \in \mathcal{V}$ for $\forall \alpha \in \mathcal{F}$ and $\mathbf{x} \in \mathcal{V}$, closure for scalar multiplication
- (M2) $(\alpha\beta)\mathbf{x} = \alpha(\beta\mathbf{x})$ for $\forall \alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$
- (M3) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$ for every $\alpha \in \mathcal{F}$ and $\forall \mathbf{x}, \mathbf{y} \in \mathcal{V}$
- (M4) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$ for $\forall \alpha, \beta \in \mathcal{F}$ and every $\mathbf{x} \in \mathcal{V}$
- (M5) $1\mathbf{x} = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{V}$

- ▶ The set $\mathbb{R}^{m \times n}$ of $m \times n$ real matrices is a vector space over \mathbb{R} ;
- ▶ The set $\mathbb{C}^{m \times n}$ of $m \times n$ complex matrices is a vector space over \mathbb{C} ;
- ▶ The real coordinate spaces

$$\mathbb{R}^n = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, x_i \in \mathbb{R} \right\},$$

will be the main focus of our course.

Given an n -dimensional vector $\mathbf{x} \in \mathbb{R}^n$, the non-negative norm $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ is defined such that

1. $\|\mathbf{x}\| > 0$ when $\mathbf{x} \neq \mathbf{0}$ and $\|\mathbf{x}\| = 0 \iff \mathbf{x} = \mathbf{0}$

2. $\|k\mathbf{x}\| = |k|\|\mathbf{x}\|, \forall k \in \mathbb{R}$

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

- used to measure the length of a vector;
- also used to measure the distance of two vectors, specifically, via $\|\mathbf{x} - \mathbf{y}\|$ where \mathbf{x}, \mathbf{y} are two vectors.

Examples of norm:

► **2-norm** or Euclidean norm: $\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^n |x_i|^2} = (\mathbf{x}^T \mathbf{x})^{1/2}$

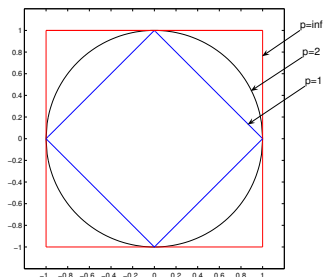
► **1-norm**, taxicab norm or Manhattan norm: $\|\mathbf{x}\|_1 = \sum_{i=1}^n |x_i|$

► **∞ -norm**: $\|\mathbf{x}\|_\infty = \max_{i=1, \dots, n} |x_i|$

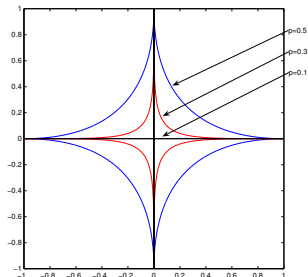
► **p -norm**, $p \geq 1$: $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$

Let

$$f_p(\mathbf{x}) = \left(\sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}, \quad p > 0.$$



(a) Contour plot of $f_p(\mathbf{x}) = 1$, $p \geq 1$



(b) Contour plot of $f_p(\mathbf{x}) = 1$, $p < 1$

- ▶ f_p is *not* a norm for $0 < p < 1$, **triangular inequality does not hold**
- ▶ ℓ_1 norm (1-norm) is widely used to get sparse results (cf. Figure(a))

Similar as the vector norm, a matrix norm is a function $\|\cdot\|: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ that satisfies

1. $\|\mathbf{A}\| > 0$ and $\|\mathbf{A}\| = 0 \iff \mathbf{A} = \mathbf{0}$
2. $\|k\mathbf{A}\| = |k|\|\mathbf{A}\|, \forall k \in \mathbb{R}$
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|, \forall \mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$
4. $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$ for all conformable matrices

Naturally, the Frobenius norm given by

$$\begin{aligned}\|A\|_F^2 &= \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \\ &= \sum_{i=1}^m \|\mathbf{A}_{i*}\|_2^2 \\ &= \sum_{j=1}^n \|\mathbf{A}_{*j}\|_2^2 \\ &= \text{tr}(\mathbf{A}^T \mathbf{A})\end{aligned}$$

satisfies the above definition. **Are there more?**

Induced by the aforementioned vector p -norm,

$$\|\mathbf{A}\|_p = \max_{\mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_p}{\|\mathbf{x}\|_p} = \max_{\|\mathbf{x}\|_p=1} \|\mathbf{Ax}\|_p$$

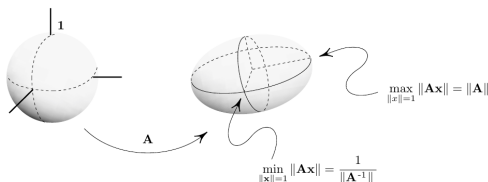
► $p = 1$

$$\begin{aligned}\|\mathbf{A}\|_1 &= \max_{\|\mathbf{x}\|_1=1} \|\mathbf{Ax}\|_1 = \max_j \sum_{i=1}^m |a_{ij}| \\ &= \text{the largest absolute column sum.}\end{aligned}$$

How to prove?

► $p = 2$

$$\|\mathbf{A}\|_2 = \max_{\|\mathbf{x}\|_2=1} \|\mathbf{Ax}\|_2$$



- also called spectral norm
- details will be discussed later in this course

► $p = \infty$

$$\begin{aligned}\|\mathbf{A}\|_\infty &= \max_{\|\mathbf{x}\|_\infty=1} \|\mathbf{Ax}\|_\infty = \max_i \sum_{j=1}^n |a_{ij}| \\ &= \text{the largest absolute row sum.}\end{aligned}$$

Let \mathcal{S} be a nonempty subset of a vector space \mathcal{V} over \mathcal{F} (i.e., $\mathcal{S} \subset \mathcal{V}$). If \mathcal{S}

- ▶ is also a vector space over \mathcal{F} ,
- ▶ uses the same addition and scalar multiplication operations

then \mathcal{S} is said to be a subspace of \mathcal{V}

Simplified definition

A nonempty subset \mathcal{S} of a vector space \mathcal{V} is a subspace of \mathcal{V} if and only if

1. $\mathbf{x}, \mathbf{y} \in \mathcal{S} \implies \mathbf{x} + \mathbf{y} \in \mathcal{S}$
2. $\mathbf{x} \in \mathcal{S} \implies \alpha \mathbf{x} \in \mathcal{S}$ for $\forall \alpha \in \mathcal{F}$