Lecture 0: Mathematical Preliminaries

Direction and motivation. It can be proven that the linear system

$$\dot{\boldsymbol{x}} = \boldsymbol{A}(t)\boldsymbol{x}, \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{1}$$

possesses a unique solution corresponding to each initial state $\boldsymbol{x}_0$. A nonlinear system of the general form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, t), \qquad \boldsymbol{x}(0) = \boldsymbol{x}_0, \tag{2}$$

may not even have a solution. If $\boldsymbol{f}(\boldsymbol{x}, t)$ is *continuous* in $\boldsymbol{x}$ and t, then at least one solution exists. However, there may be several possible solutions. We will spend a bit of time developing the vocabulary and tools needed for nonlinear analysis, paying particular attention to the question of existence and uniqueness of solutions and their dependence on initial conditions and system parameters. One important tool is the Banach Fixed Point Theorem. The proof of the theorem requires the notion of a *complete, normed linear space*, or *Banach space*. We begin by developing this notion from more basic ideas.

Groups, Fields, and Vector Spaces

Definition: A **group** G is a set together with a binary operation ($\circ$) called "group multiplication" with the following properties

1. *Closure:* $a \circ b \in G$ for all $a, b \in G$.
2. *Associativity:* $(a \circ b) \circ c = a \circ (b \circ c)$ for all $a, b, c \in G$.
3. An *identity* element: There exists an element $e \in G$ such that $a \circ e = e \circ a = a$ for every $a \in G$.
4. An *inverse* element: For each $a \in G$, there exists an element $a^{-1} \in G$ such that $a \circ a^{-1} = a^{-1} \circ a = e$.

A group G is called **abelian** or **commutative** if group multiplication is commutative, i.e., if $a \circ b = b \circ a$ for every $a, b \in G$.

Example 1: $\mathbb{R}$ with scalar addition.

1. $\mathbb{R}$ is closed because the sum of any two real numbers is real.
2. Scalar addition is associative.
3. The identity element is the number 0.
4. The inverse of an element a is $-a$.

This group is abelian because scalar addition is commutative.

Example 2: $\mathbb{R}^n$ with $n \in \{1, 2, \ldots\}$ and vector (element-wise) addition. This group is also abelian, since $\mathbb{R}$ is abelian.

Definition: A **field** $\mathbb{K}$ is a set with two binary operations, addition ('+') and multiplication ('$\circ$'), such that

1. $\mathbb{K}$ is an abelian group under addition with additive identity 0.
2. $\mathbb{K} - \{0\}$ is an abelian group under multiplication with multiplicative identity 1.
3. Multiplication distributes over addition: $a \circ (b + c) = a \circ b + a \circ c$ for all $a, b, c \in \mathbb{K}$.

Note: A **ring** is a slightly more general object than a field in the sense that $\mathbb{K} - \{0\}$ need not be a group under multiplication.

Example 1: $\mathbb{R}$ with the usual scalar addition and multiplication.

Example 2: $\mathbb{C} = \{x + jy \mid x, y \in \mathbb{R},\ j = \sqrt{-1}\}$ with the usual scalar addition and multiplication.

Definition: A **vector space** (sometimes referred to as a **linear space**) over a field $\mathbb{K}$ is a set with two operations, vector addition and scalar multiplication, for which vector addition is commutative and scalar multiplication distributes over vector addition.

To summarize, a vector space V satisfies all of the following criteria. (See Appendix B in [2], [1], or any textbook on mathematical analysis.)

CLOSURE AXIOMS

1. $\boldsymbol{a} + \boldsymbol{b} \in V$ for all $\boldsymbol{a}, \boldsymbol{b} \in V$.

2. $\alpha\boldsymbol{a} \in V$ for all $\alpha \in \mathbb{R}$ and $\boldsymbol{a} \in V$.

AXIOMS FOR VECTOR ADDITION

3. *Commutativity:* $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{b} + \boldsymbol{a}$ for all $\boldsymbol{a}, \boldsymbol{b} \in V$.

4. *Associativity:* $(\boldsymbol{a} + \boldsymbol{b}) + \boldsymbol{c} = \boldsymbol{a} + (\boldsymbol{b} + \boldsymbol{c})$ for all $\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{c} \in V$.

5. *An additive identity:* $\mathbf{0} \in V$ such that $\boldsymbol{a} + \mathbf{0} = \mathbf{0} + \boldsymbol{a} = \boldsymbol{a}$ for all $\boldsymbol{a} \in V$.

6. *An additive inverse:* An element $-\boldsymbol{a} \in V$ for any $\boldsymbol{a} \in V$ such that $\boldsymbol{a} + (-\boldsymbol{a}) = (-\boldsymbol{a}) + \boldsymbol{a} = \mathbf{0}$.

AXIOMS FOR SCALAR MULTIPLICATION

7. *Associativity:* $\alpha(\beta\boldsymbol{a}) = (\alpha\beta)\boldsymbol{a}$ for all $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{a} \in V$.

8. *Distributive law for vector addition:* $\alpha(\boldsymbol{a} + \boldsymbol{b}) = \alpha\boldsymbol{a} + \alpha\boldsymbol{b}$ for all $\alpha \in \mathbb{R}$ and $\boldsymbol{a}, \boldsymbol{b} \in V$.

9. *Distributive law for scalar addition:* $(\alpha + \beta)\boldsymbol{a} = \alpha\boldsymbol{a} + \beta\boldsymbol{a}$ for all $\alpha, \beta \in \mathbb{R}$ and $\boldsymbol{a} \in V$.

10. *Scalar multiplicative identity:* $1 \in \mathbb{R}$ such that $1 \cdot \boldsymbol{a} = \boldsymbol{a} \cdot 1 = \boldsymbol{a}$ for all $\boldsymbol{a} \in V$.

Example 1: $\mathbb{R}^n$ over the field $\mathbb{R}$ for $n \in \{1, 2, \ldots\}$. This is just n-dimensional Euclidean space, the vector space of real n-vectors.

Example 2: $C[a, b]$ over the field $\mathbb{R}$, the vector space of continuous, real-valued functions on a closed interval $[a, b]$.

Definition: The elements $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ in a vector space V are said to be **linearly independent** if

$$\sum_{i=1}^{n} \alpha_i \boldsymbol{v}_i = 0 \Rightarrow \alpha_i = 0 \qquad \forall\ i \in \{1, \ldots, n\}.$$

Equivalently, the elements are linearly independent if no single element can be written as a linear combination of the others.

Definition: A **basis** for a vector space V is a set of linearly independent vectors $\{\bar{\boldsymbol{v}}_1, \bar{\boldsymbol{v}}_2, \ldots\}$ which spans the vector space. That is to say, any element $\boldsymbol{v} \in V$ can be written as a sum

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \bar{\boldsymbol{v}}_i.$$

If the basis is finite, then V is called **finite dimensional**. Otherwise, V is called **infinite dimensional**.

Example 1: The unit vectors

$$\{\boldsymbol{e_1}, \boldsymbol{e_2}, \boldsymbol{e_3}\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}$$

form a basis for the three-dimensional vector space $\mathbb{R}^3$.

Example 2:

$$\{\cos \frac{(n-1)\pi x}{l}, \sin \frac{n\pi x}{l}\} \qquad \text{for} \qquad n \in \{1, 2, \ldots\}$$

is a basis for $C[-l, l]$, the space of continuous functions on the interval $[-l, l]$. The vector space $C[-l, l]$ is infinite dimensional.

Definition: A **norm** $\|\cdot\|$ on a vector space V satisfies three properties

1. Positive definiteness:
$$\begin{array}{rclccrcl} \|\boldsymbol{v}\| & > & 0 & & \text{for all} & \boldsymbol{v} & \neq & \boldsymbol{0} \\ \|\boldsymbol{v}\| & = & 0 & & \text{when} & \boldsymbol{v} & = & \boldsymbol{0} \end{array}$$
2. $\|\alpha \boldsymbol{v}\| = |\alpha| \|\boldsymbol{v}\|$.
3. The *triangle inequality*: $\|\boldsymbol{v}_1 + \boldsymbol{v}_2\| \leq \|\boldsymbol{v}_1\| + \|\boldsymbol{v}_2\|$

A vector space V together with a norm $\|\cdot\|$ is known as a **normed vector space** or a **normed linear space**. One may often define a variety of norms for a given vector space. For example, for an n-dimensional vector space V (with $n < \infty$), one may define the p-norm

$$\|\boldsymbol{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}$$

for any finite scalar $p \geq 1$.

- The 1-norm is simply the sum of absolute values of the components of $\boldsymbol{v}$.
- The 2-norm is the conventional Euclidean norm measuring "straight line distance from the origin."
- In the limit $p \to \infty$, one obtains the ∞-norm. This is also known as the "max" norm because it returns the absolute value of the largest component of $\boldsymbol{v}$.

All norms on $\mathbb{R}^n$ are "equivalent" in the sense that there exist positive constants k_1 through k_4 such that, for any two norms $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on $\mathbb{R}^n$,

$$k_1 \|\boldsymbol{v}\|_\alpha \leq \|\boldsymbol{v}\|_\beta \leq k_2 \|\boldsymbol{v}\|_\alpha \qquad \text{and} \qquad k_3 \|\boldsymbol{v}\|_\beta \leq \|\boldsymbol{v}\|_\alpha \leq k_4 \|\boldsymbol{v}\|_\beta$$

for every $\boldsymbol{v} \in \mathbb{R}^n$. (See Appendix A in [2].) Briefly, this means that the comparative distance between points in $\mathbb{R}^n$ is independent of the distance measure used. (If point A is "closer" than point B to point C, as measured by the 2-norm, it is closer under any norm....) It does *not* mean that the measured distances are the same!

Definition: An **inner product** on a vector space V is an operation $< \cdot, \cdot >: \; V \times V \rightarrow \mathbb{R}$ satisfying

1. *Positivity:*
$$< \boldsymbol{v}, \boldsymbol{v} > \geq \; 0$$
with equality if and only if $\boldsymbol{v} = \mathbf{0}$.
2. *Bi-linearity:*
$$< \alpha \boldsymbol{v}, \boldsymbol{w} + \boldsymbol{x} > = \; \alpha < \boldsymbol{v}, \boldsymbol{w} > + \alpha < \boldsymbol{v}, \boldsymbol{x} > .$$
3. *Conjugate symmetry:*
$$< \boldsymbol{v}, \boldsymbol{w} > = \; \overline{< \boldsymbol{w}, \boldsymbol{v} >}.$$

Example: Consider the space $\mathbb{C}^n$ of n-dimensional complex vectors. One inner product on this space is

$$< \boldsymbol{v}, \boldsymbol{w} >= \overline{\boldsymbol{v}}^T \boldsymbol{w}$$

A vector space V with an associated inner product defines an **inner product space**. Note that properties (1) and (2) define a *norm* on V. Thus, any inner product space is a normed linear space. (But not vice versa!)

An important property of an inner product (what Strang [3] calls "probably the most important inequality in mathematics") is the **Cauchy-Schwarz inequality**.

Proposition: Consider a norm $\| \cdot \|$ induced by an inner product on a given vector space V. For any two elements $\boldsymbol{v}, \boldsymbol{w} \in V$,

$$| < \boldsymbol{v}, \boldsymbol{w} > | \; \leq \; \|\boldsymbol{v}\| \|\boldsymbol{w}\|.$$

Proof (for real vector spaces): The triangle inequality gives

$$\|\boldsymbol{v} + \boldsymbol{w}\| \leq \|\boldsymbol{v}\| + \|\boldsymbol{w}\|.$$

It follows that

$$\|\boldsymbol{v} + \boldsymbol{w}\|^2 \leq \|\boldsymbol{v}\|^2 + 2\|\boldsymbol{v}\|\|\boldsymbol{w}\| + \|\boldsymbol{w}\|^2. \tag{3}$$

But from bi-linearity and symmetry of the inner product, we have

$$\begin{aligned} \|\boldsymbol{v} + \boldsymbol{w}\|^2 &= \; < (\boldsymbol{v} + \boldsymbol{w}), (\boldsymbol{v} + \boldsymbol{w}) > \\ &= \; \|\boldsymbol{v}\|^2 + 2 < \boldsymbol{v}, \boldsymbol{w} > + \|\boldsymbol{w}\|^2 \end{aligned}$$

Substituting into (3), we find that

$$< \boldsymbol{v}, \boldsymbol{w} > \leq \|\boldsymbol{v}\| \|\boldsymbol{w}\|.$$

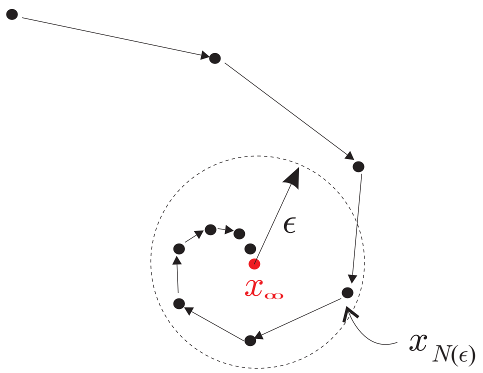
A similar series of arguments, starting with $\|\boldsymbol{v} - \boldsymbol{w}\|^2$ shows that

$$- < \boldsymbol{v}, \boldsymbol{w} > \leq \|\boldsymbol{v}\| \|\boldsymbol{w}\|.$$

The proposition follows immediately. □

Definition. A sequence of vectors $\{\boldsymbol{x}_n\}$ for $n = 1, 2, 3, \ldots$ in a normed linear space is said to **converge** to a particular element $\boldsymbol{x}_\infty$ if, for every $\epsilon > 0$, there is a number $N(\epsilon)$ such that

$$\|\boldsymbol{x}_n - \boldsymbol{x}_\infty\| < \epsilon$$



for all $n \geq N$.

Definition. A **Cauchy sequence** is one for which, given any $\epsilon > 0$, there is an $N(\epsilon)$ such that

$$\|\boldsymbol{x}_n - \boldsymbol{x}_m\| < \epsilon$$

for all $n, m \geq N$.

Example: Consider the sequence $\{x_n\}$ defined by setting $x_n = \frac{2n}{n+1}$:

$$\{x_n\} = \{1, \frac{4}{3}, \frac{3}{2}, \frac{8}{5}, \ldots\}.$$

This sequence converges to the number 2 as $n \to \infty$. To prove this, let $N(\epsilon) = \frac{2}{\epsilon}$. Then for $n > N$, we have

$$\left|\frac{2n}{n+1} - 2\right| = \left|-\frac{2}{n+1}\right| < \left|\frac{2}{n}\right| < \left|\frac{2}{N}\right| = |\epsilon| = \epsilon.$$

This sequence is also Cauchy. Given n and m,

$$\left|\frac{2n}{n+1} - \frac{2m}{m+1}\right| \leq \left|\frac{2n}{n+1}\right| + \left|\frac{2m}{m+1}\right|.$$

Let $N = \frac{4}{\epsilon}$ and apply the same series of arguments above to find that

$$\left|\frac{2n}{n+1} - \frac{2m}{m+1}\right| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

In fact, *every* convergent sequence is Cauchy. It is *not* true, however, that every Cauchy sequence converges! *If* every Cauchy sequence in a given normed linear space converges to a point in that space, then we call the space a **complete** normed linear space or a **Banach space**. (If every Cauchy sequence in a given inner product space converges, then we call the space a complete inner product space or a **Hilbert space**.)

Examples of Banach Spaces:

- $\mathbb{R}^n$ with any norm.
- $C^n[a, b]$ with

  $$\|\boldsymbol{x}(\cdot)\|_C = \max_{t \in [a,b]} \|\boldsymbol{x}(t)\|$$

  where $\|\cdot\|$ denotes any norm on $\mathbb{R}^n$.

- $L_p^n[a,b]$, the vector space of "p-integrable" functions on $[a,b]$, with

  $$\|\boldsymbol{f}(\cdot)\|_p = \left(\int_a^b \|\boldsymbol{f}(t)\|^p dt\right)^{\frac{1}{p}}$$

  where $\|\cdot\|$ denotes any norm on $\mathbb{R}^n$.

Example of an incomplete normed linear space: Given the vector space of rational (scalar) numbers, consider the following function:

$$f(x) = x^2 - 2$$

The Newton iteration sequence

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

clearly maps rational numbers to rational numbers. Moreover, assuming that the sequence starts sufficiently close to one of the two roots of $f(x)$, the sequence will converge to that root. However, the two roots ($\pm\sqrt{2}$) are irrational. Here, then, is a convergent (and therefore Cauchy) sequence which does not converge to a number in the vector space. The set of rational numbers is thus *not* a complete normed linear space.

Theorem (Banach Fixed Point Theorem) Let $(V, \|\cdot\|)$ be a Banach space and consider a closed subset $S \subset V$. Suppose that there is a map $T : S \to S$ such that

$$\|T(\boldsymbol{x}) - T(\boldsymbol{y})\| \leq \rho\|\boldsymbol{x} - \boldsymbol{y}\|$$

for all $\boldsymbol{x}, \boldsymbol{y} \in S$ with $\rho \in (0,1)$. (Such a map is called a **contraction map**.) Then there exists a unique fixed point $\boldsymbol{x}^* \in S$:

$$T(\boldsymbol{x}^*) = \boldsymbol{x}^*.$$

Proof: Define the sequence $\boldsymbol{x}_{n+1} = T(\boldsymbol{x}_n)$ by iteratively applying the map T. Then

$$\begin{aligned}
\|\boldsymbol{x}_{n+1} - \boldsymbol{x}_n\| = \|T(\boldsymbol{x}_n) - T(\boldsymbol{x}_{n-1})\| &\leq \rho\|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| \\
&\leq \rho^2\|\boldsymbol{x}_{n-1} - \boldsymbol{x}_{n-2}\| \\
&\vdots \\
&\leq \rho^n\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|
\end{aligned}$$

Now consider two points $\boldsymbol{x}_n$ and $\boldsymbol{x}_m$ in the sequence. Without loss of generality, assume that $n \geq m$ and observe that

$$\begin{aligned}
\|\boldsymbol{x}_n - \boldsymbol{x}_m\| &= \|\boldsymbol{x}_n - \boldsymbol{x}_{n-1} + \boldsymbol{x}_{n-1} - \boldsymbol{x}_{n-2} + \boldsymbol{x}_{n-2} - \cdots + \boldsymbol{x}_{m+1} - \boldsymbol{x}_m\| \\
&\leq \|\boldsymbol{x}_n - \boldsymbol{x}_{n-1}\| + \|\boldsymbol{x}_{n-1} - \boldsymbol{x}_{n-2}\| + \cdots + \|\boldsymbol{x}_{m+1} - \boldsymbol{x}_m\| \\
&\leq \rho^{n-1}\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| + \rho^{n-2}\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| + \cdots + \rho^m\|\boldsymbol{x}_1 - \boldsymbol{x}_0\| \\
&= \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|\rho^m \sum_{i=0}^{n-m-1} \rho^i \\
&\leq \|\boldsymbol{x}_1 - \boldsymbol{x}_0\|\frac{\rho^m}{1-\rho}.
\end{aligned}$$

The last inequality follows by recognizing the geometric series

$$\frac{1}{1-\rho} = 1 + \rho + \rho^2 + \cdots,$$

which converges for $|\rho| < 1$. Now,

$$\frac{\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|}{1-\rho}$$

is just a number – a constant. Because $\rho < 1$, the number ρ^m can be made arbitrarily small by choosing m large enough. Thus, for any $\epsilon > 0$, we can choose N large enough that for all $n \geq m \geq N$, we have $\|\boldsymbol{x}_n - \boldsymbol{x}_m\| < \epsilon$. The sequence is therefore Cauchy. Since *every* Cauchy sequence contained in a Banach space converges, *this particular* sequence converges and we call its limit $\boldsymbol{x}^*$. For any element $x_k = T(x_{k-1})$ in the sequence,

$$\begin{aligned} \|\boldsymbol{x}^* - T(\boldsymbol{x}^*)\| &= \|\boldsymbol{x}^* - \boldsymbol{x}_k + \boldsymbol{x}_k - T(\boldsymbol{x}^*)\| \\ &\leq \|\boldsymbol{x}^* - \boldsymbol{x}_k\| + \|\boldsymbol{x}_k - \boldsymbol{x}^*\| \\ &= \|\boldsymbol{x}^* - \boldsymbol{x}_k\| + \|T(\boldsymbol{x}_{k-1}) - T(\boldsymbol{x}^*)\| \\ &\leq \|\boldsymbol{x}^* - \boldsymbol{x}_k\| + \rho\|\boldsymbol{x}_{k-1} - \boldsymbol{x}^*\| \end{aligned}$$

In the limit that $k \to \infty$, the right-hand side goes to zero and we conclude that $T(\boldsymbol{x}^*) = \boldsymbol{x}^*$. That is, $\boldsymbol{x}^*$ is a fixed point of the map T. This fixed point is unique because any other fixed point $\boldsymbol{x}^{**}$ would satisfy

$$\|\boldsymbol{x}^* - \boldsymbol{x}^{**}\| = \|T(\boldsymbol{x}^*) - T(\boldsymbol{x}^{**})\| \leq \rho\|\boldsymbol{x}^* - \boldsymbol{x}^{**}\|.$$

Since $\rho < 1$, it follows that $\|\boldsymbol{x}^* - \boldsymbol{x}^{**}\| = 0$ which implies that $\boldsymbol{x}^{**} = \boldsymbol{x}^*$. $\square$

References

[1] T. M. Apostol. *Calculus: Volume II.* John Wiley & Sons, second edition, 1969.

[2] H. K. Khalil. *Nonlinear Systems.* Prentice-Hall, third edition, 2002.

[3] G. Strang. *Linear Algebra and Its Applications.* Saunders (Harcourt, Brace, Jovanovich) College Publishing, third edition, 1986.