ELE6202E - Multivariable Systems

Lecture 2: Mathematics Review

Fall 2026

Instructor: Bowen Yi

https://bwyi.github.io

POLYTECHNIQUE MONTRÉAL

Rudolf Kalman (1930-2016)

"Once you get the physics right, the rest is mathematics."

Rudolf Kalman

Contents

1. Why Control Engineering Needs Mathematics?

A Lesson from X -15 Flight 3-65-97 Disaster

Early adaptive-control designs could adjust gains online as flight conditions changed. In the 1960s, practical experience advanced faster than the tools available for rigorous stability and robustness analysis.

What can mathematics guarantee, and what assumptions are required for that guarantee?

Lessons Learned

Video: https://www.youtube.com/watch?v=2YIXwOfOiDI

- First hypersonic vehicles pushed the boundaries of aerospace technology and research. High-performance missions involving the NASA's X-15 in 1950s.
- On November 15, 1967, a fatal accident: Shortly after the aircraft reached its peak altitude, the X-15-3 began a sharp descent, and entered a Mach 5 spin. Although the pilot recovered from the spin, the adaptive controller began a limit cycle oscillation, which prevented it from reducing the pitch gain to the appropriate level. Consequently, the pilot was unable to pitch up, and the aircraft continued to dive. Encountering rapidly increasing dynamic pressures, the X-15-3 broke apart about 65,000 feet above sea level.

What can we learn from the X-15 disaster?

Lessons Learned

- It is certainly possible that, under an alternative set of flight conditions, one or more of the effects not included in the model might become significant.
- A sudden change in actuator effectiveness, which could have been caused by the electrical disturbance, causes the dynamics to depart significantly from those represented in the model and therefore in the control design.

The MH-96 (adaptive control system) lacked an analytically based proof of stability, which was highlighted by the fatal crash in 1967. After four decades, the theoretical groundwork for applying adaptive control has now made it possible to design adaptive controllers that offer high performance as well as stability guarantees in the presence of uncertainties.

2. Function and Map

Function and Map

Definition 1 (Cartesian Product) The Cartesian product of two given sets X and Y is defined as

$$X \times Y = \{(x, y) : x \in X, y \in Y\},$$

i.e., the set of all ordered pairs (x, y).

The set of all ordered n-tuple of real (complex) numbers is denoted by $\mathbb{R}^n$ ($\mathbb{C}^n$).

Definition 2 (Function and Map) Given two sets X and Y, by $f : X \to Y$ we mean that to every $x \in X$, f assigns a unique $f(x) \in Y$ called value of f at x. A more complete specification for a map is

$$f : X \to Y : x \mapsto f(x),$$

where $x \mapsto f(x)$ means that f sends $x \in X$ to $f(x) \in Y$.

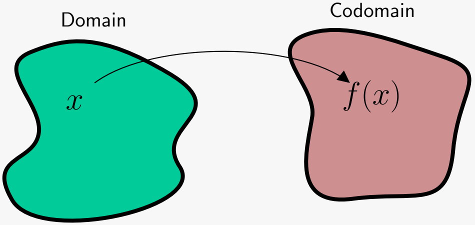


Figure 1: f maps X into Y

- Here X is the domain, and Y is the codomain;
- Function vs map: for the function, the codomain is $\mathbb{R}$ (or $\mathbb{C}$);
- The range of f:

$$f(X) := \{f(x)|x \in X\}.$$

3. Groups, Rings and Fields

Definition 3 (Group) A group $(G, *)$ is a set G with operation $*$

$$G \times G \to G$$
$$(x, y) \mapsto x * y \in G$$

satisfying the following properties:

1. Associativity: $\forall x, y, z \in G$, $x * (y * z) = (x * y) * z$;
2. Existence of a neutral element: $\exists e \in G$, s.t. $\forall x \in G$, $e * x = x * e = x$;
3. Existence of an inverse: $\forall x \in G$, $\exists x^{-1} \in G$, $x * x^{-1} = x^{-1} * x = e$.

If only (1) holds, we call it a semi-group.

Example 4

- $(\mathbb{Z}, +)$ with $e = 0, x^{-1} = -x$.
- Multiplicative group: $(\mathbb{Q} - \{0\}, \times)$ with $e = 1, x^{-1} = \frac{1}{x}$ (rational $\mathbb{Q}$)
- Not a group: $(\mathbb{N}, +)$

Definition 5 (Commutative/Abelian Group) Given a group $(G, *)$, if the law $*$ is commutative, i.e.

$$\forall x, y \in G,\ x * y = y * x,$$

then $(G, *)$ is called commutative or Abelian.

Definition 6 (Ring) $(R, *, \circ)$ is a ring if

1. $(R, *)$ is a commutative group [Addition];
2. $(R, \circ)$ is a semi-group (i.e. $\circ$ is internal and associate) [Multiplication];
3. Law $\circ$ is distributive, i.e.

$$x \circ (y * z) = (x \circ y) * (x \circ z)$$
$$(x * y) \circ z = (x \circ z) * (y \circ z).$$

If $\circ$ is commutative, we call it a commutative ring. It has an identity if $\circ$ has a neutral element. The neutral element of $(R, *)$ is called zero, and that of $(R, \circ)$ is called the identity.

Example. $\mathbb{Z}$ equipped with the usual operations of addition and multiplication $(\mathbb{Z}, +, \times)$ is a ring.

Definition 7 Field $(F, *, \circ)$ is a field if

1. $(F, +, \cdot)$ is a commutative ring with identity $1 \neq 0$;
2. every nonzero element $x \in F \setminus \{0\}$ has a multiplicative inverse, i.e., there exists $x^{-1} \in F$ such that

$$x \cdot x^{-1} = x^{-1} \cdot x = 1.$$

Example.

- $(\mathbb{Q}, +, \times)$ is a field.
- $(\mathbb{R}, +, \times)$ is also a field.
- $(\mathbb{R}^{n\times n}, +, \times)$ is] is *not* a field.

4. Vector Space

Vector Space (or Linear Space)

Definition 8 (Vector Space) A vector space $(V, \mathbb{F})$ is an object consisting of

- a set of vectors V,
- a field of scalars $\mathbb{F}$, and
- two binary operations: vector addition $+$ and scalar multiplication $\cdot$ (multiplication of vector by scalar) such that

1. Addition: $(V, +)$ is an abelian group.

   A1 Commutative: $x + y = y + x, \ \forall x, y \in V$
   A2 Associative: $(x + y) + z = x + (y + z), \ \forall x, y, z \in V$
   A3 $\exists \mathbf{0} \in V$: $x + \mathbf{0} = \mathbf{0} + x = x, \ \forall x \in V$ [zero vector]
   A4 $\exists$ inverse: $\forall x \in V, \exists$ element $(-x) \in V$ s.t. $x + (-x) = \mathbf{0}$.

(cont'd)

2. Scalar Multiplication is $\cdot : \mathbb{F} \times V \to V : (\alpha, x) \mapsto \alpha \cdot x$ and satisfies:

   SM1 $\alpha(\beta x) = (\alpha\beta)x, \ \forall \alpha, \beta \in \mathbb{F}$ and $\forall x \in V$,

   SM2 $1 \cdot x = x$ and $0 \cdot x = \mathbf{0}, \ \forall x \in V$,

   where 1 and 0 are the multiplicative and additive identities of $\mathbb{F}$, respectively.

3. Distributive laws:

$$(\alpha + \beta)x = \alpha x + \beta x, \quad \forall \alpha, \beta \in \mathbb{F}, \forall x \in V$$
$$\alpha(x + y) = \alpha x + \alpha y, \quad \forall \alpha \in \mathbb{F}, \forall x, y \in V.$$

Canonical Example. $(\mathbb{F}^n, \mathbb{F})$: the space of n-tuple in $\mathbb{F}$ over the field $\mathbb{F}$ is a vector space, e.g., $(\mathbb{R}^n, \mathbb{R})$ and $(\mathbb{C}^n, \mathbb{C})$.

Other Examples.

- $(\mathbb{R}^{n\times n}, \mathbb{R})$.
- $(C[0,T], \mathbb{R})$ is also a vector space over $\mathbb{R}$ (called real vector space), where $C[0,T]$ is the set of continuous scalar functions on $[0,T]$ with $\forall f, g \in C[0,T]$:

$$(f+g)(t) = f(t) + g(t) \qquad \text{(addition of two scalars)}$$
$$(\alpha \cdot f)(t) = \alpha f(t) \qquad \text{(multiplication of two scalars)}$$

- $(L^2[0,T], \mathbb{R})$ is another example, with

$$L^2[0,T] = \left\{ x(t) : \int_0^T |x(s)|^2 ds < +\infty \right\}.$$

The elements of a vector space are called vectors, regardless of their nature (matrices, functions or vectors in the usual geometric sense).

Definition 9 (Linear Independence and Dependence) Suppose that $(V, \mathbb{F})$ is a linear space. The set of vectors $\{v_1, v_2, \cdots, v_p\}$ where $v_i \in V$ is said to be linearly independent if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = \mathbf{0} \quad \Longrightarrow \quad \alpha_1 = \cdots = \alpha_p = 0,$$

where $\alpha_i \in \mathbb{F}$.

The set of vectors is said to be linearly dependent if there exist scalars $\alpha_1, \alpha_2, \cdots, \alpha_p \in \mathbb{F}$ not all zero such that

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_p v_p = \mathbf{0}.$$

Definition 10 (Basis) Suppose $(V, \mathbb{F})$ is a linear space. Then $B = \{b_1, b_2, \cdots, b_n\}$ is called a basis of V if

1. B spans V, i.e., $\forall x \in V$, there exist $\xi_1, \xi_2, \cdots, \xi_n \in \mathbb{F}$ such that
$$x = \xi_1 b_1 + \xi_2 b_2 + \cdots + \xi_n b_n;$$
2. B is a linearly independent set.

If a basis set is finite, V is said to be finite dimensional and

$$\text{dim } V = \#(B).$$

Remarks:

a) Any set of n linearly independent vectors of a vector space V of dimension n generates it; that is to say, it is a basis.

b) Any set of m vectors ($m > n$) of a vector space V of dimension n is necessarily dependent. If $\{v_1, v_2, \ldots, v_n\}$ is a basis for V, then for any element $v \in V$:

$$\forall v \in V, \exists! \alpha_1, \alpha_2, \ldots, \alpha_n \in F : v = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$$

can be expressed uniquely in this basis.

5. Linear Map

Linear Map

Definition 11 (Linear Map) L is a linear mapping if:

1. $\forall x, y \in V: \; L(x+y) = L(x) + L(y)$
2. $\forall a \in F: \; L(ax) = aL(x)$,

where, in condensed equivalent, if

$$L(ax + \beta y) = aL(x) + \beta L(y).$$

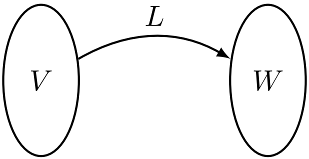


Figure 2: L maps V to W

Linear Map

Exercise. Show that if L is linear then $L(0_v) = 0_w$, with

- 0_v: Neutral Element of V
- 0_w: Neutral Element of W

Note: $L(V, W)$ represents the set of all linear maps from V to W. If $V = W$, we write $L(V)$.

Example 1: Applications in Control Systems

A system can generally be viewed as a mapping that converts the input signal into an output signal.

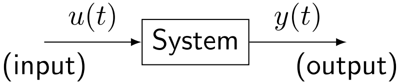
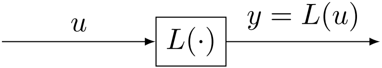


Figure 3: Representation of input/output system

In general, a system is multivarible in the sense that

$$u(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_m(t) \end{bmatrix}, \quad y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_p(t) \end{bmatrix}.$$

If the system is linear, we would have



Example 2: Finite-dimensional Vector Spaces

For *finite-dimensional* vector spaces, once bases are chosen, a linear map can be represented by a matrix.

Consider $V = \mathbb{F}^n$, $W = \mathbb{F}^m$,

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n,$$

and a linear map

$$L(x) = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \in \mathbb{F}^m,$$

where y_i is defined by

$$y_i = \sum_{j=1}^{n} a_{ij} x_j.$$

Now

$$L(x) = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \sum_{j=1}^{n} a_{2j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{mj} x_j \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Letting

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \in \mathbb{F}^{m \times n},$$

we have $y = Ax$.

Question: Does Example 1 belong to this case?

6. Matrix Representation

Matrix Representation

Linear Algebraic Equations. A matrix $A \in \mathbb{R}^{m \times n}$ is essentially a linear mapping from $\mathbb{F}^n$ to $\mathbb{F}^m$:

$$y = Ax.$$

We are interested in the solution of $x \in \mathbb{F}^n$, given $A \in \mathbb{F}^{m \times n}; b \in \mathbb{F}^m$ and the equation $Ax = b$, i.e.

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

with $A := \left[a_1 | \ldots | a_n\right]$, where a_i is the i-th column vector of A. This is related to whether b belongs to the sub-space (image) of $\mathbb{F}^m$ generated by $a_1, \ldots, a_n$.

Definition 12 (Range and Null Spaces) Given a linear map $A : U \to V$, define the range space of A as[a]

$$R(A) := \{v \in V \mid v = A(u), \text{ for some } u \in U\} \subset V;$$

and the null space of A as

$$N(A) := \{u \in U \mid A(u) = \mathbf{0}_v\} \subset U.$$

[a] Sometimes, we also write Im $(A) := R(A)$ and Ker $(A) := N(A)$.

Exercise. Prove that $R(A)$ and $N(A)$ are linear subspaces.

Example.

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} \in R\left(\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}\right) \text{ or not?}$$

Definition 13 (Rank and Nullity)

- By rank of a matrix $A \in \mathbb{F}^{m \times n}$, denoted by $\mathrm{rank}(A)$, we mean $\dim R(A)$;
- By nullity of a matrix $A \in \mathbb{F}^{m \times n}$, denoted by $\mathrm{nl}(A)$, we mean $\dim N(A)$.

We have

$$\mathrm{rank}\, A + \mathrm{nl}\, A = n.$$

If $m = n$,

$$\mathbb{F}^n = R(A^*) \oplus N(A).$$

$x \in N(A) \implies x$ is orthogonal to the columns of A^*. Therefore, $x \perp R(A^*)$, i.e.

$$N(A) = R^{\perp}(A^*).$$

Proof ($\star$)

Let $A : \mathbb{F}^n \to \mathbb{F}^m$. Since $N(A) \subset \mathbb{F}^n$, we can assume that

$$v_1, v_2, \ldots, v_k \in N(A) \quad (k \leq n)$$

form a basis for $N(A)$. Supplementing

$$v_{k+1}, \ldots, v_n \in \mathbb{F}^n$$

such that $v_1, v_2, \ldots, v_k, v_{k+1}, \ldots, v_n$ form a basis for $\mathbb{F}^n$.

Now we prove that $R(A)$ has a basis $Av_{k+1}, \ldots, Av_n$:

1. Assume not linearly independent. Then, $\exists$ not all zero scalars $\alpha_{k+1}, \ldots, \alpha_n$ s.t.

   $$\alpha_{k+1} Av_{k+1} + \ldots + \alpha_n Av_n = 0.$$

   Because A is a linear map, we have

   $$A(\alpha_{k+1} v_{k+1} + \cdots + \alpha_n v_n) = 0.$$

Now since $\alpha_{k+1}v_{k+1} + \ldots + \alpha_n v_n \in N(A)$, it can be represented by $v_1, \ldots, v_k$ as

$$\alpha_{k+1}v_{k+1} + \ldots + \alpha_n v_n = \alpha_1 v_1 + \ldots + \alpha_k v_k.$$

This contradicts the fact that $\{v_1, \ldots, v_n\}$ is a basis for $\mathbb{F}^n$. Thus, $Av_{k+1}, \ldots, Av_n$ are linearly independent.

2. For any $y \in R(A)$, $\exists v = \sum_{i=1}^{n} \alpha_i v_i \in \mathbb{F}^n$ s.t. $y = Av$. Then,

$$y = Av = A\left(\sum_{i=1}^{n} \alpha_i v_i\right) = \sum_{i=1}^{n} \alpha_i A v_i$$
$$= \sum_{i=1}^{k} \alpha_i A v_i + \sum_{i=k+1}^{n} \alpha_i A v_i = \sum_{i=k+1}^{n} \alpha_i A v_i,$$

that is, an arbitrary $y \in R(A)$ can be represented by a linear combination of Av_i $(i = k+1, \ldots, n)$. ■

Proposition 1 Let $A \in \mathbb{F}^{m \times n}$ be a matrix. Then

$$0 \leq \operatorname{rank} A \leq \min(m, n),$$

and $\operatorname{rank} A$ is equal to

- The max number of linearly independent column vectors of A;
- The max number of linearly independent row vectors of A;

Let $A \in \mathbb{F}^{m \times n}$ be a matrix.

- Row rank of A (denoted by row rank A): max number of linearly independent row vectors;
- Column rank of A (denoted by col rank A): max number of linearly independent column vectors.

Therefore,

$$\begin{aligned} \text{rank} A &= \text{row rank} A \\ &= \text{col rank} A. \end{aligned}$$

The matrix $A \in \mathbb{F}^{m \times n}$ is said to be of:

- Full row rank, if rank $A = m$;
- Full column rank, if rank $A = n$.

Examples

$$\text{rank}\ (1) = 1,\ \text{rank}\ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = 1,\ \text{rank}\ \begin{bmatrix} 1 & -2 & 0 \end{bmatrix} = 1$$

$$\text{rank}\ \begin{bmatrix} 1 & -\sqrt{2} \\ 0 & 0 \end{bmatrix} = 1$$

$$\text{rank}\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} = 2$$

$$\text{rank}\ \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix} = 2$$

Properties. Let $A \in \mathbb{F}^{m\times n}$ and $B \in \mathbb{F}^{n\times q}$.

- rank $A =$ rank A^* (i.e., number of independent columns = number of independent rows).
- rank $A +$ rank $B - n \leq$ rank $AB \leq$ min(rank A, rank B).

Returning to the equation $Ax = b$:

- If a solution exists, $b \in R(A)$.
- If $R(A) = \mathbb{F}^m$ (i.e., rank $A = m$), then a solution always exists.

  Requiring $m \leq n$ (fewer equations than unknowns).
- Special case: If $b = 0$, a solution still exists (trivial solution: $x = 0$). One might ask whether there are nontrivial solutions $x \neq 0$, which motivates the definition of null space.

Special Case: Square Matrix ($n = m$)

- If all column vectors are independent (i.e., rank$(A) = n$), A is a non-singular matrix. The solution is unique:

$$x = A^{-1}b.$$

- If not, a solution exists if and only if $b \in R(A) \subset \mathbb{F}^m$. If a solution exists, there are infinitely many solutions.
- If $\det(A) = 0$, then rank$(A) < n$, and

$$\text{dim } N(A) = n - \text{rank } (A) > 0 \quad \Longrightarrow \quad N(A) \neq \{0\}.$$

  We say that A has a nontrivial kernel.
- Any solution x can be expressed as

$$x = x_0 + v, \qquad v \in N(A)$$

  with x_0 a particular solution.

Eigenvalues and Eigenvectors

Proposition 2 Let $A \in \mathbb{C}^{n \times n}$ and let λ be a complex number. The following statements are equivalent:

1. $\chi_A(\lambda) = 0$, where
   $\chi_A(s) := \det(sI - A) = s^n + p_{n-1}s^{n-1} + \cdots + p_1 s + p_0$;
2. There exists a non-zero vector $v \in \mathbb{C}^n$ s.t. $Av = \lambda v$;
3. There exists a non-zero vector $u \in \mathbb{C}^n$ s.t. $u^* A = \lambda u^*$;

- Here λ is an eigenvalue of A, v is a right eigenvector, and u^* is a left eigenvector associated with λ.
- The spectrum of A is the set of its distinct eigenvalues, denoted by $\sigma(A) := \{\lambda_1, \ldots, \lambda_\ell\}$.

Diagonalization (Similarity Transformation)

If the eigenvalues of a matrix A, denoted as $\lambda_1, \ldots, \lambda_n$, are distinct, then the eigenvectors $v_1, \ldots, v_n$ are linearly independent, and A is diagonalizable. In other words, there exists a nonsingular matrix P such that

$$P^{-1}AP = \mathrm{diag}(\lambda) = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

where P is given by

$$P = \begin{pmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_n \\ | & | & & | \end{pmatrix}.$$

Property. Two similar matrices have the same spectrum, i.e. the eigenvalues are unaffected by such change of basis.

$$\lambda(P^{-1}AP) = \lambda(A).$$

- Not all matrices are diagonalizable.
  - A matrix $A \in \mathbb{C}^{n \times n}$ is diagonalizable if and only if it has n linearly independent eigenvectors.
  - Distinct eigenvalues are sufficient, but not necessary.
- Every complex square matrix nevertheless has a Jordan form.

Proposition 3 *(Jordan Form)* For a matrix $A \in \mathbb{C}^{n\times n}$, $\exists$ a nonsingular matrix $T \in \mathbb{C}^{n\times n}$ s.t. $A = TJT^{-1}$, where $T^*T = I$

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & J_q \end{bmatrix}, \quad J_i = \begin{bmatrix} J_i^1 & & & \\ & J_i^2 & & \\ & & \ddots & \\ & & & J_i^{\beta_i} \end{bmatrix} \in \mathbb{C}^{d_i\times d_i}$$

and each J_j $(j = 1, \ldots, q)$ is of the form

$$J_i^k = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_i \end{bmatrix},$$

where $\lambda_1, \ldots, \lambda_q$ are the eigenvalues of A. The characteristic polynomial of A is given by: $\chi_A(s) = (s-\lambda_1)^{d_1}(s-\lambda_2)^{d_2}\ldots(s-\lambda_\sigma)^{d_q}$.

Quadratic Form and Symmetric Matrix

Definition 14

- Any square matrix $A \in \mathbb{R}^{n \times n}$ determines a quadratic form

$$q_A(x_1, \ldots, x_n) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = x^\mathsf{T} A x.$$

- A matrix $Q \in \mathbb{R}^{n \times n}$ is called symmetric if

$$Q^\top = Q, \qquad \text{i.e. } Q_{ij} = Q_{ji}.$$

For any matrix $A \in \mathbb{R}^{n \times n}$, we have the decomposition

$$A = \underbrace{\frac{1}{2}(A + A^\top)}_{\text{symmetric part}} + \underbrace{\frac{1}{2}(A - A^\top)}_{\text{anti-symmetric part}} := A_s + A_{as}.$$

Since $x^\top A_{as} x = 0$, we have $x^\top A x = x^\top A_s x$.

Definition 15 (Positive Definite) We say that $Q^\top = Q$ is positive definite, denoted as $Q \succ 0$, if

$$x^\top Q x > 0, \qquad \forall x \in \mathbb{R}^n \setminus \{0\}.$$

Similarly, $Q^\top = Q$ is positive semidefinite, denoted by $Q \succeq 0$, if

$$x^\top Q x \geq 0, \qquad \forall x \in \mathbb{R}^n.$$

Proposition 4 Given a symmetric matrix $Q \in \mathbb{R}^{n \times n}$, we have

- eigenvalues of a symmetric matrix are always real;
- $Q \succ 0 \iff \lambda_i(Q) > 0$;
- $Q \succeq 0 \iff \lambda_i(Q) \geq 0$.

For a symmetric matrix, its eigenvalues can therefore be ordered from the smallest to the largest, defined as $\lambda_{\mathtt{min}}(Q)$ and $\lambda_{\mathtt{max}}(Q)$, respectively.

Rayleigh-Ritz Inequality

$$\lambda_{\mathtt{min}}(Q)x^\top x \leq x^\top Q x \leq \lambda_{\mathtt{max}}(Q)x^\top x, \qquad \forall x \in \mathbb{R}^n.$$

7. Normed Linear Space

Norms in Vector Spaces

Definition 16 (Norm on a Vector Space) A norm on a vector space V over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ is a function $\| \cdot \| : V \to \mathbb{R}$ satisfying:

1. $\|x\| \geq 0$ for all $x \in V$, and $\|x\| = 0$ if and only if $x = 0$.
2. $\|\lambda x\| = |\lambda| \cdot \|x\|$ for all $\lambda \in \mathbb{F}$ and $x \in V$.
3. $\|x + y\| \leq \|x\| + \|y\|$ for all $x, y \in V$.

Examples on $\mathbb{R}^n$.

- Euclidean norm: $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_n^2}$
- 1-norm: $\|x\|_1 = |x_1| + |x_2| + \ldots + |x_n|$
- p-norm $(1 \leq p < \infty)$:

$$\|x\|_p = (|x_1|^p + |x_2|^p + \ldots + |x_n|^p)^{\frac{1}{p}}$$

- ∞-norm:

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

- Examples for matrices $(\mathbb{F}^{m\times n}, \mathbb{F})$
    1. $||A||_a = \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$, entrywise absolute sum norm;
    2. $||A||_F = \left(\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2\right)^{\frac{1}{2}} = \sqrt{\mathsf{Tr}(AA^\dagger)}$, Frobenius norm;
    3. $||A||_b = \mathsf{max}_{i,j}|a_{ij}|$.

- Examples for functions $f \in C([t_0, t_1], \mathbb{F}^n)$
    1. $||f||_1 = \int_{t_0}^{t_1} \|f(t)\| dt,$, $\|f(t)\|$ can be any norm in $(\mathbb{F}^n, \mathbb{F}), L_1$;
    2. $||f||_2 = \left(\int_{t_0}^{t_1} \|f(t)\|^2 dt\right)^{\frac{1}{2}}$, L_2;
    3. $||f||_\infty = \mathsf{max}_{[t_0,t_1]}\|f(t)\|$, L_∞.

Definition 17 (Equivalent Norms) Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $(V, \mathbb{F})$ are said to be equivalent if $\exists\, m_l > 0$ and $m_r > 0$ such that

$$m_l\|v\|_a \leq \|v\|_b \leq m_r\|v\|_a.$$

Proposition 5

For $(\mathbb{F}^n, \mathbb{F})$, show that

1. $\|x\|_\infty \leq \|x\|_1 \leq n\|x\|_\infty$;
2. $\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n}\|x\|_\infty$;
3. $\frac{1}{\sqrt{n}}\|x\|_1 \leq \|x\|_2 \leq \|x\|_1$.

Proposition 6 We conclude that l_1, l_2, and l_∞ norms are all equivalent.

Induced Norm

Definition 18 (Induced norm) Let $\mathcal{A} : (U, \mathbb{F}) \to (V, \mathbb{F})$ be a linear operator. Let U and V be endowed with norms $\|\cdot\|_U$ and $\|\cdot\|_V$, respectively. Then the induced norm of $\mathcal{A}$ is defined by

$$\|\mathcal{A}\|_{\mathrm{i}} = \sup_{u \neq 0} \frac{\|\mathcal{A}u\|_V}{\|u\|_U} = \sup_{\|u\|=1} \|\mathcal{A}u\|_V.$$

Theorem 19 *Let $(U, \|\cdot\|_U)$, $(V, \|\cdot\|_V)$, and $(W, \|\cdot\|_W)$ be normed linear spaces, and $\mathcal{A} : V \to W$, $\mathcal{B} : V \to W$, and $\mathcal{C} : U \to V$. Then,*

1. $\|\mathcal{A}v\|_W \leq \|\mathcal{A}\|_{\mathrm{i}} \cdot \|v\|_V$;
2. $\|\alpha\mathcal{A}\|_{\mathrm{i}} = |\alpha| \, \|\mathcal{A}\|_{\mathrm{i}}, \qquad \forall\, \alpha \in \mathbb{F}$;
3. $\|\mathcal{A} + \mathcal{B}\|_{\mathrm{i}} \leq \|\mathcal{A}\|_{\mathrm{i}} + \|\mathcal{B}\|_{\mathrm{i}}$;
4. $\|\mathcal{A}\|_{\mathrm{i}} = 0 \quad \iff \quad \mathcal{A} = 0$;
5. $\|\mathcal{A}\mathcal{C}\|_{\mathrm{i}} \leq \|\mathcal{A}\|_{\mathrm{i}} \, \|\mathcal{C}\|_{\mathrm{i}}$.

Proof ($\star$)

1. $\|\mathcal{A}v\|_W \leq \|\mathcal{A}\|_{\mathrm{i}} \cdot \|v\|_V$:

$$\|\mathcal{A}\|_{\mathrm{i}} = \sup_{x \neq 0} \frac{\|\mathcal{A}x\|_W}{\|x\|_V} \geq \frac{\|\mathcal{A}v\|_W}{\|v\|_V}.$$

Therefore

$$\|\mathcal{A}v\|_W \leq \|\mathcal{A}\|_{\mathrm{i}} \cdot \|v\|_V.$$

3. $\|\mathcal{A} + \mathcal{B}\|_{\mathrm{i}} \leq \|\mathcal{A}\|_{\mathrm{i}} + \|\mathcal{B}\|_{\mathrm{i}}$:

$$\begin{aligned}
\|\mathcal{A} + \mathcal{B}\|_{\mathrm{i}} &= \sup_{\|x\|=1} \|(\mathcal{A} + \mathcal{B})x\|_W \\
&\leq \sup_{\|x\|=1} \left(\|\mathcal{A}x\|_W + \|\mathcal{B}x\|_W\right) \\
&\leq \sup_{\|x\|=1} \|\mathcal{A}x\|_W + \sup_{\|x\|=1} \|\mathcal{B}x\|_W \\
&= \|\mathcal{A}\|_{\mathrm{i}} + \|\mathcal{B}\|_{\mathrm{i}}.
\end{aligned}$$

5. $\|\mathcal{AC}\|_\mathrm{i} \le \|\mathcal{A}\|_\mathrm{i} \cdot \|\mathcal{C}\|_\mathrm{i}$:
Suppose that $u^* = \arg\sup \frac{\|\mathcal{AC}u\|_W}{\|u\|_U}$. We then have

$$\|\mathcal{AC}\|_\mathrm{i} = \sup_{u \neq 0} \frac{\|\mathcal{AC}u\|_W}{\|u\|_U} = \frac{\|\mathcal{AC}u^*\|_W}{\|u^*\|_U}.$$

From 1), we have

$$\|\mathcal{AC}u^*\|_W \le \|\mathcal{A}\|_\mathrm{i} \cdot \|\mathcal{C}u^*\|_V.$$

Then,

$$\begin{aligned} \|\mathcal{AC}\|_\mathrm{i} &\le \frac{\|\mathcal{A}\|_\mathrm{i} \cdot \|\mathcal{C}u^*\|_V}{\|u^*\|_U} \le \|\mathcal{A}\|_\mathrm{i} \sup_{u \neq 0} \frac{\|\mathcal{C}u\|_V}{\|u\|_U} \\ &= \|\mathcal{A}\|_\mathrm{i} \cdot \|\mathcal{C}\|_\mathrm{i}. \end{aligned}$$

■

Theorem 20 *Consider $A : \mathbb{F}^n \to \mathbb{F}^m$. Define*

$$\|A\|_{p,\mathrm{i}} = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}.$$

1. $\|A\|_{1,\mathrm{i}} =$ *max*$_j \{\sum_{i=1}^m |a_{ij}|\}$ – *max column sum;*
2. $\|A\|_{2,\mathrm{i}} =$ *max*$_j \sqrt{\lambda_j(A^\dagger A)}$ – *max singular value;*
3. $\|A\|_{\infty,\mathrm{i}} =$ *max*$_i \{\sum_{j=1}^n |a_{ij}|\}$ – *max row sum.*

As a comparison:

1. $\|A\|_a := \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|$;
2. $\|A\|_F := \sqrt{AA^\dagger}$ – Frobenius norm;
3. $\|A\|_b := \max_{i,j} |a_{ij}|$.

Proof ($\star$)

1) From definition, we have

$$\|\mathcal{A}\|_{1,\mathrm{i}} = \sup_{x\neq 0} \frac{\|\mathcal{A}x\|_1}{\|x\|_1}.$$

Then

$$\begin{aligned}\|\mathcal{A}x\|_1 &= \sum_{i=1}^{m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|\,|x_j| = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} |a_{ij}| \right) |x_j| \\ &\leq \left(\mathsf{max}_j \sum_{i=1}^{m} |a_{ij}| \right) \sum_{j=1}^{n} |x_j| = \left(\mathsf{max}_j \sum_{i=1}^{m} |a_{ij}| \right) \|x\|_1,\end{aligned}$$

and

$$\frac{\|\mathcal{A}x\|_1}{\|x\|_1} \leq \mathsf{max}_j \sum_{i=1}^{m} |a_{ij}|.$$

Find j^* such that

$$\max_j \sum_{i=1}^{m} |a_{ij}| = \sum_{i=1}^{m} |a_{ij^*}|.$$

Let $x = e_{j^*}$ (j^*-th column in the identity matrix), we have

$$\|\mathcal{A}x\|_1 = \|\mathcal{A}e_{j^*}\|_1 = \sum_{i=1}^{m} |a_{ij^*}|,$$

and

$$\|x\|_1 = \sum_{j=1}^{n} |x_j| = 1.$$

2) From definition, we have

$$\|\mathcal{A}\|_{2,\mathrm{i}} = \sup_{x \neq 0} \frac{\|\mathcal{A}x\|_2}{\|x\|_2},$$

where

$$\|x\|_2 = \sqrt{x^\dagger x} \qquad \text{and} \qquad \|\mathcal{A}x\|_2 = \sqrt{x^\dagger \mathcal{A}^\dagger \mathcal{A} x}.$$

For the Hermitian matrix $\mathcal{A}^\dagger \mathcal{A}$, $\exists$ a unitary matrix U $(UU^\dagger = I)$ such that

$$\mathcal{A}^\dagger \mathcal{A} = U^\dagger \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} U, \quad \text{where} \quad \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix},$$

and

$$0 < \lambda_r \leq \cdots \leq \lambda_1.$$

Therefore,

$$\left[\frac{\|\mathcal{A}x\|_2}{\|x\|_2}\right]^2 = \frac{x^\dagger U^\dagger \begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix} Ux}{x^\dagger x} \overset{y=Ux}{=} \frac{\sum_{j=1}^{r} \lambda_j |y_j|^2}{y^\dagger y} \leq \lambda_1.$$

We can choose x such that $y = Ux = e_1 = [1, 0, \ldots, 0]^\top$, which in turn yields that

$$\|\mathcal{A}x\|_2 = \sqrt{x^\dagger U^\dagger \begin{bmatrix}\Lambda & 0\\ 0 & 0\end{bmatrix} Ux} = \sqrt{\lambda_1},$$

and

$$\|x\|_2 = \sqrt{x^\dagger x} = \sqrt{e_1^\dagger U U^\dagger e_1} = \sqrt{e_1^\dagger e_1} = 1.$$

3) From definition, we have

$$\|\mathcal{A}\|_{\infty,\mathrm{i}} = \sup_{x\neq 0} \frac{\|\mathcal{A}x\|_\infty}{\|x\|_\infty}.$$

Then

$$\|\mathcal{A}x\|_\infty = \mathsf{max}_i \left|\sum_{j=1}^{n} a_{ij}x_j\right| \leq \mathsf{max}_i \sum_{j=1}^{n} |a_{ij}|\,|x_j|$$
$$\leq \mathsf{max}_i \sum_{j=1}^{n} |a_{ij}| \cdot \mathsf{max}_j |x_j| = \mathsf{max}_i \sum_{j=1}^{n} |a_{ij}| \cdot \|x\|_\infty,$$

and

$$\frac{\|\mathcal{A}x\|_\infty}{\|x\|_\infty} \leq \mathsf{max}_i \sum_{j=1}^{n} |a_{ij}|.$$

8. Inner Product Space

Inner Product in Linear Spaces

Definition 21 Consider a linear space $(H, \mathbb{F})$, with $\mathbb{F}$ either $\mathbb{R}$ or $\mathbb{C}$. The function

$$\begin{aligned} \langle \cdot, \cdot \rangle : & H \times H \to \mathbb{F} \\ & (x, y) \mapsto \langle x, y \rangle \end{aligned}$$

is called an inner product if

1. $\langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle$ for all $x, z, y \in H$;
2. $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$, for all $\alpha \in \mathbb{F}$;
3. $\langle x, x \rangle \geq 0$, with equality if and only if $x = \mathbf{0}_H$;
4. $\langle x, y \rangle = \overline{\langle y, x \rangle}$, where $\overline{(\cdot)}$ denotes complex conjugation.

A linear space equipped with an inner product is an inner product space. Its induced norm is $\|x\| = \sqrt{\langle x, x \rangle}$. With the convention above, the second argument is conjugate-linear: $\langle x, \alpha y \rangle = \overline{\alpha} \langle x, y \rangle$.

Examples:

- For $\mathbb{R}^n$: $\langle x, y\rangle = x_1 y_1 + \ldots + x_n y_n$
- For $L^2([0,T],\mathbb{C})$: $\langle f, g\rangle = \frac{1}{T}\int_0^T f(t)\overline{g(t)}\,dt$

- Note that for any scalar product, we can define the standard (induced) norm:
$$\|v\| = \sqrt{\langle v, v\rangle}.$$
- The definition of the standard norm allows us to discuss continuity, limits, and convergence of scalar product sequences because:
$$\langle x, y\rangle = \|x\|\|y\|\cos(\angle(x,y))$$
- We say that x is orthogonal to y (i.e. $x \perp y$) if $\langle x, y\rangle = 0$.
- If V is a real vector space, and W is sub-space of V, then we define
$$W^\perp := \{x \in V : x \perp z,\ \forall z \in W\}.$$

Exercise

Let V be a finite-dimensional inner product space. Show that every subspace W gives the orthogonal decomposition $V = W \oplus W^{\perp}$.

Note: $V = W \oplus W^{\perp}$. This means that every element of V can be decomposed uniquely as follows:

$$\forall z \in V,\ \exists! x \in W, y \in W^{\perp}:\ z = x + y.$$

Exercise. Show that if

1. $\{v_1, \ldots, v_n\}$ is an orthogonal system (i.e. $v_i \perp v_j;\ i \neq j$), then it is a basic of $\mathbb{R}^n$.
2. $\{v_1, \ldots, v_n\}$ is a system $\perp$, then components of any vector $x \in \mathbb{R}^n$ in this basis are given by

$$\alpha_i = \frac{\langle x, v_i \rangle}{\|v_i\|^2}.$$

Schwarz's Inequality

Theorem 22 *Let $(H, \mathbb{F}, \langle \cdot, \cdot \rangle)$ be an inner prodcut space. We have*

$$|\langle x, y\rangle| \leq \sqrt{\langle x, x\rangle} \cdot \sqrt{\langle y, y\rangle} = \|x\| \cdot \|y\|, \quad \forall x, y \in H.$$

Proof.

Choose $\alpha \in \mathbb{F}$ with $|\alpha| = 1$ such that

$$\alpha\langle x, y\rangle = |\langle x, y\rangle|.$$

Then for all $\lambda \in \mathbb{R}$, we obtain that

$$\begin{aligned} 0 &\leq \|\lambda x + \alpha y\|^2 = \langle \lambda x + \alpha y, \lambda x + \alpha y\rangle \\ &= \langle x, x\rangle \lambda^2 + |\alpha|^2 \langle y, y\rangle + \lambda\bar{\alpha}\langle y, x\rangle + \bar{\lambda}\alpha\langle x, y\rangle \\ &= \|x\|^2\lambda^2 + 2|\langle x, y\rangle|\lambda + \|y\|^2. \end{aligned}$$

■

9. Hilbert Space

Hilbert Space

Definition 23 An inner product space $(H, \mathbb{F}, \langle \cdot, \cdot \rangle)$ that is complete in the norm defined by the inner product is called a Hilbert space.[a]

[a] Here complete means that very Cauchy sequence in H converges in H.

Examples

- $(\mathbb{F}^n, \mathbb{F}, \langle \cdot, \cdot \rangle)$ under the inner product $\langle x, y \rangle = \sum_{i=1}^{n} \bar{x}_i y_i =: x^\dagger y$ (Hermitian transpose $\dagger$)
- $L_2([t_0, t_1], \mathbb{F}^n)$ – space of square integrable $\mathbb{F}^n$-valued functions on $[t_0, t_1]$ under the inner product

$$\langle f, g \rangle := \int_{t_0}^{t_1} f^\dagger(t) g(t) dt$$

Adjoint

Definition 24 (Adjoint) Let $(U, \mathbb{F}, \langle \cdot, \cdot \rangle)$ and $(V, \mathbb{F}, \langle \cdot, \cdot \rangle)$ be Hilbert spaces. Let $\mathcal{A} : U \to V$ be continuous and linear. Then the adjoint of $\mathcal{A}$, denoted as $\mathcal{A}^*$, is the map $\mathcal{A}^* : V \to U$ such that

$$\langle v, \mathcal{A}u \rangle_V = \langle \mathcal{A}^* v, u \rangle_U .$$

Example

Let $g(\cdot) \in C([t_0, t_1], \mathbb{R}^n)$ and define

$$\mathcal{A} : C([t_0, t_1], \mathbb{R}^n) \to \mathbb{R} : f(\cdot) \mapsto \langle g(\cdot), f(\cdot) \rangle.$$

Find the adjoint map of $\mathcal{A}$.

Solution

From definition, we know that $\mathcal{A}^* : \mathbb{R} \to C([t_0, t_1], \mathbb{R}^n)$ such that

$$\langle v, \mathcal{A} f(\cdot) \rangle_{\mathbb{R}} = \langle \mathcal{A}^* v, f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)},$$

where $v \in \mathbb{R}$ and $f(\cdot) \in C([t_0, t_1], \mathbb{R}^n)$. We then have

$$\begin{aligned} \langle v, \mathcal{A} f(\cdot) \rangle_{\mathbb{R}} &= \left\langle v, \langle g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)} \right\rangle_{\mathbb{R}} \\ &= v \langle g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)} = \langle v g(\cdot), f(\cdot) \rangle_{C([t_0, t_1], \mathbb{R}^n)}. \end{aligned}$$

Comparing the above equations, we have $\mathcal{A}^* : v \mapsto v g(\cdot)$.

10. Laplace and z Transforms

One-Sided Laplace Transform

Following [Kailath, Sec. 1.2], we will use the one-sided Laplace transform to help solve time-continuous ODEs, which is defined by

$$X_-(s) = \mathcal{L}_-[x(t)] := \int_{0^-}^{\infty} x(t)e^{-st}dt.$$

For a signal of exponential order, the region of convergence is a right half-plane $\{s : \mathrm{Re}(s) > \alpha\}$; convergence on the boundary must be checked separately.

The inverse is given by the contour integral

$$x(t) = \mathcal{L}_-^{-1}[X(s)] = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{c-iT}^{c+iT} X(s)e^{st}\, ds, \qquad c > \alpha.$$

Table of Laplace Transform

Table 1: Short list of Laplace and z transforms.

$f(t), t \geq 0$	$F(s)$	$f(k), k \geq 0$	$F(z)$
$\delta(t)$	1	$\delta(k)$	1
1	$\frac{1}{s}$	1	$\frac{z}{z-1}$
t	$\frac{1}{s^2}$	k	$\frac{z}{(z-1)^2}$
$\frac{t^{q-1}}{(q-1)!}$	$\frac{1}{s^q}$	$\binom{k}{r-1}$	$\frac{z}{(z-1)^r}$
$e^{\lambda t}$	$\frac{1}{s-\lambda}$	λ^k	$\frac{z}{z-\lambda}$
$\frac{t^{q-1}}{(q-1)!}e^{\lambda t}$	$\frac{1}{(s-\lambda)^q}$	$\binom{k}{r-1}\lambda^{k+1-r}$	$\frac{z}{(z-\lambda)^r}$

$$\mathcal{L}_{-}[\dot{x}(t)] = sX_{-}(s) - x(0_{-}).$$

Z Transform for Discrete-Time Signals ($\star$)

If $f_k := f[k] \in \mathbb{R}^n$, $k \geq 0$, we define the Z-transform as

$$\mathcal{Z}\{f\}(z) = \sum_{k=0}^{\infty} f_k z^{-k}, \quad \text{for } z \in \mathbb{C}.$$

If f_k is of exponential order, its region of convergence contains the exterior of a circle: $|z| > \beta$ for some $\beta \geq 0$.

Example. $\delta_k = 1$ if $k = 0, 0$ otherwise.

$$\mathcal{Z}\{\delta\}(z) = 1$$

Properties

$$\mathcal{Z}\left\{f_{k+1}\right\}(z) = z\mathcal{Z}\{f_k\}(z) - zf(0)$$

$$\mathcal{Z}\left\{f_{k-1}\right\}(z) = \frac{1}{z}\mathcal{Z}\{f_k\}. \quad \text{if } f_k = 0 \text{ for } k < 0.$$

Proof.

$$\sum_{k=0}^{\infty} f_{k+1} z^{-k} = \sum_{k=0}^{\infty} f_{k+1} z^{-(k+1)} z = z \sum_{k=1}^{\infty} f_k z^{-k}$$

$$= z(\mathcal{Z}\{f_k\} - f(0)).$$

$$\sum_{k=0}^{\infty} f_{k-1} z^{-k} = \sum_{j=-1}^{\infty} f_j z^{-(j+1)} = \frac{1}{z} \sum_{j=0}^{\infty} f_j z^{-j} = \frac{1}{z}\mathcal{Z}\{f_k\},$$

■

11. Reading Materials ⋆

Matrix-Valued Functions

Let $I \subseteq \mathbb{R}$ be a time interval. A time-varying vector or matrix can be viewed as a map

$$x : I \to \mathbb{R}^n, \qquad A : I \to \mathbb{R}^{m\times n}.$$

Most vector and matrix concepts are interpreted pointwise in time:

- $\|x(t)\|$ is the Euclidean norm of $x(t)$ at time t;
- $A(t)$ is invertible if $A^{-1}(t)$ exists at that time;
- $\lambda_i(A(t))$ and $\|A(t)\|$ are scalar functions of time;
- for symmetric $Q(t)$,

  $$Q(t) \succ 0$$

  means that $Q(t)$ is positive definite for every $t \in I$.

The statement "$A(t)$ is invertible $\forall t$" concerns each matrix $A(t)$ individually. It does *not* mean that the map $A : I \to \mathbb{R}^{n\times n}$ is invertible.

Matrix Calculus

- Differentiation and integration of a matrix-valued function are defined entry by entry:

$$\frac{d}{dt}A(t) = \left[\frac{d}{dt}a_{ij}(t)\right], \qquad \int_{t_0}^{t} A(\sigma)\,d\sigma = \left[\int_{t_0}^{t} a_{ij}(\sigma)\,d\sigma\right].$$

- Product rule remains valid: $\frac{d}{dt}\big[A(t)B(t)\big] = \dot{A}(t)B(t) + A(t)\dot{B}(t)$.
- For time-varying integration limits, the Leibniz rule becomes

$$\frac{d}{dt}\int_{f(t)}^{g(t)} A(t,\sigma)\,d\sigma = A\big(t, g(t)\big)\dot{g}(t) - A\big(t, f(t)\big)\dot{f}(t) + \int_{f(t)}^{g(t)} \frac{\partial A}{\partial t}(t,\sigma)\,d\sigma.$$

Caution: matrices need not commute. In general,

$$\frac{d}{dt}A^2(t) = \dot{A}(t)A(t) + A(t)\dot{A}(t) \neq 2A(t)\dot{A}(t).$$

Convergence

Definition 25 (Convergence of a vector sequence) A sequence $\{x_k\}_{k=0}^{\infty} \subset \mathbb{R}^n$ converges to $x^\star \in \mathbb{R}^n$ if, for every $\varepsilon > 0$, there exists an integer $K(\varepsilon)$ such that

$$\|x_k - x^\star\| < \varepsilon, \qquad k \geq K(\varepsilon).$$

We write $\lim_{k\to\infty} x_k = x^\star$.

The norm reduces vector convergence to scalar convergence:

$$x_k \to x^\star \quad \Longleftrightarrow \quad \|x_k - x^\star\| \to 0.$$

For a sequence of functions

$$x_k : [t_0, t_1] \to \mathbb{R}^n,$$

two notions of convergence are important.

Pointwise vs. Uniform Convergence

Pointwise convergence: for every fixed $t \in [t_0, t_1]$ and every $\varepsilon > 0$, there exists $K(\varepsilon, t)$ such that

$$\|x_k(t) - x^\star(t)\| < \varepsilon, \qquad k \geq K(\varepsilon, t).$$

The index K may depend on the selected time t.

Uniform convergence The sequence $\{x_k\}$ converges uniformly to $x^\star$ on $[t_0, t_1]$ if, for every $\varepsilon > 0$, there exists an integer $K(\varepsilon)$ such that

$$\|x_k(t) - x^\star(t)\| < \varepsilon$$

for every $t \in [t_0, t_1]$ and every $k \geq K(\varepsilon)$.

Equivalently,

$$\sup_{t \in [t_0, t_1]} \|x_k(t) - x^\star(t)\| \longrightarrow 0.$$

The key distinction is the order of the quantifiers.

Pointwise: $\forall t,\ \forall \varepsilon > 0,\ \exists K(\varepsilon, t),$

Uniform: $\forall \varepsilon > 0,\ \exists K(\varepsilon)$ such that the estimate holds for all t.

For uniform convergence, the *same* $K(\varepsilon)$ works over the entire interval.

Example. Consider

$$x_k(t) = t^k, \qquad t \in [0, 1].$$

Then $x_k(t)$ converges pointwise, but not uniformly, because the convergence becomes arbitrarily slow near $t = 1$.

Series of Vector-Valued Functions

Consider a series of vector-valued functions

$$\sum_{j=0}^{\infty} x_j(t), \qquad x_j : [t_0, t_1] \to \mathbb{R}^n,$$

with partial sums $s_k(t) = \sum_{j=0}^{k} x_j(t)$. The series converges pointwise or uniformly when the sequence $\{s_k\}$ converges pointwise or uniformly.

Theorem 26 *(Uniform limit theorem) Suppose that every s_k is continuous on $[t_0, t_1]$ and that*

$$s_k \longrightarrow x^\star$$

uniformly on $[t_0, t_1]$. Then $x^\star$ is continuous on $[t_0, t_1]$.

- Uniform convergence allows properties of the approximating functions to pass to the limit. Pointwise convergence alone is not sufficient.
- Uniform convergence does *not*, by itself, justify term-by-term differentiation. Additional conditions on the derivative series are required.

Differentiation and the Weierstrass M-Test

Theorem 27 *(Term-by-term differentiation) Suppose that every x_j is continuously differentiable on $[t_0, t_1]$, and that $\sum_{j=0}^{\infty} x_j(t)$ converges uniformly to $x^\star(t)$. If the derivative series $\sum_{j=0}^{\infty} \dot{x}_j(t)$ also converges uniformly, then $\dot{x}^\star(t) = \sum_{j=0}^{\infty} \dot{x}_j(t)$.*

Theorem 28 *(Weierstrass M-test) Suppose that there exist constants $\alpha_j \geq 0$ such that*

$$\|x_j(t)\| \leq \alpha_j, \ \forall t \in [t_0, t_1],$$

and that $\sum_{j=0}^{\infty} \alpha_j < \infty$. Then

$$\sum_{j=0}^{\infty} x_j(t)$$

converges uniformly and absolutely on $[t_0, t_1]$.

Sources and Further Reading

- Algebraic foundations and linear maps: The definitions of sets and maps, groups, rings, fields, vector spaces, bases, and linear maps draw primarily on the mathematical preliminaries in (Callier and Desoer).
- Matrix theory: The material follows primarily (Callier and Desoer) and Rugh.
- Normed and inner-product spaces: It draws mainly on the mathematical appendices of (Callier and Desoer).
- Laplace and z transforms: The one-sided Laplace-transform convention follows Kailath, Sec. 1.2, while the basic Laplace- and z-transform identities are also consistent with Rugh, pp. 16.
- The X-15 case study: The adaptive-control perspective is based on Dydek, Annaswamy, and Lavretsky, while the accident reconstruction follows NASA's later NESC assessment.[1]

Selected proofs, exercises, and control-oriented interpretations were adapted and reorganized by the instructor.

[1] Z. T. Dydek, A. M. Annaswamy, and E. Lavretsky, "Adaptive Control and the NASA X-15-3 Flight Revisited," *IEEE Control Systems Magazine*, vol. 30, no. 3, pp. 32–48, 2010; NASA Engineering and Safety Center, *A Comprehensive Analysis of the X-15 Flight 3-65 Accident*, NESC-RP-14-00957, 2014.