

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 2.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Metric Spaces

Topology

Vector Spaces

Normed Spaces

Inner Product Spaces

Notions of Convergence

Orthogonality

Linear Operators

Projections

Gram Schmidt Orthogonalization

## Section 1

### Metric Spaces

# Spaces

- ▶ One of the objectives of this course is to develop tools that work in a wide variety of settings.
- ▶ We will mostly focus on finite dimensional Hilbert spaces, which include:
  - ▶  $\mathbb{R}^n, \mathbb{C}^n, \mathbb{C}^{m \times n}$ ,
  - ▶ the set of all functions with finite integral,
  - ▶ the set of all finitely summable sequences,
  - ▶ binary vectors, binary sequences.
- ▶ But does not include important objects like
  - ▶ rotations matrices, quaternions, homogeneous transformations.
- ▶ To make things clear, we will develop the theory systematically in the following order:
  1. Metric space
  2. Norm space / Banach space
  3. Inner product space / Hilbert space

# Metric Spaces

## Definition (Metric Space)

A metric space is a pair  $(\mathbb{X}, d)$  where  $\mathbb{X}$  is a set and

$$d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$$

is a metric defined over  $\mathbb{X}$ .

A metric is a measure of distance between elements in a set.

# Metric Spaces

## Definition (Metric)

Let  $\mathbb{X}$  be a set. Then  $d : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$  is a metric if:

$$(M1) \quad d(x, y) = d(y, x), \quad \forall x, y \in \mathbb{X}$$

$$(M2) \quad d(x, y) \geq 0, \quad \forall x, y \in \mathbb{X}$$

$$(M3) \quad d(x, y) = 0, \quad \iff x = y$$

$$(M4) \quad d(x, z) \leq d(x, y) + d(y, z), \quad \forall x, y, z \in \mathbb{X}$$

(M4) is called the Triangle inequality.

# Examples of Metric Spaces

## Example (E1)

$(\mathbb{R}, d)$  where  $d(x, y) \stackrel{\Delta}{=} |x - y|$  is a metric space.

Note that

- ▶ (M1)  $|x - y| = |y - x|, \forall x, y \in \mathbb{R}$ .
- ▶ (M2)  $|x - y| \geq 0, \forall x, y \in \mathbb{R}$ .
- ▶ (M3)  $|x - y| = 0, \text{ if } x = y$ .
- ▶ (M4)  $|x - z| \leq |x - y| + |y - z| \forall x, y, z \in \mathbb{R}$ .

To convince yourself (M4), draw a picture. Note, a picture is not a proof.

# Examples of Metric Spaces

## Example (E2)

$(\mathbb{R}^n, d)$  where

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

where  $x = (x_1, \dots, x_n)^\top$  and  $y = (y_1, \dots, y_n)^\top$ .

Verify that  $d(\cdot, \cdot)$  satisfies (M1)-(M4).

# Examples of Metric Spaces

## Example (E3)

$(\mathbb{R}^n, d)$  where

$$d(x, y) \triangleq \left( \sum_{i=1}^n |x_i - y_i|^p \right)^{\frac{1}{p}}$$

where  $p \geq 1$ .

For general  $p \geq 1$ , the triangle inequality is a nontrivial and famous result.

# Examples of Metric Spaces

## Example (E4 bounded sequence space)

Let  $\ell^\infty$  be the set of all sequences of complex numbers where each number is bounded, i.e.,

$$x = (x_1, x_2, x_3, \dots) \in \ell$$

if  $x_i \in \mathbb{C}$  and  $|x_i| < \infty$ .

$(\ell, d)$  is a metric space where

$$d(x, y) = \sup_{j \in \mathbb{N}} |x_j - y_j|.$$

Verify (M1)-(M4).

# Examples of Metric Spaces

## Example (E5 continuous function space)

- ▶ Let  $C[a, b]$  be the set of all continuous functions on  $[a, b]$ , i.e., i.e.  $x \in C[a, b] \Rightarrow x(t)$  is continuous on  $[a, b]$ .

Let

$$d(x, y) = \max_{t \in [a, b]} |x(t) - y(t)|$$

then  $(C[a, b], d)$  is a metric space.

- ▶ This is a different perspective than calculus. In calculus you consider one function at a time. In this class, a function is one point in a larger metric space.

## Examples of Metric Spaces

### Example (E6 discrete metric space)

Let  $\mathbb{X}$  be any set, e.g., the set of three legged dogs, and let

$$d(x, y) = \begin{cases} 1 & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases}.$$

Then  $(\mathbb{X}, d)$  is a metric space since

- ▶ (M1)  $d(x, y) = d(y, x), \forall x, y \in \mathbb{X}.$
- ▶ (M2)  $d(x, y) \geq 0, \forall x, y \in \mathbb{X}.$
- ▶ (M3)  $d(x, y) = 0, \text{ if } x = y.$
- ▶ (M4)  $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in \mathbb{X}.$

## Examples of Metric Spaces

### Example (E7 binary vector space)

Let  $\mathbb{X} = \{0, 1\}^n$  be the set of binary vectors, i.e  
 $x \in \mathbb{X} \Rightarrow x = (x_1, x_2, \dots, x_n)$  where  $x_i \in \{0, 1\}$ . Let

$$d(x, y) = \sum_{i=1}^n h(x_i - y_i)$$

where

$$h(w) = \begin{cases} 1 & \text{if } w \neq 0 \\ 0 & \text{otherwise} \end{cases}.$$

$h$  is called the hamming distance, and simply counts the number of elements in  $x$  and  $y$  that are different.

## Metric Spaces / Norm Spaces / Inner Product Spaces

- ▶ Later in the chapter, we will later introduce the concepts of a norm and a norm space, and an inner product and inner product spaces.
- ▶ Many of the metric spaces introduced above are also norm spaces and inner product spaces, but not all.
- ▶ Metric spaces are the most general of the three.
- ▶ Before introducing the concept of a norm and a normed space, we develop general tools that also work for metric spaces.

## Section 2

### Topology

# Topology

- ▶ In this next section, we develop a set of tools that fall under that category of topology.
- ▶ These tools hold for metric spaces (including norm and inner product spaces).
- ▶ WARNING: There are a lot of definitions. These definitions will help talk formally about things in the future.

# Topology: Open and Closed Sets

## Definition (Ball)

Given a metric space  $(\mathbb{X}, d)$  a  $\delta$ -ball around  $x_0$  is defined to be  
 $B(x_0, \delta) = \{x \in \mathbb{X} : d(x, x_0) < \delta\}$

## Definition (Interior Point)

A point  $x_o \in \mathbb{X}$  is interior to  $S \subset \mathbb{X}$  if  
 $\exists \delta > 0$  such that  $B(x_o, \delta) \subset S$ .

## Definition (Open Set)

A set  $\mathbb{X}$  is open if all points in  $\mathbb{X}$  are interior.

## Definition (Closed Set)

A set  $S$  is closed in  $\mathbb{X}$  if  $\mathbb{X} \setminus S$  is open.

# Topology: Convergence

Let  $(\mathbb{X}, d)$  be a metric space.

## Definition (Convergence)

Given a sequence  $\{x_n\}_{n=1}^{\infty}$ , where  $x_n \in \mathbb{X}$ , the following are equivalent

- ▶  $\lim_{n \rightarrow \infty} x_n = x^*$
- ▶  $x_n \rightarrow x^*$
- ▶  $\forall \epsilon > 0, \exists N(\epsilon) \text{ such that } n \geq N \Rightarrow d(x_n, x^*) < \epsilon$

A sequence  $\{x_n\}_{n=1}^{\infty}$  in  $\mathbb{X}$  with a limit  $x^* \in \mathbb{X}$  is said to converge.

# Topology: Convergence

Note that a limit may not always exist (similar to min, max)

For example,  $\lim_{t \rightarrow \infty} \sin(t)$  does not exist.

## Definition ( $\limsup$ )

Define  $\limsup$  as the largest limit (possibly infinity) of any subsequence.

## Definition ( $\liminf$ )

Define  $\liminf$  is the smallest limit of all possible subsequences.

## Example

- ▶  $\limsup_{t \rightarrow \infty} \sin(t) = 1$  since the subsequence  
 $t_n = \frac{k\pi}{2}, k = 1, 5, 9, \dots$  converges to 1
- ▶  $\liminf_{t \rightarrow \infty} \sin(t) = -1$  since the subsequence  
 $t_n = \frac{k\pi}{2}, k = 3, 7, 11, \dots$  converges to -1

# Topology: Cauchy Sequence

## Definition (Cauchy Sequence)

A sequence  $\{x_n\}_{n=1}^{\infty}$  in a metric space  $(\mathbb{X}, d)$  is said to be a Cauchy sequence if  $\forall \epsilon > 0, \exists N(\epsilon) > 0$  such that  $n, m > N \Rightarrow d(x_n, x_m) < \epsilon$

A sequence is Cauchy if elements in its tail get increasingly closer together. Note that we have not said anything about an element of convergence.

## Theorem

*If a sequence  $\{x_n\}_{n=1}^{\infty}$  in  $\mathbb{X}$  converges to an element  $x^* \in \mathbb{X}$  then it is a Cauchy sequence.*

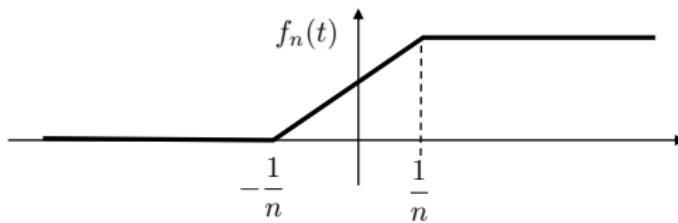
The converse is not true!! I.e., not all Cauchy sequences converge.

# Topology: Cauchy Sequence

Example (from book)

Let  $\mathbb{X} = C[-1, 1]$  and  $d(f, g) = \left( \int_{-1}^1 (f(t) - g(t))^2 dt \right)^{\frac{1}{2}}$

let  $f_n :$



By integration we get:

$$d(f_n, f_n) = \frac{1}{6m^3n}(m^3 + 4m^2n + mn^2 + 2n^3)$$

$\rightarrow 0$  for  $n, m$  large ( $m > n$ )

but  $f_n$  converges to a discontinuous function which is not in  $\mathbb{X}$ .

This is undesirable

# Topology: Complete Metric Space

## Definition (Complete metric space)

A metric space  $(\mathbb{X}, d)$  is complete if every Cauchy sequence in  $\mathbb{X}$  converges to a value in  $\mathbb{X}$ .

## Implication

$C[a, b]$  with metric  $(\int_a^b |f - g|^2 dt)^{1/2}$  is not complete.

- ▶ Banach spaces are complete normed spaces (discussed later).
- ▶ Hilbert spaces (extremely important in signal processing and control) are complete inner product spaces (discussed later).
- ▶ The importance of  $L_p$  and  $\ell_p$  are that they are complete spaces.

## Section 3

### Vector Spaces

# Vector Spaces

A field is a set of scalars with well defined addition and multiplication operations.

## Example of fields:

- ▶  $\mathbb{R}$  with normal addition and multiplication operations
- ▶  $\mathbb{C}$  with complex addition and complex multiplication
- ▶ The set of quaternions, with addition and quaternion multiplication
- ▶ Binary numbers  $\{0, 1\}$  where addition is the “or” operator and multiplication is the “and” operator.

# Vector Spaces

## Definition (Linear Vector Space)

A linear vector space is a pair  $(\mathbb{X}, \mathbb{F})$ , where  $\mathbb{X}$  is a set of objects, and  $\mathbb{F}$  is a field, this is closed under addition and scalar multiplication. i.e.,

- ▶  $x \in \mathbb{X}, \alpha \in \mathbb{F} \Rightarrow \alpha x \in \mathbb{X}$
- ▶  $x, y \in \mathbb{X} \Rightarrow x + y \in \mathbb{X}$ .

By implication

- ▶  $x \in \mathbb{X}, \alpha, \beta \in \mathbb{F} \Rightarrow (\alpha + \beta)x = \alpha x + \beta x \in \mathbb{X}$
- ▶  $x, y \in \mathbb{X}, \alpha \in \mathbb{F} \Rightarrow \alpha(x + y) = \alpha x + \alpha y \in \mathbb{X}$
- ▶  $x, y \in \mathbb{X}, \alpha, \beta \in \mathbb{F} \Rightarrow \alpha x + \beta y \in \mathbb{X}$ .

# Vector Spaces: Subspace

## Definition (Subspace)

A subspace  $V \subset \mathbb{X}$  is a subset of  $\mathbb{X}$  that is also a linear vector space, in particular it contains zero.

**Important property:** A vector space contains a zero element.

## Vector Spaces: Examples

The following are vector spaces:

- ▶  $(\mathbb{R}^n, \mathbb{R})$ ,  $(\mathbb{C}^n, \mathbb{C})$ ,  $(\mathbb{R}^{m \times n}, \mathbb{R})$ ,  $(C[a, b], \mathbb{R})$ ,  $(\ell^\infty, \mathbb{R})$ ,  $(L^\infty, \mathbb{R})$ .

The following are NOT vector spaces:

- ▶ The set  $\mathbb{X} = \mathbb{R} \times [0, 2\pi]$ , (a cylinder) is not a vector space for any field  $\mathbb{F}$ . This is the state space for an inverted pendulum.
- ▶ The set of rotation matrices is not a vector space for any field  $\mathbb{F}$ . This is in the configuration space for robots and satellites.
- ▶ The set of unit quaternions is not a vector space for any field. Quaternions are used extensively in robotics, quantum mechanics, and computer graphics.
- ▶ There are many useful spaces that are NOT linear vector spaces.

## Vector Spaces: Linear Independence

Let  $S$  be a vector space and let  $T \subset S$ . ( $T$  may have uncountable infinite members).  $T$  is linearly independent if for each finite nonempty subset of  $T$ . i.e.,  $\{p_1, \dots, p_n\}$  where  $p_i \in T$ , we have that

$$c_1p_1 + \cdots + c_np_n = 0 \iff c_1 = c_2 = \cdots = c_n = 0.$$

Otherwise  $T$  is linearly dependent.

# Vector Spaces: Linear Independence

## Example

Let  $S = \mathbb{R}^3$  then the set  $T = \{(1, 0, 0)^\top, (0, 1, 0)^\top\} \subset \mathbb{R}^3$  is linearly independent since

$$c_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

if and only if  $c_1 = c_2 = 0$ .

However, the set  $T = \{(1, 1, 0)^\top, (2, 2, 0)^\top\} \subset \mathbb{R}^3$  is linearly dependent since

$$c_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + c_2 \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix} = \begin{pmatrix} c_1 + 2c_2 \\ c_1 + 2c_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

when  $c_1 = -2$  and  $c_2 = 1$  (as only on example).

# Vector Spaces: Span

## Definition (Span)

Let  $S$  be a vector space, then  $\text{span}(T)$  is the set of all linear combinations of  $T \subseteq S$ .

## Example

$$\text{span} \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} \alpha \\ \alpha \end{pmatrix} \mid \alpha \in \mathbb{R} \right\}$$

## Example

$$\text{span} \left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\} = \left\{ \begin{pmatrix} \alpha \\ \beta \end{pmatrix} : \alpha, \beta \in \mathbb{R} \right\} = \mathbb{R}^2.$$

# Vector Spaces: Basis

## Definition (Basis)

$T$  is a basis for the vector space  $S$  if  $T$  is linearly independent and  $\text{span}(T) = S$ .

## Definition (Dimension)

The dimension of the vector space  $S$  is the smallest number of linearly independent vectors needed to span  $S$ .

## Example

One possible basis for  $\mathbb{R}^n$  is given by

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \right\}.$$

Therefore  $\dim(\mathbb{R}^n) = n$ .

## Vector Spaces: Basis

### Example

One possible basis for  $\ell^\infty$  is given by

$$\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ \vdots \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ \vdots \end{pmatrix}, \dots, \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \end{pmatrix}, \dots \right\}$$

Therefore  $\dim(\ell^\infty) = \infty$ .

# Vector Spaces: Basis

## Example

The set of all polynomials  $P$  is a vector space with basis

$$\{1, t, t^2, \dots\}$$

Therefore  $\dim(P) = \infty$ .

## Example

The set of all polynomials of degree  $\leq q$   $P^q$  is a vector space with basis

$$\{1, t, t^2, \dots, t^q\}$$

Therefore  $\dim(P^q) = q$ .

## Section 4

### Normed Spaces

# Norms and Normed Spaces

## Definition (Norm)

Let  $S$  be a vector space,  $\|x\|$  is a norm if:

$$(N1) \quad \|x\| \geq 0 \quad \forall x \in S$$

$$(N2) \quad \|x\| = 0 \quad \Leftrightarrow x = 0$$

$$(N3) \quad \|\alpha x\| = |\alpha| \|x\|$$

$$(N4) \quad \|x + y\| \leq \|x\| + \|y\| \quad (\text{triangle inequality})$$

## Differences between norms and metrics:

- ▶ Norms only have one argument (the length of a vector), where metrics are distances between elements of a set.
- ▶ Norms are only defined for vector spaces!  
(i.e. there is no norm for rotation matrices, but there are metrics!)
- ▶ Norms scale with the vector (N3)  
(there are metrics that don't scale), e.g.  
$$d(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$
- ▶ Every norm is also a metric

$$\|x - y\| = d(x, y)$$

$$\|x\| = d(x, 0)$$

# Definition: Normed Space

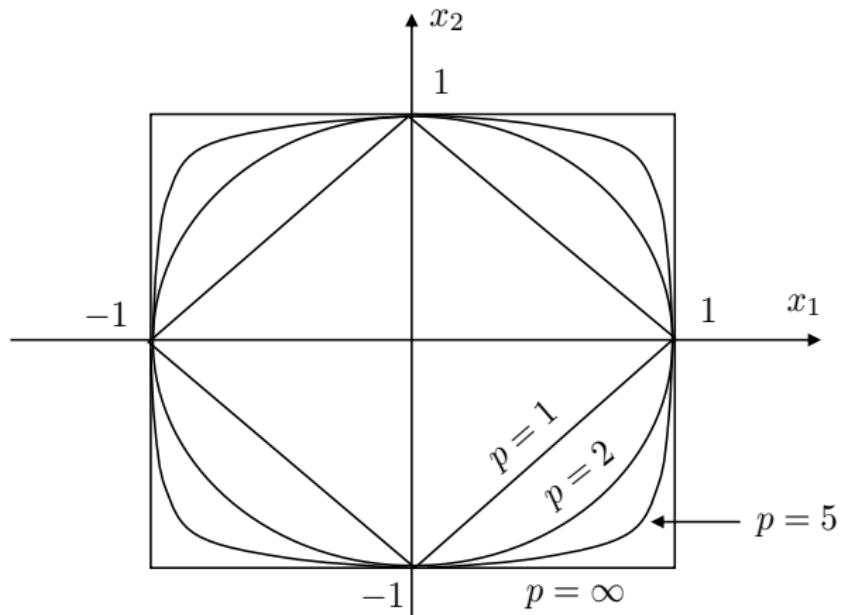
A normed space is a pair  $(\mathbb{X}, \|\cdot\|)$  where  $\mathbb{X}$  is a vector space and  $\|\cdot\|$  is a norm.

## Example (Normed Spaces)

$\mathbb{R}^n$  is a vector space. All of the following norms are valid:

- ▶ one-norm  $\|x\|_1 = \sum_{i=1}^n |x_i|$  (power vectors)
- ▶ two-norm  $\|x\|_2 = \left( \sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$  (energy vectors)
- ▶ infinity-norm  $\|x\|_\infty = \max_{i=1,\dots,n} |x_i|$  (bounded vectors)
- ▶ p-norm  $\|x\|_p = \left( \sum_{i=1}^n |x_i|^p \right)^{\frac{1}{p}}$

# Unit Circle in $\mathbb{R}^2$



## Normed Space Example: Sequence Spaces

Let  $\ell$  be the set of sequences:  $x = (x_1, x_2, x_3, \dots)$ . The following normed vector spaces can be defined:

- ▶  $\ell_1$ : (power sequences) If  $\|x\|_{\ell_1} = \sum_{i=1}^{\infty} |x_i|$  then  
 $\ell_1 \stackrel{\triangle}{=} \{x \in \ell : \|x\|_{\ell_1} < \infty\}$
- ▶  $\ell_2$ : (energy sequences) If  $\|x\|_{\ell_2} = (\sum_{i=1}^{\infty} |x_i|^2)^{1/2}$  then  
 $\ell_2 \stackrel{\triangle}{=} \{x \in \ell : \|x\|_{\ell_2} < \infty\}$
- ▶  $\ell_{\infty}$ : (bounded sequences) If  $\|x\|_{\ell_{\infty}} = \sup_{j \in \mathbb{N}} |x_j|$  then  
 $\ell_{\infty} \stackrel{\triangle}{=} \{x \in \ell : \|x\|_{\ell_{\infty}} < \infty\}$
- ▶  $\ell_p$ : If  $\|x\|_{\ell_p} = (\sum_{i=1}^{\infty} |x_i|^p)^{1/p}$  then  $\ell_p \stackrel{\triangle}{=} \{x \in \ell : \|x\|_{\ell_p} < \infty\}$   
for  $1 \leq p \leq \infty$

# Normed Space Examples

## Example

Consider the sequence  $x = (1, 1, 1, \dots)$ :

- ▶  $x \in \ell_\infty$ , but
- ▶  $x \notin \ell_p$  for  $1 \leq p < \infty$ .

## Example

Consider the sequence  $x = (1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$

- ▶  $x \notin \ell_1$  (prove this), but
- ▶  $x \in \ell_p \quad p > 1$  (prove this)

## Normed Space Example: Function Spaces

Let  $L^n(\Omega)$  be the set of functions on  $\Omega$ .  $x \in L^n(\Omega)$  is an equivalent classes of functions, i.e. equal except on a set of measure zero. (picture) The following norms are valid:

- ▶  $L_1^n(\Omega)$  (power signals). If  $\|x\|_{L_1^n(\Omega)} = \int_{\Omega} \|x(t)\| dt$ , then  $L_1^n(\Omega) = \{x \in L^n(\Omega) | \|x\|_{L_1^n(\Omega)} < \infty\}$ .
- ▶  $L_2^n(\Omega)$  (energy signals). If  $\|x\|_{L_2^n(\Omega)} = \left( \int_{\Omega} \|x(t)\|^2 dt \right)^{1/2}$ , then  $L_2^n(\Omega) = \{x \in L^n(\Omega) | \|x\|_{L_2^n(\Omega)} < \infty\}$ .
- ▶  $L_p^n(\Omega)$ . If  $\|x\|_{L_p^n(\Omega)} = \left( \int_{\Omega} \|x(t)\|^p dt \right)^{1/p}$ , then  $L_p^n(\Omega) = \{x \in L^n(\Omega) | \|x\|_{L_p^n(\Omega)} < \infty\}$ ,  $1 \leq p \leq \infty$ .
- ▶  $L_{\infty}^n(\Omega)$  (bounded signals). If  $\|x\|_{L_{\infty}^n(\Omega)} = \sup_{t \in \Omega} \|x(t)\|$ , then  $L_{\infty}^n(\Omega) = \{x \in L^n(\Omega) | \|x\|_{L_{\infty}^n(\Omega)} < \infty\}$ .

## Section 5

### Inner Product Spaces

# Inner Product Spaces

## Definition (Inner Product)

Let  $S$  be a vector space over  $\mathbb{R}$ . An inner product  $\langle \cdot, \cdot \rangle: S \times S \rightarrow \mathbb{R}$  has the following properties:

$$(IP1) \quad \langle x, y \rangle = \overline{\langle y, x \rangle}$$

$$(IP2) \quad \langle \alpha x, y \rangle = \alpha \langle x, y \rangle$$

$$(IP3) \quad \langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$$

$$(IP4) \quad \langle x, x \rangle > 0 \quad \text{if } x \neq 0, \langle x, x \rangle = 0 \Leftrightarrow x = 0$$

## Definition (Inner Product Space)

A vector space with an inner product defined is called an inner-product space.

## Definition (Hilbert Space)

A complete inner-product space is called a Hilbert space.

## Inner Product Spaces: Examples

- ▶  $\mathbb{R}^n$ :  $\langle x, y \rangle = \sum_{i=1}^n x_i y_i = x^T y$  is called the Euclidean inner product.
- ▶  $\mathbb{C}^n$ :  $\langle x, y \rangle = \sum_{i=1}^n x_i \overline{y_i} = y^H x$
- ▶  $\mathbb{R}^n$  with the Euclidean inner product is a Hilbert space .
- ▶  $\mathbb{C}^n$  with the Euclidean inner product is a Hilbert space.
- ▶ All finite-dimensional inner-product spaces are Hilbert spaces.

## Inner Product Spaces: Examples

- ▶ Real sequences  $\ell_2$ :  $\langle x, y \rangle_{\ell_2} = \sum_{i=1}^{\infty} x_i y_i$
- ▶ Complex sequences  $\ell_2$ :  $\langle x, y \rangle_{\ell_2} = \sum_{i=1}^{\infty} x_i \overline{y_i}$
- ▶ Both of these examples are Hilbert spaces.

## Inner Product Spaces: Examples

- ▶ Complex function space  $L_2^n(\Omega)$  with inner product:

$$\langle x, y \rangle = \int_{-\infty}^{\infty} y^H(t)x(t) dt$$

is a Hilbert space, but

- ▶ Continuous function  $C[a, b]$  with the same inner product is NOT a Hilbert space.

## Norms vs Inner Products

Every inner product defines a norm (but not vice-versa)

$$\|x\| = \langle x, x \rangle^{\frac{1}{2}}$$

where  $\|\cdot\|$  is called the norm induced by the inner product  $\langle \cdot, \cdot \rangle$ .

## Examples of induced norms

$$\|\cdot\|_2 : \langle x, x \rangle^{1/2} = \left( \sum_{i=1}^n x_i^2 \right)^{1/2} = \|x\|_2$$

$$\|\cdot\|_{\ell_2} : \langle x, x \rangle^{1/2} = \left( \sum_{i=1}^{\infty} x_i^2 \right)^{1/2} = \|x\|_{\ell_2}$$

$$\|\cdot\|_{L_2} : \langle x, x \rangle^{1/2} = \left( \int_{\Omega} x^T(t) x(t) dt \right)^{1/2} = \left( \int_{\Omega} \|x(t)\|_2^2 dt \right)^{1/2} = \|x\|_{L_2}$$

Note that induced norms are all 2-norms.

# Cauchy-Schwartz Inequality

## Theorem (Cauchy-Schwartz)

Let  $S$  be any inner product space (doesn't need to be Hilbert) and let  $\|\cdot\| = \langle \cdot, \cdot \rangle^{1/2}$  then  $\forall x, y \in S$

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

with equality iff  $y = \alpha x$  where  $\alpha \in \mathbb{F}$  is any scalar in the field  $\mathbb{F}$ .

## Cauchy-Schwartz Inequality: Proof

The inequality clearly holds if either  $x = 0$  or  $y = 0$ . Therefore assume that  $x \neq 0$  and  $y \neq 0$ . Then

$$\begin{aligned}\|x - \alpha y\|^2 &= \langle x - \alpha y, x - \alpha y \rangle \\&= \langle x, x \rangle - \alpha \langle y, x \rangle - \langle x, \alpha y \rangle + \langle \alpha y, \alpha y \rangle \\&= \langle x, x \rangle - \alpha \langle y, x \rangle - \overline{\langle x, \alpha y \rangle} + \alpha \overline{\langle y, \alpha y \rangle} \\&= \langle x, x \rangle - \alpha \langle y, x \rangle - \overline{\langle \alpha y, x \rangle} + \alpha \overline{\langle \alpha y, y \rangle} \\&= \langle x, x \rangle - \alpha \langle y, x \rangle - \overline{\alpha} \overline{\langle y, x \rangle} + \alpha \overline{\alpha} \overline{\langle y, \alpha y \rangle} \\&= \langle x, x \rangle - \alpha \langle y, x \rangle - \overline{\alpha} \langle x, y \rangle + |\alpha|^2 \langle y, y \rangle \\&= \|x\|^2 - \alpha \langle y, x \rangle - \overline{\alpha} \langle x, y \rangle + |\alpha|^2 \|y\|^2\end{aligned}$$

## Cauchy-Schwartz Inequality: Proof

Recall the technique of completing the square:

$$\begin{aligned} ax^2 + bx + c &= a\left(x^2 + \frac{b}{a}x\right) + c \\ &= a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2}{4a} + c. \end{aligned}$$

Complete the square in  $\alpha$ :

$$\begin{aligned} \|x - \alpha y\|^2 &= \|y\|^2 \left( \alpha \bar{\alpha} - \alpha \frac{\langle x, y \rangle}{\|y\|^2} - \bar{\alpha} \frac{\langle x, y \rangle}{\|y\|^2} \right) + \|x\|^2 \\ &= \|y\|^2 \left( \alpha - \frac{\langle x, y \rangle}{\|y\|^2} \right) \left( \bar{\alpha} - \frac{\langle x, y \rangle}{\|y\|^2} \right) - \frac{|\langle x, y \rangle|^2}{\|y\|^2} + \|x\|^2 \end{aligned}$$

## Cauchy-Schwartz Inequality: Proof

Let  $\alpha^* = \frac{\langle x, y \rangle}{\|y\|^2}$  to get

$$\begin{aligned} 0 &\leq \|x - \alpha^* y\|^2 = \|x\|^2 - \frac{|\langle x, y \rangle|^2}{\|y\|^2} \\ \Rightarrow |\langle x, y \rangle|^2 &\leq \|x\|^2 \|y\|^2 \\ \Rightarrow |\langle x, y \rangle| &\leq \|x\| \|y\| \end{aligned}$$

## Section 6

### Notions of Convergence

# Notions of Convergence

Definition (Strong Convergence/ Convergence in norm)

$x_n$  converges strongly to  $x$ , i.e.  $x_n \xrightarrow{s} x$  iff

$$\|x_n - x\| \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty$$

Definition (Weak Convergence / Convergence in inner product)

$x_n$  converges weakly to  $x$ , i.e.  $x_n \xrightarrow{w} x$  iff

$$\langle x_n, y \rangle \rightarrow \langle x, y \rangle, \forall y \in S,$$

Note that this must hold for all  $y \in S$ , therefore Example 2.4.4 in the book is bogus!

## Notions of Convergence (cont.)

### Theorem (Strong vs. Weak Convergence)

Let  $(x_n)$  be a sequence in a normed space  $\mathbb{X}$ . Then

- A. Strong convergence  $\Rightarrow$  weak convergence with the same limit
- B. The converse of (A.) is not generally true
- C. If  $\dim \mathbb{X} < \infty$ , then weak convergence  $\Rightarrow$  strong convergence.

## Proof:

(A) By definition of strong convergence,

$$x_n \xrightarrow{s} x^* \Rightarrow \|x_n - x^*\| \rightarrow 0$$

so let  $y$  be any element in  $\mathbb{X}$  then

$$|\langle x_n, y \rangle - \langle x^*, y \rangle| = |\langle x_n - x^*, y \rangle| \leq \|x_n - x^*\| \|y\|$$

but the RHS  $\rightarrow 0$  which implies that the LHS  $\rightarrow 0$  which implies weak convergence.

## Proof:

(B) Before proving part (B) lets first understand what is wrong with Example 2.4.4 in the book.

$$x_n = (0, 0, 0, \dots, 1, 0, \dots)$$

$$y = (1, 1/2, 1/4, 1/8, \dots)$$

Then  $\langle x_n, y \rangle \rightarrow 0$  but this does not imply weak convergence since it must hold for all  $y \in \mathbb{X}$ .

## Proof:

To prove part (B) we need a counter example. Again let  $x_n = (0, 0, \dots, 0, 1, 0, \dots)$  and let  $\mathbb{X} = \ell_2$  i.e.

$$y \in \mathbb{X} \Rightarrow \left( \sum_{i=1}^{\infty} |y_i|^2 \right)^{\frac{1}{2}} < \infty$$
$$\Rightarrow y_i \rightarrow 0 \quad \text{as} \quad i \rightarrow \infty$$

so

$$\langle x_n, y \rangle = y_n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty \quad \forall y \in \mathbb{X}$$
$$\Rightarrow \{x_n\} \xrightarrow{w} 0$$

but there is no  $x^*$  such that  $\|x_n - x^*\| \rightarrow 0$ .

## Proof:

(C) Suppose that  $x_n \xrightarrow{w} x$  and  $\dim(\mathbb{X}) = k$  then

$$\forall y \in \mathbb{X} \quad \langle x_n, y \rangle \rightarrow \langle x, y \rangle.$$

Let  $\{e_1, \dots, e_k\}$  be an orthonormal basis for  $\mathbb{X}$ , i.e.  $\langle e_i, e_j \rangle = \delta_{ij}$ , then

$$x_n = a_1^{(n)} e_1 + \cdots + a_k^{(n)} e_k$$

$$x = a_1 e_1 + \cdots + a_k e_k.$$

## Proof:

Then since  $\langle x_n, y \rangle \rightarrow \langle x, y \rangle \forall y$ , let  $y = e_j$

$$\Rightarrow \left\langle a_1^{(n)}e_1 + \cdots + a_k^{(n)}e_k, e_j \right\rangle = a_j^{(n)}$$

and

$$\langle a_1e_1 + \cdots + a_ke_k, e_j \rangle = a_j$$

so

$$\langle x_n, e_j \rangle \rightarrow \langle x, e_j \rangle \Rightarrow a_j^{(n)} \rightarrow a_j \quad \forall j = 1, \dots, k$$

Also,

$$\begin{aligned} \|x_n - x\| &= \left\| \sum_{j=1}^k a_j^{(n)}e_j - \sum_{j=1}^k a_j e_j \right\| = \left\| \sum_{j=1}^k (a_j^{(n)} - a_j)e_j \right\| \\ &\leq \sum_{j=1}^k |a_j^{(n)} - a_j| \|e_j\| \rightarrow 0 \end{aligned}$$

$\Rightarrow$  strong convergence

# Equivalence of Norms

## Theorem

Let  $\dim(\mathbb{X}) = k$  and let  $\|\cdot\|$  and  $\|\cdot\|_0$  be two different norms on  $\mathbb{X}$  then  $\exists a, b$  such that

$$a \|\mathbf{x}\|_0 \leq \|\mathbf{x}\| \leq b \|\mathbf{x}\|_0$$

## Proof.

(in book page 96)



**Implication:** For convergence proofs, it doesn't matter which norm you use, therefore, use the one that simplifies the proof.

## Section 7

### Orthogonality

# Orthogonality

Let  $x, y \in \mathbb{X}$  where  $\mathbb{X}$  is an inner product space. Then the angle between  $x$  and  $y$  is

$$\theta = \cos^{-1} \left( \frac{\langle x, y \rangle}{\|x\| \|y\|} \right).$$

i.e.

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

## Orthogonality, cont.

### Definition (Colinear)

Two vectors  $x, y \in \mathbb{X}$  are said to be colinear if

$$\theta = 180 * n \quad n = 0, \pm 1, \pm 2, \dots$$

### Definition (Orthogonal)

Two vectors  $x, y \in \mathbb{X}$  are said to be orthogonal if

$$\theta = 90 * n \quad n = \pm 1, \pm 3, \pm 5, \dots$$

i.e.,  $\langle x, y \rangle = 0$ .

If  $\langle x, y \rangle = 0$  we write  $x \perp y$ .

## Orthogonality, cont.

### Example (Vectors in $L_2[0, 2\pi]$ )

The functions  $x = \sin(t)$  and  $y = \cos(t)$  are orthogonal since

$$\langle x, y \rangle = \int_0^{2\pi} \sin(t) \cos(t) dt = 0.$$

### Example (Vectors in $\ell$ )

The sequences

$$x = (1, 1, 1, 1, 0, 0, \dots)$$

$$y = (1, -1, 1, -1, 1, \dots)$$

are orthogonal since

$$\langle x, y \rangle = \sum_{i=1}^{\infty} x_i y_i = 0.$$

## Other useful inner products and norms: Weighting

### Definition (Positive Definite Matrix)

A matrix  $W : \mathbb{R}^k \rightarrow \mathbb{R}^k$  is positive definite (PD) if

$$\forall x \in \mathbb{R}^k \quad x^T W x > 0$$

- ▶  $W$  is positive semi-definite (PSD) if  $x^T W x \geq 0$
- ▶  $W$  is negative definite (ND) if  $x^T W x < 0 \quad \forall x \in \mathbb{R}^k$
- ▶  $W$  is negative semi-definite (NSD) if  $x^T W x \leq 0 \quad \forall x \in \mathbb{R}^k$
- ▶ Otherwise it is indefinite

## Examples of positive definiteness

- ▶  $W = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  is PD since

$$x^T W x = x_1^2 + x_2^2 > 0 \quad \forall x \in \mathbb{R}^2$$

- ▶  $W = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$  is PSD since

$$x^T W x = x_1^2 = 0 \quad \forall x = \begin{pmatrix} 0 \\ \alpha \end{pmatrix} \neq 0$$

- ▶  $W = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$  is indefinite since

$$x^T W x = -x_1^2 + x_2^2$$

which can be positive or negative depending on  $x$ .

## Examples of Inner Products

### Weighted Inner Products and Norms

If  $W > 0$  then  $\langle x, y \rangle_W = x^H W y$  is a valid inner product which induces the weighted norm  $\|x\|_W = (x^H W x)^{\frac{1}{2}}$

We can define weighted inner products for functions:

$$\langle f, g \rangle_W = \int f(t)g(t)w(t)dt$$

where  $w(t) > 0$  except on a set of measure zero.

# Examples of Inner Products

## Definition (Expectation)

Expectation is a weighted inner product with weight  $f_{\mathbb{X}\mathbb{Y}}(x, y)$

$$\langle \mathbb{X}, \mathbb{Y} \rangle = \int xy f_{\mathbb{X}\mathbb{Y}}(x, y) dx dy = E[\mathbb{X}\mathbb{Y}]$$

if  $\mathbb{X}$  is a zero mean then

$$\langle x, x \rangle = \text{var}(x)$$

is the norm induced by  $E[\cdot]$

## Examples of Inner Products

- ▶ Let  $\mathbb{I}(m, n)$  be the set of grayscale images with  $m \times n$  pixels, each valued between  $[0, 255]$ .
- ▶ A valid inner on  $\mathbb{I}(m, n)$  is given by

$$\langle I, J \rangle = \sum_{u=1}^m \sum_{v=1}^n I(u, v)J(u, v), \quad \forall I, J \in \mathbb{I}(m, n).$$

# Orthogonal Subspaces

## Definition (Orthogonal Subspaces)

Let  $V, W$  be subspaces of  $S$ .  $V \perp W$  if

$$\forall v \in V \text{ and } \forall w \in W, \quad \langle v, w \rangle = 0$$

## Definition (Orthogonal Complement)

$V^\perp$ , called the orthogonal complement of  $V$ , is the set

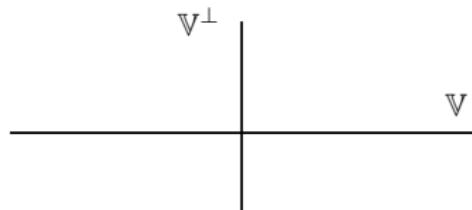
$$V^\perp = \{x \in S : \forall v \in V, \langle x, v \rangle = 0\}$$

# Orthogonal Subspaces, cont.

## Example

Let  $S = \mathbb{R}^2$  and  $V = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \alpha \in \mathbb{R} \right\}$  then

$$V^\perp = \left\{ \begin{pmatrix} 0 \\ \alpha \end{pmatrix}, \alpha \in \mathbb{R} \right\}$$



## Orthogonal Subspaces, cont.

### Theorem

Let  $V$  and  $W$  be subspaces of an inner product space  $S$  (not necessarily Hilbert). Then

1.  $V^\perp$  is a closed subspace of  $S$
2.  $V \subset V^{\perp\perp}$  ( $V = V^{\perp\perp}$  if  $S$  is complete)
3. If  $V \subset W$  then  $W^\perp \subset V^\perp$
4.  $V^{\perp\perp\perp} = V^\perp$
5. If  $x \in V \cap V^\perp$  then  $x = 0$
6.  $\{0\}^\perp = S$  and  $S^\perp = \{0\}$

Prove one in class.

# Inner Sum and Direct Sum

## Definition (Inner Sum)

If  $V$  and  $W$  are linear subspaces then

$$V + W = \{x : x = v + w, v \in V \text{ and } w \in W\}$$

is the inner sum.

## Definition (Orthogonal Sum)

If  $V$  and  $W$  are orthogonal subspaces then the sum

$$V \oplus W = \{x : x = v + w, v \in V \text{ and } w \in W\}$$

is called the orthogonal sum.

## Definition (Disjoint Subspaces)

Two subspaces are said to be disjoint if

$$V \cap W = \{0\}$$

## Inner Sum and Direct Sum, cont.

### Lemma

Let  $V + W$  be subspaces of  $S$  and let  $x \in V + W$  then the representation  $x = v + w$  is unique iff  $V + W$  are disjoint.

### Proof.

( $\Leftarrow$ ) Assume  $V, W$  are disjoint but  $x = v + w$  is not unique i.e.  $x = v_1 + w_1 = v_2 + w_2$  where  $v_1 \neq v_2$  and  $w_1 \neq w_2$ . This implies that  $v_1 - v_2 = w_2 - w_1$  but  $v_1 - v_2 \in V$  and  $w_2 - w_1 \in W$  since  $V, W$  are subspaces. Since  $V \cap W = \{0\}$  we must have that  $v_1 - v_2 = w_2 - w_1 = 0$  or  $v_1 = v_2$  and  $w_1 = w_2$  which is a contradiction. □

# Inner Sum and Direct Sum, cont.

## Lemma

If  $V$  and  $W$  are orthogonal subspaces then the representation of  $x \in V \oplus W$  is unique (i.e.  $x = v + w$ , where  $v \in V$  and  $w \in W$ ).

## Example

Let  $S = \mathbb{R}^2$ , let  $V = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} : \alpha \in \mathbb{R} \right\}$ , let

$W = \left\{ \begin{pmatrix} 0 \\ \alpha \end{pmatrix} : \alpha \in \mathbb{R} \right\}$  Then

$$\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 5 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

is a unique decomposition.

## Difference between a Hamel basis and a Complete basis.

### Definition

An orthonormal set of basis vectors  $T = \{p_1, p_2, \dots\}$  is said to be a complete basis for a Hilbert space  $S$  if every  $x \in S$  can be represented as

$$x = \sum_{j=1}^{\infty} c_j p_j$$

Examples of complete bases: Fourier functions:  $e^{j\omega t}$   
Legendre & Chebyshev polynomials

Difference: A Hamel basis  $\Rightarrow$  every  $x$  can be represented by a finite representation. A complete basis allows functions through a limiting process.

# Section 8

## Linear Operators

# Operators and Transformations

## Definition (Linear Operator)

Let  $\mathcal{L} : \mathbb{X} \rightarrow \mathbb{Y}$  be an operator from  $\mathbb{X}$  to  $\mathbb{Y}$ .  $\mathcal{L}$  is a linear operator if

1.  $\mathcal{L}[\alpha x] = \alpha \mathcal{L}[x] \quad \forall x \in \mathbb{X} \quad \forall \alpha \in \mathbb{F}$
2.  $\mathcal{L}[x_1 + x_2] = \mathcal{L}[x_1] + \mathcal{L}[x_2], \quad \forall x_1, x_2 \in \mathbb{X}$

# Examples of Linear Operators

## Example (Matrices)

Operators from  $\mathbb{C}^n$  to  $\mathbb{C}^m$  are  $m \times n$  matrices.

$$A(\alpha x + \beta y) = \alpha Ax + \beta Ay$$

$A$  is a linear operator.

## Example (Differential Equations with no input)

The differential equation  $\dot{x} = Ax; \quad x(0) = x_0$  defines a linear operator from  $\mathbb{R}^n$  to  $L_2[0, T]$

$$y(t) = \mathcal{L}[x_0] \text{ where } \mathcal{L}[x_0] = e^{At}x_0$$

$\mathcal{L}$  is linear since

$$e^{At}(\alpha x_{01} + \beta x_{02}) = \alpha e^{At}x_{01} + \beta e^{At}x_{02}$$

# Examples of Linear Operators

## Example (Convolution)

Convolution is a linear operator from  $L_\infty$  to  $L_\infty$  if  $h(t) \in L_1[-\infty, \infty]$ , i.e.

$$y(t) = \mathcal{L}[x(t)] = \int_{-\infty}^{\infty} h(\tau)x(t - \tau)d\tau$$

(Recall: for a system to be BIBO stable required that  $\int_{-\infty}^{\infty} |h(\tau)|d\tau < \infty$  i.e.  $h(t) \in L_1[-\infty, \infty]$ )

# Examples of Linear Operators

## Example (Fourier Transform)

(E4) The Fourier transform defines a linear operator from  $L_2[-\infty, \infty]$  to  $L_2[-\infty, \infty]$ .

$$X(j\omega) = \mathcal{L}[x(t)] \triangleq \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$$

There are many examples of linear operators!

# Range and Null Space of an Operator

## Definition (Range Space)

Let  $\mathcal{L} : \mathbb{X} \rightarrow \mathbb{Y}$  be a linear operator. The range space (or image) of  $\mathcal{L}$  is

$$\mathcal{R}(\mathcal{L}) = \{y \in \mathbb{Y} : y = \mathcal{L}[x] \text{ and } x \in \mathbb{X}\} \subseteq \mathbb{Y}$$

## Definition (Null Space)

The Null space or kernel of  $\mathcal{L}$  is

$$\mathcal{N}(\mathcal{L}) = \{x \in \mathbb{X} : \mathcal{L}[x] = 0\} \subseteq \mathbb{X}$$

## Example of Range and Null Space

- ▶ Consider the matrix  $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$  which defines a linear operator from  $\mathbb{R}^3$  to  $\mathbb{R}^2$ .
- ▶ Note that  $y = Ax = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$ .
- ▶ Therefore, the range space is

$$\mathcal{R}(A) = \left\{ \begin{pmatrix} \alpha \\ 0 \end{pmatrix} : \alpha \in \mathbb{R} \right\} \subset \mathbb{R}^2.$$

- ▶ Similarly, the null space is

$$\mathcal{N}(A) = \left\{ \begin{pmatrix} 0 \\ \alpha \\ \beta \end{pmatrix} : \alpha, \beta \in \mathbb{R} \right\} \subset \mathbb{R}^3.$$

## Section 9

### Projections

# Projections

- ▶ Suppose that  $V$  and  $W$  are disjoint subspaces of  $S$  such that  $V + W = S$ , i.e.

$$x \in S \Rightarrow x = v + w$$

where  $v \in V$  and  $w \in W$  is a unique decomposition.

- ▶ Define the linear operator  $P : S \rightarrow V \subset S$  as

$$Px = P(v + w) = v$$

- ▶ Note that  $P(Px) = Pv = v$

## Projections, cont.

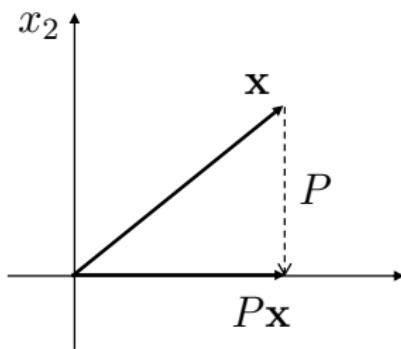
### Definition (Projection Operator)

Let  $P : S \rightarrow S$  such that  $P^2 = P$ , then  $P$  is called a projection operator or idempotent.

### Example

Let  $P = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ , then  $P \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ 0 \end{pmatrix}$

i.e.  $P$  projects elements of  $P$  onto the  $x_1$  axis:



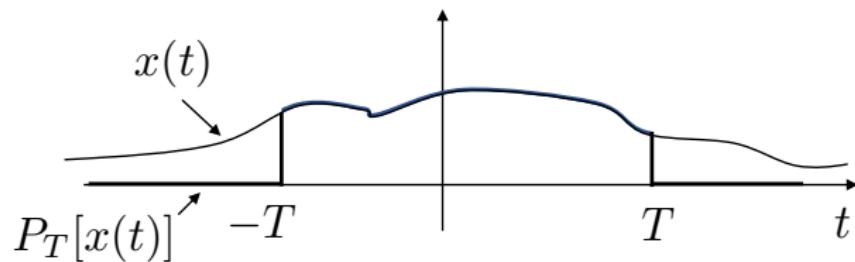
# Projections, cont.

## Example

Truncation: let

$$(P_T x)(t) = \begin{cases} x(t), & -T \leq t \leq T \\ 0, & \text{otherwise} \end{cases}$$

Then  $P_T$  projects  $x(t)$  onto its truncated function:



## Projections, cont.

Theorem (Moon 2.7)

Let  $P : S \rightarrow S$  be a projection operator, then

$$S = R(P) + N(P)$$

Proof.

Homework problem.



# Projections, cont.

## Theorem

If  $P : S \rightarrow S$  is a projection operator then so is  $(I - P) : S \rightarrow S$

Proof.

$$\begin{aligned}(I - P)^2 &= (I - P)(I - P) = \\&= I - P - P + P^2 \\&= I - P - P + P \\&= I - P\end{aligned}$$



## Projections, cont.

- ▶ Note that if  $P : S \rightarrow V$  and  $I - P : S \rightarrow W$  then  $V$  and  $W$  are disjoint and  $S = V + W$  since

$$x = \underbrace{Px}_{\in V} + \underbrace{(I - P)x}_{\in W}.$$

- ▶  $V$  and  $W$  are disjoint. If not, then  $\exists x_0 (\neq 0) \in S$  such that

$$Px_0 = (I - P)x_0 = x_0 - Px_0$$

$$2Px_0 = x_0$$

$$\Rightarrow Px_0 = \frac{1}{2}x_0$$

$$\text{and } P^2x_0 = \frac{1}{4}x_0 = \frac{1}{2}x_0 \Leftrightarrow x_0 = 0$$

## Projections, cont.

Definition (Orthogonal Projection)

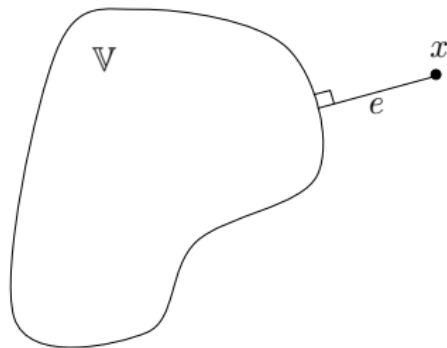
If  $V$  and  $W$  are orthogonal then  $P$  is an orthogonal projection

Theorem

$P$  is an orthogonal projection iff  $R(P) \perp N(P)$

## Applications to Engineering

Given a point  $x \in S$ , suppose that we want to approximate  $x$  by a point in  $\mathbb{V} \subset S$  (assuming  $x \notin \mathbb{V}$ ) then we want to find the point in  $\mathbb{V}$  that is closest to  $x$ .



This is given by the orthogonal projection of  $x$  onto  $\mathbb{V}$ . i.e.  
 $\langle e, v \rangle = 0 \quad \forall v \in \mathbb{V}$

# Applications to Engineering

Let  $\mathbf{n}$  be a unit vector in  $\mathbb{R}^3$  (i.e.,  $\|\mathbf{n}\| = 1$ ), then

$$\Pi_{\mathbf{n}}^{\perp} \triangleq \mathbf{n}\mathbf{n}^T$$

is a projection operation. Geometrically  $\Pi_{\mathbf{n}}^{\perp}x = \mathbf{n}\mathbf{n}^T x$  find the projection of  $x$  along the unit vector  $\mathbf{n}$

Also

$$\Pi_{\mathbf{n}} = I - \mathbf{n}\mathbf{n}^T$$

is a projection operator. Geometrically,  $\Pi_{\mathbf{n}}x$  projections  $x$  onto the 2D space that is orthogonal to  $\mathbf{n}$ .

# The Projection Theorem

## Theorem

Let  $\mathbb{S}$  be a Hilbert space and let  $\mathbb{V}$  be a closed subspace of  $\mathbb{S}$ . For any  $x \in \mathbb{S}$  there exists a unique  $v_0 \in \mathbb{V}$  closest to  $x$ ; i.e.

$$\|x - v_0\| \leq \|x - v\| \quad \forall v \in \mathbb{V}.$$

Furthermore  $v_0$  minimizes  $\|x - v_0\|$  iff  $x - v_0$  is orthogonal to  $\mathbb{V}$ .

# The Projection Theorem, proof

## Step 1. Show that $v_0$ exists.

Assume  $x \notin \mathbb{V}$  and let  $\delta = \inf_{v \in \mathbb{V}} \|x - v\|$ . We need to show that in fact  $\exists v_0 \in \mathbb{V}$  such that  $\|x - v_0\| = \delta$ .

Let  $\{v_i\}$  be a sequence in  $\mathbb{V}$  such that  $\|x - v_i\| \rightarrow \delta$  and show that  $\{v_i\}$  is Cauchy.

Need parallelogram law:

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Consider

$$\begin{aligned} & \| (v_j - x) + (x - v_i) \|^2 + \| (v_j - x) - (x - v_i) \|^2 \\ & \quad = 2\| (v_j - x) \|^2 + 2\| x - v_i \|^2 \\ \Rightarrow & \| v_j - v_i \|^2 = 2\| v_j - x \|^2 + 2\| x - v_i \|^2 - 4 \left\| \frac{(v_j + v_i)}{2} - x \right\|^2 \end{aligned}$$

## The Projection Theorem, proof

$$\begin{aligned} v_i, v_j \in \mathbb{V} &\Rightarrow \frac{v_j + v_i}{2} \in \mathbb{V} \Rightarrow \left\| \frac{(v_j - v_i)}{2} - x \right\|^2 \geq \delta^2 \\ &\Rightarrow \|v_j - v_i\|^2 \leq 2\|v_j - x\|^2 + 2\|v_i - x\|^2 - 4\delta^2 \\ \text{But } \|v_j - x\| &\rightarrow \delta \\ &\Rightarrow \|v_j - v_i\| \rightarrow 0 \end{aligned}$$

and is therefore Cauchy.

Since  $\mathbb{V}$  is a Hilbert space

$$v_i \rightarrow v_0 \in \mathbb{V}.$$

Note that this proof is not constructive, i.e. it doesn't tell you how to construct the sequence  $\{v_i\}$ .

## The Projection Theorem, proof

**Step 2. Show that**  $v_0 = \arg \min_{v \in \mathbb{V}} \|x - v\| \Rightarrow x - v_0 \perp \mathbb{V}$ .

Proof by contradiction. Suppose that  $x - v_0$  is not perpendicular to  $\mathbb{V}$ . Then there exists a  $v \in \mathbb{V}$  such that

$$\langle x - v_0, v \rangle = \delta \neq 0$$

and w.l.o.g. (why?) let  $\|v\| = 1$

Let  $z = v_0 + \delta v \in \mathbb{V}$  then

$$\begin{aligned}\|x - z\|^2 &= \|x - v_0 - \delta v\|^2 = \|x - v_0\|^2 - 2\operatorname{Re} \langle x - v_0, \delta v \rangle + \|\delta v\|^2 \\ &= \|x - v_0\|^2 - 2\delta^2 + \delta^2 < \|x - v_0\|^2\end{aligned}$$

which is a contradiction since  $v_0$  is the minimizer.

## The Projection Theorem, proof

**Step 3.** Suppose  $(x - v_0) \perp \mathbb{V}$  then  $\forall v \in \mathbb{V}$  such that  $v \neq v_0$

$$\begin{aligned}\|x - v\|^2 &= \|x - v_0 + v_0 - v\|^2 \\&= \|x - v_0\|^2 + 2\operatorname{Re} \langle x - v_0, v_0 - v \rangle + \|v_0 - v\|^2 \\&= \|x - v_0\|^2 + \|v_0 - v\|^2 \\&> \|x - v_0\|^2\end{aligned}$$

**Step 4. Uniqueness** Same as proof on page 25 of notes.

# Closed Subspace

Theorem (Moon Theorem 2.10)

Let  $\mathbb{V}$  be a closed subspace of a Hilbert space  $\mathbb{S}$ , then

$$\mathbb{S} = \mathbb{V} \oplus \mathbb{V}^\perp$$

$$\mathbb{V} = \mathbb{V}^{\perp\perp}$$

Proof.

In book. □

## Section 10

### Gram Schmidt Orthogonalization

## Application: Gram Schmidt Orthogonalization

Given a set  $T = \{p_1, \dots, p_n\}$

Find a set  $T' = \{q_1, \dots, q_{n'}\}$        $n' \leq n$  such that

$$\text{span}(T') = \text{span}(T) \text{ and } \langle q_i, q_j \rangle = \delta_{ij}$$

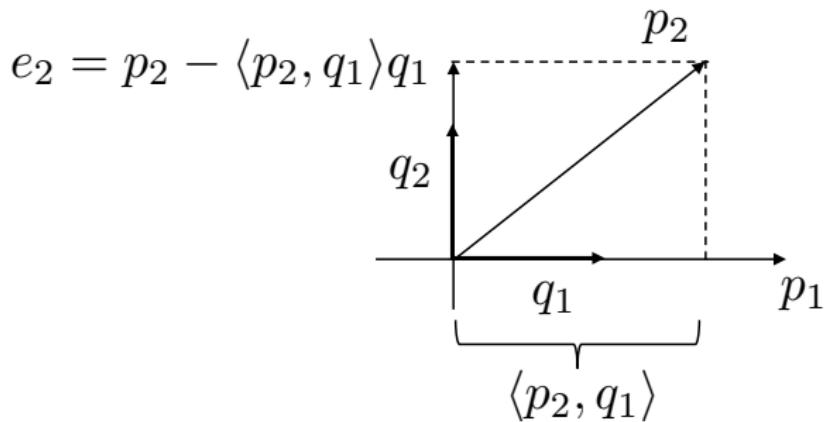
Step 1. Normalize the First Vector

$$q_1 = \frac{p_1}{\|p_1\|} \quad (\text{i.e. } \langle q_1, q_1 \rangle = 1)$$

## Application: Gram Schmidt Orthogonalization, cont

Step 2. Let  $e_2$  be  $p_2$  minus the projection of  $p_2$  on  $q_1$  i.e.

$$e_2 = p_2 - \langle p_2, q_1 \rangle q_1$$



Then normalize  $e_2$ :

$$q_2 = \frac{e_2}{\|e_2\|}$$

## Application: Gram Schmidt Orthogonalization, cont

Step 3. Let  $e_k$  be  $p_k$  minus the projection of  $p_k$  on  $q_1, \dots, q_{k-1}$ :

$$e_k = p_k - \sum_{j=1}^{k-1} \langle p_k, q_j \rangle q_j \Rightarrow q_k = \frac{e_k}{\|e_k\|}$$

## Example: Gram Schmidt Orthogonalization

Given the set

$$T = \{p_1, p_2, p_3\} \triangleq \left\{ \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$$

find a set  $T' = \{q_1, q_2, q_3\}$  where the vectors in  $T'$  are orthonormal and  $\text{span}(T) = \text{span}(T')$ .

$$q_1 = \frac{p_1}{\|p_1\|} = \frac{\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}}{\sqrt{4+0+0}} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}.$$

## Example: Gram Schmidt Orthogonalization, cont.

$$e_2 = p_2 - \langle p_2, q_1 \rangle q_1$$

$$= \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}^\top \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix} - 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$$

$$\text{Therefore } q_2 = \frac{e_2}{\|e_2\|} = \frac{\begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}}{\sqrt{4}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

## Example: Gram Schmidt Orthogonalization, cont.

$$e_3 = p_3 - \langle p_3, q_1 \rangle q_1 - \langle p_3, q_2 \rangle q_2$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}^\top \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}^\top \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - 1 \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} - 2 \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

$$\text{Therefore } q_3 = \frac{e_3}{\|e_3\|} = \frac{\begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}}{3} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

## Example: Gram Schmidt Orthogonalization, cont.

Therefore, the Gram Schmidt orthonormalization of

$$T = \{p_1, p_2, p_3\} \triangleq \left\{ \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \right\}$$

is

$$T' = \{q_1, q_2, q_3\} = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Note that  $\text{span}(T) = \text{span}(T')$ .

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 3.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Approximation Theory

Dual Approximation

Underdetermined Problems

Generalized Fourier Series

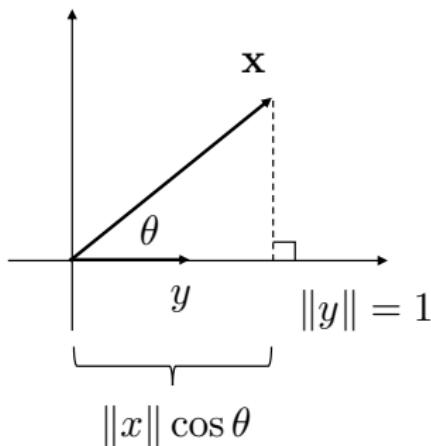
## Section 1

# Approximation Theory

# Projection and Inner Product

- ▶ How does inner product represent a projection?
- ▶ Recall that

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



- ▶ In 2-D  $\langle x, y \rangle$  represents the length of the projection of  $x$  in the direction of  $y$ .
- ▶ In general, inner products represent (non-orthogonal) projection of one vector onto another.

## Approximation Problem

- ▶ Let  $\mathbb{S}$  be a Hilbert space, and let  $x \in \mathbb{S}$ .
- ▶ Let  $\{p_1, \dots, p_n\}$  be a set of vectors, all in  $\mathbb{S}$ .
- ▶ Find  $\hat{x} \in \text{span}\{p_1, \dots, p_n\}$  that minimizes  $\|x - \hat{x}\|$ .

## Approximation Problem, cont

- ▶ Let  $\hat{x} = c_1 p_1 + \dots + c_n p_n \in \text{span}\{p_1, \dots, p_n\}$ .
- ▶ By the projection theorem, the error

$$\begin{aligned} e &= x - \hat{x} \\ &= x - c_1 p_1 - \dots - c_n p_n \end{aligned}$$

is minimized if

$$e \perp \text{span}\{p_1, \dots, p_n\}.$$

## Approximation Problem, cont

$$e \perp \text{span}\{p_1, \dots, p_n\}.$$

iff

$$\langle e, p_1 \rangle = 0$$

$$\langle e, p_2 \rangle = 0$$

⋮

$$\langle e, p_n \rangle = 0$$

iff

$$\langle x - c_1 p_1 - \dots - c_n p_n, p_1 \rangle = 0$$

⋮

$$\langle x - c_1 p_1 - \dots - c_n p_n, p_n \rangle = 0$$

## Approximation Problem, cont

By properties of the inner product we can write this as

$$\langle x, p_1 \rangle - c_1 \langle p_1, p_1 \rangle - \cdots - c_n \langle p_n, p_1 \rangle = 0$$

⋮

$$\langle x, p_n \rangle - c_1 \langle p_1, p_n \rangle - \cdots - c_n \langle p_n, p_n \rangle = 0$$

or in matrix notation,

$$\underbrace{\begin{pmatrix} \langle p_1, p_1 \rangle & \cdots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \cdots & \langle p_n, p_n \rangle \end{pmatrix}}_R \underbrace{\begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}}_c = \underbrace{\begin{pmatrix} \langle x, p_1 \rangle \\ \vdots \\ \langle x, p_n \rangle \end{pmatrix}}_p$$

$R$  is called the Grammian of the set  $\{p_1, \dots, p_n\}$ .

# The Grammian of a set

## Definition (Grammian)

Given a set  $\{p_1, \dots, p_n\}$  of vectors in  $\mathbb{S}$ , the Grammian of the set is the matrix

$$R = \begin{pmatrix} \langle p_1, p_1 \rangle & \cdots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \cdots & \langle p_n, p_n \rangle \end{pmatrix}$$

Note that  $R^H = R$

We also have the following theorem:

## Theorem (Moon, Theorem 3.1)

*The Grammian  $R$  is positive definite iff the set of vectors  $\{p_1, \dots, p_n\}$  are linearly independent.*

## Proof

Let  $y \in \mathbb{S}$  then

$$\begin{aligned} y^H R y &= (\bar{y}_1 \cdots \bar{y}_n) \begin{pmatrix} \langle p_1, p_1 \rangle & \dots & \langle p_n, p_1 \rangle \\ \vdots & & \vdots \\ \langle p_1, p_n \rangle & \dots & \langle p_n, p_n \rangle \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \left( \sum_{i=1}^n \bar{y}_i \langle p_1, p_i \rangle \dots \bar{y}_i \sum_{i=1}^n \langle p_n, p_i \rangle \right) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \\ &= \sum_{j=1}^n \sum_{i=1}^n \bar{y}_i y_j \langle p_j, p_i \rangle \\ &= \left\langle \sum y_j p_j, \sum y_i p_i \right\rangle = \left\| \sum y_i p_i \right\|^2 \geq 0 \end{aligned}$$

Therefore  $R$  is always positive semi-definite.

## Proof, cont.

( $\Rightarrow$ ): Suppose that  $R$  is pd then

$$y^H R y = \left\| \sum y_i p_i \right\|^2 > 0$$

$\Rightarrow \sum y_i p_i \neq 0$  for all nonzero  $y \in \mathbb{S}$

$\Rightarrow \{p_1, \dots, p_n\}$  is linearly independent

( $\Leftarrow$ ): Conversely suppose  $\{p_1, \dots, p_n\}$  is linearly independent, but  $R$  is only psd.  $R$  is psd implies that  $\exists y \neq 0$  such that

$$y^H R y = \left\| \sum y_i p_i \right\|^2 = 0$$

$$\Rightarrow \sum y_i p_i = 0$$

$\Rightarrow \{p_1, \dots, p_n\}$  is linearly dependent.

Which contradicts the assumption that  $R$  is psd.

# Orthogonality Theorem

Theorem (Moon, Theorem 3.2)

Let  $p_1, p_2, \dots, p_n$  be data vectors (or basis vectors) in a Hilbert space  $\mathbb{S}$ . Let  $x \in \mathbb{S}$ . Let  $e$  be defined as

$$e \stackrel{\triangle}{=} x - \hat{x} = x - \sum_{j=1}^n c_j p_j,$$

then  $e$  is minimized when it is orthogonal to each of the data vectors, i.e.

$$\langle e, p_j \rangle = 0 \quad j = 1, \dots, n$$

Equivalently

$$R\mathbf{c} = \mathbf{p}.$$

Proof.

Follows directly from projection theorem. □

## Calculus-Based Approach (Alternative proof)

Rather than using the projection theorem, we can derive the same result using calculus.

**Problem Statement:** Let  $\mathbf{e} = x - \sum_{i=1}^n c_i p_i$ . Find  $\mathbf{c} = (c_1, \dots, c_n)^\top$  that minimizes  $\|\mathbf{e}\|$ .

**Solution:** First note that minimizing  $\|\mathbf{e}\|^2$  is equivalent to minimizing  $\|\mathbf{e}\|$ . Also note that

$$\begin{aligned}\|\mathbf{e}\|^2 &= \left\langle x - \sum c_j p_j, x - \sum c_j p_j \right\rangle \\ &= \|x\|^2 - 2\operatorname{Re}\left\{\sum_{i=1}^n \bar{c}_i \langle x, p_i \rangle\right\} + \sum \sum c_j \bar{c}_i \langle p_j, p_i \rangle \\ &= \|x\|^2 - 2\operatorname{Re}\{\mathbf{c}^H \mathbf{p}\} + \mathbf{c}^H R \mathbf{c}.\end{aligned}$$

## Calculus-Based Approach, cont.

To minimize

$$\|e\|^2 = \|x\|^2 - 2\operatorname{Re}\{\mathbf{c}^H \mathbf{p}\} + \mathbf{c}^H R \mathbf{c}$$

differentiate with respect to  $\mathbf{c}$  and set to zero. This will be a local minima if the second derivative is psd.

## Calculus-Based Approach, cont.

From Moon Appendix we have

$$\frac{\partial}{\partial \bar{\mathbf{c}}} \operatorname{Re}\{\mathbf{c}^H \mathbf{p}\} = \frac{1}{2} \mathbf{p}$$
$$\frac{\partial}{\partial \bar{\mathbf{c}}} \mathbf{c}^H R \mathbf{c} = R \mathbf{c}$$

Therefore

$$\frac{\partial \|e\|^2}{\partial \bar{\mathbf{c}}} = -\mathbf{p} + R \mathbf{c} = 0 \quad \Rightarrow \quad R \mathbf{c} = \mathbf{p}$$

In addition, we have that

$$\frac{\partial^2 \|e\|^2}{\partial \bar{\mathbf{c}}} = R \geq 0.$$

Therefore the solution of  $R \mathbf{c} = \mathbf{p}$  minimize  $\|e\|$ .

$R \mathbf{c} = \mathbf{p}$  is the same equation we obtained using the projection theorem.

# Matrix Representation

- ▶ Stack the vectors  $\{p_1, \dots, p_n\}$  in a matrix

$$A = (p_1 \quad p_2 \quad \dots \quad p_n)$$
$$\mathbf{c} = (c_1 \quad c_2 \quad \dots \quad c_n)^\top$$

- ▶ Then  $\hat{x} = \sum c_j p_j = A\mathbf{c}$ .
- ▶ Therefore  $\mathbf{e} = x - \hat{x} = x - A\mathbf{c}$ .

## Matrix Representation, cont.

- ▶ Project  $\mathbf{e}$  onto  $\{p_1 \dots p_n\}$ :

$$\langle x - A\mathbf{c}, p_1 \rangle = p_1^H(x - A\mathbf{c}) = 0$$

⋮

$$\langle x - A\mathbf{c}, p_n \rangle = p_n^H(x - A\mathbf{c}) = 0$$

$$A^H = \begin{bmatrix} p_1^H \\ \vdots \\ p_n^H \end{bmatrix}.$$

- ▶ Note that  $A^H = \begin{bmatrix} p_1^H \\ \vdots \\ p_n^H \end{bmatrix}$ .
- ▶ Rewrite as

$$\begin{aligned} A^H(x - A\mathbf{c}) &= 0 \\ \Rightarrow \underbrace{A^H A}_{R} \mathbf{c} &= \underbrace{A^H x}_{\mathbf{p}} \end{aligned}$$

## Matrix Representation, cont.

- ▶ If  $\{p_1, \dots, p_n\}$  are linearly independent then  $R > 0$  which implies that  $R^{-1}$  exists, so

$$\mathbf{c} = (A^H A)^{-1} A^H x$$

- ▶ Since  $\hat{x} = A\mathbf{c}$  we have that

$$\hat{x} = A(A^H A)^{-1} A^H x$$

is the best approximation to  $x$  in  $\text{span}\{p_1, \dots, p_n\}$ .

- ▶ **Fact:**  $P_A = A(A^H A)^{-1} A^H$  is a projection operator from  $S$  to  $\text{span}\{p_1, \dots, p_n\}$

## Application: Polynomial Approximation

- ▶ Suppose you are given a real continuous function  $f(t)$  and you would like to approximate it by an  $m^{th}$  order polynomial on the interval  $[a, b]$ .
- ▶ Let the basis vectors be  $\{1, t, \dots, t^m\}$ .
- ▶ Then  $\hat{f}(t) = c_1 + c_2 t + \dots + c_{m+1} t^m$
- ▶ Define the inner product as  $\langle f, g \rangle = \int_a^b f(t)g(t)dt$

## Application: Polynomial Approximation, cont.

Then the orthogonality theorem implies that the “best” approximation is given by

$$\langle f - \hat{f}, 1 \rangle = 0$$

⋮

$$\langle f - \hat{f}, t^m \rangle = 0$$

or

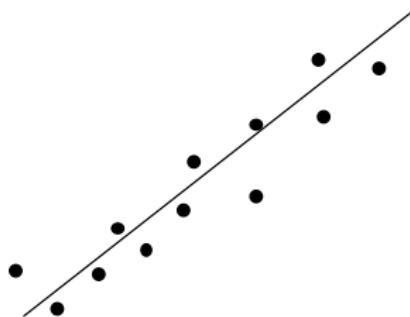
$$\underbrace{\begin{pmatrix} \langle 1, 1 \rangle & \cdots & \langle t^m, 1 \rangle \\ \vdots & & \vdots \\ \langle 1, t^m \rangle & \cdots & \langle t^m, t^m \rangle \end{pmatrix}}_{\text{Grammian Matrix}} \begin{pmatrix} c_1 \\ \vdots \\ c_{m+1} \end{pmatrix} = \begin{pmatrix} \langle f, 1 \rangle \\ \vdots \\ \langle f, t^m \rangle \end{pmatrix}$$

or

$$R\mathbf{c} = \mathbf{p}.$$

## Application: Linear Regression

- ▶ Suppose you have a number of data points that you are trying to fit to a line.



- ▶ Given  $(x_i, y_i) \quad i = 1, \dots, N$
- ▶ The equation for a line is  $y = ax + b$
- ▶ **Problem:** Find  $a$  and  $b$  that minimizes the mean squared error  $\sum_{i=1}^N |y_i - ax_i - b|^2$

## Application: Linear Regression, cont.

- ▶ For each data point we have

$$e_i = y_i - ax_i - b$$

where  $e_i$  is the error for the  $i^{th}$  data point.

- ▶ Let

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_N \end{pmatrix}, \quad A = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_N & 1 \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} a \\ b \end{pmatrix}$$

- ▶ Then  $\mathbf{e} = \mathbf{y} - A\mathbf{c}$ .

## Application: Linear Regression, cont.

- ▶ Project the error  $\mathbf{e}$  on the data vector (columns of  $A$ ) and set to zero:

$$A^H \mathbf{e} = A^H(\mathbf{y} - A\mathbf{c}) = 0$$

- ▶ Therefore

$$A^H A \mathbf{c} = A^H \mathbf{y}$$

- ▶ Giving the minimum least squares solution

$$\mathbf{c} = (A^H A)^{-1} A^H \mathbf{y}.$$

## Section 2

# Dual Approximation

## Dual Approximation

This section develops an approach that allows approximation in infinite dimensional spaces with finite constraints.

For matrices, we will solve the problem

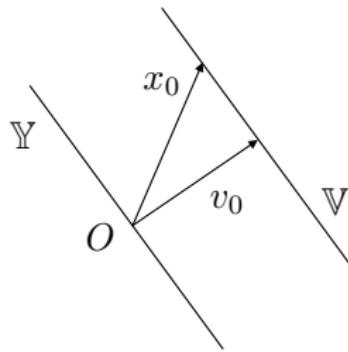
$$\min \|x\|$$

$$\text{s.t. } Ax = b$$

## Dual Approximation, cont.

### Definition (Affine Space)

Let  $\mathbb{Y}$  be a subspace of  $\mathbb{S}$  and let  $x_0 \in \mathbb{S}$ . The set  $\mathbb{V} = x_0 + \mathbb{Y}$  is called a linear variety or an affine space.



The projection theorem says that there exists a  $v_0 \in \mathbb{V}$  such that  $v_0 = \arg \min_{v \in \mathbb{V}} \|v\|$  such that  $v_0 \perp \mathbb{Y}$ .

## Dual Approximation, cont.

Let  $M = \text{span}\{y_1, \dots, y_m\}$  then  $\dim(M) < \infty$ .

If  $\dim(\mathbb{S}) = \infty$  then  $\dim(M^\perp) = \infty$  where  $M^\perp$  is the set of all  $x \in \mathbb{S}$  such that

$$\langle x, y_1 \rangle = 0$$

⋮

$$\langle x, y_m \rangle = 0$$

## Dual Approximation, cont.

Now suppose that there are  $m$  inner product constraints:

$$\langle x, y_1 \rangle = a_1$$

⋮

$$\langle x, y_m \rangle = a_n$$

If  $\exists x_0$  that satisfies the constraints then so does  $x_0 + v$  where  $v \in M^\perp$  since

$$\begin{aligned}\langle x_0 + v, y_j \rangle &= \langle x_0, y_j \rangle + \langle v, y_j \rangle \\ &= \langle x_0, y_j \rangle \\ &= a_j\end{aligned}$$

Therefore all solutions are in the (infinite dimensional) affine space

$$v = x_0 + M^\perp$$

## Dual Approximation, cont.

Theorem (Moon Theorem 3.4)

Let  $\{y_1, \dots, y_m\}$  be linearly independent in a Hilbert space  $\mathbb{S}$ , and let  $M = \text{span}\{y_1, \dots, y_m\}$ . The solution of the problem

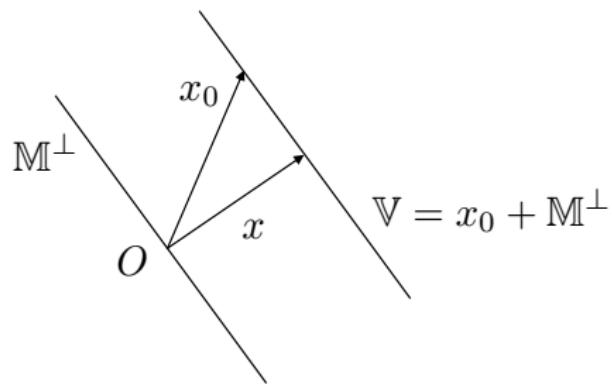
$$\begin{aligned} \min_{x \in \mathbb{S}} \quad & \|x\|^2 \\ \text{s.t.} \quad & \langle x, y_1 \rangle = \alpha_1, \\ & \vdots, \\ & \langle x, y_m \rangle = \alpha_m \end{aligned}$$

is an element of  $M$ , i.e.,  $\hat{x} = \arg \min_{x \in \mathbb{S}} \|x\|^2 = \sum_{i=1}^m c_i y_i$ , where  $\mathbf{c}$  satisfies  $R\mathbf{c} = \boldsymbol{\alpha}$ , where  $R$  is the Grammian and

$$\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_m)^\top.$$

## Proof:

From the previous discussion, the solution lies in the affine space  $\mathbb{V} = x_0 + M^\perp$  for some  $x_0 \in \mathbb{S}$ .



The minimum norm solution is orthogonal to  $M^\perp$  i.e.  
 $\hat{x} \perp M^\perp \Rightarrow \hat{x} \in M^{\perp\perp} = M$

So  $\hat{x}$  is of the form  $\hat{x} = \sum_{j=1}^m c_j y_j$

## Proof, cont.

Now projecting  $x$  onto  $M$  gives

$$\langle \hat{x}, y_1 \rangle = \left\langle \sum c_j y_j, y_1 \right\rangle = \sum c_j \langle y_j, y_1 \rangle = \alpha_1$$

$$\vdots = \vdots = \vdots = \vdots$$

$$\langle \hat{x}, y_m \rangle = \left\langle \sum c_j y_j, y_m \right\rangle = \sum c_j \langle y_j, y_m \rangle = \alpha_m$$

rewriting in matrix notation gives

$$R\mathbf{c} = \boldsymbol{\alpha}$$

## Dual Approximation, Example

Given the differential equation

$$\ddot{y} + 6\dot{y} + 8y = 4\dot{u} + 10u, \quad y(0) = \dot{y}(0) = 0$$

Solve the following optimal control problem:

$$\begin{aligned} & \min_{u \in L_2} \|u\|^2 \\ \text{s.t. } & y(1) = 1, \\ & \int_0^1 y(t) dt = 0 \end{aligned}$$

## Dual Approximation, Example, cont.

The corresponding transfer function is

$$\begin{aligned}H(s) &= \frac{4s + 10}{s^2 + 6s + 8} = \frac{1}{s+2} + \frac{3}{s+4} \\ \Rightarrow h(t) &= e^{-2t} + 3e^{-4t} \\ \Rightarrow y(t) &= \int_0^t \left[ e^{-2(t-\tau)} + 3e^{-4(t-\tau)} \right] u(\tau) d\tau\end{aligned}$$

Define the following inner product

$$\langle f(t), g(t) \rangle = \int_0^1 f(\tau)g(\tau) d\tau$$

then  $y(1) = 1$  can be written as

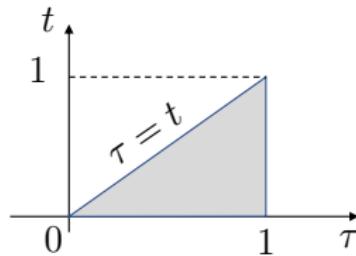
$$\int_0^1 \left[ e^{-2(1-\tau)} + 3e^{-4(1-\tau)} \right] u(\tau) d\tau = \langle u, y_1 \rangle = 1$$

where  $y_1(t) = e^{-1(1-t)} + 3e^{-4(1-t)}$

## Dual Approximation, Example, cont.

The second constraint is of the form

$$\int_0^1 y(t)dt = \int_{t=0}^{t=1} \int_{\tau=0}^{\tau=t} h(t-\tau)u(\tau)d\tau dt = 0$$



Changing order of integration gives

$$= \int_{\tau=0}^1 \left[ \int_{t=\tau}^1 h(t-\tau)dt \right] u(\tau)d\tau.$$

## Dual Approximation, Example, cont.

Letting  $\sigma = t - \tau \Rightarrow t = \sigma + \tau \Rightarrow dt = d\sigma$  gives

$$\begin{aligned} &= \int_{\tau=0}^1 \left[ \int_{\sigma=0}^{\sigma=1-\tau} h(\sigma) d\sigma \right] u(\tau) d\tau \\ &= \int_{\tau=0}^1 \left( \frac{5}{4} - \frac{3}{4} e^{-4(1-\tau)} - \frac{1}{2} e^{-2(1-\tau)} \right) u(\tau) d\tau \\ &= \langle u, y_2 \rangle = 0 \end{aligned}$$

where

$$y_2(t) = \frac{5}{4} - \frac{3}{4} e^{-4(1-t)} - \frac{1}{2} e^{-2(1-t)}$$

so we have that

$$\begin{aligned} \langle u, y_1 \rangle &= 1 \\ \langle u, y_2 \rangle &= 0 \end{aligned}$$

and we want to minimize  $\|u\|_{L_2[0,1]}^2$

## Dual Approximation, Example, cont.

Let  $M = \text{span}\{y_1, y_2\}$ .

By Theorem 3.4

$$u \in M \Rightarrow u(t) = c_1 y_1(t) + c_2 y_2(t)$$

where

$$\begin{pmatrix} \langle y_1, y_1 \rangle & \langle y_2, y_1 \rangle \\ \langle y_1, y_2 \rangle & \langle y_2, y_2 \rangle \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

## Section 3

### Underdetermined Problems

## Section 3.15: Underdetermined Problems

Given  $Ax = b$  where  $A$  is fat, i.e. fewer equations than unknowns, solve the following problem:

$$\begin{aligned} & \min && \|x\|_2 \\ & \text{s.t.} && Ax = b \end{aligned}$$

where  $A = \begin{pmatrix} y_1^H \\ \vdots \\ y_m^H \end{pmatrix}$ ,  $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ ,  
and  $y_i \in \mathbb{C}^n$  and  $b \in \mathbb{C}^m$ .

## Section 3.15: Underdetermined Problems, cont.

$Ax = b$  is a set of inner product constraints

$$y_1^H x = b_1$$

⋮

$$y_m^H x = b_m$$

Let  $M = \text{span}\{y_1, \dots, y_m\}$ .

Theorem 3.4 implies that  $x_0 = \arg \min \|x\| \in M$

$$\Rightarrow x_0 = \sum c_j y_j = A^H c$$

and that  $c$  satisfies

$$Rc = \mathbf{b} \text{ where } R = AA^H$$

if  $\{y_1, \dots, y_m\}$  are linearly independent then

$$\mathbf{c} = (AA^H)^{-1}\mathbf{b} \quad \Rightarrow \quad x_0 = \underbrace{A^H(AA^H)^{-1}}_{\text{pseudo-inverse}} \mathbf{b}$$

## Section 4

### Generalized Fourier Series

## Section 3.17: Generalized Fourier Series

Topic of interest:  $L_2$  function approximation

### Definition (Complete Basis)

An orthonormal set  $\{p_i, i = 1, \dots, \infty\}$  in a Hilbert space  $\mathbb{S}$  is a complete basis or total basis if  $\forall x \in \mathbb{S}$

$$x = \sum_{i=1}^{\infty} \langle x, p_i \rangle p_i$$

Note that if  $x = \sum_{i=1}^{\infty} c_i p_i$  and  $\langle p_i, p_j \rangle = \delta_{ij}$  then

$$\langle x, p_j \rangle = \sum_{i=1}^{\infty} c_i \langle p_i, p_j \rangle = c_j$$

$$\Rightarrow c_j = \langle x, p_j \rangle$$

## Generalized Fourier Series, cont.

Therefore we can write

$$x = \sum_{i=1}^{\infty} \langle x, p_i \rangle p_i.$$

Most common example: standard Fourier basis

$$P_n(t) = \frac{1}{\sqrt{T}} e^{j(\frac{2\pi}{T})nt}$$

Any function  $f \in L_2[0, T]$  can be written as

$$f(t) = \sum_{n=-\infty}^{\infty} c_n \frac{1}{T} e^{j(\frac{2\pi}{T})nt}$$

where the coefficients are given as

$$c_n = \left\langle f, \frac{1}{\sqrt{T}} e^{j(\frac{2\pi}{T})nt} \right\rangle \triangleq \frac{1}{\sqrt{T}} \int_0^T f(t) e^{j(\frac{2\pi}{T})nt} dt$$

## Generalized Fourier Series, cont.

Actually it is common to place the  $\frac{1}{\sqrt{T}}$ 's together letting

$$f(t) = \sum_{n=-\infty}^{\infty} b_n e^{j(\frac{2\pi}{T})nt} \text{ where}$$

$$b_n = \left\langle f(t), \frac{1}{T} e^{j(\frac{2\pi}{T})nt} \right\rangle = \frac{1}{T} \int_0^T f(t) e^{-j(\frac{2\pi}{T})nt} dt$$

Generalized Fourier series hold for any complete basis, i.e.

$$x = \sum_{j=1}^{\infty} \langle x, p_j \rangle p_j$$

## Generalized Fourier Series, cont.

There are two important relationships between a function and its Fourier transform.

### Theorem (Bessel's Inequality)

Suppose  $\{p_1, p_2, \dots\}$  is orthonormal but not necessarily complete and let

$$c = \{\langle x, p_1 \rangle, \langle x, p_2 \rangle, \dots\} = \{c_1, c_2, \dots\}$$

then

$$\|c\|_{\ell_2} \leq \|x\|_{L_2}$$

## Proof:

$$\begin{aligned} 0 \leq \left\| x - \sum c_j p_j \right\|_{L_2}^2 &= \left\langle x - \sum c_j p_j, x - \sum c_j p_j \right\rangle_{L_2} \\ &= \langle x, x \rangle_{L_2} - \sum \bar{c}_j \langle x, p_j \rangle_{L_2} \\ &\quad - \sum c_j \langle x, \bar{p}_j \rangle_{L_2} + \sum \sum c_j \bar{c}_k \langle p_j, p_k \rangle_{L_2} \\ &= \|x\|_{L_2}^2 - \sum \bar{c}_j c_j - \sum c_j \bar{c}_j + \sum c_j \bar{c}_j \\ &= \|x\|_{L_2}^2 - \sum_{j=1}^{\infty} |c_j|^2 \\ &= \|x\|_{L_2}^2 - \|c\|_{\ell_2}^2 \\ \Rightarrow \|c\|_{\ell_2}^2 &\leq \|x\|_{L_2}^2 \end{aligned}$$

## Generalized Fourier Series, cont.

Theorem (Parseval's Equality)

If  $T = \{p_1, p_2, \dots\}$  is complete then

$$\|x\|_{L_2}^2 = \|c\|_{\ell_2}^2$$

Proof.

If  $T$  is complete then

$$\left\| x - \sum c_j p_j \right\|^2 = 0$$

and the result follows from the proof of Bessel's inequality .

□

## Significance of Parseval's Equality

$\|x\|_{L_2}^2 = \|c\|_{\ell_2}^2$  says that the energy in a signal (i.e.  $\|x\|_{L_2}$ ) is equal to the energy in the Fourier coefficients (i.e.  $\|c\|_{\ell_2}^2$ ).

This relationship between  $x$  and its transform  $c$  is written as

$$x \xleftrightarrow{\mathcal{F}} c.$$

## Significance of Parseval's Equality, cont.

**Lemma (Moon Lemma 3.1)**

If  $x \xleftrightarrow{\mathcal{F}} c$  and  $y \xleftrightarrow{\mathcal{F}} b$  for the same complete basis  $\{p_1, p_2, \dots\}$  then

$$\langle x, y \rangle_{L_2} = \langle c, b \rangle_{\ell_2}.$$

**Proof.**

Let  $x = \sum_{i=1}^{\infty} c_i p_i$ , and  $y = \sum_{i=1}^{\infty} b_i p_i$  then

$$\begin{aligned}\langle x, y \rangle_{L_2} &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} c_i \bar{b}_j \langle p_i, p_j \rangle \\ &= \sum_{i=1}^{\infty} c_i \bar{b}_i \\ &= \langle c, b \rangle_{\ell_2}\end{aligned}$$

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 4.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Linear Operators

Neumann Expansion

Matrix Norms

Adjoint Operators

Fundamental Subspaces

Matrix Inverses

Matrix Condition Number

Schur Complement and the Matrix Inversion Lemma

Recursive Least Squares Filtering

# Section 1

## Linear Operators

# Linear Operators

Recall from Chapter 3 the definition of a Linear operator:

## Definition

Let  $\mathbb{X}$  and  $\mathbb{Y}$  be vector spaces, then  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  is a linear operator if

$$\mathcal{A}[\alpha_1 x_1 + \alpha_2 x_2] = \alpha_1 \mathcal{A}[x_1] + \alpha_2 \mathcal{A}[x_2]$$

$\forall x_1, x_2 \in \mathbb{X}$  and  $\forall \alpha_1, \alpha_2 \in \mathbb{F}$

See chapter 2 notes (slides 79–83) for examples of linear operators.

# Norm of a Linear Operator

An important concept is the norm of an operator. There are several ways to define norms for operators. The most important is the “induced” or “subordinate” norm.

## Definition

Let  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  then

$$\begin{aligned}\|\mathcal{A}\| &= \sup_{x \neq 0} \frac{\|\mathcal{A}[x]\|_{\mathbb{Y}}}{\|x\|_{\mathbb{X}}} \\ &= \sup_{\|x\|_{\mathbb{X}}=1} \|\mathcal{A}[x]\|_{\mathbb{Y}}\end{aligned}$$

Different norms on  $\mathcal{A}$  are defined by taking different norms in  $\mathbb{X}$  and  $\mathbb{Y}$ .

# Norm of a Linear Operator, Examples

## Example

Let  $\mathcal{A} : L_2 \rightarrow L_2$  then

$$\begin{aligned}\|\mathcal{A}\|_2 &= \sup_{x \neq 0} \frac{\|\mathcal{A}[x]\|_{L_2}}{\|x\|_{L_2}} \\ &= \sup_{\|x\|_{L_2}=1} \|\mathcal{A}[x]\|_{L_2}\end{aligned}$$

## Example

Let  $\mathcal{A} : L_\infty \rightarrow L_\infty$  then

$$\begin{aligned}\|\mathcal{A}\|_\infty &= \sup_{x \neq 0} \frac{\|\mathcal{A}[x]\|_{L_\infty}}{\|x\|_{L_\infty}} \\ &= \sup_{\|x\|_{L_\infty}=1} \|\mathcal{A}[x]\|_{L_\infty}\end{aligned}$$

# Norm of a Linear Operator, Examples

## Example

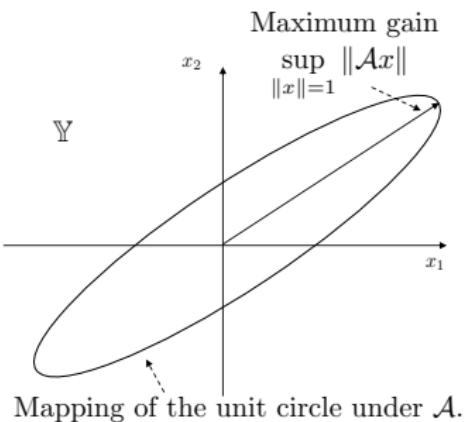
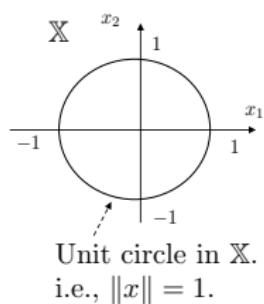
Let  $\mathcal{A} : L_p \rightarrow L_p$  then

$$\begin{aligned}\|\mathcal{A}\|_p &= \sup_{x \neq 0} \frac{\|\mathcal{A}[x]\|_{L_p}}{\|x\|_{L_p}} \\ &= \sup_{\|x\|_{L_p}=1} \|\mathcal{A}[x]\|_{L_p}\end{aligned}$$

Why is it called the induced or subordinate norm? The norm on the operator is induced by the vector norm.

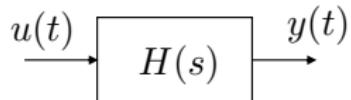
# Norm of a Linear Operator, Geometric Interpretation

$$\|A\| = \sup_{\|x\|=1} \|Ax\|$$



# Norm of a Linear Operator, System Interpretation

Given a linear system



The norm of the system  $H(s)$  is the maximum gain of the system.

## Norm of BIBO System

Let  $\mathcal{A} : L_\infty \rightarrow L_\infty$  be an LTI system that is BIBO stable with impulse response  $h(t)$ , then

$$\begin{aligned}y(t) &= \int_0^t h(t-\tau)u(\tau)d\tau \\&\triangleq \mathcal{A}[u]\end{aligned}$$

Find  $\|\mathcal{A}\|_\infty$ .

# Norm of BIBO System, cont

Lemma

$$\begin{aligned}\|\mathcal{A}\|_{\infty} &= \|h\|_{L_1[0,\infty]} \\ &\triangleq \int_0^{\infty} |h(t)| dt\end{aligned}$$

Proof.

We need to prove two things

1.  $\|\mathcal{A}\|_{\infty} \leq \int_0^{\infty} |h(t)| dt$

2.  $\int_0^{\infty} |h(t)| dt \leq \|\mathcal{A}\|_{\infty}$



# Norm of BIBO System, Proof

## Proof of 1.

$$\begin{aligned}\sup_{\|x\|_\infty=1} \|\mathcal{A}[u]\|_\infty &= \sup_{\|u\|_\infty=1} \left\| \int_0^t h(t-\tau)u(\tau)d\tau \right\|_\infty \\&= \sup_{\|u\|_\infty=1} \left[ \sup_{t>0} \left| \int_0^t h(t-\tau)u(\tau)d\tau \right| \right] \\&\leq \sup_{\|u\|_\infty=1} \left[ \sup_{t>0} \int_0^t |h(t-\tau)u(\tau)| d\tau \right] \\&\leq \sup_{\|u\|_\infty=1} \left[ \|u\|_\infty \sup_{t>0} \int_0^t |h(t-\tau)| d\tau \right] \\&\leq \int_0^\infty |h(\tau)| d\tau = \|h\|_{L_1[0,\infty]}$$

## Norm of BIBO System, Proof

### Proof of 2.

$$\text{Let } \hat{u}_t(\tau) = \begin{cases} 1 & \text{if } h(t - \tau) \geq 0 \\ -1 & \text{otherwise} \end{cases}.$$

Note that  $\|\hat{u}_t\|_\infty = 1 \ \forall t > 0$ , we have that

$$\int_0^t h(t - \tau) \hat{u}_t(\tau) d\tau = \int_0^t |h(t - \tau)| d\tau.$$

Therefore for this particular choice of  $\hat{u}_t$  we have that

$$\sup_{t>0} \left[ \int_0^t |h(t - \tau)| d\tau \right] = \|A\hat{u}_\infty\|_\infty = \int_0^\infty |h(\tau)| d\tau.$$

By definition of sup

$$\int_0^\infty |h(\tau)| d\tau = \|A\hat{u}_\infty\|_\infty \leq \sup_{\|u\|=1} \|Au\|_\infty.$$

# Operator Norm: Proof Technique

The proof technique shown here is the general approach to show that the norm of an operator is some value.

Suppose that you would like to prove that

$$\|\mathcal{A}\| = M.$$

You need to show two things

1.  $\|\mathcal{A}\| \leq M$
2.  $M \leq \|\mathcal{A}\|$ .

## Operator Norm: Proof Technique

To show (1) use triangle and other inequalities to show that

$$\|\mathcal{A}x\| \leq M \|x\|$$

which implies that

$$\sup_{\|x\|=1} \|\mathcal{A}x\| \leq \sup_{\|x\|=1} M \|x\| = M$$

To show (2), construct a specific  $\hat{x}$  such that

$$\|\hat{x}\| = 1 \text{ and } \|\mathcal{A}\hat{x}\| = M.$$

This implies that

$$M \leq \sup_{\|x\|=1} \|\mathcal{A}x\| = \|\mathcal{A}\|.$$

# Properties of Linear Operators

## Lemma

*For any induced operator norm,*

$$\|\mathcal{A}x\| \leq \|\mathcal{A}\| \|x\|.$$

Proof.

$$\|\mathcal{A}\| = \sup_{x \neq 0} \frac{\|\mathcal{A}x\|}{\|x\|}.$$

Therefore for any  $x \neq 0$  we must have that

$$\begin{aligned}\|\mathcal{A}\| &\geq \frac{\|\mathcal{A}x\|}{\|x\|} \\ \Rightarrow \|\mathcal{A}x\| &\leq \|\mathcal{A}\| \|x\|.\end{aligned}$$



# Properties of Linear Operators, cont

## Lemma

All induced operator norms satisfy the “submultiplicative property,” i.e.,

$$\|\mathcal{A}\mathcal{B}\| \leq \|\mathcal{A}\| \|\mathcal{B}\|$$

## Proof.

$$\begin{aligned}\|\mathcal{A}\mathcal{B}\| &= \sup_{\|x\|=1} \|\mathcal{A}\mathcal{B}x\| \\ &\leq \sup_{\|x\|=1} \|\mathcal{A}\| \|\mathcal{B}x\| \\ &\leq \sup_{\|x\|=1} \|\mathcal{A}\| \|\mathcal{B}\| \|x\| \\ &= \|\mathcal{A}\| \|\mathcal{B}\|\end{aligned}$$

# Properties of Linear Operators, cont

## Definition

An operator  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  is bounded if  $\|\mathcal{A}\| < \infty$

## Definition

The following three statements are equivalent

1.  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  is continuous
2.  $x_n \rightarrow x^* \Rightarrow \mathcal{A}[x_n] \rightarrow \mathcal{A}[x^*]$  for all convergent sequences in  $\mathbb{X}$
3.  $\forall \epsilon > 0, \exists \delta > 0$  such that

$$\|x - y\| \leq \delta \Rightarrow \|\mathcal{A}[x] - \mathcal{A}[y]\| < \epsilon \quad \forall x, y \in \mathbb{X}$$

## Properties of Linear Operators, cont

Theorem (Moon Theorem 4.1)

A linear operator is bounded iff it is continuous.

Proof.



( $\Rightarrow$ ) Suppose  $\|\mathcal{A}\| = M < \infty$ , let  $\{x_n\}$  be any convergent sequence with limit  $x^* \in \mathbb{X}$ , then

$$\begin{aligned}\|\mathcal{A}x_n - \mathcal{A}x^*\| &= \|\mathcal{A}(x_n - x^*)\| \leq \|\mathcal{A}\| \|x_n - x^*\| \\ &= M \|x_n - x^*\| \rightarrow 0 \Rightarrow \|\mathcal{A}x_n - \mathcal{A}x^*\| \rightarrow 0.\end{aligned}$$

Therefore  $\mathcal{A}$  is continuous.

## Proof, cont

( $\Leftarrow$ ) Assume  $\mathcal{A}$  is continuous and let  $\epsilon = 1$  and  $y = 0$  then  $\exists \delta$  such that  $\|x\| \leq \delta \Rightarrow \|\mathcal{A}x\| < 1$

Now let  $0 \neq x \in \mathbb{X}$  be arbitrary, then

$$\left\| \frac{\delta x}{\|x\|} \right\| = \frac{\delta}{\|x\|} \|x\| = \delta \leq \delta$$

implies that

$$\left\| \mathcal{A} \left( \frac{\delta x}{\|x\|} \right) \right\| = \frac{\delta}{\|x\|} \|\mathcal{A}x\| < 1$$

which implies that

$$\|\mathcal{A}x\| \leq \frac{1}{\delta} \|x\|$$

Therefore  $\mathcal{A}$  is bounded.

## Properties of Linear Operators, cont

### Theorem (Moon Theorem 4.2)

Let  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  be a linear operator. If  $\mathbb{X}$  is a finite dimensional Hilbert space, then  $\mathcal{A}$  is bounded.

Proof.

□

Let  $\dim(\mathbb{X}) = n$  and let  $\{p_1, \dots, p_n\}$  be an orthonormal basis for  $\mathbb{X}$ , then

$$x = \sum_{k=1}^n \langle x, p_k \rangle p_k$$

## Proof, cont.

Define  $D = \max\{\|\mathcal{A}p_1\|, \|\mathcal{A}p_2\|, \dots, \|\mathcal{A}p_n\|\}$  then

$$\begin{aligned}\|\mathcal{A}x\| &= \left\| \mathcal{A} \left( \sum_{k=1}^n \langle x, p_k \rangle p_k \right) \right\| \\ &\leq \sum_{k=1}^n |\langle x, p_k \rangle| \|\mathcal{A}p_k\| \\ &\leq D \sum_{k=1}^n |\langle x, p_k \rangle| \\ &\leq D \sum_{k=1}^n \|x\| \|p_k\| \quad (\text{Cauchy-Schwartz}) \\ &= Dn \|x\|\end{aligned}$$

Therefore  $\mathcal{A}$  is bounded.

## Section 2

### Neumann Expansion

## Geometric Series

One of the most important series in analysis is the geometric series

$$S = 1 + x + x^2 + \dots = \sum_{i=0}^{\infty} x^i$$

Noting that

$$\begin{aligned} 1 + xS &= 1 + x + x^2 + \dots = S \\ \Rightarrow S(1 - x) &= 1 \end{aligned}$$

Therefore

$$S = \sum_{i=0}^{\infty} x^i = \frac{1}{1-x} = (1-x)^{-1}$$

The series converges if  $|x| < 1$ .

# Geometric Series for Operators (Neumann Expansion)

For operators we have a similar expression:

**Theorem (Moon Theorem 4.3)**

*Suppose  $\|\cdot\|$  is a norm satisfying the submultiplicative property and  $\|\mathcal{A}\| < 1$ . Then  $(I - \mathcal{A})^{-1}$  exists and*

$$(I - \mathcal{A})^{-1} = \sum_{i=0}^{\infty} \mathcal{A}^i = I + \mathcal{A} + \mathcal{A}^2 + \mathcal{A}^3 + \dots$$

where

$$\mathcal{A}^2 = \mathcal{A}\mathcal{A}$$

$$\mathcal{A}^3 = \mathcal{A}\mathcal{A}^2$$

$$\mathcal{A}^k = \mathcal{A}\mathcal{A}^{k-1}.$$

## Neumann Expansion, Proof

Suppose that  $(I - \mathcal{A})^{-1}$  does not exist. Then  $\mathcal{N}(I - \mathcal{A})$  is non-trivial.

Therefore,  $\exists x \neq 0$  such that

$$\begin{aligned}(I - \mathcal{A})x = 0 &\iff x = \mathcal{A}x \\ &\iff \|x\| = \|\mathcal{A}x\| \leq \|\mathcal{A}\| \|x\| < \|x\|,\end{aligned}$$

which is a contradiction.

Therefore  $(I - \mathcal{A})^{-1}$  exists.

## Neumann Expansion, cont.

Note that  $\|\mathcal{A}^k\| \leq \|\mathcal{A}\|^k$  since  $\|\cdot\|$  satisfies the submultiplication property.

Since  $\|\mathcal{A}\| < 1$

$$\lim_{k \rightarrow \infty} \|\mathcal{A}^k\| = 0 \iff \lim_{k \rightarrow \infty} \mathcal{A}^k = 0$$

Note that

$$(I - \mathcal{A})(I + \mathcal{A} + \mathcal{A}^2 + \cdots + \mathcal{A}^{k-1}) = I - \mathcal{A}^k$$

$k \rightarrow \infty$  gives

$$(I - \mathcal{A}) \left( \sum_{i=0}^{\infty} \mathcal{A}^i \right) = I$$

Therefore

$$\sum_{i=0}^{\infty} \mathcal{A}^i = (I - \mathcal{A})^{-1}.$$

## Section 3

### Matrix Norms

# Matrix Norms

For matrices  $A : \mathbb{C}^m \rightarrow \mathbb{C}^n$  we have the following induced norm:

$$\|A\|_{\infty} = \max_{\|x\|_{\infty}=1} \|Ax\|_{\infty}$$

(Why max not sup?)

## Lemma

$$\|A\|_{\infty} = \max_{i=1:m} \sum_{j=1:n} |a_{ij}|$$

i.e., the largest row sum.

## Proof

First show that  $\|A\|_\infty \leq \max_{i=1:m} \sum_{j=1:n} |a_{ij}|$ :

$$\begin{aligned}\|A\|_\infty &= \max_{\|x\|_\infty=1} \left\| \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right\|_\infty \\ &= \max_{\|x\|_\infty=1} \left[ \max \begin{pmatrix} \left| \sum_{j=1}^n a_{1j} x_j \right| \\ \vdots \\ \left| \sum_{j=1}^n a_{mj} x_j \right| \end{pmatrix} \right] \\ &\leq \max_{\substack{x \text{ s.t.} \\ \max|x_i|=1}} \left[ \max \left( \sum_{j=1}^n |a_{1j}| |x_j|, \dots, \sum_{j=1}^n |a_{mj}| |x_j| \right) \right] \\ &\leq \max_{\|x\|_\infty=1} \left[ \max \left( \|x\|_\infty \sum_{j=1}^n |a_{1j}|, \dots, \|x\|_\infty \sum_{j=1}^m |a_{mj}| \right) \right] \\ &= \max_{i=1:m} \sum_{j=1}^m |a_{ij}|\end{aligned}$$

## Proof, cont.

Now we need to show that  $\max_{i=1:m} \sum_{j=1:n} |a_{ij}| \leq \|A\|_\infty$ :

Let  $k = \arg \max_{i=1:m} \sum_{j=1:n} |a_{ij}|$

and let  $\hat{x}$  be such that

$$\hat{x}_j = \begin{cases} 1 & \text{if } a_{kj} \geq 0 \\ -1 & \text{otherwise} \end{cases}$$

then  $\|\hat{x}\|_\infty = 1$  and then

$$\|A\hat{x}\|_\infty = \max_{i=1:m} \sum_{j=1:n} |a_{ij}| \leq \max_{\|x\|_\infty=1} \|Ax\|_\infty = \|A\|_\infty.$$

# Other Matrix Norms

## Lemma

$$\begin{aligned}\|A\|_1 &= \max_{\|x\|_1=1} \|Ax\|_1 \\ &= \max_{j=1:n} \sum_{i=1}^m |a_{ij}| \quad (\text{largest column sum})\end{aligned}$$

## Lemma

$$\|A\|_2 = \max_i \sqrt{\lambda_i(A^H A)} = \text{largest singular value of } A$$

More discussion of this in Chapter 7.

# Norm of $A^{-1}$

## Theorem

For induced matrix norms, where  $A^{-1}$  exists we have

$$\|A^{-1}\| = \frac{1}{\min_{\substack{x \neq 0 \\ \|x\|=1}} \frac{\|Ax\|}{\|x\|}} = \frac{1}{\min_{\|x\|=1} \|Ax\|}$$

## Proof.

Let  $Ax = b \Rightarrow x = A^{-1}b$  then

$$\begin{aligned}\|A^{-1}\| &= \max_{b \neq 0} \frac{\|A^{-1}b\|}{\|b\|} = \max_{x \neq 0} \frac{\|x\|}{\|Ax\|} = \max_{x \neq 0} \frac{1}{\frac{\|Ax\|}{\|x\|}} \\ &= \frac{1}{\min_{\substack{x \neq 0 \\ \|x\|=1}} \frac{\|Ax\|}{\|x\|}} = \frac{1}{\min_{\|x\|=1} \|Ax\|}\end{aligned}$$

# Frobenius Norm

## Definition

The Frobenius norm of a matrix is given by

$$\begin{aligned}\|A\|_F &= \left( \sum_{i=1}^m \sum_{j=1}^m |a_{ij}|^2 \right)^{\frac{1}{2}} \\ &= \sqrt{\text{tr}(A^H A)}\end{aligned}$$

**Fact:** The Frobenius norm is NOT an induced norm.

## Matrix Convergence

For matrices: convergence in any norm implies convergence in any other norm. In particular

$$\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2$$

$$\max |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max |a_{ij}|$$

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

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

## Section 4

### Adjoint Operators

# Adjoint Operator

## Definition

Let  $\mathcal{A} : \mathbb{X} \rightarrow \mathbb{Y}$  be a bounded linear operator from Hilbert space  $\mathbb{X}$  to Hilbert space  $\mathbb{Y}$ , then the adjoint of  $\mathcal{A}$  ( $\mathcal{A}^*$ ) is the linear operator  $\mathcal{A}^* : \mathbb{Y} \rightarrow \mathbb{X}$  such that

$$\langle \mathcal{A}x, y \rangle_{\mathbb{Y}} = \langle x, \mathcal{A}^*y \rangle_{\mathbb{X}}$$

$\forall x \in \mathbb{X}$  and  $\forall y \in \mathbb{Y}$ .

$\mathcal{A}$  is self-adjoint if  $\mathcal{A}^* = \mathcal{A}$

# Adjoint Operator, Example

Example (Complex matrices)

$$A : \mathbb{C}^n \rightarrow \mathbb{C}^m$$

What is  $A^*$ ?

By definition:

$$\begin{aligned}\langle Ax, y \rangle_{\mathbb{C}^m} &= \langle x, A^*y \rangle_{\mathbb{C}^n} \\ \iff y^H A x &= y^H (A^*)^H x \\ \iff A^* &= A^H\end{aligned}$$

Note  $A^H : \mathbb{C}^m \rightarrow \mathbb{C}^n$

# Adjoint Operator, Example

Example (Real matrices)

$$A : \mathbb{R}^n \rightarrow \mathbb{R}^m$$

What is  $A^*$ ?

By definition,

$$\begin{aligned}\langle Ax, y \rangle_{\mathbb{R}^m} &= \langle x, A^*y \rangle_{\mathbb{R}^n} \\ \iff x^\top A^\top y &= x^\top A^*y \\ \iff A^* &= A^\top\end{aligned}$$

# Adjoint Operator, Example

## Example (Convolution)

$$\mathcal{A} : L_2 \rightarrow L_2$$

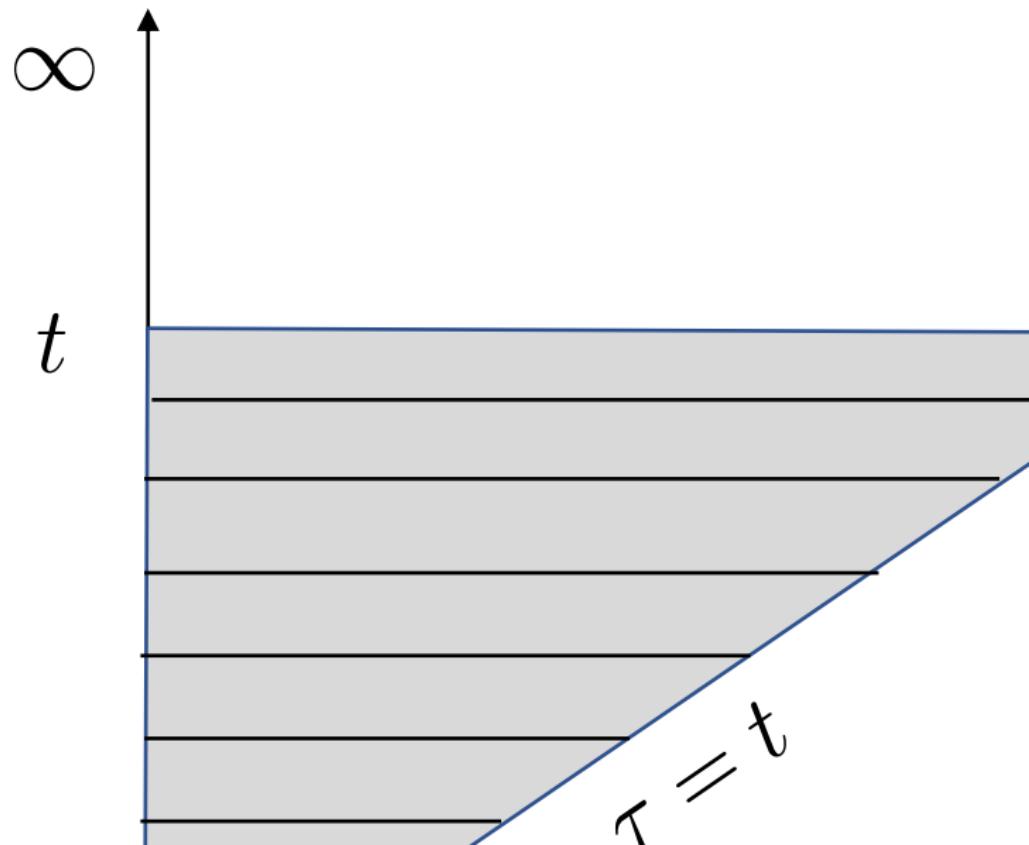
$$\mathcal{A}[x](t) = \int_0^{\top} h(t - \tau)x(\tau)d\tau$$

Let  $x \in L_2[0, \infty]$  and  $y \in L_2[0, \infty]$  then  $\mathcal{A}^*$  is defined by

$$\langle \mathcal{A}x, y \rangle_{L_2} = \langle x, \mathcal{A}^*y \rangle_{L_2}$$

$$\iff \int_{t=0}^{\infty} \left[ \int_{\tau=0}^t h(t - \tau)x(\tau)d\tau \right] y(t)dt = \int_0^{\infty} x(t)\mathcal{A}^*[y](t)dt$$

## Adjoint Operator, Example, Convolution, cont.



# Adjoint Operator, Example

Example (linear ode's)

$$\dot{x} = Fx \quad ; \quad x(0) = x_0$$

The solution is  $x(t) = e^{Ft}x_0$

Let  $\mathcal{A}[x_0](t) = e^{Ft}x_0$ , then

$$\mathcal{A} : \mathbb{R}^n \rightarrow L_{2[0,T]}$$

What is  $\mathcal{A}^*$ ?

## Adjoint Operator, Example, linear ODE, cont.

Let  $x \in \mathbb{R}^n$  and let  $y \in L_2[0, T]$  then by definition,

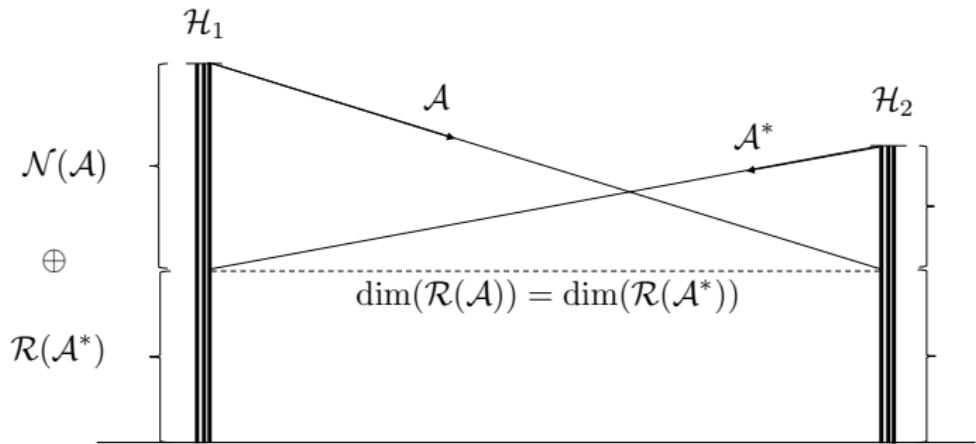
$$\begin{aligned}\langle \mathcal{A}[x_0], y \rangle_{L_2[0, T]} &= \langle x_0, \mathcal{A}^*y \rangle_{\mathbb{R}^n} \\ \iff \int_0^T x_0^\top (e^{Ft})^\top y(t) dt &= x_0^\top \mathcal{A}^*y \\ \iff x_0^\top \int_0^T e^{F^\top t} y(t) dt &= x_0^\top \mathcal{A}^*y \\ \Rightarrow \boxed{\mathcal{A}^*[y] = \int_0^T e^{F^\top t} y(t) dt}\end{aligned}$$

## Section 5

### Fundamental Subspaces

# Fundamental Subspaces

Let  $\mathcal{A} : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  where  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are Hilbert spaces.  
Then  $\mathcal{A}^* : \mathcal{H}_2 \rightarrow \mathcal{H}_1$  and we have the following picture:



# Fundamental Subspaces, cont.

## Lemma

1.  $\mathcal{H}_1 = \mathcal{N}(\mathcal{A}) \oplus \mathcal{R}(\mathcal{A}^*)$
2.  $\mathcal{H}_2 = \mathcal{N}(\mathcal{A}^*) \oplus \mathcal{R}(\mathcal{A})$
3.  $\dim(\mathcal{H}_1) = \dim(\mathcal{N}(\mathcal{A})) + \dim(\mathcal{R}(\mathcal{A}^*))$
4.  $\dim(\mathcal{H}_2) = \dim(\mathcal{N}(\mathcal{A}^*)) + \dim(\mathcal{R}(\mathcal{A}))$
5.  $\dim(\mathcal{R}(\mathcal{A})) = \dim(\mathcal{R}(\mathcal{A}^*))$

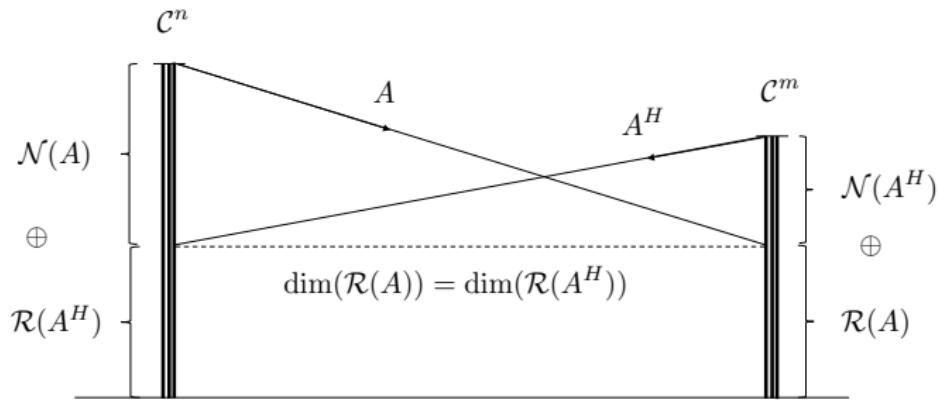
Proofs to follow.

# Fundamental Subspaces for Matrices

For matrices, the picture looks as follows:

$$A : \mathbb{C}^n \rightarrow \mathbb{C}^m$$

$$A^* = A^H : \mathbb{C}^m \rightarrow \mathbb{C}^n$$



$$\dim(\mathcal{R}(A^H)) = \dim(\mathcal{R}(A))$$

## Fundamental Subspaces, cont

### Theorem (Moon Theorem 4.5)

Let  $\mathcal{A} : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  be bounded and let  $\mathcal{H}_1$  and  $\mathcal{H}_2$  be Hilbert spaces and let  $\mathcal{R}(\mathcal{A})$  and  $\mathcal{R}(\mathcal{A}^*)$  be closed, then

1.  $[\mathcal{R}(\mathcal{A})]^\perp = \mathcal{N}(\mathcal{A}^*)$
2.  $[\mathcal{R}(\mathcal{A}^*)]^\perp = \mathcal{N}(\mathcal{A})$

## Theorem 4.5, Proof

(1): To show that  $[\mathcal{R}(\mathcal{A})]^\perp = \mathcal{N}(\mathcal{A}^*)$  we need to show that  $\mathcal{N}(\mathcal{A}^\perp) \subseteq [\mathcal{R}(\mathcal{A})]^\perp$  and  $[\mathcal{R}(\mathcal{A})]^\perp \subseteq \mathcal{N}(\mathcal{A}^*)$ .

We first show that  $\mathcal{N}(\mathcal{A}^*) \subseteq [\mathcal{R}(\mathcal{A})]^\perp$ :

Select any  $y \in \mathcal{N}(\mathcal{A}^*)$  and any  $\hat{y} \in \mathcal{R}(\mathcal{A})$ . Then  $\exists \hat{x} \in \mathcal{H}_1$  such that  $\hat{y} = \mathcal{A}\hat{x}$ . Therefore

$$\begin{aligned}\langle \hat{y}, y \rangle &= \langle \mathcal{A}\hat{x}, y \rangle \\ &= \langle \hat{x}, \mathcal{A}^*y \rangle \\ &= \langle \hat{x}, 0 \rangle = 0 \\ \Rightarrow \quad y &\in [\mathcal{R}(\mathcal{A})]^\perp \\ \Rightarrow \quad \mathcal{N}(\mathcal{A}^*) &\subseteq [\mathcal{R}(\mathcal{A})]^\perp\end{aligned}$$

## Theorem 4.5, Proof, cont.

We first show that  $[\mathcal{R}(\mathcal{A})]^\perp \subseteq \mathcal{N}(\mathcal{A}^*)$ :

Select any  $y \in [\mathcal{R}(\mathcal{A})]^\perp$ . For every  $\hat{x} \in \mathcal{H}_1$  we have  $\hat{y} = \mathcal{A}\hat{x} \in \mathcal{R}(\mathcal{A})$ , and therefore

$$\langle \hat{y}, y \rangle = \langle \mathcal{A}\hat{x}, y \rangle = 0$$

By definition of the adjoint, we therefore have that

$$\langle \hat{x}, \mathcal{A}^*y \rangle = 0$$

Since this is true for every  $\hat{x} \in \mathcal{H}_1$  it must be that  $\mathcal{A}^*y = 0$ .

Therefore

$$y \in \mathcal{N}(\mathcal{A}^*),$$

which implies that

$$[\mathcal{R}(\mathcal{A})]^\perp \subseteq \mathcal{N}(\mathcal{A}^*).$$

Item (2) is shown similarly.

## Fundamental Subspaces, cont

Theorem 2.10 states that if  $\mathcal{H}$  is a Hilbert space and if  $\mathbb{V}$  a closed subspace in  $\mathcal{H}$  then

$$\mathcal{H} = \mathbb{V} \oplus \mathbb{V}^\perp$$

Therefore Theorem 4.5 implies that

$$\mathcal{H}_1 = \mathcal{R}(\mathcal{A}^*) \oplus \mathcal{N}(\mathcal{A})$$

$$\mathcal{H}_2 = \mathcal{R}(\mathcal{A}) \oplus \mathcal{N}(\mathcal{A}^*)$$

Which also implies that

$$\dim(\mathcal{H}_1) = \dim(\mathcal{R}(\mathcal{A}^*)) + \dim(\mathcal{N}(\mathcal{A}))$$

$$\dim(\mathcal{H}_2) = \dim(\mathcal{R}(\mathcal{A})) + \dim(\mathcal{N}(\mathcal{A}^*))$$

# Fundamental Subspaces, cont

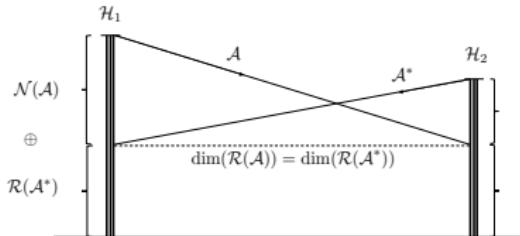
## Lemma

- ▶  $\mathcal{R}(A) = \mathcal{R}(AA^*)$
- ▶  $\mathcal{R}(A^*) = \mathcal{R}(A^*A)$

## Proof.

We will prove (1) by showing that:

- (a)  $\mathcal{R}(A) \subseteq \mathcal{R}(AA^*)$
- (b)  $\mathcal{R}(AA^*) \subseteq \mathcal{R}(A)$



## Fundamental Subspaces, cont

### Proof (cont.)

(a) Let  $y \in \mathcal{R}(\mathcal{A}) \Rightarrow \exists x \in \mathcal{H}_1$  such that  $y = \mathcal{A}x$   
Since  $\mathcal{H}_1 = \mathcal{R}(\mathcal{A}^*) \oplus \mathcal{N}(\mathcal{A})$ ,  $x = x_n + x_r$  where

$$x_n \in \mathcal{N}(\mathcal{A}) \text{ and } x_r \in \mathcal{R}(\mathcal{A}^*)$$

$$\Rightarrow \exists \hat{y} \in \mathcal{H}_2 \text{ such that } x_r = \mathcal{A}^* \hat{y}$$

so

$$y = \mathcal{A}x = \mathcal{A}(x_n + x_r) = \mathcal{A}\mathcal{A}^* \hat{y}$$

$$\Rightarrow y \in \mathcal{R}(\mathcal{A}\mathcal{A}^*)$$

(b) let  $y \in \mathcal{R}(\mathcal{A}\mathcal{A}^*) \Rightarrow \exists \hat{y} \in \mathcal{H}_2$  such that

$$y = \mathcal{A}\mathcal{A}^* \hat{y} \Rightarrow y = \mathcal{A}\hat{x} \text{ where } \hat{x} \in \mathcal{H}_1$$

$$\Rightarrow y \in \mathcal{R}(\mathcal{A}).$$

# Fundamental Subspaces, cont

Theorem

$$\dim(\mathcal{R}(\mathcal{A})) = \dim(\mathcal{R}(\mathcal{A}^*))$$

Proof.

We need to show that

- (a)  $\dim(\mathcal{R}(\mathcal{A})) \leq \dim(\mathcal{R}(\mathcal{A}^*))$
- (b)  $\dim(\mathcal{R}(\mathcal{A}^*)) \leq \dim(\mathcal{R}(\mathcal{A}))$

## Fundamental Subspaces, cont

### Proof (cont.)

(a) Let  $P = \{p_1, p_2, \dots\}$  be a Hamel basis for  $\mathcal{R}(\mathcal{A})$  so  $\dim(\mathcal{R}(\mathcal{A})) = \text{cardinality of } P$ .

$$p_i \in \mathcal{R}(\mathcal{A}) \Rightarrow \exists \hat{q}_i \in \mathcal{H}_1 \text{ such that } p_i = \mathcal{A}\hat{q}_i$$

$$\mathcal{H}_1 = \mathcal{R}(\mathcal{A}^*) \oplus \mathcal{N}(\mathcal{A}) \Rightarrow \hat{q}_i = q_{i,n} + q_i$$

where  $q_{i,n} \in \mathcal{N}(\mathcal{A})$  and  $q_i \in \mathcal{R}(\mathcal{A}^*)$

$$\Rightarrow p_i = \mathcal{A}q_i,$$

let

$$Q = \{q_1, q_2, \dots\}$$

we will show that  $Q$  is linearly independent  $\Rightarrow$  any Hamel basis of  $\mathcal{R}(A^*)$  contains  $Q \Rightarrow \dim(\mathcal{R}(A^*)) \geq \dim(\mathcal{R}(A))$ ,

## Fundamental Subspaces, cont

### Proof (cont.)

$P$  is a Hamel basis  $\Rightarrow$  all finite subsets of  $P$  are linearly independent, i.e.

$$\sum_{i \in I} c_i p_i = 0 \iff c_i = 0, i \in I$$

where  $I$  is a finite index set. But,

$$\sum_I c_i p_i = 0 \iff \sum_I c_i \mathcal{A} q_i = 0 \iff \mathcal{A}(\sum_I c_i q_i) = 0$$

but  $\sum_I c_i q_i \in \mathcal{R}(\mathcal{A}^*) \perp \mathcal{N}(\mathcal{A})$

so

$$\iff \sum_I c_i q_i = 0 \iff c_i = 0, i \in I$$

$\Rightarrow Q$  is linearly independent

(b) Substitute  $\mathcal{A}$  for  $\mathcal{A}^*$  and  $\mathcal{A}^*$  for  $\mathcal{A}$  is above argument.

# Solution of Operator Equations

We turn to solutions to the linear operator equation

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

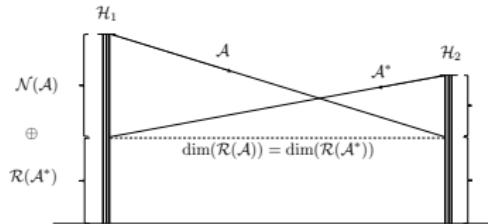
where  $\mathcal{A} : \mathcal{H}_1 \rightarrow \mathcal{H}_2$  is bounded,  $\mathcal{H}_1$  and  $\mathcal{H}_2$  are Hilbert and  $\mathcal{R}(\mathcal{A})$  is closed.

**Fact 1.**  $\mathcal{A}x = y$  has a solution

$$\iff y \in \mathcal{R}(\mathcal{A})$$

**Fact 2.**  $\mathcal{A}x = y$  has a solution

$$\iff y \perp \mathcal{N}(\mathcal{A}^*)$$

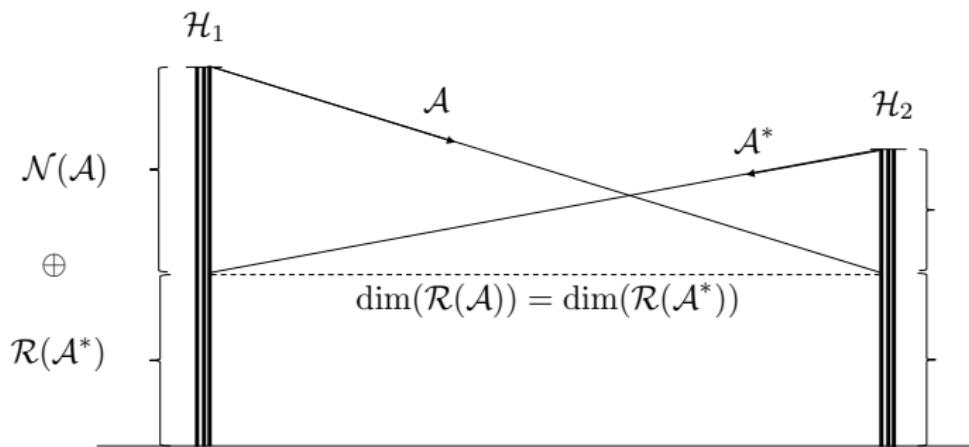


# Solution of Operator Equations

Fact 3. If  $\mathcal{A}x = y$  has a solution then it is unique  
 $\iff \mathcal{N}(\mathcal{A}) = \{0\}$

Fact 4. If  $\mathcal{N}(\mathcal{A}) \neq \{0\}$  and  $y \in \mathcal{R}(\mathcal{A})$  then  $\mathcal{A}x = y$  has an infinite number of solutions.

Fact 5 .  $\mathcal{A}^{-1}$  exists  $\Rightarrow \mathcal{N}(\mathcal{A}) = \{0\}$  (otherwise can't get back to all of  $\mathcal{H}$ ).



# Matrix Rank

## Definition (Row Rank)

The row rank of  $A \in \mathbb{C}^{m \times n}$  is the number of linearly independent rows.

## Definition (Column Rank)

The column rank of  $A \in \mathbb{C}^{m \times n}$  is the number of linearly independent columns.

- ▶ Since  $\mathcal{R}(A) = \text{span}\{\text{columns of } A\}$  we have that  
 $\dim(\mathcal{R}(A)) = \text{column rank}$
- ▶ Since  $\mathcal{R}(A^H) = \text{span}\{\text{rows of } A\}$  we have that  
 $\dim(\mathcal{R}(A^*)) = \text{row rank}$
- ▶ Therefore  $\dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^H))$  implies that  
column rank = row rank

# Matrix Rank

## Definition

The rank of  $A$  is the number of linearly independent rows or columns.

## Lemma

$$\text{rank}(A) = \text{rank}(A^H)$$

## Definition

$A : \mathbb{C}^n \rightarrow \mathbb{C}^m$  is full rank if  $\text{rank}(A) = \min(n, m)$

# Sylvester's Inequality

Lemma (Sylvester's Inequality)

Let  $A \in \mathbb{C}^{q \times n}$  and  $B \in \mathbb{C}^{n \times p}$  then

$$\text{rank}(A) + \text{rank}(B) - n \leq \text{rank}(AB) \leq \min(\text{rank}(A), \text{rank}(B)).$$

Example

Let  $x \in \mathbb{R}^m$  and  $y \in \mathbb{R}^n$  then

$$\text{rank}(xy^\top) = 1$$

## Section 6

### Matrix Inverses

# Matrix Inverses

## Definition

$A \in \mathbb{C}^{m \times n}$  has a left inverse if  $\exists B \in \mathbb{C}^{n \times m}$  such that

$$\begin{matrix} B \\ n \times m \end{matrix} \quad \begin{matrix} A \\ m \times n \end{matrix} = \begin{matrix} I \\ n \times n \end{matrix}$$

## Definition

$A \in \mathbb{C}^{m \times n}$  has a right inverse if  $\exists D \in \mathbb{C}^{n \times m}$  such that

$$\begin{matrix} A \\ m \times n \end{matrix} \quad \begin{matrix} C \\ n \times m \end{matrix} = \begin{matrix} I \\ m \times m \end{matrix}$$

## Matrix Inverses, cont

### Example

The matrix

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 7 & 0 \end{pmatrix}.$$

has an infinite number of right inverses, namely

$$C = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{7} \\ c_1 & c_2 \end{pmatrix} \quad \forall c_1, c_2 \in \mathbb{R}$$

since

$$AC = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

## Matrix Inverses, cont

- ▶ Suppose  $A$  has a left inverse, then

$$Ax = b \iff BAx = Bb \iff x = Bb$$

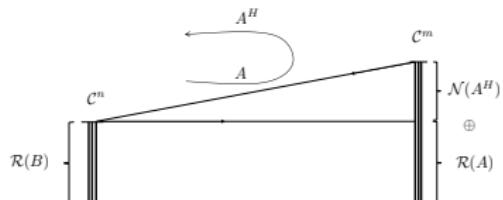
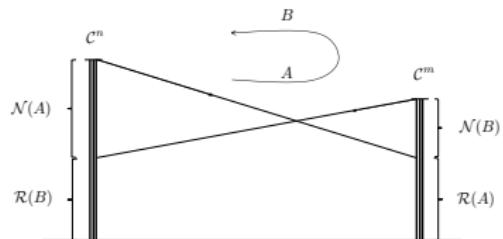
- ▶ Suppose  $A$  has a right inverse, then let

$$x = Cb \Rightarrow Ax = ACb = b$$

so  $x = Cb$  is a solution.

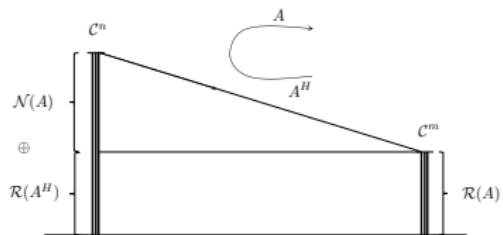
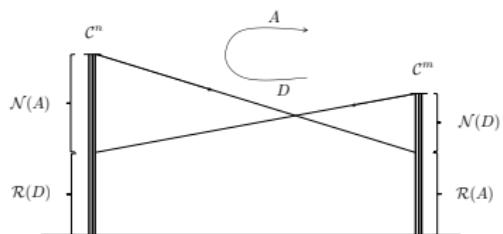
# Left Inverse

- ▶ Let  $B$  be a left inverse of  $A$ .
- ▶ Then  $BA = I : \mathbb{C}^n \rightarrow \mathbb{C}^n$ .
- ▶ Of necessity we must have that  $\mathcal{N}(A) = \{0\}$ , otherwise there are vectors  $x \in \mathcal{N}(A) \subseteq \mathbb{C}^n$  such that  $BAx = B0 = 0 \neq x$ , i.e.,  $BA \neq I$ .
- ▶ Therefore  $Ax = b$  has at most one solution  
(since  $b$  may not be in  $\mathcal{R}(A)$ ).



# Right Inverse

- ▶ Let  $D$  be a right inverse of  $A$ .
- ▶ Then  $AD = I : \mathbb{C}^m \rightarrow \mathbb{C}^m$ .
- ▶ Of necessity we must have that  $\mathcal{N}(A^H) = \{0\}$ , otherwise  $D^H A^H = I$  is impossible.
- ▶  $\mathcal{N}(A)$  may be nontrivial therefore if  $\hat{x}$  is a solution so is  $\hat{x} + x_n$  where  $x_n \in \mathcal{N}(A)$  since  $A(\hat{x} + x_n) = A\hat{x} = b$ . Therefore, there is at least one solution.



# Right and Left Inverses

## Lemma

1. If  $A$  has a left inverse then  $Ax = b$  has at most one solution.
2. If  $A$  has a right inverse then  $Ax = b$  has at least one solution.

## Regular Inverse

If  $A \in \mathbb{C}^{n \times n}$  when the following statements are equivalent:

1.  $A^{-1}$  exists
2.  $\mathcal{N}(A) = \{0\}$  and  $\mathcal{N}(A^H) = \{0\}$ .
3.  $\text{rank}(A) = n$
4.  $\det(A) \neq 0$
5. (right inverse of  $A$ ) = (left inverse of  $A$ ) =  $A^{-1}$
6. there are no zero eigenvalues of  $A$
7.  $A^H A$  is positive definite
8.  $A$  is nonsingular

## Regular Inverse, cont.

If  $A^{-1}$  exists then

$$A^{-1} = \frac{\text{adj}(A)}{\det(A)}$$

where  $\text{adj}(A)$  is the adjugate of  $A$  where  $\text{adj}(A) = [B_{ij}]^\top$  and  $B_{ij} = (-1)^{i+j} \det(M_{ij})$  and  $M_{ij}$  is the  $(i,j)^{\text{th}}$  minor of  $A$ .

### Example

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

$$\text{adj}(A) = \begin{pmatrix} (-1)^2|d| & (-1)^3|c| \\ (-1)^3|b| & (-1)^4|a| \end{pmatrix} = \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}$$

$$\text{so } A^{-1} = \frac{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}}{\det(A)} = \frac{\begin{pmatrix} d & -c \\ -b & a \end{pmatrix}}{ad - cb}$$

# Matrix Rank

## Lemma

Let  $A : \mathbb{C}^n \rightarrow \mathbb{C}^m$  then

$$\text{rank}\left(\begin{matrix} A \\ m \times n \end{matrix}\right) = \text{rank}\left(\begin{matrix} A^H \\ n \times m \end{matrix}\right) = \text{rank}\left(\begin{matrix} A^H A \\ n \times n \end{matrix}\right) = \text{rank}\left(\begin{matrix} AA^H \\ m \times m \end{matrix}\right)$$

Proof.

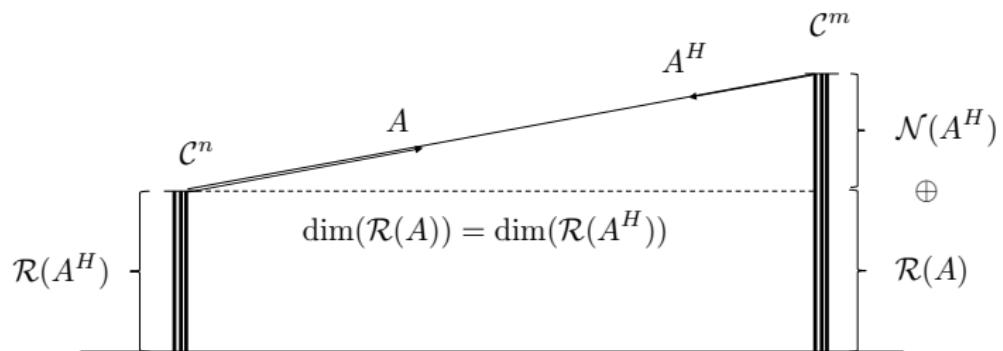
$$\begin{aligned}\text{rank}(B) &= \# \text{ of linearly independent columns} = \dim(\mathcal{R}(B)) \\ &= \# \text{ of linearly independent rows} = \dim(\mathcal{R}(B^H)).\end{aligned}$$

Therefore

$$\begin{aligned}\text{rank}(A) &= \dim(\mathcal{R}(A)) = \dim(\mathcal{R}(A^H)) = \text{rank}(A^H) \\ &= \dim(\mathcal{R}(AA^H)) = \text{rank}(AA^H) \text{ Since } \mathcal{R}(A^*) = \mathcal{R}(AA^*) \\ &= \dim(\mathcal{R}(A^H A)) = \text{rank}(A^H A) \text{ Since } \mathcal{R}(A) = \mathcal{R}(A^*A)\end{aligned}$$

## Left Inverse: Least Squares

- ▶ Consider the solution of  $Ax = b$  where  $m > n$ , i.e.,  $A$  is tall.
- ▶ Assume  $A$  is full rank, i.e.,  $\text{rank}(A) = n$ .
- ▶ Assume  $b \in \mathcal{R}(A)$



- ▶ Map  $b$  to  $\mathcal{R}(A^*) : A^H b = A^H A x$
- ▶ Since  $\text{rank}(A) = n \iff \text{rank}(A^H A) = n$  so  $(A^H A)^{-1}$  exists

$$\Rightarrow x = (A^H A)^{-1} A^H b$$

## Left Inverse: Least Squares, cont.

What if  $b \notin \mathcal{R}(A)$ ? This is the least squares problem, e.g.

$$\underbrace{\begin{pmatrix} a_1 & 1 \\ a_2 & 1 \\ \vdots & \vdots \\ a_n & 1 \end{pmatrix}}_A \underbrace{\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}}_x = \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}}_b$$

linear regression

Since there is no solution, it is reasonable to find  $x$  that minimizes  $\|e\|_2$  where

$$e = Ax - b$$

## Left Inverse: Least Squares, cont.

- ▶ Note that  $b = b_r + b_n$  where  $b_r \in \mathcal{R}(A)$  and  $b_n \in \mathcal{N}(A^H)$  so  $e = Ax - b_r - b_n$ .
- ▶ Since  $Ax - b_r \in \mathcal{R}(A) \perp \mathcal{N}(A^H)$  the best we can do is make  $Ax = b_r \Rightarrow e = b_n$ .
- ▶ Since  $b_n \in \mathcal{N}(A^H)$  we have

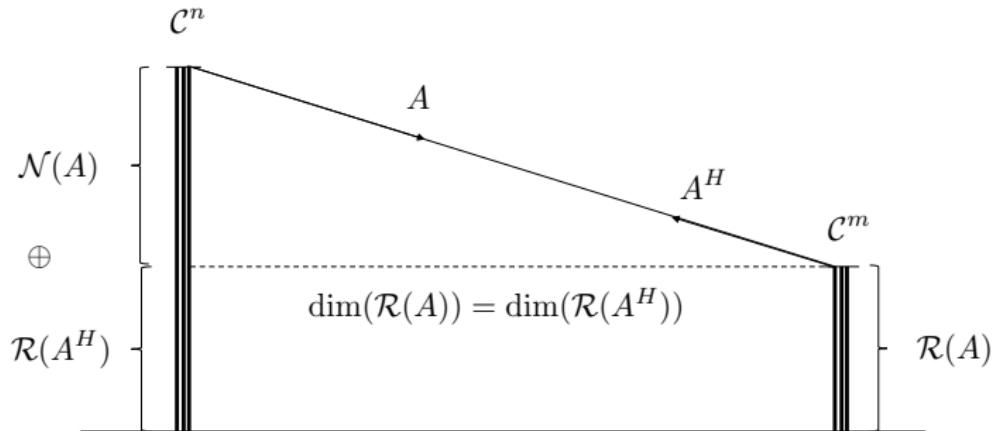
$$0 = A^H A x - A^H b_r$$
$$\Rightarrow \underbrace{A^H A x}_{\text{projection of } x \text{ onto } \mathcal{R}(A^H)} = A^H b_r = \underbrace{A^H b}_{\text{projection of } b \text{ onto } \mathcal{R}(A^H)}$$

- ▶ Since  $\text{rank}(A^H A) = \text{rank}(A) = n$  we have

$$\underbrace{x = (A^H A)^{-1} A^H b}_{\text{least square solution}}$$

## Right Inverse: Min-Norm Solution

- ▶ Consider the solution of  $Ax = b$  where  $m < n$ , i.e.,  $A$  is fat.
- ▶ Assume  $A$  is full rank, i.e.,  $\text{rank}(A) = m$ .



We would like to solve  $Ax = b$  note that since  $x = x_r + x_n$  where  $x_r \in \mathcal{R}(A^H)$  and  $x_n \in \mathcal{N}(A)$  and  $\mathcal{N}(A) \neq \{0\}$  there are an infinite number of solutions (i.e. add any thing in  $\mathcal{N}(A)$  to a solution). The minimum norm solution will be the element of  $\mathcal{R}(A^H)$  that satisfies  $Ax_r = b$ .

## Right Inverse: Min-norm Solution, cont.

$$x_r \in \mathcal{R}(A^H) \Rightarrow x_r = A^H y \text{ where } y \in \mathbb{C}^m$$

so we need to solve

$$\left( \begin{array}{cc} A & A^H \\ m \times n & n \times m \end{array} \right)_{m \times 1} y = \begin{array}{c} b \\ m \times 1 \end{array}$$

Since  $\text{rank}(A) = \text{rank}(AA^H) = m$ ,  $(AA^H)^{-1}$  exists.

$$\Rightarrow y = (AA^H)^{-1}b$$

$$\Rightarrow \boxed{x_r = A^H(AA^H)^{-1}b}$$

Note that this is the same solution as

$$\min \|x\|_2$$

$$\text{s.t. } Ax = b$$

# Right and Left Inverses

## Lemma

If  $A \in \mathbb{C}^{m \times n}$  where  $m > n$  and  $A$  is full rank, then  $(A^H A)^{-1} A^H$  is a left inverse of  $A$ .

## Proof.

$$(A^H A)^{-1} A^H A = I_n$$



## Lemma

If  $A \in \mathbb{C}^{m \times n}$  where  $m < n$  and  $A$  is full rank, then  $A^H (A A^H)^{-1} b$  is a right inverse of  $A$ .

## Proof.

$$A A^H (A A^H)^{-1} = I_m$$



- ▶ Both are examples of pseudo-inverses.
- ▶  $A^H (A A^H)^{-1}$  is called the Moore-Penrose pseudo-inverse.
- ▶ In Matlab type `pinv(A)`.

## Section 7

### Matrix Condition Number

# Matrix Condition Number

- ▶ Suppose that  $A \in \mathbb{C}^{n \times n}$  is full rank and  $A^{-1}$  is to be computed numerically. How reliable is the computation?
- ▶  $Ax = b$  can be written as

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

- ▶ Therefore, the solution  $x$  is the intersection of  $n$ -hyperplanes:

$$a_{11}x_1 + \dots + a_{1n}x_n = b_1$$

⋮

$$a_{n1}x_1 + \dots + a_{nn}x_n = b_n$$

## Matrix Condition Number, cont.

- ▶ The problem comes when these hyperplanes are almost parallel.
- ▶ In two dimensions we have two lines

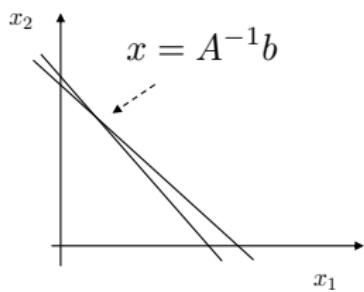
$$a_{11}x_1 + a_{12}x_2 = b_1$$

$$a_{21}x_1 + a_{22}x_2 = b_2$$

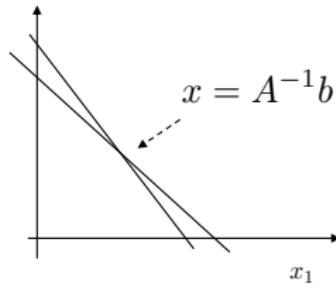
which can be rewritten as

$$x_2 = -\frac{a_{11}}{a_{12}}x_1 + \frac{b_1}{a_{12}}$$
$$x_2 = \underbrace{-\frac{a_{21}}{a_{22}}}_{\text{slope}} x_1 + \underbrace{\frac{b_2}{a_{22}}}_{\text{x-intercept}}$$

## Matrix Condition Number, cont.



Small change in  
 $y$ -intercept of  
second line has  
large impact on  
solution.



If the two lines are almost parallel then small changes in the slope or  $x_2$ -intercept of either line will result in large changes in  $x = A^{-1}b$ .

## Matrix Condition Number, cont.

- ▶ Since computers must represent numbers to finite precision, representation errors could significantly change the numerical solution to the equation  $Ax = b$ .
- ▶ The condition number quantifies this effect.

### Definition

The condition number of a square matrix is defined to be

$$\mathcal{K}(A) = \|A\| \|A^{-1}\|$$

where  $\|\cdot\|$  is an induced matrix norm usually taken to be the induced 2-norm.

## Matrix Condition Number: Derivation

- ▶ Given the two equations  $Ax = b$  and  $(A + \epsilon E)x = b$  where  $\epsilon E$  is a “small” perturbation of  $A$  (introduced by finite machine precision of  $A$ )
- ▶ Let  $x_0 = A^{-1}b$  and

$$\begin{aligned}x_E &= (A + \epsilon E)^{-1}b \\&= [A(I + \epsilon A^{-1}E)]^{-1}b \\&= (I + \epsilon A^{-1}E)^{-1}A^{-1}b \\&= \underbrace{(I + \epsilon A^{-1}E)^{-1}}_{\text{perturbation}} x_0\end{aligned}$$

## Matrix Condition Number: Derivation, cont.

Using the Neumann expansion gives

$$(I + \epsilon A^{-1}E)^{-1} = \sum_{i=0}^{\infty} (-\epsilon A^{-1}E)^i. \text{ Therefore}$$

$$\begin{aligned}x_E &= (I + \epsilon A^{-1}E)^{-1}A^{-1}b \\&= (I - \epsilon A^{-1}E)A^{-1}b + O(\|\epsilon E\|^2 x_0) \\&= A^{-1}b - \epsilon A^{-1}EA^{-1}b + O(\|\epsilon E\|^2 x_0) \\&= x_0 - \epsilon A^{-1}Ex_0 + O(\|\epsilon E\|^2 x_0)\end{aligned}$$

Therefore

$$\underbrace{\frac{\|x_E - x_0\|}{\|x_0\|}}_{\text{relative change in the solution}} \leq \underbrace{\epsilon \|A^{-1}\| \|E\|}_{\text{want to relate to relative change in } A} + O(\|\epsilon E\|^2)$$

## Matrix Condition Number: Derivation, cont.

What is the relative change in  $A$ ?

$$\frac{\|A - (A + \epsilon E)\|}{\|A\|} = \frac{\epsilon \|E\|}{\|A\|} \triangleq \rho$$

Therefore

$$\frac{\|x_E - x_0\|}{\|x_0\|} \leq \rho \underbrace{\|A^{-1}\| \|A\|}_{\mathcal{K}(A)} + O(\|\epsilon E\|^2)$$

The condition number  $\mathcal{K}(A)$  relates (approximately) the relative change in  $A$  to the relative change in the solution  $x_0$ .

## Matrix Condition Number: Implication

### Rule of Thumb:

If the solution is computed to  $n$  digits then only

$$n - \log_{10} \mathcal{K}(A)$$

can be considered to be accurate.

## Section 8

# Schur Complement and the Matrix Inversion Lemma

# Schur Complement

## Definition

Consider the partitioned matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

- When  $A_{11}$  is non-singular,

$$S_{ch}(A_{11}) \stackrel{\triangle}{=} A_{22} - A_{21}A_{11}^{-1}A_{12}$$

is called the Schur Complement of  $A_{11}$  in  $A$ .

- When  $A_{22}$  is non-singular,

$$S_{ch}(A_{22}) \stackrel{\triangle}{=} A_{11} - A_{12}A_{22}^{-1}A_{21}$$

is called the Schur Complement of  $A_{22}$  in  $A$ .

## Schur Complement, cont.

### Lemma

When  $A_{11}$  is nonsingular,  $A$  is nonsingular if and only if  $S_{ch}(A_{11})$  is nonsingular, in which case

$$A^{-1} = \begin{bmatrix} A_{11}^{-1} + A_{11}^{-1}A_{12}S_{ch}^{-1}(A_{11})A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}S_{ch}^{-1}(A_{11}) \\ -S_{ch}^{-1}(A_{11})A_{21}A_{11}^{-1} & S_{ch}^{-1}(A_{11}) \end{bmatrix}$$

### Lemma

When  $A_{22}$  is nonsingular,  $A$  is nonsingular if and only if  $S_{ch}(A_{22})$  is nonsingular, in which case

$$A^{-1} = \begin{bmatrix} S_{ch}^{-1}(A_{22}) & -S_{ch}^{-1}(A_{22})A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{12}S_{ch}^{-1}(A_{22}) & A_{22}^{-1} + A_{22}^{-1}A_{21}S_{ch}^{-1}(A_{22})A_{12}A_{22}^{-1} \end{bmatrix}$$

### Proof.

By direct manipulation.

# Matrix Inversion Lemma

Lemma (Matrix Inversion Lemma)

If  $A \in \mathbb{R}^{n \times n}$  and  $R \in \mathbb{R}^{m \times m}$  are invertible, and  $X \in \mathbb{R}^{n \times m}$  and  $Y \in \mathbb{R}^{m \times n}$  then

$$(A + XRY)^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$$

Proof.

Equate the (2, 2) elements of  $A^{-1}$  in the previous slide, and re-label matrices.



## Matrix Inversion Lemma, cont.

- ▶ A special case of this matrix inversion lemma is the formula

$$(A + xy^H)^{-1} = A^{-1} - \frac{A^{-1}xy^HA^{-1}}{1 + y^HA^{-1}x}$$

where  $x$  and  $y$  are vectors.

- ▶ Sylvester's inequality gives

$$\text{rank}(x) + \text{rank}(y) - 1 \leq \text{rank}(xy^H) \leq \min(\text{rank}(x), \text{rank}(y)).$$

But

$$\text{rank}(x) + \text{rank}(y) - 1 = 1$$

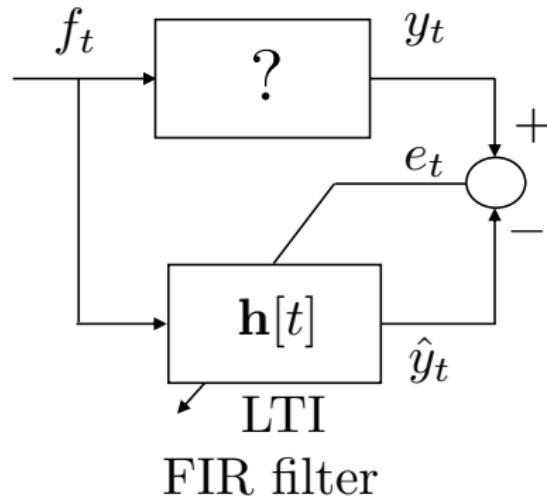
$$\min(\text{rank}(x), \text{rank}(y)) = 1$$

- ▶ Therefore  $\text{rank}(xy^H) = 1$

## Section 9

### Recursive Least Squares Filtering

# Least Squares Filtering Problem



**Problem Statement:** Given the input data  $f_t$  and  $y_t$ , find the FIR filter coefficients  $\mathbf{h}[t]$  that minimize the running least squared error  $e_t$ .

# Least Squares Filtering Problem

## Definition (Least Squares Filtering Problem)

Given the filter

$$\hat{y}_t = \sum_{i=1}^m h_i f_{t-i}$$

where the inputs  $f_t$  are known and we measure the actual outputs  $y_t$ , find the coefficients  $h_i$  such that the mean squared error

$$E = \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

is minimized.

# Batch Least Squares Filtering

If we assume  $f_t = 0, t \leq 0$  we get

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} f_1 & 0 & \cdots & \cdots & 0 \\ f_2 & f_1 & 0 & \cdots & 0 \\ \vdots & & & \ddots & \\ f_m & f_{m-1} & \cdots & \cdots & f_1 \\ f_{m+1} & f_m & f_{m-1} & \cdots & f_2 \\ \vdots & & & \ddots & \\ f_N & f_{N-1} & \cdots & \cdots & f_{N-m+1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix}$$

## Batch Least Squares Filtering, cont.

Define

$$\begin{aligned}\mathbf{q}_i &= (f_i \quad f_{i-1} \quad \dots \quad f_{i-m+1})^H \\ \mathbf{y}_N &= (\bar{y}_1 \quad \bar{y}_2 \quad \dots \quad \bar{y}_N)^H \\ \mathbf{h}[N] &= (\bar{h}_1[N] \quad \bar{h}_2[N] \quad \dots \quad \bar{h}_m[N])^H \\ A_N &= \begin{pmatrix} \mathbf{q}_1^H \\ \vdots \\ \mathbf{q}_m^H \end{pmatrix},\end{aligned}$$

then the least squares problem reduces to

$$\mathbf{e}_N = \mathbf{y}_N - \underbrace{A_N \mathbf{h}[N]}_{\hat{\mathbf{y}}_N}$$

where  $\mathbf{e}_N$  is the error to be minimized. From the projection theorem,  $\|\mathbf{e}\|_2$  is minimized when

$$\mathbf{h}[N] = \left( \begin{matrix} A_N^H \\ \vdots \\ A_N^H \end{matrix} \right)_{m \times 1}^{-1} \left( \begin{matrix} \mathbf{y}_N \\ \vdots \\ \mathbf{y}_N \end{matrix} \right)_{m \times NN \times 1}.$$

## Batch Least Squares Filtering

- ▶ Note that the size of  $y_N$  and  $A_N$  grow linearly with time  $N$ .
- ▶ Therefore, each time step requires more computation than the last step. This is obviously problematic as  $N \rightarrow \infty$ .
- ▶ For some  $N$ , batch least squares is no longer a real-time algorithm.
- ▶ Note that at time  $N + 1$  the data include new samples, but includes all of the data available at time  $N$ .

??? Is it possible to design an algorithm with fixed computational cost at each time step, that produces the same least squares solution?

# Recursive Least Squares Filtering

Define

$$\begin{aligned}\mathbf{q}_t &= (f_i \quad f_{i-1} \quad \dots \quad f_{i-m+1})^H \\ \mathbf{y}_t &= (\bar{y}_1 \quad \bar{y}_2 \quad \dots \quad \bar{y}_t)^H \\ \mathbf{h}[t] &= (\bar{h}_1[t] \quad \bar{h}_2[t] \quad \dots \quad \bar{h}_m[t])^H \\ A_t &= \begin{pmatrix} \mathbf{q}_1^H \\ \vdots \\ \mathbf{q}_t^H \end{pmatrix}.\end{aligned}$$

Then at time  $t$  we have  $\mathbf{e}_t = \mathbf{y}_t - A_t \mathbf{h}[t]$ . From the projection theorem, the error is minimized when

$$\mathbf{h}[t] = (A_t^H A_t)^{-1} A_t^H \mathbf{y}_t.$$

## Recursive Least Squares Filtering, cont.

Let

$$R_{t-1} \stackrel{\triangle}{=} A_{t-1}^H A_{t-1} = (\mathbf{q}_1 \quad \cdots \quad \mathbf{q}_{t-1}) \begin{pmatrix} \mathbf{q}_1^H \\ \vdots \\ \mathbf{q}_{t-1}^H \end{pmatrix}$$
$$= \sum_{i=1}^{t-1} \mathbf{q}_i \mathbf{q}_i^H$$

be the associated Grammian when there are  $t - 1$  samples.

Suppose that we receive new data  $q_t$  and  $y_t$  at time  $t$ .

Then

$$R_t = \sum_{i=1}^t \mathbf{q}_i \mathbf{q}_i^H$$
$$= \sum_{i=1}^{t-1} \mathbf{q}_i \mathbf{q}_i^H + \mathbf{q}_t \mathbf{q}_t^H$$
$$= R_{t-1} + \mathbf{q}_t \mathbf{q}_t^H.$$

## Recursive Least Squares Filtering, cont.

In the solution  $\mathbf{h}_t = (A_t^H A_t)^{-1} A_t^H \mathbf{y}_t$ , we need  $R_t^{-1} \triangleq (A_t^H A_t)^{-1}$ . Note that

$$R_t^{-1} = (\underbrace{R_{t-1}}_A + \underbrace{q_t}_{X} \underbrace{R=1}_{Y} \underbrace{q_t^H}_{Y})^{-1}$$

and recall the matrix inversion lemma:

$$(A + XRY)^{-1} = A^{-1} - A^{-1}X(R^{-1} + YA^{-1}X)^{-1}YA^{-1}$$

Therefore

$$R_t^{-1} = R_{t-1}^{-1} - R_{t-1}^{-1} \mathbf{q}_t (1 + \mathbf{q}_t^H R_{t-1}^{-1} \mathbf{q}_t)^{-1} \mathbf{q}_t^H R_{t-1}^{-1}.$$

## Recursive Least Squares Filtering, cont.

Defining  $P_t = R_t^{-1}$  gives

$$P_t = P_{t-1} - \frac{P_{t-1}\mathbf{q}_t\mathbf{q}_t^H P_{t-1}}{1 + \mathbf{q}_t^H P_{t-1} \mathbf{q}_t}.$$

Define the (Kalman) gain as

$$\mathbf{k}_t = \frac{P_{t-1}\mathbf{q}_t}{1 + \mathbf{q}_t^H P_{t-1} \mathbf{q}_t}$$

Then

$$P_t = P_{t-1} - \mathbf{k}_t \mathbf{q}_t^H P_{t-1}.$$

Note that we have found a fixed computational scheme to update

$$P_t = (A_t^H A_t)^{-1}$$

using old data  $P_{t-1}$  and new data  $\mathbf{q}_t$ .

## Recursive Least Squares Filtering, cont.

In the solution  $\mathbf{h}[t] = (A_t^H A_t)^{-1} A_t^H \mathbf{y}_t$ , we have found a clever way to update  $P_t = (A_t^H A_t)^{-1}$  recursively. Define

$$\mathbf{z}_t \triangleq A_t^H \mathbf{y}_t.$$

We need a recursive update for  $\mathbf{z}_t$ .

Toward that end note that

$$\begin{aligned}\mathbf{z}_t &= A_t^H \mathbf{y}_t \\ &= \sum_{i=1}^t \mathbf{q}_i y_i \\ &= \sum_{i=1}^{t-1} \mathbf{q}_i y_i + \mathbf{q}_t y_t \\ &= \mathbf{z}_{t-1} + \mathbf{q}_t y_t\end{aligned}$$

## Recursive Least Squares Filtering, cont.

Therefore

$$\begin{aligned}\mathbf{h}_t &= (A_t^H A_t)^{-1} A_t^H \mathbf{y}_t \\&= P_t \mathbf{z}_t \\&= (P_{t-1} - \mathbf{k}_t \mathbf{q}_t^H P_{t-1})(\mathbf{z}_{t-1} + \mathbf{q}_t y_t) \\&= P_{t-1} \mathbf{z}_{t-1} - \mathbf{k}_t \mathbf{q}_t^H P_{t-1} \mathbf{z}_{t-1} + P_{t-1} \mathbf{q}_t y_t - \mathbf{k}_t \mathbf{q}_t^H P_{t-1} \mathbf{q}_t y_t \\&= \mathbf{h}_{t-1} - \mathbf{k}_t \mathbf{q}_t^H \mathbf{h}_{t-1} + \underbrace{\left( P_{t-1} - \mathbf{k}_t \mathbf{q}_t^H P_{t-1} \right)}_{P_t} \mathbf{q}_t y_t \\&= \mathbf{h}_{t-1} + \mathbf{k}_t (y_t - \mathbf{q}_t^H \mathbf{h}_{t-1}) \\&\implies \mathbf{h}_t = \mathbf{h}_{t-1} + \mathbf{k}_t (y_t - \hat{y}),\end{aligned}$$

where we have used the fact that  $P_t q_t = \mathbf{k}_t$ .

Note that  $\hat{y}_t = \mathbf{q}_t^H \mathbf{h}_{t-1}$  is the predicted output, and  $e_t = y_t - \hat{y}_t$  is the quantity that is being minimized.

## Summary: Recursive Least Squares Filtering

At time  $t = 0$  initialize algorithm with

$$P_0 = \alpha I, \text{ where } \alpha > 0 \text{ is a large number}$$
$$\mathbf{h}_0 = 0.$$

At time  $t$ , get  $y_t$ ,  $f_t$ , and compute  $\mathbf{q}_t$  from  $f_t$ . Update the least squares estimate using

$$\mathbf{k}_t = \frac{P_{t-1}\mathbf{q}_t}{1 + \mathbf{q}_t^H P_{t-1} \mathbf{q}_t}$$
$$P_t = P_{t-1} - \mathbf{k}_t \mathbf{q}_t^H P_{t-1}$$
$$\mathbf{h}_t = \mathbf{h}_{t-1} + \mathbf{k}_t(y_t - \mathbf{q}_t^H \mathbf{h}_{t-1}).$$

This is equivalent to a discrete time Kalman filter with stationary dynamics.

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 5.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

LU Factorization

Cholesky Factorization

QR Factorization

# Section 1

## LU Factorization

## LU Factorization

- ▶ Suppose that  $A \in \mathbb{C}^{n \times n}$  is full rank. What is a numerically efficient method for computing the solution to  $Ax = b$ , i.e.  $x = A^{-1}b$ ?
- ▶ An explicit formula is:

$$x = \frac{\text{adj}(A)b}{\det(A)}$$

but this requires numerical computation of determinants.

- ▶ LU factorization is more efficient.

## LU Factorization: Basic Idea

- ▶ Find a permutation matrix  $P$ , a lower diagonal matrix with ones on the diagonal  $L$ , and an upper diagonal matrix  $U$  such that

$$PA = LU.$$

- ▶ How? Will illustrate by example:

## LU Factorization: cont.

Let

$$A = \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}$$

The idea is to perform row reductions to get a triangular matrix.

**Key Idea:** Reduce the row with the largest element.

## LU Factorization: cont.

First, permute  $A$  to get the third row on top:

$$\begin{aligned} P_{13}A &= \underbrace{\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}}_{P_{13}} \begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix} \\ &= \begin{pmatrix} 7 & -8 & 9 \\ -4 & 5 & -6 \\ 1 & -2 & 3 \end{pmatrix} \end{aligned}$$

The idea is that you always want to divide by the largest element (in absolute value) in the row to avoid numerical problems.

## LU Factorization: cont.

Now zero out the  $-4$  and  $1$  by multiplying the first row by  $+\frac{4}{7}$  and adding to the second row and multiplying the first row by  $-\frac{1}{7}$  and adding to the third row:

$$\begin{aligned} E_1 P_{13} A &= \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ \frac{4}{7} & 1 & 0 \\ -\frac{1}{7} & 0 & 1 \end{pmatrix}}_{E_1} \begin{pmatrix} 7 & -8 & 9 \\ -4 & 5 & -6 \\ 1 & -2 & 3 \end{pmatrix} \\ &= \begin{pmatrix} 7 & -8 & 9 \\ 0 & 0.4286 & -5.4286 \\ 0 & -0.8571 & 2.8571 \end{pmatrix} \end{aligned}$$

## LU Factorization: cont.

Now permute (or “pivot”) to get the largest (in absolute value) number in the second column in the second row:

$$\begin{aligned} P_{23}E_1P_{13}A &= \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{P_{23}} \begin{pmatrix} 7 & -8 & 9 \\ 0 & 0.4286 & -5.4286 \\ 0 & -0.8571 & 2.8571 \end{pmatrix} \\ &= \begin{pmatrix} 7 & -8 & 9 \\ 0 & -0.8571 & 2.8571 \\ 0 & 0.4286 & -5.4286 \end{pmatrix} \end{aligned}$$

## LU Factorization: cont.

Zero out the 0.4286 by multiplying the second row by  $\frac{0.4286}{0.8571}$  and adding to the third row:

$$\begin{aligned} E_2 P_{23} E_1 P_{13} A &= \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{0.4286}{0.8571} & 1 \end{pmatrix}}_{E_2} \begin{pmatrix} 7 & -8 & 9 \\ 0 & -0.8571 & 2.8571 \\ 0 & 0.4286 & -5.4286 \end{pmatrix} \\ &= \begin{pmatrix} 7 & -8 & 9 \\ 0 & -0.8571 & 2.8571 \\ 0 & 0 & -4 \end{pmatrix} \\ &= U \end{aligned}$$

Therefore

$$\begin{aligned} A &= (E_2 P_{23} E_1 P_{13})^{-1} U \\ &= P_{13}^{-1} E_1^{-1} P_{23}^{-1} E_2^{-1} U \end{aligned}$$

## LU Factorization: cont.

Note that if  $E_1 = \begin{pmatrix} 1 & 0 & 0 \\ \frac{4}{7} & 1 & 0 \\ -\frac{1}{7} & 0 & 1 \end{pmatrix}$ , then  $E_1^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{4}{7} & 1 & 0 \\ \frac{1}{7} & 0 & 1 \end{pmatrix}$

since

$$\begin{pmatrix} 1 & 0 & 0 \\ \frac{4}{7} & 1 & 0 \\ -\frac{1}{7} & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ -\frac{4}{7} & 1 & 0 \\ \frac{1}{7} & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

So the inverse of any lower diagonal matrix formed by multiplying and adding rows is found by negating the off-diagonal terms.

Therefore  $E_1^{-1}$  and  $E_2^{-1}$  are easy to compute.

## LU Factorization: cont.

Also note that for permutation matrices

$$P_{ij}^{-1} = P_{ji}$$

since

$$\underbrace{P_{ij}}_{\text{switch } ij \text{ rows}} \quad \underbrace{P_{ij}^{-1}}_{\text{switch back}} = I.$$

For example

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

## LU Factorization: cont.

So we have  $A = VU$  where

$$V = P_{13}E_1^{-1}P_{23}E_2^{-1} = \begin{pmatrix} 0.1429 & 1 & 0 \\ -0.5714 & -0.5 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

Note that  $V$  is not lower triangular but

$$\begin{aligned} L &= P_{23}P_{13}V = P_{23} \begin{pmatrix} 1 & 0 & 0 \\ -0.5714 & -0.5 & 1 \\ 0.1429 & 1 & 0 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0.1429 & 1 & 0 \\ -0.5714 & -0.5 & 1 \end{pmatrix} \end{aligned}$$

is, so  $P_{23}P_{13}A = P_{23}P_{13}VU$ . Therefore

$$PA = LU$$

where  $P = P_{23}P_{13}$ .

## LU Factorization: cont.

For our example we have

$$\underbrace{\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}}_P \underbrace{\begin{pmatrix} 1 & -2 & 3 \\ -4 & 5 & -6 \\ 7 & -8 & 9 \end{pmatrix}}_A = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0.1429 & 1 & 0 \\ -0.5714 & -0.5 & 1 \end{pmatrix}}_L \underbrace{\begin{pmatrix} 7 & -8 & 9 \\ 0 & -0.8571 & 2.8571 \\ 0 & 0 & -4 \end{pmatrix}}_U$$

How do we solve the equation  $Ax = b$ ?

Suppose  $b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$

Note that

$$PAx = Pb = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

So that

$$LUx = Pb.$$

## LU Factorization: cont.

Let  $y = Ux$  then

$$Ly = Pb$$

$$\Rightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0.1429 & 1 & 0 \\ -0.5714 & -0.5 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$$

$$\Rightarrow \begin{cases} y_1 = 3 \\ y_2 = 1 - 0.1429y_1 \\ y_3 = 2 + 0.5714y_1 + 0.5y_2 \end{cases} \quad (\text{easy to solve})$$

$$\Rightarrow \begin{cases} y_1 = 3 \\ y_2 = 0.5741 \\ y_3 = 4 \end{cases} \quad (\text{easy to solve})$$

## LU Factorization: cont.

Next solve  $Ux = y$  for  $x$ :

$$Ux = y$$

$$\Rightarrow \begin{pmatrix} 7 & -8 & 9 \\ 0 & -0.8571 & 2.8571 \\ 0 & 0 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 3 \\ 0.5714 \\ 4 \end{pmatrix}$$

$$\Rightarrow \begin{cases} -4x_3 = 4 \\ -0.8571x_2 = 0.5714 - 2.8571x_2 \\ 7x_1 = 3 + 8x_2 - x_3 \end{cases} \quad (\text{easy to solve})$$

$$\Rightarrow \begin{cases} x_1 = -4 \\ x_2 = -4 \\ x_3 = -1 \end{cases}$$

## LU Factorization: cont.

In Matlab:

```
A = [1, 2, 3; 4, 5, 6; 7, 8, 0];
[L, U, P] = lu(A)
```

In Python:

```
import numpy as np
import scipy.linalg as linalg

A = np.array([[1, 2, 3], [4, 5, 6], [7, 8, 9]])
P, L, U = linalg.lu(A)
```

**Homework problem:** Write your own custom `lu` function and compare to the built in `lu` function on 100 randomly generated matrices.

## Section 2

### Cholesky Factorization

## Square Root of a Matrix

- ▶ If  $B = B^H > 0$  then we can compute the “square root” of  $B$  as  $B = QQ^H$  where  $Q = B^{\frac{1}{2}}$  is the square root of  $B$ .
- ▶ In general, the square root of a matrix is not unique!

### Example

Let  $B = \begin{pmatrix} 9 & 0 \\ 0 & 0 \end{pmatrix}$

We can write

$$B = \begin{pmatrix} 3 \\ 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$$

So both  $Q = \begin{pmatrix} 3 \\ 0 \end{pmatrix}$  and  $Q = \begin{pmatrix} 3 & 0 \\ 0 & 0 \end{pmatrix}$  are square roots of  $B$ .

# Cholesky Factorization

## Definition

The Cholesky factorization of  $B$  is a square lower triangular square root  $L \in \mathbb{C}^{n \times n}$  of  $B$ , where

$$B = LL^H.$$

Note that this can also be written as

$$B = U^H U$$

where  $U = L^H$  is upper triangular.

## Cholesky Factorization: Numerical Algorithm

Let  $B = \begin{pmatrix} \alpha & \mathbf{v}^H \\ \mathbf{v} & B_1 \end{pmatrix}$ . Then factor  $B$  as

$$\begin{aligned} B &= \begin{pmatrix} \alpha & \mathbf{v}^H \\ \mathbf{v} & B_1 \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{\mathbf{v}}{\sqrt{\alpha}} & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & B_1 - \frac{\mathbf{v}\mathbf{v}^H}{\alpha} \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & \frac{\mathbf{v}^H}{\sqrt{\alpha}} \\ 0 & I_{n-1} \end{pmatrix} \end{aligned}$$

## Cholesky Factorization: Numerical Algorithm, cont.

(RECURSIVE ALGORITHM)

Now find the Cholesky factorization of  $B_1 - \frac{\mathbf{w}^H}{\alpha} \triangleq G_1 G_1^H$ , so that

$$\begin{aligned} B &= \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{\mathbf{v}}{\sqrt{\alpha}} & I_{n-1} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & G_1 G_1^H \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & \frac{\mathbf{v}^H}{\sqrt{\alpha}} \\ 0 & I_{n-1} \end{pmatrix} \\ &= \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{\mathbf{v}}{\sqrt{\alpha}} & G_1 \end{pmatrix} \begin{pmatrix} \sqrt{\alpha} & \frac{\mathbf{v}^H}{\sqrt{\alpha}} \\ 0 & G_1^H \end{pmatrix} \end{aligned}$$

which implies that the Cholesky factor is

$$L = \begin{pmatrix} \sqrt{\alpha} & 0 \\ \frac{\mathbf{v}}{\sqrt{\alpha}} & G_1 \end{pmatrix}.$$

## Cholesky Factorization: Example

Let  $B = \begin{pmatrix} 1 & 2 & 4 & 1 \\ 2 & 13 & 17 & 8 \\ 4 & 17 & 29 & 16 \\ 1 & 8 & 16 & 30 \end{pmatrix}$ . Then

$$B = \begin{pmatrix} \alpha_1 & \mathbf{v}_1^\top \\ \mathbf{v}_1 & B_1 \end{pmatrix},$$

where

$$\alpha_1 = 1$$

$$\mathbf{v}_1 = (2 \quad 4 \quad 1)^\top$$

$$B_1 = \begin{pmatrix} 13 & 17 & 8 \\ 17 & 29 & 16 \\ 8 & 16 & 30 \end{pmatrix}.$$

## Cholesky Factorization: Example, cont.

Therefore

$$\begin{aligned} B &= \begin{pmatrix} \sqrt{\alpha_1} & 0^\top \\ \frac{\mathbf{v}_1}{\sqrt{\alpha_1}} & G_1 \end{pmatrix} \begin{pmatrix} \sqrt{\alpha_1} & \frac{\mathbf{v}_1^\top}{\sqrt{\alpha_1}} \\ 0 & G_1^\top \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & & & \\ 4 & & G_1 & \\ 1 & & & \end{pmatrix} \begin{pmatrix} 1 & 2 & 4 & 1 \\ 0 & & & \\ 0 & & G_1^\top & \\ 0 & & & \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} G_1 G_1^\top &= B_1 - \frac{\mathbf{v}_1 \mathbf{v}_1^\top}{\alpha_1} \\ &= \begin{pmatrix} 13 & 17 & 8 \\ 17 & 29 & 16 \\ 8 & 16 & 30 \end{pmatrix} - \frac{1}{1} \begin{pmatrix} 2 \\ 4 \\ 1 \end{pmatrix} (2 \quad 4 \quad 1) \\ &= \begin{pmatrix} 9 & 9 & 6 \\ 9 & 13 & 12 \\ 6 & 12 & 29 \end{pmatrix}. \end{aligned}$$

## Cholesky Factorization: Example, cont.

Therefore

$$\begin{aligned} G_1 G_1^\top &= \begin{pmatrix} \sqrt{\alpha_2} & 0^\top \\ \frac{\mathbf{v}_2}{\sqrt{\alpha_2}} & G_2 \end{pmatrix} \begin{pmatrix} \sqrt{\alpha_2} & \frac{\mathbf{v}_2^\top}{\sqrt{\alpha_2}} \\ 0 & G_2^\top \end{pmatrix} \\ &= \begin{pmatrix} 3 & 0 & 0 \\ 3 & & \\ 2 & & G_2 \end{pmatrix} \begin{pmatrix} 3 & 3 & 2 \\ 0 & & \\ 0 & & G_2^\top \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} G_2 G_2^\top &= B_2 - \frac{\mathbf{v}_2 \mathbf{v}_2^\top}{\alpha_2} \\ &= \begin{pmatrix} 13 & 12 \\ 12 & 29 \end{pmatrix} - \frac{1}{9} \begin{pmatrix} 9 \\ 6 \end{pmatrix} \begin{pmatrix} 9 & 6 \end{pmatrix} \\ &= \begin{pmatrix} 4 & 6 \\ 6 & 25 \end{pmatrix}. \end{aligned}$$

## Cholesky Factorization: Example, cont.

Therefore

$$\begin{aligned} G_2 G_2^\top &= \begin{pmatrix} \sqrt{\alpha_3} & 0^\top \\ \frac{\mathbf{v}_3}{\sqrt{\alpha_3}} & G_3 \end{pmatrix} \begin{pmatrix} \sqrt{\alpha_3} & \frac{\mathbf{v}_3^\top}{\sqrt{\alpha_3}} \\ 0 & G_3^\top \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 3 & G_3 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & G_3^\top \end{pmatrix} \end{aligned}$$

where

$$\begin{aligned} G_3 G_3^\top &= B_3 - \frac{\mathbf{v}_3 \mathbf{v}_3^\top}{\alpha_3} \\ &= 25 - \frac{1}{4} 3 \cdot 3 \\ &= 16 \end{aligned}$$

Therefore  $G_3 = 4$ .

## Cholesky Factorization: Example, cont.

Combining gives

$$\begin{aligned}L &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & & & \\ 4 & & G_1 & \\ 1 & & & \end{pmatrix} \\&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 3 & & \\ 1 & 2 & & G_2 \end{pmatrix} \\&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 3 & 2 & 0 \\ 1 & 2 & 3 & G_3 \end{pmatrix} \\&= \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 3 & 0 & 0 \\ 4 & 3 & 2 & 0 \\ 1 & 2 & 3 & 4 \end{pmatrix}.\end{aligned}$$

## Applications of Cholesky Factorization: Quadratic Forms

The quadratic form

$$x^H Q x = \|x\|_Q^2$$

where  $Q = Q^H$ , can be written as

$$x^H Q x = x^H U^H U x = \|Ux\|_2^2$$

where  $Q = U^H U = LL^H$

In other words, work with the regular 2-norm as opposed to the  $Q$  norm.

## Applications of Cholesky Factorization: Simulating a random vector

Suppose you want to generate in Matlab/Simulink/Python/etc. a Gaussian random vector with covariance  $R = R^T > 0$ .

The Matlab `randn([m, 1])` command returns an  $m \times 1$  random vector which is normally distributed with zero mean and co-variance  $I$  ( $\mathcal{N}(0, I)$ ).

To generate  $\mathcal{N}(0, R)$  let  $R = LL^T$  and let  $z = Lx$  where  $x \sim \mathcal{N}(0, I)$ .

Then

$$\begin{aligned} E\{zz^T\} &= E\{Lxx^T L^T\} = LE\{xx^T\}L^T = LL^T = R \\ \Rightarrow z &\sim \mathcal{N}(0, R). \end{aligned}$$

## Applications of Cholesky Factorization: Solving normal equations

Normal equations are given by

$$R\mathbf{c} = \mathbf{b}$$

where  $R = R^H$  is the Grammian and full rank if the data vectors are linearly independent.

Let  $R = LL^H$ , then  $LL^H\mathbf{c} = \mathbf{b}$

First solve

$$L\mathbf{y} = \mathbf{b}$$

by forward substitution, and then solve

$$L^H\mathbf{c} = \mathbf{y}$$

by backward substitution.

## Applications of Cholesky Factorization: Kalman filtering

In Kalman filtering we propagate two items; The estimate  $\hat{x}(k)$  and the error covariance  $P(k)$  where  $P(k) = P^T(k) > 0$ .

If implemented directly, numerical error can cause  $P(k)$  to become indefinite introducing large errors into the estimate  $\hat{x}(k)$ .

To avoid this problem a “square root” Kalman filter is usually implemented where  $P(k) = L(k)L^T(k)$  and  $L(k)$  is propagated instead of  $P(k)$ . Then even with numerical errors in  $L(k)$ ,  $P(k)$  is still symmetric positive definite.

## Cholesky Factorization: cont.

In Matlab:

```
L1 = [2, 0, 0; 3, 4, 0; 5, 6, 7];  
A = L1 * L1';  
L = chol(A)'
```

In Python:

```
import numpy as np  
import scipy.linalg as linalg  
  
L1 = np.array([[2, 0, 0], [3, 4, 0], [5, 6, 7]])  
A = L1 @ L1.T  
L = linalg.cholesky(A)
```

$L$  should equal  $L_1$ .

Note that both Matlab and Python return an upper triangular matrix.

**Homework problem:** Write your own custom `cholesky` function and compare to the built in `cholesky` function on 100 randomly generated symmetric matrices.

## Section 3

### QR Factorization

# Unitary and Orthogonal Matrices

## Definition

$Q \in \mathbb{C}^{m \times m}$  is unitary if

$$Q^H Q = QQ^H = I$$

Equivalently  $Q^{-1} = Q^H$ .

Equivalently, the rows of  $Q$  form an orthonormal set.

Equivalently, the columns of  $Q$  form an orthonormal set.

## Definition

$Q \in \mathbb{R}^{m \times m}$  is orthogonal if

$$Q^T Q = QQ^T = I.$$

Rotation matrices are examples of orthogonal matrices.

# Hermitian Matrices

## Definition

$Q \in \mathbb{C}^{m \times m}$  is Hermitian if  $Q^H = Q$

Hermitian matrices are like real numbers, i.e.,  $\bar{z} = z$ .

Unitary matrices correspond to the unit circle

$$|z|^2 = \bar{z}z = 1$$

Bilinear transformation

$z = \frac{1+jr}{1-jr}$  maps the real line to the unit circle

For matrices this becomes Cayley's formula

$$U = (I + jR)(I - jR)^{-1}$$

which maps Hermitian (analogous to real #'s) to unitary matrices (analogous to complex unit circle).

## Unitary Matrices, cont

Lemma (Moon Lemma 5.1)

Let  $Q \in \mathbb{C}^{m \times m}$  then  $\|Qx\|_2 = \|x\|_2, \forall x \in \mathbb{C}^m$  iff  $Q$  is unitary.

Proof.

If  $Q$  is unitary then

$$\|Qx\|_2 = \langle Qx, Qx \rangle^{\frac{1}{2}} = (x^H Q^H Q x)^{\frac{1}{2}} = (x^H x)^{\frac{1}{2}} = \|x\|_2$$

Conversely if  $\|Qx\|_2 = \|x\|_2, \forall x \in \mathbb{C}^m$  then

$$\begin{aligned} x^H Q^H Q x &= x^H x & \forall x \in \mathbb{C}^m \\ \iff x^H (Q^H Q - I) x &= 0 & \forall x \in \mathbb{C}^m \\ \iff Q^H Q &= I. \end{aligned}$$

Therefore  $Q$  is unitary. □

## Unitary Matrices, cont

Lemma (Moon Lemma 5.2)

If  $Y = QX$  where  $Q$ -unitary then

$$\|Y\|_F = \|X\|_F$$

# Unitary Matrices, cont.

## Lemma

If  $Q_1$  and  $Q_2$  are unitary then  $Q_2 Q_1$  is unitary.

Proof.

$$(Q_2 Q_1)^H (Q_2 Q_1) = Q_1^H Q_2^H Q_2 Q_1 = Q_1^H Q_1 = I.$$



# QR - Factorization

## Definition

Let  $A \in \mathbb{C}^{m \times n}$ . The QR factorization of  $A$  is given by

$$A = QR$$

where  $Q \in \mathbb{C}^{m \times m}$  is unitary and  $R \in \mathbb{C}^{m \times n}$  is upper triangular.

## Lemma

*Every matrix  $A \in \mathbb{C}^{m \times n}$  has a QR factorization.*

## QR - Factorization, cont.

### Example

$$A = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \end{pmatrix} = \underbrace{\begin{pmatrix} -0.1961 & -0.9806 \\ -0.9806 & 0.1961 \end{pmatrix}}_Q \underbrace{\begin{pmatrix} -5.0090 & -6.2757 & -7.4524 & -8.6291 \\ 0 & -0.7845 & -1.5689 & -2.3534 \end{pmatrix}}_R$$

In Matlab:

```
[Q, R] = qr(A)
```

In Python:

```
Q, R = scipy.linalg.qr(A)
```

## QR - Factorization, cont.

### Example

$$A = \begin{pmatrix} 1 & 5 \\ 2 & 6 \\ 3 & 7 \\ 4 & 8 \end{pmatrix}$$
$$= \underbrace{\begin{pmatrix} -0.1826 & -0.8165 & -0.4001 & -0.3741 \\ -0.3651 & -0.4082 & 0.2546 & 0.797 \\ -0.5477 & 0 & 0.6910 & -0.4717 \\ -0.7303 & 0.4082 & -0.5455 & 0.0488 \end{pmatrix}}_Q \underbrace{\begin{pmatrix} -5.4772 & -12.7 \\ 0 & -3.266 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}}_R$$

## Application: Full rank least squares

If  $A \in \mathbb{C}^{m \times n}$  is full rank and  $m > n$ , find

$$\hat{x} = \arg \min \|Ax - b\|_2$$

Recall that the solution is  $\hat{x} = (A^H A)^{-1} A^H b$  but

$$\mathcal{K}(A^H A) = (\mathcal{K}(A))^2$$

Therefore computing the inverse of  $A^H A$  with LU or Cholesky factorization may be ill-advised.

Use QR factorization instead.

## Application: Full rank least squares, cont.

Let  $A = QR = Q \begin{bmatrix} R_1 \\ 0 \end{bmatrix}$  where  $Q \in \mathbb{C}^{m \times m}$  and  $R_1 \in \mathbb{C}^{n \times n}$  is upper triangular.

Let  $Q^H \mathbf{b} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}$  where  $\mathbf{c} \in \mathbb{C}^n$  and  $\mathbf{d} \in \mathbb{C}^{m-n}$ .

Then

$$\begin{aligned}\|A\mathbf{x} - \mathbf{b}\|_2^2 &= \|QR\mathbf{x} - \mathbf{b}\|_2^2 \\&= \left\|Q(R\mathbf{x} - Q^H \mathbf{b})\right\|_2^2 \quad (\text{since } QQ^H = I) \\&= \left\|\begin{bmatrix} R_1 \\ 0 \end{bmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}\right\|_2^2 \quad (\text{by lemma 5.1}) \\&= \|R_1 \mathbf{x} - \mathbf{c}\|_2^2 + \|\mathbf{d}\|_2^2 \quad (\text{by definition of 2-norm})\end{aligned}$$

so  $\hat{\mathbf{x}} = \arg \min \|A\mathbf{x} - \mathbf{b}\|_2^2$  satisfies  $R_1 \hat{\mathbf{x}} = \mathbf{c}$  where  $\hat{\mathbf{x}}$  is easily found by forward-substitution.

## Application: Full rank least squares, cont.

Note that we don't actually need to compute all of  $Q$  since

$$Q^H \mathbf{b} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}.$$

Let

$$Q = \begin{pmatrix} Q_1 & Q_2 \\ m \times n & m \times (m-n) \end{pmatrix}$$

then

$$Q^H \mathbf{b} = \begin{pmatrix} Q_1^H \\ Q_2^H \end{pmatrix} \mathbf{b} = \begin{pmatrix} Q_1^H \mathbf{b} \\ Q_2^H \mathbf{b} \end{pmatrix} = \begin{pmatrix} \mathbf{c} \\ \mathbf{d} \end{pmatrix}$$

so  $\mathbf{c} = Q_1^H \mathbf{b}$ .

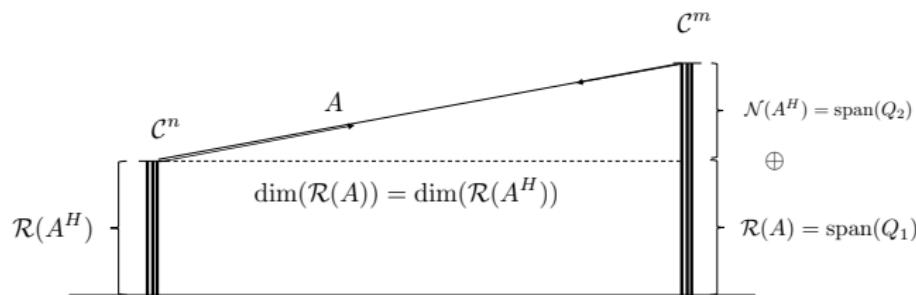
Therefore, we only need the first  $n$  columns of  $Q$ .

# QR Factorization and Fundamental Subspaces

If  $A$  is tall then

$$A = QR$$

$$= \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} R_1 \\ \mathbf{0} \end{pmatrix}$$



# Computational Methods for QR Factorization

We will discuss two methods for computing the QR Factorization:

- ▶ Given rotation.
- ▶ Householder transformation.

## QR Factorization using Given Rotation

The basic idea is to diagonalize  $A$  one element at a time: So find

$$Q_1 \text{ such that } Q_1 A = \begin{pmatrix} x & x \\ 0 & x \\ x & x \end{pmatrix}$$

$$\text{Then find } Q_2 \text{ such that } Q_2 Q_1 A = \begin{pmatrix} x & x \\ 0 & x \\ 0 & x \end{pmatrix}$$

$$\text{Then find } Q_3 \text{ such that } Q_3 Q_2 Q_1 A = \begin{pmatrix} x & x \\ 0 & x \\ 0 & 0 \end{pmatrix}$$

Then

$$\begin{aligned} A &= (Q_3 Q_2 Q_1)^{-1} R \\ &= (Q_3 Q_2 Q_1)^H R \quad (\text{since } (Q_3 Q_2 Q_1) \text{ is unitary}) \\ &= \underbrace{Q_1^H Q_2^H Q_3^H}_{\hat{Q}} R \\ &= QR. \end{aligned}$$

# QR Factorization using Givens Rotations

Consider the  $2 \times 2$  rotation matrix

$$G(\theta) = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}$$

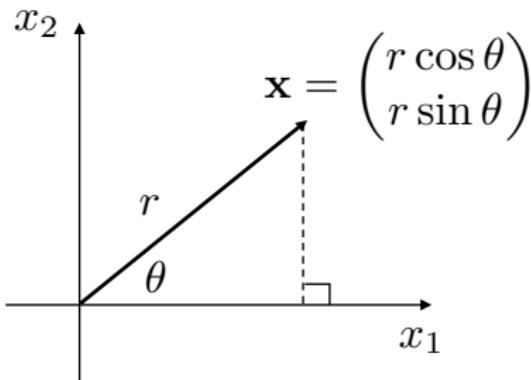
Note that

$$G^{-1}(\theta) = G^T(\theta) = \begin{pmatrix} \cos \theta & -\sin \theta \\ +\sin \theta & \cos \theta \end{pmatrix}$$

Therefore,  $G(\theta)$  is orthogonal and hence unitary.

## QR Factorization using Givens Rotations

Let  $x = \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \in \mathbb{R}^2$ :



Then

$$\begin{aligned} G(\theta)x &= \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix} \\ &= \begin{pmatrix} r \cos^2(\theta) + r \sin^2(\theta) \\ -r \sin \theta \cos \theta + r \cos \theta \sin \theta \end{pmatrix} \\ &= \begin{pmatrix} r \\ 0 \end{pmatrix}. \end{aligned}$$

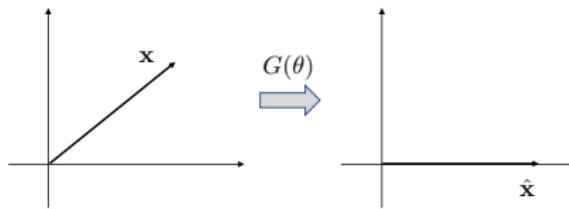
# QR Factorization using Givens Rotations

Therefore  $G(\theta)$  rotated

$$x = \begin{pmatrix} r \cos \theta \\ r \sin \theta \end{pmatrix}$$

to

$$\hat{x} = \begin{pmatrix} r \\ 0 \end{pmatrix}.$$



## QR Factorization using Givens Rotations

Note that if  $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$  then  $\theta = \tan^{-1} \left( \frac{x_2}{x_1} \right)$  and

$$\cos \theta = \cos \left( \tan^{-1} \left( \frac{x_2}{x_1} \right) \right) = \frac{x_1}{\sqrt{x_1^2 + x_2^2}}$$

$$\sin \theta = \sin \left( \tan^{-1} \left( \frac{x_2}{x_1} \right) \right) = \frac{x_2}{\sqrt{x_1^2 + x_2^2}}$$

Therefore

$$G_x(\theta) = \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2 + x_2^2}} & \frac{x_2}{\sqrt{x_1^2 + x_2^2}} \\ -\frac{x_2}{\sqrt{x_1^2 + x_2^2}} & \frac{x_1}{\sqrt{x_1^2 + x_2^2}} \end{pmatrix}.$$

Note that each term in  $G_x(\theta)$  decreases as a result of dividing by  $\frac{1}{\sqrt{x_1^2 + x_2^2}}$  so even if  $x_1$  and  $x_2$  are small, this is numerically stable.

## QR Factorization using Givens Rotations: Example

Let

$$A = \begin{pmatrix} 1 & 6 & 7 & 12 \\ 2 & 5 & 8 & 11 \\ 13 & 4 & 9 & 10 \end{pmatrix}$$

Letting  $x_1 = 1$  and  $x_2 = 2$  and

$$Q_1 = \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2+x_2^2}} & \frac{x_2}{\sqrt{x_1^2+x_2^2}} & 0 \\ -\frac{x_2}{\sqrt{x_1^2+x_2^2}} & \frac{x_1}{\sqrt{x_1^2+x_2^2}} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0.4472 & 0.8944 & 0. \\ -0.8944 & 0.4472 & 0. \\ 0 & 0 & 1 \end{pmatrix}$$

gives

$$Q_1 A = \begin{pmatrix} 2.2360 & 7.1554 & 10.2859 & 15.2052 \\ 0 & -3.1304 & -2.6832 & -5.8137 \\ 13 & 4 & 9 & 10 \end{pmatrix}.$$

## QR Factorization using Givens Rotations: Example

$$Q_1 A = \begin{pmatrix} 2.2360 & 7.1554 & 10.2859 & 15.2052 \\ 0 & -3.1304 & -2.6832 & -5.8137 \\ 13 & 4 & 9 & 10 \end{pmatrix}.$$

Letting  $x_1 = 2.2360$  and  $x_2 = 13$  and

$$Q_2 = \begin{pmatrix} \frac{x_1}{\sqrt{x_1^2+x_2^2}} & 0 & \frac{x_2}{\sqrt{x_1^2+x_2^2}} \\ 0 & 1 & 0 \\ -\frac{x_2}{\sqrt{x_1^2+x_2^2}} & 0 & \frac{x_1}{\sqrt{x_1^2+x_2^2}} \end{pmatrix} = \begin{pmatrix} 0.1695 & 0 & 0.9855 \\ 0 & 1 & 0 \\ -0.9855 & 0 & 0.1695 \end{pmatrix}$$

gives

$$Q_2 Q_1 A = \begin{pmatrix} 13.1909 & 5.1550 & 10.6133 & 12.4328 \\ 0 & -3.1304 & -2.6832 & -5.8137 \\ 0 & -6.3737 & -8.6114 & -13.2900 \end{pmatrix}.$$

## QR Factorization using Givens Rotations: Example

$$Q_2 Q_1 A = \begin{pmatrix} 13.1909 & 5.1550 & 10.6133 & 12.4328 \\ 0 & -3.1304 & -2.6832 & -5.8137 \\ 0 & -6.3737 & -8.6114 & -13.2900 \end{pmatrix}.$$

Letting  $x_1 = -3.1304$  and  $x_2 = -6.3737$  and

$$Q_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{x_1}{\sqrt{x_1^2+x_2^2}} & \frac{x_2}{\sqrt{x_1^2+x_2^2}} \\ 0 & -\frac{x_2}{\sqrt{x_1^2+x_2^2}} & \frac{x_1}{\sqrt{x_1^2+x_2^2}} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -0.44084797 & -0.8975 \\ 0 & 0.89758179 & -0.4408 \end{pmatrix}$$

gives

$$Q_3 Q_2 Q_1 A = \begin{pmatrix} 13.1909 & 5.1550 & 10.6133 & 12.4328 \\ 0 & 7.101 & 8.912 & 14.4918 \\ 0 & 0 & 1.3878 & 0.6405 \end{pmatrix}.$$

## QR Factorization using Givens Rotations: Example

Therefore

$$\underbrace{\begin{pmatrix} 1 & 6 & 7 & 12 \\ 2 & 5 & 8 & 11 \\ 13 & 4 & 9 & 10 \end{pmatrix}}_A = \underbrace{\begin{pmatrix} 0.0967 & 0.9077 & -0.4082 \\ 0.4834 & 0.3157 & 0.816 \\ 0.8701 & -0.2763 & 0.4082 \end{pmatrix}}_Q \underbrace{\begin{pmatrix} 13.1909 & 5.1550 & 10.6133 & 12.4328 \\ 0 & 7.101 & 8.912 & 14.4918 \\ 0 & 0 & 1.3878 & 0.6405 \end{pmatrix}}_R$$

where

$$\begin{aligned} Q &= Q_1^H Q_2^H Q_3^H \\ &= \begin{pmatrix} 0.0758 & 0.7899 & -0.6085 \\ 0.1516 & 0.5940 & 0.7900 \\ 0.9855 & -0.1521 & -0.0747 \end{pmatrix}. \end{aligned}$$

## QR Factorization using Givens Rotations: Example

```
import numpy as np

def Q_givens(x1, x2, size, m, n):
    Q = np.eye(size)
    cos_theta = x1 / np.sqrt(x1**2 + x2**2)
    sin_theta = x2 / np.sqrt(x1**2 + x2**2)
    Q[n-1, n-1] = cos_theta
    Q[m-1, m-1] = cos_theta
    Q[m-1, n-1] = -sin_theta
    Q[n-1, m-1] = sin_theta
    return Q

A = np.array([[1, 6, 7, 12],
              [2, 5, 8, 11],
              [13, 4, 9, 10]])
```

## QR Factorization using Givens Rotations: Example

```
Q1 = Q_givens(x1=1, x2=2, size=3, m=2, n=1)
R1 = Q1 @ A
```

```
Q2 = Q_givens(x1=R1[0,0], x2=R1[2,0], size=3,
                m=3, n=1)
```

```
R2 = Q2 @ Q1 @ A
```

```
Q3 = Q_givens(x1=R2[1,1], x2=R2[2,1], size=3,
                m=3, n=2)
```

```
R3 = Q3 @ Q2 @ Q1 @ A
```

```
Q = Q1.conj().T @ Q2.conj().T @ Q3.conj().T
```

```
print("Q=", Q)
```

```
print("R=", R3)
```

# QR Factorization using Householder Transformation

The basic idea is to diagonalize  $A$  one column at a time using unitary matrices.

## Lemma

If  $Q_1$  and  $Q_2$  are unitary then  $Q_2 Q_1$  is unitary.

## Proof.

$$(Q_2 Q_1)^H (Q_2 Q_1) = Q_1^H Q_2^H Q_2 Q_1 = Q_1^H Q_1 = I.$$



## QR Factorization using Householder Transformation

So find  $Q_1$  such that  $Q_1 A = \begin{pmatrix} x & x & x \\ 0 & x & x \\ 0 & x & x \\ 0 & x & x \end{pmatrix}$

Then find  $Q_2$  such that  $Q_2 Q_1 A = \begin{pmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & x \end{pmatrix}$

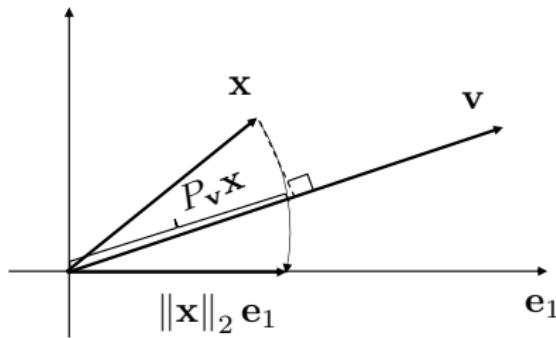
Then find  $Q_3$  such that  $Q_3 Q_2 Q_1 A = \begin{pmatrix} x & x & x \\ 0 & x & x \\ 0 & 0 & x \\ 0 & 0 & 0 \end{pmatrix}$

Then

$$\begin{aligned} A &= (Q_3 Q_2 Q_1)^{-1} R \\ &= (Q_3 Q_2 Q_1)^H R \quad (\text{since } (Q_3 Q_2 Q_1) \text{ is unitary}) \\ &= \underbrace{Q_1^H Q_2^H Q_3^H}_{\hat{Q}} R \end{aligned}$$

# QR Factorization using Householder Transformation

Geometrically what do we want?



We would like to rotate  $x$  down to  $e_1$ . This can be thought of as a reflection of  $x$  about some vector  $v$

We need an operator that transforms  $x$  to  $y = \|x\|_2 e_1$

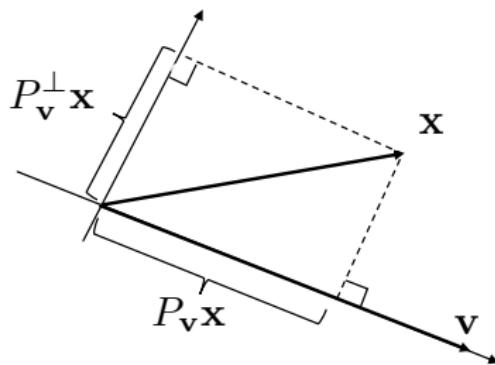
# QR Factorization using Householder Transformation

Let

$$P_v = \frac{vv^H}{v^H v}$$

be the projection matrix that projects onto the vector  $v$  and let

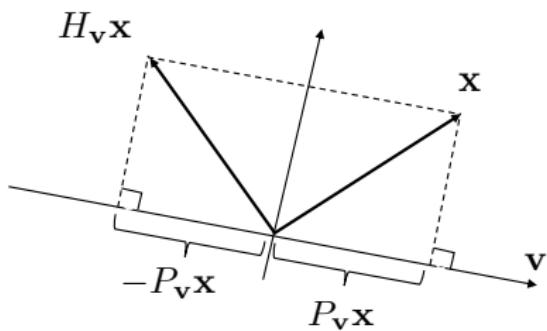
$$P_v^\perp = I - P_v$$



# QR Factorization using Householder Transformation

The Householder transformation is

$$H_v = I - 2P_v$$



$H_v \mathbf{x}$  reflects  $\mathbf{x}$  about the vector that is orthogonal to  $\mathbf{v}$ , and in the same hyperplane as both  $\mathbf{x}$  and  $\mathbf{v}$ .

# QR Factorization using Householder Transformation

Lemma

$H_v$  is unitary.

Proof.

$$\begin{aligned} H_v^H H_v &= (I - 2P_v^H)^H(I - 2P_v) \\ &= I - 2P_v - 2P_v + 4P_v^2 \\ &= I - 4P_v + 4P_v \quad (\text{since } P_v^2 = P_v) \\ &= I \end{aligned}$$



# QR Factorization using Householder Transformation

Lemma

$$H_v v = -v$$

Proof.

$$H_v v = v - 2P_v v = v - 2v = -v$$



Lemma

If  $z \perp v$  then  $H_v z = z$ .

Proof.

$$H_v z = z - zP_v z = z$$



# QR Factorization using Householder Transformation

Find  $v$  so that

$$H_v \mathbf{x} = \begin{pmatrix} \pm \|\mathbf{x}\|_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \pm \|\mathbf{x}\|_2 \mathbf{e}_1$$

i.e. the Householder transformation compresses all of the energy in  $\mathbf{x}$  into the first component.

$$H_v \mathbf{x} = \mathbf{x} - \frac{2\mathbf{v}\mathbf{v}^H}{\mathbf{v}^H\mathbf{v}}\mathbf{x} = \mathbf{x} - 2\frac{\mathbf{v}^H\mathbf{x}}{\mathbf{v}^H\mathbf{v}}\mathbf{v} = \pm \|\mathbf{x}\|_2 \mathbf{e}_1$$

Therefore

$$\left(2\frac{\mathbf{v}^H\mathbf{x}}{\mathbf{v}^H\mathbf{v}}\right)\mathbf{v} = \mathbf{x} \pm \|\mathbf{x}\|_2 \mathbf{e}_1$$

which implies that  $\mathbf{v}$  is a scalar multiple of  $\mathbf{x} \pm \|\mathbf{x}\|_2 \mathbf{e}_1$ .

## QR Factorization using Householder Transformation

- ▶ Let  $\mathbf{v} = \mathbf{x} \pm \|\mathbf{x}\|_2 \mathbf{e}_1$ .
- ▶ Numerically we would like  $\mathbf{v}$  to be large so that dividing by  $\frac{1}{\mathbf{v}^H \mathbf{v}}$  does not cause problems.
- ▶ Selecting  $\mathbf{v} = \mathbf{x} + sign(x_1) \|\mathbf{x}\|_2 \mathbf{e}_1$  implies that

$$\|\mathbf{v}\| = \|\mathbf{x} + sign(x_1) \|\mathbf{x}\|_2 \mathbf{e}_1\| \geq \|\mathbf{x}\|$$

(Since we only change the first element and the magnitude of that element always increases we can use  $\geq$ ).

- ▶ Therefore, if  $\mathbf{v} = \mathbf{x} + sign(x_1) \|\mathbf{x}\|_2 \mathbf{e}_1$  then  $H_{\mathbf{v}} \mathbf{x} = \begin{pmatrix} \|\mathbf{x}\|_2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$

and  $H_{\mathbf{v}}$  is numerically well conditioned, i.e. we are not dividing by small numbers.

## QR Factorization using Householder Transformation

Suppose that  $A = (a_1 \cdots a_n)$ .

Letting  $Q_1 = H_{v_1}$  where  $v_1 = a_1 + sign(a_{11}) \|a_1\|_2 e_1$  implies that

$$Q_1 A = \begin{pmatrix} \|a_1\|_2 & * & \cdots & * \\ 0 & & & \\ \vdots & \tilde{a}_2 & \cdots & \tilde{a}_n \\ 0 & & & \end{pmatrix}.$$

### Lemma

If  $S$  is unitary then

$$Q = \begin{pmatrix} I & 0 \\ 0 & S \end{pmatrix}$$

is unitary.

Proof.

$$QQ^H = \begin{pmatrix} I & 0 \\ 0 & S^H \end{pmatrix} \begin{pmatrix} I & 0 \\ 0 & S \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & S^H S \end{pmatrix} = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix} = I.$$

## QR Factorization using Householder Transformation

Let  $Q_2 = \begin{pmatrix} I & 0 \\ 0 & H_{v_2} \end{pmatrix}$  where  $v_2 = \tilde{a}_2 + \text{sign}(\tilde{a}_{21}) \|\tilde{a}_2\|_2 e_2$

Could also write as:

$$Q_2 = I - 2 \frac{\tilde{v}_2 \tilde{v}_2^H}{\tilde{v}_2^H \tilde{v}_2} \quad \text{where } \tilde{v}_2 = \begin{pmatrix} 0 \\ v_2 \end{pmatrix}$$

Then

$$Q_2 Q_1 A = \begin{pmatrix} \|\tilde{a}_1\|_2 & * & * & \cdots & * \\ 0 & \|\tilde{a}_2\| & * & \cdots & * \\ 0 & 0 & & & \\ \vdots & \vdots & \tilde{a}_3 & \cdots & \tilde{a}_n \\ 0 & 0 & & & \end{pmatrix}$$

The process is repeated until an upper triangular matrix is obtained on the right.

## QR Factorization using Householder Transformation: Example

$$\text{Let } A = \begin{pmatrix} 1 & -2 & 13 \\ -6 & 5 & -4 \\ 7 & -8 & 9 \\ -12 & 11 & -10 \end{pmatrix}$$

$$\text{Let } v_1 = \begin{pmatrix} 1 \\ -6 \\ 7 \\ -12 \end{pmatrix} + sign(1) \left\| \begin{pmatrix} 1 \\ -6 \\ 7 \\ -12 \end{pmatrix} \right\| e_1 = \begin{pmatrix} 6.1657 \\ -6 \\ 7 \\ -12 \end{pmatrix} \text{ and}$$

$Q_1 = I - 2 \frac{v_1 v_1^H}{v_1^H v_1}$ . Then

$$Q_1 A = \begin{pmatrix} -15.1657 & 14.5063 & -14.5063 \\ 0 & -1.1264 & 6.2091 \\ 0 & -0.8525 & -2.9106 \\ 0 & -1.2528 & 10.4182 \end{pmatrix}.$$

## QR Factorization using Householder Transformation: Example

$$Q_1 A = \begin{pmatrix} -15.1657 & 14.5063 & -14.5063 \\ 0 & -1.1264 & 6.2091 \\ 0 & -0.8525 & -2.9106 \\ 0 & -1.2528 & 10.4182 \end{pmatrix}.$$

Let

$$v_2 = \begin{pmatrix} 0 \\ -1.1264 \\ -0.8525 \\ -1.2528 \end{pmatrix} + sign(-1.1264) \left\| \begin{pmatrix} 0 \\ -1.1264 \\ -0.8525 \\ -1.2528 \end{pmatrix} \right\| e_2 = \begin{pmatrix} 0 \\ -3.0146 \\ -0.8525 \\ -1.2528 \end{pmatrix}$$

which implies that

$$Q_2 = I - 2 \frac{v_2 v_2^H}{v_2^H v_2} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -0.5965 & -0.4514 & -0.6635 \\ 0 & -0.4514 & 0.8723 & -0.1876 \\ 0 & -0.6635 & -0.1876 & 0.72424585 \end{pmatrix}.$$

# QR Factorization using Householder Transformation: Example

Then

$$Q_2 Q_1 A = \begin{pmatrix} -15.1657 & 14.5063 & -14.5063 \\ 0 & 1.88810 & -9.30270 \\ 0 & 0 & -7.29720 \\ 0 & 0 & 3.97160 \end{pmatrix}.$$

$$v_3 = \begin{pmatrix} 0 \\ 0 \\ -7.29720 \\ 3.97160 \end{pmatrix} + sign(-7.29720) \left\| \begin{pmatrix} 0 \\ -7.29720 \\ 3.97160 \end{pmatrix} \right\| e_3 = \begin{pmatrix} 0 \\ 0 \\ -15.6053 \\ 3.971 \end{pmatrix}$$

which implies that

$$Q_3 = I - 2 \frac{v_3 v_3^H}{v_3^H v_3} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -0.8783 & 0.47804 \\ 0 & 0 & 0.4780 & 0.8783 \end{pmatrix}.$$

## QR Factorization using Householder Transformation: Example

Then

$$Q_3 Q_2 Q_1 A = \begin{pmatrix} -15.1657 & 14.5063 & -14.5063 \\ 0 & 1.888 & -9.3027 \\ 0 & 0 & 8.3080 \\ 0 & 0 & 0 \end{pmatrix}.$$

$$\begin{aligned} Q &= Q_1^H Q_2^H Q_3^H \\ &= \begin{pmatrix} -0.0659 & -0.5526 & 0.8308 & 0 \\ 0.3956 & -0.3914 & -0.2289 & -0.79862957 \\ -0.4615 & -0.6907 & -0.4961 & 0.25219881 \\ 0.7912 & -0.2532 & -0.1056 & 0.54643076 \end{pmatrix} \end{aligned}$$

# QR Factorization using Householder Transformation: Example

```
import numpy as np

def Q_householder(A, column):
    (m, n) = A.shape
    x = A[column-1:m, column-1:column]
    e = np.zeros(x.shape)
    e[0,0]=1
    v = x + np.sign(x[0, 0])
        * np.linalg.norm(x) * e
    H = np.eye(m)
    H[(column-1):m, (column-1):m]
        = np.eye(m-(column-1))
        - 2 * v @ v.T / (v.T @ v)
return H
```

# QR Factorization using Householder Transformation: Example

```
A = np.array ([[1, -2, 13],  
              [-6, 5, -4],  
              [7, -8, 9],  
              [-12, 11, -10]])  
  
Q1 = Q_householder(A, column=1)  
R1 = Q1 @ A  
Q2 = Q_householder(R1, column=2)  
R2 = Q2 @ Q1 @ A  
Q3 = Q_householder(R2, column=3)  
R3 = Q3 @ Q2 @ Q1 @ A  
Q = Q1.conj().T @ Q2.conj().T @ Q3.conj().T  
print("Q=", Q)  
print("R=", R3)
```

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 6.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Eigenvalues and Eigenvectors

Jordan Form

Cayley-Hamilton Theorem

Self Adjoint Matrices

Invariant Subspaces

Quadratic Forms

Eigenfilters

# Section 1

## Eigenvalues and Eigenvectors

# Eigenpair

Let  $A \in \mathbb{C}^{n \times n}$ .

## Definition

- ▶  $(\lambda, x)$  is a right eigen-pair if  $Ax = \lambda x$  and  $x \neq 0$ .
- ▶  $(\lambda, x)$  is a left eigen-pair if  $x^H A = \lambda x^H$  and  $x \neq 0$ .

Note that  $Ax = \lambda x$  can be written as

$$(\lambda I - A)x = 0.$$

Therefore for  $x$  to be an eigenvector (associated with  $\lambda$ ) then  $x \in \mathcal{N}(\lambda I - A)$ , and

$$x \neq 0 \Rightarrow \mathcal{N}(\lambda I - A) \neq \{0\} \Rightarrow \det(\lambda I - A) = 0$$

This formula can be used to find the eigenvalues and eigenvectors of a matrix.

## Eigenpair: Example

Let  $A = \begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix}$ . Find the eigenvalues and eigenvectors.

Eigenvalues:

$$\det(\lambda I - A) = \det \begin{pmatrix} \lambda & -1 \\ 2 & \lambda + 2 \end{pmatrix} = \lambda^2 + 2\lambda + 2 = 0$$

implies that

$$\lambda = -1 \pm \sqrt{1 - 2} = -1 \pm j$$

so that

$$\lambda_1 = -1 + j \quad \lambda_2 = -1 - j.$$

Which one is larger? Note, there is no possible ordering among complex numbers.

## Eigenpair: Example

Eigenvectors: The eigenvectors can be found from the formula  
 $(\lambda I - A)x = 0$ .

$$\lambda_1 : \begin{pmatrix} -1+j & -1 \\ 2 & 1+j \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Note that the rows are linearly dependent since

$$\begin{aligned} & (-1+j)(-1+j - 1) + (2 - 1+j) \\ &= (-2 - (1+j)) + (2 - 1+j) \\ &= 0. \end{aligned}$$

Therefore, solving  $(-1+j)x_{11} - x_{12} = 0$  gives

$$x_{12} = (-1+j)x_{11}$$

Let  $x_{11} = 1$  then  $x_{12} = -1+j$ .

So

$$x_1 = \begin{pmatrix} 1 \\ -1+j \end{pmatrix}$$

is an eigenvector.

## Eigenpair: Example

$$\lambda_2 : \begin{pmatrix} -1-j & -1 \\ 2 & 1-j \end{pmatrix} \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Again the rows are linearly dependent so solve to get  
 $(-1-j)x_{21} = x_{22}$  Let  $x_{21} = 1$  then  $x_{22} = -1-j$ .

So

$$x_2 = \begin{pmatrix} 1 \\ -1-j \end{pmatrix}$$

is an eigenvector.

# Characteristic Polynomial

## Definition

The polynomial

$$\chi_A(\lambda) = \det(\lambda I - A)$$

is called the characteristic polynomial of  $A$ . The eigenvalues of  $A$  are the roots of  $\chi_A(\lambda) = 0$ . The set of roots of  $\chi_A(\lambda) = 0$  is called the spectrum of  $A$ , denoted  $\lambda(A)$ .

# Relationship between transfer function and state space models

Given a state space system:

$$\dot{x} = Ax + Bu$$

$$y = Cx$$

where  $x \in \mathbb{R}^n$ ,  $u \in \mathbb{R}^m$ ,  $y \in \mathbb{R}^p$ , what is the transfer function?

Take the Laplace transform to get

$$sX(s) = AX(s) + BU(s)$$

$$Y(s) = CX(s)$$

From the first equation we get

$$X(s) = (sl - A)^{-1}BU(s)$$

From the second equation we get

$$Y(s) = \underbrace{C(sl - A)^{-1}B}_{(p \times m) \text{ transfer matrix}} U(s)$$

## Relationship between transfer function and state space models

What are the poles of the system?

$$\begin{aligned}Y(s) &= C(sl - A)^{-1}BU(s) \\&= \frac{C\text{adj}(sl - A)B}{\det(sl - A)}U(s)\end{aligned}$$

Therefore, the poles are when

$$\det(sl - A) = 0,$$

i.e. when  $s$  is an eigenvalue of  $A$ .

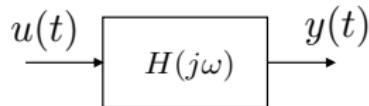
The poles of an LTI system and the eigenvalues of  $A$  are equivalent!

# Generalized Eigenvalues

Eigenvalues and eigenvectors can be defined for more general operators than just matrices.

## Example

Let  $h(t)$  be the impulse response of an LTI system with Fourier transform  $H(j\omega)$ .



Recall that if  $u(t) = e^{j\omega_0 t}$  then

$$\begin{aligned}y(t) &= |H(j\omega_0)| e^{j(\omega_0 t + \angle H(j\omega_0))} \\&= |H(j\omega_0)| e^{j\angle H(j\omega_0)} e^{j\omega_0 t}\end{aligned}$$

i.e. if a sinusoid goes in then the output will be a sinusoid of the same frequency but different magnitude and phase.

# Generalized Eigenvalues

Lemma

Let  $\mathcal{A}[u] = \int_0^{\top} h(t - \tau)u(\tau)d\tau$  then

$$(\lambda, x) = \left( H(j\omega) e^{j\angle H(j\omega)}, e^{j\omega t} \right)$$

is an eigenpair of  $\mathcal{A}$ .

Proof.

$$\mathcal{A}[e^{j\omega t}] = \left( |H(j\omega)| e^{j\angle H(j\omega)} \right) e^{j\omega t}.$$

Note that for  $\mathcal{A}$  there are an uncountable infinite number of eigenpairs. □

# Geometric and Algebraic Multiplicity

## Definition

Factor the characteristic polynomial as follows:

$$\chi_A(\lambda) = (\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_p)^{m_p}$$

$m_i$  is the algebraic multiplicity of eigenvalue  $\lambda_i$ .

## Definition

The geometric multiplicity of eigenvalue  $\lambda_i$  is defined as

$$q_i = \dim(\mathcal{N}(\lambda_i I - A)).$$

## Geometric and Algebraic Multiplicity: Example

Let  $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$  then

$$\chi_A(\lambda) = \det(\lambda I - A) = \det \begin{pmatrix} \lambda - 1 & 0 \\ 0 & \lambda - 1 \end{pmatrix} = (\lambda - 1)^2.$$

Therefore, the algebraic multiplicity of  $\lambda_1 = 1$  is  $m_1 = 2$ .  
What is the geometric multiplicity?

$$q_1 = \dim(\mathcal{N}(\lambda_1 I - A)) = \dim \left( \mathcal{N} \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \right) = \dim(\mathbb{R}^2) = 2.$$

Note that the eigenvectors  $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$  and  $\begin{pmatrix} 0 \\ \beta \end{pmatrix}$  are linearly independent!

## Geometric and Algebraic Multiplicity: Example

Let  $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$  then

$$\chi_A(\lambda) = \det \begin{pmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 1 \end{pmatrix} = (\lambda - 1)^2$$

so the algebraic multiplicity of  $\lambda_1 = 1$  is  $m_1 = 2$ .

The geometric multiplicity is

$$\begin{aligned} q_1 &= \dim(\mathcal{N}(I - A)) = \dim\left(\mathcal{N}\begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}\right) \\ &= \dim(\{x \in \mathbb{R}^2 \mid x_2 = 0\}) = 1 \neq m_1 \end{aligned}$$

What are the eigenvectors associated with  $A$ ?

$$(\lambda I - A)x = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = \begin{pmatrix} x_{12} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \Rightarrow x_{12} = 0$$

so  $\begin{pmatrix} \alpha \\ 0 \end{pmatrix}$  are the eigenvectors associated with  $\lambda_1$ . There are not two linearly independent eigenvectors.

# Linearly Independent Eigenvectors

In general we have,

## Lemma

*Let  $A \in \mathbb{C}^{n \times n}$ , then there are  $n$ -linearly independent eigenvectors if and only if*

$$\text{algebraic multiplicity} = \text{geometric multiplicity}$$

*for each eigenvalue of  $A$ .*

## Linearly Independent Eigenvectors: Proof

Proof.

First prove that if  $\lambda_i \neq \lambda_j$  then

$$\mathcal{N}(\lambda_i I - A) \cap \mathcal{N}(\lambda_j I - A) = \{0\}.$$

To prove the claim, suppose not, then

$$\exists x \neq 0 \text{ such that } x \in \mathcal{N}(\lambda_i I - A) \text{ and } x \in \mathcal{N}(\lambda_j I - A)$$

$$\begin{aligned} &\iff Ax = \lambda_i x \text{ and } Ax = \lambda_j x \\ &\iff \lambda_i x = \lambda_j x \\ &\iff (\lambda_i - \lambda_j)x = 0 \\ &\iff \lambda_i = \lambda_j \end{aligned}$$

which is a contradiction.

## Linearly Independent Eigenvectors: Proof

Note that the number of linearly independent eigenvectors associated with  $\lambda_i$  is the geometric multiplicity  $q_i$  since we can find  $q_i$  linearly independent vectors that span  $\mathcal{N}(\lambda_i - A)$ .

The previous claim shows that if  $x_i \in \mathcal{N}(\lambda_i I - A)$  then  $x_i \notin \mathcal{N}(\lambda_j I - A)$  which implies that there are  $\sum q_i$  linearly independent eigenvectors of  $A$ . Since  $\sum m_i = n$ , the lemma follows.

Note that if the eigenvalues are all distinct then  $m_i = 1$ . Also since  $1 \leq q_i \leq m_i$ , for each  $i$ , we must have that the algebraic multiplicity equals the geometric multiplicity.



## Linearly Independent Eigenvectors

Suppose that there are  $n$ -linearly independent eigenvectors (where some of the eigenvalues might be repeated), then we can write

$$\begin{aligned} (Ax_1 & \quad Ax_2 & \cdots & \quad Ax_n) = (\lambda_1 x_1 & \quad \lambda_2 x_2 & \cdots & \quad \lambda_n x_n) \\ \iff A \underbrace{(x_1 & \quad x_2 & \cdots & \quad x_n)}_S &= \underbrace{(x_1 & \quad x_2 & \cdots & \quad x_n)}_S \underbrace{\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & & \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}}_{\Lambda} \\ \iff AS &= S\Lambda \end{aligned}$$

Since the eigenvectors are linearly independent,  $S$  is invertible.  
Therefore

$$\begin{aligned} A &= S\Lambda S^{-1} \\ \iff \Lambda &= S^{-1}AS \end{aligned}$$

Therefore, we say that  $S$  diagonalizes  $A$ .

# Similarity Transformation

## Definition

Two matrices,  $A$  and  $B$  are said to be similar if  $\exists$  an invertible  $T$  such that

$$A = TBT^{-1}.$$

## Lemma

*Similar matrices have the same eigenvalues.*

## Proof.

Let  $(\lambda, x)$  be an eigenpair of  $A$ , then

$$Ax = \lambda x$$

$$\iff TBT^{-1}x = \lambda x$$

$$\iff BT^{-1}x = \lambda T^{-1}x$$

$$\iff By = \lambda y \text{ where } y = T^{-1}x$$

$\iff (\lambda, y)$  is an eigenpair of  $B$

## Section 2

### Jordan Form

## Jordan Form

What if the algebraic multiplicity does not equal the geometric multiplicity? (i.e.,  $q_i \neq m_i$  for some eigenvalue  $\lambda_i$  of  $A$ )?

Then we cannot diagonalize  $A$  using a similarity transformation.  
However we can “almost” diagonalize  $A$ .

The resulting “almost diagonal” matrix is called the Jordan form of  $A$ .

## Jordan Form, cont.

Suppose the algebraic multiplicity of  $\lambda_1$  is  $m_1 > 1$  but the geometric multiplicity is  $q_1 = 1$ .

Then  $\exists$  one linearly independent eigenvector  $x_1$  s.t.  $Ax_1 = \lambda_1 x_1$ .

Now form the following chain:

$$A\xi_{11} = \lambda_1 \xi_{11} + x_1$$

$$A\xi_{12} = \lambda_1 \xi_{12} + \xi_{12}$$

⋮

$$A\xi_{1,m_1} = \lambda_1 \xi_{1,m_1} + \xi_{1,(m_1-1)}$$

$\xi_{11}, \dots, \xi_{1,m_1}$  are called the “generalized eigenvectors” associated with  $x_1$ .

## Jordan Form, cont.

Note that we can write the generalized eigenvector equations as

$$A \begin{pmatrix} x_1 & \xi_{11} & \cdots & \xi_{1,m_1} \end{pmatrix} = \begin{pmatrix} x_1 & \xi_{11} & \cdots & \xi_{1,m_1} \end{pmatrix} \underbrace{\begin{pmatrix} \lambda_1 & 1 & \cdots & & \\ & \lambda_1 & 1 & 0 & \\ \cdots & & \lambda_1 & \cdots & \cdots \\ & 0 & & \ddots & 1 \\ & & \ddots & & \lambda_1 \end{pmatrix}}_{\text{This is called a Jordan block}}$$

### Lemma

If the geometric multiplicity of  $\lambda_i$  is  $q_i = 1$  then the associated  $m_1 - 1$  generalized eigenvectors are linearly independent of the other eigenvectors.

## Jordan Form, cont.

If  $1 < q_i < m_i$  then the problem is slightly more complicated.

There are precisely  $q_i$  linearly independent eigenvectors associated with  $\lambda_i$  and there will be  $q_i$  Jordan blocks associated with  $\lambda_i$ . What are the sizes of the Jordan blocks? For example, suppose  $m_i = 4$  and  $q_i = 2$ , the possible Jordan blocks are:

$$\begin{pmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{pmatrix} \text{ and } (\lambda_1) \text{ i.e., } \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 0 \\ 0 & 0 & 0 & \lambda_1 \end{pmatrix} \text{ or } \begin{pmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{pmatrix}$$

or

$$\begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix} \text{ and } \begin{pmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{pmatrix} \text{ i.e., } \begin{pmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 0 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{pmatrix}$$

Which option is correct?

## Jordan Form, cont.

To decide, generate the generalized eigenvector for each eigenvector and pick the linearly independent ones.

Example: Let

$$A = \begin{pmatrix} 1 & 1 & -1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Since  $\det(\lambda I - A) = (\lambda - 1)^4$  we have  $\lambda_1 = 1$  and  $m_1 = 4$ .

$$q_1 = \dim(\mathcal{N} \begin{pmatrix} 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}) = 2$$

since there are 2 linearly independent rows.

## Jordan Form, cont.

So there are two linearly independent eigenvectors:

$$(\lambda_1 I - A)x_1 = \begin{pmatrix} 0 & -1 & 1 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \\ x_{14} \end{pmatrix} = \begin{pmatrix} -x_{12} + x_{13} - x_{14} \\ -x_{14} \\ -x_{14} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$
$$\implies x_{14} = 0 \text{ and } -x_{12} + x_{13} - x_{14} = 0$$

so

$$x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \text{ is an eigenvector, and so is } x_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

## Jordan Form, cont.

Find the possible generalized eigenvector associated with eigenvector  $x_1$ :

$$A\xi_{11} = \xi_{11} + x_1 \Rightarrow (\lambda_1 I - A)\xi_{11} = -x_1$$

$$\text{i.e. } -\xi_{112} + \xi_{113} - \xi_{114} = 1 \quad \xi_{114} = 0$$

$$\xi_{112} = \xi_{113} + 1 \quad \text{so} \quad \xi_{11} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} \text{ is valid}$$

$$(\lambda_1 I - A)\xi_{12} = \xi_{12} \text{ so } \begin{pmatrix} -\xi_{122} + \xi_{123} - \xi_{124} \\ -\xi_{124} \\ -\xi_{124} \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \\ 0 \end{pmatrix} \leftarrow \text{can't use.}$$

## Jordan Form, cont.

**Note:** There are an infinite number of possibilities of generalized eigenvectors from each true eigenvector, but you can only pick ones that are linearly independent. This second eigenvector forms a linearly dependent subset of one of the real eigenvectors.

Therefore, one Jordan block is of size 2.

Also solve  $(\lambda_1 I - A)\xi_{21} = x_2$  i.e.

$$\begin{pmatrix} -\xi_{212} + \xi_{213} - \xi_{214} \\ -\xi_{214} \\ -\xi_{214} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix} \Rightarrow \xi_{214} = 1, \xi_{213} = \xi_{212} + 1$$

$$\text{so } \xi_{21} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 1 \end{pmatrix}.$$

## Jordan Form, cont.

In summary

$$A \underbrace{\begin{pmatrix} x_1 & \xi_{11} & x_2 & \xi_{21} \end{pmatrix}}_S = \underbrace{\begin{pmatrix} x_1 & \xi_{11} & x_2 & \xi_{21} \end{pmatrix}}_S \underbrace{\begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix}}_J$$

or

$$A = SJS^{-1}$$

$J$  is called the “Jordan” form of  $A$

If the eigenvalues are distinct or  $q_i = m_i$  for each  $i$  then  $J = \Lambda$  (is diagonal).

Otherwise  $J$  is block diagonal with Jordan blocks along the diagonal ( $q_i$  Jordan blocks for each eigenvalue).

## Jordan Form, cont.

Example: suppose there are 3 eigenvalues with  $\lambda_1 = 1, \lambda_2 = 2, \lambda_3 = 3$ , and  $m_1 = 1, m_2 = 2, m_3 = 3$ , and  $q_1 = 1, q_2 = 1, q_3 = 2$ . There are two possible Jordan forms:

$$\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & 1 & 0 \\ & & \lambda_2 & \\ & & & \lambda_3 & 1 \\ 0 & & & & \lambda_3 \\ & & & & & \lambda_3 \end{pmatrix} \text{ or } \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & 1 & 0 \\ & & \lambda_2 & \\ & & & \lambda_3 \\ 0 & & & & \lambda_3 \\ & & & & & \lambda_3 \end{pmatrix}$$

## Section 3

### Cayley-Hamilton Theorem

# Functions of Matrices

## Lemma

A square matrix can always be put in Jordan form.

This implies that we can always write

$$A = SJS^{-1}$$

This implies that

$$\begin{aligned} A^k &= \underbrace{AA \cdots A}_{k \text{ times}} \\ &= SJS^{-1}SJS^{-1} \cdots SJS^{-1} \\ &= SJ^kS^{-1} \end{aligned}$$

This is particularly simple if  $J = \Lambda$  since

$$A^k = S\Lambda^kS^{-1} \text{ where } \Lambda^k = \begin{pmatrix} \lambda_1^k & & 0 \\ & \ddots & \\ 0 & & \lambda_n^k \end{pmatrix}$$

## Functions of Matrices, cont.

For square matrices we can define analytic functions of matrices. Analytic functions are functions that can be expanded as a Taylor series, e.g.

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \dots$$

$$\cos(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \dots$$

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots$$

The corresponding Matrix function is defined in terms of its Taylor series, e.g.,

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \dots$$

$$\cos(A) = I - \frac{A^2}{2!} + \frac{A^4}{4!} - \frac{A^6}{6!} + \dots$$

$$\sin(A) = A - \frac{A^3}{3!} + \frac{A^5}{5!} - \frac{A^7}{7!} + \dots$$

# Cayley-Hamilton Theorem

Computing infinite series of matrices is a pain. Fortunately we have the following theorem:

**Theorem (Cayley-Hamilton Theorem)**

*Every matrix satisfies its own characteristic polynomial, i.e.*

$$\chi_A(A) = 0.$$

## Cayley-Hamilton Theorem, proof

The proof holds for all  $A$  but we will only prove the case when  $q_i = m_i$  for each  $\lambda_i$ . In this case  $A = S\Lambda S^{-1}$ . Note that for each eigenvalue  $\chi_A(\lambda_i) = 0$  since  $\chi_A(\lambda_i) = \det(\lambda_i I - A)$

Let

$$\chi_A(\lambda) = \lambda^n + a_{n-1}\lambda^{n-1} + \cdots + a_1\lambda + a_0$$

then

$$\begin{aligned}\chi_A(A) &= A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0I \\ &= S\Lambda^n S^{-1} + a_{n-1}S\Lambda^{n-1}S^{-1} + \cdots + a_1S\Lambda S^{-1} + a_0SS^{-1} \\ &= S(\Lambda^n + a_{n-1}\Lambda^{n-1} + \cdots + a_1\Lambda + a_0I)S^{-1}\end{aligned}$$

Note that the matrix

$$\Lambda^n + a_{n-1}\Lambda^{n-1} + \cdots + a_1\Lambda + a_0I$$

is diagonal with each element on the diagonal equal to

$$\lambda_i^n + a_{n-1}\lambda_i^{n-1} + \cdots + a_1\lambda_i + a_0 = 0.$$

Therefore

$$\Lambda^n + a_{n-1}\Lambda^{n-1} + \cdots + a_1\Lambda + a_0I = 0.$$

# Cayley-Hamilton Theorem, implications

Recall polynomial division:

$$\frac{f(x)}{q(x)} = a(x) + \frac{r(x)}{q(x)}$$
$$\Rightarrow \underbrace{f(x)}_{\text{degree } m} = \underbrace{a(x)}_{\text{degree } (m-n)} \overbrace{q(x)}^{\text{degree } n} + \underbrace{r(x)}_{\text{degree } < n}$$

Application to infinite series like  $e^x$  gives

$$\underbrace{e^x}_{\text{degree } \infty} = \underbrace{a(x)}_{\text{degree } \infty} \overbrace{\chi_A(x)}^{\text{degree } n} + \underbrace{r(x)}_{\text{degree } < n}$$

Since  $\chi_A(A) = 0$ ,

$$e^A = \underbrace{r(A)}_{\text{degree } < n} = \cdots b_{n-1} A^{n-1} + \cdots + b_1 A + b_0 I$$

Since  $e^{\lambda_i} = r(\lambda_i)$  the coefficients can be found from

$$e^{\lambda_i} = \cdots b_{n-1} \lambda^{n-1} + \cdots + b_1 \lambda + b_0.$$

## Cayley-Hamilton Theorem, example

Find  $e^A$  where  $A = \begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix}$ .

$$\det(\lambda I - A) = \lambda^2 + 2\lambda + 2 \Rightarrow \lambda_{1,2} = -1 \pm j$$

$$\Rightarrow e^A = b_1 A + b_0 I = \begin{pmatrix} 0 & b_1 \\ -2b_1 & -2b_1 \end{pmatrix} + \begin{pmatrix} b_0 & 0 \\ 0 & b_0 \end{pmatrix} = \begin{pmatrix} b_0 & b_1 \\ -2b_1 & -2b_1 - b_0 \end{pmatrix}$$

where  $b_0$  and  $b_1$  satisfy

$$e^{-1+j} = b_1(-1+j) + b_0$$

$$e^{-1-j} = b_1(-1-j) + b_0$$

$$\Rightarrow \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} = \begin{pmatrix} 1 & -1+j \\ 1 & -1-j \end{pmatrix}^{-1} \begin{pmatrix} e^{-1+j} \\ e^{-1-j} \end{pmatrix} = \begin{pmatrix} 0.5083 \\ 0.3096 \end{pmatrix}$$

$$\Rightarrow e^A = \begin{pmatrix} 0.5083 & 0.3096 \\ -0.6191 & -0.1108 \end{pmatrix}$$

## Section 4

### Self Adjoint Matrices

# Self Adjoint Matrices

## Definition

A matrix  $A \in \mathbb{C}^{n \times n}$  is said to be self adjoint (also called Hermitian) if

$$A = A^H.$$

## Lemma (Moon 6.2)

If  $A = A^H$  then the eigenvalues of  $A$  are real.

## Proof.

Let  $(\lambda, x)$  be a right eigen-pair, then

$$Ax = \lambda x, \text{ and } x^H A^H = \bar{\lambda} x^H.$$

Therefore

$$\begin{aligned}x^H A x &= \lambda x^H x, \text{ and } x^H A^H x = \bar{\lambda} x^H x \\ \implies \lambda x^H x &= \bar{\lambda} x^H x \implies \bar{\lambda} = \lambda, \\ \implies \lambda &\text{ is real.}\end{aligned}$$

# Self Adjoint Matrices

Lemma (Moon 6.3)

If  $A = A^H$  and the eigenvalues are distinct, then the eigenvectors are orthogonal.

Proof.

Let  $(\lambda_1, x_1)$  and  $(\lambda_2, x_2)$  be distinct eigenpairs, i.e.  $\lambda_1 \neq \lambda_2$ , then

$$x_2^H A x_1 = \lambda_1 x_2^H x_1$$

$$\text{and } x_2^H A^H x_1 = \lambda_2 x_2^H x_1$$

Therefore  $(\lambda_1 - \lambda_2)x_2^H x_1 = 0$ . Because  $\lambda_1 \neq \lambda_2$  we must have that

$$x_2^H x_1 = 0$$

which implies that  $x_1$  and  $x_2$  are orthogonal. □

Note the eigenvectors can always be chosen to be orthonormal.

# Self Adjoint Matrices

Theorem (Moon Theorem 6.2 (Special Theorem))

If  $A \in \mathbb{C}^{n \times n}$  is Hermitian, then  $q_i = m_i$  for each eigenvalue  $\lambda_i$ .

Corollary

If  $A = A^H$ , then  $\exists$  a unitary  $U$  and real diagonal  $\Lambda$  such that

$$A = U\Lambda U^H.$$

# Eigenvalues and Rank

## Lemma (Moon Lemma 6.5)

*Let  $A \in \mathbb{C}^{m \times m}$  be of rank  $r < m$ . Then at least  $m - r$  of the eigenvalues of  $A$  are equal to zero*

## Section 5

### Invariant Subspaces

# Invariant Subspaces

## Definition

Let  $A$  be a square matrix. If  $\mathbb{S} \subset \mathcal{R}(A)$  is such that  
 $x \in \mathbb{S} \implies Ax \in \mathbb{S}$  then  $\mathbb{S}$  is an invariant subspace of  $A$ .

## Example

An eigenvector forms an invariant subspace i.e.

$$\mathbb{S} = \{\alpha x \mid x \text{ is an eigenvector}\}$$

is invariant since  $\hat{x} \in \mathbb{S} \Rightarrow A\hat{x} = \lambda\hat{x} \in \mathbb{S}$ .

# Invariant Subspaces

## Example

The span of any subset of eigenvectors is invariant: Let  $x_1 \dots x_p$  be eigenvectors with associated eigenvalues  $\lambda_1 \dots \lambda_p$ .

Let

$$\mathbb{S} = \text{span}\{x_1 \dots x_p\}$$

then

$$\hat{x} \in \mathbb{S}$$

$$\implies \hat{x} = \alpha_1 x_1 + \dots + \alpha_p x_p$$

$$\implies A\hat{x} = \alpha_1 A x_1 + \dots + \alpha_p A x_p$$

$$\implies A\hat{x} = \alpha_1 \lambda_1 x_1 + \dots + \lambda_p \alpha_p x_p$$

$$\implies A\hat{x} \in \mathbb{S}.$$

# Applications to Differential Equations

Consider the differential equation  $\dot{x} = Ax$  with initial condition  $x(0) = x_0$ .

## Lemma

The solution is given by  $x(t) = e^{At}x_0$  where  
 $e^{At} = I + At + \frac{1}{2!}A^2t^2 + \frac{1}{3!}A^3t^3 + \dots$ .

## Proof.

Plug into equation

$$\begin{aligned}\frac{dx(t)}{dt} &= \frac{d}{dt}(e^{At})x_0 + e^{At}\frac{d}{dt}(x_0) = \frac{d}{dt}e^{At}x_0 \\ &= \frac{d}{dt}\left(I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \frac{A^4t^4}{4!} + \dots\right)x_0 \\ &= \left(A + A^2t + \frac{A^3t^2}{2!} + \frac{A^4t^3}{3!} + \dots\right)x_0 \\ &= A\left(I + At + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \dots\right)x_0 \\ &= Ae^{At}x_0 = Ax(t)\end{aligned}$$

so  $x(t) = e^{At}x_0$  satisfies  $\dot{x} = Ax$  with initial condition  $x_0$ .

# Applications to Differential Equations

## Lemma

If  $\mathbb{S}$  is an invariant subspace of  $A$  then  $\mathbb{S}$  is an invariant subspace of  $e^{At}$

## Proof.

Let  $x_0 \in \mathbb{S}$  then

$$\begin{aligned} Ax_0 \in \mathbb{S} &\implies tAx_0 \in \mathbb{S} \\ &\implies A(Ax_0) \in \mathbb{S} \\ &\implies \frac{A^2t^2}{2!}x_0 \in \mathbb{S} \end{aligned}$$

...

Therefore

$$x(t) = Ix_0 + Atx_0 + \frac{A^2t^2}{2!}x_0 + \frac{A^3t^3}{3!}x_0 + \cdots \in \mathbb{S}.$$

## Applications to Differential Equations: Example

Consider the differential equation

$$\dot{x} = \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} x.$$

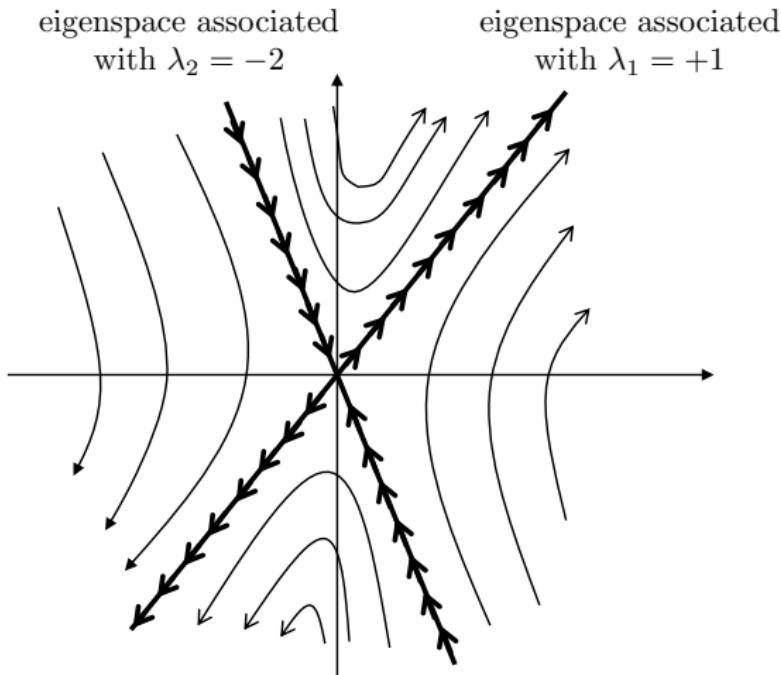
The eigenvalues of  $A$  are given by

$$\det(\lambda I - \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix}) = (\lambda - 1)(\lambda + 2) = 0 \text{ and so } \lambda_1 = 1 \text{ and } \lambda_2 = -2.$$

The associated eigenvector are

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } x_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

## Applications to Differential Equations: Example



- ▶ If the initial condition is on  $\text{span}\{x_1\}$ , then the solution remains on  $\text{span}\{x_1\}$ .
- ▶ If the initial condition is on  $\text{span}\{x_2\}$ , then the solution remains on  $\text{span}\{x_2\}$ .

## Applications to Difference Equations: Example

RWB: Change system to eigenvalues in unit circle. Show that eigenspaces are invariant. Provide example.

Consider the differential equation

$$x[k+1] = \begin{pmatrix} 0 & 1 \\ 2 & -1 \end{pmatrix} x[k].$$

Again the eigenvalues of  $A$  are  $\lambda_1 = 1$  and  $\lambda_2 = -2$ , and the eigenvectors are

$$x_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad x_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}.$$

# Section 6

## Quadratic Forms

# Quadratic Forms

## Definition

A real square matrix is symmetric if  $A^T = A$

## Definition

A real square matrix is skew-symmetric if  $A^T = -A$

# Quadratic Forms

## Lemma

Any real square matrix  $B \in \mathbb{R}^{n \times n}$  can be written as

$$B = B_s + B_{ss}$$

where  $B_s$  is symmetric and  $B_{ss}$  is skew-symmetric.

Proof.

$$B = \frac{B + B^T}{2} + \frac{B - B^T}{2} \triangleq B_s + B_{ss}$$

where

$$B_s^T = \left( \frac{B + B^T}{2} \right)^T = \frac{B^T - B}{2} = \frac{B + B^T}{2} = B_s$$

$$B_{ss}^T = \left( \frac{B - B^T}{2} \right)^T = \frac{B^T - B}{2} = -\left( \frac{B - B^T}{2} \right) = -B_{ss}$$

# Quadratic Forms

## Lemma

For any real square matrix  $A$  and for all  $y$

$$y^\top A y = y^\top A_s y$$

where  $A_s$  is the symmetric part of  $A$ .

Proof.

$$y^\top A y = y^\top A_s y + y^\top A_{ss} y$$

but

$$y^\top A_{ss} y = y^\top \left( \frac{A - A^\top}{2} \right) y = \frac{1}{2} y^\top A y - \frac{1}{2} y^\top A^\top y.$$

But since

$$y^\top A^\top y = (y^\top A^\top y)^\top = y^\top A y \implies y^\top A_{ss} y = 0.$$

# Quadratic Forms

## Definition

A quadratic form of a real square matrix  $A$  is  $Q_A(\mathbf{y}) = \mathbf{y}^\top A \mathbf{y}$ .

w.l.o.g.  $A$  can be assumed to be symmetric. If not, we can always limit our attention to the symmetric part of  $A$  since

$$\mathbf{y}^\top A \mathbf{y} = \mathbf{y}^\top A_s \mathbf{y}.$$

Quadratic forms show up in numerous places. For example, the pdf for a Gaussian random variable is

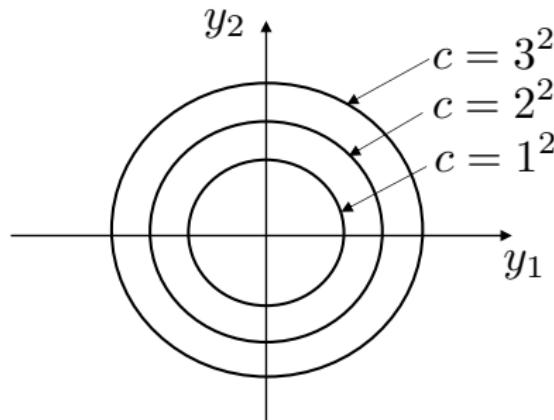
$$p(\mathbf{x}) = \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right).$$

# Quadratic Forms

## Example

Let

$$Q_A(\mathbf{y}) = \mathbf{y}^\top \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mathbf{y} = y_1^2 + y_2^2 = c$$



The level curves of  $Q_A(\mathbf{y})$  are circles of radius  $\sqrt{c}$ .

# Quadratic Forms

## Example

Consider the quadratic equation

$$f(x) = 2y_1^2 + 3y_1y_2 + 4y_2^2,$$

and note that

$$\begin{aligned} f(x) &= 2y_1^2 + 3y_1y_2 + 4y_2^2 \\ &= \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} 2y_1 + \frac{3}{2}y_2 \\ 4y_2 + \frac{3}{2}y_1 \end{pmatrix} \\ &= \begin{pmatrix} y_1 & y_2 \end{pmatrix} \begin{pmatrix} 2 & \frac{3}{2} \\ \frac{3}{2} & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \\ &= \mathbf{y}^\top A \mathbf{y} \end{aligned}$$

Any quadratic equation in  $n$  variables can be written in the form  $\mathbf{y}^\top A \mathbf{y}$ .

## Quadratic Forms

By the spectral theorem,  $A$  is diagonalizable. In other words, there exists an invertible  $U$  so that  $A = U\Lambda U^\top$ .

From Moon Lemma 6.2 the eigenvalues are real so we can order them as

$$\Lambda = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

with

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n.$$

# Quadratic Forms

## Lemma

*Level curves of the quadratic form*

$$Q_A(x - x_0) = (x - x_0)^\top A(x - x_0) = c$$

*are hyper-ellipsoids with the length of the axes given by  $\frac{1}{\sqrt{\lambda_i}}$ .*

## Proof.

Let  $z = U^\top y$  then

$$\begin{aligned} Q_A(y) &= y^\top A y = y^\top U \Lambda U^\top y = z^\top \Lambda z \\ &= (z_1 \quad \cdots \quad z_n) \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix} \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \\ &= \lambda_1 z_1^2 + \cdots + \lambda_n z_n^2 \end{aligned}$$

## Quadratic Forms

Note that in the variable  $z$ , the quadratic form is an ellipsoid:

$$Q_A(\mathbf{y}) = (\sqrt{\lambda_1})^2 z_1^2 + (\sqrt{\lambda_2})^2 z_2^2 + \cdots + (\sqrt{\lambda_n})^2 z_n^2 = 1$$

or

$$Q_A(\mathbf{y}) = \frac{z_1^2}{\left(\frac{1}{\sqrt{\lambda_1}}\right)^2} + \frac{z_2^2}{\left(\frac{1}{\sqrt{\lambda_2}}\right)^2} + \cdots + \frac{z_n^2}{\left(\frac{1}{\sqrt{\lambda_n}}\right)^2} = 1$$

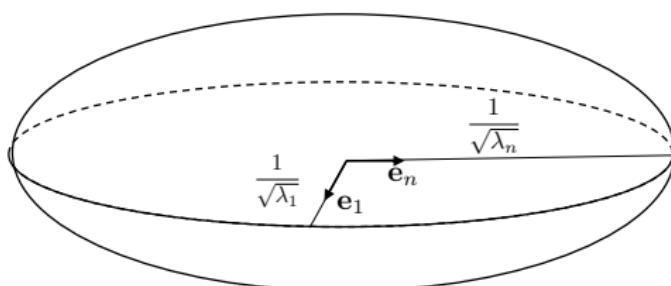
Either of these are the general equation for an ellipsoid with minor axis  $\mathbf{e}_1 = (1 \ 0 \ \cdots \ 0)^\top$  and major axis  $\mathbf{e}_n = (0 \ \cdots \ 0 \ 1)^\top$

# Quadratic Forms

In the original space, what is  $\mathbf{e}_1$ ?

Note that along  $\mathbf{e}_1$ , the stretching is

$$(\sqrt{\lambda_1})^2 z_1^2 = 1 \Rightarrow z_1 = \frac{1}{\sqrt{\lambda_1}}$$



$$\begin{aligned}\mathbf{e}_1 &= U^\top \mathbf{y} = \begin{pmatrix} \mathbf{u}_1^\top \\ \vdots \\ \mathbf{u}_n^\top \end{pmatrix} \mathbf{y} \\ &= \begin{pmatrix} \mathbf{u}_1^\top \mathbf{y} \\ \vdots \\ \mathbf{u}_n^\top \mathbf{y} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.\end{aligned}$$

Therefore  $\mathbf{y} = \mathbf{u}_1$  since  $U$  is orthogonal.  
i.e.  $\mathbf{u}_i = U\mathbf{e}_i$ .

## Quadratic Forms

Therefore the major axis is given by the eigenvector associated with the smallest eigenvalue, and the minor axis is given by the eigenvector associated with the largest eigenvalue.

**Question:** What is the geometric picture associated with

$$(x - x_0)^\top A(x - x_0) = c$$

where  $c$  is a constant and  $A$  is symmetric and positive definite?

**Answer:** An ellipsoid of radius  $\sqrt{c}$  centered at  $x_0$  with axes along the eigenvectors of  $A$  and stretching along each axis given by  $\frac{1}{\sqrt{\lambda_i}}$ .

## Quadratic Forms

**Question:** What if we would like to maximize

$$Q_A(y) = y^\top A y \text{ where } \|y\| = 1.$$

Which axis provides the most bang-for-the-buck?

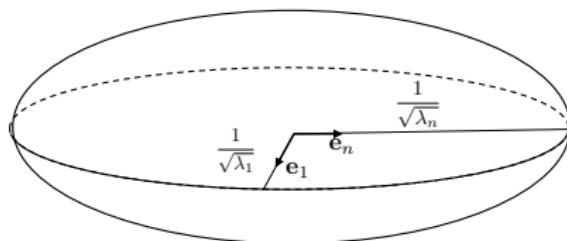
**Answer:** The major axis! i.e. the axis associated with the largest eigenvalue.

# Quadratic Forms

Rather than drawing

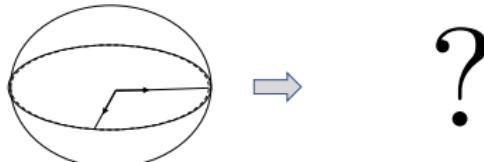
$$\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_n z_n^2 = 1$$

which is



lets draw the mapping of the unit circle through  $\mathbf{y}^\top A \mathbf{y}$  i.e.

$$\{\|\mathbf{y}\| = 1\} \xrightarrow{Q_A(\mathbf{y})} \{\mathbf{y}^\top A \mathbf{y}\}$$



## Quadratic Forms

If  $A = A^\top$  then  $A = U\Lambda U^\top$  where  $U$  is orthogonal, i.e.,  $UU^\top = U^\top U = I$ . Then

$$\max_{\|\mathbf{y}\|=1} \mathbf{y}^\top A \mathbf{y} = \max_{\|\mathbf{y}\|=1} \mathbf{y}^\top U \Lambda U^\top \mathbf{y}.$$

Let  $\mathbf{z} = U^\top \mathbf{y}$  and note that  $\|\mathbf{z}\| = \|U^\top \mathbf{y}\| = \|\mathbf{y}\|$  since  $U$  is orthogonal. Then

$$\begin{aligned} \max_{\|\mathbf{y}\|=1} \mathbf{y}^\top U \Lambda U^\top \mathbf{y} &= \max_{\|\mathbf{z}\|=1} \mathbf{z}^\top \Lambda \mathbf{z} \\ &= \max_{\|\mathbf{z}\|=1} (\lambda_1 z_1^2 + \lambda_2 z_2^2 + \cdots + \lambda_n z_n^2) \end{aligned}$$

where  $\Lambda$  is arranged such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

# Quadratic Forms

The maximum is therefore

$$\mathbf{z}^* = \begin{pmatrix} 1 & 0 & \vdots & 0 \end{pmatrix}^\top$$

where it is clear that  $\|\mathbf{z}\| = 1$ .

Furthermore

$$\max_{\|\mathbf{z}\|=1} \mathbf{z}^\top \Lambda \mathbf{z} = \lambda_1,$$

which implies that

$$\mathbf{y}^* = U\mathbf{z}^* = (\mathbf{u}_1 \quad \cdots \quad \mathbf{u}_n) \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \mathbf{u}_1.$$

## Quadratic Forms

This mapping also forms an ellipsoid but with a different effect.  
Let  $\mathbf{y} = \mathbf{u}_1 \implies \|\mathbf{y}\| = 1$  to get

$$Q_A(\mathbf{y}) = \lambda_1$$

**Question:** Is it possible to pick a  $\hat{\mathbf{y}}$  where  $\|\hat{\mathbf{y}}\| = 1$  such that

$$Q_A(\hat{\mathbf{y}}) > Q_A(\mathbf{u}_1)?$$

**Answer:** No.

## Quadratic Forms

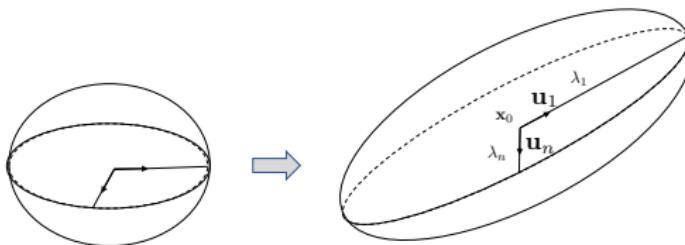
Explanation: Recall that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$  and

$$\|\hat{\mathbf{y}}\|^2 = y_1^2 + \dots + y_n^2 = 1.$$

Therefore

$$\begin{aligned} Q_A(\hat{\mathbf{y}}) &= \lambda_1 y_1^2 + \dots + \lambda_n y_n^2 \\ &\leq \lambda_1 y_1^2 + \lambda_1 y_2^2 + \dots + \lambda_1 y_n^2 \\ &= \lambda_1 \|\hat{\mathbf{y}}\|^2 \\ &= Q_A(\mathbf{u}_1) \end{aligned}$$

So the mapping of the unit circle looks like



# Quadratic Forms

We have essentially proved the following theorem:

**Theorem (Moon Theorem 6.5)**

*For a positive semi-definite Hermitian matrix  $A$ , the maximum*

$$\lambda_1 = \max_{\|\mathbf{x}\|_2=1} \mathbf{x}^H A \mathbf{x}$$

*where  $\lambda_1$  is the largest eigenvalue of  $A$ , and the maximizing  $\mathbf{x}$  is  $\mathbf{x} = \mathbf{u}_1$ , the associated eigenvector.*

*Furthermore if we maximize  $\mathbf{x}^H A \mathbf{x}$  subject to the constraints*

$$\langle \mathbf{x}, \mathbf{u}_i \rangle = 0 \quad i = 1, \dots, k-1,$$

$$\|\mathbf{x}\|_2 = 1$$

*then the maximum is  $\lambda_k$  and  $\mathbf{x}_{\max} = \mathbf{u}_k$ .*

## Quadratic Forms

Note that if  $A$  is positive semi-definite Hermitian then

$$\|A\|_2 = \sup_{\|\mathbf{x}\|_2 \neq 0} \frac{\|A\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \max_{\|\mathbf{x}\|_2=1} \sqrt{\mathbf{x}^H A^H A \mathbf{x}} = \sqrt{\lambda_1 \mathbf{u}_1^H \mathbf{u}_1} = \sqrt{\lambda_1}$$

where  $\lambda_1$  is the largest eigenvalue of  $A^H A$ .

More generally,

$$R(\mathbf{x}) = \frac{\mathbf{x}^\top A \mathbf{x}}{\mathbf{x}^\top \mathbf{x}}$$

is called a Rayleigh quotient and

$$\max_{\|\mathbf{x}\| \neq 0} R(\mathbf{x}) = \lambda_1$$

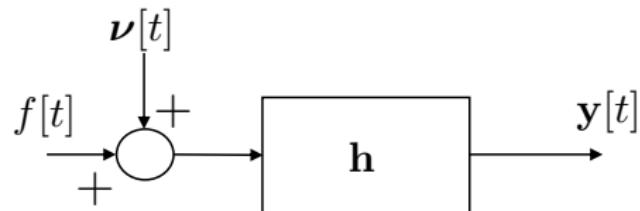
$$\min_{\|\mathbf{x}\| \neq 0} R(\mathbf{x}) = \lambda_n.$$

## Section 7

### Eigenfilters

# Eigenfilters for Random Signals

Problem: Given



LTI  
FIR filter

where  $\nu$  is white noise with variance  $\sigma^2$ , and  $f$  is a stationary, zero-mean random process.

Find  $h$  to maximize the signal-to-noise ratio.

# Eigenfilters for Random Signals

Let

$$\mathbf{f}(t) = \begin{pmatrix} f(t) \\ f(t-1) \\ \vdots \\ f(t-(m-1)) \end{pmatrix}$$

then

$$y(t) = \mathbf{h}^H \mathbf{f}(t).$$

The output power due to the signal  $\mathbf{f}$  is

$$\begin{aligned} P_0 &= E|y(t)|^2 = E|\mathbf{h}^H \mathbf{f}(t)|^2 = E\{\mathbf{h}^H \mathbf{f}(t) \mathbf{h}^H \mathbf{f}(t)\} \\ &= E\{\mathbf{h}^H \mathbf{f}(t) \mathbf{f}^H(t) \mathbf{h}\} = \mathbf{h}^H E\{\mathbf{f}(t) \mathbf{f}^H(t)\} \mathbf{h} \\ &= \mathbf{h}^H R \mathbf{h} \end{aligned}$$

where  $R = E\{\mathbf{f}(t) \mathbf{f}^H(t)\}$

# Eigenfilters for Random Signals

Let

$$\nu(t) = \begin{pmatrix} v(t) \\ v(t-1) \\ \vdots \\ v(t-m+1) \end{pmatrix}$$

Then the output due to the noise is

$$\mathbf{h}\nu(t)$$

and the average noise power is

$$N_0 = E\{\mathbf{h}^H \nu(t) \nu^H(t) \mathbf{h}\} = \sigma^2 \mathbf{h}^H \mathbf{h}$$

# Eigenfilters for Random Signals

The signal-to-noise ratio is

$$\begin{aligned} SNR &= \frac{P_0}{N_0} \\ &= \frac{1}{\sigma^2} \cdot \underbrace{\frac{\mathbf{h}^H R \mathbf{h}}{\sigma^2 \mathbf{h}^H \mathbf{h}}}_{\text{Rayleigh quotient}} \end{aligned}$$

Therefore

$$SNR_{max} = \frac{\lambda_1}{\sigma^2}$$

where  $\lambda_1$  is the largest eigenvalue of  $R$  and  $\mathbf{h} = q_1$  the largest eigenvector of  $R$ .

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 7.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Singular Value Decomposition

Pseudo Inverse and the SVD

SVD and Numerically Sensitive Problems

Rank Reducing Approximations

Applications

## Section 1

# Singular Value Decomposition

# Singular Value Decomposition

## Theorem (Moon Theorem 7.1)

*Every matrix  $A \in \mathbb{C}^{m \times n}$  can be factored as  $A = U\Sigma V^H$  where  $U \in \mathbb{C}^{m \times m}$  and  $V \in \mathbb{C}^{n \times n}$  are unitary and  $\Sigma \in \mathbb{R}^{m \times n}$  is diagonal with diagonal elements  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ .*

The diagonal elements are called the singular values of  $A$ . If  $A$  is real then  $U$  and  $V$  are real and orthogonal.

## Singular Value Decomposition, Proof

Note that the  $A^H A$  is Hermitian, and positive definite because  $x^H A^H A x = \|Ax\|^2 \geq 0 \quad \forall x \in \mathbb{C}^n$ .

So, from Chapter 6 we know that the eigenvalues of  $A^H A$  are real with  $m_i = q_i$  for each  $\lambda_i$ .

Let  $(\lambda_i, \mathbf{v}_i)$  be an eigenpairs of  $A^H A$  then

$$A^H A V = V \Lambda \quad V\text{-unitary}$$

where

$$V = (\mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n), \quad \Lambda = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix},$$

with

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n.$$

## Singular Value Decomposition, Proof

Since the  $\text{rank}(A^H A) \leq \min(m, n) = p$ , then number of non-zero eigenvalues is  $r \leq p$ .

For  $1 \leq i \leq r$  let  $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}$ .

Then

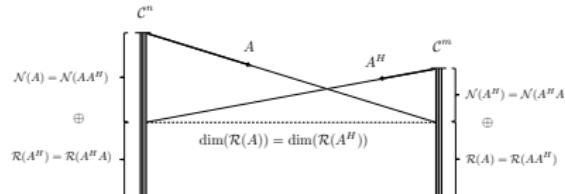
$$\begin{aligned}\langle \mathbf{u}_i, \mathbf{u}_j \rangle &= \left\langle \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}, \frac{A\mathbf{v}_j}{\sqrt{\lambda_j}} \right\rangle = \frac{1}{\sqrt{\lambda_i \lambda_j}} \mathbf{v}_i^H A^H A \mathbf{v}_j \\ &= \frac{\lambda_j}{\sqrt{\lambda_i \lambda_j}} \mathbf{v}_i^H \mathbf{v}_j = \delta_{ij}\end{aligned}$$

Use Gram-Schmidt to extend  $\mathbf{u}_1, \dots, \mathbf{u}_r$  to  $[\mathbf{u}_1, \dots, \mathbf{u}_m]$  such that  $U = [\mathbf{u}_1, \dots, \mathbf{u}_m]$  is unitary.

# Singular Value Decomposition, Proof

## Lemma

If  $(\lambda_i, \mathbf{v}_i)$  is an eigenpair of  $A^H A$ , then  $\mathbf{u}_i = \frac{A\mathbf{v}_i}{\sqrt{\lambda_i}}$  are eigenvectors of  $AA^H$ .



## Proof.

Note that since  $\mathbf{u}_i = \frac{1}{\sqrt{\lambda_i}} A\mathbf{v}_i \quad i = 1, \dots, p$  then

$$\mathbf{u}_i \in \mathcal{R}(A) \quad i = 1, \dots, p$$

$$\implies \mathbf{u}_i \in \mathcal{N}(A^H) \quad i = p+1, \dots, m$$

$$\implies \mathbf{u}_i \in \mathcal{N}(AA^H) \quad i = p+1, \dots, m$$

$$\implies AA^H \mathbf{u}_i = 0 \cdot \mathbf{u}_i = 0$$

$$\implies (0, \mathbf{u}_i) \text{ is an eigenpair of } AA^H \quad i = p+1, \dots, m$$

## Singular Value Decomposition, Proof

Now lets look at

$$U^H A V = \begin{pmatrix} \mathbf{u}_1^H \\ \vdots \\ \mathbf{u}_m^H \end{pmatrix} A \begin{pmatrix} \mathbf{v}_1 & \cdots & \mathbf{v}_n \end{pmatrix} = \begin{pmatrix} \mathbf{u}_1^H A \mathbf{v}_1 & \cdots & \mathbf{u}_1^H A \mathbf{v}_n \\ \vdots & \ddots & \vdots \\ \mathbf{u}_m^H A \mathbf{v}_1 & \cdots & \mathbf{u}_m^H \mathbf{v}_n \end{pmatrix}.$$

The  $(i, j)^{th}$  element of  $U^H A V$  is  $\mathbf{u}_i^H A \mathbf{v}_j$ .

If  $i \leq p$  then

$$\begin{aligned} \mathbf{u}_i^H A \mathbf{v}_j &= \frac{1}{\sqrt{\lambda_i}} \mathbf{v}_i^H A^H A \mathbf{v}_j \\ &= \frac{\lambda_j}{\sqrt{\lambda_i}} \mathbf{v}_i^H \mathbf{v}_j = \sqrt{\lambda_j} \delta_{ij} \end{aligned}$$

## Singular Value Decomposition, Proof

If  $i > p$ , then

$$\begin{aligned}\mathbf{u}_i \in \mathcal{N}(A^H) &\Rightarrow A^H \mathbf{u}_i = 0 \\ &\Rightarrow \mathbf{u}_i^H A = 0 \\ &\Rightarrow \mathbf{u}_i^H A \mathbf{v}_j = 0\end{aligned}$$

Therefore

$$U^H A V = \Sigma$$

where  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p)$  is real and diagonal, where  $\sigma_j = 0$  when  $j > p$ . Therefore

$$A = U \Sigma V^H$$

as required. □

# Singular Value Decomposition

Note that the singular values of  $A$  are the square root of the eigenvalues of  $A^H A$  and  $AA^H$ .

Also note that we can write

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix}$$

where

$$\Sigma_1 = \underbrace{\text{diag}(\sigma_1, \dots, \sigma_p)}_{\mathbb{R}^{r \times r}}$$

$$\Sigma_2 = 0$$

# Singular Value Decomposition

Then

$$\begin{aligned} A &= (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} \\ &= \underbrace{U_1}_{m \times p} \underbrace{\Sigma_1}_{p \times p} \underbrace{V_1^H}_{n \times p} \quad \leftarrow \text{alternate form of SVD} \\ &= \sum_{i=1}^p \sigma_i \mathbf{u}_i \mathbf{v}_i^H \quad \leftarrow \text{alternate form of SVD} \end{aligned}$$

where  $\mathbf{u}_i$ 's are orthonormal and  $\mathbf{v}_i$ 's are orthonormal.

# Singular Value Decomposition and Matrix Norm

Note that

$$\begin{aligned}\|A\|_2 &= \sup_{\|x\|_2=1} \|Ax\|_2 = \sup_{\|x\|_2=1} \sqrt{x^H A^H A x} \\&= \sup_{\|x\|_2=1} \sqrt{x^H V_1 \Sigma_1 U_1^H U_1 \Sigma_1 V_1^H x} \\&= \sup_{\|x\|_2=1} \sqrt{x^H V_1 \Sigma_1^2 V_1^H x} \\&= \sup_{\|x\|_2=1} \sqrt{\begin{pmatrix} x^H \mathbf{v}_1 & \cdots & x^H \mathbf{v}_r \end{pmatrix} \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^H x \\ \vdots \\ \mathbf{v}_p^H x \end{pmatrix}} \\&= \sigma_1,\end{aligned}$$

where the minimizer is  $x = \mathbf{v}_1$ .

# Singular Value Decomposition and Rank

## Lemma

If  $A \in \mathbb{C}^{m \times n}$ , then  $\text{rank}(A) = p$  where  $p$  is the number of non-zero singular values.

## Proof.

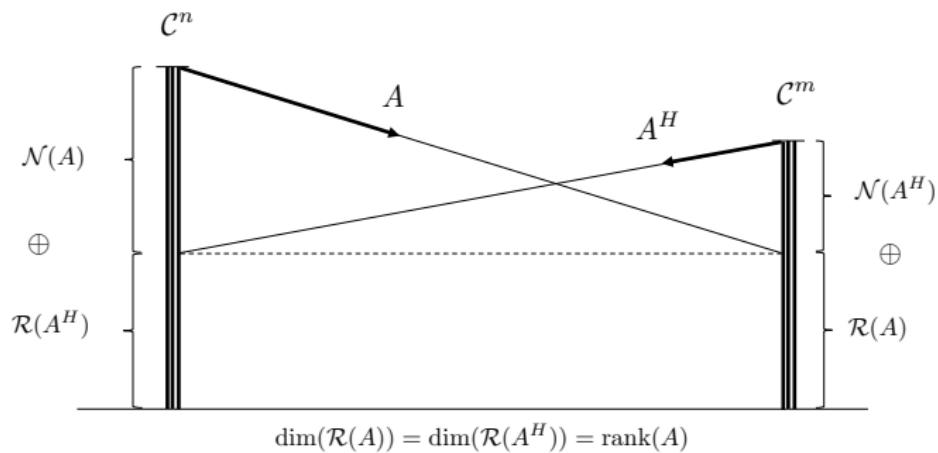
$$\text{rank}(A) = \text{rank}(U\Sigma V^H) = \text{rank}(\Sigma)$$

since  $U$  and  $V$  are both full rank. Clearly  $\text{rank}(\Sigma) = p$ .

□

# Singular Value Decomposition and Fundamental Subspaces

Fundamental subspace diagram:



**Question:** What information does the SVD provide?

**Answer:** The SVD completely characterizes all of the spaces.

# Singular Value Decomposition and Fundamental Subspaces

Given that

$$A = (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 \Sigma_1 V_1^H.$$

Let  $y \in \mathcal{R}(A)$ , then  $\exists x \in \mathbb{C}^n$  such that  $y = Ax$ . Which implies that

$$\begin{aligned} y &= U_1 \Sigma_1 V_1^H x \\ &= U_1 z \text{ where } z = \Sigma_1 V_1^H x \\ &= [\mathbf{u}_1 \cdots \mathbf{u}_p] \begin{pmatrix} z_1 \\ \vdots \\ z_p \end{pmatrix} = \mathbf{u}_1 z_1 + \cdots + \mathbf{u}_p z_p \end{aligned}$$

$$\implies y \in \text{span}\{\mathbf{u}_1, \dots, \mathbf{u}_p\}$$

$$\implies \boxed{\mathcal{R}(A) = \text{span}(U_1)}$$

# Singular Value Decomposition and Fundamental Subspaces

Since the columns of  $U_2$  are orthonormal to  $U_1$  and  $\text{span}(U) = \mathbb{C}^m$  and  $\mathcal{R}(A) \oplus \mathcal{N}(A^H) = \mathbb{C}^m$  we must have that

$$\mathcal{N}(A^H) = \text{span}(U_2)$$

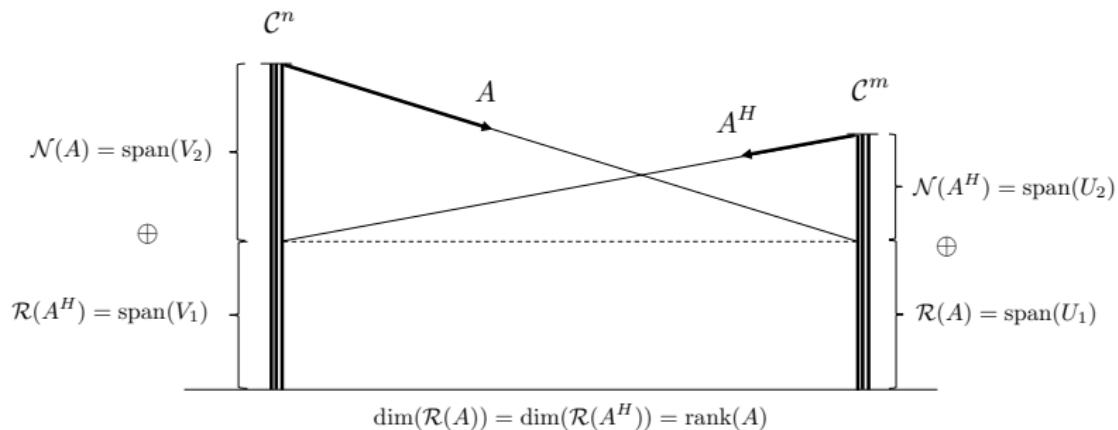
A similar argument shows that

$$\mathcal{R}(A^H) = \text{span}(V_1)$$

$$\mathcal{N}(A) = \text{span}(V_2)$$

# Singular Value Decomposition and Fundamental Subspaces

Therefore, the fundamental subspace diagram becomes



## Section 2

### Pseudo Inverse and the SVD

## Pseudo Inverses of $A$

Least squares solution to  $Ax = b$  (i.e.  $\min \|Ax - b\|_2$ ) where  $A$ -tall is

$$\hat{x} = (A^H A)^{-1} A^H b \triangleq A^\dagger b.$$

Minimum norm solution to  $Ax = b$  (i.e.  $\min \|x\|$  for  $Ax = b$ ) where  $A$ -fat is

$$x = A^H (A A^H)^{-1} b \triangleq A^\dagger b.$$

How does the SVD help compute the pseudo inverse. We will consider both when  $A$  is full rank, and when  $A$  is not full rank.

## SVD and Least Squared: Full Rank $A$

Assume  $A \in \mathbb{C}^{m \times n}$  is tall, i.e.,  $m > n$ , and that  $\text{rank}(A) = n$ . Then

$$A = (U_1 \quad U_2) \begin{pmatrix} \Sigma \\ 0 \end{pmatrix} V^H = U_1 \Sigma V^H$$

where  $U_1 \in \mathbb{C}^{m \times n}$ ,  $\Sigma \in \mathbb{R}^{n \times n}$ , and  $V \in \mathbb{C}^{n \times n}$ .

Then

$$\begin{aligned} (A^H A)^{-1} A^H &= (V \Sigma U_1^H U_1 \Sigma V^H)^{-1} V \Sigma U_1^H \\ &= (V \Sigma^2 V^H)^{-1} V \Sigma U_1^H \\ &= V \Sigma^{-2} V^H V \Sigma U_1^H \\ &= V \Sigma^{-1} U_1^H \end{aligned}$$

where  $\Sigma^{-1} = \text{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_n})$ .

## SVD and min Norm: Full Rank $A$

Assume  $A \in \mathbb{C}^{m \times n}$  is fat, i.e.,  $m < n$ , and that  $\text{rank}(A) = m$ . Then

$$\begin{aligned} A &= U \begin{pmatrix} \Sigma & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} \\ &= U \Sigma V_1^H \end{aligned}$$

where  $U \in \mathbb{C}^{m \times m}$ ,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_m)$ ,  $V_1 \in \mathbb{C}^{n \times m}$ .

Then

$$\begin{aligned} A^H (AA^H)^{-1} &= V_1 \Sigma U^H (U \Sigma V_1^H V_1 \Sigma U^H)^{-1} \\ &= V_1 \Sigma U^H (U \Sigma^2 U^H)^{-1} \\ &= V_1 \Sigma U^H U \Sigma^{-2} U^H \\ &= V_1 \Sigma^{-1} U^H \end{aligned}$$

where  $\Sigma^{-1} = \text{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_m})$

## SVD and Pseudo Inverse: Not Full Rank $A$

Assume  $A \in \mathbb{C}^{m \times n}$  and that  $\text{rank}(A) = p < \min(m, n)$ . Then

$$A = (U_1 \quad U_2) \begin{pmatrix} \Sigma & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} = U_1 \Sigma V_1^H$$

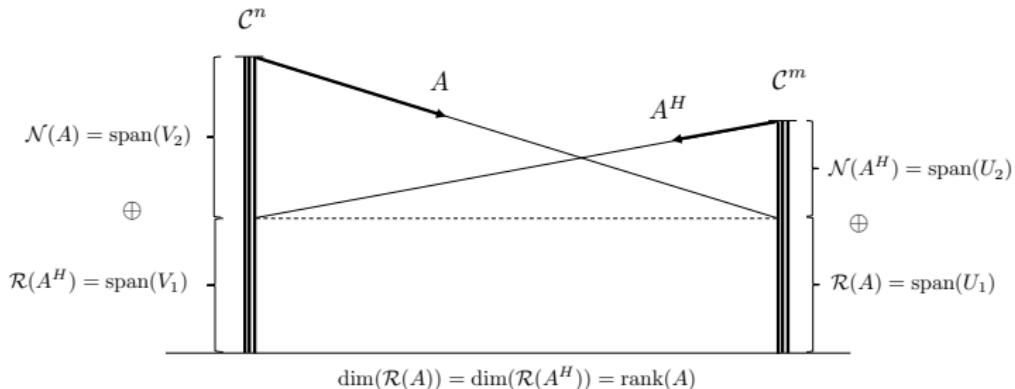
where  $U_1 \in \mathbb{C}^{m \times p}$ ,  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{p \times p}$ ,  $V_1 \in \mathbb{C}^{n \times p}$ . Consider the least squares problem

$$\begin{aligned}\hat{x} &= (A^H A)^{-1} A^H b \\ &= (V_1 \Sigma_1 U_1^H U_1 \Sigma_1 V_1^H)^{-1} V_1 \Sigma_1 U_1^H b \\ &= (V_1 \Sigma_1^2 V_1^H)^{-1} V_1 \Sigma_1 U_1^H b \\ &= V_1 \Sigma_1^{-2} V_1^H V_1 \Sigma_1 U_1^H b \\ &= V_1 \Sigma_1^{-1} U_1^H b\end{aligned}$$

where  $\Sigma_1 = \text{diag}(\frac{1}{\sigma_1}, \dots, \frac{1}{\sigma_p})$ .

## SVD and Pseudo Inverse: Not Full Rank $A$

So we can compute it, but what did we do? How do we interpret the solution since the inverse of  $A^H A$  does not exist?



Find a solution to  $Ax = b$  where  $b \in \mathcal{R}(A)$ . But  $\mathcal{N}(A) \neq \{0\}$  implies that there are more than one solution.

Therefore, find the minimum norm  $x$  that minimizes  $\|Ax - b\|_2$ .

## SVD and Pseudo Inverse: Not Full Rank $A$

Note the following:

$$\underbrace{U_1}_{m \times p} : \mathbb{C}^p \rightarrow \mathcal{R}(A) \subset \mathbb{C}^m$$

so that

$$U_1^* = U_1^H : \mathbb{C}^m \rightarrow \mathbb{C}^p.$$

Also,

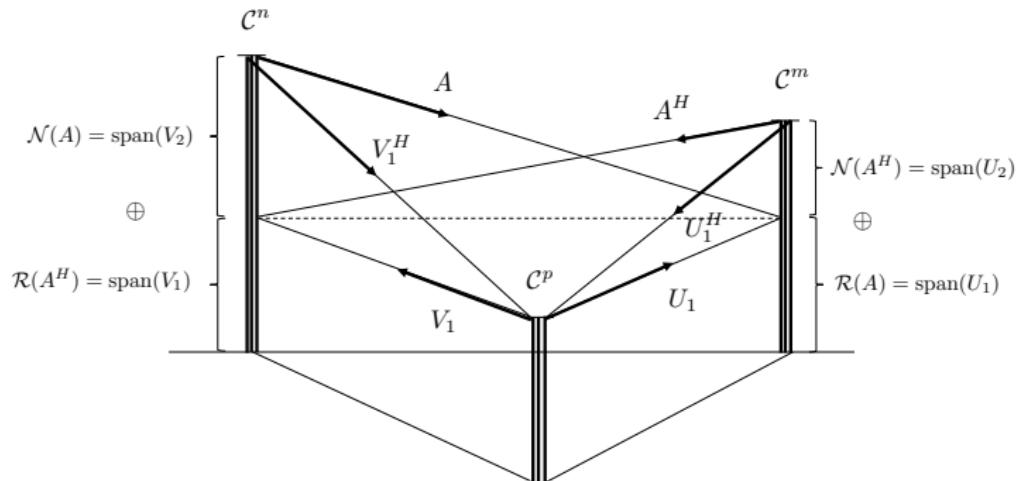
$$\underbrace{V_1}_{n \times p} : \mathbb{C}^p \rightarrow \mathcal{R}(A^H) \subset \mathbb{C}^n$$

so that

$$V_1^H : \mathbb{C}^n \rightarrow \mathbb{C}^p.$$

# SVD and Pseudo Inverse: Not Full Rank $A$

So we have the following:

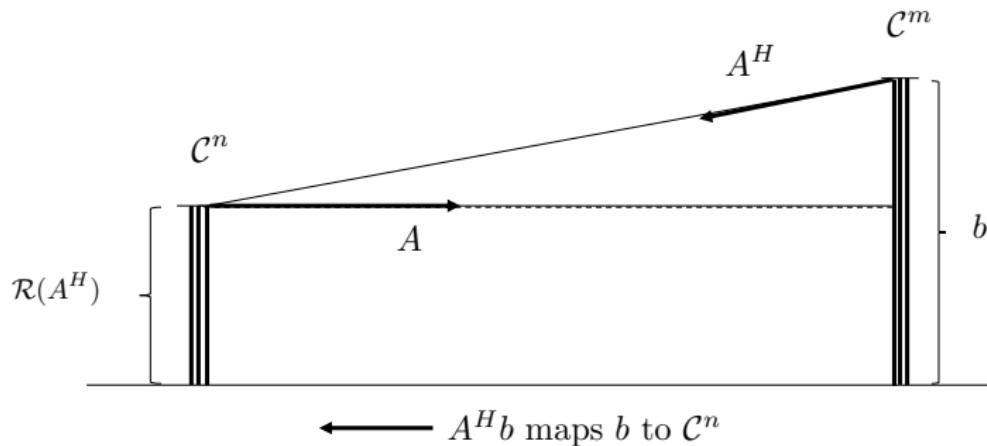


Since  $\text{rank}(A) = p$  we can only take inverses in  $\mathbb{C}^p$ . Therefore instead of solving  $Ax = b$  directly in  $\mathbb{C}^n$  and  $\mathbb{C}^m$  we go indirectly through  $\mathbb{C}^p$ .

# SVD and Pseudo Inverse: Not Full Rank $A$

## Step 1: Least Squares

Recall that to solve  $\min \|Ax - b\|_2$  when  $A$  is full rank:

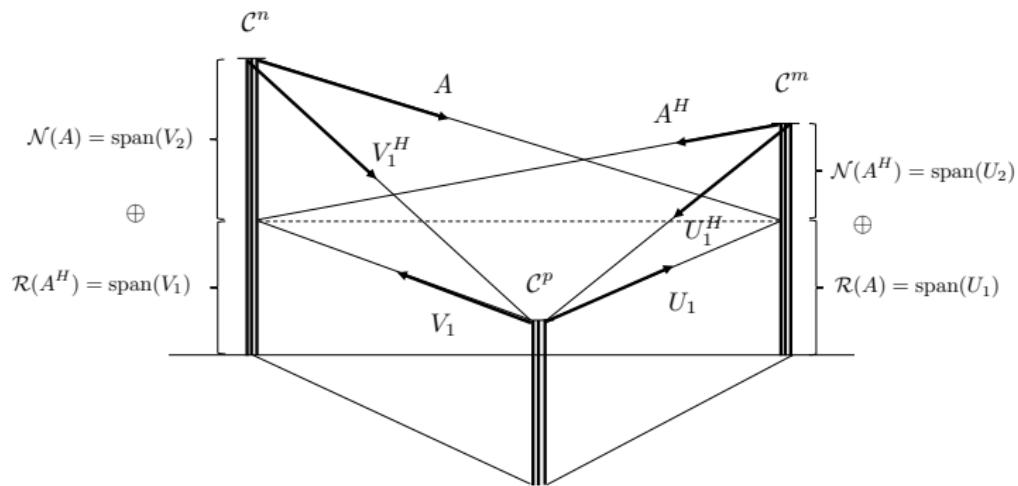


where we can invert things, i.e.

$$\begin{aligned} A^H A x &= A^H b \\ \implies \hat{x} &= (A^H A)^{-1} A^H b. \end{aligned}$$

# SVD and Pseudo Inverse: Not Full Rank $A$

So we have the following:



Now instead of  $A^H$  we use  $U_1^H$  to map to  $\mathbb{C}^p$ , i.e., given

$$Ax = b$$

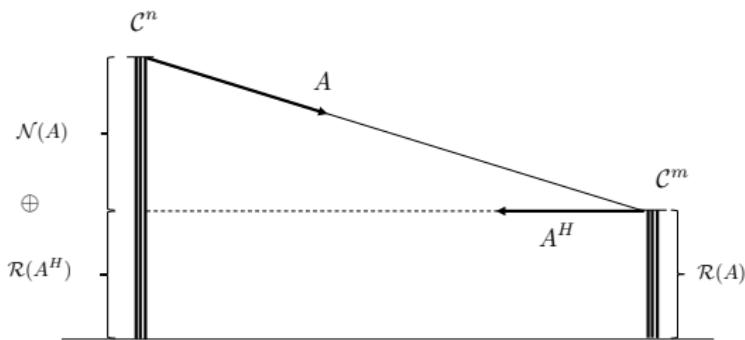
map to  $\mathbb{C}^p$  using  $U_1^H$  to get:

$$U_1^H Ax = U_1^H b \quad \in \mathbb{C}^p.$$

# SVD and Pseudo Inverse: Not Full Rank $A$

## Step 2: Minimum Norm

Recall that to min  $\|x\|$  such that  $Ax = b$ ,  $A$ -full rank,



to minimize  $\|x\|$  we zero out the part that is in the null space of  $A$ , i.e. let

$$x = A^H z \text{ where } z \in \mathbb{C}^m$$

then

$$AA^H z = b \quad \Rightarrow \quad z = (AA^H)^{-1}b$$

so that

$$\hat{x} = A^H (AA^H)^{-1}b.$$

# SVD and Pseudo Inverse: Not Full Rank A

In fact,

In our case, again pick  $x$  to zero  
the portion in the null space of  
 $A$ . Let

$$x = V_1 z \quad \text{where} \quad z \in \mathbb{C}^p$$

so that

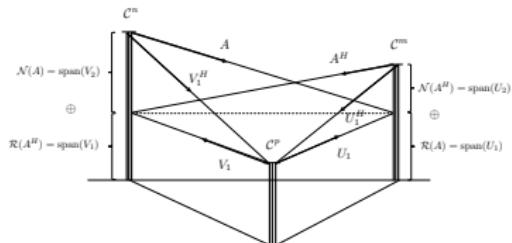
$$U_1^H A x = (U_1 A V_1) z = U_1^H b.$$

Note that

$$U_1 A V_1 : \mathbb{C}^p \rightarrow \mathbb{C}^p.$$

$U_1^H A V_1 = U_1^H U_1 \Sigma_1 V_1^H V_1 = \Sigma_1$ .  
so we have

$$\begin{aligned}\Sigma_1 z &= U_1^H b \\ \implies z &= \Sigma_1^{-1} U_1^H b \\ \implies \hat{x} &= V_1 \Sigma_1^{-1} U_1^H b\end{aligned}$$



## Section 3

# SVD and Numerically Sensitive Problems

## Numerically Sensitive Problems

Suppose that we would like to solve

$$Ax = b$$

where  $A \in \mathbb{R}^{n \times n}$  and  $\text{rank}(A) = n$  but the condition number  $\mathcal{K}(A)$  is large. Let  $A = U\Sigma V^H$ , then

$$\begin{aligned} A^{-1} &= V\Sigma^{-1}U^H \\ &= \sum_{j=1}^n \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} \end{aligned}$$

so the solution to  $Ax = b$  is

$$x = A^{-1}b = \sum_{j=1}^n \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j}.$$

## Numerically Sensitive Problems

Recall that  $\mathcal{K}(A) = \|A\| \|A^{-1}\|$  where  $\|A\| = \sigma_{\max}(A)$  and  $\|A^{-1}\| = \frac{1}{\min_{\|x\|} \|Ax\|} = \frac{1}{\sigma_{\min}(A)}$ . Therefore

$$\mathcal{K}(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

Therefore a large  $\mathcal{K}(A)$  implies there is significant difference between the largest and smallest singular values.

## Numerically Sensitive Problems

For example  $\sigma_{\min}(A)$  may be very small, therefore given

$$x = \sum_{j=1}^n \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} b$$

$x$  is very sensitive to small change in  $b$  due to the terms in the sum that have very small singular values.

**Solution:** Zero out small singular values to get the approximate solution

$$Ax = (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} x \approx (U_1 \quad U_2) \begin{pmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} x$$

so

$$x = V_1 \Sigma_1^{-1} U_1^H b$$

is an approximate solution that is numerically stable.

## Numerically Sensitive Problems

- ▶ Moon Example 7.4.1 shows that if  $\sigma_j$ -small then the vector  $\mathbf{u}_j \in \mathbb{R}^m$  defines a sensitive direction for  $b$ . i.e. if  $b$  is almost parallel with  $\mathbf{u}_j$  then  $x = \frac{\mathbf{v}_j \mathbf{u}_j^H}{\sigma_j} b$  is clearly sensitive to small changes in  $b$ . If  $b$  is perpendicular to  $\mathbf{u}_j$  then  $\mathbf{u}_j^H b = 0$  and we are ok.
- ▶ If  $A$  comes from noisy data (almost always) then  $A$  will usually be full rank, even if the original data that produced  $A$  would have resulted in a lower rank  $A$  if it wasn't corrupted by noise.
- ▶ But the nonzero singular values added by noise will usually be small.
- ▶ Therefore, an effective way to reduce the rank of  $A$  to get rid of the effect of noise is to zero the “small” singular values.

## Section 4

### Rank Reducing Approximations

# Rank Reducing Approximations

**Problem:** Given  $A$  with  $\text{rank}(A) = r$ , find a matrix  $B$  that is “close” to  $A$  in some sense, but with lower rank.

**Theorem (Moon Theorem 7.2)**

Given  $A \in \mathbb{C}^{m \times n}$  with  $\text{rank}(A) = r$ , then

$$A = U_1 \Sigma_1 V_1^H = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^H$$

Let  $k < r$  and let

$$A_k \triangleq \sum_{j=1}^k \sigma_j \mathbf{u}_j \mathbf{v}_j^H \quad (\text{rank}(A_k) = k)$$

Then  $\|A - A_k\|_2 = \sigma_{k+1}$  and  $A_k$  is the nearest rank  $k$  matrix to  $A$ , in the matrix 2-norm, i.e.

$$A_k = \arg \min_{\text{rank}(B)=k} \|A - B\|_2.$$

## Rank Reducing Approximations, Proof

**Remark:** In the previous section, we saw that we could reduce the rank by zeroing small singular values. This theorem shows that this is the best way to reduce the rank in the matrix 2-norm sense.

Proof.

$$\begin{aligned}\|A - A_k\|_2 &= \left\| \sum_{j=k+1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^H \right\|_2 \\ &= \max_{\|\mathbf{x}\|=1} \left\| \sum_{j=k+1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^H \mathbf{x} \right\|_2\end{aligned}$$

Note that we maximize by letting  $\mathbf{x}^* = \mathbf{v}_{k+1}$  since any other  $\mathbf{x}$  will be a linear combination of smaller singular values.

# Rank Reducing Approximations, Proof

Therefore

$$\|A - A_k\| = \|\sigma_{k+1} \mathbf{u}_{k+1}\| = \sigma_{k+1}$$

since  $\|\mathbf{u}_{k+1}\| = 1$ .

Because  $\|A - A_k\|_2 = \sigma_{k+1}$  we know that

$$\min_{\text{rank}(B)=k} \|A - B\| \leq \sigma_{k+1}.$$

To complete the proof we need to show that

$$\sigma_{k+1} \leq \min_{\text{rank}(B)=k} \|A - B\|.$$

## Rank Reducing Approximations, Proof

Let  $B$  be any rank- $k$  matrix. Then

$$\text{rank}(B) = k \implies \dim(\mathcal{N}(B)) = n - k.$$

Therefore, there exists  $\{x_{k+1}, \dots, x_n\}$  such that

$$\mathcal{N}(B) = \text{span}\{x_{k+1}, \dots, x_n\}$$

The columns of  $V_1$  are  $\{\mathbf{v}_1 \dots \mathbf{v}_k, \mathbf{v}_{k+1} \dots \mathbf{v}_r\}$  where  $\mathbf{v}_i \in \mathbb{C}^n$ . Let

$$z \in \underbrace{\text{span}\{\underbrace{x_{k+1}, \dots, x_n}_{\text{dim}=n-k}\}}_{\text{dimension at least one since there are } n+1 \text{ vectors}} \cap \underbrace{\text{span}\{\underbrace{\mathbf{v}_1, \dots, \mathbf{v}_{k+1}}_{\text{dim}=k+1}\}}$$

Therefore  $z \neq 0$ .

## Rank Reducing Approximations, Proof

Let

$$\begin{aligned}\|A - B\|_2 &= \max_{\|x\| \neq 0} \frac{\|(A - B)x\|}{\|x\|} \leq \frac{\|(A - B)z\|}{\|z\|} \\ &= \frac{\|Ax\|}{\|z\|} \text{ since } z \in \mathcal{N}(B) \\ &= \frac{\left\| \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^H z \right\|}{\|z\|} = \frac{\left\| \sum_{j=1}^{k+1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H z \right\|}{\|z\|}\end{aligned}$$

Since  $z \perp \text{span}\{\mathbf{v}_{k+2}, \dots, \mathbf{v}_r\}$  the smallest we can make the numerator is  $\sigma_{k+1}$  by a choice of  $z = \mathbf{v}_{k+1}$ . So

$$\|A - B\|_2 \geq \frac{\|\sigma_{k+1} \mathbf{v}_{k+1}\|}{\|\mathbf{v}_{k+1}\|} = \sigma_{k+1}$$

for any  $B$  such that  $\text{rank}(B) = k$  so that

$$\min_{\text{rank}(B)=k} \|A - B\|_2 \geq \sigma_{k+1}.$$

## Section 5

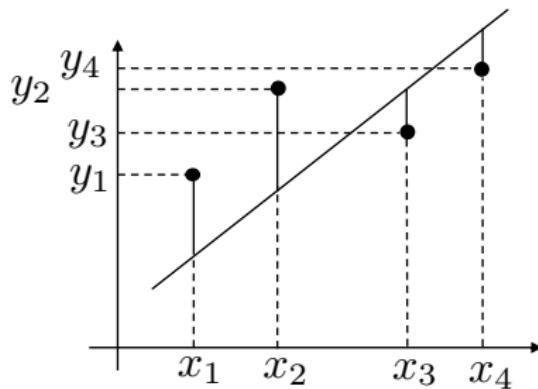
### Applications

## Application: Total least squares

If we are trying to fit a line to

$$y_i = ax_i$$

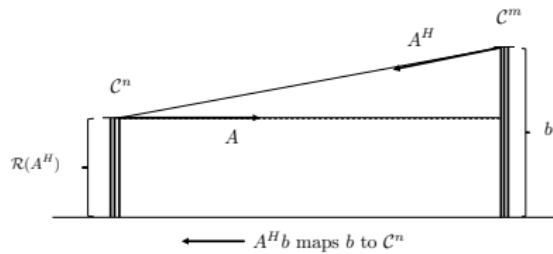
where  $(y_i, x_i)$  are measured. The least squares solution minimizes  $e_i = y_i - ax_i$ . Therefore  $y_i - e_i = ax_i$ .



In other words: fix the  $x_i$ 's and play with  $a$  to minimize the error.

## Application: Total least squares

For the general problem  $\min \|Ax - b\|$  we assume  $A$  is perfect and that the imperfection is completely in  $b$



Recall  $A^H Ax = A^H b$ . When we premultiply by  $A^H$  we zero everything in  $b$  that was in the null space of  $A^H$  (i.e. we get rid of the bad parts of  $b$ ).

## Application: Total least squares

However  $A$  often comes from noisy data as well (like when fitting a line to data) e.g. if  $\mathbf{u}_i = ax_i + b$ , then

$$\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix}$$

noisy                                    perfect

## Application: Total least squares

Another interpretation of least squares is to find the smallest perturbation of  $b$ , i.e.,  $\delta b$  such that

$$Ax = b + \delta b$$

where  $b + \delta b \in \mathcal{R}(A)$ .

The total least squares problem is to find the smallest perturbation of  $b$  and  $A$ , denoted  $\delta b$ ,  $\delta A$  such that

$$(A + \delta A)x = (b + \delta b)$$

supposing that  $(A \quad b)$  is full rank.

## Application: Total least squares

This can be written as

$$(A \quad b) \begin{pmatrix} x \\ -1 \end{pmatrix} + (\delta A \quad \delta b) \begin{pmatrix} x \\ -1 \end{pmatrix} = 0$$

or

$$[(A \quad b) + (\delta A \quad \delta b)] \begin{pmatrix} x \\ -1 \end{pmatrix} = 0.$$

Define

$$C \stackrel{\triangle}{=} (A \quad b) \text{ and } \Delta = (\delta A \quad \delta b)$$

then

$$(C + \Delta) \begin{pmatrix} x \\ -1 \end{pmatrix} = 0.$$

## Application: Total least squares

So  $\begin{pmatrix} x \\ -1 \end{pmatrix} \in \mathcal{N}(C + \Delta)$  which implies that  $C + \Delta$  is not full rank.

The problem is then to find the smallest perturbation  $\Delta$  such that  $C + \Delta$  loses rank.

Note that since  $C = (A \ b) \in \mathbb{C}^{m \times (n+1)}$ , for  $C$  to be full rank, we must have that  $m > n$ . Therefore we can write

$$C = \sum_{j=1}^{n+1} \sigma_j \mathbf{u}_j \mathbf{v}_j^H.$$

## Application: Total least squares

Hence, the smallest  $\Delta$  that reduces the rank of  $C$  is

$$\Delta = -\sigma_{n+1} \mathbf{u}_{n+1} \mathbf{v}_{n+1}^H.$$

Note that  $\mathbf{v}_{n+1} \in \mathcal{N}(C + \Delta)$  since

$$(C + \Delta)\mathbf{v}_{n+1} = \sum_{j=1}^r \sigma_j \mathbf{u}_j \mathbf{v}_j^H \mathbf{v}_{n+1} = 0$$

since  $\mathbf{v}_i \mathbf{v}_j = \delta_{ij}$ .

## Application: Total least squares

Therefore

$$\begin{pmatrix} x \\ -1 \end{pmatrix} = \alpha \mathbf{v}_{n+1} = \alpha \begin{pmatrix} \mathbf{v}_{n+1}(n : 1) \\ \mathbf{v}_{n+1}(n + 1) \end{pmatrix}$$

Letting  $\alpha = -\frac{1}{\mathbf{v}_{n+1}(n+1)}$  gives

$$x = \alpha \mathbf{v}_{n+1}(n : 1)$$

This is valid if  $\mathbf{v}_{n+1}(n + 1) \neq 0$ . Note that if  $\sigma_{n+1}$  is not a unique minimum singular value, i.e.  $\sigma_{n+1} = \sigma_n = \dots = \sigma_{k+1}$  then we want to find the smallest norm  $x$  such that

$$\begin{pmatrix} x \\ -1 \end{pmatrix} \in \text{span}\{\mathbf{v}_{k+1}, \dots, \mathbf{v}_{n+1}\}$$

## Application: Homography Matrix

## Application: MIMO Communication

Consider the MIMO communication system modeled by

$$\underbrace{Y(j\omega)}_{p \times 1} = \underbrace{H(j\omega)}_{1 \times m} \underbrace{X(j\omega)}_{m \times 1}$$

What is the maximum gain of the system?

$$\|Y(j\omega)\| = \|H(j\omega)X(j\omega)\| \leq \|H(j\omega)\| \|X(j\omega)\|$$

Therefore, the maximum gain is given by

$$\begin{aligned}\gamma_{\max}(j\omega) &= \max_{X(j\omega) \neq 0} \frac{\|H(j\omega)X(j\omega)\|}{\|X(j\omega)\|} \\ &= \|H(j\omega)\| \\ &= \bar{\sigma}(H(j\omega)),\end{aligned}$$

where  $\bar{\sigma}(H(j\omega))$  is the maximum singular value of  $H(j\omega)$ .

## Application: MIMO Communication

How do you achieve this gain? Since

$$H(j\omega) = \Sigma \sigma_k(j\omega) \mathbf{u}_k(j\omega) \mathbf{v}_k^H(j\omega),$$

letting

$$X(j\omega) = \mathbf{v}_1(j\omega)$$

maximizes the gain in the system over the set  $\|X(j\omega)\| = 1$ .

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 14.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Gradient Descent

Gradient Descent: Multivariable Case

Application: LMS Adaptive Filtering

Gauss-Newton Optimization

Levenberg-Marquardt Optimization

# Section 1

## Gradient Descent

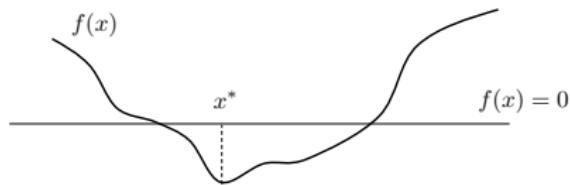
# Gradient Descent

The topic for the remainder of the course is minimization and maximization of functions.

In particular we will constrain our attention to continuously differentiable functions.

# Gradient Descent

Suppose we have a function of the form



and we would like to find  $x^*$ , what should we do?

## Gradient Descent

The basic idea of gradient descent is to pick any  $x^{[0]}$  and then move “downward”. To move down, we look at the slope of  $f$ .

If  $\frac{\partial f}{\partial x}(x^{[k]})$  is positive, chose  $x^{[k+1]} < x^{[k]}$ .

If  $\frac{\partial f}{\partial x}(x^{[k]})$  is negative, choose  $x^{[k+1]} > x^{[k]}$

i.e.

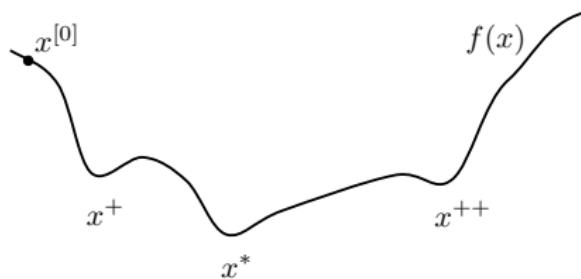
$$x^{[k+1]} = x^{[k]} - \alpha \frac{\partial f}{\partial x}(x^{[k]}),$$

where  $\alpha$  is the step size.

# Gradient Descent

Before moving to the multivariable case, let's consider the potential problems with this approach.

**Problem 1: Local Minima.** If  $f$  looks like this:



then if the initial condition is at  $x^{[0]}$ , the iteration

$$x^{[k+1]} = x^{[k]} - \alpha \frac{\partial f}{\partial x}(x^{[k]})$$

will converge to  $x^+$ , if  $\alpha$  is small enough.

# Gradient Descent

Other initial conditions will result in  $x^{++}$  while others will give  $x^*$ , the true minimum.

This is a fundamental problem with any method that relies on derivative information. There are no completely satisfactory solutions to the problem. However there are many ad-hoc fixes.

## Example

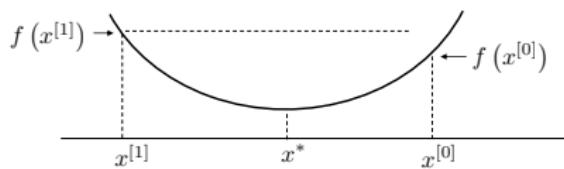
- ▶ Execute from numerous “random” initial conditions and pick the lowest solution.
- ▶ Occasionally introduce random jumps in  $x$ .
- ▶ etc...

# Gradient Descent

**Problem 2: Step Size.** The selection of  $\alpha$  can have a major effect on the convergence of the sequence

$$x^{[k+1]} = x^{[k]} - \alpha \frac{\partial f}{\partial x}(x^{[k]})$$

For example,



Note  $f$  is very steep on sides, so  $\alpha \frac{\partial f}{\partial x}(x^{[k]})$  could be large. This could cause  $x^{[1]}$  to overshoot the minimum. This could cause (1) instability, (2) limit cycles, (3) extremely slow and oscillatory convergence

# Gradient Descent

**Lesson:** Don't make  $\alpha$  too large.

However if  $\alpha$  is too small, then convergence will be very slow.

Most implementations adapt the size of  $\alpha$ .

## Section 2

### Gradient Descent: Multivariable Case

# Gradient Descent: Multivariable Case

Let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be a multivariable function.

## Example

If  $x \in \mathbb{R}^n$  then  $f(x) = x_1^2 + x_2^2 + \cdots + x_n^2$  maps  $\mathbb{R}^n \rightarrow \mathbb{R}$ .

The gradient of a multivariable function is

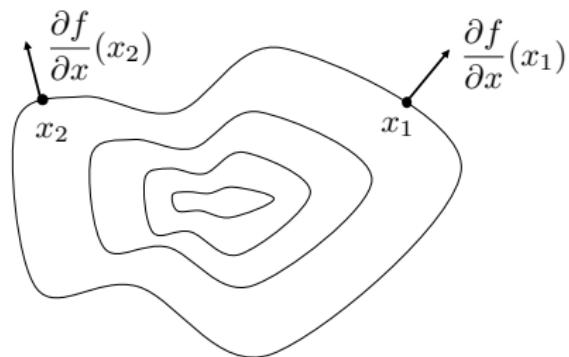
$$\frac{\partial f}{\partial x} = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

and maps  $\mathbb{R}^n \rightarrow \mathbb{R}^n$ .

# Gradient Descent: Multivariable Case

## Example

If  $f(x) = x_1^2 + \cdots + x_n^2$  then  $\frac{\partial f}{\partial x} = \begin{pmatrix} 2x_1 \\ 2x_2 \\ \vdots \\ 2x_n \end{pmatrix}$



The gradient points perpendicular to the level curves of  $f$ .

## Gradient Descent: Multivariable Case

Theorem (Moon Theorem 14.5)

*Let  $f : \mathbb{R}^m \rightarrow \mathbb{R}$  be a differentiable function in some open set  $D$ .  
The gradient  $\frac{\partial f}{\partial x}(x)$  points in the direction of the maximum  
increase of  $f$  at the point  $x$ .*

## Gradient Descent: Multivariable Case

Proof.

Expand  $f(x + \lambda y)$  in a Taylor series as

$$f(x + \lambda y) = f(x) + \lambda \frac{\partial f^T}{\partial x}(x)y + \text{Higher Order Terms (HOT)}$$

where HOT. are  $O(\lambda^2)$ , i.e.,

$$\lim_{\lambda \rightarrow 0} \frac{\text{H.O.T.}}{\lambda} = 0.$$

We would like to find  $y$  that maximizes  $f(x + \lambda y)$  as  $\lambda \rightarrow 0$ .

By Cauchy-Schwartz,  $\frac{\partial f^T}{\partial x}y$  is maximized when  $y = \frac{\partial f}{\partial x}$ . □

## Gradient Descent: Multivariable Case

For multivariable functions, the gradient descent formula is

$$x^{[k+1]} = x^{[k]} - \alpha_k \frac{\partial f}{\partial x}(x^{[k]})$$

Again, the selection of the step size is very important. If  $\alpha_k$  is too small convergence will be slow.

If  $\alpha_k$  is too large, algorithm could be unstable.

How to pick the right  $\alpha$ ?

## Gradient Descent: Multivariable Case

Locally around a min or max, every smooth function can be approximated by a quadratic (Taylor series).

We can gain insight about the selection of  $\alpha$  by studying quadratic functions.

Let  $f(x) = x^T R x - 2b^T x$  where  $x \in \mathbb{R}^m, b \in \mathbb{R}^m, R = R^T > 0$ .

Taking the gradient we get

$$\frac{\partial f}{\partial x} = 2Rx - 2b.$$

## Gradient Descent: Multivariable Case

So the gradient descent algorithm is

$$x^{[k+1]} = x^{[k]} - 2\alpha(Rx^{[k]} - b).$$

Let  $x^*$  satisfy  $Rx^* = b$  then

$$x^{[k+1]} - x^* = x^{[k]} - x^* - 2\alpha(Rx^{[k]} - Rx^*)$$

Define  $y^{[k]} = x^{[k]} - x^*$  and  $\mu = 2\alpha$ , then

$$\begin{aligned}y^{[k+1]} &= y^{[k]} - \mu Ry^{[k]} \\&= (I - \mu R)y^{[k]} \\ \implies y^{[k]} &= (I - \mu R)^k y^{[0]}. \end{aligned}$$

## Gradient Descent: Multivariable Case

Since  $R$  is symmetric positive definite

$$R = Q\Lambda Q^T$$

where  $Q$ -orthogonal. Therefore,

$$\begin{aligned}y^{[k]} &= (QQ^T - \mu Q\Lambda Q^T)^k y^{[0]} \\&= Q(I - \mu\Lambda)^k Q^T y^{[0]}\end{aligned}$$

Letting  $z = Q^T y$ ,

$$z^{[k]} = (I - \mu\Lambda)^k z^{[0]} \tag{1}$$

$$\implies z_i^{[k]} = (1 - \mu\lambda_i)^k z_i^{[0]} \tag{2}$$

which converges if  $|1 - \mu\lambda_i| < 1$ ,  $i = 1, \dots, m$ .

## Gradient Descent: Multivariable Case

Therefore, convergence happens when

$$\begin{aligned} -1 &< 1 - \mu\lambda_i < 1 \\ \iff -2 &< -\mu\lambda_i < 0 \\ \iff 0 &< \mu\lambda_i < 2 \\ \iff 0 &< \mu < \frac{2}{\lambda_i} \end{aligned}$$

Recall that  $\lambda_i > 0$  when  $R$  is positive definite, so if

$$0 < \alpha < \frac{1}{\lambda_{\max}(R)}$$

then steepest descent converges for quadratic functions.

## Gradient Descent: Multivariable Case

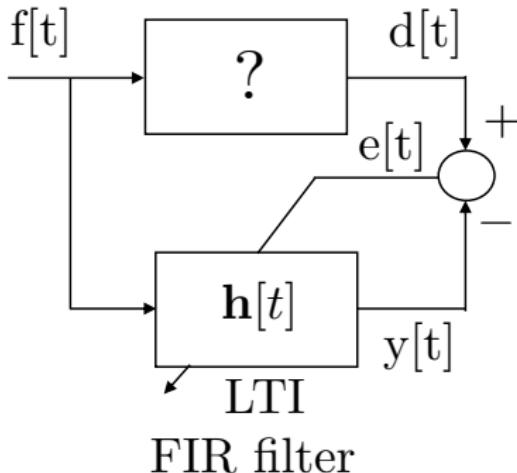
Note that the convergence along each eigenaxis is determined by  $\frac{1}{\lambda_i}$ .

Therefore if  $R$  is ill-conditioned, i.e.,  $\frac{\lambda_{\max}}{\lambda_{\min}}$  is large, then convergence for gradient descent will be much slower along some axes than others.

## Section 3

### Application: LMS Adaptive Filtering

# LMS Adaptive Filtering



Recall the RLS adaptive filter algorithm.

The objective is to minimize the error

$$J(\mathbf{h}) = (d[t] - y[t])^2.$$

- ▶ The RLS minimizes the squared error of all past outputs, but LMS only minimizes the squared error of the current output.
- ▶ The RLS algorithm was derived using the projection theorem.
- ▶ LMS is derived using gradient descent.

# LMS Adaptive Filtering

Assume that the output of the adaptive filter is

$$y[t] = \sum_{\ell=0}^{m-1} h[\ell]f[t - \ell] = \mathbf{f}^\top[t]\mathbf{h}$$

where

$$\mathbf{f}[t] = \begin{pmatrix} f[t] \\ f[t-1] \\ \vdots \\ f[t-m+1] \end{pmatrix} \text{ and } \mathbf{h} = \begin{pmatrix} h[0] \\ h[1] \\ \vdots \\ h[m-1] \end{pmatrix}$$

# LMS Adaptive Filtering

Then

$$\begin{aligned} J(\mathbf{h}) &= (d[t] - y[t])^2 \\ &= (d[t] - \mathbf{f}^\top[t]\mathbf{h})^2 \\ &= d^2[t] - d[t]\mathbf{f}^\top[t]\mathbf{h} - d[t]\mathbf{h}^\top\mathbf{f}[t] + \mathbf{h}\mathbf{f}[t]\mathbf{f}^\top[t]\mathbf{h} \end{aligned}$$

where

$$\frac{\partial J}{\partial \mathbf{h}} = 2\mathbf{f}[t]\mathbf{f}^\top[t]\mathbf{h} - 2d[t]\mathbf{f}[t]$$

# LMS Adaptive Filtering

So let

$$\mathbf{h}[t+1] = \mathbf{h}[t] - \alpha \frac{\partial J}{\partial \mathbf{h}}(\mathbf{h}[t])$$

gives

$$\begin{aligned}\mathbf{h}[t+1] &= \mathbf{h}[t] - 2\alpha(\mathbf{f}[t]\mathbf{f}^\top[t]\mathbf{h}[t] - d[t]\mathbf{f}[t]) \\ &= \mathbf{h}[t] + \mu\mathbf{f}[t](d[t] - \mathbf{f}^\top[t]\mathbf{h}[t])\end{aligned}$$

$$\boxed{\mathbf{h}[t+1] = \mathbf{h}[t] + \mu\mathbf{f}[t]e[t]}$$

This is known as the LMS adaptive filter.

Compare to RLS...

For discussion on convergence, consult Moon Chap 14...

## Section 4

# Gauss-Newton Optimization

## Least Squares as a Gradient Descent Problem

Consider the least squares problem

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2^2$$

where  $A \in \mathbb{R}^{m \times n}$  is tall. We know that the solution is

$$x^* = (A^\top A)^{-1} A^\top b.$$

Can we pose this as a gradient descent problem?

## Least Squares as a Gradient Descent Problem

Define the residual as

$$\mathbf{r}(x) = \begin{pmatrix} r_1(x) \\ \vdots \\ r_m(x) \end{pmatrix} = Ax - b$$

and define the sum-of-squares error as

$$\begin{aligned} S(x) &= \frac{1}{2} \mathbf{r}^\top(x) \mathbf{r}(x) \\ &= \frac{1}{2} \sum_{j=1}^m r_j^2(x) \\ &= \frac{1}{2} (Ax - b)^\top (Ax - b) \\ &= \frac{1}{2} \|Ax - b\|_2^2. \end{aligned}$$

The least squares problem is to find  $x$  that minimizes  $S(x)$ .

## Least Squares as a Gradient Descent Problem

The gradient of  $S$  is given by

$$\begin{aligned}\frac{\partial S}{\partial x} &= \frac{\partial \mathbf{r}^\top}{\partial x}(x) \mathbf{r}(x) \\ &= A^\top(Ax - b) = A^\top Ax - A^\top b.\end{aligned}$$

So the gradient descent algorithm gives

$$x^{[k+1]} = x^{[k]} - \alpha \left( A^\top Ax^{[k]} - A^\top b \right)$$

In general, we might allow  $\alpha > 0$  to be a positive definite matrix  $\mathcal{A} > 0$ :

$$x^{[k+1]} = x^{[k]} - \mathcal{A} \left( A^\top Ax^{[k]} - A^\top b \right).$$

# Least Squares as a Gradient Descent Problem

Selecting

$$\mathcal{A} = (A^\top A)^{-1}$$

gives

$$\begin{aligned}x^{[k+1]} &= x^{[k]} - (A^\top A)^{-1} \left( A^\top A x^{[k]} - A^\top b \right) \\&= x^{[k]} - (A^\top A)^{-1} (A^\top A) x^{[k]} + (A^\top A)^{-1} A^\top b \\&= (A^\top A)^{-1} A^\top b,\end{aligned}$$

which is the optimal solution.

Noting that  $A = \frac{\partial \mathbf{r}}{\partial x}$ , we have shown that the iteration

$$x^{[k+1]} = x^{[k]} - \left( \frac{\partial \mathbf{r}^\top}{\partial x}(x^{[k]}) \frac{\partial \mathbf{r}}{\partial x}(x^{[k]}) \right)^{-1} \frac{\partial \mathbf{r}^\top}{\partial x}(x^{[k]}) \mathbf{r}(x^{[k]})$$

converges to the optimal in one step when  $\mathbf{r}(x) = Ax - b$ .

## Nonlinear Least Squares

Let  $r_j(x)$ ,  $j = 1, \dots, m$  be a general set of residual function to be minimized. In other words, suppose we wish to solve

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \mathbf{r}^\top(x) \mathbf{r}(x).$$

Let  $\mathbf{J}(x) \triangleq \frac{\partial \mathbf{r}}{\partial x}(x)$ . Then the Gauss-Newton (GN) iteration algorithm is given by

$$x^{[k+1]} = x^{[k]} - \left( \mathbf{J}^\top(x^{[k]}) \mathbf{J}(x^{[k]}) \right)^{-1} \mathbf{J}^\top(x^{[k]}) \mathbf{r}(x^{[k]})$$

We know that the GN method converges in one step for the linear least squares problem.

## Section 5

### Levenberg-Marquardt Optimization

## Nonlinear Least Squares

The downside of GN is that the matrix  $J^\top(x)J(x)$  may be ill-conditioned at some states  $x$ .

For the general nonlinear least squares problem, we have

$$\frac{\partial \frac{1}{2}\mathbf{r}^\top(x)\mathbf{r}(x)}{\partial x} = \frac{\partial \mathbf{r}^\top}{\partial x}(x)\mathbf{r}(x) = \mathbf{J}^\top(x)\mathbf{r}(x).$$

Therefore we have

Gradient Descent  $x^{[k+1]} = x^{[k]} - \alpha \mathbf{J}^\top(x^{[k]})\mathbf{r}(x^{[k]})$

Gauss-Newton  $x^{[k+1]} = x^{[k]} - \left( \mathbf{J}^\top(x^{[k]})\mathbf{J}(x^{[k]}) \right)^{-1} \mathbf{J}^\top(x^{[k]})\mathbf{r}(x^{[k]}).$

Note that there is no inverse for Gradient Descent, but it may converge slowly, even for linear residuals.

## Nonlinear Least Squares

The Levenberg-Marquardt (LM) iteration is a combination of gradient descent and Gauss-Newton:

$$x^{[k+1]} = x^{[k]} - \left( \lambda I + \mathbf{J}^\top(x^{[k]}) \mathbf{J}(x^{[k]}) \right)^{-1} \mathbf{J}^\top(x^{[k]}) \mathbf{r}(x^{[k]}),$$

where  $\lambda = 1/\alpha$ .

Note that  $\lambda I + \mathbf{J}^\top \mathbf{J}$  is guaranteed to be full rank and well conditioned for large  $\lambda$ .

Standard practice:

- ▶ For the first iteration make  $\lambda$  large (e.g.,  $\approx 10^4$ )
- ▶ If squared error decreases, decrease  $\lambda$  for next iteration (e.g., by half).
- ▶ If squared error increases, increase  $\lambda$  for next iteration (e.g., by 2x).

# Weighted Nonlinear Least Squares

If  $W = W^\top > 0$  is a weighting matrix, then

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \mathbf{r}^\top(x) W \mathbf{r}(x).$$

results in

$$(GD) \quad x^{[k+1]} = x^{[k]} - \lambda^{-1} \mathbf{J}^\top W \mathbf{r}|_{x^{[k]}}$$

$$(GN) \quad x^{[k+1]} = x^{[k]} - \left( \mathbf{J}^\top W \mathbf{J} \right)^{-1} \mathbf{J}^\top W \mathbf{r}|_{x^{[k]}}$$

$$(LM) \quad x^{[k+1]} = x^{[k]} - \left( \lambda I + \mathbf{J}^\top W \mathbf{J} \right)^{-1} \mathbf{J}^\top W \mathbf{r}|_{x^{[k]}}.$$

# ECEn 671: Mathematics of Signals and Systems

## Moon: Chapter 18.

Randal W. Beard

Brigham Young University

August 29, 2023

# Table of Contents

Constrained Optimization

General Constrained Optimization

Equality Constraints: Lagrange Multipliers

Sufficient Conditions

Inequality Constraints: Kuhn-Tucker Conditions

## Section 1

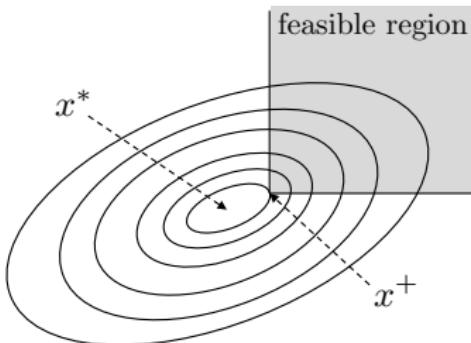
# Constrained Optimization

# Constrained Optimization

In Chapter 14 we studied unconstrained minimization of continuously differentiable functions.

In Chapter 18 we focus on constrained optimization problems.

For example, given the level curves,



Note that the constrained optimum  $x^+$  does not equal the unconstrained optimum  $x^*$ .

The unconstrained optimum is  $x^*$ ; the constrained optimum is  $x^+$ .

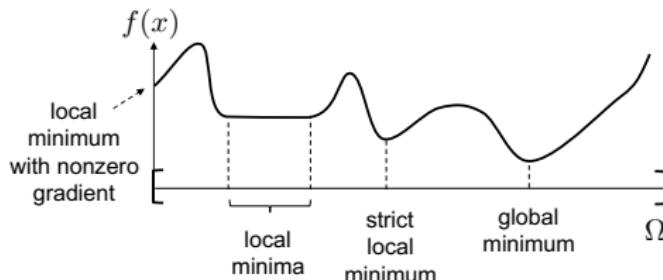
# Constrained Optimization

## Definition

Let  $\Omega \subseteq \mathbb{R}^n$  be the feasible region. Then  $x^* \in \Omega$  is a local minimum of  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  over  $\Omega$  if  $\exists \epsilon > 0$  such that

$$x \in \Omega \cap \{y \in \mathbb{R}^n : |u - x^*| < \epsilon\} \implies f(x) \geq f(x^*).$$

If  $f(x) > f(x^*)$  then  $x^*$  is a strict local minimum. If true for all  $\epsilon > 0$  then  $x^*$  is a global minimum.



# Constrained Optimization

## Definition

Let  $x \in \Omega$  and  $d \in \mathbb{R}^n$ , then

$$y = x + \alpha d$$

is a feasible point if  $y \in \Omega$ .

## Definition

The vector  $d$  is a feasible direction at  $x$ , if  $\exists \epsilon_0 > 0$  such that

$$x + \epsilon d \in \Omega$$

for every  $0 \leq \epsilon \leq \epsilon_0$ .

# Constrained Optimization

Recall, if  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  then the gradient vector is

$$\frac{\partial f}{\partial x} = \nabla_x f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

and the Hessian matrix is

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

If  $\Omega = \mathbb{R}^n$  then a necessary condition for  $x^*$  to be a local minima is that  $\nabla_x f(x^*) = 0$ . What about constrained optimization problems?

# Constrained Optimization

## Theorem (Moon Theorem 18.1)

Let  $\Omega \subseteq \mathbb{R}^n$  and let  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  be  $\mathcal{C}^1$  (continuously differentiable) on  $\Omega$ .

1. If  $x^*$  is a local minimum of  $f$  over  $\Omega$ , then for any feasible direction  $d \in \mathbb{R}^n$  at  $x^*$

$$[\nabla_x f(x^*)]^\top d \geq 0$$

2. If  $x^*$  is an interior point of  $\Omega$ , then

$$\nabla f(x^*) = 0.$$

3. If in addition,  $f \in \mathcal{C}^2$  and  $\nabla_x f(x^*)^\top d = 0$ , then

$$d^\top \nabla^2 f(x^*) d \geq 0$$

Note that this is a weaker condition than psd Hessian.

## Proof of Theorem 18.1

1. By Taylor series expansion,

$$\begin{aligned} f(x^* + \epsilon d) &= f(x^*) + \epsilon \nabla_x f(x^*)^\top d + O(\epsilon) \\ \implies f(x^* + \epsilon d) - f(x^*) &= \epsilon \nabla_x f(x^*)^\top d + O(\epsilon) \end{aligned}$$

Since  $x^*$  is a local minimum, for  $\epsilon$  sufficiently small we must have that

$$\begin{aligned} \implies f(x^* + \epsilon d) - f(x^*) &\geq 0 \\ \implies \nabla_x f(x^*)^\top d &\geq 0. \end{aligned}$$

## Proof of Theorem 18.1, cont.

2. If  $x^*$  is an interior point then every  $d \in \mathbb{R}^n$  is feasible at  $x^*$ , i.e.

$$\langle \nabla_x f(x^*), d \rangle_{\mathbb{R}^n} = 0, \quad \forall d \in \mathbb{R}^n.$$

Therefore,

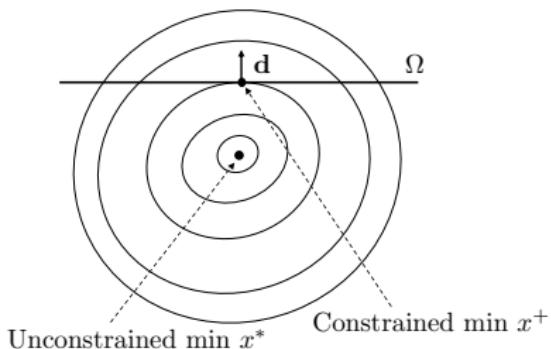
$$\begin{aligned} \nabla_x f(x^*)^\top d &\geq 0 \quad \text{and} \quad \nabla_x f(x^*)^\top (-d) \geq 0 \\ \implies \nabla_x f(x^*)^\top d &= 0, \quad \forall d \in \mathbb{R}^n \\ \implies \nabla_x f(x^*) &= 0 \end{aligned}$$

since  $\mathbb{R}^n$  is a finite dimensional vector space .

## Proof of Theorem 18.1, cont.

3. If  $\nabla_x f(x^*)^\top d = 0$  then the Taylor series for  $f$  is

$$\begin{aligned}f(x^* + \epsilon d) &= f(x^*) + \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2) \\ \implies 0 \leq f(x^* + \epsilon d) - f(x^*) &= \epsilon^2 d^\top \nabla^2 f(x^*) d + O(\epsilon^2) \\ \implies d^\top \nabla^2 f(x^*) d &\geq 0.\end{aligned}$$



Note: Any feasible  $d$  points uphill.

Note: The function is concave in feasible region.

# Constrained Optimization: Sufficient Conditions

Are there sufficient conditions?

First, suppose that the constraints are not active, i.e.  $x^*$  is an interior point of  $\Omega$ . (We will consider the active constraint case later.)

Theorem (Moon Theorem 18.2)

Let  $f \in C^2$  on  $\Omega$  and let  $x^*$  be an interior point of  $\Omega$ . If  $\nabla f(x^*) = 0$  and  $\nabla^2 f(x^*)$  is positive definite then  $x^*$  is a strict local minimum of  $f$ .

## Constrained Optimization: Sufficient Conditions: Proof

Proof.

Let  $d$  be any unit vector in  $\mathbb{R}^n$  then

$$\begin{aligned}f(x^* + \epsilon d) &= f(x^*) + \epsilon \nabla f(x^*)^\top d + \frac{\epsilon^2}{2} d^\top \nabla^2 f(x^*) d + O(\epsilon^3) \\ \implies f(x^* + \epsilon d) - f(x^*) &= \frac{\epsilon^2}{2} d^\top \nabla^2 f(x^*) d + O(\epsilon^3)\end{aligned}$$

Since  $\nabla^2 f(x^*)$  is positive definite, it follows that for  $\epsilon$  sufficiently small

$$f(x^* + \epsilon d) - f(x^*) > 0,$$

which implies that  $x^*$  is a strict local minimum. □

**Note:** we cannot generalize this theorem to the case when  $\nabla^2 f(x^*)$  is p.s.d.. Why?

## Section 2

### General Constrained Optimization

# Constrained Optimization

In general we have two types of constraints:

1. Equality constraints of the form

$$h_i(x) = 0$$

For example:

$$h_1(x) \stackrel{\triangle}{=} x_1^2 + x_1x_2x_3 + \tan(x_3)\cos(x_2) = 0$$

2. Inequality constraints of the form

$$g_i(x) \leq 0$$

For example

$$x_1 \geq 0,$$

$$x_2 \geq 0$$

$$\Rightarrow g_1(x) \stackrel{\triangle}{=} -x_1 \leq 0,$$

$$g_2(x) \stackrel{\triangle}{=} -x_2 \leq 0$$

# Constrained Optimization

In fact a region  $\Omega \subset \mathbb{R}^n$  can always be described by inequality constraints.

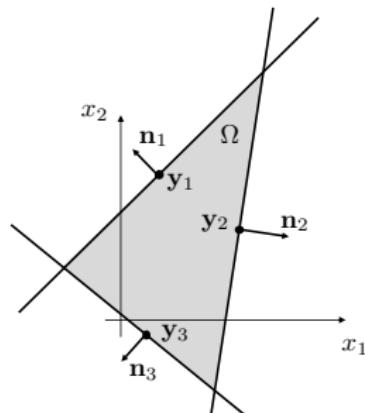
## Example

Feasible Region  $\Omega$ :

$$(x - \mathbf{y}_1)^\top \mathbf{n}_1 \leq 0$$

$$(x - \mathbf{y}_2)^\top \mathbf{n}_2 \leq 0$$

$$(x - \mathbf{y}_3)^\top \mathbf{n}_3 \leq 0$$



Where  $\mathbf{n}_i$  is a vector normal to the linear constraint.

# Constrained Optimization

A general constrained optimization problem can be written as

$$\begin{aligned} & \min_{x \in \Omega} f(x) \\ \text{s.t. } & h_1(x) = 0, \\ & \vdots, \\ & h_m(x) = 0, \\ & g_1(x) \leq 0, \\ & \vdots, \\ & g_p(x) \leq 0 \end{aligned}$$

# Constrained Optimization

Letting

$$\mathbf{h} = (h_1 \dots h_m)^\top$$
$$\mathbf{g} = (g_1 \dots g_p)^\top,$$

we have

$$\begin{aligned} & \min_{x \in \Omega} f(x) \\ \text{s.t. } & \mathbf{h}(x) = 0, \\ & \mathbf{g}(x) \leq 0 \end{aligned}$$

Equality constraints are easier to deal with than inequality constraints.

We will first treat equality constraints, then inequality constraints.

## Section 3

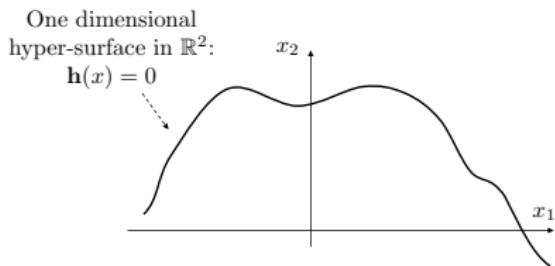
### Equality Constraints: Lagrange Multipliers

# Equality Constraints: Lagrange Multipliers

Several geometric insights help:

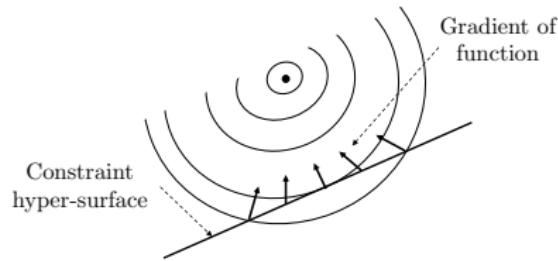
## Insight #1

- ▶ Geometrically what do the constraints  $\mathbf{h}(x) = 0$  look like?
- ▶ What if  $\mathbf{h}$  is linear, i.e.  $\mathbf{h}(x) = Hx = 0$  where  $H : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .
- ▶ The constraint implies that  $x$  must be in the null space of  $H$ , which is a linear space of dimension  $n - m$ , i.e. an  $n - m$  dimensional hyperplane.
- ▶ In general,  $\mathbf{h}(x) = 0$  is an  $n - m$  dimensional hypersurface in  $\mathbb{R}^n$ .



# Equality Constraints: Lagrange Multipliers

## Insight #2

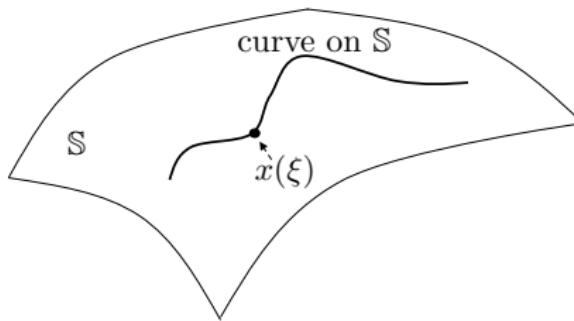


At a constrained minimum the gradient is orthogonal to the hypersurface.

# Equality Constraints: Lagrange Multipliers

To formalize, we need some definitions.

Let  $\mathbb{S}$  be a hyper-surface of dimension  $n - m$ . Let  $x$  be a curve on  $\mathbb{S}$  continuously parameterized by  $\xi \in [a, b]$ , i.e.  $x(\xi) \in \mathbb{S}$ .



The derivative of the curve at  $x(\xi_0)$  is

$$\dot{x}(\xi_0) = \frac{d}{d\xi}x(\xi_0).$$

# Equality Constraints: Lagrange Multipliers

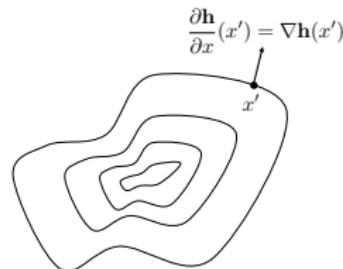
## Definition

The tangent plane to a surface  $\mathbb{S}$  at  $x \in \mathbb{S}$  is the span of the derivatives of all the differentiable curves on  $\mathbb{S}$  at  $x$ .

The problem with this definition is that it is not constructive, i.e. it doesn't give us a good way to actually construct the tangent plane.

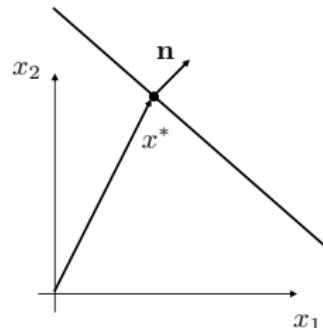
## Equality Constraints: Lagrange Multipliers

To construct the tangent plane, recall that the gradient of  $\mathbf{h}(x)$  is orthogonal to level curves of  $\mathbf{h}(x)$ :



Also recall that the formula for the plane is

$$\left\{ y \in \mathbb{R}^n \mid \mathbf{n}^\top (y - x^*) = 0 \right\}.$$



# Equality Constraints: Lagrange Multipliers

Therefore the tangent plane of  $h(x)$  at  $x^*$  is given by

$$P = \{y \in \mathbb{R}^n \mid \nabla \mathbf{h}^\top(x^*)(y - x^*) = 0\}$$

where  $\nabla \mathbf{h}^\top(x^*)$  defines an  $n - m$  dimensional plane if the rows are linearly independent.

## Definition

When the gradient vectors  $\nabla h_1, \nabla h_2, \dots, \nabla h_m$  are linearly independent at  $x^*$ ,  $x^*$  is called a regular point.

We will always assume “regularity” of the constraints.

# Equality Constraints: Lagrange Multipliers

Lemma (Moon Lemma 18.1)

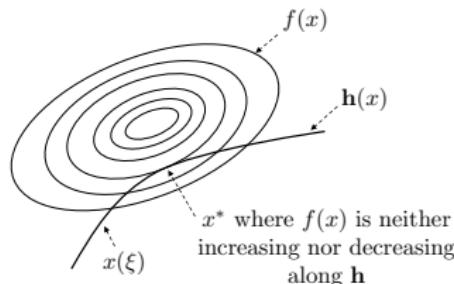
Let  $x(\xi)$  be a curve on  $\mathbf{h}(x) = 0$  such that

$$x(\xi)|_{\xi=0} = x^*,$$

is a constrained local minimum of  $f$ . Then

$$\frac{d}{d\xi} f(x(\xi)) \Big|_{\xi=0} = 0.$$

Geometry: at point  $p$ ,  $f$  is neither increasing nor decreasing along  $x(\xi)$ .



# Equality Constraints: Lagrange Multipliers: Proof

Proof.

Expanding  $f(x(\xi))$  in a Taylor series:

$$f(x(\xi)) = f(x(0)) + \xi \frac{d}{d\xi} f(x(\xi)) \Big|_{\xi=0} + O(|\xi|)$$

If  $f(x(0))$  is a local minimum then for  $|\xi|$ -small

$$\xi \frac{d}{d\xi} f(x(\xi)) \Big|_{\xi=0} \geq 0$$

for all  $\xi$  both positive and negative. Therefore

$$\frac{d}{d\xi} f(x(\xi)) \Big|_{\xi=0} = 0$$

where  $\frac{d}{d\xi} f(x(\xi)) = \nabla^\top f(x(\xi)) \dot{x}(\xi)$  and  $\dot{x}(\xi)$  is an element of the tangent plane.

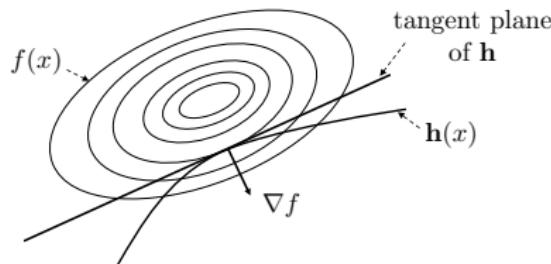
# Equality Constraints: Lagrange Multipliers

Lemma (Moon Lemma 18.2)

If  $x^*$  is a regular point of  $\mathbf{h}(x) = 0$  and a local constrained minimum, then

$$\nabla \mathbf{h}^\top(x^*)y = 0 \Rightarrow \nabla f^\top(x^*)y = 0$$

i.e. if  $y$  is in the tangent plane of  $\mathbf{h}$  at  $x^*$ , then  $y$  is orthogonal to the gradient of  $f$ .



## Proof of Lemma 18.2

**Proof.**

Translate the coordinate system such that  $x^* = 0$ . Regularity implies that the tangent plane is given by

$$P(x^*) \stackrel{\triangle}{=} \{z \in \mathbb{R}^n \mid \nabla \mathbf{h}^\top(x^*) z = 0\}$$

Let  $y \in P(x^*)$  then

$$\nabla \mathbf{h}^\top(x^*) y = 0.$$

Now choose a smooth curve  $x(\xi)$  on  $\mathbf{h}(x) = 0$  such that  $x(0) = x^*$  and  $\dot{x}(0) = y$ .

From Lemma 18.1,  $\nabla f^\top(x^*) y = 0$ . □

## Key Insight

At a constrained local minimum  $\nabla f(x^*)$  and the columns of  $\nabla \mathbf{h}(x^*)$  are parallel, i.e., there is some scalar  $\mu_i$  such that

$$\begin{aligned}\nabla f(x^*) &= \mu_i \nabla h_i(x^*) \quad i = 1, \dots, m \\ \implies m \nabla f(x^*) &= \sum_{i=1}^m \mu_i \nabla h_i(x^*) \\ \implies \nabla f(x^*) - \sum_{i=1}^m \frac{\mu_i}{m} \nabla h_i(x^*) &= 0 \\ \implies \boxed{\nabla f(x^*) + \nabla \mathbf{h}(x^*) \lambda} &= 0\end{aligned}$$

where

$$\lambda = \left( -\frac{\mu_1}{m} \quad \dots \quad -\frac{\mu_m}{m} \right)^\top.$$

The vector  $\lambda \in \mathbb{R}^m$  is called the Lagrange Multiplier.

## Necessary Conditions

Theorem (Moon Theorem 18.3 (Necessary conditions for equality constraints))

Let  $x^*$  be a local extremum of  $f$  subject to the constraints  $h(x) = 0$ , and let  $x^*$  be a regular point. Then there is a  $\lambda \in \mathbb{R}^n$  such that

$$\nabla f(x^*) + \nabla h(x^*)\lambda = 0.$$

## Corollary

Let

$$L(x, \lambda) = f(x) + h^\top(x)\lambda.$$

Then if  $x^*$  is a regular local extremum, then

$$\nabla_x L(x^*, \lambda^*) = \frac{\partial L}{\partial x}(x^*, \lambda^*) = 0$$

and

$$\nabla_\lambda L(x^*, \lambda^*) = \frac{\partial L}{\partial \lambda}(x^*, \lambda^*) = 0.$$

## Proof of Theorem 18.3

Proof.

$$\frac{\partial L}{\partial x} = \nabla f(x^*) + \nabla h(x^*)\lambda^* = 0 \quad (1)$$

$$\frac{\partial L}{\partial \lambda} = h(x^*) = 0 \quad (2)$$

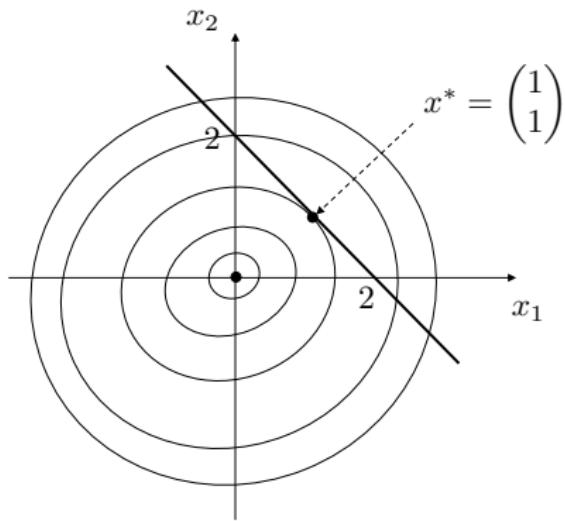
Equation (2) implies that the constraints are satisfied.

Equation (1) comes from Theorem 18.3. □

## Lagrange Multipliers: Example 1

$$\min \quad x_1^2 + x_2^2$$

$$\text{s.t.} \quad x_1 + x_2 = 2$$



The Lagrangian is

$$L = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 2).$$

## Lagrange Multipliers: Example 1, cont.

The necessary conditions for a minimum are

$$\frac{\partial L}{\partial x_1} = 2x_1 + \lambda = 0 \implies x_1 = -\frac{\lambda}{2}$$

$$\frac{\partial L}{\partial x_2} = 2x_2 + \lambda = 0 \implies x_2 = -\frac{\lambda}{2}$$

$$\frac{\partial L}{\partial \lambda} = x_1 + x_2 - 2 = 0 \implies -\lambda - 2 = 0.$$

Therefore  $\lambda = -2$ ,  $x_1 = 1$ ,  $x_2 = 1$ , which implies that

$$x^* = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

as expected.

## Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

Let  $\mathbb{X}$  be a random variable with probability mass function

$$p(\mathbb{X} = x_i) = p_i \quad i = 1, \dots, m,$$

Then, the entropy of  $\mathbb{X}$  is defined as

$$H = - \sum_{i=1}^m p_i \log p_i$$

Question: Which pmf has maximum entropy?

To answer, let's solve the optimization problem:

$$\max H$$

$$\text{s.t. } \begin{aligned} \sum p_i &= 1, \\ p_i &\geq 0 \end{aligned}$$

## Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

Lets ignore the inequality constraint for now and go back later.

The Lagrangian is

$$L = - \sum p_i \log p_i + \lambda(\sum p_i - 1)$$

The necessary conditions are

$$\frac{\partial L}{\partial p_i} = 0 \quad i = 1, \dots, m$$

$$\frac{\partial L}{\partial \lambda} = 0$$

where

$$\frac{\partial L}{\partial p_i} = -\log p_i - 1 + \lambda = 0 \quad i = 1, \dots, m$$

$$\implies \lambda = 1 + \log p_i$$

$$\implies \log p_i = \lambda - 1.$$

## Lagrange Multipliers: Example 18.4.7 (Maximum Entropy)

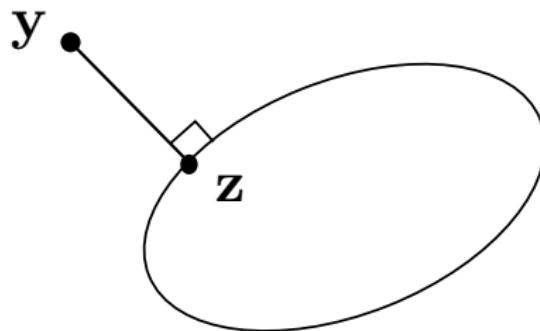
So  $p_i$  must be constant for all  $i$ . The constant  $\sum p_i = 1 \Rightarrow p_i = \frac{1}{n} \geq 0$ , so  $p_i$  satisfies the inequality constraint.

In other words: the uniform pmf maximizes entropy.

In other words: all possibilities are equally likely.

## Lagrange Multipliers: Example: Constrained Least Squares

Given an ellipsoid and a point  $y$  outside the ellipsoid, find the point  $z$  in the ellipsoid nearest to  $y$ .



The equation for an ellipsoid is given by

$$E = \{z \in \mathbb{R}^n : z^\top L^\top L z \leq 1\}.$$

## Lagrange Multipliers: Example: Constrained Least Squares

So we need to solve the following constrained optimization problem:

$$\begin{aligned} \min_{z \in \mathbb{R}^n} \quad & \|z - y\| \\ \text{s.t.} \quad & z^\top L^\top L z = 1 \end{aligned}$$

The Lagrangian is

$$L = (y - z)^\top (y - z) + \lambda(z^\top L^\top L z - 1)$$

## Lagrange Multipliers: Example: Constrained Least Squares

The necessary conditions are

$$\begin{aligned}\frac{\partial L}{\partial z} &= -2(y - z) + 2\lambda L^\top L z = 0 \\ \frac{\partial L}{\partial \lambda} &= z^\top L^\top L z - 1 = 0\end{aligned}$$

The first equations gives

$$\begin{aligned}(I + \lambda L^\top L)z &= y \\ \implies z &= (I + \lambda L^\top L)^{-1}y.\end{aligned}$$

Therefore,  $\lambda$  must satisfy

$$g(\lambda) = y^\top (I + \lambda L^\top L)^{-1} L^\top L (I + \lambda L^\top L)^{-1} y = 1$$

which must be solved numerically using a root finding technique using e.g. Newton's method.

## Lagrange Multipliers: Example

Consider the following optimization problem:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^\top A x \\ \text{s.t.} \quad & Bx = c \end{aligned}$$

where  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{m \times n}$ ,  $c \in \mathbb{R}^m$ .

The Lagrangian is

$$L = \frac{1}{2} x^\top A x + \lambda^\top (Bx - c)$$

## Lagrange Multipliers: Example

The necessary conditions are

$$\frac{\partial L}{\partial x} = Ax + B^\top \lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = Bx - c = 0$$

If  $A$  is invertible, then

$$\begin{aligned} x &= -A^{-1}B^\top \lambda \\ \implies -BA^{-1}B^\top \lambda &= c. \end{aligned}$$

If  $BA^{-1}B^\top$  is invertible then

$$\begin{aligned} \lambda &= -(BA^{-1}B^\top)^{-1}c \\ \implies x &= A^{-1}B^\top(BA^{-1}B^\top)^{-1}c \end{aligned}$$

which is the weighted norm pseudo-inverse.

## Section 4

### Sufficient Conditions

## Lagrange Multipliers: Sufficient Conditions

The necessary conditions tell us where a local extremum might exist, but not whether it is a local min, max, or saddle point.

Are there sufficient conditions for constrained optimization problems?

## Lagrange Multipliers: Sufficient Conditions

For the unconstrained problem, we look at the Hessian of  $f$  for sufficient conditions. For unconstrained problems we look at the second derivative of  $L$  with respect to  $x$ .

$$\underbrace{L}_{1 \times 1} = \underbrace{f}_{1 \times 1} + \underbrace{\lambda^\top}_{1 \times m} \underbrace{h}_{m \times 1} = f + \sum_{i=1}^m \lambda_i h_i$$

so

$$\underbrace{\nabla_x L}_{n \times 1} = \underbrace{\nabla_x f}_{n \times 1} + \underbrace{\nabla h}_{n \times m} \underbrace{\lambda}_{m \times 1} = \nabla_x f + \sum_{i=1}^m \lambda_i \nabla_x h_i$$

$$\nabla_{xx}^2 L = \nabla_{xx}^2 f + \sum_{i=1}^m \lambda_i \nabla_{xx}^2 h_i$$

We will drop the  $xx$  notation (unless not obvious) to get

$$\boxed{\nabla^2 L = \nabla^2 f + \sum_{i=1}^m \lambda_i \nabla^2 h_i}$$

# Lagrange Multipliers: Sufficient Conditions

Let  $P(x^*) = \{y \in \mathbb{R}^n \mid \nabla h(x^*)y = 0\}$  be the tangent plane at  $x^*$ .

Theorem (Moon Theorem 18.4)

Let  $f$  and  $h$  be  $C^2$

1. (Necessity) Suppose that  $x^*$  is a local constrained min of  $f$  and that  $x^*$  is regular. Then  $\exists \lambda$  such that

$$\nabla f(x^*) + \nabla h(x^*)\lambda = 0.$$

2. (Sufficiency) If

1.  $h(x^*) = 0$
2.  $\exists \lambda$  such that  $\nabla f(x^*) + \nabla h(x^*)\lambda = 0$
3.  $y^\top \nabla^2 L(x^*)y \geq 0 \quad \forall y \in P(x^*)$

then  $x^*$  is a local constrained min of  $f$ .

## Lagrange Multipliers: Sufficient Conditions: Example

### 18.5.1

$$\max \quad x_1x_2 + x_2x_3 + x_1x_3$$

$$\text{s.t.} \quad x_1 + x_2 + x_3 = 3$$

The Lagrangian is

$$L = x_1x_2 + x_2x_3 + x_1x_3 + \lambda(x_1 + x_2 + x_3 - 3)$$

The necessary conditions are therefore

$$\nabla_x L = \begin{pmatrix} x_2 + x_3 + \lambda \\ x_1 + x_3 + \lambda \\ x_2 + x_1 + \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\nabla_\lambda L = x_1 + x_2 + x_3 - 3 = 0$$

## Lagrange Multipliers: Sufficient Conditions: Example

### 18.5.1

Therefore, we must solve

$$\begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 3 \end{pmatrix}.$$

The solution is:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \lambda \end{pmatrix}^* = \begin{pmatrix} 1 \\ 1 \\ 1 \\ -2 \end{pmatrix}$$

## Lagrange Multipliers: Sufficient Conditions: Example

### 18.5.1

Is the solution a local max?

$$\nabla^2 L = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} + \lambda \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

Note that

$$\text{eig } \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix} = -1, -1, 2$$

and so  $\nabla^2 L$  is indefinite.

However, the sufficient condition requires that we restrict attention to  $P(x^*)$ .

## Lagrange Multipliers: Sufficient Conditions: Example

### 18.5.1

Note that  $\nabla h = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ ,  $\forall x$  and so

$$P(x^*) = \{x \in \mathbb{R}^n \mid x_1 + x_2 + x_3 = 0\}$$

Therefore

$$x \in P \implies x = \begin{pmatrix} x_1 \\ x_2 \\ -(x_1 + x_2) \end{pmatrix}.$$

Restricting attention to  $P$  gives

$$x^\top \nabla^2 L x = -x_1^2 - x_3^2 - (x_1 + x_2)^2 \leq 0.$$

Therefore  $x^*$  is local maximum.

## Lagrange Multipliers: Sufficient Conditions: Example

### 18.5.1

What did we do in this example to check the negative definite condition? We first projected the  $x$  on to the null space of  $\nabla h(x^*)$   
In general we can check condition (3)

$$y^\top \nabla^2 L(x^*) y \geq 0 \quad \forall y \in P(x^*)$$

as follows:

Let  $E$  be an orthonormal basis for  $\mathcal{N}(\nabla h(x^*))$ , then

$$y^\top \nabla^2 L(x^*) y \geq 0 \quad \forall y \in P(x^*) \iff E^\top \nabla^2 L(x^*) E \geq 0.$$

# Lagrange Multipliers: Sufficient Conditions: Example

## 18.5.1

Using Matlab:

```
>> E = null([1,1,1])
```

$$\gg E = \begin{pmatrix} -0.5774 & -0.5774 \\ 0.7887 & -0.2113 \\ -0.2113 & 0.7887 \end{pmatrix}$$

Therefore

$$E^\top \nabla^2 L(x^*) E = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \leq 0,$$

verifying the sufficient condition.

## Lagrange Multipliers

Question: Is there a physical interpretation of Lagrange Multipliers?

In the book (Section 18.6) it is shown that for the optimization problem

$$\min f(x)$$

$$\text{s.t. } h(x) = c$$

where  $c \neq 0$ , and the solution is given by  $x^*(c)$ . If we let  $x^*$  be a function of  $c$  and  $x^* = x^*(0)$ , then

$$\left. \frac{\partial f}{\partial c}(x^*(c)) \right|_{c=0} = -\lambda$$

In other words,  $\lambda$  indicates how  $f$  changes near the optimum as the constraint values are changed.

Another way of looking at it is that the Lagrange multipliers indicate the sensitivity of  $x^*$  to changes in  $h(x)$ , or the steepness of  $f$  along  $h$ .

## Section 5

### Inequality Constraints: Kuhn-Tucker Conditions

# Inequality Constraints

Lets first consider the problem with just inequality constraints, i.e.

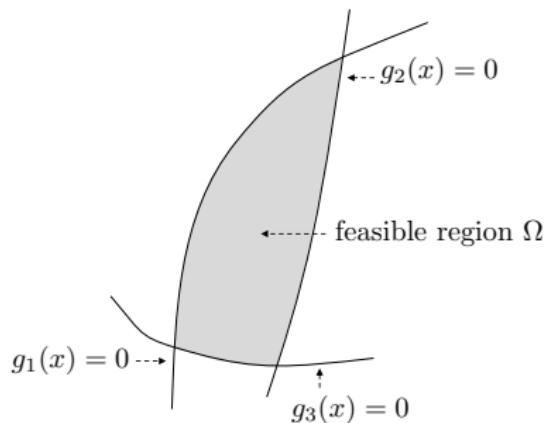
$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & \mathbf{g}(x) \leq 0 \end{aligned}$$

where  $\mathbf{g}(x) \leq 0$  means that

$$\begin{pmatrix} g_1(x) \\ \vdots \\ g_q(x) \end{pmatrix} \leq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

i.e., element-wise.

For example, let  $x \in \mathbb{R}^2$  and let  $q = 3$ .



## Inequality Constraints

Case I. If the local min is in the interior of  $\Omega$ , then clearly

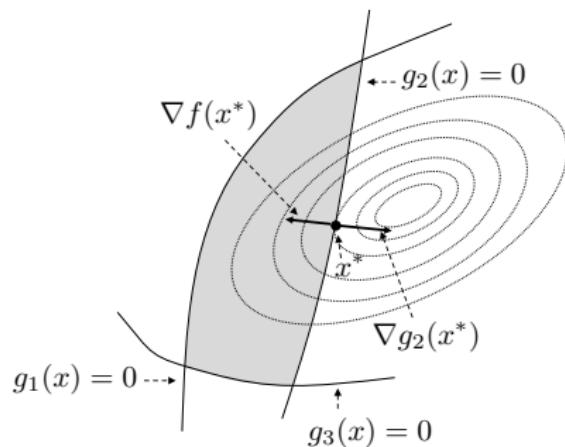
$$\nabla f(x^*) = 0$$

or

$$\nabla f(x^*) + 0 \cdot \nabla g_1(x^*) + 0 \cdot \nabla g_2(x^*) + 0 \cdot g_3(x^*) = 0.$$

# Inequality Constraints

Case II. The local minimum is on the boundary but not at a corner

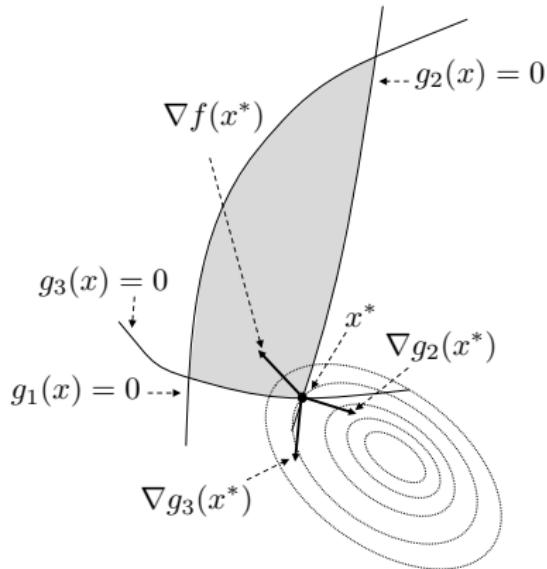


Since in this case  $g_1$  is an equality constraint, we must have that  $\nabla f(x^*) \parallel \nabla g_1(x^*)$ . In fact, in this case the two vectors point in opposite directions! Therefore

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + 0 \cdot \nabla g_2(x^*) + 0 \cdot \nabla g_3(x^*) = 0.$$

# Inequality Constraints

## Case III.



In this case,  $\nabla f(x^*)$  is in the linear span of  $\nabla g_1(x^*)$  and  $\nabla g_2(x^*)$  where the coefficients are negative. Therefore

$$\nabla f(x^*) + \mu_1 \nabla g_1(x^*) + \mu_2 \nabla g_2(x^*) + 0 \cdot g_3(x^*) = 0$$

where  $\mu_1 > 0$  and  $\mu_2 > 0$ .

## Inequality Constraints

In general, for inequality constraints at a local minimum  $x^*$  we have that

1.  $\nabla f(x^*) + \nabla \mathbf{g}(x^*)\mu = 0$
2.  $\mathbf{g}(x^*)^\top \mu = 0$
3.  $\mu \geq 0$

Conditions (1) and (3) together mean that  $\nabla f(x^*)$  is contained in the (negative) linear span of  $\{\nabla g_1(x^*), \dots, \nabla g_q(x^*)\}$ .

Condition (2): Note that if the constraint is active, i.e.  $g_i(x^*) = 0$  then  $\mu_i$  can be nonzero, but if  $g_i$  is inactive, i.e.  $g_i(x^*) < 0$  then  $\mu_i$  must be zero to satisfy (2).

# Inequality Constraints

Now lets go back to the general constrained optimization problem:

$$\begin{aligned} & \min f(x) \\ \text{s.t. } & \mathbf{h}(x) = 0, \\ & \mathbf{g}(x) \leq 0 \end{aligned}$$

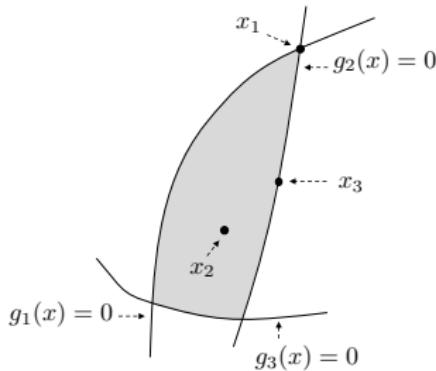
where  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $\mathbf{h}(x) : \mathbb{R}^n \rightarrow \mathbb{R}^p$ ,  $\mathbf{g}(x) : \mathbb{R}^n \rightarrow \mathbb{R}^q$ .

## Definition

$x^*$  is a regular point if  $\nabla h_i(x^*)$ ,  $i = 1, \dots, p$  and  $\nabla g_j(x^*)$  are linearly independent for all  $j = 1, \dots, q$  such that  $g_j(x^*)$  is active.

## Inequality Constraints

For example, suppose that  $\mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}$ , and  $\mathbf{g} = \begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix}$ .



Then  $x^*$  is a regular point at:

- ▶  $x_1$  if  $\{\nabla h_1(x_1), \nabla h_2(x_1), \nabla g_1(x_1), \nabla g_2(x_1)\}$  are linearly independent.
- ▶  $x_2$  if  $\{\nabla h_1(x_2), \nabla h_2(x_2)\}$  are linearly independent.
- ▶  $x_3$  if  $\{\nabla h_1(x_3), \nabla h_2(x_3), \nabla g_1(x_3)\}$  are linearly independent.

# Kuhn Tucker Conditions: Necessary Conditions

## Theorem (Moon Theorem 18.6)

Let  $x^*$  be a regular local minimum, then  $\exists \lambda \in \mathbb{R}^p$  (regular Lagrange multipliers), and  $\exists \mu \in \mathbb{R}^q$ , such that

1.  $\mu \geq 0$  (element wise)
2.  $\mathbf{g}^\top(x^*)\mu = 0$
3.  $\nabla f(x^*) + \nabla \mathbf{h}^\top(x^*)\lambda + \nabla \mathbf{g}^\top(x^*)\mu = 0.$

# Kuhn Tucker Conditions: Sufficient Conditions

## Theorem (Moon 18.7)

Suppose  $f, g, h$  are in  $C_2$ . If there exist  $\lambda \in \mathbb{R}^p, \mu \in \mathbb{R}^q$  such that at  $x^*$

1.  $\mu \geq 0$
2.  $\mathbf{g}^\top(x^*)\mu = 0$
3.  $\nabla f(x^*) + \nabla \mathbf{h}^\top(x^*)\lambda + \nabla \mathbf{g}^\top(x^*)\mu = 0$
4.  $p^\top(\nabla^2 f(x^*) + \sum_{k=1}^p \nabla^2 h_k(x^*)\lambda_k + \sum_{k=1}^q \nabla^2 g_k(x^*)\mu_k)p > 0$

for all  $p$  in the tangent plane of the active constraints, then  $x^*$  is a local constrained minimum.

## Kuhn Tucker Conditions: Example 18.9.1

$$\begin{aligned} \text{min } & 3x_1^2 + 4x_2^2 + 6x_1x_2 - 8x_2 - 6x_1 \\ \text{s.t. } & x_1^2 + x_2^2 - 9 \leq 0, \\ & 2x_1 - x_2 - 4 \leq 0 \end{aligned}$$

The necessary conditions are:

$$\begin{aligned} 6x_1 + 6x_2 - 6 + \mu_1(2x_1) + \mu_2(2) &= 0 \\ 8x_2 + 6x_1 - 8 + \mu_1(2x_2) + \mu_2(-1) &= 0 \\ \mu_1(x_1^2 + x_2^2 - 9) + \mu_2(2x_1 - x_2 - 4) &= 0 \\ \mu_1 \geq 0, \mu_2 \geq 0 \end{aligned}$$

## Kuhn Tucker Conditions: Example 18.9.1

Lets try various combinations of active constraints:

Check inequality constraints:

Case I (Both inactive) i.e.

$$\mu_1 = \mu_2 = 0$$

Therefore, must solve

$$6x_1 + 6x_2 - 6 = 0$$

$$8x_2 + 6x_1 - 8 = 0$$

i.e.,

$$\begin{pmatrix} 6 & 6 \\ 6 & 8 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 6 \\ 8 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$g_1(x) = 1 - 9 = -8 \leq 0$$

$$g_2(x) = -1 - 4 \leq 0$$

Therefore

$$x^* = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \mu^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

satisfies necessary conditions.

Sufficient condition:

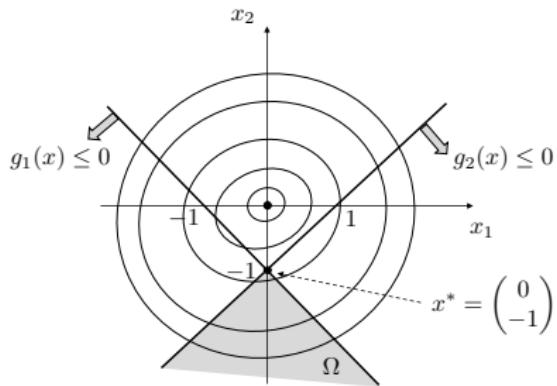
$$\nabla^2 f = \begin{pmatrix} 6 & 6 \\ 6 & 8 \end{pmatrix} > 0$$

implies local minimum.

# Kuhn Tucker Conditions: Example

$$\min \quad x_1^2 + x_2^2$$

$$\text{s.t.} \quad x_1 + x_2 + 1 \leq 0,$$
$$-x_1 + x_2 + 1 \leq 0$$



## Kuhn Tucker Conditions: Example

The necessary conditions are:

$$2x_1 + \mu_1 - \mu_2 = 0$$

$$2x_2 + \mu_1 + \mu_2 = 0$$

$$\mu_1(x_1 + x_2 + 1) + \mu_2(-x_1 + x_2 + 1) = 0$$

$$\mu_1 \geq 0, \mu_2 \geq 0$$

Try various combinations of active constraints **Case 1: (Both inactive)**

$$2x_1 = 0$$

$$2x_2 = 0$$

$$\implies x^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

However, both constraints are violated since

$$g_1^*(x^*) = 1 \geq 0$$

$$g_2(x^*) = 1 \geq 0.$$

## Kuhn Tucker Conditions: Example

Case 2:  $g_1$ -active,  $g_2$ -inactive

$$2x_1 + \mu_1 = 0 \implies x_1 = -\frac{1}{2}\mu_1$$

$$2x_2 + \mu_1 = 0 \implies x_2 = -\frac{1}{2}\mu_1$$

$$\mu_1(x_1 + x_2 + 1) = 0$$

$$\mu_1 > 0$$

Last two equations imply that

$$\mu_1\left(-\frac{1}{2}\mu_1 - \frac{1}{2}\mu_1 + 1\right) = -\mu_1^2 + \mu_1 = \mu_1(1 - \mu_1) = 0.$$

Solving for  $\mu_1$  gives  $\mu_1 = 0$  or  $\boxed{\mu_1 = 1}$ . Therefore

$$x^* = \begin{pmatrix} -\frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

## Kuhn Tucker Conditions: Example

Checking constraints:

$$g_1(x^*) = -\frac{1}{2} - \frac{1}{2} + 1 = 0 \leq 0 \quad \text{ok}$$

$$g_2(x^*) = \frac{1}{2} - \frac{1}{2} + 1 = 1 \geq 0 \quad \text{no}$$

Case 3:  $g_1$ -inactive,  $g_2$ -active Similar results to Case 2.

Case 4: Both active

$$\begin{aligned} & \mu_1\left(\frac{1}{2}\mu_2 - \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 - \frac{1}{2}\mu_1 + 1\right) \\ & + \mu_2\left(-\frac{1}{2}\mu_2 + \frac{1}{2}\mu_1 - \frac{1}{2}\mu_1 - \frac{1}{2}\mu_2 + 1\right) = 0 \\ \implies & \mu_1(1 - \mu_1) + \mu_2(1 - \mu_2) = 0 \end{aligned}$$

## Kuhn Tucker Conditions: Example

A positive solution is

$$\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} > 0$$

which gives

$$x^* = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$$

Constraints can be verified to be satisfied.

Sufficient condition:

$$\nabla^2 f + \nabla^2 g \mu = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} 1 + \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} 1 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} > 0$$

Therefore  $x^*$  is a local minimum.