Nonlinear Systems and Control
Lecture # 0
Mathematical Preliminaries

Vector Space (or Linear Space): A vector space $(\boldsymbol{\mathcal{V}}, \mathbb{F})$ consists of a set (of vectors) $\boldsymbol{\mathcal{V}}$, a field (of scalars) $\mathbb{F}$ and two operations viz. *addition* of vectors $(+)$ and *multiplication* of vectors by scalars $(\cdot)$, which obey the following axioms:

1. *Addition* is given by
   $+ : \boldsymbol{\mathcal{V}} \times \boldsymbol{\mathcal{V}} \to \boldsymbol{\mathcal{V}} : (\boldsymbol{x} + \boldsymbol{y}) \to \boldsymbol{x} + \boldsymbol{y}$; It is
   - Associative:
     $(\boldsymbol{x} + \boldsymbol{y}) + \boldsymbol{z} = \boldsymbol{x} + (\boldsymbol{y} + \boldsymbol{z}),\ \forall \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in \boldsymbol{\mathcal{V}}$
   - Commutative: $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{y} + \boldsymbol{x},\ \forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\mathcal{V}}$
   - $\exists!$ *identity* $\mathbf{0}$, (called the zero vector), s.t.
     $\boldsymbol{x} + \mathbf{0} = \mathbf{0} + \boldsymbol{x} = \boldsymbol{x},\ \forall \boldsymbol{x} \in \boldsymbol{\mathcal{V}}$
   - $\exists!$ inverse: $\forall \boldsymbol{x} \in \boldsymbol{\mathcal{V}}$, $\exists!(-\boldsymbol{x}) \in \boldsymbol{\mathcal{V}}$ s.t.
     $\boldsymbol{x} + (-\boldsymbol{x}) = \mathbf{0}$;

2. *Multiplication* by scalars is given by $\cdot : \mathbb{F} \times \boldsymbol{\mathcal{V}} \rightarrow \boldsymbol{\mathcal{V}} (\boldsymbol{\alpha}, \boldsymbol{x}) \rightarrow \boldsymbol{\alpha x}$, where $\forall \boldsymbol{x} \in \boldsymbol{\mathcal{V}}$ and $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{F}$,
   - $(\boldsymbol{\alpha\beta})\boldsymbol{x} = \boldsymbol{\alpha}(\boldsymbol{\beta x})$
   - $\mathbf{1}\boldsymbol{x} = \boldsymbol{x}$, $\mathbf{0}\boldsymbol{x} = \mathbf{0}$;
3. Addition and multiplication by scalars are related by *distributed laws* viz.
   - $\forall \boldsymbol{x} \in \boldsymbol{\mathcal{V}}$, $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{F}$ $(\boldsymbol{\alpha} + \boldsymbol{\beta})\boldsymbol{x} = \boldsymbol{\alpha x} + \boldsymbol{\beta x}$
   - $\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\mathcal{V}}$, $\forall \boldsymbol{\alpha} \in \mathbb{F}$ $\boldsymbol{\alpha}(\boldsymbol{x} + \boldsymbol{y}) = \boldsymbol{\alpha x} + \boldsymbol{\alpha y}$.

Subspace: Let $\boldsymbol{\mathcal{W}}$ be a subset of $\boldsymbol{\mathcal{V}}$. If $\boldsymbol{\mathcal{W}}$ is a vector space itself, with the same vector space operations as $\boldsymbol{\mathcal{V}}$ has, then it is a subspace of $\boldsymbol{\mathcal{V}}$.

Inner Product Space: A vector space $(\mathcal{V}, \mathbb{F})$ is an inner product space if there is a function $\langle \cdot, \cdot \rangle : \mathcal{V} \times \mathcal{V} \mapsto \mathbb{F}$ such that for every $x, y, z \in \mathcal{V}$ and $\alpha \in \mathbb{F}$ the following hold:

- $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle,$
  $\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle$
- $\langle x, \alpha y \rangle = \alpha \langle x, y \rangle \quad \forall \alpha \in \mathbb{F}$
- $\|x\|^2 = \langle x, x \rangle \geq 0$
- $\langle x, x \rangle = 0$ if and only $x = 0$.
- $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where the overbar denotes the complex conjugate operator.

The function $\langle \cdot, \cdot \rangle$ is called the inner product on $\mathcal{V}$.

The supremum or least upper bound (LUB) of a set S of real numbers is denoted by sup(S) and is defined to be the smallest real number that is greater than or equal to every number in S. Every nonempty subset of the set of real numbers that is bounded above has a supremum that is also an element of the set of real numbers.

$$\sup\{x \in \mathbb{R} : 0 < x < 1\} = \sup\{x \in \mathbb{R} : 0 \leq x \leq 1\} = 1.$$

The infimum or greatest lower bound (GLB) of a set S of real numbers is denoted by inf(S) and is defined to be the biggest real number that is smaller than or equal to every number in S. Any bounded nonempty subset of the real numbers has an infimum in the non-extended real numbers.

$$\inf\{x \in \mathbb{R} : 0 < x < 1\} = 0.$$

Vector Norms: Let $\mathcal{X}$ be a vector space, a real-valued function $\|\cdot\|$ defined on $\mathcal{X}$ is said to be a norm on $\mathcal{X}$ if it satisfies the following properties:

- $\|x\| \geq 0$ (positivity) ;
- $\|x\| = 0$ if and only if $x = 0$ (positive definiteness) ;
- $\|\alpha x\| = |\alpha| \|x\|,$ for any scalar α (homogeneity) ;
- $\|x + y\| \neq \|x\| + \|y\|$ (triangle inequality) for any $x \in \mathcal{X}$ and $y \in \mathcal{X}$.

Finite-dimensional Vector p-norm: Let $x \in \mathbb{R}^n$. Then we define the vector p-norm of x as

$$\|x\|_p := \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p} \text{ for } 1 \leq p \leq \infty.$$

In particular, when $p = 1, 2, \infty$, we have

$$\|x\|_1 := \sum_{i=1}^{n} |x_i|; \quad \|x\|_2 := \sqrt{\sum_{i=1}^{n} |x_i|^2}$$

$$\|x\|_\infty := \max_{1 \leq i \leq n} |x_i|$$

Equivalence between p-norms: All p-norms are equivalent in the sense that if $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ are two different p-norms, then there exist positive constants c_1 and c_2 such that

$$c_1\|x\|_\alpha \leq \|x\|_\beta \leq c_2\|x\|_\alpha$$

for all $x \in \mathbb{R}^n$.
For the $1-$, $2-$, and $\infty-$norms, we have

$$\boxed{\begin{array}{c} \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2 \\ \|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty \\ \|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty \end{array}}$$

Hölder Inequality: For all $x \in \mathbb{R}^n$, $y \in \mathbb{R}^n$,

$$|x^T y| \leq \|x\|_p \|y\|_q, \ \frac{1}{p} + \frac{1}{q} = 1.$$

For $p = q = 2$, Hölder's inequality results in the Cauchy-Schwarz inequality.
For $\mathcal{L}_p$ spaces with $f \in \mathcal{L}_p$, $g \in \mathcal{L}_q$ and $\frac{1}{p} + \frac{1}{q} = 1$:

$$\int |f(t)g(t)|dt \leq \left(\int |f(t)|^p dt\right)^{1/p} \left(\int |g(t)|^q dt\right)^{1/q}.$$

Matrix Induced Norms: Let $A = [a_{ij}] \in \mathbb{R}^{m \times n}$, then the matrix norm induced by a vector p-norm is defined as

$$\|A\|_p := \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p} = \max_{\|x\|_p = 1} \|Ax\|_p$$

- The induced matrix 2-norm

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

- The induced matrix 1-norm

$$\|A\|_1 = \max_j \sum_{i=1}^{m} |a_{ij}|$$

- The induced matrix ∞-norm

$$\|A\|_\infty = \max_i \sum_{j=1}^{n} |a_{ij}|$$

- Frobenius norm (not an induced norm)
$\|A\|_F := \sqrt{\text{trace}\, A^T A} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$.

Facts of Matrix Norms: Let $A \in \mathbb{R}^{m \times n}$ and $B \in \mathbb{R}^{n \times l}$, given unitary matrices $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$, i.e., $U^T U = I_m, V^T V = I_n$, we have

$$\boxed{\begin{array}{c}
\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty \\
\frac{1}{\sqrt{m}}\|A\|_1 \le \|A\|_2 \le \sqrt{n}\|A\|_1 \\
\|UAV\|_2 = \|A\|_2 \\
\|A\|_p < 1, \text{ then } \det(I - A) \neq 0 \\
\|A\|_2 \le \sqrt{\|A\|_1\|A\|_\infty} \\
\|AB\|_p \le \|A\|_p\|B\|_p \\
\|A_1 + A_2\|_p \le \|A_1\|_p + \|A_2\|_p
\end{array}}$$

Proof of $\frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$

$$\|A\|_2 := \sup_{x\neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2$$

$$\|Ax\|_\infty \le \|Ax\|_2 \le \sqrt{m}\|Ax\|_\infty$$

$$\|x\|_\infty \le \|x\|_2 \le \sqrt{n}\|x\|_\infty$$

$$\sup_{x\neq 0} \frac{\|Ax\|_\infty}{\sqrt{n}\|x\|_\infty} \le \|A\|_2 \le \sup_{x\neq 0} \frac{\sqrt{m}\|Ax\|_\infty}{\|x\|_\infty}$$

$$\Rightarrow \frac{1}{\sqrt{n}}\|A\|_\infty \le \|A\|_2 \le \sqrt{m}\|A\|_\infty$$

Proof of $\frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$

$$\|A\|_2 := \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} = \max_{\|x\|_2=1} \|Ax\|_2$$

$$\|Ax\|_2 \leq \|Ax\|_1 \leq \sqrt{m}\|Ax\|_2$$

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\|x\|_2$$

$$\sup_{x \neq 0} \frac{\|Ax\|_1}{\sqrt{m}\|x\|_1} \leq \|A\|_2 \leq \sup_{x \neq 0} \frac{\sqrt{n}\|Ax\|_1}{\|x\|_1}$$

$$\Rightarrow \frac{1}{\sqrt{m}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$$

Proof of $\|UAV\|_2 = \|A\|_2$

$$U^T U = I_m; \quad V^T V = I_n$$

$$\|UAx\|_2^2 = x^T A^T U^T U A x = x^T A^T A x = \|Ax\|_2^2$$

$$\Rightarrow \|UA\|_2 = \max_{\|x\|_2=1} \|UAx\|_2 = \max_{\|x\|_2=1} \|Ax\|_2 = \|A\|_2$$

$$\Rightarrow \|AV\|_2 = \|V^T A^T\|_2 = \|A^T\|_2 = \|A\|_2$$

$$\Rightarrow \|UAV\|_2 = \|A\|_2$$

Proof of $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$

$$\|A\|_2^2 = \lambda_{max}(A^T A) \Rightarrow A^T A x = \|A\|_2^2 x$$

$$\begin{aligned} \|A\|_2^2 \|x\|_1 &= \|A^T A x\|_1 \\ &\leq \|A^T\|_1 \|A\|_1 \|x\|_1 \\ &= \|A\|_\infty \|A\|_1 \|x\|_1 \end{aligned}$$

$$\Rightarrow \|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$$

Proof of $\|AB\|_p \leq \|A\|_p \|B\|_p$

$$
\begin{aligned}
\|AB\|_p &= \sup_{x \neq 0} \frac{\|ABx\|_p}{\|x\|_p} \\
&= \sup_{x, Bx \neq 0} \frac{\|A(Bx)\|_p}{\|Bx\|_p} \frac{\|Bx\|_p}{\|x\|_p} \\
&\leq \sup_{y \neq 0} \frac{\|Ay\|_p}{\|y\|_p} \cdot \sup_{x \neq 0} \frac{\|Bx)\|_p}{\|x\|_p} \\
&= \|A\|_p \|B\|_p
\end{aligned}
$$

$$
\Rightarrow \|AB\|_p \leq \|A\|_p \|B\|_p
$$

Proof of $\|\boldsymbol{A_1} + \boldsymbol{A_2}\|_p \leq \|\boldsymbol{A_1}\|_p + \|\boldsymbol{A_2}\|_p$

$$
\begin{aligned}
\|\boldsymbol{A_1} + \boldsymbol{A_2}\|_p &= \max_{\|\boldsymbol{x}\|_p = 1} \|(\boldsymbol{A_1} + \boldsymbol{A_2})\boldsymbol{x}\|_p \\
&= \max_{\|\boldsymbol{x}\|_p = 1} \|\boldsymbol{A_1 x} + \boldsymbol{A_2 x}\|_p \\
&\leq \max_{\|\boldsymbol{x}\|_p = 1} (\|\boldsymbol{A_1 x}\| + \|\boldsymbol{A_2 x}\|_p) \\
&\leq \max_{\|\boldsymbol{x}\|_p = 1} \|\boldsymbol{A_1 x}\|_p + \max_{\|\boldsymbol{x}\|_p = 1} \|\boldsymbol{A_2 x}\|_p \\
&= \|\boldsymbol{A_1}\|_p + \|\boldsymbol{A_2}\|_p \\
\Rightarrow \|\boldsymbol{A_1} + \boldsymbol{A_2}\|_p &\leq \|\boldsymbol{A_1}\|_p + \|\boldsymbol{A_2}\|_p
\end{aligned}
$$

Eigenvalues: Roots of the characteristic polynomial $\det(\boldsymbol{\lambda} \boldsymbol{I}_n - \boldsymbol{A})$, where $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.

Repeated Eigenvalues: The number of linearly independent eigenvectors $\boldsymbol{q}_i$ associated with an eigenvalue $\boldsymbol{\lambda}_i$ repeated with an algebraic multiplicity $\boldsymbol{m}_i$ is equal to the nullity of $(\boldsymbol{\lambda}_i \boldsymbol{I}_n - \boldsymbol{A})$. This dimension is given by $\boldsymbol{q}_i = \boldsymbol{n} -$ rank$(\boldsymbol{\lambda}_i \boldsymbol{I}_n - \boldsymbol{A})$ and is called the geometric multiplicity of $\boldsymbol{\lambda}_i$, since it is the dimension of the subspace spanned by the eigenvectors.

Hermitian Matrix: A Hermitian matrix $\boldsymbol{H} \in \mathbb{C}^{n\times n}$ (or self-adjoint matrix) is a square matrix with complex entries which is equal to its own conjugate transpose, i.e., $\boldsymbol{H} = \boldsymbol{H}^*$. If $\boldsymbol{H} \in \mathbb{R}^{n\times n}$, then $\boldsymbol{H}$ is symmetrical. Facts are:

- All the eigenvalues of a hermitian matrix $\boldsymbol{H}$ are real.
- The Jordan from representation of a hermitian matrix is diagonal.
- The eigenvectors of a hermitian matrix corresponding to different eigenvalues are orthogonal.
- Let $\boldsymbol{\lambda}_{min}$ and $\boldsymbol{\lambda}_{max}$ be the smallest and largest eigenvalues of a hermitian matrix $\boldsymbol{P}$. Then we have $\boldsymbol{\lambda}_{min}\|\boldsymbol{x}\|_2^2 \le \boldsymbol{x}^*\boldsymbol{P}\boldsymbol{x} \le \boldsymbol{\lambda}_{max}\|\boldsymbol{x}\|_2^2$ for any $\boldsymbol{x} \in \mathbb{C}^n$.

Positive Definite Matrix: A Hermitian matrix $\boldsymbol{P} \in \mathbb{C}^{\boldsymbol{n} \times \boldsymbol{n}}$ is positive definite (positive semidefinite) if and only if any one of the following conditions holds:

- All the eigenvlaues of $\boldsymbol{P}$ are positive (nonnegative).
- All the leading principle minors of $\boldsymbol{P}$ are positive (all the principal minors of $\boldsymbol{P}$ are nonnegative).
- There exists a nonsingular matrix $\boldsymbol{N}$ (a singular matrix $\boldsymbol{N}$) such that $\boldsymbol{P} = \boldsymbol{N}^* \boldsymbol{N}$.

Topological Concepts in $\mathbb{R}^n$

Convergence: A sequence $\{x_k\} \in \mathbb{R}^n$,converges to $x \in \mathbb{R}^n$ if

$$\|x_k - x\| \to 0 \text{ as } k \to \infty,$$

which is equivalent to saying that, $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}$ such that

$$\|x_k - x\| < \varepsilon, \ \forall k \geq N.$$

Open Set: A set $S \subset \mathbb{R}^n$ is open if, for every vector $x \in S$, one can find an $\varepsilon-$neighborhood of x $N(x, \varepsilon) = \{z \in \mathbb{R}^n \mid \|z - x\| < \varepsilon\}$ such that $N(x, \varepsilon) \subset S$.

Closed Set: A set $S \subset \mathbb{R}^n$ is closed if and only if its complement in $\mathbb{R}^n$ is open. S is closed if and only if every convergent sequence with elements in S has its limit in S.

Boundedness: A set $\boldsymbol{S}$ is bounded if there is $\boldsymbol{r} > \boldsymbol{0}$ such that $\|\boldsymbol{x}\| \leq \boldsymbol{r}$ for all $\boldsymbol{x} \in \boldsymbol{S}$.
Compact: A set $\boldsymbol{S}$ is compact if it is closed and bounded.
Boundary: A point $\boldsymbol{p}$ is a boundary point of a set $\boldsymbol{S}$ if every neighborhood of $\boldsymbol{p}$ contains at least one point of $\boldsymbol{S}$ and one point not belonging to $\boldsymbol{S}$. The set of all boundary points of $\boldsymbol{S}$, denoted by $\boldsymbol{\partial S}$, is called the boundary of $\boldsymbol{S}$. A closed set contains all its boundary points. An open set contains none of its boundary points.
Interior: The interior of a set $\boldsymbol{S}$ is $\boldsymbol{S} \setminus \boldsymbol{\partial S}$.
Closure: The closure of a set $\boldsymbol{S}$, denoted by $\overline{\boldsymbol{S}}$, is the union of $\boldsymbol{S}$ and its boundary. A closed set is equal to its closure.
Connected: An open set $\boldsymbol{S}$ is connected if every pair of points in $\boldsymbol{S}$ can be joined by an arc lying in $\boldsymbol{S}$.

Region: A set S is called a region if it is the union of an open connected set with some, none, or all of its boundary points. If none of boundary points are included, the region is called an open region or domain.
Convex: A set S is convex if, for every $x, y \in S$ and every real number $\theta, 0 < \theta < 1$, the point $\theta x + (1 - \theta)y \in S$.
Continuous Functions: A function $f : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is said to be *continuous* at a point x if, given $\varepsilon > 0$, $\exists \delta > 0$ such that

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

A function f is continuous on a set S if it is continuous at every point of S, and it is uniformly continuous on S if, given $\varepsilon > 0$ there is $\delta > 0$ (depending only on ε) such that the inequality holds for all $x, y \in S$.

If f is uniformly continuous on a set S, then it is continuous on S. The converse is not true in general. However if S is a compact set, then continuity and uniform continuity on S are equivalent.

The linear combination of any two continuous functions is continuous. The composition of two continuous functions f_1 and f_2, i.e., $(f_2 \circ f_1)(\cdot) = f_2(f_1(\cdot))$ is continuous.

Image: If $f : S \longrightarrow \mathbb{R}^m$, then the set of $f(x)$ such that $x \in S$ is called the image of S under f and is denoted by $f(S)$.

If f is a continuous function defined on a compact set S, then $f(S)$ is compact; hence, continuous functions on compact sets are bounded.

If $f : S \longrightarrow \mathbb{R}$, then there are points p and q in the compact set S such that $f(p) \leq f(x) \leq f(q)$ for all $x \in S$.

If is a continuous function defined on a connected set S, then $f(S)$ is connected.
A function f defined on a set S is said to be one to one on S if whenever $x, y \in S$, and $x \neq y$, then $f(x) \neq f(y)$. If $f : S \longrightarrow \mathbb{R}^m$ is a continuous one-to-one function on a compact set $s \subset \mathbb{R}^n$, then f has a continuous inverse f^{-1} on $f(S)$, i.e., $f^{-1}(f(x)) = x$.
A function $f : R \longrightarrow \mathbb{R}^n$ is said to be piecewise continuous on an interval $J \subset \mathbb{R}$ if for every bounded subinterval $J_0 \subset J$, f is continuous for all $x \in J_0$, except, possibly, at a finite number of points where f may have discontinuities.

Implicit Function Theorem: Assume that $\boldsymbol{f} : \mathbb{R}^{\boldsymbol{n}} \times \mathbb{R}^{\boldsymbol{m}} \longrightarrow \mathbb{R}^{\boldsymbol{n}}$ is continuously differentiable [a] at each point $(\boldsymbol{x}, \boldsymbol{y})$ of an open set $\boldsymbol{S} \subset \mathbb{R}^{\boldsymbol{n}} \times \mathbb{R}^{\boldsymbol{m}}$. Let $(\boldsymbol{x_0}, \boldsymbol{y_0})$ be a point in $\boldsymbol{S}$ for which $\boldsymbol{f}(\boldsymbol{x_0}, \boldsymbol{y_0}) = \boldsymbol{0}$ and for which the Jacobian matrix $[\boldsymbol{\partial f / \partial x}](\boldsymbol{x_0}, \boldsymbol{y_0})$ is nonsingular. Then there exist neighborhood $\boldsymbol{U} \subset \mathbb{R}^{\boldsymbol{n}}$ of $\boldsymbol{x_0}$ and $\boldsymbol{V} \subset \mathbb{R}^{\boldsymbol{m}}$ of $\boldsymbol{y_0}$ such that for each $\boldsymbol{y} \in \boldsymbol{V}$ the equation $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0}$ has a unique solution $\boldsymbol{x} \in \boldsymbol{U}$. Moreover, this solution can be given as $\boldsymbol{x} = \boldsymbol{g}(\boldsymbol{y})$, where $\boldsymbol{g}$ is continuously differentiable at $\boldsymbol{y} = \boldsymbol{y_0}$.

[a] The partial derivatives exist and are continuous at the suggested point.

Mean Value Theorem: Assume that $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ is continuously differentiable at each point x of an open set $S \subset \mathbb{R}^n$. Let x and y be two points of S such that the line segment $L(x, y) := \{z \mid z = \theta x + (1 - \theta)y,\ 0 < \theta < 1\} \subset S$. Then there exists a point z of $L(x, y)$ such that

$$f(y) - f(x) = \frac{\partial f}{\partial x}\Big|_{x=z}(y - x).$$

Taylor's theorem: If f is an $n \in \mathbb{N}$ times continuously differentiable function on $[a, x]$ and $n+1$ times differentiable on (a, x), then

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n,$$

where the remainder R_n is give by

$$R_n := \frac{f^{(n+1)}(z)}{(n+1)!}(x-a)^{n+1}, \exists z \in (a, x).$$

This is a generalization of the mean value theorem.

Taylor's series:

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots$$
$$= \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k.$$

Monotonic Sequence: A sequence $\{s_n\}$ is said to be

- monotonically increasing if $s_n \leq s_{n+1}$, $n \in \mathbb{N}$;
- monotonically decreasing if $s_n \geq s_{n+1}$, $n \in \mathbb{N}$.

Theorem: Suppose $\{s_n\}$ is monotonic. Then $\{s_n\}$ converges if and only if it is bounded.

Proof: Suppose $s_n \leq s_{n+1}$. Let E be the range of $\{s_n\}$. If $\{s_n\}$ is bounded, then $\exists s$ the least upper bound of E. Then $s_n \leq s, \forall n \in \mathbb{N}$. For every $\epsilon > 0$, there is N such that $s - \varepsilon < s_N \leq s$, otherwise $s - \varepsilon$ would be an upper bound of E. Since $\{s_n\}$ increases, $n \geq N$ therefore implies $s - \varepsilon < s_n \leq s$, which shows that $\{s_n\}$ converges to s ($\Leftarrow$). $\Rightarrow$ is follows from the fact that if $\{s_n\}$ converges, then $\{s_n\}$ is bounded.

Normed Linear Space: A linear space $\mathcal{X}$ is a normed linear space if, to each vector $x \in \mathcal{X}$, there is a norm $\|x\|$.
Cauchy Sequence: A sequence $\{x_k\} \in \mathcal{X}$ is said to be a Cauchy sequence if

$$\|x_k - x_m\| \to 0 \text{ as } k, m \to \infty.$$

Every convergent sequence is Cauchy, but not vice versa.
Banach Space: A normed linear space $\mathcal{X}$ is complete if every Cauchy sequence in $\mathcal{X}$ converges to a vector in $\mathcal{X}$. A complete normed linear space is a Banach space.

Banach Spaces:

$l_p[0,\infty)$ spaces for $1 \le p < \infty$: For each p, $l_p[0,\infty)$ consists of all sequences $x = (x_0, x_1, \cdots)$ such that $\sum_{i=1}^{\infty} |x_i|^p < \infty$. The associated norm is

$$\|x\|_{l_p} := \left(\sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} .$$

$l_\infty[0,\infty)$ space consists of all bounded sequences $x = (x_0, x_1, \cdots)$, and the l_∞ norm is

$$\|x\|_{l_\infty} := \sup_i |x_i|.$$

Banach Spaces:

$\mathcal{L}_p(I)$ spaces for $1 \le p < \infty$: For each p, $\mathcal{L}_p(I)$ consists of all Lebesgue measurable functions $x(t)$ defined on an interval $I \in \mathbb{R}$ such that

$$\|x\|_{\mathcal{L}_p} := \left(\int_I |x(t)|^p dt \right)^{1/p} < \infty, \text{ for } 1 \le p < \infty.$$

and

$$\|x(t)\|_{\mathcal{L}_\infty} := \sup_{t \in I} |x(t)| < \infty.$$

$C[a, b]$ space consists of all continuous functions on the real interval $[a, b]$ with the norm defined as

$$\|x\|_C := \sup_{t \in [a,b]} |x(t)|.$$

Logic and Proofs:

Law of the Excluded Middle: Every statement must be either true or false.

Logical operators:

Statement	Notation
P and Q	$P \wedge Q$
P or Q	$P \vee Q$
If P, then Q (or P implies Q)	$P \Rightarrow Q$
P if and only if Q (or P and Q are equivalent)	$P \Leftrightarrow Q$
not P	$\neg P$

where P and Q are statements.

A statement P is a necessary condition of a statement Q if Q implies P ($Q \Rightarrow P$). A necessary condition P of a statement Q must be satisfied for the statement Q to be true.

A statement P is a sufficient condition of a statement Q if P implies Q ($P \Rightarrow Q$). A sufficient condition P is one that, if satisfied, assures the statement Q's truth.

A statement P is a necessary and sufficient condition of a statement Q if P if and only if Q ($P \Leftrightarrow Q$).

Tautologies: True statements for any cases.

Statement	True/False
$P \Rightarrow P$	True
$(P \Rightarrow Q) \Leftrightarrow (\neg Q \Rightarrow \neg P)$ (Contrapositive)	True
$P \vee \neg P$	True
$P \wedge (P \Rightarrow Q) \Rightarrow Q$	True
$(P \Rightarrow Q) \wedge (Q \Rightarrow R) \Rightarrow (P \Rightarrow R)$	True

The converse of $P \Rightarrow Q$, i.e., $Q \Rightarrow P$ is not always true.

Proof by Negation (or Contradiction): To prove $P \Rightarrow Q$,

- Assume P is true and Q is false, i.e., $P \wedge \neg Q$ is true.
- Derive a contradiction F.
- Thus our assumption $(P \wedge \neg Q)$ is false. Hence if P is true, Q must be true.

In summary, we use

$$[(P \wedge \neg Q) \Rightarrow F] \Rightarrow [P \Rightarrow Q],$$

where F is a contradiction.

Proof by Induction: To prove $P(n)$ is true for all n,

- Prove $P(1)$ is true.
- Prove for all n, if $P(n)$ is true, then $P(n+1)$ is true.
- Then for all n, $P(n)$ is true.

In summary, we use

$$P(1) \wedge \forall n [P(n) \Rightarrow P(n+1)] \Rightarrow \forall n, P(n).$$