

1 Mathematical preliminaries

1.1 Real vectors and matrices

1.1.1 Vectors

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

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

- A n -dimensional column vector and row vector,

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

- Properties,

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

1.1.2 Matrices

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

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

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

1.2 Functions

- Set $X : x \in X$.
- Function $f : X \rightarrow Y$.
- 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: $f : \mathbb{R}^3 \rightarrow \mathbb{R}$,

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

1.3 Linear independence

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

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

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

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

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \dots + \alpha_k \mathbf{a}_k.$$

- \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 $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ 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 $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k$ is called the span of the vectors,

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

- Any set of linearly independent vectors $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\} \subset \mathcal{V}$, is a basis of the subspace \mathcal{V} , if $\mathcal{V} = \text{span}[\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k]$.
- **PROPOSITION 2.2** If $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_k\}$ is a basis of \mathcal{V} , then any vector \mathbf{a} of \mathcal{V} can be represented uniquely as

$$\mathbf{a} = \alpha_1 \mathbf{a}_1 + \alpha_2 \mathbf{a}_2 + \dots + \alpha_k \mathbf{a}_k,$$

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

1.4 Rank of a matrix

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

► PROPOSITION 2.3 The rank of a matrix \mathbf{A} is invariant under the following operations:

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

- 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 $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n]$ is a linear function of each column; that is,

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

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

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

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

.

3. Let

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

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

$$\det \mathbf{I}_n = 1.$$

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

► PROPOSITION 2.4 If an $m \times n$ matrix \mathbf{A} ($m \geq n$) has a nonzero n th-order minor, then the columns of \mathbf{A} are linearly independent; that is, $\text{rank}(\mathbf{A}) = n$.

- The rank of a matrix is equal to the highest order of its nonzero minor(s).
- A square matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is nonsingular or invertible if $\text{rank } \mathbf{A} = n$ (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 $\mathbf{Ax} = \mathbf{b}$ has a solution if and only if

$$\text{rank } \mathbf{A} = \text{rank}[\mathbf{A}, \mathbf{b}].$$

♠ **THEOREM 2.2** Consider the equation $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\text{rank } \mathbf{A} = m$. A solution to $\mathbf{Ax} = \mathbf{b}$ 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
- Define the Euclidean inner product of \mathbf{x} and \mathbf{y} :

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

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

Example,

$$\langle \mathbf{Ax}, \mathbf{x} \rangle = (\mathbf{Ax})^\top \mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} = \mathbf{x}^\top (\mathbf{A}^\top \mathbf{x}) = \langle \mathbf{x}, \mathbf{A}^\top \mathbf{x} \rangle.$$

- Define the Euclidean norm of \mathbf{x} :

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

- The Euclidean norm properties,
 - $\|\mathbf{x}\| \geq 0$, $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
 - $\|\mathbf{x}\| = |r|\|\mathbf{x}\|$, $r \in \mathbb{R}$.
 - $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$.
- ♠ **THEOREM 2.3** Cauchy-Schwarz Inequality. For any two vectors \mathbf{x} and \mathbf{y} in \mathbb{R}^n , the Cauchy-Schwarz inequality holds,

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

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

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

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

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

1.6.2 Complex domain

- An complex inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ 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 $\langle \cdot, \cdot \rangle : \mathbb{C}^n \times \mathbb{C}^n \rightarrow \mathbb{C}$,
 - $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.
 - $\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle}$.
 - $\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle$.
 - $\langle r\mathbf{x}, \mathbf{y} \rangle = r\langle \mathbf{x}, \mathbf{y} \rangle$, where $r \in \mathbb{C}$.
- Deduction from above properties,

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

- Define the Complex norm of \mathbf{x} :

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

1.7 Linear transformations

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

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

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

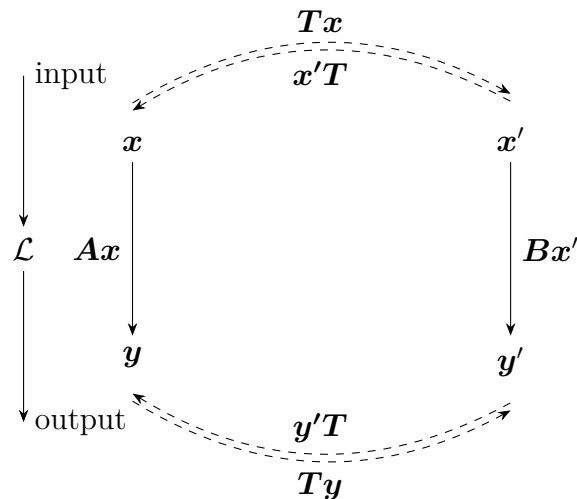
$$\begin{aligned} [\mathbf{v}]_B &= x_1 \mathbf{e}_1 + \dots + x_n \mathbf{e}_n = [\mathbf{e}_1, \dots, \mathbf{e}_n] \mathbf{x}, \\ [\mathbf{v}]_{B'} &= x'_1 \mathbf{e}'_1 + \dots + x'_n \mathbf{e}'_n = [\mathbf{e}'_1, \dots, \mathbf{e}'_n] \mathbf{x}', \\ \mathbf{x}' &= \begin{bmatrix} \mathbf{e}'_1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

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

$$\mathbf{y}' = \mathbf{T} \mathbf{y} = \mathbf{T} \mathbf{A} \mathbf{x}, \mathbf{y}' = \mathbf{A}' \mathbf{x}' = \mathbf{A}' \mathbf{T} \mathbf{x}.$$

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

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



1.8 Eigenvalues and eigenvectors

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

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

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

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

♠ THEOREM 3.2 All eigenvalues of a real symmetric matrix are real.

♠ THEOREM 3.3 If real $n \times n$ matrix \mathbf{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} ,

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

- \mathcal{V} and \mathcal{V}^\perp span \mathbb{R}^n in the sense that every vector $\mathbf{x} \in \mathbb{R}^n$ can be represented uniquely as $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$, where $\mathbf{x}_1 \in \mathcal{V}$ and $\mathbf{x}_2 \in \mathcal{V}^\perp$.
- $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$ is the orthogonal decomposition of \mathbf{x} with respect to \mathcal{V} . \mathbf{x}_1 and \mathbf{x}_2 are orthogonal projections of \mathbf{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 \mathbf{P} is an orthogonal projector onto \mathcal{V} if for all $\mathbf{x} \in \mathbb{R}^n$ we have $\mathbf{P}\mathbf{x} \in \mathcal{V}$ and $\mathbf{x} - \mathbf{P}\mathbf{x} \in \mathcal{V}^\perp$.
- Let $\mathbf{A} \in \mathbb{R}^{m \times n}$, the range, or image of \mathbf{A} is

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

The nullspace, or kernel of \mathbf{A} is

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

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

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

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

♠ THEOREM 03.05: A matrix \mathbf{P} is an orthogonal projector if and only if $\mathbf{P}^2 = \mathbf{P} = \mathbf{P}^\top$.

1.10 Symmetric matrices

- \mathbf{Q} is symmetric if $\mathbf{Q} = \mathbf{Q}^\top$.
- A symmetric matrix \mathbf{Q} is said to be positive definite if $\mathbf{x}^\top \mathbf{Q}\mathbf{x} > 0$ for **all nonzero** vectors \mathbf{x} .
- It is positive semi-definite if $\mathbf{x}^\top \mathbf{Q}\mathbf{x} \geq 0$ for **all** \mathbf{x} .
- Similarly, negative definite and negative semi-definite, if $\mathbf{x}^\top \mathbf{Q}\mathbf{x} < 0$ for **all nonzero** vectors \mathbf{x} , or $\mathbf{x}^\top \mathbf{Q}\mathbf{x} \leq 0$ for **all** \mathbf{x} , respectively.

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

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

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

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

1.11 Quadratic functions

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

$$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{Q} \mathbf{x} + \mathbf{b}^\top \mathbf{x} + c,$$

where \mathbf{Q} is **symmetric**.

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

$$\begin{aligned}\mathbf{Q}_0 = \mathbf{Q}_0^\top &= \frac{1}{2} (\mathbf{Q} + \mathbf{Q}^\top) \\ \mathbf{x}^\top \mathbf{Q} \mathbf{x} &= \mathbf{x}^\top \mathbf{Q}_0 \mathbf{x} = \mathbf{x}^\top \left(\frac{1}{2} \mathbf{Q} + \frac{1}{2} \mathbf{Q}^\top \right) \mathbf{x}\end{aligned}$$

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

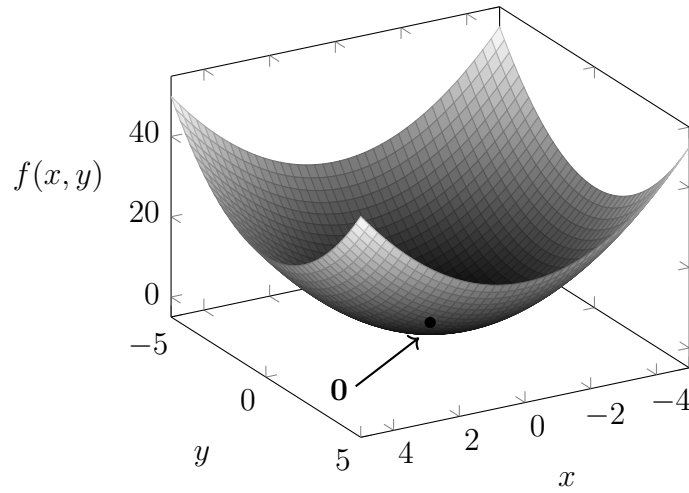
$$\Delta_1 = q_{11}, \quad \Delta_2 = \det \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}, \quad \Delta_3 = \det \begin{bmatrix} q_{11} & q_{12} & q_{13} \\ q_{21} & q_{22} & q_{23} \\ q_{31} & q_{32} & q_{33} \end{bmatrix}, \quad \dots$$

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

- If \mathbf{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 \mathbf{A} , denoted by $\|\mathbf{A}\|$, satisfies
 1. $\|\mathbf{A}\| > 0$ if $\mathbf{A} \neq \mathbf{O}$, and $\|\mathbf{O}\| = 0$, where \mathbf{O} is a matrix with all entries equal to zero.
 2. $\|c\mathbf{A}\| = |c|\|\mathbf{A}\|$, for any $c \in \mathbb{R}$.
 3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$.
- For $\mathbf{A} \in \mathbb{R}^{m \times n}$, an example of a matrix norm is the Frobenius norm, defined as

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

- 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. $\|\mathbf{AB}\| \leq \|\mathbf{A}\|\|\mathbf{B}\|$.
- The Frobenius norm satisfies condition 4, $\|\mathbf{AB}\|_F \leq \|\mathbf{A}\|_F \|\mathbf{B}\|_F$.
- 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 $\mathbf{A} \in \mathbb{R}^{m \times n}$ and any vector $\mathbf{x} \in \mathbb{R}^n$, the following inequality is satisfied:

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

- An induced matrix norm as

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

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

♠ **THEOREM 3.8:** Let

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

The matrix norm induced by this vector norm is

$$\|\mathbf{A}\| = \sqrt{\lambda_1},$$

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

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

$$\lambda_{\min}(\mathbf{P})\|\mathbf{x}\|^2 \leq \mathbf{x}^\top \mathbf{P} \mathbf{x} \leq \lambda_{\max}(\mathbf{P})\|\mathbf{x}\|^2,$$

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