Lecture 1 - Mathematical Preliminaries

The Space $\mathbb{R}^n$

- $\mathbb{R}^n$ - the set of $n$-dimensional column vectors with real components endowed with the component-wise addition operator:

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

and the scalar-vector product

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

- $e_1, e_2, \ldots, e_n$ - standard/canonical basis.
- $e$ and $0$ - all ones and all zeros column vectors.
Important Subsets of $\mathbb{R}^n$

- nonnegative orthant:
  \[ \mathbb{R}^n_+ = \{ (x_1, x_2, \ldots, x_n)^T : x_1, x_2, \ldots, x_n \geq 0 \} \]

- positive orthant:
  \[ \mathbb{R}^n_{++} = \{ (x_1, x_2, \ldots, x_n)^T : x_1, x_2, \ldots, x_n > 0 \} \]

- If $x, y \in \mathbb{R}^n$, the closed line segment between $x$ and $y$ is given by
  \[ [x, y] = \{ x + \alpha(y - x) : \alpha \in [0, 1] \} \]

- the open line segment $(x, y)$ is similarly defined as
  \[ (x, y) = \{ x + \alpha(y - x) : \alpha \in (0, 1) \} \]
  for $x \neq y$ and $(x, x) = \emptyset$

- unit-simplex:
  \[ \Delta_n = \{ x \in \mathbb{R}^n : x \succeq 0, e^T x = 1 \} \]
The Space $\mathbb{R}^{m \times n}$

- The set of all real valued matrices is denoted by $\mathbb{R}^{m \times n}$.
- $I_n$ - $n \times n$ identity matrix.
- $0_{m \times n}$ - $m \times n$ zeros matrix.
Inner Products

Definition An inner product on $\mathbb{R}^n$ is a map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ with the following properties:

1. (symmetry) $\langle x, y \rangle = \langle y, x \rangle$ for any $x, y \in \mathbb{R}^n$.
2. (additivity) $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$ for any $x, y, z \in \mathbb{R}^n$.
3. (homogeneity) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle$ for any $\lambda \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$.
4. (positive definiteness) $\langle x, x \rangle \geq 0$ for any $x \in \mathbb{R}^n$ and $\langle x, x \rangle = 0$ if and only if $x = 0$.

Examples

- the “dot product”

$$\langle x, y \rangle = x^T y = \sum_{i=1}^{n} x_i y_i$$ for any $x, y \in \mathbb{R}^n$.

- the “weighted dot product”

$$\langle x, y \rangle_w = \sum_{i=1}^{n} w_i x_i y_i,$$

where $\mathbf{w} \in \mathbb{R}^n_{++}$. 
Vector Norms

Definition. A norm $\| \cdot \|$ on $\mathbb{R}^n$ is a function $\| \cdot \| : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying

- **(Nonnegativity)** $\| x \| \geq 0$ for any $x \in \mathbb{R}^n$ and $\| x \| = 0$ if and only if $x = 0$.
- **(Positive homogeneity)** $\| \lambda x \| = |\lambda| \| x \|$ for any $x \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.
- **(Triangle inequality)** $\| x + y \| \leq \| x \| + \| y \|$ for any $x, y \in \mathbb{R}^n$.

One natural way to generate a norm on $\mathbb{R}^n$ is to take any inner product $\langle \cdot, \cdot \rangle$ defined on $\mathbb{R}^n$, and define the associated norm

$$\| x \| \equiv \sqrt{\langle x, x \rangle}, \text{ for all } x \in \mathbb{R}^n,$$

The norm associated with the dot-product is the so-called **Euclidean norm** or $l_2$-norm:

$$\| x \|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} \text{ for all } x \in \mathbb{R}^n.$$
\( l_p \)-norms

- The \( l_p \)-norm \((p \geq 1)\) is defined by
  \[
  \|x\|_p \equiv \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}.
  \]
- The \( l_\infty \)-norm is
  \[
  \|x\|_\infty \equiv \max_{i=1,2,\ldots,n} |x_i|.
  \]
- It can be shown that
  \[
  \|x\|_\infty = \lim_{p \to \infty} \|x\|_p.
  \]

Example: \( l_{1/2} \) is not a norm. why?

The Cauchy-Schwartz Inequality

**Lemma:** For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$:

$$|\mathbf{x}^T \mathbf{y}| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

**Proof:** For any $\lambda \in \mathbb{R}$:

$$\|\mathbf{x} + \lambda \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + 2\lambda \langle \mathbf{x}, \mathbf{y} \rangle + \lambda^2 \|\mathbf{y}\|^2$$

Therefore (why?),

$$4\langle \mathbf{x}, \mathbf{y} \rangle^2 - 4\|\mathbf{x}\|^2 \|\mathbf{y}\|^2 \leq 0,$$

establishing the desired result.
Matrix Norms

Definition. A norm \( \| \cdot \| \) on \( \mathbb{R}^{m \times n} \) is a function \( \| \cdot \| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R} \) satisfying

1. (Nonnegativity) \( \| A \| \geq 0 \) for any \( A \in \mathbb{R}^{m \times n} \) and \( \| A \| = 0 \) if and only if \( A = 0 \).

2. (Positive homogeneity) \( \| \lambda A \| = |\lambda| \| A \| \) for any \( A \in \mathbb{R}^{m \times n} \) and \( \lambda \in \mathbb{R} \).

3. (Triangle inequality) \( \| A + B \| \leq \| A \| + \| B \| \) for any \( A, B \in \mathbb{R}^{n} \).
Induced Norms

- Given a matrix $A \in \mathbb{R}^{m \times n}$ and two norms $\| \cdot \|_a$ and $\| \cdot \|_b$ on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, the induced matrix norm $\|A\|_{a,b}$ (called $(a,b)$-norm) is defined by

$$\|A\|_{a,b} = \max_{x} \{ \|Ax\|_b : \|x\|_a \leq 1 \}.$$ 

- conclusion:

$$\|Ax\|_b \leq \|A\|_{a,b} \|x\|_a$$

- An induced norm is a norm (satisfies nonnegativity, positive homogeneity and triangle inequality).

- We refer to the matrix-norm $\| \cdot \|_{a,b}$ as the $(a, b)$-norm. When $a = b$, we will simply refer to it as an $a$-norm.
Matrix Norms Contd

- **spectral norm:** If $\| \cdot \|_a = \| \cdot \|_b = \| \cdot \|_2$, the induced $(2,2)$-norm of a matrix $A \in \mathbb{R}^{m \times n}$ is the maximum singular value of $A$

$$\|A\|_2 = \|A\|_{2,2} = \sqrt{\lambda_{\text{max}}(A^T A)} \equiv \sigma_{\text{max}}(A),$$

This norm is called the **spectral norm**.

- **$l_1$-norm:** when $\| \cdot \|_a = \| \cdot \|_b = \| \cdot \|_1$, the induced $(1,1)$-matrix norm of a matrix $A \in \mathbb{R}^{m \times n}$ is given by

$$\|A\|_1 = \max_{j=1,2,\ldots,n} \sum_{i=1}^m |A_{i,j}|.$$

- **$l_\infty$-norm:** when $\| \cdot \|_a = \| \cdot \|_b = \| \cdot \|_\infty$, the induced $(\infty, \infty)$-matrix norm of a matrix $A \in \mathbb{R}^{m \times n}$ is given by

$$\|A\|_\infty = \max_{i=1,2,\ldots,m} \sum_{j=1}^n |A_{i,j}|.$$
The Frobenius norm

\[ ||A||_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A^2_{ij}}, \quad A \in \mathbb{R}^{m \times n} \]

The Frobenius norm is **not** an induced norm.
Why is it a norm?
Eigenvalues and Eigenvectors

▶ Let $A \in \mathbb{R}^{n \times n}$. Then a nonzero vector $v \in \mathbb{R}^n$ is called an eigenvector of $A$ if there exists a $\lambda \in \mathbb{C}$ for which

$$Av = \lambda v.$$ 

The scalar $\lambda$ is the eigenvalue corresponding to the eigenvector $v$.

▶ In general, real-valued matrices can have complex eigenvalues, but when the matrix is symmetric the eigenvalues are necessarily real.

▶ The eigenvalues of a symmetric $n \times n$ matrix $A$ are denoted by

$$\lambda_1(A) \geq \lambda_2(A) \geq \ldots \geq \lambda_n(A).$$

▶ The maximum eigenvalue is also denote by $\lambda_{\text{max}}(A)(= \lambda_1(A))$ and the minimum eigenvalue is also denote by $\lambda_{\text{min}}(A)(= \lambda_n(A))$. 
The Spectral Factorization Theorem

**Theorem.** Let $A \in \mathbb{R}^{n \times n}$ be an $n \times n$ symmetric matrix. Then there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ ($U^T U = U U^T = I$) and a diagonal matrix $D = \text{diag}(d_1, d_2, \ldots, d_n)$ for which

$$U^T A U = D.$$

- The columns of the matrix $U$ constitute an orthogonal basis comprising eigenvectors of $A$ and the diagonal elements of $D$ are the corresponding eigenvalues.
- A direct result is that $\text{Tr}(A) = \sum_{i=1}^{n} \lambda_i(A)$ and $\det(A) = \prod_{i=1}^{n} \lambda_i(A)$.
Basic Topological Concepts

- the open ball with center \( c \in \mathbb{R}^n \) and radius \( r \):
  \[
  B(c, r) = \{ x : \| x - c \| < r \}.
  \]

- the closed ball with center \( c \) and radius \( r \):
  \[
  B[c, r] = \{ x : \| x - c \| \leq r \}.
  \]

Definition. Given a set \( U \subseteq \mathbb{R}^n \), a point \( c \in U \) is called an interior point of \( U \) if there exists \( r > 0 \) for which \( B(c, r) \subseteq U \).

The set of all interior points of a given set \( U \) is called the interior of the set and is denoted by \( \text{int}(U) \):

\[
\text{int}(U) = \{ x \in U : B(x, r) \subseteq U \text{ for some } r > 0 \}.
\]

Examples.

\[
\begin{align*}
\text{int}(\mathbb{R}^n) &= \mathbb{R}^n, \\
\text{int}(B[c, r]) &= B(c, r) \quad (c \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \\
\text{int}([x, y]) &= ?
\end{align*}
\]
open and closed sets

- an open set is a set that contains only interior points. Meaning that $U = \text{int}(U)$.
- examples of open sets are open balls (hence the name...) and the positive orthant $\mathbb{R}^n_{++}$.

Result: a union of any number of open sets is an open set and the intersection of a finite number of open sets is open.

- a set $U \subseteq \mathbb{R}^n$ is closed if it contains all the limits of convergent sequences of vectors in $U$, that is, if $\{x_i\}_{i=1}^{\infty} \subseteq U$ satisfies $x_i \to x^*$ as $i \to \infty$, then $x^* \in U$.
- a known result states that $U$ is closed iff its complement $U^c$ is open.
- examples of closed sets are the closed ball $B[c, r]$, closed lines segments, the nonnegative orthant $\mathbb{R}^n_+$ and the unit simplex $\Delta_n$.

What about $\mathbb{R}^n \cap \emptyset$?
Boundary Points

Definition. Given a set $U \subseteq \mathbb{R}^n$, a boundary point of $U$ is a vector $x \in \mathbb{R}^n$ satisfying the following: any neighborhood of $x$ contains at least one point in $U$ and at least one point in its complement $U^c$.

- The set of all boundary points of a set $U$ is denoted by $\text{bd}(U)$.

Examples:

\[(c \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \text{bd}(B(c, r)) = \]
\[(c \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \text{bd}(B[c, r]) = \]
\[\text{bd}(\mathbb{R}_{++}^n) = \]
\[\text{bd}(\mathbb{R}_{++}^n) = \]
\[\text{bd}(\mathbb{R}_+^n) = \]
\[\text{bd}(\mathbb{R}_+^n) = \]
\[\text{bd}(\mathbb{R}^n) = \]
\[\text{bd}(\Delta_n) = \]
Closure

- the closure of a set $U \subseteq \mathbb{R}^n$ is denoted by $\text{cl}(U)$ and is defined to be the smallest closed set containing $U$:

$$ \text{cl}(U) = \bigcap \{ T : U \subseteq T, \text{ } T \text{ is closed} \}.$$

- another equivalent definition of $\text{cl}(U)$ is:

$$ \text{cl}(U) = U \cup \text{bd}(U).$$

Examples.

$$ \text{cl}(\mathbb{R}^n_{++}) = \text{...}, $$

$$(c \in \mathbb{R}^n, r \in \mathbb{R}_{++}), \text{cl}(B(c, r)) = \text{...},$$

$$(x \neq y), \text{cl}((x, y)) = \text{...}$$
Boundedness and Compactness

- A set $U \subseteq \mathbb{R}^n$ is called **bounded** if there exists $M > 0$ for which $U \subseteq B(0, M)$.
- A set $U \subseteq \mathbb{R}^n$ is called **compact** if it is closed and bounded.
- Examples of compact sets: closed balls, unit simplex, closed line segments.
Directional Derivatives and Gradients

Definition. Let \( f \) be a function defined on a set \( S \subseteq \mathbb{R}^n \). Let \( x \in \text{int}(S) \) and let \( d \in \mathbb{R}^n \). If the limit

\[
\lim_{t \to 0^+} \frac{f(x + td) - f(x)}{t}
\]

exists, then it is called the directional derivative of \( f \) at \( x \) along the direction \( d \) and is denoted by \( f'(x; d) \).

\[\Rightarrow\] For any \( i = 1, 2, \ldots, n \), if the limit

\[
\lim_{t \to 0} \frac{f(x + te_i) - f(x)}{t}
\]

exists, then its value is called the \( i \)-th partial derivative and is denoted by \( \frac{\partial f}{\partial x_i}(x) \).

\[\Rightarrow\] If all the partial derivatives of a function \( f \) exist at a point \( x \in \mathbb{R}^n \), then the gradient of \( f \) at \( x \) is

\[
\nabla f(x) = \left( \frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x), \ldots, \frac{\partial f}{\partial x_n}(x) \right).
\]
Continuous Differentiability

A function $f$ defined on an open set $U \subseteq \mathbb{R}^n$ is called continuously differentiable over $U$ if all the partial derivatives exist and are continuous on $U$. In that case,

$$f'(x; d) = \nabla f(x)^T d, \quad x \in U, d \in \mathbb{R}^n$$

**Proposition** Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that $f$ is continuously differentiable over $U$. Then

$$\lim_{d \to 0} \frac{f(x + d) - f(x) - \nabla f(x)^T d}{\|d\|} = 0 \text{ for all } x \in U.$$ 

Another way to write the above result is as follows:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + o(\|y - x\|),$$

where $o(\cdot) : \mathbb{R}_+^n \to \mathbb{R}$ is a one-dimensional function satisfying $\frac{o(t)}{t} \to 0$ as $t \to 0^+$. 
Twice Differentiability

- The partial derivatives \( \frac{\partial f}{\partial x_i} \) are themselves real-valued functions that can be partially differentiated. The \((i, j)\)-partial derivatives of \( f \) at \( x \in U \) (if exists) is defined by

\[
\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial}{\partial x_i} \left( \frac{\partial f}{\partial x_j} \right)(x).
\]

- A function \( f \) defined on an open set \( U \subseteq \mathbb{R}^n \) is called twice continuously differentiable over \( U \) if all the second order partial derivatives exist and are continuous over \( U \). In that case, for any \( i \neq j \) and any \( x \in U \):

\[
\frac{\partial^2 f}{\partial x_i \partial x_j}(x) = \frac{\partial^2 f}{\partial x_j \partial x_i}(x).
\]
The Hessian of $f$ at a point $x \in U$ is the $n \times n$ matrix:

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

For twice continuously differentiable functions, the Hessian is a symmetric matrix.
Linear Approximation Theorem

**Theorem.** Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that $f$ is twice continuously differentiable over $U$. Let $x \in U$ and $r > 0$ satisfy $B(x, r) \subseteq U$. Then for any $y \in B(x, r)$ there exists $\xi \in [x, y]$ such that:

$$f(y) = f(x) + \nabla f(x)^T (y - x) + \frac{1}{2} (y - x)^T \nabla^2 f(\xi)(y - x).$$
Theorem. Let $f : U \to \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Suppose that $f$ is twice continuously differentiable over $U$. Let $x \in U$ and $r > 0$ satisfy $B(x, r) \subseteq U$. Then for any $y \in B(x, r)$:

\[ f(y) = f(x) + \nabla f(x)^T(y - x) + \frac{1}{2} (y - x)^T \nabla^2 f(x)(y - x) + o(\|y - x\|^2). \]