Linhui Xie
August 1, 2022

1 Mathematical preliminaries

1.1 Real vectors and matrices

1.1.1 Vectors

- $\alpha, \beta, \gamma, ...$: scalars.
- $a_1, ..., a_n, x_1, ..., x_n, y_1, ..., y_n$: real numbers, components of a vector, element of a set.
- $\mathbb{R}$: set of real numbers.
- $\mathbb{R}^n$: set of real column vectors.

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad x_i \in \mathbb{R}$$

- A n-dimensional column vector and row vector,

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}, \quad \boldsymbol{a}^\top = [a_1, a_2, \ldots, a_n].$$

- Properties,

$$\begin{aligned} \boldsymbol{a} + \boldsymbol{b} &= \boldsymbol{b} + \boldsymbol{a} & \alpha(\boldsymbol{a} + \boldsymbol{b}) &= \alpha\boldsymbol{a} + \alpha\boldsymbol{b} \\ (\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} &= \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c}) & \alpha(\beta\boldsymbol{a}) &= (\alpha\beta)\boldsymbol{a} \\ \boldsymbol{0} &= [0, 0, \ldots, 0]^\top & 1\boldsymbol{a} &= \boldsymbol{a} \\ \boldsymbol{a} + \boldsymbol{0} &= \boldsymbol{0} + \boldsymbol{a} = \boldsymbol{a} & \alpha\boldsymbol{0} &= \boldsymbol{0} = 0\boldsymbol{a} \end{aligned}$$

1.1.2 Matrices

- $\mathbb{R}^{m \times n}$: set of $m \times n$ real matrices,

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}.$$

- $\mathbb{R}^{n \times 1}$ and $\mathbb{R}^n$ as equivalent.
- $\boldsymbol{A}^\top$: transpose of $\boldsymbol{A}$.

1.2 Functions

- Set $X : x \in X$.
- Function $f : X \to Y$.
- f takes values in X and gives values in Y.
  - $f(x)$ is the value of f at x, where $x \in X$.
- A symbol $:=$ denotes arithmetic assignment; $x := y$, means"x becomes y".
- A symbol $\triangleq$ means "equals by definition".
- Example: $f : \mathbb{R}^3 \to \mathbb{R}$,

$$f(\boldsymbol{x}) = \frac{x_1^2 + 2\log(x_2 x_3) + x_1 x_2 x_3}{x_2}.$$

1.3 Linear independence

- A set of vectors $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ is said to be linearly independent if

$$\alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k = \mathbf{0},$$

  implies that all the scalar coefficients $\alpha_i, i = 1, \ldots, k$, are equal to zero.

- A vector $\boldsymbol{a}$ is said to be a linear combination of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ if there are scalars $\alpha_1, \ldots, \alpha_k$ such that

$$\boldsymbol{a} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k.$$

- $\mathcal{V}$: a subspace of $\mathbb{R}^n$, if $\mathcal{V}$ is closed for vector addition and scalar multiplication.

▶ *Proposition* 2.1 A set of vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is linearly dependent if and only if one of the vectors from the set is a linear combination of the remaining vectors.

- The set of all linear combinations of $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ is called the span of the vectors,

$$\operatorname{span}[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k] = \left\{ \sum_{i=1}^{k} \alpha_i \boldsymbol{a}_i : \alpha_1, \ldots, \alpha_k \in \mathbb{R} \right\}.$$

- Any set of linearly independent vectors $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\} \subset \mathcal{V}$, is a basis of the subspace $\mathcal{V}$, if $\mathcal{V} = \operatorname{span}[\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k]$.

▶ *Proposition* 2.2 If $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k\}$ is a basis of $\mathcal{V}$, then any vector $\boldsymbol{a}$ of $\mathcal{V}$ can be represented uniquely as

$$\boldsymbol{a} = \alpha_1 \boldsymbol{a}_1 + \alpha_2 \boldsymbol{a}_2 + \cdots + \alpha_k \boldsymbol{a}_k,$$

where $\alpha_i \in \mathbb{R}, i = 1, 2, \ldots, k$.

1.4 Rank of a matrix

- The maximal number of linearly independent columns of $\boldsymbol{A}$ is called the rank of the matrix $\boldsymbol{A}$, denoted rank $\boldsymbol{A}$.

▶ *Proposition* 2.3 The rank of a matrix $\boldsymbol{A}$ is invariant under the following operations:

  - $\operatorname{rank}[\boldsymbol{a}_1, \ldots, \alpha \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n] = \operatorname{rank}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n], \alpha \neq 0.$
  - $\operatorname{rank}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_l, \ldots, \boldsymbol{a}_n] = \operatorname{rank}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_l, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n].$
  - $\operatorname{rank}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k + (\alpha_1 \boldsymbol{a}_1 + \ldots + \alpha_n \boldsymbol{a}_n), \ldots, \boldsymbol{a}_n] = \operatorname{rank}[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n].$

- The determinant of the square matrix $\mathbf{A}$, denoted $\det \mathbf{A}$ or $|\mathbf{A}|$. The determinant of a square matrix is a function of its columns,

  1. The determinant of the matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_n]$ is a linear function of each column; that is,

$$\begin{aligned} &\det\left[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k-1}, \alpha \boldsymbol{a}_k^{(1)} + \beta \boldsymbol{a}_k^{(2)}, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n\right] \\ =\ &\alpha \det\left[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_k^{(1)}, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n\right] + \beta \det\left[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{k-1}, \boldsymbol{a}_k^{(2)}, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n\right] \end{aligned}$$

  for each $\alpha, \beta \in \mathbb{R}, \boldsymbol{a}_k^{(1)}, \boldsymbol{a}_k^{(2)} \in \mathbb{R}^n$.

  2. If for some k we have $\boldsymbol{a}_k = \boldsymbol{a}_{k+1}$, then

$$\det \boldsymbol{A} = \det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_{k+1}, \ldots, \boldsymbol{a}_n] = \det[\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k, \boldsymbol{a}_k, \ldots, \boldsymbol{a}_n] = 0$$

  3. Let

$$\boldsymbol{I}_n = [\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

  where $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ is the natural basis for $\mathbb{R}^n$. Then

$$\det \boldsymbol{I}_n = 1$$

- A pth-order minor of an $m \times n$ matrix $\boldsymbol{A}$, with $p \leq \min\{m, n\}$, is the determinant of a $p \times p$ matrix obtained from $\boldsymbol{A}$ by deleting $m - p$ rows and $n - p$ columns.

▶ *Proposition* 2.4 If an $m \times n$ matrix $A (m \geq n)$ has a nonzero nth-order minor, then the columns of A are linearly independent; that is, $\operatorname{rank}(A) = n$.

- The rank of a matrix is equal to the highest order of its nonzero minor(s).

- A square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is nonsingular or invertible if rank $\boldsymbol{A} = n$ (full rank).

- A matrix is nonsingular if and only if its determinant is nonzero.

1.5 Linear equations

♠ *Theorem* 2.1 The system of equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has a solution if and only if

$$\operatorname{rank} \boldsymbol{A} = \operatorname{rank}[\boldsymbol{A}, \boldsymbol{b}].$$

♠ *Theorem* 2.2 Consider the equation $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$, where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and rank $\boldsymbol{A} = m$. A solution to $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ can be obtained by assigning arbitrary values for $n - m$ variables and solving for the remaining ones.

1.6 Inner product and norm

1.6.1 Real domain

- Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
- Define the Euclidean inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^\top \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i.$$

- The inner product is a real-valued function $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$,
  - $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0, \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \mathbf{0}$.
  - $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$.
  - $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.
  - $\langle r\boldsymbol{x}, \boldsymbol{y} \rangle = r\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for every $r \in \mathbb{R}$.

  Example,

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{x} \rangle = (\boldsymbol{A}\boldsymbol{x})^\top \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{x} = \boldsymbol{x}^\top \left(\boldsymbol{A}^\top \boldsymbol{x}\right) = \left\langle \boldsymbol{x}, \boldsymbol{A}^\top \boldsymbol{x} \right\rangle.$$

- Define the Euclidean norm of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

- The Euclidean norm properties,
  - $\|\boldsymbol{x}\| \geq 0, \|\boldsymbol{x}\| = 0$ if and only if $\boldsymbol{x} = 0$.
  - $\|\boldsymbol{x}\| = |r|\|\boldsymbol{x}\|, r \in \mathbb{R}$.
  - $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

♠ *Theorem* 2.3 Cauchy-Schwarz Inequality. For any two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$, the Cauchy-Schwarz inequality holds,

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|.$$

- The Euclidean norm is often referred to as the 2-norm, and denoted $\|\boldsymbol{x}\|_2$. The norms above are special cases of the **$\boldsymbol{p}$-norm**, given by

$$\|\boldsymbol{x}\|_p = \begin{cases} \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p} & \text{if } 1 \leq p < \infty \\ \max\left\{|x_1|, \ldots, |x_n|\right\} & \text{if } p = \infty \end{cases}$$

- $\boldsymbol{x}$ and $\boldsymbol{y}$ are orthogonal if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.
- A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is continuous at x if for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\|\boldsymbol{y} - \boldsymbol{x}\| < \delta \Rightarrow \|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\| < \varepsilon.$$

1.6.2 Complex domain

- An complex inner product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ to be $\sum_{i=1}^{n} x_i \bar{y}_i$ in complex vector space $\mathbb{C}^n$, where the bar denotes complex conjugation.
- The inner product on $\mathbb{C}^n$ is a complex-valued function $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$,
  - $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$ if and only if $\boldsymbol{x} = \boldsymbol{0}$.
  - $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$.
  - $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.
  - $\langle r\boldsymbol{x}, \boldsymbol{y} \rangle = r\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, where $r \in \mathbb{C}$.
- Deduction from above properties,

$$\langle \boldsymbol{x}, r_1\boldsymbol{y} + r_2\boldsymbol{z} \rangle = \bar{r}_1\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \bar{r}_2\langle \boldsymbol{x}, \boldsymbol{z} \rangle.$$

- Define the Complex norm of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle} = \sqrt{\boldsymbol{x}^\top \boldsymbol{x}} = \sqrt{\sum_{i=1}^{n} x_i \bar{x}_i}.$$

1.7 Linear transformations

- A function $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ is called a linear transformation if
  - $\mathcal{L}(a\boldsymbol{x}) = a\mathcal{L}(\boldsymbol{x})$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and $a \in \mathbb{R}$;
  - $\mathcal{L}(\boldsymbol{x} + \boldsymbol{y}) = \mathcal{L}(\boldsymbol{x}) + \mathcal{L}(\boldsymbol{y})$ for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$.
- Let $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ and $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n\}$ be two bases for $\mathbb{R}^n$. Define the transformation matrix $\boldsymbol{T}$ from $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ to $\{\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n\}$

$$\begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix} = \begin{bmatrix} \boldsymbol{e}'_1 & \boldsymbol{e}'_2 & \cdots & \boldsymbol{e}'_n \end{bmatrix} \boldsymbol{T}.$$

- Given a vector $\boldsymbol{v}$, $\boldsymbol{x}$ is the coordinates of the vector with respect to a base $\boldsymbol{B} = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ and $\boldsymbol{x}'$ be the coordinates of the same vector with respect to a base $\boldsymbol{B}' = \{\boldsymbol{e}'_1, \boldsymbol{e}'_2, \ldots, \boldsymbol{e}'_n\}$.

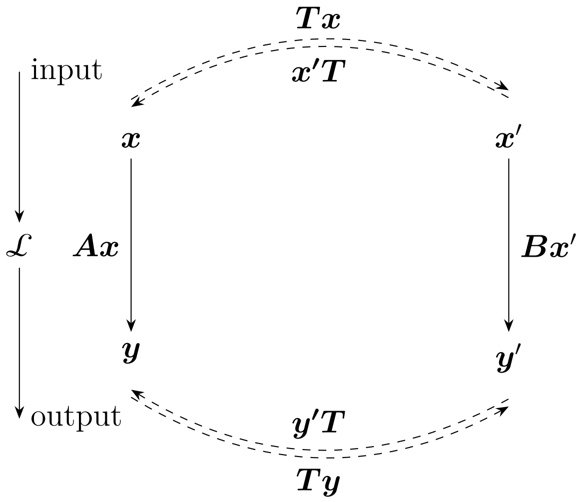
$$[\boldsymbol{v}]_B = x_1\boldsymbol{e}_1 + \cdots + x_n\boldsymbol{e}_n = [\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n]\,\boldsymbol{x}$$
$$[\boldsymbol{v}]_{B'} = x'_1\boldsymbol{e}'_1 + \cdots + x'_n\boldsymbol{e}'_n = [\boldsymbol{e}'_1, \ldots, \boldsymbol{e}'_n]\,\boldsymbol{x}'$$
$$\boldsymbol{x}' = \begin{bmatrix} \boldsymbol{e}'_1 & \boldsymbol{e}'_2 & \cdots & \boldsymbol{e}'_n \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \cdots & \boldsymbol{e}_n \end{bmatrix} \boldsymbol{x} = \boldsymbol{T}\boldsymbol{x}$$

- Example, let $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ and $\boldsymbol{y}' = \boldsymbol{A}'\boldsymbol{x}'$. Therefore,

$$\boldsymbol{y}' = \boldsymbol{T}\boldsymbol{y} = \boldsymbol{T}\boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}' = \boldsymbol{A}'\boldsymbol{x}' = \boldsymbol{A}'\boldsymbol{T}\boldsymbol{x}$$

  and hence $\boldsymbol{T}\boldsymbol{A} = \boldsymbol{A}'\boldsymbol{T}$, or $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{A}'\boldsymbol{T}$.

- Two $n \times n$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ are similar if there exists a nonsingular matrix $\boldsymbol{T}$ such that $\boldsymbol{A} = \boldsymbol{T}^{-1}\boldsymbol{B}\boldsymbol{T}$.



1.8 Eigenvalues and eigenvectors

- Let $\boldsymbol{A}$ be an $n \times n$ square matrix.
- A scalar λ (possibly complex) and a nonzero vector $\boldsymbol{v}$ satisfying the equation $\boldsymbol{A}\boldsymbol{v} = \lambda\boldsymbol{v}$ are said to be, respectively, an eigenvalue and eigenvector of $\boldsymbol{A}$.
- λ is an eigenvalue of $\boldsymbol{A}$ if and only if $\lambda\boldsymbol{I} - \boldsymbol{A}$ is singular (i.e., $\det[\lambda\boldsymbol{I} - \boldsymbol{A}] = 0$).
- $\det[\lambda\boldsymbol{I} - \boldsymbol{A}]$ is called the characteristic polynomial of $\boldsymbol{A}$,

$$\det[\lambda\boldsymbol{I} - \boldsymbol{A}] = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0.$$

♠ *Theorem* 3.1 Suppose that the characteristic equation $\det[\lambda\boldsymbol{I} - \boldsymbol{A}] = 0$ has n distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then, there exist n linearly independent vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n$ such that

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i\boldsymbol{v}_i, \quad i = 1, 2, \ldots, n.$$

♠ *Theorem* 3.2 All eigenvalues of a real symmetric matrix are real.

♠ *Theorem* 3.3 If real $n \times n$ matrix $\boldsymbol{A}$ is symmetric, then a set of its eigenvectors forms an orthogonal basis for $\mathbb{R}^n$.

1.9 Orthogonal projections

- If $\mathcal{V}$ is a subspace of $\mathbb{R}^n$, then the orthogonal complement of $\mathcal{V}$, denoted by $\mathcal{V}^\perp$, consists of all vectors that are orthogonal to every vector in $\mathcal{V}$,

$$\mathcal{V}^\perp = \left\{\boldsymbol{x} : \boldsymbol{v}^\top\boldsymbol{x} = 0 \text{ for all } \boldsymbol{v} \in \mathcal{V}\right\}.$$

- $\mathcal{V}$ and $\mathcal{V}^\perp$ span $\mathbb{R}^n$ in the sense that every vector $\boldsymbol{x} \in \mathbb{R}^n$ can be represented uniquely as $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$, where $\boldsymbol{x}_1 \in \mathcal{V}$ and $\boldsymbol{x}_2 \in \mathcal{V}^\perp$.

- $\boldsymbol{x} = \boldsymbol{x}_1 + \boldsymbol{x}_2$ is the orthogonal decomposition of $\boldsymbol{x}$ with respect to $\mathcal{V}$. $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are orthogonal projections of $\boldsymbol{x}$ onto the subspaces $\mathcal{V}$ and $\mathcal{V}^{\perp}$, respectively.

- We write $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^{\perp}$ and say that $\mathbb{R}^n$ is a direct sum of $\mathcal{V}$ and $\mathcal{V}^{\perp}$.

- A linear transformation $\boldsymbol{P}$ is an orthogonal projector onto $\mathcal{V}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$ we have $\boldsymbol{P}\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \in \mathcal{V}^{\perp}$.

- Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the range, or image of $\boldsymbol{A}$ is

$$\mathcal{R}(\boldsymbol{A}) \triangleq \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\}. \quad \text{That's column space.}$$

  The nullspace, or kernel of $\boldsymbol{A}$ is

$$\mathcal{N}(\boldsymbol{A}) \triangleq \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\}.$$

  $\mathcal{R}(\boldsymbol{A})$ and $\mathcal{N}(\boldsymbol{A})$ are subspaces.

♠ *Theorem* 3.4 $\mathcal{R}(\boldsymbol{A})^{\perp} = \mathcal{N}\left(\boldsymbol{A}^{\top}\right)$ and $\mathcal{N}(\boldsymbol{A})^{\perp} = \mathcal{R}\left(\boldsymbol{A}^{\top}\right)$ (That's row space. Together, four fundamental spaces in Linear Algebra.)

- If $\boldsymbol{P}$ is an orthogonal projector onto $\mathcal{V}$, then $\boldsymbol{P}\boldsymbol{x} = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{V}$, and $\mathcal{R}(\boldsymbol{P}) = \mathcal{V}$.

♠ *Theorem* 03.05: A matrix $\boldsymbol{P}$ is an orthogonal projector if and only if $\boldsymbol{P}^2 = \boldsymbol{P} = \boldsymbol{P}^T$.

1.10 Symmetric matrices

- $\boldsymbol{Q}$ is symmetric if $\boldsymbol{Q} = \boldsymbol{Q}^{\top}$.

- A symmetric matrix $\boldsymbol{Q}$ is said to be positive definite if $\boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} > 0$ for all nonzero vectors $\boldsymbol{x}$.

- It is positive semi-definite if $\boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \geq 0$ for all $\boldsymbol{x}$.

- Similarly, negative definite and negative semi-definite, if $\boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} < 0$ for all nonzero vectors $\boldsymbol{x}$, or $\boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \leq 0$ for all $\boldsymbol{x}$, respectively.

- For an $n \times n$ symmetric real matrix $\boldsymbol{Q}$,

$$\begin{aligned}
\boldsymbol{Q} \text{ positive-definite} &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} > 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\} \\
\boldsymbol{Q} \text{ positive semi-definite} &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \geq 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^n \\
\boldsymbol{Q} \text{ negative-definite} &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} < 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^n \backslash \{\boldsymbol{0}\} \\
\boldsymbol{Q} \text{ negative semi-definite} &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \leq 0 \text{ for all } \boldsymbol{x} \in \mathbb{R}^n
\end{aligned}$$

- For an $n \times n$ Hermitian complex matrix $\boldsymbol{Q}$,

$$\begin{aligned}
\boldsymbol{Q} \text{ positive-definite }, \boldsymbol{Q} \succ 0 &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} > 0 \text{ for all } \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\} \\
\boldsymbol{Q} \text{ positive semi-definite }, \boldsymbol{Q} \succeq 0 &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \geq 0 \text{ for all } \boldsymbol{x} \in \mathbb{C}^n \\
\boldsymbol{Q} \text{ negative-definite }, \boldsymbol{Q} \prec 0 &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} < 0 \text{ for all } \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\} \\
\boldsymbol{Q} \text{ negative semi-definite }, \boldsymbol{Q} \preceq 0 &\iff \boldsymbol{x}^{\top}\boldsymbol{Q}\boldsymbol{x} \leq 0 \text{ for all } \boldsymbol{x} \in \mathbb{C}^n
\end{aligned}$$

1.11 Quadratic functions

- $f : \mathbb{R}^n \to \mathbb{R}$ is a quadratic function if

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} + \boldsymbol{b}^T \boldsymbol{x} + c,$$

where $\boldsymbol{Q}$ is symmetric.

- If the matrix $\boldsymbol{Q}$ is not symmetric, we can always replace it with the symmetric

$$\boldsymbol{Q}_0 = \boldsymbol{Q}_0^T = \frac{1}{2}\left(\boldsymbol{Q} + \boldsymbol{Q}^T\right)$$
$$\boldsymbol{x}^T \boldsymbol{Q} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{Q}_0 \boldsymbol{x} = \boldsymbol{x}^T \left(\frac{1}{2}\boldsymbol{Q} + \frac{1}{2}\boldsymbol{Q}^T\right) \boldsymbol{x}$$

- The leading principal minors of matrix $\boldsymbol{Q}$ are,

$$\Delta_1 = q_{11}, \quad \Delta_2 = \det \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}, \quad \ldots$$

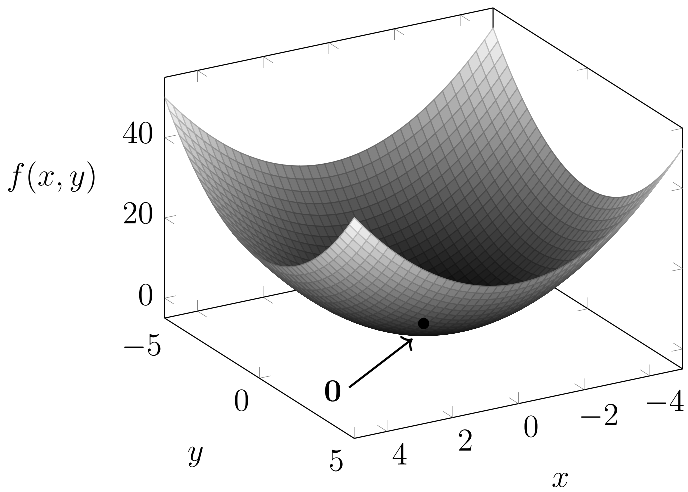
♠ *Theorem* 3.6 Sylvester's Criterion. A quadratic form ($\boldsymbol{Q}$ is symmetric) $\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}, \boldsymbol{Q} = \boldsymbol{Q}^\top$, is positive definite if and only if the leading principal minors of $\boldsymbol{Q}$ are positive.

- A necessary condition for a real quadratic form to be positive semi-definite is that the leading principal minors be nonnegative. However, it is not a sufficient condition.

- A real quadratic form is positive semi-definite if and only if all principal minors are nonnegative.

♠ *Theorem* 3.7 A symmetric matrix $\boldsymbol{Q}$ is positive definite (or positive semidefinite) if and only if all eigenvalues of $\boldsymbol{Q}$ are positive (or nonnegative).

- If $\boldsymbol{Q}$ is positive definite, then f is a parabolic "bowl".



- Quadratics simplify optimization, offering a clear structure for minimum or maximum solutions.

- Near optimal points, objective functions often resemble quadratics.

- Algorithms are more transparent when tested on quadratics.

- Insights from quadratic algorithm analysis extend to broader algorithmic applications.

1.12 Matrix norm

- The norm of a matrix $\boldsymbol{A}$, denoted by $\|\boldsymbol{A}\|$, satisfies

  1. $\|\boldsymbol{A}\| > 0$ if $\boldsymbol{A} \neq \boldsymbol{O}$, and $\|\boldsymbol{O}\| = 0$, where $\boldsymbol{O}$ is a matrix with all entries equal to zero.
  2. $\|c\boldsymbol{A}\| = |c|\|\boldsymbol{A}\|$, for any $c \in \mathbb{R}$.
  3. $\|\boldsymbol{A} + \boldsymbol{B}\| \leq \|\boldsymbol{A}\| + \|\boldsymbol{B}\ |$.

- For $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, an example of a matrix norm is the Frobenius norm, defined as

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

- Note that the Frobenius norm is equivalent to the Euclidean norm on $\mathbb{R}^{m\times n}$.

- For this course, only consider matrix norms satisfying the addition condition:

  4. $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\|\|\boldsymbol{B}\|$.

- The Frobenius norm satisfies condition 4, $\|\boldsymbol{AB}\|_F \leq \|\boldsymbol{A}\|_F\|\boldsymbol{B}\|_F$.

- Let $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ be vector norms on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. The matrix norm is induced by, or is compatible with, the given vector norms if for any matrix $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and any vector $\boldsymbol{x} \in \mathbb{R}^n$, the following inequality is satisfied:

$$\|\boldsymbol{Ax}\|_{(m)} \leq \|\boldsymbol{A}\|\|\boldsymbol{x}\|_{(n)}.$$

- An induced matrix norm as

$$\|\boldsymbol{A}\| = \max_{\|\boldsymbol{x}\|_{(n)}=1}\|\boldsymbol{Ax}\|_{(m)}.$$

- For each matrix $\boldsymbol{A}$ the maximum $\max_{\|\boldsymbol{x}\|=1}\|\boldsymbol{Ax}\|$ is attainable; that is, a vector $\boldsymbol{x}_0$ exists such that $\|\boldsymbol{x}_0\| = 1$ and $\|\boldsymbol{Ax}_0\| = \|\boldsymbol{A}\|$.

♠ *Theorem* 3.8: Let

$$\|\boldsymbol{x}\| = \left(\sum_{k=1}^{n}|x_k|^2\right)^{1/2} = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}$$

the matrix norm induced by this vector norm is

$$\|\boldsymbol{A}\| = \sqrt{\lambda_1}$$

where λ_1 is the largest eigenvalue of the matrix $\boldsymbol{A}^\top\boldsymbol{A}$.

- Rayleigh's Inequality: If an $n\times n$ matrix $\boldsymbol{P}$ is real symmetric positive definite, then

$$\lambda_{\min}(\boldsymbol{P})\|\boldsymbol{x}\|^2 \leq \boldsymbol{x}^T\boldsymbol{Px} \leq \lambda_{\max}(\boldsymbol{P})\|\boldsymbol{x}\|^2$$

where $\lambda_{\min}(\boldsymbol{P})$ denotes the smallest eigenvalue of $\boldsymbol{P}$, and $\lambda_{\max}(\boldsymbol{P})$ denotes the largest eigenvalue of $\boldsymbol{P}$.

[Ref]:

Edwin K.P. Chong, Stanislaw H. Żak, “PART I MATHEMATICAL REVIEW” in “An introduction to optimization”, 4th Edition, John Wiley and Sons, Inc. 2013.