MA4230 Matrix Computation

Chapter 1: Preliminaries

1.1 Matrices

- Basic operations
- Connection to linear maps
- Range and nullspace
- Invertible matrices
- Orthogonality

Definition (Matrices and vectors)

For $m, n \in \mathbb{N} = \{1, 2, \dots\}$, define the set of real $m \times n$ **matrices** by

$$\mathbb{R}^{m\times n} := \left\{ \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \middle| \, a_{ij} \in \mathbb{R} \; \forall \, 1 \le i \le m, 1 \le j \le n \right\}.$$

Define $\mathbb{R}^m := \mathbb{R}^{m\times 1}$, the set of real (column) m-**vectors**.

Notation:

- $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ with $a_{ij} \in \mathbb{R}$ denoting entry in row i, column j,
- $A = (a_1|a_2|\cdots|a_n) \in \mathbb{R}^{m\times n}$ with $a_1, \dots, a_n \in \mathbb{R}^m$ the columns of A.

Example: $A = \begin{pmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{pmatrix} \in \mathbb{R}^{2\times 3}$. $\quad A = (a_{ij})$ with entries

$$a_{11} = 1,\; a_{12} = 3,\; a_{13} = 5,\; a_{21} = 2,\; a_{22} = 4,\; a_{23} = 6$$

and $A = (a_1|a_2|a_3)$ with columns $a_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, a_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}, a_3 = \begin{pmatrix} 5 \\ 6 \end{pmatrix}$.

1.1.1 Basic operations

For $A = (a_{ij}) \in \mathbb{R}^{m\times n}$, $B = (b_{ij}) \in \mathbb{R}^{m\times n}$, $C = (c_{ij}) \in \mathbb{R}^{n\times l}$, and $\alpha \in \mathbb{R}$:

- addition: $A + B \in \mathbb{R}^{m\times n}$,

$$(A+B)_{ij} := a_{ij} + b_{ij},$$

- scalar multiplication: $\alpha A \in \mathbb{R}^{m\times n}$,

$$(\alpha A)_{ij} := \alpha\, a_{ij},$$

- transposition: $A^{\mathrm{T}} \in \mathbb{R}^{n\times m}$,

$$(A^{\mathrm{T}})_{ij} := a_{ji},$$

- matrix multiplication: $AC \in \mathbb{R}^{m\times l}$,

$$(AC)_{ij} := \sum_{k=1}^{n} a_{ik} c_{kj}.$$

Matrix-vector product of $A = (a_{ij}) \in \mathbb{R}^{m\times n}$ and $x = (x_1, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$: $Ax \in \mathbb{R}^m$ with entries $(Ax)_i = \sum_{k=1}^{n} a_{ik} x_k$.

Matrix-vector and matrix-matrix products

(i) For $A = (a_1|\cdots|a_n) \in \mathbb{R}^{m\times n}$ and $x = (x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$:

$$Ax = (a_1|\cdots|a_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{k=1}^{n} x_k a_k \in \operatorname{span}(a_1, \ldots, a_n) \subseteq \mathbb{R}^m.$$

$\Longrightarrow$ Regard Ax not only as "A acts on x", but also as "x acts on A".

(ii) For $A = (a_1|\cdots|a_n) \in \mathbb{R}^{m\times n}$ and $C = (c_1|\cdots|c_l) \in \mathbb{R}^{n\times l}$:

$$AC = A(c_1|\cdots|c_n) = (Ac_1|\cdots|Ac_l) \in \mathbb{R}^{m\times l}.$$

Note colmuns of AC belong to $\operatorname{span}(a_1, \ldots, a_n) \subseteq \mathbb{R}^m$.

1.1.2 Connection to linear maps

Definition (Linear maps)

Let $m, n \in \mathbb{N}$. A map $f : \mathbb{R}^n \to \mathbb{R}^m$ is called linear iff

$$f(\alpha x + y) = \alpha f(x) + f(y) \qquad \forall x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}.$$

We denote the set of all linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ by $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$.

For $A \in \mathbb{R}^{m\times n}$, define $L_A : \mathbb{R}^n \to \mathbb{R}^m, x \mapsto Ax$ (**associated linear map**).
Note for $A, B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{n\times l}$ and $\alpha \in \mathbb{R}$:

$$L_{A+B} = L_A + L_B, \quad L_{\alpha A} = \alpha L_A, \quad L_{AC} = L_A \circ L_C.$$

Theorem (Characterization of linear maps)

There holds $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^m) = \{L_A : A \in \mathbb{R}^{m\times n}\}$.

"$\supseteq$": Associated linear maps are indeed linear.
"$\subseteq$": Let $f \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$. For any $x = \sum_{i=1}^n x_i e_i \in \mathbb{R}^n$, have
$f(x) = \sum_{i=1}^n x_i f(e_i) = Ax$ with $A = (f(e_1)|\cdots|f(e_n)) \in \mathbb{R}^{m\times n}$. $\square$

1.1.3 Range and nullspace

Let $A = (a_1|\cdots|a_n) \in \mathbb{R}^{m\times n}$. We define its

(i) **range** (or **column space**) $\mathcal{R}(A) := \{y \in \mathbb{R}^m | \exists\, x \in \mathbb{R}^n : y = Ax\} = \operatorname{span}(a_1, \ldots, a_n)$,

(ii) **nullspace** $\mathcal{N}(A) := \{x \in \mathbb{R}^n \,|\, Ax = 0\}$,

(iii) **rank** $\operatorname{rk}(A) := \dim(\mathcal{R}(A))$,

(iv) **nullity** $\operatorname{nullity}(A) := \dim(\mathcal{N}(A))$.

Theorem (Properties of rank)

Let $A, B \in \mathbb{R}^{m\times n}$ and $C \in \mathbb{R}^{n\times l}$. Then the following assertions hold.

(i) $0 \leq \operatorname{rk}(A) = \operatorname{rk}(A^{\mathrm{T}}) \leq \min\{m, n\}$ *("column rank equals row rank")*,

(ii) $\operatorname{rk}(A) + \operatorname{nullity}(A) = n$ *(rank-nullity theorem)*,

(iii) $\operatorname{rk}(A) + \operatorname{rk}(C) - n \leq \operatorname{rk}(AC) \leq \min\{\operatorname{rk}(A), \operatorname{rk}(C)\}$ *(Sylvester ineqy)*,

(iv) $\operatorname{rk}(A + B) \leq \operatorname{rk}(A) + \operatorname{rk}(B)$,

(v) $\operatorname{rk}(A^{\mathrm{T}}A) = \operatorname{rk}(A) = \operatorname{rk}(AA^{\mathrm{T}})$.

We say $A \in \mathbb{R}^{m\times n}$ has **full rank** iff $\operatorname{rk}(A) = \min\{m, n\}$ (otherwise **rank-deficient**).

Theorem (Characterization of full-rank tall matrices)

Let $A = (a_1 | \cdots | a_n) \in \mathbb{R}^{m \times n}$*,* $m \geq n$*. Then, the following are equivalent:*

(i) A *is of full rank, i.e.,* $\mathrm{rk}(A) = n$*.*

(ii) $a_1, \ldots, a_n$ *are linearly independent.*

(iii) L_A *is injective.*

Proof: (i)⇒(ii): If $\mathrm{rk}(A) = \dim(\mathrm{span}(a_1, \ldots, a_n)) = n$, then clearly $a_1, \ldots, a_n$ are linearly independent.

(ii)⇒(iii): Suppose $a_1, \ldots, a_n$ are linearly independent, and let $x = (x_1, \ldots, x_n)^{\mathrm{T}}, y = (y_1, \ldots, y_n)^{\mathrm{T}} \in \mathbb{R}^n$ such that $L_A(x) = L_A(y)$, i.e., $Ax = Ay$. Then $A(x - y) = \sum_{i=1}^{n} (x_i - y_i) a_i = 0 \in \mathbb{R}^m$ and hence, $x_i - y_i = 0$ for all $1 \leq i \leq n$, i.e., $x = y$.

¬(i)⇒ ¬(iii): Suppose that A is not of full rank. Then, $\mathrm{rk}(A) = \dim(\mathrm{span}(a_1, \ldots, a_n)) < n$ and hence, $a_1, \ldots, a_n$ are linearly dependent. Then, there exists $c = (c_1, \ldots, c_n)^{\mathrm{T}} \in \mathbb{R}^n \backslash \{0\}$ such that $L_A(c) = \sum_{i=1}^{n} c_i a_i = 0$ and we conclude that L_A is not injective. □

1.1.4 Invertible matrices

Definition (Invertible matrix)

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be **invertible** (or **non-singular**) iff there exists a matrix $A^{-1} \in \mathbb{R}^{n\times n}$, called the inverse of A, such that

$$AA^{-1} = A^{-1}A = I_n.$$

Here, I_n denotes the $n \times n$ **identity matrix**

$$I_n := (e_1|e_2|\cdots|e_n) := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Note that for $A = (a_1|\cdots|a_n) \in \mathbb{R}^{n\times n}$ invertible, $b \in \mathbb{R}^n$:

Writing $x = A^{-1}b = (x_1, \ldots, x_n)^{\mathrm{T}}$: $b = Ax = \sum_{k=1}^{n} x_k a_k, \; A^{-1}b = \sum_{k=1}^{n} x_k e_k.$

$\Longrightarrow$ Left-multiplication by A^{-1} is a change of basis operation.

Characterization of invertible matrices:

Theorem (Characterization of invertibility)

For $A \in \mathbb{R}^{n \times n}$, the following are equivalent:

(i) *A is invertible.*

(ii) *L_A is an invertible linear map.*

(iii) *A has full rank, i.e., $\mathrm{rk}(A) = n$.*

(iv) *$\mathscr{R}(A) = \mathbb{R}^n$ (or equivalently, L_A is surjective).*

(v) *$\mathscr{N}(A) = \{0\}$ (or equivalently, L_A is injective).*

(vi) *$\det(A) \neq 0$.*

(vii) *$0 \notin \Lambda(A)$.*

Here, $\Lambda(A) := \{\lambda \in \mathbb{C} : \det(A - \lambda I_n) = 0\}$ is the **spectrum** of $A \in \mathbb{R}^{n \times n}$.

Theorem (Properties for inverse)

Let $A, C \in \mathbb{R}^{n\times n}$ invertible, $\alpha \in \mathbb{R}\backslash\{0\}$. Then, $A^{-1}, AC, \alpha A, A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$ are invertible and we have

(i) $(A^{-1})^{-1} = A$, $(AC)^{-1} = C^{-1}A^{-1}$, $(\alpha A)^{-1} = \frac{1}{\alpha}A^{-1}$, $(A^{\mathrm{T}})^{-1} = (A^{-1})^{\mathrm{T}}$.

(ii) $\mathrm{rk}(A^{-1}) = \mathrm{rk}(A) = n, \quad \det(A^{-1}) = \frac{1}{\det(A)}$.

Compare with transposition: For $A, B \in \mathbb{R}^{m\times n}$, $C \in \mathbb{R}^{n\times l}$, $\alpha \in \mathbb{R}$:

(i) $(A^{\mathrm{T}})^{\mathrm{T}} = A, \quad (AC)^{\mathrm{T}} = C^{\mathrm{T}}A^{\mathrm{T}}, \quad (\alpha A)^{\mathrm{T}} = \alpha A^{\mathrm{T}}$.

(ii) $\mathrm{rk}(A^{\mathrm{T}}) = \mathrm{rk}(A), \quad \det(A^{\mathrm{T}}) = \det(A)$.

Important classes of matrices:

Definition (Symmetric matrix, orthogonal matrix)

(i) A matrix $A \in \mathbb{R}^{n\times n}$ is said to be **symmetric** iff $A^{\mathrm{T}} = A$.

(ii) A matrix $Q \in \mathbb{R}^{n\times n}$ is said to be **orthogonal** iff $QQ^{\mathrm{T}} = Q^{\mathrm{T}}Q = I_n$, i.e., iff Q is invertible and $Q^{-1} = Q^{\mathrm{T}}$.

1.1.5 Orthogonality

Definition (Euclidean inner product and Euclidean norm)

Let $x, y \in \mathbb{R}^n$. We define

(i) the **Euclidean inner product** $\langle x, y\rangle := x^{\mathrm{T}} y \in \mathbb{R}$, and

(ii) the **Euclidean norm** $\|x\|_2 := \sqrt{\langle x, x\rangle} \in \mathbb{R}$.

Properties:

- $\langle x, y\rangle = \|x\|_2 \|y\|_2 \cos(\theta_{x,y})$ with $\theta_{x,y}$ angle between x and y.
- $\langle \cdot, \cdot\rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is bilinear (linear in both arguments) and symmetric ($\langle x, y\rangle = \langle y, x\rangle \ \forall x, y \in \mathbb{R}^n$)

Definition (Orthogonal vectors and subsets)

(i) $x, y \in \mathbb{R}^n$ are **orthogonal** ($x \perp y$), iff $\langle x, y\rangle = 0$.

(ii) $X, Y \subseteq \mathbb{R}^n$ are orthogonal ($X \perp Y$), iff $x \perp y \ \forall x \in X, y \in Y$.

(iii) $S \subseteq \mathbb{R}^n \backslash \{0\}$ is orthogonal iff $\forall x, y \in S : x \neq y \Longrightarrow x \perp y$.

(iv) $S \subseteq \mathbb{R}^n \backslash \{0\}$ is **orthonormal** iff S is orthogonal and $\|x\|_2 = 1 \ \forall x \in S$.

Theorem (Vectors in orthogonal set linearly independent)

The vectors in an orthogonal set $S \subseteq \mathbb{R}^n \backslash \{0\}$ are linearly independent. In particular, any orthogonal set $S \subseteq \mathbb{R}^n \backslash \{0\}$ containing n vectors is a basis for $\mathbb{R}^n$.

Proof.

Let $S = \{v_1, \ldots, v_N\} \subseteq \mathbb{R}^n \backslash \{0\}$ orthogonal set, and suppose its elements were linearly dependent. Then, $\exists\, v_k \in S : v_k = \sum_{i \in \{1,\ldots,N\} \backslash \{k\}} c_i v_i$ for some $\{c_i\} \subseteq \mathbb{R}$. Have

$$\|v_k\|_2^2 = \langle v_k, v_k \rangle = \sum_{i \in \{1,\ldots,N\} \backslash \{k\}} c_i \langle v_i, v_k \rangle = 0.$$

$\implies v_k = 0$, contradicting $v_k \in S \subseteq \mathbb{R}^n \backslash \{0\}$. □

Decomposing a vector into orthogonal components

Given $x \in \mathbb{R}^n$, orthonormal set $\{q_1, q_2, \ldots, q_N\} \subseteq \mathbb{R}^n \backslash \{0\}$, $N \leq n$. Write

$$x = \sum_{k=1}^{N} \langle x, q_k \rangle q_k + r = \sum_{k=1}^{N} (q_k q_k^{\mathrm{T}}) x + r.$$

Then $\{r\} \perp \{q_1, \ldots, q_N\}$ as

$$\langle r, q_i \rangle = \langle x, q_i \rangle - \sum_{k=1}^{N} \langle x, q_k \rangle \langle q_k, q_i \rangle = \langle x, q_i \rangle - \langle x, q_i \rangle = 0 \qquad \forall 1 \leq i \leq N,$$

$\implies r$ is the part of x orthogonal to the subspace $\mathrm{span}(q_1, \ldots, q_N) \subseteq \mathbb{R}^n$, and $(q_k q_k^{\mathrm{T}}) x$ is the part of x in direction q_k for $1 \leq k \leq N$.

Later: $P_q := q q^{\mathrm{T}}$ *orthogonal projector* onto $\mathrm{span}(q)$.

Observation: if $N = n$, have $\{q_1, \ldots, q_n\}$ is basis of $\mathbb{R}^n$ and hence $r = 0$.

Let $Q = (q_1 | \cdots | q_n) \in \mathbb{R}^{n\times n}$ orthogonal matrix. Then,

- $\{q_1, \ldots, q_n\} \subseteq \mathbb{R}^n$ orthonormal basis ($Q^\mathrm{T} Q = I_n$ yields $q_i^\mathrm{T} q_j = \delta_{ij}$).
- $\forall x, y \in \mathbb{R}^n$: $\langle Qx, Qy\rangle = x^\mathrm{T} Q^\mathrm{T} Q y = x^\mathrm{T} y = \langle x, y\rangle$ and $\|Qx\|_2 = \|x\|_2$. "Euclidean inner product is invariant under orthogonal transformations".
- $|\det(Q)| = 1$ $\big(1 = \det(I_n) = \det(Q^\mathrm{T} Q) = \det(Q^\mathrm{T})\det(Q) = (\det(Q))^2\big)$
- L_Q is an orthogonal transformation preserving the inner product on $\mathbb{R}^n$, and corresponds to a rigid rotation (when $\det(Q) = 1$) or a reflection (when $\det(Q) = -1$) of the space.

2D: An orthogonal matrix $Q \in \mathbb{R}^{2\times 2}$ with $\det(Q) = 1$ can be written as

$$Q = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \theta \in [0, 2\pi),$$

with L_Q rotating plane anticlockwise by the angle θ.
An orthogonal matrix $Q \in \mathbb{R}^{2\times 2}$ with $\det(Q) = -1$ can be written as

$$Q = \begin{pmatrix} \cos(\beta) & \sin(\beta) \\ \sin(\beta) & -\cos(\beta) \end{pmatrix}, \quad \beta \in [0, 2\pi),$$

with L_Q reflecting plane across $y = \tan(\frac{\beta}{2})x$ if $\beta \neq \pi$ (else, across $x = 0$).

1.2 Norms

- Vector norms
- Induced matrix norms
- Frobenius norm
- Orthogonal invariance

What is a norm?

Definition (Norm)

Let V be a vector space over $\mathbb{R}$. A map $\|\cdot\| : V \to [0, \infty)$ is called a **norm** on V iff there holds

(i) definiteness: $\forall v \in V : \|v\| = 0 \Longrightarrow v = 0$,

(ii) absolute homogeneity: $\|\alpha v\| = |\alpha|\|v\| \quad \forall v \in V, \alpha \in \mathbb{R}$,

(iii) triangle inequality: $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\| \quad \forall v_1, v_2 \in V$.

If $V = \mathbb{R}^n$, say $\|\cdot\|$ is a **vector norm**. If $V = \mathbb{R}^{m \times n}$, say $\|\cdot\|$ is a **matrix norm**.

Important vector norms: the p-norms $\|\cdot\|_p$ (Euclidean norm for $p = 2$).

1.2.1 Vector norms

Definition (The p-norms)

For $p \in [1, \infty)$, define the **p-norm** $\| \cdot \|_p : \mathbb{R}^n \to [0, \infty)$,

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \qquad x = (x_1, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^n.$$

Define the ∞**-norm** (or **maximum norm**) $\| \cdot \|_\infty : \mathbb{R}^n \to [0, \infty)$,

$$\|x\|_\infty := \max_{1 \le i \le n} |x_i|, \qquad x = (x_1, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^n.$$

Observation: In 1D ($n = 1$) have $\| \cdot \|_p = | \cdot | \quad \forall p \in [1, \infty) \cup \{\infty\}$.

Theorem (Hölder's inequality)

Let $p, q \in (1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = 1$. *Then, for any* $x, y \in \mathbb{R}^n$*:*

$$|\langle x, y\rangle| = \left|\sum_{i=1}^{n} x_i y_i\right| \leq \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}} = \|x\|_p \|y\|_q.$$

The case $p = q = 2$ *is also known as the* **Cauchy–Schwarz inequality***.*

Lemma (Young's inequality)

For $p, q \in (1, \infty)$ *with* $\frac{1}{p} + \frac{1}{q} = 1$ *have* $\forall a, b \geq 0$*:* $ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q$*.*

Pf: Assume $a, b > 0$ (claim trivial if $a = 0$ or $b = 0$). Key observation: $\exp : \mathbb{R} \to \mathbb{R}$ is convex, i.e., for any $\alpha \in [0, 1]$ and $x, y \in \mathbb{R}$ we have $e^{\alpha x + (1-\alpha) y} \leq \alpha e^x + (1 - \alpha) e^y$. Hence,

$$ab = e^{\log(ab)} = e^{\frac{1}{p}(p \log(a)) + (1 - \frac{1}{p})(q \log(b))} \leq \frac{1}{p} e^{p \log(a)} + (1 - \frac{1}{p}) e^{q \log(b)}$$

$$\implies ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q.$$

□

Let us now prove Hölder's inequality $|\langle x, y\rangle| \leq \|x\|_p \|y\|_q$:

Proof of Hölder's inequality.

Assume $x, y \in \mathbb{R}^n \backslash \{0\}$ (claim trivial if $x = 0$ or $y = 0$). Then,

$$\frac{|\langle x, y\rangle|}{\|x\|_p \|y\|_q} \leq \sum_{i=1}^{n} \frac{|x_i|}{\|x\|_p} \frac{|y_i|}{\|y\|_q} \leq \frac{1}{p} \frac{\sum_{i=1}^{n} |x_i|^p}{\|x\|_p^p} + \frac{1}{q} \frac{\sum_{i=1}^{n} |y_i|^q}{\|y\|_q^q} = \frac{1}{p} + \frac{1}{q} = 1.$$

□

Rk: We also have $|\langle x, y\rangle| \leq \|x\|_1 \|y\|_\infty$ for any $x, y \in \mathbb{R}^n$ as

$$|\langle x, y\rangle| = \left| \sum_{i=1}^{n} x_i y_i \right| \leq \sum_{i=1}^{n} |x_i| \, |y_i| \leq \|y\|_\infty \sum_{i=1}^{n} |x_i| = \|y\|_\infty \|x\|_1 .$$

Are the p-norms really norms? Yes:

Theorem (p-norms are norms)

The map $\|\cdot\|_p : \mathbb{R}^n \to [0,\infty)$ *is indeed a norm for any* $p \in [1,\infty) \cup \{\infty\}$.

Proof.

Let us only show the triangle inequality for $p \in (1,\infty)$. Key: Hölder.
Let $p \in (1,\infty)$. Set $q := \frac{p}{p-1}$ (then $\frac{1}{p} + \frac{1}{q} = 1$). For any $x, y \in \mathbb{R}^n$:

$$\begin{aligned}
\|x+y\|_p^p &= \sum_{i=1}^n |x_i + y_i||x_i + y_i|^{p-1} \\
&\le \sum_{i=1}^n |x_i||x_i + y_i|^{p-1} + \sum_{i=1}^n |y_i||x_i + y_i|^{p-1} \\
&\le (\|x\|_p + \|y\|_p) \left(\sum_{i=1}^n |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}} \\
&= (\|x\|_p + \|y\|_p) \|x+y\|_{(p-1)q}^{p-1} = (\|x\|_p + \|y\|_p) \|x+y\|_p^{p-1},
\end{aligned}$$

and hence, $\|x+y\|_p \le \|x\|_p + \|y\|_p$. □

Equivalence of vector norms

Theorem (Equivalence of vector norms)

Let $\| \cdot \|, ||| \cdot ||| : \mathbb{R}^n \to [0, \infty)$ *be norms on* $\mathbb{R}^n$*. Then,* $\| \cdot \|$ *and* $||| \cdot |||$ *are equivalent, that is, there exist constants* $C_1, C_2 > 0$ *such that*

$$C_1 \|x\| \leq |||x||| \leq C_2 \|x\| \qquad \forall x \in \mathbb{R}^n.$$

Actually, any two norms on a finite dimensional space are equivalent.

1.2.2 Induced matrix norms

First observation: note that for $A = (a_1|\cdots|a_n) \in \mathbb{R}^{m\times n}$, we have

$$\mathrm{vec}(A) := \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^{mn}, \qquad \text{(note } a_i \in \mathbb{R}^m \ \forall 1 \leq i \leq n\text{)}$$

and we can use the aforementioned vector norms to measure its size. However, it is more useful to view $A \in \mathbb{R}^{m\times n}$ in terms of the associated linear operator $L_A \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^m)$ and use the operator norm induced by given vector norms on $\mathbb{R}^n$ and $\mathbb{R}^m$.

Definition (induced matrix norm)

Consider the normed vector spaces $(\mathbb{R}^n, \|\cdot\|_{(n)})$ and $(\mathbb{R}^m, \|\cdot\|_{(m)})$. Then we define the **induced matrix norm** $\|\cdot\|_{(m,n)} : \mathbb{R}^{m\times n} \to [0,\infty)$ by

$$\|A\|_{(m,n)} := \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_{(m)}}{\|x\|_{(n)}} = \sup_{\substack{x\in\mathbb{R}^n \\ \|x\|_{(n)}=1}} \|Ax\|_{(m)}, \qquad A \in \mathbb{R}^{m\times n}.$$

In the case that $\|\cdot\|_{(n)} = \|\cdot\|_{(m)} = \|\cdot\|_p$ for $p \in [1,\infty) \cup \{\infty\}$, we call

$$\|A\|_p := \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_p}{\|x\|_p} = \sup_{\substack{x \in \mathbb{R}^n \\ \|x\|_p = 1}} \|Ax\|_p, \qquad A \in \mathbb{R}^{m \times n}$$

the p-norm of A.

Theorem (induced norm is a norm)

The map $\|\cdot\|_{(m,n)} : \mathbb{R}^{m\times n} \to [0,\infty)$ is a norm on $\mathbb{R}^{m\times n}$ for any choice of vector norms $\|\cdot\|_{(n)}$ on $\mathbb{R}^n$ and $\|\cdot\|_{(m)}$ on $\mathbb{R}^m$.

Proof: Exercise. □

Observe: The number $\|A\|_{(m,n)}$ is the smallest constant $C \geq 0$ such that

$$\|L_A(x)\|_{(m)} = \|Ax\|_{(m)} \leq C\|x\|_{(n)} \qquad \forall x \in \mathbb{R}^n,$$

i.e., it is the greatest factor by which L_A can stretch a vector in $\mathbb{R}^n$.

Induced matrix norms are submultiplicative

For $n_1, n_2, n_3 \in \mathbb{N}$ let $\|\cdot\|_{(n_k)}$ be a norm on $\mathbb{R}^{n_k}$, and let $A \in \mathbb{R}^{n_1 \times n_2}$ and $C \in \mathbb{R}^{n_2 \times n_3}$. Then,

$$\|AC\|_{(n_1,n_3)} \leq \|A\|_{(n_1,n_2)} \|C\|_{(n_2,n_3)}, \text{ i.e.,}$$

$$\sup_{x \in \mathbb{R}^{n_3} \setminus \{0\}} \frac{\|ACx\|_{(n_1)}}{\|x\|_{(n_3)}} \leq \left[\sup_{x \in \mathbb{R}^{n_2} \setminus \{0\}} \frac{\|Ax\|_{(n_1)}}{\|x\|_{(n_2)}} \right] \left[\sup_{x \in \mathbb{R}^{n_3} \setminus \{0\}} \frac{\|Cx\|_{(n_2)}}{\|x\|_{(n_3)}} \right].$$

(Warning: not every matrix norm is submultiplicative! (exercise))

Proof: For any $x \in \mathbb{R}^{n_3}$:

$$\|A \underbrace{Cx}_{\in \mathbb{R}^{n_2}} \|_{(n_1)} \leq \|A\|_{(n_1,n_2)} \|Cx\|_{(n_2)} \leq \|A\|_{(n_1,n_2)} \|C\|_{(n_2,n_3)} \|x\|_{(n_3)}. \quad \square$$

Let's do some examples for computing matrix norms:

Example 1: p-norms of a diagonal matrix

$$A := \operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n) := \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix} \in \mathbb{R}^{n\times n}.$$

Then, $\|A\|_p = \max_{1\le i\le n}|\alpha_i|$ for all $p \in [1,\infty) \cup \{\infty\}$.

Proof for $p \in [1,\infty)$: Write $m := \max_{1\le i\le n}|\alpha_i|$. For any $x \in \mathbb{R}^n$:

$$\|Ax\|_p^p = \sum_{i=1}^{n}|\alpha_i x_i|^p \le \left(\max_{1\le i\le n}|\alpha_i|^p\right)\sum_{i=1}^{n}|x_i|^p = m^p\|x\|_p^p, \quad \|Ax\|_p \le m\|x\|_p,$$

$\implies \|A\|_p \le m$. Converse inequality:

$$\|A\|_p = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_p}{\|x\|_p} \ge \frac{\|Ae_i\|_p}{\|e_i\|_p} = \frac{\|\alpha_i e_i\|_p}{\|e_i\|_p} = |\alpha_i| \qquad \forall 1 \le i \le n,$$

$\implies \|A\|_p \ge m$. □

$$A := \operatorname{diag}(\alpha_1, \alpha_2, \ldots, \alpha_n) := \begin{pmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_n \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Then, $\|A\|_p = \max_{1 \le i \le n} |\alpha_i|$ for all $p \in [1, \infty) \cup \{\infty\}$.

<u>Proof for $p = \infty$</u>: Write $m := \max_{1 \le i \le n} |\alpha_i|$. For any $x \in \mathbb{R}^n$:

$$\|Ax\|_\infty = \max_{1 \le i \le n} |\alpha_i x_i| \le \left(\max_{1 \le i \le n} |\alpha_i| \right) \left(\max_{1 \le i \le n} |x_i| \right) = m \|x\|_\infty,$$

$\implies \|A\|_\infty \le m$. Converse inequality:

$$\|A\|_\infty = \sup_{x \in \mathbb{R}^n \setminus \{0\}} \frac{\|Ax\|_\infty}{\|x\|_\infty} \ge \frac{\|Ae_i\|_\infty}{\|e_i\|_\infty} = \frac{\|\alpha_i e_i\|_\infty}{\|e_i\|_\infty} = |\alpha_i| \qquad \forall 1 \le i \le n,$$

$\implies \|A\|_\infty \ge m$. □

Example 2: ∞-norm and 1-norm of a matrix

For $A = (a_1|\cdots|a_n) = (b_1|\cdots|b_m)^{\mathrm{T}} \in \mathbb{R}^{m\times n}$:

$$\|A\|_\infty = \max_{1\le i\le m} \|b_i\|_1, \qquad \|A\|_1 = \max_{1\le j\le n} \|a_j\|_1,$$

i.e., $\|A\|_\infty$ is "maximum row sum", $\|A\|_1$ "maximum column sum" of A.

Proof for 1-norm: Write $m := \max_{1\le j\le n} \|a_j\|_1$. For any $x \in \mathbb{R}^n$:

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

$\implies \|A\|_1 \le m$. Converse inequality:

$$\|A\|_1 = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_1}{\|x\|_1} \ge \frac{\|Ae_i\|_1}{\|e_i\|_1} = \frac{\|a_i\|_1}{1} = \|a_i\|_1 \qquad \forall 1 \le i \le n,$$

$\implies \|A\|_1 \ge m$. □

Proof for ∞-norm: Exercise.

Example 3: Matrix 2-norm of a row vector

Consider a row vector

$$A := a^{\mathrm{T}} \in \mathbb{R}^{1\times n} \qquad (a \in \mathbb{R}^n).$$

Then, $\|A\|_2 = \|a\|_2$ (lhs: matrix 2-norm, rhs: vector 2-norm).

<u>Proof:</u> For any $x \in \mathbb{R}^n$:

$$\|Ax\|_2 = \|a^{\mathrm{T}}x\|_2 = |\langle a, x\rangle| \le \|a\|_2\|x\|_2$$

$\Longrightarrow \|A\|_2 \le \|a\|_2$. Converse inequality:

If $a = 0 \in \mathbb{R}^n$, then $\|A\|_2 \le 0$ which yields $\|A\|_2 = 0 = \|a\|_2$.
If $a \ne 0 \in \mathbb{R}^n$, then

$$\|A\|_2 \ge \frac{\|Aa\|_2}{\|a\|_2} = \frac{|\langle a, a\rangle|}{\|a\|_2} = \|a\|_2.$$

$\Longrightarrow \|A\|_2 \ge \|a\|_2$. $\square$

Example 4: Matrix 2-norm of outer product

Let $u \in \mathbb{R}^m$, $v \in \mathbb{R}^n$. Consider the **outer product**

$$A := uv^{\mathrm{T}} \in \mathbb{R}^{m \times n}.$$

Then, $\|A\|_2 = \|u\|_2 \|v\|_2$.

<u>Proof:</u> For any $x \in \mathbb{R}^n$:

$$\|Ax\|_2 = \|uv^{\mathrm{T}} x\|_2 = \|u\|_2 |\langle v, x \rangle| \le \|u\|_2 \|v\|_2 \|x\|_2$$

$\Longrightarrow \|A\|_2 \le \|u\|_2 \|v\|_2$. Converse inequality:

If $v = 0$, then $\|A\|_2 \le 0$ which yields $\|A\|_2 = 0 = \|u\|_2 \|v\|_2$.
If $v \neq 0$, then

$$\|A\|_2 \ge \frac{\|Av\|_2}{\|v\|_2} = \frac{\|uv^{\mathrm{T}} v\|_2}{\|v\|_2} = \frac{\|u\|_2 |\langle v, v \rangle|}{\|v\|_2} = \|u\|_2 \|v\|_2$$

$\Longrightarrow \|A\|_2 \ge \|u\|_2 \|v\|_2$. □

The matrix 2-norm is also called the **spectral norm**.

Later: For $A \in \mathbb{R}^{m \times n}$:

$$\|A\|_2 = \sqrt{\lambda_{\max}(A^{\mathrm{T}}A)},$$

where $\lambda_{\max}(A^{\mathrm{T}}A)$ largest eigenvalue of $A^{\mathrm{T}}A$.

Next: Most important example of a norm which is not induced by vector norms: the Frobenius norm.

1.2.3 Frobenius norm

Definition (Frobenius norm)

The map $\|\cdot\|_F : \mathbb{R}^{m\times n} \to [0,\infty)$ given by

$$\|A\|_F := \sqrt{\operatorname{tr}(A^{\mathrm{T}}A)} = \sqrt{\operatorname{tr}(AA^{\mathrm{T}})} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2}, \quad A = (a_{ij}) \in \mathbb{R}^{m\times n}$$

is called the **Frobenius norm**.

$\operatorname{tr}(B)$ denotes the trace of a square matrix B, i.e., sum of diagonal entries.

Theorem (submultiplicativity of Frobenius norm)

The map $\|\cdot\|_F$ is a norm on $\mathbb{R}^{m\times n}$. Further, it is submultiplicative:

$$\|AC\|_F \leq \|A\|_F\|C\|_F \qquad \forall A \in \mathbb{R}^{m\times n}, C \in \mathbb{R}^{n\times l}.$$

Proof: Exercise.

Frobenius inner product

$\|\cdot\|_F$ is induced by **Frobenius inner product** $\langle\cdot,\cdot\rangle_F : \mathbb{R}^{m\times n} \times \mathbb{R}^{m\times n} \to \mathbb{R}$:

$$\langle A, B\rangle_F := \mathrm{tr}(A^{\mathrm{T}}B) = \mathrm{tr}(BA^{\mathrm{T}}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}b_{ij}, \qquad A, B \in \mathbb{R}^{m\times n},$$

i.e., $\|A\|_F = \sqrt{\langle A, A\rangle_F}$ for any $A \in \mathbb{R}^{m\times n}$.

We have

$$|\langle A, B\rangle_F| \le \|A\|_F\|B\|_F \qquad \forall A, B \in \mathbb{R}^{m\times n}.$$

(Cauchy–Schwarz inequality for inner product spaces)

Equivalence of matrix norms

Theorem (equivalence of matrix norms)

Let $\|\cdot\|, |||\cdot||| : \mathbb{R}^{m\times n} \to [0,\infty)$ *be norms on* $\mathbb{R}^{m\times n}$*. Then,* $\|\cdot\|$ *and* $|||\cdot|||$ *are equivalent, that is, there exist constants* $C_1, C_2 > 0$ *such that*

$$C_1\|A\| \leq |||A||| \leq C_2\|A\| \qquad \forall A \in \mathbb{R}^{m\times n}.$$

(recall: any two norms on a finite dimensional space are equivalent)

1.2.4 Orthogonal invariance

The spectral norm $\|\cdot\|_2$ and the Frobenius norm $\|\cdot\|_F$ are invariant under multiplication by orthogonal matrices:

Theorem (Orthogonal invariance of spectral norm and Frobenius norm)

Let $A \in \mathbb{R}^{m\times n}$. Let $U \in \mathbb{R}^{m\times m}$, $V \in \mathbb{R}^{n\times n}$ orthogonal matrices. Then,

(i) $\|UA\|_2 = \|A\|_2, \qquad \|AV\|_2 = \|A\|_2,$

(ii) $\|UA\|_F = \|A\|_F, \qquad \|AV\|_F = \|A\|_F.$

Proof of (i): Note $\|Vx\|_2 = \|x\|_2$, $\|Uy\|_2 = \|y\|_2$ $\forall x \in \mathbb{R}^n, y \in \mathbb{R}^m$. Have

$$\|UA\|_2 = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|UAx\|_2}{\|x\|_2} = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ax\|_2}{\|x\|_2} = \|A\|_2.$$

Also,

$$\|AV\|_2 = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|AVx\|_2}{\|x\|_2} = \sup_{x\in\mathbb{R}^n\setminus\{0\}} \frac{\|AVx\|_2}{\|Vx\|_2} = \sup_{\tilde{x}\in\mathbb{R}^n\setminus\{0\}} \frac{\|A\tilde{x}\|_2}{\|\tilde{x}\|_2} = \|A\|_2$$

using that $L_V : \mathbb{R}^n \to \mathbb{R}^n$ is a bijection (as V is invertible). □

Theorem (Orthogonal invariance of spectral norm and Frobenius norm)

Let $A \in \mathbb{R}^{m \times n}$. *Let* $U \in \mathbb{R}^{m \times m}$, $V \in \mathbb{R}^{n \times n}$ *orthogonal matrices. Then,*

(i) $\|UA\|_2 = \|A\|_2, \qquad \|AV\|_2 = \|A\|_2,$

(ii) $\|UA\|_F = \|A\|_F, \qquad \|AV\|_F = \|A\|_F.$

Proof of (ii): Recall that

$$\|B\|_F^2 = \mathrm{tr}(B^{\mathrm{T}}B) = \mathrm{tr}(BB^{\mathrm{T}}) \qquad \forall B \in \mathbb{R}^{m \times n}.$$

Hence,

$$\|UA\|_F^2 = \mathrm{tr}((UA)^{\mathrm{T}}(UA)) = \mathrm{tr}(A^{\mathrm{T}}U^{\mathrm{T}}UA) = \mathrm{tr}(A^{\mathrm{T}}A) = \|A\|_F^2,$$

and

$$\|AV\|_F^2 = \mathrm{tr}((AV)(AV)^{\mathrm{T}}) = \mathrm{tr}(AVV^{\mathrm{T}}A^{\mathrm{T}}) = \mathrm{tr}(AA^{\mathrm{T}}) = \|A\|_F^2.$$

□

End of “Chapter 1: Preliminaries”.