EEEN60151 – Machine Learning & Optimization Techniques

Week One

Dr Khairi A Hamdi

26^{th} September, 2024

Review of Essential Mathematical Optimisation Problem

- Vectors and vector space
- Matrices
- Quadratic Forms
- Differentiation of functions of several variables
    - Gradient vectors
    - Hessian matrices

References

Recommended Readings for preliminaries and Mathematical Background

(no need to study in details)

(will not be included in the final exam)

1. Chong, Lu and Zak, An Introduction to Optimization, **chapters 2 - 5**
2. Grippo and Sciandrone, Introduction to Methods for Nonlinear Optimization, **Chapters 27, 28**
3. Calafiore and El Ghaoui, *Optimization Models*, Cambridge, **Chapter 2**

Vectors Space

- An $n-vector$ or vector in $\mathbb{R}^n$ is an ordered $n-tuple$

$$\boldsymbol{x} = (x_1, x_2, \dots, x_n)$$

- It is often convenient to think of an n-vector as a column vector:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

- The number $x_1, x_2, \dots, x_n \in \mathbb{R}$ are components (elements, or entries) of the vector $\boldsymbol{x}$,
- $\mathbb{R}$ is the set of all real numbers, i.e., $\infty < x_i < +\infty$ for $i = 1, 2, \dots, n.$
- We say that $\boldsymbol{x} \in \mathbb{R}^n$ where $\mathbb{R}^n$ is the set of all column n-vectors with real components. .

Vectors Space (contd.)

The transpose of a column vector x is a row n- vector, denoted by x^T

$$x^T = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}^T = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}$$

$$x = (x_1, x_2, \dots, x_n) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^T$$

Addition (or subtraction) of any two vectors **of same size**

$$x + y = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}$$

Vectors Space (contd.)

Multiplication by a real number $\lambda \in R$

$$\lambda \mathbf{x} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}$$

Linearly Dependent and Independent Vectors

A set of vectors $\{v_1, v_2, \dots, v_q\}$ is called **linearly dependent** if there exist scalars $\alpha_1, \alpha_2, \dots, \alpha_q$ (not ALL zeros) such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \dots, \alpha_q v_q = 0$$

(i,e., there is at least one scalar $\alpha_i \neq 0$.)
Alternatively, a set of vectors is said to be **linearly independent**.

Example

Check if the following sets of vectors are linearly independent?

1. $x_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 0 \\ 5 \\ 6 \end{pmatrix}$
2. $x_1 = \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$, $x_2 = \begin{pmatrix} 0 \\ 5 \\ 6 \end{pmatrix}$ and $x_3 = \begin{pmatrix} 2 \\ 9 \\ 6 \end{pmatrix}$

$$a_1 \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} + a_2 \begin{pmatrix} 0 \\ 5 \\ 6 \end{pmatrix} = 0 \Rightarrow \begin{array}{cc} a_1 + 0 = 0 & \Rightarrow a_1 = 0 \\ 2a_1 + 5a_2 = 0 & \\ 0 + 6a_2 = 0 & \Rightarrow a_2 = 0 \end{array}$$

Therefore x_1 and x_2 are linearly independent

It can be verified that $2x_1 + x_2 = x_1$. Therefore the set $\{x_1, x_2, x_1\}$ is linearly dependent.

The Subspace

- A subset $\mathcal{S}$ of $\mathbb{R}^n$ is called a **subspace** if $\mathcal{S}$ is **closed** under vector addition and scalar multiplication. Note that every subspace contains the zero vector $\mathbf{0}$.
- **Linear Span (Linear Hull):** For a given set of vectors $\mathcal{S} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\} \in \mathbb{R}^n$, their linear span is the set of all their linear combinations

$$\operatorname{span}(\mathcal{S}) = \left\{ \sum_{i=1}^{k} \alpha_i \mathbf{x}_i \colon \alpha_1, \alpha_2, \dots, \alpha_k \in \mathbb{R} \right\}$$

- **Affine Span (Affine Hull)**

$$\operatorname{span}(\mathcal{S}) = \left\{ \alpha_0 + \sum_{i=1}^{k} \alpha_i \mathbf{x}_i \colon \alpha_0, \alpha_1, \dots, \alpha_k \in \mathbb{R} \right\}$$

- **Basis and Dimensions:** A set of vectors $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \in \mathcal{S}$ is a **basis** for $\mathcal{S}$ if it spans $\mathcal{S}$ and is linearly independent.

The Subspace (contd.)

- The **dimension** is the number of vectors in a basis.
- There are infinitely many bases for $\mathbb{R}^n$. The **standard basis** for $\mathbb{R}^n$ is

$$\mathbf{e}_1,\ \mathbf{e}_2, \dots\ , \mathbf{e}_n$$

where $\mathbf{e}_i$ is a vector in $\mathbb{R}^n$ with all zero entries except for the ith entry

Example

The standard basis for $\mathbb{R}^3$ is given by

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Inner (Dot) Product of Vectors

The inner product of two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ is a function (mapping) $\mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

Example

Find the dot product of $\mathbf{x}_1 = \begin{pmatrix} 1 \\ 5 \\ 7 \end{pmatrix}$ and $\mathbf{x}_2 = \begin{pmatrix} -3 \\ 4 \\ 1 \end{pmatrix}$

$$\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = (1)\,(-3) + (5)\,(4) + (7)\,(1) = 24$$

Inner (Dot) Product of Vectors (contd.)

Properties of Inner Product

For all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in \mathcal{S}$ and all $\alpha \in \mathbb{R}$

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle = \sum_{i=1}^{n} u_i v_i$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle = \sum_{i=1}^{n} u_i w_i + \sum_{i=1}^{n} v_i w_i$
3. $\langle \mathbf{u}, \mathbf{w} + \mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \alpha \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u}, \alpha \mathbf{v} \rangle = \alpha \langle \mathbf{v}, \mathbf{u} \rangle = \alpha \sum_{i=1}^{n} u_i v_i$
5. $\langle \mathbf{u}, \mathbf{u} \rangle \geqslant 0$
6. $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$

Inner (Dot) Product of Vectors (contd.)

The Norm (Euclidean length): The norm of vector $x \in \mathcal{S}$, denoted by $\|x\|$ is defined as Euclidean length of the vector x

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

Angle Between Vectors: the angle θ between vectors x and y can be found using

$$\langle x, y \rangle = \|x\| \left\|y\right\| \cos\theta$$

Orthogonal Vectors: Vectors x and y are said to be orthogonal, $x \perp y$, if angle between them is $\pm 90^\circ$,

$$\langle x, y \rangle = \mathbf{0}.$$

Note that mutually orthogonal vectors are linearly independent (look at the standard basis of $\mathbb{R}^n$)

Inner (Dot) Product of Vectors (contd.)

Cauchy-Schwarz Inequality: For any two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then

$$\langle \mathbf{x}, \mathbf{y} \rangle \leqslant \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle \ \langle \mathbf{y}, \mathbf{y} \rangle}$$

or

$$\left(\sum_{i=1}^{n} x_i y_i \right)^2 \leqslant \left(\sum_{i=1}^{n} x_i^2 \right) \left(\sum_{i=1}^{n} y_i^2 \right)$$

Distance between two vectors: The Euclidean distance between two vectors $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^n$ is

$$d(\mathbf{x}, \mathbf{y}) = \left\| \mathbf{x} - \mathbf{y} \right\| = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

Inner (Dot) Product of Vectors (contd.)

Some Important results

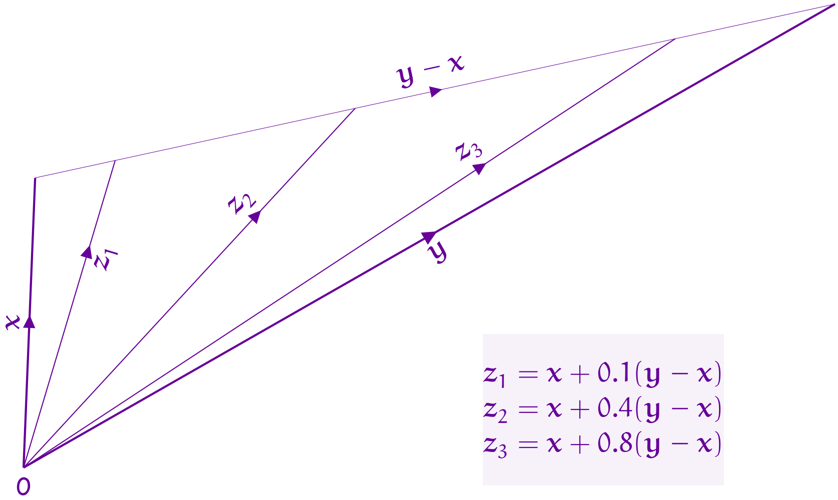
1. $\left\|x+y\right\|^2 = \|x\|^2 + \left\|y\right\|^2 + 2\langle x, y\rangle$
2. **Parallelogram law:** $\left\|x+y\right\|^2 + \left\|x-y\right\|^2 = 2\|x\|^2 + 2\left\|y\right\|^2$
3. **Pythagoras theorem:** If $x \perp y$ then $\left\|x+y\right\|^2 = \left\|x-y\right\|^2 = \|x\|^2 + \left\|y\right\|^2$
4. **Triangle inequality:** $\left\|x+y\right\| \leqslant \|x\| + \left\|y\right\|$

Line Segment

The line segment between two points x and y in $\mathbb{R}^n$ is the set of points on the straight line joining point x and y. If z lies on the line segment between points x and y, then

$$
\begin{aligned}
z &= x + \lambda\left(y - x\right) \qquad 0 \leqslant \lambda \leqslant 1 \\
\Rightarrow \quad z &= \lambda x + (1 - \lambda)\, y
\end{aligned}
$$

Line Segment (contd.)



$$z_1 = x + 0.1(y - x)$$
$$z_2 = x + 0.4(y - x)$$
$$z_3 = x + 0.8(y - x)$$

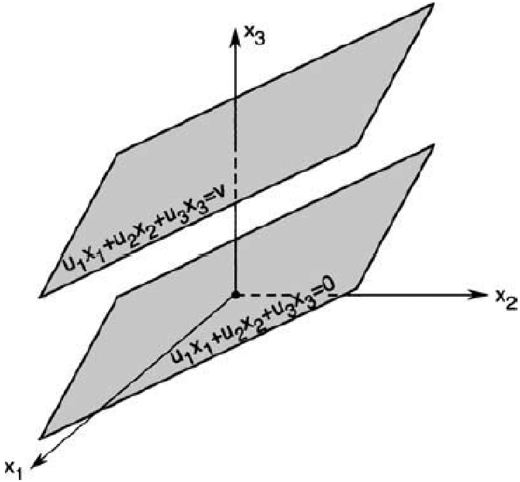
Hyperplanes

- The set of all points x that satisfy the linear equation

$$u_1 x_1 + y_2 x_2 + \cdots + u_n x_n = v$$

  is called a **hyperplane** of the space $\mathbb{R}^n$
- When hyperplanes contains the origin, then can be regarded as a subspace in $\mathbb{R}^{n-1}$ (dimension of hyperplanes is $n - 1$).
- In $\mathbb{R}^2$, hyperplanes are straight lines
- In $\mathbb{R}^3$, hyperplanes are ordinary planes

Hyperplanes (contd.)



Hyperplanes (contd.)

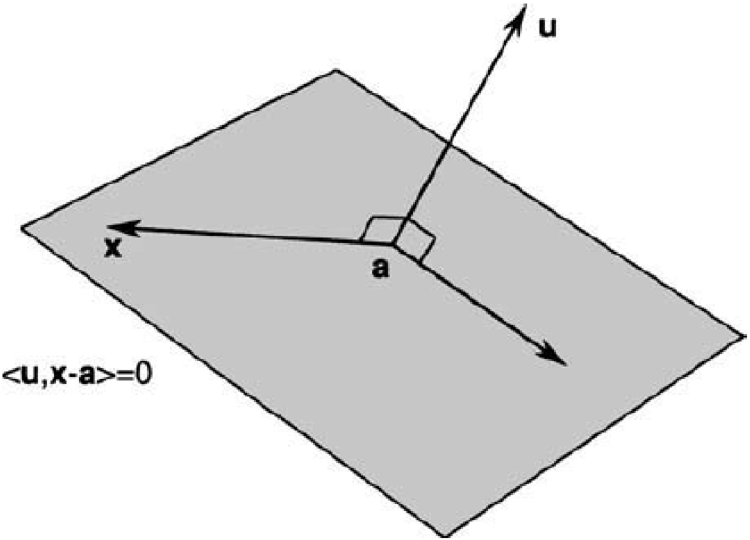
Let x_1 and x_2 be two arbitrary points of the hyperplane $H = \{x \in \mathbb{R}^n : \mathbf{u}^T x = v\}$.
Therefore

$$\mathbf{u}^T x_1 = v, \quad \mathbf{u}^T x_2 = v \quad \Rightarrow \mathbf{u}^T (x_1 - x_2) = 0 \Rightarrow \mathbf{u} \perp (x_1 - x_2)$$

where $(x_1 - x_2)$ is a vector on the hyperplane.

Therefore vector u is perpendicular to the hyperplane H.

Hyperplanes (contd.)



Matrices

A matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ of m raws and n columns (called $m \times n$ matrix) has the form $\mathbb{R}^{m\times n}$

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

The identity matrix

$$\mathbf{I}_m = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Inverse of a square matrix

for a square $m \times m$ matrix $\mathbf{A}$, its inverse is defined as

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}_m$$

Matrices (contd.)

Rank of a matrix

The rank is the maximum number of linearly independent columns of a matrix

Determinant of a square matrix A

$$\det(\boldsymbol{A}) = |\boldsymbol{A}|$$

you should now how to calculate at least $\det(\boldsymbol{A})$ **and** A^{-1} **for a** 3×3 **matrix**

Matrices (contd.)

Transpose of a matrix

$$\mathbf{A}^{\top} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}^{\top} = \begin{pmatrix} a_{11} & a_{21} & \vdots & a_{m1} \\ a_{12} & a_{22} & \vdots & a_{m2} \\ \dots & \dots & \ddots & \dots \\ a_{1n} & a_{2n} & \vdots & a_{mn} \end{pmatrix}$$

Symmetric Matrix

$$\mathbf{A}^{\top} = \mathbf{A}$$

Transpose Rules

- $(\mathbf{A}^{\top})^{\top} = \mathbf{A}$
- $(\mathbf{A} + \mathbf{B})^{\top} = \mathbf{A}^{\top} + \mathbf{B}^{\top}$
- $(k\mathbf{A})^{\top} = k\mathbf{A}^{\top}$
- $(\mathbf{AB})^{\top} = \mathbf{B}^{\top}\mathbf{A}^{\top}$
- $(\mathbf{A}^{\top})^{-1} = (\mathbf{A}^{-1})^{\top}$

Matrices (contd.)

Linear Transformations

The matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ transforms a vector $\boldsymbol{x} \in \mathbb{R}^n$ into a new vector $\boldsymbol{y} \in \mathbb{R}^m$

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$$

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

Eigenvalues and Eigenvectors : The eignevalue for a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ is a scalar λ such that

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v}$$

and $\boldsymbol{v}$ is called the eigenvector.

- Equation says that the transformation $\boldsymbol{A}\boldsymbol{v}$ does not change the direction of $\boldsymbol{v}$
- Eigenvectors of real symmetric matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ are mutually orthogonal.

Quadratic Forms

Let $\mathbf{A}$ be a $n \times n$ matrix A. Then the Quadratic form associated with A is a function $f(\mathbf{y}) = \mathbf{y} \cdot \mathbf{A}\mathbf{y}$ on $\mathcal{R}^n$

$$f(\mathbf{y}) = \mathbf{y} \cdot \mathbf{A}\mathbf{y} = \mathbf{y}^T \mathbf{A}\mathbf{y} = \sum_{i=1}^{n} \sum_{j}^{n} a_{ij} y_i y_j$$

Example

What is the quadratic form associated with 3×3 matrix $\mathbf{A}$

$$\mathbf{A} = \begin{pmatrix} 2 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 5 \end{pmatrix}$$

Quadratic Forms (contd.)

Let $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix}^T$

$$\begin{aligned}\mathbf{y} \cdot \mathbf{A}\mathbf{y} &= \begin{pmatrix} y_1 & y_2 & y_3 \end{pmatrix} \begin{pmatrix} 2 & -1 & 2 \\ -1 & 3 & 0 \\ 2 & 0 & 5 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \\ &= 2y_1^2 + 3y_2^2 + 5y_3^2 - 2y_1y_2 + 4y_1y_3.\end{aligned}$$

Quadratic Forms

Example

Find the matrix associated with the following quadratic function

$$q(y) = y_1^2 - y_2^2 + 4y_3^2 - 2y_1y_2 + 4y_2y_3$$

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 0 \\ 0 & -1 & 4 \\ 0 & 0 & 4 \end{pmatrix} \qquad \text{not symmetric}$$

$$\mathbf{Q} = \frac{1}{2}\left(\mathbf{A} + \mathbf{A}^T\right) = \frac{1}{2}\begin{pmatrix} 1 & -2 & 0 \\ 0 & -1 & 4 \\ 0 & 0 & 4 \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 \\ -2 & -1 & 0 \\ 0 & 4 & 4 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & -1 & 0 \\ -1 & -1 & 2 \\ 0 & 2 & 4 \end{pmatrix} \qquad \text{symmetric}$$

Quadratic Forms (contd.)

$$q(\mathbf{y}) = \mathbf{y} \cdot \mathbf{A}\mathbf{y} = \mathbf{y} \cdot \mathbf{Q}\mathbf{y}$$

$$\begin{aligned} q(\mathbf{y}) &= y_1^2 - y_2^2 + 4y_3^2 - 2y_1y_2 + 4y_2y_3 \\ &= y_1^2 - y_2^2 + 4y_3^2 - y_1y_2 - y_2y_1 + 2y_2y_3 + 2y_3y_2 \end{aligned}$$

Positive Definite Matrices

Definition

Suppose that A is an $n \times n$ symmetric matrix . Then $\mathbf{A}$ and its associated quadratic form $y \cdot \mathbf{A}y$ are called

1. **positive semidefinite** if $\mathbf{y} \cdot \mathbf{Ay} \geqslant 0$ for all y
2. **positive definite** if $\mathbf{y} \cdot \mathbf{Ay} > 0$ for all y
3. **negative semidefinite** if $\mathbf{y} \cdot \mathbf{Ay} \leqslant 0$ for all y
4. **negative definite** if $\mathbf{y} \cdot \mathbf{Ay} < 0$ for all y
5. **indefinite** if $\mathbf{y} \cdot \mathbf{Ay} > 0$ for some $y \in \mathrm{R}^n$ and $y \cdot \mathbf{A}y < 0$ other y

There are different methods to check the definiteness of a quadratic form. One method which is useful to check the positive definiteness of the matrix is shown next.

Positive Definite Matrices

Theorem (Sylvester's criterion)

A **symmetric** $n \times n$ *matrix* Q *with* **leading principal minors** $\Delta_i, \quad i = 1, 2, \ldots, n$. *Then*

1. Q *is* **positive definite** *if an only if all* $\Delta_i > 0, \quad i = 1, 2, \ldots, n.$
2. *If* $\Delta_i > 0 \; i = 1, 2, \ldots, n-1$ *and* $\Delta_n = 0$ *then* Q *is* **positive semidefinite**.

Theorem (Eigenvalue Test)

A **symmetric** $n \times n$ *matrix* Q *is*

1. **positive definite** *if and only if all eigenvalues are* **strictly positive**.
2. **positive semidefinite** *if and only if eigenvalues are* **non-negative.**

Positive Definite Matrices

Example

Determine the forms of the following matrices

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$$

for matrix $\mathbf{A}$: $\Delta_1 = 2 > 0$, $\Delta_2 = \begin{vmatrix} 2 & 0 \\ 0 & 4 \end{vmatrix} = (2 \times 4) - 0 = 8 > 0$,

$\Delta_3 = \begin{vmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{vmatrix} = 24 > 0$. Therefore $\mathbf{A}$ is positive definite

For matrix $\mathbf{B}$, $\Delta_1 = -1 < 0$. Therefore $\mathbf{B}$ is not positive definite.

Positive Definite Matrices

Example

Determine the forms of the following matrices

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} 2-\lambda & 0 & 0 \\ 0 & 4-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{vmatrix} = 0 \Rightarrow \lambda = 2, 3, 4$$

$$|\mathbf{B} - \lambda \mathbf{I}| = \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0 \Rightarrow \lambda - 2\lambda + \lambda^2 = 0 \Rightarrow \lambda = 0, 2$$

Therefore $\mathbf{A}$ is positive definite and matrix $\mathbf{B}$ is positive semidefinite

Differentiation

The directional derivative $Df(\mathbf{x}, \mathbf{d})$ of a function $f = f(x_1, x_2, \dots, x_n) : \mathbb{R}^n \to \mathbb{R}$ along the direction $\mathbf{d} \in \mathbb{R}^n$ is

$$Df(\mathbf{x}, \mathbf{d}) = \lim_{t \to 0^+} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t} \tag{1}$$

Partial Derivative: The partial derivative $\frac{\partial}{\partial x_j} f(\mathbf{x})$ with respect to the variable x_j is

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \lim_{t \to 0} \frac{f\left(x_1, x_2, \dots, x_j + t, \dots, x_n\right) - f(\mathbf{x})}{t}$$

$$\frac{\partial f(\mathbf{x})}{\partial x_j} = \lim_{t \to 0} \frac{f\left(\mathbf{x} + t\mathbf{e}_j\right) - f(\mathbf{x})}{t}$$

where $\mathbf{e}_j = \begin{pmatrix} 0 & 0 & \dots & 1 & 0 \end{pmatrix}^T$ is the unit vector with a "1" in the jth position

Differentiation (contd.)

The Gradient $\nabla f(\boldsymbol{x})$: The gradient of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at a point $\boldsymbol{x}$ is the **column vector** of the partial derivatives with respect to $x_1, x_2, \dots, x_n$

$$\nabla f(\boldsymbol{x}) = \begin{pmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_2} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{pmatrix}$$

Differentiation (contd.)

Example (Gradient of the distance function)

The distance function from a pint $\mathbf{p} \in \mathbb{R}^n$ to another point $\mathbf{x} \in \mathbb{R}^n$ is defined as

$$\rho(\mathbf{x}) = \left\|\mathbf{x} - \mathbf{p}\right\| = \sqrt{\sum_{i=1}^{n} (x_i - p_i)^2}$$

Solution

The function is differentiable at all $\mathbf{p} \neq \mathbf{x}$

$$\nabla\rho(\mathbf{x}) = \frac{1}{\left\|\mathbf{x} - \mathbf{p}\right\|}(\mathbf{x} - \mathbf{p})$$

Differentiation (contd.)

Example (Gradient of the log-sum-exp function)

The log-sum-exp function $\mathrm{lse}\colon \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathrm{lse}\left(\boldsymbol{x}\right) = \ln\left(\sum_{i=1}^{n} \exp\left(x_i\right)\right)$$

Solution

$$\nabla \mathrm{lse}\left(\boldsymbol{x}\right) = \frac{1}{\sum_{i=1}^{n} \exp\left(x_i\right)} \begin{pmatrix} \exp\left(x_1\right) \\ \exp\left(x_2\right) \\ \vdots \\ \exp\left(x_n\right) \end{pmatrix} = \frac{\boldsymbol{z}}{Z}$$

where $\boldsymbol{z} = \begin{pmatrix} \exp\left(x_1\right) & \exp\left(x_2\right) & \cdots & \exp\left(x_n\right) \end{pmatrix}^{\mathsf{T}}$ *and* $Z = \sum_{i=1}^{n} z_i$ *with* $z_i = \exp\left(x_i\right)$

The Hessian Matrix

The Hessian $\nabla^2 f(z)$ (a matrix of second derivatives) is defined as

$$\nabla^2 f(x) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & & \vdots \\ \vdots & & \ddots & \\ \vdots & & & \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

Theorem (Mean-Value Theorem)

Let $f(x)$ *be a continuous function in the interval* $[a, b]$ *. Then there exist a point* $a < z < b$ *such that*

$$\frac{f(b) - f(a)}{b - a} = f'(z)$$

$$\boxed{f(b) = f(a) + (b - a) f'(z)}$$

where $a < z < b$

Note that the point z is between a and b (**but unknown**).

Functions of a Single Variable (contd.)

Theorem (Taylor Series - Remainder Form)

$$f(b) = f(a) + (b-a) f'(a) + \frac{1}{2}(b-a)^2 f''(z)$$

where $a < z < b$.

- Note that we can always express z in terms of the distance between the two point a and b as follows: $z = a + \lambda(b-a)$ with $0 < \lambda < 1$.

Theorem (Mean-Value)

Suppose that a *and* a *are two points in* $\mathbb{R}^n$ *and that* $f(x)$ *is a function of* n *variables. Then*

$$f(b) = f(a) + \nabla f(z) \cdot (b - a)$$

$$f(b) = f(a) + \nabla f(a) \cdot (b - a) + \frac{1}{2}(b - a) \cdot \nabla^2 f(z)(b - a)$$

where z *is in the line joining* a *and* b. $z = a + \lambda(b - a), \quad 0 < \lambda < 1$

Examples

Exercise

Consider the function

$$f(\mathbf{x}) = \left(\mathbf{a}^{\mathsf{T}}\mathbf{x}\right)\left(\mathbf{b}^{\mathsf{T}}\mathbf{x}\right)$$

1. *find the gradiant* $\nabla f(\mathbf{x})$
2. *the Hessian* $\nabla^2 f(\mathbf{x})$

Solution

1. $\nabla f(\mathbf{x}) = \left(\mathbf{a}\mathbf{b}^{\mathsf{T}} + \mathbf{b}\mathbf{a}^{\mathsf{T}}\right)\mathbf{x}$
2. $\nabla^2 f(\mathbf{x}) = \mathbf{a}\mathbf{b}^{\mathsf{T}} + \mathbf{b}\mathbf{a}^{\mathsf{T}}$