Linear algebra and applications

Jitkomut Songsiri

Department of Electrical Engineering
Faculty of Engineering
Chulalongkorn University

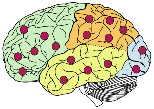
CUEE

Ja 5

Outline

How to read this handout

1. the note is used with lecture in EE205 (you cannot master this topic just by reading this note) – class activities include
   - graphical concepts, math derivation of details/steps in between
   - computer codes to illustrate examples
2. always read 'textbooks' after lecture
3. pay attention to the symbol ✎; you should be able to prove such ✎ result
4. each chapter has a list of references; find more formal details/proofs from in-text citations
5. almost all results in this note can be Googled; readers are encouraged to 'stimulate neurons' in your brain by proving results without seeking help from the Internet first
6. typos and mistakes can be reported to `jitkomut@gmail.com`



System of linear equations

System of linear equations

a linear system of m equations in n variables

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

in matrix form: $Ax = b$

problem statement: given A, b, find a solution x (if exists)

Example: solving ordinary differential equations

given $y(0) = 1, \dot{y}(0) = -1, \ddot{y}(0) = 0$, solve

$$\dddot{y} + 6\ddot{y} + 11\dot{y} + 6y = 0$$

the closed-form solution is

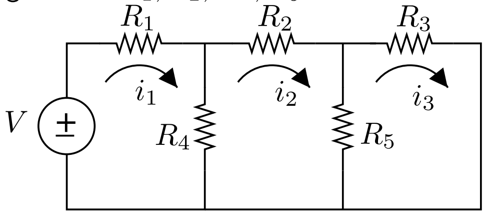
$$y(t) = C_1 e^{-t} + C_2 e^{-2t} + C_3 e^{-3t}$$

C_1, C_2 and C_3 can be found by solving a set of linear equations

$$\begin{array}{rcccl} 1 & = & y(0) & = & C_1 + C_2 + C_3 \\ -1 & = & \dot{y}(0) & = & -C_1 - 2C_2 - 3C_3 \\ 0 & = & \ddot{y}(0) & = & C_1 + 4C_2 + 9C_3 \end{array}$$

Example: linear static circuit

given V, $R_1, R_2, \ldots, R_5$, find the currents in each loop



$$
\begin{aligned}
V &= (R_1 + R_4)i_1 - R_4 i_2 \\
0 &= -R_4 i_1 + (R_2 + R_4 + R_5)i_2 - R_5 i_3 \\
0 &= -R_5 i_2 + (R_3 + R_5)i_3
\end{aligned}
$$

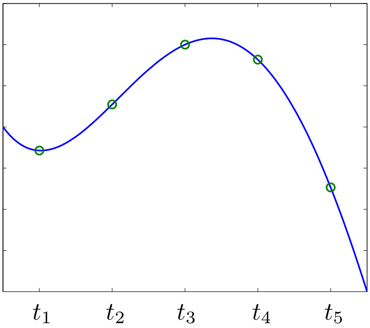
by KVL, we obtain a set of linear equations

Example: polynomial interpolation

fit a polynomial

$$p(t) = x_1 + x_2 t + x_3 t^2 + \cdots + x_n t^{n-1}$$

through n points $(t_1, y_1), \ldots, (t_n, y_n)$



write out the conditions on x:

$$\begin{aligned}
p(t_1) &= x_1 + x_2 t_1 + x_3 t_1^2 + \cdots + x_n t_1^{n-1} \\
p(t_2) &= x_1 + x_2 t_2 + x_3 t_2^2 + \cdots + x_n t_2^{n-1} \\
&\vdots \\
p(t_n) &= x_1 + x_2 t_n + x_3 t_n^2 + \cdots + x_n t_n^{n-1}
\end{aligned}$$

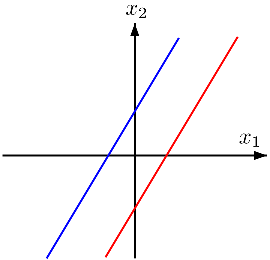
problem data (parameters): $(t_1, y_1), (t_2, y_2), \ldots, (t_n, y_n)$

problem variables: find $x_1, \ldots, x_n$ such that $p(t_i) = y_i$ for all i

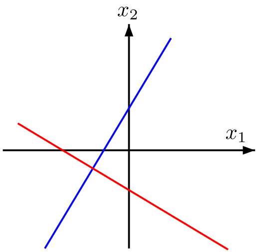
Special case: two variables

Examples:

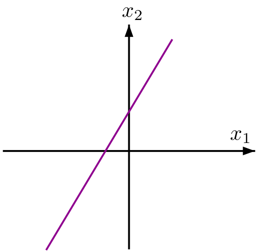
$$\begin{aligned} 2x_1 - x_2 &= -1 \\ 4x_1 - 2x_2 &= 2 \end{aligned} \qquad \begin{aligned} 2x_1 - x_2 &= -1 \\ x_1 + x_2 &= -1 \end{aligned} \qquad \begin{aligned} 2x_1 - x_2 &= -1 \\ 4x_1 - 2x_2 &= -2 \end{aligned}$$



(a) no solution



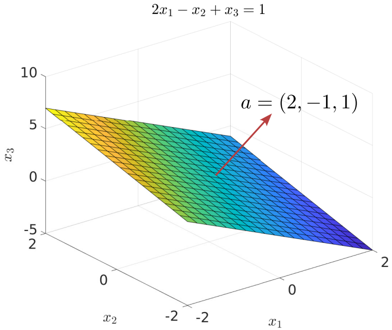
(b) one solution



(c) many solutions

- no solution if two lines are parallel but different interceptions on x_2-axis
- many solutions if the two lines are identical

Geometrical interpretation



the set of solutions to a linear equation

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = b$$

can be interpreted as a hyperplane on $\mathbf{R}^n$

a solution to m linear equations is an **intersection** of m hyperplanes

Three types of linear equations

- **square** if $m = n$ (A is square)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

- **underdetermined** if $m < n$ (A is fat)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

- **overdetermined** if $m > n$ (A is skinny)

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

Existence and uniqueness of solutions

given a system of linear equations **existence:**

- no solution (the linear system is **inconsistent**)
- a solution exists (the linear system is **consistent**)

  uniqueness:

  - the solution is unique
  - there are infinitely many solutions

every system of linear equations has zero, one, or infinitely many solutions

there are no other possibilities

no solution

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + 2x_2 &= 0 \end{aligned} \qquad \begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + x_2 &= -1 \\ x_1 - x_2 &= 2 \end{aligned}$$

unique solution

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 - x_2 &= 0 \end{aligned} \Rightarrow x = (1/3, 2/3) \qquad \begin{aligned} x_1 + x_2 &= 0 \\ 2x_1 + x_2 &= -1 \\ x_1 - x_2 &= -2 \end{aligned} \Rightarrow x = (-1, 1)$$

infinitely many solutions

$$\begin{aligned} x_1 + x_2 &= 1 \\ 2x_1 + 2x_2 &= 2 \end{aligned} \qquad \begin{aligned} x_1 - x_2 + 2x_3 &= 1 \\ -x_1 + x_3 &= -1 \\ 3x_1 - 2x_2 + 3x_3 &= 3 \end{aligned}$$

$$x = (1 - t, t), \qquad x = (1 - t, 3t, t), \qquad t \in \mathbf{R}$$

Elementary row operations

define the **augmented matrix** of the linear equations on page 5 as

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{array}\right]$$

the following operations on the row of the augmented matrix:

1. multiply a row through by a nonzero constant
2. interchange two rows
3. add a constant times one row to another

do not alter the solution set and yield a simpler system

these are called **elementary row operations** on a matrix

Example

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 2 \\ -x_1 + x_2 + x_3 &= -1 \\ 2x_1 - x_2 - 2x_3 &= 3 \end{aligned} \quad \overset{\text{augmented matrix}}{\Longrightarrow} \quad \begin{bmatrix} 1 & 3 & 2 & 2 \\ -1 & 1 & 1 & -1 \\ 2 & -1 & -2 & 3 \end{bmatrix}$$

add the first row to the second $(R_1 + R_2 \to R_2)$

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 2 \\ 4x_2 + 3x_3 &= 1 \\ 2x_1 - x_2 - 2x_3 &= 3 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 4 & 3 & 1 \\ 2 & -1 & -2 & 3 \end{bmatrix}$$

add -2 times the first row to the third $(-2R_1 + R_3 \to R_3)$

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 2 \\ 4x_2 + 3x_3 &= 1 \\ -7x_2 - 6x_3 &= -1 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 4 & 3 & 1 \\ 0 & -7 & -6 & -1 \end{bmatrix}$$

multiply the second row by $1/4$ ($R_2/4 \to R_2$)

$$\begin{array}{rcl} x_1 + 3x_2 + 2x_3 & = & 2 \\ x_2 + \frac{3}{4}x_3 & = & \frac{1}{4} \\ -7x_2 - 6x_3 & = & -1 \end{array} \implies \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 3/4 & 1/4 \\ 0 & -7 & -6 & -1 \end{bmatrix}$$

add 7 times the second row to the third ($7R_2 + R_3 \to R_3$)

$$\begin{array}{rcl} x_1 + 3x_2 + 2x_3 & = & 2 \\ x_2 + \frac{3}{4}x_3 & = & \frac{1}{4} \\ -\frac{3}{4}x_3 & = & \frac{3}{4} \end{array} \implies \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 3/4 & 1/4 \\ 0 & 0 & -3/4 & 3/4 \end{bmatrix}$$

multiply the third row by $-4/3$ ($-4R_3/3 \to R_3$)

$$\begin{array}{rcl} x_1 + 3x_2 + 2x_3 & = & 2 \\ x_2 + \frac{3}{4}x_3 & = & \frac{1}{4} \\ x_3 & = & -1 \end{array} \implies \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 3/4 & 1/4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

add $-3/4$ times the third row to the second $(R_2 - (3/4)R_3 \rightarrow R_2)$

$$\begin{aligned} x_1 + 3x_2 + 2x_3 &= 2 \\ x_2 &= 1 \\ x_3 &= -1 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 3 & 2 & 2 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

add -3 times the second row to the first $(R_1 - 3R_2 \rightarrow R_1)$

$$\begin{aligned} x_1 + 2x_3 &= -1 \\ x_2 &= 1 \\ x_3 &= -1 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 0 & 2 & -1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

add -2 times the third row to the first $(R_1 - 2R_2 \rightarrow R_1)$

$$\begin{aligned} x_1 &= 1 \\ x_2 &= 1 \\ x_3 &= -1 \end{aligned} \quad \Longrightarrow \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}$$

Gaussian elimination

- a systematic procedure for solving systems of linear equations
- based on performing row operations of the augmented matrix
- simplifies the system of equations into an easy form where a solution can be obtained by inspection

Row echelon form

definition: a matrix is in **row echelon form** if

1. a row does not consist entirely of zeros, then the first nonzero number in the row is a 1 (called a **leading 1**)
2. all nonzero rows are above any rows of all zeros
3. in any two successive rows that do not consist entirely of zeros, the leading 1 in the lower row occurs farther to the right than the leading 1 in the higher row

examples:

$$\begin{bmatrix} 1 & 4 & -3 & 5 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 1 & 2 \end{bmatrix}, \quad \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 2 & 5 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Reduced row echelon form

definition: a matrix is in **reduced row echelon form** if

- it is in a row echelon form and
- every leading 1 is the only nonzero entry in its column

examples:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & -2 & 0 & 1 \\ 0 & 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Facts about echelon forms

1. every matrix has a *unique* reduced row echelon form
2. row echelon forms are not unique

$$\text{example:} \quad \left[\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

3. all row echelon forms of a matrix have the same number of zero rows
4. the leading 1's always occur in the same positions in the row echelon forms of a matrix A
5. the columns that contain the leading 1's are called **pivot columns** of A
6. **rank** of A is defined as

$$\text{the number of nonzero rows of (reduced) echelon form of } A$$

Inspecting a solution

- simplify the augmented matrix to the *reduced echelon form*
- read the solution from the reduced echelon form

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \implies 0 \cdot x_3 = 1 \quad \text{(no solution)}$$

$$\begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 5 \end{bmatrix} \implies x_1 = -2, \;\; x_2 = -1, \;\; x_3 = 5 \quad \text{(unique solution)}$$

$$\begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \implies x_1 = 2, \; x_2 = 1 \quad \text{(unique solution)}$$

Leading and free variables

$$\begin{bmatrix} 1 & 0 & 3 & -2 \\ 0 & 1 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{aligned} x_1 + 3x_2 &= -2 \\ x_2 - x_3 &= 1 \end{aligned}$$

definition:

- the corresponding variables to the leading 1's are called **leading variables**
- the remaining variables are called **free variables**

here x_1, x_2 are leading variables and x_3 is a free variable

let $x_3 = t$ and we obtain

$$x_1 = -3t - 2, \quad x_2 = t + 1, \quad x_3 = t$$

(many solutions)

General solution

$$\begin{bmatrix} 1 & -5 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \implies x_1 - 5x_2 + x_3 = 4$$

x_1 is the leading variable, x_2 and x_3 are free variables
let $x_2 = s$ and $x_3 = t$ we obtain

$$\begin{aligned} x_1 &= 5s - t + 4 \\ x_2 &= s \\ x_3 &= t \end{aligned}$$

(many solutions)

by assigning values to s and t, a set of parametric equations:

$$\begin{aligned} x_1 &= 5s - t + 4 \\ x_2 &= s \\ x_3 &= t \end{aligned}$$

is called a **general solution** of the system

Solution to a linear system

solving $b = Ax$ with $A \in \mathbf{R}^{m \times n}$ has only three possibilities

1. **no solution**: if $\mathbf{rank}([A|b]) \neq \mathbf{rank}(A)$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 2 \end{array}\right], \quad \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{array}\right]$$

2. **unique solution**: if $\mathbf{rank}([A|b]) = \mathbf{rank}(A) = n$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right], \quad \left[\begin{array}{cc|c} 1 & 0 & 2 \\ 0 & 2 & 3 \end{array}\right]$$

3. **infinitely many solution**: if $\mathbf{rank}([A|b]) = \mathbf{rank}(A) < n$

$$\left[\begin{array}{ccc|c} 1 & 1 & 3 & 0 \\ 0 & 1 & 2 & -1 \end{array}\right]$$

Gaussian-Jordan elimination

- simplify an augmented matrix to the reduced row echelon form
- inspect the solution from the reduced row echelon form
- the algorithm consists of two parts:
  - **forward phase:** zeros are introduced below the leading 1's
  - **backward phase:** zeros are introduced above the leading 1's

Example

$$\begin{array}{rcl} x_1 + x_2 + 2x_3 & = & 8 \\ -x_1 - 2x_2 + 3x_3 & = & 1 \\ 3x_1 - 7x_2 + 4x_3 & = & 10 \end{array} \implies \begin{bmatrix} 1 & 1 & 2 & 8 \\ -1 & -2 & 3 & 1 \\ 3 & -7 & 4 & 10 \end{bmatrix}$$

use row operations

$$R_1 + R_2 \to R_2 \qquad -3R_1 + R_3 \to R_3 \qquad (-1)\cdot R_2 \to R_2$$

$$\begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 3 & -7 & 4 & 10 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & -1 & 5 & 9 \\ 0 & -10 & -2 & -14 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & -10 & -2 & -14 \end{bmatrix}$$

$$10R_2 + R_3 \to R_3 \qquad R_3/(-52) \to R_3$$

$$\begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & -52 & -104 \end{bmatrix} \quad \begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & -5 & -9 \\ 0 & 0 & 1 & 2 \end{bmatrix}$$

(a row echelon form)

we have added zero below the leading 1's (forward phase) continue performing row operations

$$
\begin{array}{ccc}
5R_3 + R_2 \rightarrow R_2 & -R_2 + R_1 \rightarrow R_1 & -2R_3 + R_1 \rightarrow R_1 \\
\begin{bmatrix} 1 & 1 & 2 & 8 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 2 & 7 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix} & \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}
\end{array}
$$

(reduced echelon form)

we have added zero above the leading 1's (backward phase)

from the reduced echelon form, $\mathbf{rank}([A|b]) = \mathbf{rank}(A) = n$

the system has a unique solution

$$x_1 = 3, \quad x_2 = 1, \quad x_3 = 2$$

Homogeneous linear systems

definition:

a system of linear equations is said to be **homogeneous** if b_j's are all zero

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ \vdots \quad &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0 \end{aligned}$$

- $x_1 = x_2 = \cdots = x_n = 0$ is the **trivial** solution to $Ax = 0$
- if $(x_1, x_2, \ldots, x_n)$ is a solution, so is $(\alpha x_1, \alpha x_2, \ldots, \alpha x_n)$ for any $\alpha \in \mathbf{R}$
- hence, if a solution exists, then the system has infinitely many solutions (by choosing α arbitrarily)
- if z and w are solutions to $Ax = 0$, so is $z + \alpha w$ for any $\alpha \in \mathbf{R}$

example

$$\begin{aligned} x_1 - x_2 + 2x_3 - x_4 &= 0 \\ 2x_1 + x_2 - 2x_3 - 2x_4 &= 0 \\ -x_1 + 2x_2 - 4x_3 + x_4 &= 0 \\ 3x_1 - 3x_4 &= 0 \end{aligned} \implies \begin{bmatrix} 1 & -1 & 2 & -1 & 0 \\ 2 & 1 & -2 & -2 & 0 \\ -1 & 2 & -4 & 1 & 0 \\ 3 & 0 & 0 & -3 & 0 \end{bmatrix}$$

the reduced echelon form is

$$\begin{bmatrix} 1 & 0 & 0 & -1 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \implies \begin{aligned} x_1 - x_4 &= 0 \\ x_2 - 2x_3 &= 0 \end{aligned}$$

define $x_3 = s, x_4 = t$, the parametric equation is

$$x_1 = t, \quad x_2 = 2s, \quad x_3 = s, \quad x_4 = t$$

there are two nonzero rows, so we have two $(n - 2 = 2)$ free variables

Properties of homogeneous linear system

more properties:

- the last column of the augmented matrix is entirely zero (and hence, can be neglected in the augmented matrix)
- if the reduced row echelon form has r *nonzero* rows, then the system has $n - r$ free variables
- a homogeneous linear system with more unknowns than equations has infinitely many solutions

Range space of A

range space of $A \in \mathbf{R}^{m\times n}$ is

$$\begin{aligned} \mathcal{R}(A) &= \{\, y \in \mathbf{R}^m \mid y = Ax, \ \text{ for } x \in \mathbf{R}^n \,\} \\ \mathbf{rank}(A) &\triangleq \text{number of leading 1's in row echelon form of } A \end{aligned}$$

- $y \in \mathcal{R}(A)$ if and only if y is a linear combination of columns in A:

$$y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

- a linear system $y = Ax$ has a solution if and only if $y \in \mathcal{R}(A)$ (existence)
- equivalently, $y = Ax$ has a solution if and only if $\mathbf{rank}(A) = \mathbf{rank}([A \mid y])$

Nullspace of A

nullspace of A is

$$\mathcal{N}(A) = \{ \, x \in \mathbf{R}^n \mid Ax = 0 \, \}$$

example:

$$A = \begin{bmatrix} 2 & -5 & 3 & 0 \\ -2 & -1 & 3 & -1 \\ 5 & -1 & -3 & 2 \end{bmatrix}, \quad \Longrightarrow R = \begin{bmatrix} 1 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 1/4 \\ 0 & 0 & 1 & 1/12 \end{bmatrix}, \quad x = x_4 \begin{bmatrix} -1/2 \\ -1/4 \\ -1/12 \\ 1 \end{bmatrix}, x_4 \in \mathbf{R}$$

uniqueness of solution:

- if the linear system has a solution, the solution is unique if and only if $\mathcal{N}(A) = \{0\}$
- if x_p is a solution to $Ax = b$, and $\mathcal{N}(A) \neq \{0\}$ then a general solution to $Ax = b$ can be expressed as $x = x_p + z$ where $z \in \mathcal{N}(A)$ (infinitely many solutions)

Summary of solving linear systems

for $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^{m \times n}$, the linear system $Ax = b$ has a solution if and only if

$$b \in \mathcal{R}(A) \quad \Longleftrightarrow \quad \mathbf{rank}([A|b]) = \mathbf{rank}(A)$$

if $Ax = b$ has a solution, the uniqueness of the solution in three cases:

- **square A:** the solution is unique $\Leftrightarrow \mathcal{N}(A) \neq \{0\} \Leftrightarrow$ no zero rows in reduced echelon form of A
- **tall A:** the solution is unique $\Leftrightarrow \mathcal{N}(A) \neq \{0\}$
- **fat A:** since $\mathcal{N}(A) \neq \{0\}$ (always), the solutions are never unique

References

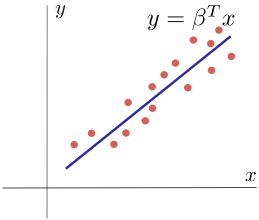
1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011
3. S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least squares*, Cambridge, 2018

Applications of linear equations

Outline

- least-squares problem
- least-norm problem
- numerical methods in solving linear equations

Least-squares problem



setting: find a linear relationship between y_i and $x_{i,k}$

$$y = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p \triangleq x^T \beta$$

given data as y_i and $x_{i1}, x_{i2}, \ldots, x_{ip}$ for $i = 1, 2, \ldots, N$

the data equation in a matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \triangleq \quad y = X\beta$$

problem: given $X \in \mathbf{R}^{m \times n}, y \in \mathbf{R}^m$, solve the linear system for $\beta \in \mathbf{R}^n$

Least-squares: problem statement

overdetermined linear equations:

$$X\beta = y, \quad X \text{ is } m \times n \text{ with } m > n$$

for most y, we cannot solve for β ✎ recall the existence of a solution?

linear least-squares formulation:

$$\underset{\beta}{\text{minimize}} \quad \|y - X\beta\|_2^2 = \sum_{i=1}^{m} (\sum_{j=1}^{n} X_{ij}\beta_j - y_i)^2$$

- $r = y - X\beta$ is called **the residual error**
- β with smallest residual norm $\|r\|$ is called **the least-squares solution**
- it generalizes solving an overdetermined linear system that cannot be solved *exactly* by allowing the system to have the smallest residual

Least-squares: solution

the zero gradient condition of LS objective is

$$\frac{d}{d\beta}\|y - X\beta\|_2^2 = -X^T(y - X\beta) = 0$$

which is equivalent to the **normal equation**

$$X^T X\beta = X^T y$$

if X is **full rank**, it can be shown that $X^T X$ is invertible:

- least-squares solution can be found by solving the normal equations
- n equations in n variables with a positive definite coefficient matrix
- the closed-form solution is $\beta = (X^T X)^{-1} X^T y$
- $(X^T X)^{-1} X^T$ is the **left inverse** of X

Least-squares: data fitting

given data points $\{(t_i, y_i)\}_{i=1}^N$, we aim to approximate y using a function $g(t)$

$$y = g(t) := \beta_1 g_1(t) + \beta_2 g_2(t) + \cdots + \beta_n g_n(t)$$

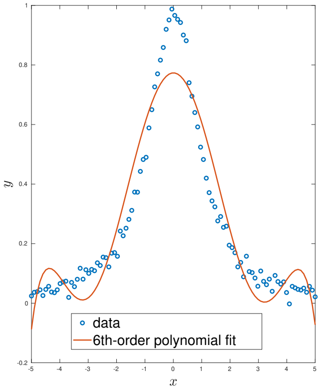
- $g_k(t) : \mathbf{R} \to \mathbf{R}$ is a basis function
  - polynomial functions: $1, t, t^2, \ldots, t^n$
  - sinusoidal functions: $\cos(\omega_k t), \sin(\omega_k t)$ for $k = 1, 2, \ldots, n$
- the linear regression model can be formulated as

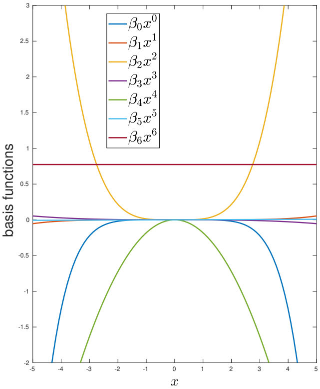
$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} g_1(t_1) & g_2(t_1) & \cdots & g_n(t_1) \\ g_1(t_2) & g_2(t_2) & \cdots & g_n(t_2) \\ \vdots & & & \vdots \\ g_1(t_m) & g_2(t_m) & \cdots & g_n(t_m) \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \triangleq \quad y = X\beta$$

- often have $m \gg n$, *i.e.*, explaining y using a few parameters in the model

Example

fitting a 6th-order polynomial to data points generated from $f(t) = 1/(1+t^2)$





- (right) the weighted sum of basis functions (x^k) is the fitted polynomial
- the ground-truth function f is nonlinear, but can be decomposed as a sum of polynomials

Least-squares: Finite Impulse Response model

given input/output data: $\{(y(t), u(t))\}_{t=0}^{m}$, we aim to estimate FIR model parameters

$$y(t) = \sum_{k=0}^{n-1} h(k)u(t-k)$$

determine $h(0), h(1), \ldots, h(n-1)$ that gives FIR model output closest to y

$$\begin{bmatrix} y(n-1) \\ y(n) \\ \vdots \\ y(m) \end{bmatrix} = \begin{bmatrix} u(n-1) & u(n-2) & \ldots & u(0) \\ u(n) & u(n-1) & \ldots & u(1) \\ \vdots & \vdots & \vdots & \vdots \\ u(m) & u(m-1) & \ldots & u(m-n+1) \end{bmatrix} \begin{bmatrix} h(0) \\ h(1) \\ \vdots \\ h(n-1) \end{bmatrix}$$

- $y(t)$ is a response to $u(t), u(t-1), \ldots, u(t-(n-1))$
- we did not use initial outputs $y(0), y(1), \ldots, y(n-2)$ since there are no historical input data for those outputs

FIR: example

setting: $y(t+1) = ay(t) + bu(t)$, $y(0) = 0$

- relationship between y and u: write the equation recursively

$$\begin{aligned} y(t) &= a^t y(0) + a^{t-1} bu(0) + a^{t-2} bu(1) + \cdots + bu(t-1) \\ &= a^t y(0) + \sum_{\tau=0}^{t-1} a^{t-1-\tau} bu(\tau) \end{aligned}$$

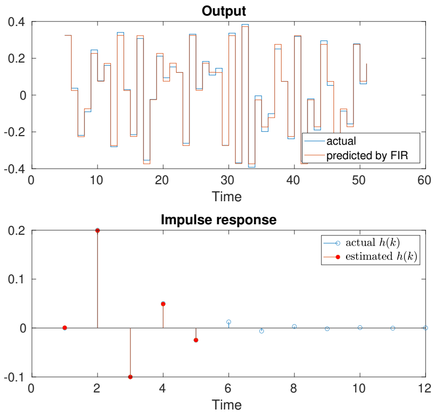
- relate it with the convolution equation: $y(t) = \sum_{k=0}^{\infty} h(k)u(t-k)$

$$h(0) = 0, \quad h(1) = b, \quad h(2) = ab, \quad h(3) = a^2 b, \ldots, \quad h(k) = a^{k-1} b$$

- the actual $h(k)$ decays as k increases but we estimate the first n sequences, *i.e.*, $\hat{h}(0), \hat{h}(1), \ldots, \hat{h}(n-1)$

FIR: example

setting: $a = -0.5, b = 0.2, m = 50, n = 5,$ randomize $u(t) \in \{-1, 1\}$

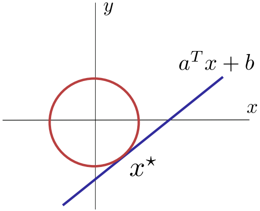


- actual $h(k)$ decays to zero, the first n sequences of $\hat{h}(k)$ are close to actual values
- the predicted output by FIR model is close to the actual output
- $\hat{h}(k)$ is estimated by `A\y` in MATLAB, which returns the least-squares solution

Least-norm problem

setting: given $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ where $m < n$ and A is full row rank

(✎ by assumption, the system $Ax = b$ has many solutions)



the least-norm problem is

$$\underset{x}{\text{minimize}} \quad \|x\|_2 \quad \text{subject to} \quad Ax = b$$

- find a point on hyperplane $Ax = b$ that has the minimum 2-norm
- it extends from solving an underdetermined system that has many solutions but we specifically aim to find the solution with smallest norm

Least-norm solution

the least-norm solution is

$$x^\star = A^T(AA^T)^{-1}y$$

- since A is full rank, it can be shown that AA^T is invertible
- $x^\star$ is linear in y and the coefficient is the **right inverse** of A

Proof. let x be any solution to $Ax = b$

- $x - x^\star$ is always orthogonal to x; by using $A(x - x^\star) = 0$

$$(x - x^\star)^T x^\star = (x - x^\star)^T A^T (AA^T)^{-1} y = (A(x - x^\star))^T (AA^T)^{-1} y = 0$$

- $\|x\|$ is always greater than $\|x^\star\|$, hence $x^\star$ is optimal

$$\|x\|^2 = \|x^\star + x - x^\star\|^2 = \|x^\star\|^2 + \underbrace{(x - x^\star)^T x^\star}_{0} + \|x - x^\star\|^2 \geq \|x^\star\|^2$$

Least-norm application: control system

a first-order dynamical system

$$x(t+1) = ax(t) + bu(t), \quad x \text{ is state, } u \text{ is input}$$

problem: given $a, b \in \mathbf{R}$ with $|a| < 1$ and $x(0)$, find

$$\mathbf{u} = (u(0), u(1), \ldots, u(T-1))$$

such that the values of $x(T), x(T-1)$ are as desired and $\mathbf{u}$ has the minimum 2-norm

background: write $x(t)$ recursively, we found that $x(t)$ is linear in $\mathbf{u}$

$$x(t) = a^t x(0) + a^{t-1} bu(0) + a^{t-2} bu(1) + \cdots + bu(t-1) = a^t x(0) + \sum_{\tau=0}^{t-1} a^{t-1-\tau} bu(\tau)$$

Least-norm application: control system

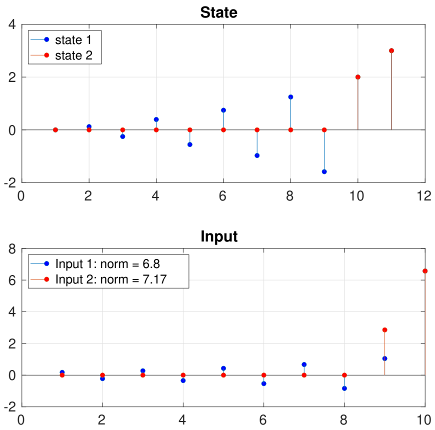
formulate the problem of design $\mathbf{u}$ to drive the state $x(t)$ as desired ✎ verify

$$\begin{bmatrix} x(T) - a^T x(0) \\ x(T-1) - a^{T-1}x(0) \end{bmatrix} = \begin{bmatrix} a^{T-1}b & a^{T-2}b & \cdots & ab & b \\ a^{T-2}b & a^{T-3}b & \cdots & b & 0 \end{bmatrix} \begin{bmatrix} u(0) \\ u(1) \\ \vdots \\ u(T-2) \\ u(T-1) \end{bmatrix} \triangleq y = A\mathbf{u}$$

- regulating the state is a problem of solving an underdetermined system
- A is full row rank, so a solution of $y = A\mathbf{u}$ exists and there are many
- we can try two choices of $\mathbf{u}$:
  1. least-norm solution
  2. any other solution to $y = A\mathbf{u}$

Least-norm application: control system

setting: $a = -0.8, b = 0.7, x(0) = 0, x(T-1) = 2, x(T) = 3$



- different sequences of input drive the state to different paths, but the values of $x(T), x(T-1)$ are as desired
- the least-norm input has the minimum norm – solved by `pinv(A)*y`
- the second choice of input is obtained from `A\y` in MATLAB, which sets many zeros to $\mathbf{u}$ (not the least-norm solution)

Numerical methods in solving linear systems

- solving linear systems by factorization approach
- solving linear systems using softwares
  - square system
  - underdetermined system
  - overdetermined system

Permutation system

a **permutation** matrix P is a square matrix that has exactly one entry of 1 in each row and each column and has zero elsewhere

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

facts: ✎

- P is obtained by interchanging any two rows (or columns) of an identity matrix
- PA results in permuting rows in A, and AP gives permuting columns in A
- $P^TP = I$, so $P^{-1} = P^T$ (simple)
- solving a permuatation system has no cost: $Px = b \Longrightarrow x = P^Tb$

Diagonal system

solve $Ax = b$ when A is diagonal with no zero elements

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

algorithm:

$$\begin{aligned} x_1 &:= b_1/a_{11} \\ x_2 &:= b_2/a_{22} \\ x_3 &:= b_3/a_{33} \\ &\vdots \\ x_n &:= b_n/a_{nn} \end{aligned}$$

cost: n flops

Forward substitution

solve $Ax = b$ when A is lower triangular with nonzero diagonal elements

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

algorithm:

$$\begin{aligned} x_1 &:= b_1/a_{11} \\ x_2 &:= (b_2 - a_{21}x_1)/a_{22} \\ x_3 &:= (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33} \\ &\vdots \\ x_n &:= (b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1})/a_{nn} \end{aligned}$$

cost: $1 + 3 + 5 + \cdots + (2n-1) = n^2$ flops

Back substitution

solve $Ax = b$ when A is upper triangular with nonzero diagonal elements

$$\begin{bmatrix} a_{11} & \cdots & a_{1,n-1} & a_{1n} \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \cdots & a_{n-1,n-1} & a_{n-1,n} \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_{n-1} \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix}$$

algorithm:

$$\begin{aligned} x_n &:= b_n/a_{nn} \\ x_{n-1} &:= (b_{n-1} - a_{n-1,n}x_n)/a_{n-1,n-1} \\ x_{n-2} &:= (b_{n-2} - a_{n-2,n-1}x_{n-1} - a_{n-2,n}x_n)/a_{n-2,n-2} \\ &\vdots \\ x_1 &:= (b_1 - a_{12}x_2 - a_{13}x_3 - \cdots - a_{1n}x_n)/a_{11} \end{aligned}$$

cost: n^2 flops

Factor-solve approach

to solve $Ax = b$, first write A as a product of 'simple' matrices

$$A = A_1 A_2 \cdots A_k$$

then solve $(A_1 A_2 \cdots A_k)x = b$ by solving k equations

$$A_1 z_1 = b, \qquad A_2 z_2 = z_1, \quad \ldots, \quad A_{k-1} z_{k-1} = z_{k-2}, \qquad A_k x = z_{k-1}$$

complexity of factor-solve method: flops $= f + s$

- f is cost of factoring A as $A = A_1 A_2 \cdots A_k$ (factorization step)
- s is cost of solving the k equations for z_1, z_2, ...z_{k-1}, x (solve step)
- usually $f \gg s$

LU decomposition

for a nonsingular A, it can be factorized as (with row pivoting)

$$A = PLU$$

factorization:

- P permutation matrix, L unit lower triangular, U upper triangular
- **factorization cost**: $(2/3)n^3$ if A has order n
- not unique; there may be several possible choices for P, L, U
- interpretation: permute the rows of A and factor P^TA as $P^TA = LU$
- also known as *Gaussian elimination with partial pivoting* (GEPP)

Not every matrix has an LU factor

without row pivoting, LU factor may not exist even when A is invertible

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Rightarrow \quad LU = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

from this example,

- if A could be factored as LU, it would require that $l_{11}u_{11} = a_{11} = 0$
- one of L or U would be singular, contradicting to the fact that $A = LU$ is nonsingular

Solving a linear system with LU factor

solving linear system: $(PLU)x = b$ in three steps

- permutation: $z_1 = P^T b$ (0 flops)
- forward substitution: solve $Lz_2 = z_1$ (n^2 flops)
- back substitution: solve $Ux = z_2$ (n^2 flops)

total cost: $(2/3)n^3 + 2n^2$ flops, or roughly $(2/3)n^3$

Softwares (MATLAB)

1. `A\b`
   - square system: it gives the solution: $x = A^{-1}b$
   - overdetermined system: it gives the solution in the least-square sense
   - underdetermined system: it gives the solution to $Ax = b$ where there are K nonzero elements in x when K is the rank of A
2. `rref(A)`: find the reduced row echelon of A
3. `null(A)`: find independent vectors in the nullspace of A
4. `[L,U,P] = lu(A)`: find LU factorization of A

Softwares (Python)

1. `numpy.linalg.solve`: solves a square system (same for scipy)
2. `numpy.linalg.lstsq`: solves a linear system in least-square sense (same for scipy)
3. `sympy.Matrix`: sympy library for symbolic mathematics
4. `scipy.linalg.null_space`: find independent vectors in the nullspace of A
5. `scipy.linalg.lu`: find LU factorization of A

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011
3. S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least squares*, Cambridge, 2018
4. Lecture notes of EE236, S. Boyd, Stanford
   `https://see.stanford.edu/materials/lsoeldsee263/08-min-norm.pdf`

Matrices

Vector notation

n-vector x:

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

- also written as $x = (x_1, x_2, \ldots, x_n)$
- set of n-vectors is denoted $\mathbf{R}^n$ (Euclidean space)
- x_i: ith **element** or **component** or **entry** of x
- it is common to denote x as a column vector
- $x^T = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}$ is then a row vector

Special vectors

standard unit vector in $\mathbf{R}^n$ is a vector with all zero element except one element which is equal to one

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \quad e_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad e_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

ones vector is the n-vector with all its elements equal to one, denoted as $\mathbf{1}$

stacked vectors: if b, c, d are vectors (can be different sizes)

$$a = \begin{bmatrix} b \\ c \\ d \end{bmatrix}, \quad \text{or} \quad a = (b, c, d)$$

is the *stacked (or concatenated) vector* of b, c, d

Linear combination of vectors

if $a_1, a_2, \ldots, a_m$ are n-vectors, and $\alpha_1, \ldots, \alpha_m$ are scalars, the n-vector

$$\beta_1 a_1 + \beta_2 a_2 + \cdots + \beta_m a_m$$

is called a **linear combination** of the vectors $a_1, \ldots, a_m$

special linear combinations

- any n-vector a can be expressed as $a = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$
- the linear combination with $\beta_1 = \cdots = \beta_m = 1$ given by $a_1 + \cdots + a_m$ is the **sum** of the vectors
- the linear combination with $\beta_1 = \cdots = \beta_m = 1/m$ given by $(a_1 + \cdots + a_m)/m$ is the **average** of the vectors
- when the coefficients are non-negative and sum to one, *i.e.*, $\beta_1 + \cdots + \beta_m = 1$, the linear combination is called a **convex combination** or **weighted average**

Inner products

definition: the inner product of two n-vectors x, y is

$$x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

also known as the **dot product** of vectors x, y

notation: x^Ty

properties ✎

- $(\alpha x)^Ty = \alpha(x^Ty)$ for scalar α
- $(x+y)^Tz = x^Tz + y^Tz$
- $x^Ty = y^Tx$

Examples

- unit vector: $e_i^T a = a_i$ the inner product of a vector with e_i gives the ith element of a
- sum: $\mathbf{1}^T a = a_1 + a_2 + \cdots + a_n$
- average: $(\mathbf{1}/n)^T a = (a_1 + \cdots + a_n)/n$
- sum of squares: $a^T a = a_1^2 + a_2^2 + \cdots + a_n^2$
- selective sum: let b be a vector all of whose entries are either 0 or 1; then $b^T a$ is the sum of elements in a for which $b_i = 1$

$$b = (0, 1, 0, 0, 1), \quad b^T a = a_2 + a_5$$

- polynomial evaluation: let c be the n-vector represents the coefficients of polynomial p with degree $n - 1$

$$p(x) = c_1 + c_2 x + \cdots + c_{n-1} x^{n-2} + c_n x^{n-1}$$

let t be a number and $z = (1, t, t^2, \ldots, t^{n-1})$ then $c^T z = p(t)$

Euclidean norm

$$\|x\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} = \sqrt{x^T x}$$

properties

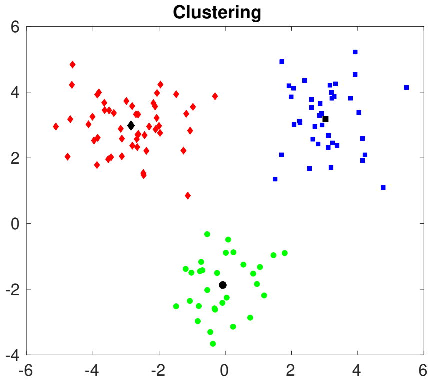
- also written $\|x\|_2$ to distinguish from other norms
- $\|\alpha x\| = |\alpha| \|x\|$ for scalar α
- $\|x + y\| \leq \|x\| + \|y\|$ (triangle inequality)
- $\|x\| \geq 0$ and $\|x\| = 0$ only if $x = 0$

interpretation

- $\|x\|$ measures the *magnitude* or length of x
- $\|x - y\|$ measures the *distance* between x and y

Cluster centroid

given three clusters of data points



it can be shown that the representative is in fact, the **centroid** of the group

$$z_j = \text{argmin}_z \ \ \|x_1 - z\|^2 + \cdots + \|x_N - z\|^2$$

$$z_j = \text{centroid} = \frac{1}{N} \sum_{i \in \text{Group } j} x_i$$

(the average of all points in group G_j)

the black marker is the representative of a cluster, defined by the point that has the smallest sum of distance to all points in a cluster

Inner product and norm of stacked vectors

inner product of stacked vectors

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix}^T \begin{bmatrix} a \\ b \\ c \end{bmatrix} = x^T a + y^T b + z^T c$$

norm of a stacked vector

$$\left\| \begin{bmatrix} x \\ y \\ z \end{bmatrix} \right\|^2 = \|x\|^2 + \|y\|^2 + \|z\|^2$$

norm of a distance

$$\|x - y\|^2 = (x - y)^T (x - y) = \|x\|^2 + \|y\|^2 - 2x^T y$$

Cauchy-Schwarz inequality

for $a, b \in \mathbf{R}^n$

$$|a^T b| \leq \|a\|_2 \|b\|_2$$

example: for $a_1, \ldots, a_n \in \mathbf{R}$ with $a_1 + \cdots + a_n = 1$ show that

$$a_1^2 + a_2^2 + \cdots + a_n^2 \geq \frac{1}{n}$$

CS-inequality can be used to verify the triangle inequality

$$\|a + b\|^2 = \|a\|^2 + 2a^T b + \|b\|^2 \leq \|a\|^2 + 2\|a\|\|b\| + \|b\|^2 = (\|a + b\|)^2$$

angle between vectors: gives a similarity degree of two vectors

$$\cos \theta = \frac{a^T b}{\|a\|\|b\|}$$

Matrix notation

an $m \times n$ matrix A is defined as

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \text{or} \quad A = [a_{ij}]_{m \times n}$$

- a_{ij} are the **elements**, or **coefficients**, or **entries** of A
- set of $m \times n$-matrices is denoted $\mathbf{R}^{m \times n}$
- A has m rows and n columns (m, n are the **dimensions**)
- the (i, j) entry of A is also commonly denoted by A_{ij}
- A is called a **square** matrix if $m = n$

Special matrices

zero matrix: $A = 0$

$$A = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$a_{ij} = 0$, for $i = 1, \ldots, m, j = 1, \ldots, n$

identity matrix: $A = I$

$$A = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

a square matrix with $a_{ii} = 1, a_{ij} = 0$ for $i \neq j$

diagonal matrix: a square matrix with $a_{ij} = 0$ for $i \neq j$

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix}$$

triangular matrix: a square matrix with zero entries in a triangular part

upper triangular

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

lower triangular

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Multiplication

product of $m \times r$-matrix A with $r \times n$-matrix B:

$$(AB)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ir}b_{rj} = \sum_{k=1}^{r} a_{ik}\, b_{kj}$$

dimensions must be compatible: $\#$ of columns in $A = \#$ of rows in B

- $(AB)_{ij}$ is the dot product of the i^{th} row of A and the j^{th} column of B
- $AB \neq BA$ in general ! (even if the dimensions make sense)
- there are exceptions, *e.g.*, $AI = IA$ for all square A
- $A(B+C) = AB + AC$

Matrix transpose

the transpose of an $m \times n$-matrix A is

$$A^T = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

properties ✎

- A^T is $n \times m$
- $(A^T)^T = A$
- $(\alpha A + B)^T = \alpha A^T + B^T, \quad \alpha \in \mathbf{R}$
- $(AB)^T = B^T A^T$
- a square matrix A is called **symmetric** if $A = A^T$, *i.e.*, $a_{ij} = a_{ji}$

Block matrix notation

example: 2×2-block matrix A

$$A = \begin{bmatrix} B & C \\ D & E \end{bmatrix}$$

for example, if B, C, D, E are defined as

$$B = \begin{bmatrix} 2 & 1 \\ 3 & 8 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 7 \\ 1 & 9 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} 0 & 1 \end{bmatrix}, \quad E = \begin{bmatrix} -4 & 1 & -1 \end{bmatrix}$$

then A is the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 & 1 & 7 \\ 3 & 8 & 1 & 9 & 1 \\ 0 & 1 & -4 & 1 & -1 \end{bmatrix}$$

note: dimensions of the blocks must be compatible

Column and Row partitions

write an $m \times n$-matrix A in terms of its columns or its rows

$$A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} b_1^T \\ b_2^T \\ \vdots \\ b_m^T \end{bmatrix}$$

- a_j for $j = 1, 2, \ldots, n$ are the columns of A
- b_i^T for $i = 1, 2, \ldots, m$ are the rows of A

example: $A = \begin{bmatrix} 1 & 2 & 1 \\ 4 & 9 & 0 \end{bmatrix}$

$$a_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}, \quad a_2 = \begin{bmatrix} 2 \\ 9 \end{bmatrix}, \quad a_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad b_1^T = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}, \quad b_2^T = \begin{bmatrix} 4 & 9 & 0 \end{bmatrix}$$

Matrix-vector product

product of $m \times n$-matrix A with n-vector x

$$Ax = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \end{bmatrix}$$

- dimensions must be compatible: $\#$ columns in $A = \#$ elements in x

if A is partitioned as $A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$, then

$$Ax = a_1x_1 + a_2x_2 + \cdots + a_nx_n$$

- Ax is a linear combination of the column vectors of A
- the coefficients are the entries of x

Product with standard unit vectors

post-multiply with a column vector

$$Ae_k = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{mk} \end{bmatrix} = \text{the } k\text{th column of } A$$

pre-multiply with a row vector

$$\begin{aligned} e_k^T A &= \begin{bmatrix} 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \\ &= \begin{bmatrix} a_{k1} & a_{k2} & \cdots & a_{kn} \end{bmatrix} = \text{the } k\text{th row of } A \end{aligned}$$

Trace

definition: trace of a square matrix A is the sum of the diagonal entries in A

$$\mathbf{tr}(A) = a_{11} + a_{22} + \cdots + a_{nn}$$

example:

$$A = \begin{bmatrix} 2 & 1 & 4 \\ 0 & -1 & 5 \\ 3 & 4 & 6 \end{bmatrix}$$

trace of A is $2 - 1 + 6 = 7$

properties ✎

- $\mathbf{tr}(A^T) = \mathbf{tr}(A)$
- $\mathbf{tr}(\alpha A + B) = \alpha\, \mathbf{tr}(A) + \mathbf{tr}(B)$
- $\mathbf{tr}(AB) = \mathbf{tr}(BA)$

Inverse of matrices

definition: a *square* matrix A is called **invertible** or **nonsingular** if there exists B s.t.

$$AB = BA = I$$

- B is called an **inverse** of A
- it is also true that B is invertible and A is an inverse of B
- if no such B can be found A is said to be **singular**

assume A is invertible

- an inverse of A is unique
- the inverse of A is denoted by A^{-1}

Facts about invertible matrices

assume A, B are invertible

facts ✎

- $(\alpha A)^{-1} = \alpha^{-1} A^{-1}$ for nonzero α
- A^T is also invertible and $(A^T)^{-1} = (A^{-1})^T$
- AB is invertible and $(AB)^{-1} = B^{-1} A^{-1}$
- $(A + B)^{-1} \neq A^{-1} + B^{-1}$

✌ **Theorem:** for a square matrix A, the following statements are equivalent

1. A is invertible
2. $Ax = 0$ has only the trivial solution ($x = 0$)
3. the reduced echelon form of A is I
4. A is invertible if and only if $\det(A) \neq 0$

Inverse of 2×2 matrices

the matrix

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

is invertible if and only if

$$ad - bc \neq 0$$

and its inverse is given by

$$A^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

example:

$$A = \begin{bmatrix} 2 & 1 \\ -1 & 3 \end{bmatrix}, \quad A^{-1} = \frac{1}{7} \begin{bmatrix} 3 & -1 \\ 1 & 2 \end{bmatrix}$$

Elementary matrices

Definition: a matrix obtained by performing a *single* row operation on the identity matrix I_n is called an **elementary** matrix

examples:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix}$$ add k times the first row to the third row of I_3

$$\begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix}$$ multiply a nonzero k with the second row of I_2

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$ interchange the second and the third rows of I_3

an elementary matrix is often denoted by E

Inverse operations

row operations on E that produces I and vice versa

$I \to E$	$E \to I$
add k times row i to row j	add $-k$ times row i to row j
multiply row i by $k \neq 0$	multiply row i by $1/k$
interchange row i and j	interchange row i and j

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -k & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} \implies \begin{bmatrix} 1 & 0 \\ 0 & 1/k \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \implies \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Facts ✌

- every elementary matrix is invertible
- the inverse is also an elementary matrix

from the examples in page 87

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ k & 0 & 1 \end{bmatrix} \implies E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -k & 0 & 1 \end{bmatrix}$$

$$E = \begin{bmatrix} 1 & 0 \\ 0 & k \end{bmatrix} \implies E^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/k \end{bmatrix}$$

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \implies E^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Row operations by matrix multiplication

assume A is $m \times n$ and E is obtained by performing a row operation on I_m

$$EA = \text{the matrix obtained by performing this same row operation on } A$$

example:

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

- add -2 times the third row to the second row of A

$$E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \quad EA = \begin{bmatrix} 1 & 2 & 3 \\ -2 & -1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

- multiply 2 with the first row of A

$$E = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad EA = \begin{bmatrix} 2 & 4 & 6 \\ 0 & 1 & -1 \\ 1 & 1 & 0 \end{bmatrix}$$

- interchange the first and the third rows of A

$$E = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad EA = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & -1 \\ 1 & 2 & 3 \end{bmatrix}$$

Inverse via row operations

assume A is invertible

- A is reduced to I by a finite sequence of row operations

$$E_1, E_2, \dots, E_k$$

  such that

$$E_k \cdots E_2 E_1 A = I$$

- the reduced echelon form of A is I
- the inverse of A is therefore given by the product of elementary matrices

$$A^{-1} = E_k \cdots E_2 E_1$$

Example

write the augmented matrix $\left[A \mid I\right]$

$$\left[\begin{array}{ccc|ccc} 2 & 4 & 3 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 1 & 0 & 4 & 0 & 0 & 1 \end{array}\right]$$

and apply row operations until the left side is reduced to I

$$\begin{array}{c} -2R_2 + R_1 \to R_1 \\ -R_2 + R_3 \to R_3 \end{array} \quad \left[\begin{array}{ccc|ccc} 0 & 0 & 1 & 1 & -2 & 0 \\ 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & -2 & 3 & 0 & -1 & 1 \end{array}\right]$$

$$R_1 \leftrightarrow R_2 \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 0 \\ 0 & -2 & 3 & 0 & -1 & 1 \end{array}\right]$$

$$-3R_2 + R_3 \to R_3 \quad \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 0 \\ 0 & -2 & 0 & -3 & 5 & 1 \end{array}\right]$$

$$R_3/(-2) \to R_3 \qquad \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & -2 & 0 \\ 0 & 1 & 0 & \frac{3}{2} & -\frac{5}{2} & -\frac{1}{2} \end{array}\right]$$

$$R_2 \leftrightarrow R_3 \qquad \left[\begin{array}{ccc|ccc} 1 & 2 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & \frac{3}{2} & -\frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 1 & -2 & 0 \end{array}\right]$$

$$-2R_2 + R_1 \to R_1 \qquad \left[\begin{array}{ccc|ccc} 1 & 0 & 1 & -3 & 6 & 1 \\ 0 & 1 & 0 & \frac{3}{2} & -\frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 1 & -2 & 0 \end{array}\right]$$

$$-R_3 + R_1 \to R_1 \qquad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -4 & 8 & 1 \\ 0 & 1 & 0 & \frac{3}{2} & -\frac{5}{2} & -\frac{1}{2} \\ 0 & 0 & 1 & 1 & -2 & 0 \end{array}\right]$$

the inverse of A is

$$\begin{bmatrix} -4 & 8 & 1 \\ \frac{3}{2} & -\frac{5}{2} & -\frac{1}{2} \\ 1 & -2 & 0 \end{bmatrix}$$

Inverse of diagonal matrix

$$A = \begin{bmatrix} a_1 & 0 & \cdots & 0 \\ 0 & a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & a_n \end{bmatrix}$$

a diagonal matrix is invertible iff the diagonal entries are all nonzero

$$a_{ii} \neq 0, \quad i = 1, 2, \ldots, n$$

the inverse of A is given by

$$A^{-1} = \begin{bmatrix} 1/a_1 & 0 & \cdots & 0 \\ 0 & 1/a_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & 1/a_n \end{bmatrix}$$

the diagonal entries in A^{-1} are the inverse of the diagonal entries in A

Inverse of triangular matrix

upper triangular

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

$a_{ij} = 0$ for $i \geq j$

lower triangular

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

$a_{ij} = 0$ for $i \leq j$

a triangular matrix is invertible iff the diagonal entries are all nonzero

$$a_{ii} \neq 0, \quad \forall i = 1, 2, \ldots, n$$

- product of lower (upper) triangular matrices is lower (upper) triangular
- the inverse of a lower (upper) triangular matrix is lower (upper) triangular

Inverse of symmetric matrix

symmetric matrix: $A = A^T$

✎

- for any square matrix A, AA^T and A^TA are always symmetric
- if A is symmetric and invertible, then A^{-1} is symmetric
- if A is invertible, then AA^T and A^TA are also invertible

for a general A, the inverse of A^T is $(A^{-1})^T$ please verify ✎

Determinants

the determinant is a *scalar value* associated with a square matrix A

commonly denoted by $\det(A)$ or $|A|$

determinants of 2×2 matrices:

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

determinants of 3×3 matrices: let $A = \{a_{ij}\}$

$$\det \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

$$- (a_{31}a_{22}a_{13} + a_{32}a_{23}a_{11} + a_{33}a_{21}a_{12})$$

How to find determinants

for a square matrix of any order, it can be computed by

- cofactor expansion
- performing elementray row operations

Minor and Cofactor

Minor of entry a_{ij}: denoted by M_{ij}

- the determinant of the resulting submatrix after deleting the ith row and jth column of A

Cofactor of entry a_{ij}: denoted by C_{ij}

- $C_{ij} = (-1)^{(i+j)} M_{ij}$

example:

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 5 & 0 & 2 \\ 1 & -1 & 2 \end{bmatrix}, \quad M_{23} = \begin{vmatrix} 3 & 1 \\ 1 & -1 \end{vmatrix} = -4, \quad C_{23} = (-1)^{(2+3)} M_{23} = 4$$

Determinants by Cofactor Expansion

Theorem: the determinant of an $n \times n$-matrix A is given by

$$\begin{aligned} \det(A) &= a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj} \\ \det(A) &= a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in} \end{aligned}$$

regardless of which row or column of A is chosen

example: pick the first row to compute $\det(A)$

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 5 & 0 & 2 \\ 1 & -1 & 2 \end{bmatrix}, \quad \det(A) = a_{11}C_{11} + a_{12}C_{12} + a_{13}C_{13}$$

$$\begin{aligned} \det(A) &= 3(-1)^2 \begin{vmatrix} 0 & 2 \\ -1 & 2 \end{vmatrix} + 1(-1)^3 \begin{vmatrix} 5 & 2 \\ 1 & 2 \end{vmatrix} - 2(-1)^4 \begin{vmatrix} 5 & 0 \\ 1 & -1 \end{vmatrix} \\ &= 3(1)(2) + (-1)(8) - 2(1)(-5) = 8 \end{aligned}$$

Basic properties of determinants

✌ let A, B be any square matrices

1. $\det(A) = \det(A^T)$
2. if A has a row of zeros or a column of zeros, then $\det(A) = 0$
3. $\det(\alpha A) = \alpha^n \det(A), \quad \alpha \neq 0$
4. If A has two rows (columns) that are equal, then $\det(A) = 0$
5. $\det(A + B) \neq \det(A) + \det(B)$!
6. $\det(AB) = \det(A) \det(B)$
7. $\det(A^{-1}) = 1/\det(A)$
8. A is invertible if and only if $\det(A) \neq 0$

Basic properties of determinants

suppose the following is true

- A and B are equal except for the entries in their kth row (column)
- C is defined as that matrix identical to A and B except that its kth row (column) is the sum of the kth rows (columns) of A and B

then we have

$$\det(C) = \det(A) + \det(B)$$

example:

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 1 & 2 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 3 & 0 & 2 \end{bmatrix}, \quad C = \begin{bmatrix} 1 & 0 & 1 \\ 2 & 1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

$$\det(A) = 0, \quad \det(B) = -1, \quad \det(C) = -1$$

Determinants of special matrices

- the determinant of a diagonal or triangular matrix is given by the product of the diagonal entries
- $\det(I) = 1$

(these properties can be proved from the def. of cofactor expansion)

Determinants under row operations

- multiply k to a row or a column

$$\begin{vmatrix} ka_{11} & ka_{12} & ka_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = k \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

- interchange between two rows or two columns

$$\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = - \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

- add k times the ith row (column) to the jth row (column)

$$\begin{vmatrix} a_{11} + ka_{21} & a_{12} + ka_{22} & a_{13} + ka_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

Example

B is obtained by performing the following operations on A

$$R_2 + 3R_1 \to R_2, \quad R_3 \leftrightarrow R_1, \quad -4R_1 \to R_1$$

$$A = \begin{bmatrix} 2 & 3 & -2 \\ 3 & 1 & 0 \\ -3 & -3 & 3 \end{bmatrix} \implies \det(B) = (-4) \cdot (-1) \cdot 1 \cdot \det(A)$$

the changes of det. under elementary operations lead to obvious facts ✎

- $\det(\alpha A) = \alpha^n \det(A), \quad \alpha \neq 0$
- If A has two rows (columns) that are equal, then $\det(A) = 0$

Determinants of elementary matrices

let B be obtained by performing a row operation on A then

$$B = EA \quad \text{and} \quad \det(B) = \det(EA)$$

$$E = \begin{bmatrix} k & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \det(B) = k\det(A) \quad (\det(E) = k)$$

$$E = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \det(B) = -\det(A) \quad (\det(E) = -1)$$

$$E = \begin{bmatrix} 1 & k & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \det(B) = \det(A) \quad (\det(E) = 1)$$

conclusion: $\det(EA) = \det(E)\det(A)$

Determinants of product and inverse

✌ let A, B be $n \times n$ matrices

- A is invertible if and only if $\det(A) \neq 0$
- if A is invertible, then $\det(A^{-1}) = 1/\det(A)$
- $\det(AB) = \det(A)\det(B)$

Adjugate formula

the adjugate of A is the transpose of the matrix of cofactors from A

$$\mathrm{adj}(A) = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}$$

if A is invertible then

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

Proof.

- the cofactor expansion using the cofactors from different row is zero

$$a_{i1}C_{k1} + a_{i2}C_{k2} + \ldots + a_{in}C_{kn} = 0, \quad \text{for } i \neq k$$

- $A\,\mathrm{adj}(A) = \det(A) \cdot I$

Cramer's rule

consider a linear system $Ax = b$ when A is **square**

if A is invertible then the solution is unique and given by

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

each component of x can be calculated by using the Cramer's rule

Cramer's rule

$$x_1 = \frac{|A_1|}{|A|}, \quad x_2 = \frac{|A_2|}{|A|}, \quad \ldots \quad , \quad x_n = \frac{|A_n|}{|A|}$$

where A_j is the matrix obtained by replacing b in the jth column of A

(its proof is left as an exercise)

Example

$$A = \begin{bmatrix} 3 & 1 & -2 \\ 5 & 0 & 2 \\ 1 & -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$$

since $\det(A) = 8$, A is invertible and the solution is

$$x = A^{-1}b = \frac{1}{8} \begin{bmatrix} 2 & 0 & 2 \\ -8 & 8 & -16 \\ -5 & 4 & -5 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 1 \\ -5 \\ -2 \end{bmatrix}$$

using Cramer's rule gives

$$x_1 = \frac{1}{8} \begin{vmatrix} 2 & 1 & -2 \\ 1 & 0 & 2 \\ 2 & -1 & 2 \end{vmatrix}, \quad x_2 = \frac{1}{8} \begin{vmatrix} 3 & 2 & -2 \\ 5 & 1 & 2 \\ 1 & 2 & 2 \end{vmatrix}, \quad x_3 = \frac{1}{8} \begin{vmatrix} 3 & 1 & 2 \\ 5 & 0 & 1 \\ 1 & -1 & 2 \end{vmatrix}$$

which yields

$$x_1 = 1, \quad x_2 = -5, \quad x_3 = -2$$

Pseudo-inverse (Penrose Theorem)

one can have a notion of 'inverse' for a non-square matrix

Penrose's Theorem: given $A \in \mathbf{R}^{m \times n}$, there is exactly one $n \times m$ matrix B such that

1. $ABA = A$ and $BAB = B$
2. both AB and BA are symmetric

definition: the **pseudo inverse** of $A \in \mathbf{R}^{m \times n}$ is the unique $n \times m$ matrix $A^\dagger$ such that

1. $AA^\dagger A = A$ and $A^\dagger AA^\dagger = A^\dagger$
2. both $AA^\dagger$ and $A^\dagger A$ are symmetric

Pseudo-inverse

consider a full rank matrix $A \in \mathbf{R}^{m \times n}$ in three cases

- **tall matrix:** A is full rank $\Leftrightarrow$ columns of A are LI $\Leftrightarrow$ $A^T A$ is invertible

$$((A^T A)^{-1} A^T) A = (A^T A)^{-1} (A^T A) = I$$

  the **pseudo-inverse** of A (or left-inverse) is $A^\dagger = (A^T A)^{-1} A^T$

- **wide matrix:** A is full rank $\Leftrightarrow$ row of A are LI $\Leftrightarrow$ AA^T is invertible

$$A(A^T (AA^T)^{-1}) = (AA^T)(AA^T)^{-1} = I$$

  the **pseudo-inverse** of A (or right-inverse) is $A^\dagger = A^T (AA^T)^{-1}$

- **square matrix:** A is full rank $\Leftrightarrow$ A is invertible and both formula of pseudo-inverses reduce to the ordinary inverse A^{-1}

✎ the pseudo inverses of the three cases have the same dimension ?

Example

$$A = \begin{bmatrix} 0 & 2 & 1 \\ -2 & 1 & -2 \end{bmatrix}, \quad A^\dagger = A^T(AA^T)^{-1} = \begin{bmatrix} 0 & -2/9 \\ 2/5 & 1/9 \\ 1/5 & -2/9 \end{bmatrix}$$

$$A = \begin{bmatrix} -2 & -1 \\ 2 & -1 \\ -1 & 0 \end{bmatrix}, \quad A^\dagger = (A^TA)^{-1}A^T = \begin{bmatrix} -2/9 & 2/9 & 1/9 \\ -1/2 & -1/2 & 0 \end{bmatrix}$$

however, when rentangular A has low rank, we can use SVD to find the pseudo inverse

Softwares (MATLAB)

1. `eye(n)` creates an identity matrix of size n
2. `inv(A)` finds the inverse of A (not used for large dimension)
3. `A\eye(n)` finds the inverse of a square matrix A
4. `pinv(A)` gives a pseudoinverse of A, denoted by $A^\dagger$
   - if A is square, a pseudoinverse is the inverse of A
   - if A is tall, $A^\dagger = (A^TA)^{-1}A^T$ is a left inverse of A
   - if A is fat, $A^\dagger = A^T(AA^T)^{-1}$ is a right inverse of A
5. `x = pinv(A)*b` solves the linear system $Ax = b$
   - if A is square, $x = A^{-1}b$
   - if A is tall, x is the solution to the least-square problem: minimize $\|Ax - b\|_2$
   - if A is fat, x is the least-norm solution that satisfies $Ax = b$
6. `det(A)` finds the determinant of A

Softwares (Python)

1. `numpy.eye` creates an identity matrix
2. `numpy.linalg.inv` finds the inverse of a square matrix A
3. `numpy.linalg.pinv` gives a pseudoinverse of A
4. `numpy.linalg.det` find the determinants of A

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least squares*, Cambridge, 2018
3. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011

Eigenvalues and eigenvectors

Linear Independence

Definition: a set of vectors $\{v_1, v_2, \dots, v_n\}$ is **linearly independent** if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

equivalent conditions:

- coefficients of $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ are uniquely determined, i.e.,

  $$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n$$

  implies $\alpha_k = \beta_k$ for $k = 1, 2, \dots, n$
- no vector v_i can be expressed as a linear combination of the other vectors

Examples

- $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\1\\0\end{bmatrix}$ are independent
- $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}$ are independent
- $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\1\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\1\end{bmatrix}, \begin{bmatrix}4\\2\\0\end{bmatrix}$ are not independent
- $\begin{bmatrix}1\\2\\1\end{bmatrix}, \begin{bmatrix}3\\1\\0\end{bmatrix}, \begin{bmatrix}2\\-1\\-1\end{bmatrix}$ are not independent

Linear span

Definition: the linear span of a set of vectors

$$\{v_1, v_2, \dots, v_n\}$$

is the set of all linear combinations of $v_1, \dots, v_n$

$$\mathrm{span}\{v_1, v_2, \dots, v_n\} = \{a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \mid a_1, \dots, a_n \in \mathbf{R}\}$$

example:

$$\mathrm{span}\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$$ is the hyperplane on $x_1 x_2$ plane

Eigenvalues

$\lambda \in \mathbf{C}$ is called an **eigenvalue** of $A \in \mathbf{C}^{n \times n}$ if

$$\det(\lambda I - A) = 0$$

equivalent to:

- there exists nonzero $x \in \mathbf{C}^n$ s.t. $(\lambda I - A)x = 0$, *i.e.*,

  $$Ax = \lambda x$$

  any such x is called an **eigenvector** of A (associated with eigenvalue λ)
- there exists nonzero $w \in \mathbf{C}^n$ such that

  $$w^T A = \lambda w^T$$

  any such w is called a **left eigenvector** of A

Computing eigenvalues

- $\mathcal{X}(\lambda) = \det(\lambda I - A)$ is called the **characteristic polynomial** of A
- $\mathcal{X}(\lambda) = 0$ is called the **characteristic equation** of A
- eigenvalues of A are the root of characteristic polynomial

Computing eigenvalues

- $\mathcal{X}(\lambda) = \det(\lambda I - A)$ is called the **characteristic polynomial** of A
- $\mathcal{X}(\lambda) = 0$ is called the **characteristic equation** of A

the characteristic equation provides a way to compute the eigenvalues of A

$$A = \begin{bmatrix} 5 & 3 \\ -6 & -4 \end{bmatrix}$$

$$\mathcal{X}(\lambda) = \begin{vmatrix} \lambda - 5 & -3 \\ 6 & \lambda + 4 \end{vmatrix} = \lambda^2 - \lambda - 2 = 0$$

solving the characteristic equation gives

$$\lambda = 2, -1$$

Computing eigenvectors

for each eigenvalue of A, we can find an associated eigenvector from

$$(\lambda I - A)x = 0$$

where x is a **nonzero** vector

for A in page 123, let's find an eigenvector corresponding to $\lambda = 2$

$$(\lambda I - A)x = \begin{bmatrix} -3 & -3 \\ 6 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \quad \Longrightarrow \quad x_1 + x_2 = 0$$

the equation has many solutions, so we can form the set of solutions by

$$\left\{ x \in \mathbf{R}^2 \;\middle|\; x = \begin{bmatrix} x_1 \\ -x_1 \end{bmatrix} \right\} = \operatorname{span} \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix} \right\}$$

this set is called the **eigenspace** of A corresponding to $\lambda = 2$

Eigenspace

eigenspace of A corresponding to λ is defined as the nullspace of $\lambda I - A$

$$\mathcal{N}(\lambda I - A)$$

equivalent definition: solution space of the homogeneous system

$$(\lambda I - A)x = 0$$

- an eigenspace is a vector space (by definition)
- 0 is in every eigenspace but it is not an eigenvector
- the *nonzero* vectors in an eigenspace are the eigenvectors of A

from page 124, any nonzero vector lies in the eigenspace is an eigenvector of A, *e.g.*, $x = \begin{bmatrix} -1 & 1 \end{bmatrix}^T$

same way to find an eigenvector associated with $\lambda = -1$

$$(\lambda I - A)x = \begin{bmatrix} -6 & -3 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \quad \Longrightarrow \quad 2x_1 + x_2 = 0$$

so the eigenspace corresponding to $\lambda = -1$ is

$$\left\{ x \;\middle|\; x = \begin{bmatrix} x_1 \\ -2x_1 \end{bmatrix} \right\} = \mathrm{span} \left\{ \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}$$

and $x = \begin{bmatrix} 1 & -2 \end{bmatrix}^T$ is an eigenvector of A associated with $\lambda = -1$

Properties

- if A is $n \times n$ then $\mathcal{X}(\lambda)$ is a polynomial of order n
- if A is $n \times n$ then there are n eigenvalues of A
- even when A is real, eigenvalues and eigenvectors can be complex, *e.g.*,

$$A = \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}, \quad A = \begin{bmatrix} -2 & 0 & 1 \\ -6 & -2 & 0 \\ 19 & 5 & -4 \end{bmatrix}$$

- if A and λ are real, we can choose the associated eigenvector to be real
- if A is real then eigenvalues must occur in complex conjugate pairs
- if x is an eigenvector of A, so is αx for any $\alpha \in \mathbf{C}$, $\alpha \neq 0$
- an eigenvector of A associated with λ lies in $\mathcal{N}(\lambda I - A)$

Important facts

denote $\lambda(A)$ an eigenvalue of A

- $\lambda(\alpha A) = \alpha\lambda(A)$ for any $\alpha \in \mathbf{C}$
- $\mathbf{tr}(A)$ is the sum of eigenvalues of A
- $\det(A)$ is the product of eigenvalues of A
- A and A^T share the same eigenvalues ✎
- $\lambda(\overline{A^T}) = \overline{\lambda(A)}$ ✎
- $\lambda(A^m) = (\lambda(A))^m$ for any integer m
- A is invertible if and only if $\lambda = 0$ is not an eigenvalue of A ✎

Matrix powers

the mth power of a matrix A for a nonnegative integer m is defined as

$$A^m = \prod_{k=1}^{m} A$$

(the multiplication of m copies of A)

and A^0 is defined as the identity matrix, *i.e.*, $A^0 = I$

✌ **Facts:** if λ is an eigenvalue of A with an eigenvector v then

- λ^m is an eigenvalue of A^m
- v is an eigenvector of A^m associated with λ^m

Invertibility and eigenvalues

A is not invertible if and only if there exists a nonzero x such that

$$Ax = 0, \quad \text{or} \quad Ax = 0 \cdot x$$

which implies 0 is an eigenvalue of A

another way to see this is that

$$A \text{ is not invertible} \quad \Longleftrightarrow \quad \det(A) = 0 \quad \Longleftrightarrow \quad \det(0 \cdot I - A) = 0$$

which means 0 is a root of the characteristic equation of A

conclusion ✎ the following statements are equivalent

- A is invertible
- $\mathcal{N}(A) = \{0\}$
- $\lambda = 0$ is not an eigenvalue of A

Eigenvalues of special matrices

diagonal matrix:

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

eigenvalues of D are the diagonal elements, *i.e.*, $\lambda = d_1, d_2, \ldots, d_n$

triangular matrix:

$$U = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \quad L = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

eigenvalues of L and U are the diagonal elements, *i.e.*, $\lambda = a_{11}, \ldots, a_{nn}$

Similarity transform

two $n \times n$ matrices A and B are said to be **similar** if

$$B = T^{-1}AT$$

for some invertible matrix T

T is called a **similarity transform**

✌ **invariant** properties under similarity transform:

- $\det(B) = \det(A)$
- $\mathbf{tr}(B) = \mathbf{tr}(A)$
- A and B have the same eigenvalues

$$\det(\lambda I - B) = \det(\lambda T^{-1}T - T^{-1}AT) = \det(\lambda I - A)$$

Diagonalization

an $n \times n$ matrix A is **diagonalizable** if there exists T such that

$$T^{-1}AT = D$$

is *diagonal*

- similarity transform by T *diagonalizes* A
- A and D are similar, so the entries of D must be the eigenvalues of A

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

- computing A^k is simple because $A^k = (TDT^{-1})^k = TD^kT^{-1}$

Eigenvalue decomposition

if A is diagonalizable then A admits the decomposition

$$A = TDT^{-1}$$

- D is diagonal containing the eigenvalues of A
- columns of T are the corresponding eigenvectors of A
- note that such decomposition is not unique (up to scaling in T)

Theorem: $A \in \mathbf{R}^{n\times n}$ is diagonalizable if and only if all n eigenvectors of A are independent

- a diagonalizable matrix is called a **simple** matrix
- if A is not diagonalizable, sometimes it is called *defective*

Proof (necessity)

suppose $\{v_1, \ldots, v_n\}$ is a *linearly independent* set of eigenvectors of A

$$Av_i = \lambda_i v_i \quad i = 1, \ldots, n$$

we can express this equation in the matrix form as

$$A \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

define $T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$ and $D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$, so

$$AT = TD$$

since T is invertible ($v_1, \ldots, v_n$ are independent), finally we have

$$T^{-1}AT = D$$

Proof (sufficiency)

conversely, if there exists $T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ that diagonalizes A

$$T^{-1}AT = D = \mathbf{diag}(\lambda_1, \ldots, \lambda_n)$$

then $AT = TD$, or

$$Av_i = \lambda_i v_i, \quad i = 1, \ldots, n$$

so $\{v_1, \ldots, v_n\}$ is a linearly independent set of eigenvectors of A

Example

find T that diagonalizes

$$A = \begin{bmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{bmatrix}$$

the characteristic equation is

$$\det(\lambda I - A) = \lambda^3 - 11\lambda^2 + 39\lambda - 45 = 0$$

the eigenvalues of A are $\lambda = 5, 3, 3$
an eigenvector associated with $\lambda_1 = 5$ can be found by

$$(5 \cdot I - A)x = \begin{bmatrix} 1 & 0 & -1 \\ -2 & 2 & -2 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \quad \Longrightarrow \quad \begin{aligned} x_1 - x_3 &= 0 \\ x_2 - 2x_3 &= 0 \end{aligned}$$

x_3 is a free variable

an eigenvector is $v_1 = \begin{bmatrix} 1 & 2 & 1 \end{bmatrix}^T$

next, find an eigenvector associated with $\lambda_2 = 3$

$$(3 \cdot I - A)x = \begin{bmatrix} -1 & 0 & -1 \\ -2 & 0 & -2 \\ -1 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = 0 \quad \Longrightarrow \quad \begin{array}{c} x_1 + x_3 = 0 \\ x_2, x_3 \text{ are free variables} \end{array}$$

the eigenspace can be written by

$$\left\{ x \;\middle|\; x = x_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} = \mathrm{span} \left\{ \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}$$

hence we can find two *independent* eigenvectors

$$v_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad v_3 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$$

corresponding to the repeated eigenvalue $\lambda_2 = 3$

easy to show that v_1, v_2, v_3 are linearly independent

we form a matrix T whose columns are v_1, v_2, v_n

$$T = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$$

then v_1, v_2, v_3 are linearly independent if and only if T is invertible

by a simple calculation, $\det(T) = 2 \neq 0$, so T is invertible

hence, we can use this T to diagonalize A and it is easy to verify that

$$T^{-1}AT = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Not all matrices are diagonalizable

example: $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$

characteristic polynomial is $\det(\lambda I - A) = s^2$, so 0 is the only eigenvalue

eigenvector satisfies $Ax = 0 \cdot x$, *i.e.*,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0 \quad \Longrightarrow \quad \begin{array}{c} x_2 = 0 \\ x_1 \text{ is a free variable} \end{array}$$

so all eigenvectors has form $x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}$ where $x_1 \neq 0$

thus A cannot have *two independent* eigenvectors

Distinct eigenvalues

Theorem: if A has distinct eigenvalues, *i.e.*,

$$\lambda_i \neq \lambda_j, \quad i \neq j$$

then a set of corresponding eigenvectors are *linearly independent*

which further implies that A is diagonalizable

the converse is *false* – A can have repeated eigenvalues but still be diagonalizable

example: all eigenvalues of I are 1 (repeated eigenvalues) but I is diagonal

Proof by contradiction

assume the eigenvectors are dependent

(simple case) let $Ax_k = \lambda_k x_k$, $k = 1, 2$

suppose there exists $\alpha_1, \alpha_2 \neq 0$

$$\alpha_1 x_1 + \alpha_2 x_2 = 0 \qquad (1)$$

multiplying (1) by A: $\quad \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_2 x_2 = 0$

multiplying (1) by λ_1: $\quad \alpha_1 \lambda_1 x_1 + \alpha_2 \lambda_1 x_2 = 0$

subtracting the above from the previous equation

$$\alpha_2 (\lambda_2 - \lambda_1) x_2 = 0$$

since $\lambda_1 \neq \lambda_2$, we must have $\alpha_2 = 0$ and consequently $\alpha_1 = 0$

the proof for a general case is left as an exercise

Algebraic and Geometric multiplicities

algebraic multiplicity of an eigenvalue λ_k is defined as the multiplicity of the root λ_k of the characteristic polynomial

example: the characteristic polynomial of A is

$$\mathcal{X}(\lambda) = (\lambda - \lambda_1)(\lambda - \lambda_2)^2(\lambda - \lambda_3)^5$$

the multiplicity of λ_1, λ_2 and λ_3 are $1, 2$ and 5 respectively

geometric multiplicity of an eigenvalue λ_k is defined as

$$\dim \mathcal{N}(\lambda_k I - A)$$

(the dimension of the corresponding eigenspace)

example: $A = I_n$; the geometric multiplicity of 1 is n

let λ be an eigenvalue of a matrix A $(n \times n)$

Theorem ✌

- the geometric multiplicity of λ is the number of linearly independent eigenvectors associated with λ
- algebraic and geometric multiplicities need not be equal
- let r be the algebraic multiplicity of λ

$$\dim \mathcal{N}(\lambda I - A) \leq r$$

  (the geometric multiplicity is less than or equal to the algebraic multiplicity)
- A is diagonalizable if and only if the geometric multiplicity of every eigenvalue is equal to the algebraic multiplicity

Matrix Power

the mth power of a matrix A for a *nonnegative* m is defined as

$$A^m = \prod_{k=1}^{m} A$$

and define $A^0 = I$

property: $A^r A^s = A^s A^r = A^{r+s}$

a *negative* power of A is defined as

$$A^{-n} = (A^{-1})^n$$

n is a nonnegative integer and A is invertible

Matrix polynomial

a **matrix polynomial** is a polynomial with matrices as variables

$$p(A) = a_0 I + a_1 A + \cdots + a_n A^n$$

for example $A = \begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}$

$$\begin{aligned} p(A) = 2I - 6A + 3A^2 &= 2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 6\begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix} + 3\begin{bmatrix} 2 & 1 \\ 0 & -1 \end{bmatrix}^2 \\ &= \begin{bmatrix} 2 & -3 \\ 0 & 11 \end{bmatrix} \end{aligned}$$

Fact ✎ any two polynomials of A commute, *i.e.*, $p(A)q(A) = q(A)p(A)$

Matrix exponential via diagonalization

suppose A is diagonalizable, *i.e.*, $\Lambda = T^{-1}AT \quad \Longleftrightarrow \quad A = T\Lambda T^{-1}$

where $T = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$, *i.e.*, the columns of T are eigenvectors of A

then we have $A^k = T\Lambda^k T^{-1}$

thus diagonalization simplifies the expression of a matrix polynomial

$$\begin{aligned} p(A) &= a_0 I + a_1 A + \cdots + a_n A^n \\ &= a_0 TT^{-1} + a_1 T\Lambda T^{-1} + \cdots + a_n T\Lambda^n T^{-1} \\ &= Tp(\Lambda)T^{-1} \end{aligned}$$

where

$$p(\Lambda) = \begin{bmatrix} p(\lambda_1) & 0 & \cdots & 0 \\ 0 & p(\lambda_2) & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(\lambda_n) \end{bmatrix}$$

Eigenvectors of matrix polynomial

if λ and v be an eigenvalue and corresponding eigenvector of A then

- $p(\lambda)$ is an eigenvalue of $p(A)$
- v is a corresponding eigenvector of $p(A)$

$$Av = \lambda v \quad \Longrightarrow \quad A^2 v = \lambda A v = \lambda^2 v \quad \cdots \quad \Longrightarrow \quad A^k v = \lambda^k v$$

thus

$$(a_0 I + a_1 A + \cdots + a_n A^n)v = (a_0 v + a_1 \lambda + \cdots + a_n \lambda^n)v$$

which shows that

$$p(A)v = p(\lambda)v$$

Matrix exponential

generalize the exponential function of a scalar

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

to an exponential function of a matrix

define **matrix exponential** as

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots \quad = \sum_{k=0}^{\infty} \frac{A^k}{k!}$$

for a square matrix A

the infinite series converges for all A

Example

example: $A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$

find all powers of A

$$A^2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad A^3 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \ldots, \quad A^k = A \quad \text{for } k = 2, 3, \ldots$$

so by definition,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + \sum_{k=1}^{\infty} \frac{A^k}{k!} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \sum_{k=1}^{\infty} \frac{1}{k!} \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} e & e-1 \\ 0 & 1 \end{bmatrix}$$

never compute e^A by element-wise operation !

$$e^A \neq \begin{bmatrix} e^1 & e^1 \\ e^0 & e^0 \end{bmatrix}$$

Eigenvalues of matrix exponential

✌ if λ and v be an eigenvalue and corresponding eigenvector of A then

- e^{λ} is an eigenvalue of e^A
- v is a corresponding eigenvector of e^A

since e^A can be expressed as power series of A:

$$e^A = I + A + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

multiplying v on both sides and using $A^k v = \lambda^k v$ give

$$\begin{aligned} e^A v &= v + Av + \tfrac{A^2 v}{2!} + \tfrac{A^3 v}{3!} + \cdots \\ &= \left(1 + \lambda + \tfrac{\lambda^2}{2!} + \tfrac{\lambda^3}{3!} + \cdots\right) v \\ &= e^{\lambda} v \end{aligned}$$

Properties of matrix exponential

- $e^0 = I$
- $e^{A+B} \neq e^A \cdot e^B$
- if $AB = BA$, *i.e.*, A and B commute, then $e^{A+B} = e^A \cdot e^B$
- $(e^A)^{-1} = e^{-A}$

✌ these properties can be proved by the definition of e^A

Computing e^A via diagonalization

if A is diagonalizable, *i.e.*,

$$T^{-1}AT = \Lambda = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where λ_k's are eigenvalues of A then e^A has the form

$$e^A = Te^{\Lambda}T^{-1}$$

- computing e^{Λ} is simple since Λ is diagonal
- one needs to find eigenvectors of A to form the matrix T
- the expression of e^A follows from

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \sum_{k=0}^{\infty} \frac{(T\Lambda T^{-1})^k}{k!} = \sum_{k=0}^{\infty} \frac{T\Lambda^k T^{-1}}{k!} = Te^{\Lambda}T^{-1}$$

- if A is diagonalizable, so is e^A

Example

example: compute $f(A) = e^A$ given $A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

the eigenvalues and eigenvectors of A are

$$\lambda_1 = 1,\ v_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \lambda_2 = 2,\ v_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \quad \lambda_3 = 0,\ v_3 = \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}$$

form $T = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}$ and compute $e^A = Te^{\Lambda}T^{-1}$

$$e^A = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} e & 0 & 0 \\ 0 & e^2 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 1 & 1/2 \\ 0 & 0 & 1/2 \end{bmatrix} = \begin{bmatrix} e & e^2 - e & (e^2 - 2e + 1)/2 \\ 0 & e^2 & (e^2 - 1)/2 \\ 0 & 0 & 1 \end{bmatrix}$$

Applications to ordinary differential equations

we solve the following first-order ODEs for $t \geq 0$ where $x(0)$ is given

scalar: $x(t) \in \mathbf{R}$ and $a \in \mathbf{R}$ is given

$$\dot{x}(t) = ax(t)$$

solution: $x(t) = e^{at}x(0)$, for $t \geq 0$

vector: $x(t) \in \mathbf{R}^n$ and $A \in \mathbf{R}^{n \times n}$ is given

$$\dot{x}(t) = Ax(t)$$

solution: $x(t) = e^{At}x(0)$, for $t \geq 0$ (use $\dfrac{de^{At}}{dt} = Ae^{At} = e^{At}A$)

Applications to difference equations

we solve the difference equations for $t = 0, 1, \ldots$ where $x(0)$ is given

scalar: $x(t) \in \mathbf{R}$ and $a \in \mathbf{R}$ is given

$$x(t+1) = ax(t)$$

solution: $x(t) = a^t x(0)$, for $t = 0, 1, 2, \ldots$

vector: $x(t) \in \mathbf{R}^n$ and $A \in \mathbf{R}^{n \times n}$ is given

$$x(t+1) = Ax(t)$$

solution: $x(t) = A^t x(0)$, for $t = 0, 1, 2, \ldots$

Example 1

solve the ODE

$$\ddot{y}(t) - \dot{y}(t) - 6y(t) = 0, \quad y(0) = 1, \dot{y}(0) = 0$$

solution: define

$$x(t) \triangleq \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \triangleq \begin{bmatrix} y(t) \\ \dot{y}(t) \end{bmatrix}$$

write the equation into the vector form $\dot{x}(t) = Ax(t)$

$$\begin{aligned} \dot{x}(t) &= \begin{bmatrix} \dot{y}(t) \\ \ddot{y}(t) \end{bmatrix} = \begin{bmatrix} \dot{y}(t) \\ \dot{y}(t) + 6y(t) \end{bmatrix} \\ &= \begin{bmatrix} x_2(t) \\ 6x_1(t) + x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix} x(t) \end{aligned}$$

the initial condition is

$$x(0) = \begin{bmatrix} y(0) \\ \dot{y}(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Example 1

thus it is left to compute e^{At}

$$A = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix}$$

the eigenvalues and eigenvectors of A are

$$\lambda_1 = -2,\ v_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \lambda_2 = 3,\ v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

all eigenvalues are distinct, so A is diagonalizable and

$$e^{At} = Te^{\Lambda t}T^{-1}, \quad T = \begin{bmatrix} v_1 & v_2 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$e^{At} = \begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} e^{-2t} & 0 \\ 0 & e^{3t} \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix}$$

Example 1

the closed-form expression of e^{At} is

$$e^{At} = \frac{1}{5}\begin{bmatrix} 3e^{-2t} + 2e^{3t} & -e^{-2t} + e^{3t} \\ -6e^{-2t} + 6e^{3t} & 2e^{-2t} + 3e^{3t} \end{bmatrix}$$

the solution to the vector equation is

$$\begin{aligned} x(t) = e^{At}x(0) &= \frac{1}{5}\begin{bmatrix} 3e^{-2t} + 2e^{3t} & -e^{-2t} + e^{3t} \\ -6e^{-2t} + 6e^{3t} & 2e^{-2t} + 3e^{3t} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix} 3e^{-2t} + 2e^{3t} \\ -6e^{-2t} + 6e^{3t} \end{bmatrix} \end{aligned}$$

hence the solution $y(t)$ can be obtained by

$$y(t) = x_1(t) = \begin{bmatrix} 1 & 0 \end{bmatrix} x(t) = \frac{1}{5}\left(3e^{-2t} + 2e^{3t}\right), \quad t \geq 0$$

Example 2

solve the difference equation

$$y(t+2) - y(t+1) - 6y(t) = 0, \quad y(0) = 1, y(1) = 0$$

solution: define

$$x(t) \triangleq \begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix} \triangleq \begin{bmatrix} y(t) \\ y(t+1) \end{bmatrix}$$

write the equation into the vector form $x(t+1) = Ax(t)$

$$\begin{aligned} x(t+1) &= \begin{bmatrix} y(t+1) \\ y(t+2) \end{bmatrix} = \begin{bmatrix} y(t+1) \\ y(t+1) + 6y(t) \end{bmatrix} \\ &= \begin{bmatrix} x_2(t) \\ 6x_1(t) + x_2(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix} x(t) \end{aligned}$$

the initial condition is

$$x(0) = \begin{bmatrix} y(0) \\ y(1) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Example 2

thus it is left to compute A^t

$$A = \begin{bmatrix} 0 & 1 \\ 6 & 1 \end{bmatrix}$$

the eigenvalues and eigenvectors of A are

$$\lambda_1 = -2, \; v_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \quad \lambda_2 = 3, \; v_2 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

all eigenvalues are distinct, so A is diagonalizable and

$$A^t = T\Lambda^t T^{-1}, \quad T = \begin{bmatrix} v_1 & v_2 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

$$A^t = \begin{bmatrix} 1 & 1 \\ -2 & 3 \end{bmatrix} \begin{bmatrix} (-2)^t & 0 \\ 0 & 3^t \end{bmatrix} \frac{1}{5} \begin{bmatrix} 3 & -1 \\ 2 & 1 \end{bmatrix}$$

Example 2

the closed-form expression of A^t is

$$A^t = \frac{1}{5}\begin{bmatrix} 2(3^t)+3(-2)^t & 3^t-(-2)^t \\ 2(3^{t+1})+3(-2)^{t+1} & 3^{t+1}-(-2)^{t+1} \end{bmatrix}$$

for $t = 0, 1, 2, \ldots$

the solution to the vector equation is

$$\begin{aligned} x(t) = A^t x(0) &= \frac{1}{5}\begin{bmatrix} 2(3^t)+3(-2)^t & 3^t-(-2)^t \\ 2(3^{t+1})+3(-2)^{t+1} & 3^{t+1}-(-2)^{t+1} \end{bmatrix}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \\ &= \frac{1}{5}\begin{bmatrix} 2(3^t)+3(-2)^t \\ 2(3^{t+1})+3(-2)^{t+1} \end{bmatrix} \end{aligned}$$

hence the solution $y(t)$ can be obtained by

$$y(t) = x_1(t) = \frac{1}{5}\left(2(3^t)+3(-2)^t\right), \quad t = 0, 1, 2, \ldots$$

Softwares (MATLAB)

1. `[V,D] = eig(A)` produces a diagonal matrix `D` of eigenvalues and a full matrix `V` whose columns are the corresponding eigenvectors
   - the eigenvectors are normalized to have a unit 2-norm
   - eigenvalues are not necessarily sorted by magnitude
2. eigs(A) returns the 6 largest magnitude eigenvalues
3. `expm(A)` computes the matrix exponential e^A
4. `exp(A)` computes the exponential of the entries in A

Softwares (Python)

1. `D,V = numpy.eig(A)` computes the eigenvalues and eigenvectors of A
2. `numpy.linalg.matrix_power(A, n)` computes the n power of A
3. `scipy.linalg.expm(A)` computes the matrix exponential of A
4. `numpy.exp(A)` computes the exponential of the entries of A

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011

Special matrices and applications

Special matrices

- orthogonal matrix
- projection matrix
- permutation matrix
- symmetric matrix
- positive definite matrix

Orthogonal matrix

a *real* matrix $U \in \mathbf{R}^{n\times n}$ is called **orthogonal** if

$$UU^T = U^TU = I$$

properties: ✎

- an orthogonal matrix is special case of unitary for real matrices
- an orthogonal matrix is always invertible and $U^{-1} = U^T$
- columns vectors of U are mutually orthogonal
- norm is preserved under an orthogonal transformation: $\|Ux\|_2^2 = \|x\|_2^2$

example:

$$\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}, \quad \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

Applications

1. rotation: in $\mathbf{R}^3$, rotate a vector x by the angle θ around the z-axis

$$w = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \triangleq U \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

   where U is orthogonal

2. eigenvectors of symmetric matrices are orthogonal (more detail later)
3. Q in QR decomposition is orthogonal
4. orthogonal matrices are used to whiten the data (transform correlated random vector to uncorrelated random vector)
5. discrete Fourier transform (DFT): $y = Wx$ where W is unitary (equivalence of orthogonal matrix in complex)

Unitary matrix

a *complex* matrix $U \in \mathbf{C}^{n\times n}$ is called **unitary** if

$$U^*U = UU^* = I, \qquad (U^* \triangleq \bar{U}^T)$$

example: let $z = e^{-i2\pi/3}$

$$U = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & z & z^2 \\ 1 & z^2 & z^4 \end{bmatrix} = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & 1 & 1 \\ 1 & e^{-i2\pi/3} & e^{-i4\pi/3} \\ 1 & e^{-i4\pi/3} & e^{-i8\pi/3} \end{bmatrix}$$

facts: ✎

- a unitary matrix is always invertible and $U^{-1} = U^*$
- columns vectors of U are mutually orthogonal
- 2-norm is preserved under a unitary transformation: $\|Ux\|_2^2 = (Ux)^*(Ux) = \|x\|_2^2$

Example: Discrete Fourier transform (DFT)

DFT of the length-N time-domain sequence $x[n]$ is defined by

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] e^{-i2\pi kn/N}, \quad 0 \leq k \leq N-1$$

define $z = e^{-i2\pi/N}$, we can write the DFT in a matrix form as

$$\begin{bmatrix} X[0] \\ X[1] \\ X[2] \\ \vdots \\ X[N-1] \end{bmatrix} = \frac{1}{\sqrt{N}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & z^1 & z^2 & \cdots & z^{N-1} \\ 1 & z^2 & z^4 & \cdots & z^{2(N-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & z^{N-1} & z^{2(N-1)} & \cdots & z^{(N-1)(N-1)} \end{bmatrix} \begin{bmatrix} x[0] \\ x[1] \\ x[2] \\ \vdots \\ x[N-1] \end{bmatrix}$$

or $\mathbf{X} = \mathbf{D}\mathbf{x}$ where $\mathbf{D}$ is called the **DFT matrix** and is unitary ($\therefore \mathbf{x} = \mathbf{D}^* \mathbf{X}$)

Unitary property of DFT

the columns of DFT matrix are of the form:

$$\phi_k = (1/\sqrt{N}) \begin{bmatrix} 1 & e^{-i2\pi k/N} & e^{-i2\pi k\cdot 2/N} & \cdots & e^{-i2\pi k(N-1)/N} \end{bmatrix}^T$$

use $\langle \phi_l, \phi_k \rangle = \phi_k^* \phi_l$ and apply the sum of geometric series:

$$\langle \phi_l, \phi_k \rangle = \frac{1}{N} \sum_{n=0}^{N-1} e^{i2\pi(k-l)n/N} = \frac{1}{N} \cdot \frac{1 - e^{i2\pi(k-l)}}{1 - e^{i2\pi(k-l)/N}}$$

the columns of DFT matrix are therefore *orthogonal*

$$\langle \phi_l, \phi_k \rangle = \begin{cases} 1, & \text{for } k = l + rN, \quad r = 0, 1, 2, \ldots \\ 0, & \text{for } k \neq l \end{cases}$$

Projection matrix

$P \in \mathbf{R}^{n \times n}$ is said to be a **projection** matrix if $P^2 = P$ (aka **idempotent**)

- P is a linear transformation from $\mathbf{R}^n$ to a subspace of $\mathbf{R}^n$, denoted as S
- columns of P are the projections of standard basis vectors and S is the range of P
- if P is applied twice on a vector in S, it gives the same vector

examples: identity and

$$\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \quad \begin{bmatrix} 3 & -6 \\ 1 & -2 \end{bmatrix}, \quad I - X(X^TX)^{-1}X^T \quad \text{(in regression)}$$

properties: ✎

- eigenvalues of P are all equal to 0 or 1
- $I - P$ is also idempotent
- if $P \neq I$, then P is singular

Orthogonal projection matrix

a matrix $P \in \mathbf{R}^{n\times n}$ is called an **orthogonal projection** matrix if

$$P^2 = P = P^T$$

properties:

- P is bounded, *i.e.*, $\|Px\| \leq \|x\|$

$$\|Px\|_2^2 = x^T P^T P x = x^T P^2 x = x^T P x \leq \|Px\|\|x\|$$

- if P is an orthogonal projection onto a line spanned by a unit vector u,

$$P = uu^T$$

  (we see that $\mathbf{rank}(P) = 1$ as the dimension of a line is 1)
- another example: $P = X(X^T X)^{-1} X^T$ for any matrix X – (in regression)

Permutation

a **permutation** matrix P is a square matrix that has exactly one entry of 1 in each row and each column and has zero elsewhere

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

facts: ✎

- P is obtained by interchanging any two rows (or columns) of an identity matrix
- PA results in permuting rows in A, and AP gives permuting columns in A
- $P^TP = I$, so $P^{-1} = P^T$ (simple)
- the modulus of all eigenvalues of P is one, *i.e.*, $|\lambda_i(P)| = 1$
- a multiplication of P with vectors or matrix has no flop count (just swap rows/columns)

Linear function

given $w \in \mathbf{R}^n$ and let $x \in \mathbf{R}^n$ be a vector variable

a **linear function** $f : \mathbf{R}^n \to \mathbf{R}$ is given by

$$f(x) = w^T x = w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$$

(✎ review its linear properties, *i.e.*, superposition)

an **affine function** is a linear function plus a constant: $f(x) = w^T x + b$

- $\frac{\partial f}{\partial x_i} = w_i$ gives the rate of change of f in x_i direction
- the set $\{x \mid w^T x + b = \text{ constant }\}$ is a hyperplane in $\mathbf{R}^n$ with the normal vector w
- linear functions are used in linear regression model and linear classifier

Energy form

given a (real) square matrix A, an energy form is a quadratic function of vector x:

$$f : \mathbf{R}^n \to \mathbf{R}, \quad f(x) = x^T A x = \sum_i \sum_j a_{ij} x_i x_j$$

- $x^T A x$ is the same as the energy form using $(A + A^T)/2$ as the coefficient because

$$x^T A x = (x^T A x)^T = \frac{x^T (A + A^T) x}{2}$$

- using $A = \frac{A+A^T}{2} + \frac{A-A^T}{2}$, we can later on assume that an energy form requires only the symmetric part of A
- reverse question: given an energy form, can you determine what A is ?

$$x_1^2 + 2x_2^2 + 3x_3^2 - x_1 x_2 + 2x_2 x_3 \quad \triangleq \quad x^T A x$$

Energy form and completing the square

recall how to complete the square:

$$x_1^2 + 3x_2^2 + 14x_1x_2 = (x_1 + 7x_2)^2 - 46x_2^2$$

given these matrices, expand the energy form and complete the square

$$A = \begin{bmatrix} 4 & 6 \\ 6 & 13 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & 6 \\ 6 & 9 \end{bmatrix}, \quad C = \begin{bmatrix} 4 & 6 \\ 6 & -4 \end{bmatrix}$$

- $x^T A x =$
- $x^T B x =$
- $x^T C x =$

Quadratic function

given $P \in \mathbf{R}^{n\times n}, q \in \mathbf{R}^n, r \in \mathbf{R}$, a **quadratic** function $f : \mathbf{R}^n \to \mathbf{R}$ is of the form

$$f(x) = (1/2)x^T Px + q^T x + r$$

- $x^T Px$ is aka an energy form (due to the quadratic form that appears in the energy/power of some physical variables)

  $$\text{electrical power} = i^2 R, \quad \text{kinetic energy} = \frac{1}{2}mv^2, \quad \text{energy stored in spring} = \frac{1}{2}kx^2$$

- the contour shape of f depends on the property of P (positive definite, indefinite, magnitude of eigenvalues, direction of eigenvectors) – as we will learn shortly

Symmetric matrix

definition: a (real) square matrix A is said to be **symmetric** if $A = A^T$

notation: $A \in \mathbf{S}^n$

examples:

$$\begin{bmatrix} X & Y \\ Y^T & Z \end{bmatrix} \text{ with symmetric } X, Z, \quad A = \mathbf{E}[XX^T] \text{ (correlation matrix)}$$

✎ **basic facts:**

- for any (rectangular) matrix A, AA^T and A^TA are always symmetric
- if A is symmetric and invertible, then A^{-1} is symmetric
- if A is invertible, then AA^T and A^TA are also invertible

Properties of symmetric matrix

spectral theorem: if A is a real symmetric matrix then the following statements hold

1. all eigenvalues of A are real
2. all eigenvectors of A are orthogonal
3. A admits a decomposition

$$A = UDU^T$$

   where $U^TU = UU^T = I$ (U is unitary) and a diagonal D contains $\lambda(A)$

4. for any x, we have

$$\lambda_{\min}(A)\|x\|_2^2 \;\leq\; x^TAx \;\leq\; \lambda_{\max}(A)\|x\|_2^2$$

   the first (and second) inequalities are tight when x is the eigenvector corresponding to $\lambda_{\min}$ (and $\lambda_{\max}$ respectively)

Proofs

1. assume $Ax = \lambda x$ and λ, x could be complex, denote $x^* = \bar{x}^T$

$$\begin{aligned} (x^*Ax)^* &= x^*A^*x = x^*Ax = x^*\lambda x = \lambda x^*x \\ &= (x^*\lambda x)^* = \bar{\lambda}x^*x \end{aligned}$$

since $x^*x \neq 0$, we must have $\lambda = \bar{\lambda}$

2. assume $Ax_1 = \lambda_1 x_1$ and $Ax_2 = \lambda_2 x_2$ (now all (λ_i, x_i) are real)

$$\begin{aligned} x_2^TAx_1 &= x_2^T\lambda_1 x_1 = \lambda_1 x_2^Tx_1 \\ &= x_1^TAx_2 = x_1^T\lambda_2 x_2 = \lambda_2 x_1^Tx_2 \end{aligned}$$

equating two terms give $(\lambda_1 - \lambda_2)x_2^Tx_1 = 0$

for simple case, we can assume that λ_i's are distinct, so $x_2^Tx_1 = 0$ $(x_2 \perp x_1)$

Exercises

1. for $x, y \in \mathbf{R}^n$, are xy^T, xx^T, yx^T symmetric?
2. for a diagonal matrix D, is $D + xx^T$ symmetric?
3. if A, B are symmetric, so is $A + B$?
4. how many distinct entries in a symmetric matrix of size n?
5. if A is symmetric and B is rectangular, is BAB^T symmetric?
6. if A is symmetric and invertible, is A^{-1} symmetric?
7. find conditions on A, B, C, D so that the block matrix: $\begin{bmatrix} A & B \\ C & D \end{bmatrix}$ is symmetric

Positive definite matrix

definition: a symmetric matrix A is **positive semidefinite**, written as $A \succeq 0$ if

$$x^T A x \geq 0, \quad \forall x \in \mathbf{R}^n$$

and is said to be **positive definite**, written as $A \succ 0$ if

$$x^T A x > 0, \quad \text{for all } \textit{nonzero } x \in \mathbf{R}^n$$

✽ the curly $\succeq$ symbol is used with matrices (to differentiate it from $\geq$ for scalars)

example: $A_1 = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \succeq 0$ and $A_2 = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succ 0$ because

$$x^T A_1 x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1^2 + x_2^2 - 2x_1x_2 = (x_1 - x_2)^2 \geq 0$$

$$x^T A_2 x = (x_1 - x_2)^2 + x_2^2 > 0, \quad \forall x \neq 0$$

exercise: ✎ check positive semidefiniteness of matrices on page 178

How to test if $A \succeq 0$?

Theorem: $A \succeq 0$ if and only if all eigenvalues of A are non-negative

($A \succ 0$ if and only if $\lambda(A) > 0$)

Sylvester's criterion: if every principal minor of A (including $\det A$) is non-negative then $A \succeq 0$

proof in Horn Theorem 7.2.5

example 1: $A = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix} \succ 0$ because

- eigenvalues of A are 0.38 and 2.61 (real and positive)
- the principle minors are 1 and $\begin{vmatrix} 1 & -1 \\ -1 & 2 \end{vmatrix} = 1$ (all positive)

example 2: $A = \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \succeq 0$ because eigenvalues of A are 0 and 3

Properties of positive definite matrix

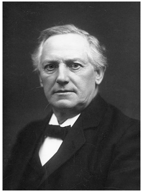
1. if $A \succeq 0$ then all the diagonal terms of A are nonnegative
2. if $A \succeq 0$ then all the leading blocks of A are positive semidefinite
3. if $A \succeq 0$ then $BAB^T \succeq 0$ for any B ✍ (exercise)
4. if $A \succeq 0$ and $B \succeq 0$, then so is $A + B$

Gram matrix

for an $m \times n$ matrix A with columns $a_1, \ldots, a_n$, the product $G = A^T A$ is called the **Gram matrix**

Gram matrix is positive semidefinite

Jørgen Pedersen Gram



$$G = A^T A = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

$$x^T G x = x^T A^T A x = \|Ax\|^2 \geq 0, \quad \forall x$$

- if A has zero nullspace then $Ax = 0 \leftrightarrow x = 0$; this implies that $A^T A \succ 0$
- let X be a data matrix, partitioned in N rows as x_k^T's; we typically encounter $G = \frac{X^T X}{N} = \frac{1}{N} \sum_{k=1}^N x_k x_k^T$ as the **sample covariance matrix**

Exercises

1. check if each of the following is positive definite

$$A_1 = \begin{bmatrix} 2 & 4 \\ 4 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} -2 & 1 \\ 1 & 3 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix}$$

2. is a diagonal matrix always positive semidefinite?
3. for $x \in \mathbf{R}^n$ and I is the identify
   1. is $I + xx^T$ positive semidefinite?
   2. is $I - xx^T$ positive semidefinite?
   3. is xx^T positive semidefinite?
4. find conditions on a, b, c so that

$$\begin{bmatrix} 2 & a & b \\ a & 1 & -1 \\ b & -1 & c \end{bmatrix}$$

is positive definite

Numerical exercises

generate each of these matrices *randomly* and check its properties

1. orthogonal: check determinant and eigenvalues
2. orthogonal projection: check eigenvalues
3. permutation: check the eigenvalues, its inverse and transpose
4. symmetric: check eigenvalues and eigenvectors
5. positive definite: check eigenvalues, eigenvalues of leading diagonal blocks,

relate what you numerically found to the properties of these matrices

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. G. Strang, *Linear Algebra and Learning from Data*, Wellesley-Cambridge Press, 2019

Matrix decomposition

Decompositions

- LU
- Cholesky
- SVD

Factor-solve approach

to solve $Ax = b$, first write A as a product of 'simple' matrices

$$A = A_1 A_2 \cdots A_k$$

then solve $(A_1 A_2 \cdots A_k)x = b$ by solving k equations

$$A_1 z_1 = b, \qquad A_2 z_2 = z_1, \quad \ldots, \quad A_{k-1} z_{k-1} = z_{k-2}, \qquad A_k x = z_{k-1}$$

complexity of factor-solve method: flops $= f + s$

- f is cost of factoring A as $A = A_1 A_2 \cdots A_k$ (factorization step)
- s is cost of solving the k equations for z_1, z_2, ...z_{k-1}, x (solve step)
- usually $f \gg s$

Forward substitution

solve $Ax = b$ when A is lower triangular with nonzero diagonal elements

$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

algorithm:

$$\begin{aligned} x_1 &:= b_1/a_{11} \\ x_2 &:= (b_2 - a_{21}x_1)/a_{22} \\ x_3 &:= (b_3 - a_{31}x_1 - a_{32}x_2)/a_{33} \\ &\vdots \\ x_n &:= (b_n - a_{n1}x_1 - a_{n2}x_2 - \cdots - a_{n,n-1}x_{n-1})/a_{nn} \end{aligned}$$

cost: $1 + 3 + 5 + \cdots + (2n - 1) = n^2$ flops

LU decomposition (w/o row pivoting)

Theorem: if A can be lower reduced (w/o row interchanged) to a row-echelon matrix U, then $A = LU$ where L is lower triangular and invertible and U is upper triangular and row-echelon

- suppose A can be reduced to $A \rightarrow E_1 A \rightarrow E_2 E_1 A \rightarrow E_k E_{k-1} \cdots E_2 E_1 A = U$
- $A = LU$ where $L = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$
  - E_j corresponds to scaling operation or $R_i + \alpha R_j \rightarrow R_i$ for $i > j$
  - E_j is lower triangular (and invertible)
  - E_j^{-1} is also lower triangular, hence, L is lower triangular

Example

find LU for $A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 1 & 2 \\ -1 & 0 & 2 \end{bmatrix}$

$$R_1/2, \quad E_1 = \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_1^{-1} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ -1 & 0 & 2 \end{bmatrix}$$

$$R_2 - R_1 \rightarrow R_2, \quad E_2 = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ -1 & 0 & 2 \end{bmatrix}$$

$$R_3 + R_1 \rightarrow R_3, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}, \quad E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \quad \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 2 & 3 \end{bmatrix}$$

$$R_2/-1 \rightarrow R_2, \quad E_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_4^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 2 & 3 \end{bmatrix}$$

$$R_3 - 2R_2 \rightarrow R_3, \quad E_5 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}, \quad E_5^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}, \quad \Rightarrow \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 5 \end{bmatrix} = U$$

we have $A = E_1^{-1}E_2^{-1}E_3^{-1}E_4^{-1}E_5^{-1}U = \begin{bmatrix} 2 & 0 & 0 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 5 \end{bmatrix}$

each column in L can be read from the leading column in A while performing Gaussian elimination

LU algorithm

let $A \in \mathbf{R}^{m \times n}$ of rank r and suppose A can be lower reduced to U (without row interchanged) then $A = LU$ where the lower triangular, invertible L is constructed as follows

1. if $A = 0$ then $L = I_m$ and $U = 0$
2. if $A \neq 0$, write $A_1 = A$ and let c_1 be the leading column of A_1
3. use c_1 to create the first leading 1 and create zero below it; denote A_2 the matrix consisting of rows 2 to m
4. if $A_2 \neq 0$ let c_2 be the leading column of A_2 and repeat step 2-3 to create A_3
5. continue until U is found where all rows below the last leading 1 consist of zeros; this happen after r steps
6. create L by placing $c_1, c_2, \ldots, c_r$ at the bottom of the first r columns of I_m

Example

find LU for $A = \begin{bmatrix} 2 & 6 & -2 & 0 & 2 \\ 3 & 9 & -3 & 3 & 1 \\ -1 & 3 & 1 & -3 & 1 \end{bmatrix}$

$$R_1/2 \quad \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 3 & 9 & -3 & 3 & 1 \\ -1 & 3 & 1 & -3 & 1 \end{bmatrix}, \qquad R_2 - 3R_1 \rightarrow R_2, R_3 + R_1 \rightarrow R_3 \quad \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 3 & -2 \\ 0 & 0 & 0 & -3 & 2 \end{bmatrix}$$

$$R_2/3 \quad \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2/3 \\ 0 & 0 & 0 & -3 & 2 \end{bmatrix}, \quad R_3 + 3R_2 \rightarrow R_3 \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = U$$

we obtain

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Is LU decomposition unique?

from the previous page

$$A = \begin{bmatrix} 2 & 6 & -2 & 0 & 2 \\ 3 & 9 & -3 & 3 & 1 \\ -1 & 3 & 1 & -3 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = L_1 U_1$$

we can make L **the unit lower triangular** (all diagonals are 1) (standard choice)

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 & 0 & 1 \\ 0 & 0 & 0 & 1 & -2/3 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 3/2 & 1 & 0 \\ -1/2 & -1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 9 & -3 & 0 & 3 \\ 0 & 0 & 0 & 3 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} = L_2 U_2$$

Not every matrix has an LU factor

without row pivoting, LU factor may not exist even when A is invertible

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \Rightarrow \quad LU = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix}$$

from this example,

- if A could be factored as LU, it would require that $l_{11}u_{11} = a_{11} = 0$
- one of L or U would be singular, contradicting to the fact that $A = LU$ is nonsingular

Existence and uniqueness

- **existence**

  Theorem: suppose A is invertible; then A has LU factorization $A = LU$ if and only if all leading principle minors are nonzero

  $$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{is non-singular but has no LU factorization}$$

- **uniqueness**

  Theorem: if an invertible A has an LU factorization then L and U are uniquely determined (if we require the diagonals of L (or U) are all 1)

(Horn, Corollary 3.5.6)

LU decomposition with row pivoting

find LU of $A = \begin{bmatrix} 0 & 0 & -1 \\ -1 & -1 & 1 \\ 2 & 1 & -2 \end{bmatrix}$

- the first row has a leading zero, so row operations require a row interchange, here choose $R_1 \Leftrightarrow R_3$ corresponding to $P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$
- note that $P^2 = I$ (permutation property), we can write

$$A = P^2 A = PPA = P \begin{bmatrix} 2 & 1 & -2 \\ -1 & -1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

- perform LU decomposition on the resulting PA

LU decomposition with row pivoting

- perform $R_1/2,\ R_2 + 2R_1 \to R_1$

$$A = P \begin{bmatrix} 2 & & \\ -1 & 1 & \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & -3/2 \\ 0 & -1/2 & -1/2 \\ 0 & 0 & -1 \end{bmatrix}$$

- perform $R_2 \times -2 \to R_2$

$$A = P \begin{bmatrix} 2 & & \\ -1 & -1/2 & \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & -3/2 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}$$

- perform $R_3 \times -1 \to R_3$

$$A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 2 & & \\ -1 & -\frac{1}{2} & \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & \frac{1}{2} & -\frac{3}{2} \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & -3 \\ 0 & -\frac{1}{2} & -\frac{1}{2} \\ 0 & 0 & -1 \end{bmatrix} \triangleq PLU$$

LU decomposition with row pivoting

same A on page 202 but swap $R_1 \Leftrightarrow R_2$ using $P = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

perform LU decomposition and we get different factors

$$A = \begin{bmatrix} 0 & 0 & -1 \\ -1 & -1 & 1 \\ 2 & 1 & -2 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 3 \\ 0 & 1 & 1/2 \\ 0 & 0 & 9/2 \end{bmatrix}$$

Common pivoting strategy

permute rows so that the largest entry of the first column is on the top left

$$A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 1 & 2 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 2 & 3 \end{bmatrix} \quad \begin{array}{l} R_1/2 \to R_1 \\ R_2 - R_1 \to R_2 \\ R_3 + R_1 \to R_3 \end{array}$$

$$= P_1 P_1 \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} P_1 P_1 \begin{bmatrix} 1 & 2 & 1 \\ 0 & -1 & 1 \\ 0 & 2 & 3 \end{bmatrix} \quad \text{(swap row 2 and 3)}, \; P_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \because P_1^2 = I$$

$$= P_1 \left(P_1 \begin{bmatrix} 2 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} P_1 \right) \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix} = P_1 \begin{bmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 2 & 3 \\ 0 & -1 & 1 \end{bmatrix}$$

$$= P_1 \begin{bmatrix} 2 & 0 & 0 \\ -1 & 2 & 0 \\ 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 \\ 0 & 1 & 3/2 \\ 0 & 0 & 5/2 \end{bmatrix} \quad \begin{array}{l} R_2/2 \to R_2 \\ R_3 + R_2 \to R_3 \end{array}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 3 \\ 0 & 0 & 5/2 \end{bmatrix}$$

Conclusion

any square matrix A can be factorized as (with row pivoting)

$$A = PLU$$

factorization:

- P permutation matrix, L unit lower triangular, U upper triangular
- **factorization cost**: $(2/3)n^3$ if A has order n
- not unique; there may be several possible choices for P, L, U
- interpretation: permute the rows of A and factor P^TA as $P^TA = LU$
- also known as *Gaussian elimination with partial pivoting* (GEPP)

Example

- a singular A (no row pivoting)

$$A = \begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 \\ 0 & 0 \end{bmatrix}$$

- nonsingular A (that requires row pivoting)

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

- nonsingular A (showing two choices of (P, L, U))

$$A = \begin{bmatrix} 2 & 4 & 2 \\ 1 & 1 & 2 \\ -1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ -1/2 & -2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & -1 & 1 \\ 0 & 0 & 5 \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ 1/2 & -1/2 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 3 \\ 0 & 0 & 5/2 \end{bmatrix}$$

Solving a linear system with LU factor

solving linear system: $(PLU)x = b$ in three steps

- permutation: $z_1 = P^T b$ (0 flops)
- forward substitution: solve $Lz_2 = z_1$ (n^2 flops)
- back substitution: solve $Ux = z_2$ (n^2 flops)

total cost: $(2/3)n^3 + 2n^2$ flops, or roughly $(2/3)n^3$

Softwares

MATLAB

- `[L,U,P] = lu(A)` find LU decomposition: $A = P^T LU$ where L is unit lower triangular and U is upper triangular

Python

- `P,L,U = scipy.linalg.lu(A)` find LU decomposition: $A = PLU$ where L is unit lower triangular and U is upper triangular

Exercises

1. find LU factorization (explain if row pivoting is required) and compare the results with coding

$$A_1 = \begin{bmatrix} 2 & 4 & 2 \\ 1 & -1 & 3 \\ -1 & 7 & -7 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 0 & -1 & 2 \\ 0 & 0 & 3 \\ -1 & 2 & 2 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 3 & 3 & 0 \\ -2 & 0 & 2 \\ 3 & 2 & -1 \end{bmatrix}$$

2. suppose we aim to solve $Ax = b^{(k)}$ for $k = 1, \ldots, 1000$ where $A \in \mathbf{R}^{2000\times 2000}$ and $b^{(k)}$'s can be randomized as examples, write computer code to solve the linear system using factor approach and measure the computation time in each process

Cholesky factorization

every positive definite matrix A can be factored as

$$A = LL^T$$

where L is lower triangular with positive diagonal elements

- **cost**: $(1/3)n^3$ flops if A is of order n
- L is called the *Cholesky factor* of A
- can be interpreted as 'square root' of a positive define matrix
- L is invertible (its diagonal elements are nonzero)
- A is invertible and

$$A^{-1} = L^{-T}L^{-1}$$

Cholesky factorization algorithm

partition matrices in $A = LL^T$ as

$$\begin{bmatrix} a_{11} & A_{21}^T \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix} \begin{bmatrix} l_{11} & L_{21}^T \\ 0 & L_{22}^T \end{bmatrix} = \begin{bmatrix} l_{11}^2 & l_{11}L_{21}^T \\ l_{11}L_{21} & L_{21}L_{21}^T + L_{22}L_{22}^T \end{bmatrix}$$

algorithm:

1. determine l_{11} and L_{21}:

$$l_{11} = \sqrt{a_{11}}, \qquad L_{21} = \frac{1}{l_{11}} A_{21}$$

2. compute L_{22} from

$$A_{22} - L_{21}L_{21}^T = L_{22}L_{22}^T$$

   this is a Cholesky factorization of order $n - 1$

Proof of Cholesky algorithm

proof that the algorithm works for positive definite A of order n

- step 1: if A is positive definite then $a_{11} > 0$
- step 2: if A is positive definite, then

$$A_{22} - L_{21}L_{21}^T = A_{22} - \frac{1}{a_{11}}A_{21}A_{21}^T$$

  is positive definite (by Schur complement)
- hence the algorithm works for $n = m$ if it works for $n = m - 1$
- it obviously works for $n = 1$; therefore it works for all n

Example of Cholesky algorithm

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{11} & l_{21} & l_{31} \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

- first column of L

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & l_{22} & 0 \\ -1 & l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & l_{22} & l_{32} \\ 0 & 0 & l_{33} \end{bmatrix}$$

- second column of L

$$\begin{bmatrix} 18 & 0 \\ 0 & 11 \end{bmatrix} - \begin{bmatrix} 3 \\ -1 \end{bmatrix} \begin{bmatrix} 3 & -1 \end{bmatrix} = \begin{bmatrix} l_{22} & 0 \\ l_{32} & l_{33} \end{bmatrix} \begin{bmatrix} l_{22} & l_{32} \\ 0 & l_{33} \end{bmatrix}$$

$$\begin{bmatrix} 9 & 3 \\ 3 & 10 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 1 & l_{33} \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 0 & l_{33} \end{bmatrix}$$

- third column of L: $10 - 1 = l_{33}^2$, *i.e.*, $l_{33} = 3$

conclusion:

$$\begin{bmatrix} 25 & 15 & -5 \\ 15 & 18 & 0 \\ -5 & 0 & 11 \end{bmatrix} = \begin{bmatrix} 5 & 0 & 0 \\ 3 & 3 & 0 \\ -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} 5 & 3 & -1 \\ 0 & 3 & 1 \\ 0 & 0 & 3 \end{bmatrix}$$

Solving equations with positive definite A

$$Ax = b \qquad (A \text{ positive definite of order } n)$$

algorithm

- factor A as $A = LL^T$
- solve $LL^Tx = b$
  - forward substitution $Lz = b$
  - back substitution $L^Tx = z$

cost: $(1/3)n^3$ flops

- factorization: $(1/3)n^3$
- forward and backward substitution: $2n^2$

Softwares

MATLAB

- `U = chol(A)` returns Cholesky decomposition $A = U^T U$ where U is upper triangular

Python

- `L = scipy.linalg.cholesky(A)` returns Cholesky decomposition $A = LL^T$ or $A = U^T U$ where L is lower (`lower=True`) and U is upper triangular

Exercises

1. find Cholesky factorization and compare the results with coding

$$A_1 = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 12 & 4 & 3 \\ 4 & 2 & -1 \\ 3 & -1 & 7 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 20 & 4 & 5 \\ 4 & 2 & 3 \\ 5 & 3 & 5 \end{bmatrix}$$

2. suggest a method to randomize A and guarantee that $A \succ 0$
3. suppose we aim to solve $Ax = b^{(k)}$ for $k = 1, \ldots, 1000$ where $A \in \mathbf{S}_{++}^{2000 \times 2000}$ (pdf) and $b^{(k)}$'s can be randomized as examples, write computer code to solve the linear system using factor approach and measure the computation time in each process

SVD decomposition

- recall that $A^T A \succeq 0$ and eigenvalues are non-negative
- singular values
- left and right singular vectors
- applications: pseudo inverse

Singular values and vectors

let $A \in \mathbf{R}^{m \times n}$, we form eigenvalue problem of A^TA

$$A^TAv_i = \sigma_i^2 v_i, \quad i = 1, 2, \ldots, n$$

- $\sigma_i = \sqrt{\lambda_i(A^TA)} > 0$ is called **singular value** of A
- v_i (orthogonal and has unit-norm) is called **right singular vector**
- fact: if rank of A is r then $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_r > 0$ and $\sigma_i = 0$ for $i > r$

rank of A is the number of non-zero singular values of A

- there exist **left singular vector** $u_1, u_2, \ldots, u_m$ that are orthogonal such that

$$Av_1 = \sigma_1 u_1, \quad Av_2 = \sigma_2 u_2, \ldots, Av_r = \sigma_r u_r, \quad Av_{r+1} = \cdots = Av_n = 0$$

Matrix form

$$Av_1 = \sigma_1 u_1, \quad Av_2 = \sigma_2 u_2, \ldots, \quad Av_r = \sigma_r u_r, \quad Av_{r+1} = \cdots = Av_n = 0$$

or in matrix form: $AV = U\Sigma$ (where U and V are orthogonal matrices)

$$A \left[\begin{array}{ccc|ccc} v_1 & \cdots & v_r & v_{r+1} & \cdots & v_n \end{array}\right] = \left[\begin{array}{ccc|ccc} u_1 & \cdots & u_r & u_{r+1} & \cdots & u_m \end{array}\right] \left[\begin{array}{ccc|c} \sigma_1 & & & 0 \\ & \ddots & & 0 \\ & & \sigma_r & 0 \\ \hline 0 & 0 & 0 & \mathbf{0} \end{array}\right]$$

it can be shown that

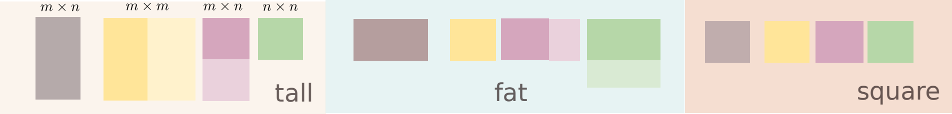
- $v_1, \ldots, v_r, v_{r+1}, \ldots, v_n$ are orthogonal (eigenvectors of $A^T A$, which is symmetric)
- $u_{r+1}, \ldots, u_m$ can be chosen such that $\{u_1, \ldots, u_m\}$ are orgothogonal
- hence, V, U are orthogonal matrices, $V^T V = I, U^T U = I$

unlike eigenvalue decomposition: $AX = X\Lambda$, SVD needs two sets of singular vectors

SVD decomposition

let $A \in \mathbf{R}^{m\times n}$ be a rectangular matrix; there exists the SVD form of A

$$A = U\Sigma V^T$$



$$A \;=\; U \quad \Sigma \quad V^T$$

- $U \in \mathbf{R}^{m\times m}, V \in \mathbf{R}^{n\times n}$ are orthogonal matrices
- $\Sigma \in \mathbf{R}^{m\times n}$ with $\Sigma_{ii} = \sigma_i \geq 0$ and $\Sigma_{ij} = 0$ for $i \neq j$
- for a rectangular A, Σ has a diagonal submatrix Σ_1 with dimension of $\min(m,n)$

$$A_{\text{tall}} = \begin{bmatrix} U_1 & | & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 \\ \hline 0 \end{bmatrix} V^T = U_1\Sigma_1 V^T, \quad A_{\text{fat}} = U \begin{bmatrix} \Sigma_1 & | & 0 \end{bmatrix} \begin{bmatrix} V_1^T \\ \hline V_2^T \end{bmatrix} = U\Sigma_1 V_1^T$$

Square A

$$\begin{bmatrix} 2 & 1 \\ -1 & -2 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \frac{1}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ -1 & -1 \end{bmatrix}^T, \mathbf{rank}(A) = 2$$

$$\begin{bmatrix} 2 & 4 & -2 \\ -2 & 0 & -2 \\ 2 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -0.94 & -0.27 & -0.20 \\ 0.11 & -0.80 & 0.59 \\ -0.31 & 0.53 & 0.78 \end{bmatrix} \begin{bmatrix} 5.10 & 0 & 0 \\ 0 & 3.46 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.53 & 0.62 & 0.58 \\ -0.80 & -0.15 & -0.58 \\ 0.27 & 0.77 & -0.58 \end{bmatrix}^T, \mathbf{rank}(A) = 2$$

$$\begin{bmatrix} -2 & 1 & 3 \\ 4 & -2 & -6 \\ 2 & -1 & -3 \end{bmatrix} = \begin{bmatrix} -0.41 & -0.91 & 0 \\ 0.82 & -0.37 & -0.45 \\ 0.41 & -0.18 & 0.89 \end{bmatrix} \begin{bmatrix} 9.17 & 0 & 9 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.53 & -0.85 & 0 \\ -0.27 & -0.17 & 0.95 \\ -0.80 & -0.51 & -0.32 \end{bmatrix}^T, \mathbf{rank}(A) = 1$$

- check the singular values and eigenvalues of $A^T A$
- confirm the rank and the number of nonzero singular values
- if A is invertible, so is Σ

Fat A

$$A_1 = \begin{bmatrix} 2 & 0 & 2 \\ 0 & -2 & 1 \end{bmatrix} = \begin{bmatrix} -0.89 & -0.45 \\ -0.45 & 0.89 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \begin{bmatrix} -0.60 & -0.45 & -0.67 \\ 0.30 & -0.89 & 0.33 \\ -0.75 & 0 & 0.67 \end{bmatrix}^T, \mathbf{rank}(A) = 2$$

$$A_2 = \begin{bmatrix} 2 & -1 & 1 & 0 \\ 2 & 0 & 1 & -2 \\ -2 & 0 & -1 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 0.42 & 0.91 & 0 \\ 0.64 & -0.30 & 0.71 \\ -0.64 & 0.30 & 0.71 \end{bmatrix} \begin{bmatrix} 4.6100 & 0 & 0 & 0 \\ 0 & 1.65 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.74 & 0.38 & 0.40 & -0.38 \\ -0.09 & -0.55 & 0.82 & 0.14 \\ 0.37 & 0.19 & 0.01 & 0.91 \\ -0.56 & 0.72 & 0.41 & 0.07 \end{bmatrix}^T, \mathbf{rank}(A) = 1$$

- A_2 is low rank, the SVD form can be reduced to $A_2 = U\Sigma V^T = U_r \Sigma_r V_r^T$ where U_r, V_r have the first r columns of U and V respectively and Σ_r is the leading r-diagonal block of Σ $\quad (r = \mathbf{rank}(A))$

Tall A

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 1 \\ -2 & -2 & 0 \\ 0 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & -1.00 \\ 0.33 & -0.63 & -0.71 & 0 \\ 0.89 & 0.46 & 0 & 0 \\ -0.33 & 0.63 & -0.71 & 0 \end{bmatrix} \begin{bmatrix} 3.080 & 0 & 0 \\ 0 & 1.59 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.58 & -0.58 & 0.58 \\ -0.79 & 0.21 & -0.58 \\ 0.21 & -0.79 & -0.58 \end{bmatrix}^T$$

- $\mathbf{rank}(A) = 2$ and there are two nonzero singular values
- A can be reduced to

$$A = U\Sigma V^T = U_r \Sigma_r V_r^T, \quad r = \mathbf{rank}(A) = 2$$

Softwares

MATLAB

- `[U,S,V] = svd(A)` returns SVD decomposition: $A = USV^T$

Python

- `U,S,Vt = scipy.linalg.svd(A)`
- `U,S,Vt = numpy.linalg.svd(A)`

  returns SVD decomposition: $A = USV^T$ where `S` is returned as a vector of singular values and `Vt` as V^T

Pseudo-inverse (Penrose Theorem)

one can have a notion of 'inverse' for a non-square matrix

Penrose's Theorem: given $A \in \mathbf{R}^{m \times n}$, there is exactly one $n \times m$ matrix B such that

1. $ABA = A$ and $BAB = B$
2. both AB and BA are symmetric

definition: the **pseudo inverse** of $A \in \mathbf{R}^{m \times n}$ is the unique $n \times m$ matrix $A^\dagger$ such that

1. $AA^\dagger A = A$ and $A^\dagger AA^\dagger = A^\dagger$
2. both $AA^\dagger$ and $A^\dagger A$ are symmetric

Pseudo-inverse

consider a full rank matrix $A \in \mathbf{R}^{m \times n}$ in three cases

- **tall matrix:** A is full rank $\Leftrightarrow$ columns of A are LI $\Leftrightarrow$ A^TA is invertible

$$((A^TA)^{-1}A^T)A = (A^TA)^{-1}(A^TA) = I$$

  the **pseudo-inverse** of A (or left-inverse) is $A^\dagger = (A^TA)^{-1}A^T$

- **wide matrix:** A is full rank $\Leftrightarrow$ row of A are LI $\Leftrightarrow$ AA^T is invertible

$$A(A^T(AA^T)^{-1}) = (AA^T)(AA^T)^{-1} = I$$

  the **pseudo-inverse** of A (or right-inverse) is $A^\dagger = A^T(AA^T)^{-1}$

- **square matrix:** A is full rank $\Leftrightarrow$ A is invertible and both formula of pseudo-inverses reduce to the ordinary inverse A^{-1}

✎ the pseudo inverses of the three cases have the same dimension ?

Example

$$A = \begin{bmatrix} 0 & 2 & 1 \\ -2 & 1 & -2 \end{bmatrix}, \quad A^{\dagger} = A^T(AA^T)^{-1} = \begin{bmatrix} 0 & -2/9 \\ 2/5 & 1/9 \\ 1/5 & -2/9 \end{bmatrix}$$

$$A = \begin{bmatrix} -2 & -1 \\ 2 & -1 \\ -1 & 0 \end{bmatrix}, \quad A^{\dagger} = (A^TA)^{-1}A^T = \begin{bmatrix} -2/9 & 2/9 & 1/9 \\ -1/2 & -1/2 & 0 \end{bmatrix}$$

however, when rentangular A has low rank, we can use SVD to find the pseudo inverse

Pseudo-inverse via SVD

the pseudo-inverse $A^\dagger$ can be computed from any SVD for $A \in \mathbf{R}^{n\times m}$

- from $A = U_{n\times n}\Sigma_{n\times m}V^T_{m\times m}$ if A has rank r then

$$\Sigma = \begin{bmatrix} \Sigma_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}, \quad \text{and that } \Sigma_r \text{ is invertible}$$

- define $\Sigma^\dagger = \begin{bmatrix} \Sigma_r^{-1} & 0 \\ 0 & 0 \end{bmatrix}_{n\times m}$ and we can verify that

$$\Sigma\Sigma^\dagger\Sigma = \Sigma, \ \ \Sigma^\dagger\Sigma\Sigma^\dagger = \Sigma^\dagger, \ \ \Sigma\Sigma^\dagger = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times m}, \ \ \Sigma^\dagger\Sigma = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times n}$$

proving that $\Sigma^\dagger$ is the pseudoinverse of Σ

Pseudo-inverse via SVD

given $A = U\Sigma V^T$, then the pseudo-inverse of A is

$$A^\dagger = V\Sigma^\dagger U^T$$

by verifying Penrose's Theorem from page 226 that

- $AA^\dagger A = (U\Sigma V^T)(V\Sigma^\dagger U^T)(U\Sigma V^T) = U\Sigma\Sigma^\dagger\Sigma V^T = U\Sigma V^T = A$
- $A^\dagger AA^\dagger = (V\Sigma^\dagger U^T)(U\Sigma V^T)(V\Sigma^\dagger U^T) = V\Sigma^\dagger\Sigma\Sigma^\dagger U^T = V\Sigma^\dagger U^T = A^\dagger$
- $AA^\dagger = U\Sigma\Sigma^\dagger U^T$ which is symmetric
- $A^\dagger A = V\Sigma^\dagger\Sigma V^T$ which is symmetric

Example

a tall full rank A

$$A = \begin{bmatrix} -2 & -1 \\ 2 & -1 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -0.6667 & -0.7071 & -0.2357 \\ 0.6667 & -0.7071 & 0.2357 \\ -0.3333 & -0.0000 & 0.9428 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 1.4142 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^T$$

$$A^\dagger = V\Sigma^\dagger U^T = V \begin{bmatrix} 0.3333 & 0 & 0 \\ 0 & 0.7071 & 0 \end{bmatrix} U^T$$

$$= \begin{bmatrix} -0.22 & 0.22 & -0.1100 \\ -0.50 & -0.50 & 0 \end{bmatrix}$$

Example

a fat low rank A

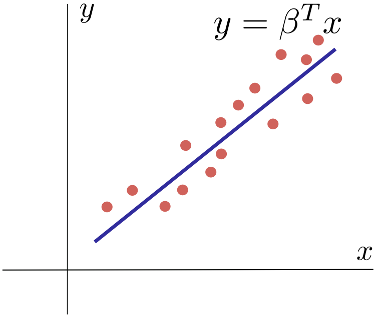
$$A = \begin{bmatrix} -2 & -1 & -3 & 0 \\ 0 & -3 & -3 & -2 \\ 2 & -2 & 0 & -2 \end{bmatrix} = \begin{bmatrix} 0.47 & 0.67 & -0.58 \\ 0.81 & -0.08 & 0.58 \\ 0.34 & -0.74 & -0.58 \end{bmatrix} \begin{bmatrix} 5.76 & 0 & 0 & 0 \\ 0 & 3.85 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} -0.05 & -0.73 & 0.51 & -0.45 \\ -0.62 & 0.27 & -0.27 & -0.68 \\ -0.67 & -0.46 & -0.25 & 0.53 \\ -0.40 & 0.43 & 0.78 & 0.23 \end{bmatrix}^T$$

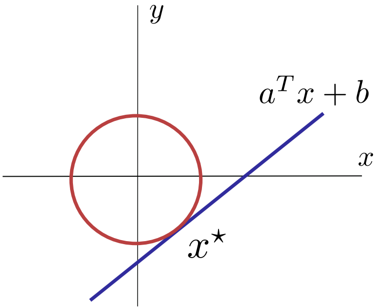
$$A^\dagger = V\Sigma^\dagger U^T = V \begin{bmatrix} 0.1736 & 0 & 0 \\ 0 & 0.2596 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} U^T$$

$$= \begin{bmatrix} -0.13 & 0.01 & 0.14 \\ 0 & -0.09 & -0.09 \\ -0.13 & -0.09 & 0.05 \\ 0.04 & -0.07 & -0.11 \end{bmatrix}$$

- $\mathbf{rank}(A) = 2 < n$ and there are two non-zero singular values
- $\Sigma \in \mathbf{R}^{3\times 4}$ and $\Sigma^\dagger \in \mathbf{R}^{4\times 3}$ with 2×2 invertible block

Applications of pseudo-inverse





- **least-square problem:** find a straight line that fit best in 2-norm sense to data points
- **least-norm problem:** find a point x on the given hyperplane that has the smallest norm

Least-square problem

given $X \in \mathbf{R}^{N \times p}, y \in \mathbf{R}^N$ where typically $N > p$, a least-square problem is

$$\underset{\beta}{\text{minimize}} \quad \|y - X\beta\|_2^2$$

- it generalizes solving an overdetermined linear system that cannot be solved exactly by allowing the system to have the smallest residual
- if X is full rank, and from zero-gradient condition, the optimal solution is

$$\beta = (X^T X)^{-1} X^T y$$

- the solution is linear in y where the coefficient is the **left inverse** of X

Least-norm problem

given $A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m$ where $m < n$ and A is full rank, the least-norm problem is

$$\underset{x}{\text{minimize}} \quad \|x\|_2 \quad \text{subject to} \quad Ax = y$$

- find a point on hyperplane $Ax = b$ while keeping the 2-norm of x smallest
- it extends from solving an under-determined system that has many solutions and we aim to find the solution with smallest norm
- it can be shown that the optimal solution is

$$x^\star = A^T(AA^T)^{-1}y, \quad \text{provided that } A \text{ is full row rank}$$

- the solution is linear in y where the coefficient is the **right inverse** of A

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least squares*, Cambridge, 2018
3. Lecture notes of EE133A, L. Vandenberhge, UCLA `https://www.seas.ucla.edu/~vandenbe/133A`

Vector space

Outline

- definition
- linear independence
- basis and dimension
- coordinate and change of basis
- range space and null space
- rank and nullity

Elements of vector space

a vector space or linear space (over **R**) consists of

- a set $\mathcal{V}$
- a vector sum $+ : \mathcal{V} \times \mathcal{V} \to \mathcal{V}$
- a scalar multiplication : $\mathbf{R} \times \mathcal{V} \to \mathcal{V}$
- a distinguished element $0 \in \mathcal{V}$

which satisfy a list of properties

properties under addition

- $x + y \in \mathcal{V} \quad \forall x, y \in \mathcal{V}$ (closed under addition)
- $x + y = y + x$, $\forall x, y \in \mathcal{V}$ ($+$ is commutative)
- $(x + y) + z = x + (y + z)$, $\forall x, y, z \in \mathcal{V}$ ($+$ is associative)
- $0 + x = x$, $\forall x \in \mathcal{V}$ (0 is additive identity)
- $\forall x \in \mathcal{V}\ \exists (-x) \in \mathcal{V}$ s.t. $x + (-x) = 0$ (existence of additive inverse)

properties under scalar multiplication

- $\alpha x \in \mathcal{V}$ for any $\alpha \in \mathbf{R}$ (closed under scalar multiplication)
- $(\alpha\beta)x = \alpha(\beta x)$, $\forall \alpha, \beta \in \mathbf{R}\ \forall x \in \mathcal{V}$ (scalar multiplication is associative)
- $\alpha(x + y) = \alpha x + \alpha y$, $\forall \alpha \in \mathbf{R}\ \forall x, y \in \mathcal{V}$ (right distributive rule)
- $(\alpha + \beta)x = \alpha x + \alpha y$, $\forall \alpha, \beta \in \mathbf{R}\ \ \forall x \in \mathcal{V}$ (left distributive rule)
- $\mathbf{1}x = x$, $\forall x \in \mathcal{V}$ (1 is multiplicative identity)

notation

- $(\mathcal{V}, \mathbf{R})$ denotes a vector space $\mathcal{V}$ over $\mathbf{R}$
- an element in $\mathcal{V}$ is called a **vector**

Theorem: let u be a vector in $\mathcal{V}$ and k a scalar; then

- $0u = 0$ (multiplication with zero gives the zero vector)
- $k0 = 0$ (multiplication with the zero vector gives the zero vector)
- $(-1)u = -u$ (multiplication with -1 gives the additive inverse)
- if $ku = 0$, then $k = 0$ or $u = 0$

roughly speaking, a vector space must satisfy the following operations

1. **vector addition**

$$x, y \in \mathcal{V} \quad \Rightarrow \quad x + y \in \mathcal{V}$$

2. **scalar multiplication**

$$\text{for any } \alpha \in \mathbf{R}, \;\; x \in \mathcal{V} \quad \Rightarrow \quad \alpha x \in \mathcal{V}$$

the second condition implies that a vector space contains the **zero vector**

$$0 \in \mathcal{V}$$

in other words, if $\mathcal{V}$ is a vector space then $0 \in \mathcal{V}$

(but the converse is *not true*)

Examples

the following sets are vector spaces (over **R**)

- $\mathbf{R}^n$
- $\{0\}$
- $\mathbf{R}^{m\times n}$
- $\mathbf{C}^{m\times n}$: set of $m \times n$-complex matrices
- $\mathbf{P}_n$: set of polynomials of degree $\leq n$

$$\mathbf{P}_n = \{p(t) \mid p(t) = a_0 + a_1 t + \cdots + a_n t^n\}$$

- $\mathbf{S}^n$: set of symmetric matrices of size n
- $C(-\infty, \infty)$: set of real-valued continuous functions on $(-\infty, \infty)$
- $C^n(-\infty, \infty)$: set of real-valued functions with continuous nth derivatives on $(-\infty, \infty)$

✎ check whether any of the following sets is a vector space (over **R**)

- $\{0, 1, 2, 3, \ldots\}$
- $\left\{ \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$
- $\left\{ x \in \mathbf{R}^2 \;\middle|\; x = \begin{bmatrix} x_1 \\ 0 \end{bmatrix}, \quad x_1 \in \mathbf{R} \right\}$
- $\{p(x) \in \mathbf{P}_2 \mid p(x) = a_1 x + a_2 x^2 \quad \text{for some } a_1, a_2 \in \mathbf{R}\}$

Subspace

- a **subspace** of a vector space is a *subset* of a vector space which is itself a vector space
- a subspace is closed under vector addition and scalar multiplication

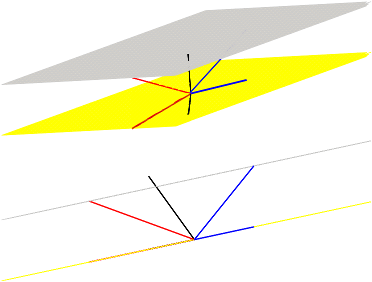
examples:

- $\{0\}$ is a subspace of $\mathbf{R}^n$
- $\mathbf{R}^{m\times n}$ is a subspace of $\mathbf{C}^{m\times n}$
- $\{x \in \mathbf{R}^2 \mid x_1 = 0\}$ is a subspace of $\mathbf{R}^2$
- $\{x \in \mathbf{R}^2 \mid x_2 = 1\}$ is not a subspace of $\mathbf{R}^2$
- $\left\{ \begin{bmatrix} 1 & 4 \\ -3 & 2 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \right\}$ is not a subspace of $\mathbf{R}^{2\times 2}$
- the solution set $\{x \in \mathbf{R}^n \mid Ax = b\}$ for $b \neq 0$ is a not subspace of $\mathbf{R}^n$

Examples of subspace

two hyperplanes; one is a subspace but the other one is not

$$2x_1 - 3x_2 + x_3 = 0 \quad \text{(yellow)}, \quad 2x_1 - 3x_2 + x_3 = 20 \quad \text{(grey)}$$



black $=$ **red** $+$ **blue**

$x = (-3, -2, 0)$ and $y = (1, -1, -5)$ are on the yellow plane, and so is $x + y$

$x = (-3, -2, 20)$ and $y = (1, -1, 15)$ are on the grey plane, but $x + y$ is not

Linear Independence

Definition: a set of vectors $\{v_1, v_2, \dots, v_n\}$ is **linearly independent** if

$$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = 0 \Longrightarrow \alpha_1 = \alpha_2 = \cdots = \alpha_n = 0$$

equivalent conditions:

- coefficients of $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n$ are uniquely determined, i.e.,

  $$\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_n v_n = \beta_1 v_1 + \beta_2 v_2 + \cdots + \beta_n v_n$$

  implies $\alpha_k = \beta_k$ for $k = 1, 2, \dots, n$
- no vector v_i can be expressed as a linear combination of the other vectors

Examples

- $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}$ are independent
- $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ are independent
- $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ 0 \end{bmatrix}$ are not independent
- $\begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ -1 \\ -1 \end{bmatrix}$ are not independent

Linear span

Definition: the linear span of a set of vectors

$$\{v_1, v_2, \ldots, v_n\}$$

is the set of all linear combinations of $v_1, \ldots, v_n$

$$\text{span}\{v_1, v_2, \ldots, v_n\} = \{a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \mid a_1, \ldots, a_n \in \mathbf{R}\}$$

example:

$$\text{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$ is the set of 2×2 symmetric matrices

Fact: if $v_1, \ldots, v_n$ are vectors in $\mathcal{V}$, $\text{span}\{v_1, \ldots, v_n\}$ is a subspace of $\mathcal{V}$

Basis and dimension

definition: set of vectors $\{v_1, v_2, \cdots, v_n\}$ is a **basis** for a vector space $\mathcal{V}$ if

- $\{v_1, v_2, \ldots, v_n\}$ is linearly independent
- $\mathcal{V} = \text{span}\ \{v_1, v_2, \ldots, v_n\}$

equivalent condition: *every* $v \in \mathcal{V}$ can be *uniquely* expressed as

$$v = \alpha_1 v_1 + \cdots + \alpha_n v_n$$

definition: the **dimension** of $\mathcal{V}$, denoted $\dim(\mathcal{V})$, is the number of vectors in a basis for $\mathcal{V}$

Theorem: the number of vectors in *any* basis for $\mathcal{V}$ is the same

(we assign $\dim\{0\} = 0$)

Examples

- $\{e_1, e_2, e_3\}$ is a standard basis for $\mathbf{R}^3$ $\qquad (\dim \mathbf{R}^3 = 3)$
- $\left\{ \begin{bmatrix} -1 \\ 3 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$ is a basis for $\mathbf{R}^2$ $\qquad (\dim \mathbf{R}^2 = 2)$
- $\{1, t, t^2\}$ is a basis for $\mathbf{P}_2$ $\qquad (\dim \mathbf{P}_2 = 3)$
- $\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ is a basis for $\mathbf{R}^{2\times 2}$ $\qquad (\dim \mathbf{R}^{2\times 2} = 4)$
- $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\}$ cannot be a basis for $\mathbf{R}^3$ why ?
- $\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 3 \end{bmatrix} \right\}$ cannot be a basis for $\mathbf{R}^2$ why ?

Example

let $\mathcal{V} = \{p \in \mathbf{P}_2 \mid p(2) = 0\}$ find a basis for $\mathcal{V}$

- ✎ verify that $\mathcal{V}$ is a subspace for $\mathbf{P}_2$
- characterize the space $\mathcal{V}$

$$p(t) = a_0 + a_1 t + a_2 t^2, \quad p(2) = a_0 + 2a_1 + 4a_2 = 0$$

therefore, any $p(t) \in \mathcal{V}$ takes the form

$$p(t) = -2a_1 - 4a_2 + a_1 t + a_2 t^2 = a_1(t-2) + a_2(t^2 - 4), \quad a_1, a_2 \in \mathbf{R}$$

- we have shown that $p(t) \in \operatorname{span}\{t-2, t^2-4\}$
- we can verify that $\{t-2, t^2-4\}$ is LI
- therefore $\{t-2, t^2-4\}$ is a basis for $\mathcal{V}$ and $\dim(\{t-2, t^2-4\}) = 2$

Standard basis for $\mathbf{S}^3$

any $A \in \mathbf{S}^3$ can be expressed as

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{bmatrix} = a_{11} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + a_{12} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + a_{13} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$+ a_{23} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} + a_{33} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\triangleq a_{11}E_{11} + a_{12}E_{12} + a_{13}E_{13} + a_{23}E_{23} + a_{33}E_{33}$$

- we have shown that $A \in \mathrm{span}\{E_{11}, E_{12}, E_{13}, E_{23}, E_{33}\}$
- verify that $\{E_{11}, E_{12}, E_{13}, E_{23}, E_{33}\}$ is LI
- hence, $\{E_{11}, E_{12}, E_{13}, E_{23}, E_{33}\}$ is a basis for $\mathbf{S}^3$ and $\dim(\mathbf{S}^3) = 5$

Review questions

✎ answer the questions and explain a reason

1. find the standard basis for $\mathbf{S}^n$
2. can $\{E_{11}, E_{12}, E_{13}, E_{21}, E_{22}, E_{23}, E_{31}, E_{32}, E_{33}\}$ be a basis for $\mathbf{S}^3$?
3. can $\{E_{11}, E_{12}, E_{13}, E_{23}, E_{33}\}$ be a basis for $\mathbf{R}^{3\times 3}$?
4. let $\mathcal{V} = \{ x \in \mathbf{R}^n \mid \sum_i x_i = 0 \}$
   - can $\{e_1, e_2, \ldots, e_n\}$ (standard basis) be a basis for $\mathcal{V}$?
   - is it possible to find two different bases for $\mathcal{V}$?

Coordinates

let $S = \{v_1, v_2, \ldots, v_n\}$ be a basis for a vector space $\mathcal{V}$

suppose a vector $v \in \mathcal{V}$ can be written as

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

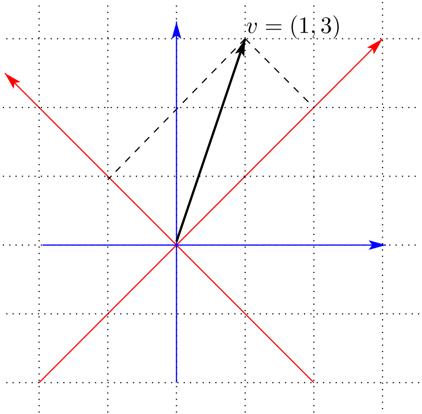
definition: the coordinate vector of v relative to the basis S is

$$[v]_S = (a_1, a_2, \ldots, a_n)$$

- linear independence of vectors in S ensures that a_k's are *uniquely* determined by S and v
- changing the basis yields a different coordinate vector

Geometrical interpretation

new coordinate in a new reference axis



$$v = \begin{bmatrix} 1 \\ 3 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2 \begin{bmatrix} 1 \\ 1 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

Examples

- $S = \{e_1, e_2, e_3\}$, $v = (-2, 4, 1)$

$$v = -2e_1 + 4e_2 + 1e_3, \quad [v]_S = (-2, 4, 1)$$

- $S = \{(-1, 2, 0), (3, 0, 0), (-2, 1, 1)\}$, $v = (-2, 4, 1)$

$$v = \begin{bmatrix} -2 \\ 4 \\ 1 \end{bmatrix} = \frac{3}{2}\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix} + 1\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}, \quad [v]_S = (3/2, 1/2, 1)$$

- $S = \{1, t, t^2\}$, $v(t) = -3 + 2t + 4t^2$

$$v(t) = -3 \cdot 1 + 2 \cdot t + 4 \cdot t^2, \quad [v]_S = (-3, 2, 4)$$

- $S = \{1, t - 1, t^2 + t\}$, $v(t) = -3 + 2t + 4t^2$

$$v(t) = -5 \cdot 1 - 2 \cdot (t - 1) + 4 \cdot (t^2 + t), \quad [v]_S = (-5, -2, 4)$$

Change of basis

let $U = \{u_1, \dots, u_n\}$ and $W = \{w_1, \dots, w_n\}$ be bases for a vector space $\mathcal{V}$

a vector $v \in \mathcal{V}$ has the coordinates relative to these bases as

$$[v]_U = (a_1, a_2, \dots, a_n), \quad [v]_W = (b_1, b_2, \dots, b_n)$$

suppose the coordinate vectors of w_k relative to U is

$$[w_k]_U = (c_{1k}, c_{2k}, \dots, c_{nk})$$

or in the matrix form as

$$\begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix}$$

the coordinate vectors of v relative to U and W are related by

$$\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \triangleq P \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

- we obtain $[v]_U$ by multiplying $[v]_W$ with P
- P is called the **transition** matrix from W to U
- the columns of P are the coordinate vectors of the basis vectors in W relative to U

Theorem ✌

P is invertible and P^{-1} is the transition matrix from U to W

Example

find $[v]_U$, given

$$U = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}, \quad W = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}, \quad [v]_W = \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

first, find the coordinate vectors of the basis vectors in W relative to U

$$\begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

from which we obtain the transition matrix

$$P = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix}$$

and $[v]_U$ is given by

$$[v]_U = \frac{1}{2} \begin{bmatrix} 3 & 1 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} -2 \\ 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} -5 \\ 1 \end{bmatrix}$$

Nullspace

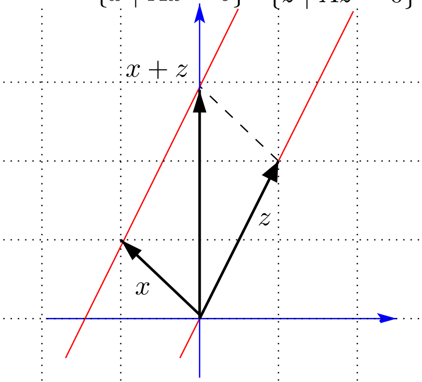
the **nullspace** of an $m \times n$ matrix is defined as

$$\mathcal{N}(A) = \{x \in \mathbf{R}^n \mid Ax = 0\}$$

- the set of all vectors that are mapped to zero by $f(x) = Ax$
- the set of all vectors that are orthogonal to the rows of A
- if $Ax = b$ then $A(x+z) = b$ for all $z \in \mathcal{N}(A)$
- also known as **kernel** of A
- $\mathcal{N}(A)$ is a subspace of $\mathbf{R}^n$

✎

Example



$$A = \begin{bmatrix} 2 & -1 \\ -4 & 2 \\ -6 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} -3 \\ 6 \\ 9 \end{bmatrix}$$

- $\mathcal{N}(A) = \{x \mid 2x_1 - x_2 = 0\}$
- the solution set of $Ax = b$ is $\{x \mid 2x_1 - x_2 = -3\}$
- the solution set of $Ax = b$ is the translation of $\mathcal{N}(A)$

Zero nullspace matrix

- A has a zero nullspace if $\mathcal{N}(A) = \{0\}$
- if A has a zero nullspace and $Ax = b$ is solvable, the solution is unique
- columns of A are independent

✌ **equivalent conditions:** $A \in \mathbf{R}^{n \times n}$

- A has a zero nullspace
- A is invertible or nonsingular
- columns of A are a basis for $\mathbf{R}^n$

Range space

the **range** of an $m \times n$ matrix A is defined as

$$\mathcal{R}(A) = \{y \in \mathbf{R}^m \mid y = Ax \text{ for some } x \in \mathbf{R}^n \}$$

- the set of all m-vectors that can be expressed as Ax
- the set of all linear combinations of the columns of $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}$

$$\mathcal{R}(A) = \{y \mid y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n, \quad x \in \mathbf{R}^n\}$$

- the set of all vectors b for which $Ax = b$ is solvable
- also known as the **column space** of A
- $\mathcal{R}(A)$ is a subspace of $\mathbf{R}^m$

Full range matrices

A has a full range if $\mathcal{R}(A) = \mathbf{R}^m$

✌ **equivalent conditions:**

- A has a full range
- columns of A span $\mathbf{R}^m$
- $Ax = b$ is solvable for *every* b
- $\mathcal{N}(A^T) = \{0\}$

Bases for $\mathcal{R}(A)$ and $\mathcal{N}(A)$

A and B are row equivalent matrices, *i.e.*,

$$B = E_k \cdots E_2 E_1 A$$

Facts ✌

- elementary row operations *do not alter* $\mathcal{N}(A)$

$$\mathcal{N}(B) = \mathcal{N}(A)$$

- columns of B are independent if and only if columns of A are
- a given set of column vectors of A forms a basis for $\mathcal{R}(A)$ if and only if the corresponding column vectors of B form a basis for $\mathcal{R}(B)$

Examples

given a matrix A and its row echelon form B:

$$A = \begin{bmatrix} -1 & 2 & 4 & 1 \\ 0 & 1 & 2 & 1 \\ 2 & 3 & 6 & 5 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

basis for $\mathcal{N}(A)$: from $\{x \mid Ax = 0\} = \{x \mid Bx = 0\}$, we read

$$x_1 + x_4 = 0, \quad x_2 + 2x_3 + x_4 = 0$$

define x_3 and x_4 as free variables, any $x \in \mathcal{N}(A)$ can be written as

$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -x_4 \\ -2x_3 - x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

(a linear combination of $(0, -2, 1, 0)$ and $(-1, -1, 0, 1)$)

hence, a basis for $\mathcal{N}(A)$ is $\left\{ \begin{bmatrix} 0 \\ -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ and $\dim \mathcal{N}(A) = 2$

basis for $\mathcal{R}(A)$: pick a set of the independent column vectors in B (here pick the 1st and the 2nd columns)

the corresponding columns in A form a basis for $\mathcal{R}(A)$:

$$\left\{ \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \right\}$$

$\dim \mathcal{R}(A) = 2$

✌ **conclusion:** if R is the row reduced echelon form of A

- the pivot column vectors of R form a basis for the range space of R
- the column vectors of A *corresponding* to the pivot columns of R form a basis for the range space of A
- $\dim \mathcal{R}(A)$ is the number of leading 1's in R
- $\dim \mathcal{N}(A)$ is the number of free variables in solving $Rx = 0$

Rank and Nullity

rank of a matrix $A \in \mathbf{R}^{m\times n}$ is defined as

$$\mathbf{rank}(A) = \dim \mathcal{R}(A)$$

nullity of a matrix $A \in \mathbf{R}^{m\times n}$ is

$$\mathbf{nullity}(A) = \dim \mathcal{N}(A)$$

Facts ✌

- $\mathbf{rank}(A)$ is maximum number of independent columns (or rows) of A

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

- $\mathbf{rank}(A) = \mathbf{rank}(A^T)$

Full rank matrices

for $A \in \mathbf{R}^{m \times n}$ we always have $\mathbf{rank}(A) \leq \min(m, n)$

we say A is **full rank** if $\mathbf{rank}(A) = \min(m, n)$

- for **square** matrices, full rank means nonsingular (invertible)
- for **skinny** matrices ($m \geq n$), full rank means columns are independent
- for **fat** matrices ($m \leq n$), full rank means rows are independent

Rank-Nullity Theorem

for any $A \in \mathbf{R}^{m \times n}$,

$$\mathbf{rank}(A) + \dim \mathcal{N}(A) = n$$

Proof:

- a homogeneous linear system $Ax = 0$ has n variables
- these variables fall into two categories
  - leading variables
  - free variables
- $\#$ of leading variables $= \#$ of leading 1's in reduced echelon form of A $= \mathbf{rank}(A)$
- $\#$ of free variables $=$ nullity of A

Softwares

MATLAB

- `rank(A)` provides an estimate of the rank of A
- `null(A)` gives normalized vectors in an orthonormal basis for $\mathcal{N}(A)$

Python

- `numpy.linalg.matrix_rank(A)` provides an estimate of the rank of A
- `scipy.linalg.null_space(A)` finds orthonormal basis for the nullspace of A

References

1. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
2. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011

Linear transformation

Outline

- linear transformation
- matrix transformation
- kernel and range
- isomorphism
- composition
- inverse transformation

Transformation

let X and Y be vector spaces

a **transformation** T from X to Y, denoted by

$$T : X \to Y$$

is an assignment taking $x \in X$ to $y = T(x) \in Y$,

$$T : X \to Y, \quad y = T(x)$$

- **domain** of T, denoted $\mathcal{D}(T)$ is the collection of all $x \in X$ for which T is defined
- vector $T(x)$ is called the **image** of x under T
- collection of all $y = T(x) \in Y$ is called the **range** of T, denoted by $\mathcal{R}(T)$

Example

example 1 define $T : \mathbf{R}^3 \to \mathbf{R}^2$ as

$$\begin{aligned} y_1 &= -x_1 + 2x_2 + 4x_3 \\ y_2 &= -x_2 + 9x_3 \end{aligned}$$

example 2 define $T : \mathbf{R}^3 \to \mathbf{R}$ as

$$y = \sin(x_1) + x_2x_3 - x_3^2$$

example 3 general transformation $T : \mathbf{R}^n \to \mathbf{R}^m$

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \ldots, x_n) \\ y_2 &= f_2(x_1, x_2, \ldots, x_n) \\ &\vdots \\ y_m &= f_m(x_1, x_2, \ldots, x_n) \end{aligned}$$

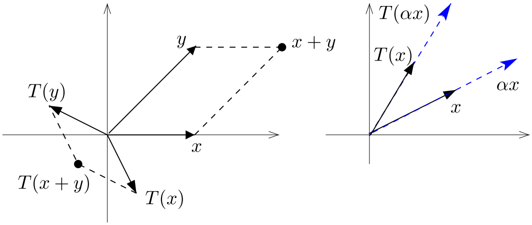
where $f_1, f_2, \ldots, f_m$ are real-valued functions of n variables

Linear transformation

let X and Y be vector spaces over **R**

Definition: a transformation $T : X \to Y$ is **linear** if

- $T(x + z) = T(x) + T(z), \quad \forall x, y \in X$ (additivity)
- $T(\alpha x) = \alpha T(x), \quad \forall x \in X, \forall \alpha \in \mathbf{R}$ (homogeneity)



Examples

✎ which of the following is a linear transformation ?

- **matrix transformation** $T : \mathbf{R}^n \to \mathbf{R}^m$

$$T(x) = Ax, \quad A \in \mathbf{R}^{m \times n}$$

- **affine transformation** $T : \mathbf{R}^n \to \mathbf{R}^m$

$$T(x) = Ax + b, \quad A \in \mathbf{R}^{m \times n}, \quad b \in \mathbf{R}^m$$

- $T : \mathbf{P}_n \to \mathbf{P}_{n+1}$

$$T(p(t)) = tp(t)$$

- $T : \mathbf{P}_n \to \mathbf{P}_n$

$$T(p(t)) = p(t+1)$$

- $T : \mathbf{R}^{m\times n} \to \mathbf{R}^{n\times m}, \quad T(X) = X^T$
- $T : \mathbf{R}^{n\times n} \to \mathbf{R}, \quad T(X) = \det(X)$
- $T : \mathbf{R}^{n\times n} \to \mathbf{R}, \quad T(X) = \mathbf{tr}(X)$
- $T : \mathbf{R}^{n} \to \mathbf{R}, \quad T(x) = \|x\| \triangleq \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$
- $T : \mathbf{R}^{n} \to \mathbf{R}^n, \quad T(x) = 0$

denote $F(-\infty, \infty)$ the set of all real-valued functions on $(-\infty, \infty)$

- $T : C^1(-\infty, \infty) \to F(-\infty, \infty)$

$$T(f) = f'$$

- $T : C(-\infty, \infty) \to C^1(-\infty, \infty)$

$$T(f) = \int_0^t f(s)ds$$

Examples of matrix transformation

$T : \mathbf{R}^n \to \mathbf{R}^m$

$$T(x) = Ax, \quad A \in \mathbf{R}^{m \times n}$$

zero transformation: $T : \mathbf{R}^n \to \mathbf{R}^m$

$$T(x) = 0 \cdot x = 0$$

T maps every vector into the zero vector

identity operator: $T : \mathbf{R}^n \to \mathbf{R}^n$

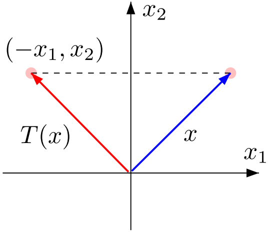
$$T(x) = I_n \cdot x = x$$

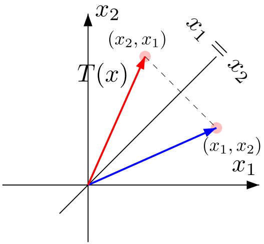
T maps a vector into itself

Reflection operator

$T : \mathbf{R}^n \to \mathbf{R}^n$

T maps each point into its symmetric image about an axis or a line



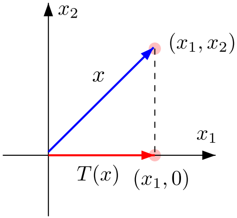


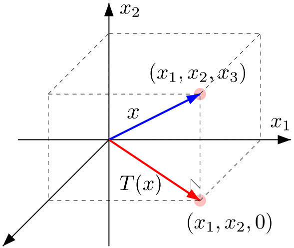
$$T(x) = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix} x \qquad T(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} x$$

Projection operator

$T : \mathbf{R}^n \to \mathbf{R}^n$

T maps each point into its orthogonal projection on a line or a plane



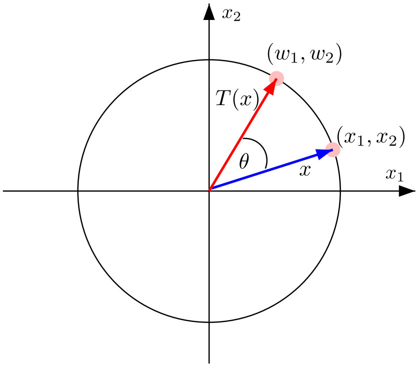


$$T(x) = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} x \qquad T(x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} x$$

Rotation operator

$T : \mathbf{R}^n \to \mathbf{R}^n$

T maps points along circular arcs

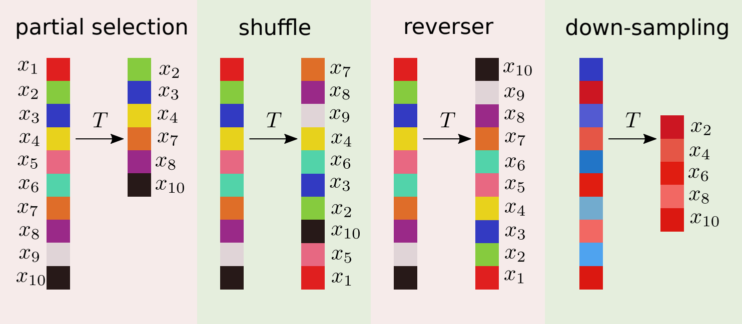


T rotates x through an angle θ

$$w = T(x) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} x$$

Selector transformations

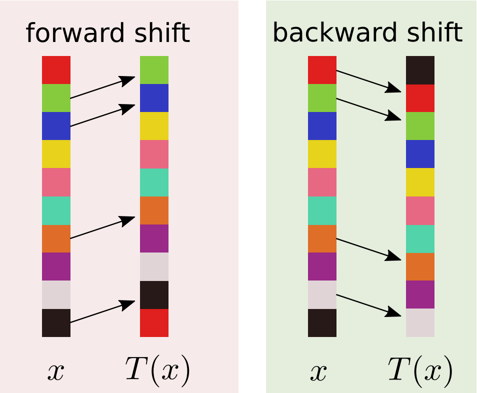
these transformations can be represented as $y = T(x) = Ax$



- partial selection: select some entries of x
- shuffle: randomize entries in x
- reverser: reverse the order of x
- down-sampling: sub-sample entries in x, *e.g.*, `x(1:2:end)`

Shift transformations

shifting sequences as a matrix transformation $T(x) = Ax$



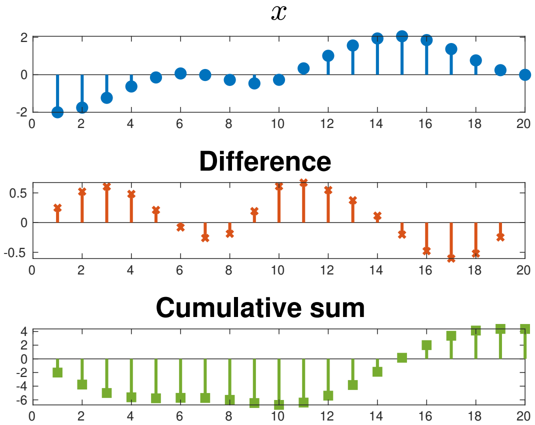
$$T_1(x) = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \\ x_1 \end{bmatrix}, \quad T_2(x) = \begin{bmatrix} x_n \\ x_1 \\ x_2 \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{bmatrix}$$

what is the associated matrix A for each transformation ?

do you notice some structure of A ?

Signal processing

differencing and cumulative sum as matrix transformations $T(x) = Ax$



$$T_1(x) = \begin{bmatrix} x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix}$$

$$T_2(x) = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ x_1 + x_2 + x_3 \\ \vdots \\ x_1 + x_2 + \cdots + x_n \end{bmatrix}$$

`diff` and `cumsum` commands in MATLAB

what is the associated matrix A for each transformation ?

Image transformation

cropping a 1200×850-pixel image to 490×430-pixel image

transformation of a matrix of $M \times N$ to the size of $m \times n$

$$T : \mathbf{R}^{M \times N} \rightarrow \mathbf{R}^{m \times n}, \quad T(X) = AXB$$

where A selects the rows of X and B selects the columns of X

Image of linear transformation

let $\mathcal{V}$ and $\mathcal{W}$ be vector spaces and a basis for $\mathcal{V}$ is

$$S = \{v_1, v_2, \ldots, v_n\}$$

let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation

the image of any vector $v \in \mathcal{V}$ under T can be expressed by

$$T(v) = a_1 T(v_1) + a_2 T(v_2) + \cdots + a_n T(v_n)$$

where $a_1, a_2, \ldots, a_n$ are coefficients used to express v, *i.e.*,

$$v = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n$$

(follow from the linear property of T)

Definition

let $T : X \to Y$ be a linear transformation from X to Y

Definitions:

kernel of T is the set of vectors in X that T maps into 0

$$\mathbf{ker}(T) = \{x \in X \mid T(x) = 0\}$$

range of T is the set of all vectors in Y that are images under T

$$\mathcal{R}(T) = \{y \in Y \mid y = T(x), \quad x \in X\}$$

Theorem ✎

- $\mathbf{ker}(T)$ is a subspace of X
- $\mathcal{R}(T)$ is a subspace of Y

Example

matrix transformation: $T : \mathbf{R}^n \to \mathbf{R}^m, \quad T(x) = Ax$

- $\mathbf{ker}(T) = \mathcal{N}(A)$: kernel of T is the nullspace of A
- $\mathcal{R}(T) = \mathcal{R}(A)$: range of T is the range (column) space of A

zero transformation: $T : \mathbf{R}^n \to \mathbf{R}^m, \quad T(x) = 0$

$$\mathbf{ker}(T) = \mathbb{R}^n, \quad \mathcal{R}(T) = \{0\}$$

identity operator: $T : \mathcal{V} \to \mathcal{V}, \quad T(x) = x$

$$\mathbf{ker}(T) = \{0\}, \quad \mathcal{R}(T) = \mathcal{V}$$

differentiation: $T : C^1(-\infty, \infty) \to F(-\infty, \infty), \quad T(f) = f'$

$\mathbf{ker}(T)$ is the set of constant functions on $(-\infty, \infty)$

Rank and Nullity

rank of a linear transformation $T : X \to Y$ is defined as

$$\mathbf{rank}(T) = \dim \mathcal{R}(T)$$

nullity of a linear transformation $T : X \to Y$ is defined as

$$\mathbf{nullity}(T) = \dim \mathbf{ker}(T)$$

(provided that $\mathcal{R}(T)$ and $\mathbf{ker}(T)$ are finite-dimensional)

redrank-Nullity theorem: suppose X is a finite-dimensional vector space

$$\mathbf{rank}(T) + \mathbf{nullity}(T) = \dim(X)$$

Proof of rank-nullity theorem

- assume $\dim(X) = n$
- assume a nontrivial case: $\dim \mathbf{ker}(T) = r$ where $1 < r < n$
- let $\{v_1, v_2, \ldots, v_r\}$ be a basis for $\mathbf{ker}(T)$
- let $W = \{v_1, v_2, \ldots, v_r\} \cup \{v_{r+1}, v_{r+2}, \ldots, v_n\}$ be a basis for X
- we can show that

$$S = \{T(v_{r+1}), \ldots, T(v_n)\}$$

  forms a basis for $\mathcal{R}(T)$ ($\therefore$ complete the proof since $\dim\ S = n - r$)

$\mathbf{span}\ S = \mathcal{R}(T)$

- for any $z \in \mathcal{R}(T)$, there exists $v \in X$ such that $z = T(v)$
- since W is a basis for X, we can represent $v = \alpha_1 v_1 + \cdots + \alpha_n v_n$
- we have $z = \alpha_{r+1} T(v_{r+1}) + \cdots + \alpha_n T(v_n)$ ($\because\ v_1, \ldots, v_r \in \mathbf{ker}(T)$)

S **is linearly independent**, *i.e.*, we must show that

$$\alpha_{r+1}T(v_{r+1}) + \cdots + \alpha_n T(v_n) = 0 \quad \Longrightarrow \quad \alpha_{r+1} = \cdots = \alpha_n = 0$$

- since T is linear

$$\alpha_{r+1}T(v_{r+1}) + \cdots + \alpha_n T(v_n) = T(\alpha_{r+1}v_{r+1} + \cdots + \alpha_n v_n) = 0$$

- this implies $\alpha_{r+1}v_{r+1} + \cdots + \alpha_n v_n \in \mathbf{ker}(T)$

$$\alpha_{r+1}v_{r+1} + \cdots + \alpha_n v_n = \alpha_1 v_1 + \alpha_2 v_2 + \cdots \alpha_r v_r$$

- since $\{v_1, \ldots, v_r, v_{r+1}, \ldots, v_n\}$ is linear independent, we must have

$$\alpha_1 = \cdots = \alpha_r = \alpha_{r+1} = \cdots = \alpha_n = 0$$

One-to-one transformation

a linear transformation $T : X \to Y$ is said to be **one-to-one** if

$$\forall x, z \in X \qquad T(x) = T(z) \quad \Longrightarrow \quad x = z$$

- T never maps distinct vectors in X to the same vector in Y
- also known as **injective** transformation

✌ **Theorem:** T is *one-to-one* if and only if $\mathbf{ker}(T) = \{0\}$, *i.e.*,

$$T(x) = 0 \quad \Longrightarrow \quad x = 0$$

- for $T(x) = Ax$ where $A \in \mathbf{R}^{n \times n}$,

$$T \text{ is one-to-one} \quad \Longleftrightarrow \quad A \text{ is } \textit{invertible}$$

Onto transformation

a linear transformation $T : X \to Y$ is said to be **onto** if

for **every** vector $y \in Y$, there exists a vector $x \in X$ such that

$$y = T(x)$$

- every vector in Y is the image of at least one vector in X
- also known as **surjective** transformation

✌ **Theorem:** T is onto if and only if $\mathcal{R}(T) = Y$

✌ **Theorem:** for a *linear operator* $T : X \to X$,

$$T \text{ is one-to-one if and only if } T \text{ is onto}$$

Examples

✎ which of the following is a one-to-one transformation ?

- $T : \mathbf{P}_n \to \mathbf{R}^{n+1}$

$$T(p(t)) = T(a_0 + a_1 t + \cdots + a_n t^n) = (a_0, a_1, \ldots, a_n)$$

- $T : \mathbf{P}_n \to \mathbf{P}_{n+1}$

$$T(p(t)) = tp(t)$$

- $T : \mathbf{R}^{m\times n} \to \mathbf{R}^{n\times m}, \quad T(X) = X^T$
- $T : \mathbf{R}^{n\times n} \to \mathbf{R}, \quad T(X) = \mathbf{tr}(X)$
- $T : C^1(-\infty, \infty) \to F(-\infty, \infty), \quad T(f) = f'$

Matrix transformation

consider a linear transformation $T : \mathbf{R}^n \to \mathbf{R}^m$,

$$T(x) = Ax, \qquad A \in \mathbf{R}^{m \times n}$$

✌ **Theorem:** the following statements are equivalent

- T is **one-to-one**
- the homogeneous equation $Ax = 0$ has only the trivial solution $(x = 0)$
- $\mathbf{rank}(A) = n$

✌ **Theorem:** the following statements are equivalent

- T is **onto**
- for every $b \in \mathbf{R}^m$, the linear system $Ax = b$ always has a solution
- $\mathbf{rank}(A) = m$

Isomorphism

a linear transformation $T : X \to Y$ is said to be an **isomorphism** if

$$T \text{ is both one-to-one and onto}$$

if there exists an isomorphism between X and Y, the two vector spaces are said to be **isomorphic**

✌ **Theorem:**

- for any n-dimensional vector space X, there always exists a linear transformation $T : X \to \mathbf{R}^n$ that is one-to-one and onto (for example, a coordinate map)
- every real n-dimensional vector space is isomorphic to $\mathbf{R}^n$

Examples

- $T : \mathbf{P}_n \rightarrow \mathbf{R}^{n+1}$

$$T(p(t)) = T(a_0 + a_1 t + \cdots + a_n t^n) = (a_0, a_1, \ldots, a_n)$$

  $\mathbf{P}_n$ is isomorphic to $\mathbf{R}^{n+1}$

- $T : \mathbf{R}^{2\times 2} \rightarrow \mathbf{R}^4$

$$T\left(\begin{bmatrix} a_1 & a_2 \\ a_3 & a_4 \end{bmatrix}\right) = (a_1, a_2, a_3, a_4)$$

  $\mathbf{R}^{2\times 2}$ is isomorphic to $\mathbf{R}^4$

in these examples, we observe that

- T maps a vector into its coordinate vector relative to a standard basis
- for any two finite-dimensional vector spaces that are isomorphic, they have the same dimension

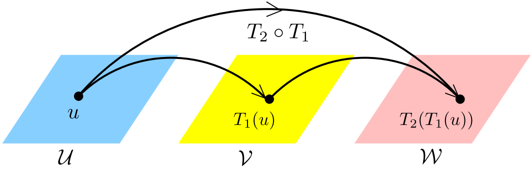
Composition of linear transformation

let $T_1 : \mathcal{U} \to \mathcal{V}$ and $T_2 : \mathcal{V} \to \mathcal{W}$ be linear transformations

the **composition** of T_2 *with* T_1 is the function defined by

$$(T_2 \circ T_1)(u) = T_2(T_1(u))$$

where u is a vector in $\mathcal{U}$



Theorem ✎ if T_1, T_2 are linear, so is $T_2 \circ T_1$

Examples

example 1: $T_1 : \mathbf{P}_1 \to \mathbf{P}_2, \quad T_2 : \mathbf{P}_2 \to \mathbf{P}_2$

$$T_1(p(t)) = tp(t), \quad T_2(p(t)) = p(2t+4)$$

then the composition of T_2 with T_1 is given by

$$(T_2 \circ T_1)(p(t)) = T_2(T_1(p(t))) = T_2(tp(t)) = (2t+4)p(2t+4)$$

example 2: $T : \mathcal{V} \to \mathcal{V}$ is a linear operator, $I : \mathcal{V} \to \mathcal{V}$ is identity operator

$$(T \circ I)(v) = T(I(v)) = T(v), \quad (I \circ T)(v) = I(T(v)) = T(v)$$

hence, $T \circ I = T$ and $I \circ T = T$

example 3: $T_1 : \mathbf{R}^n \to \mathbf{R}^m$, $T_2 : \mathbf{R}^m \to \mathbf{R}^n$ with

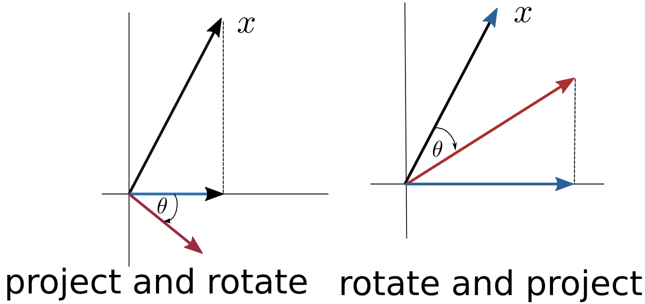
$$T_1(x) = Ax, \quad T_2(w) = Bw, \quad A \in \mathbf{R}^{m \times n}, B \in \mathbf{R}^{n \times m}$$

then $T_1 \circ T_2 = AB$ and $T_2 \circ T_1 = BA$

Order of operations matters

let $T_1, T_2 : \mathbf{R}^2 \rightarrow \mathbf{R}^2$ be the following matrix transformations

- $T_1(x)$ is the projection of x on the x_1-axis
- $T_2(x)$ is the rotation of x by θ (clockwise direction)



the composite of T_2 **with** T_1 VS the composite of T_1 **with** T_2

which is which ?

Nonlinear composite transformations

composite transformations can be defined for nonlinear mappings

many examples in applications:

- $T_1 : \mathbf{R}^n \to \mathbf{R}$ and $T_2 : \mathbf{R} \to \mathbf{R}$ **norm-squared**

$$T_1(x) = \|x\|_2, \quad T_2(x) = x^2 \quad \Rightarrow \quad (T_2 \circ T_1)(x) = \|x\|_2^2 = x^T x$$

- $T_1 : \mathbf{R}^n \to \mathbf{R}^n$ and $T_2 : \mathbf{R}^m \to \mathbf{R}$ **norm of affine**

$$T_1(x) = Ax + b, \quad T_2(x) = \|x\|_2^2 \quad \Rightarrow \quad (T_2 \circ T_1)(x) = \|Ax + b\|_2^2$$

- $T_1 : \mathbf{R}^n \to \mathbf{R}^m$ and $T_2 : \mathbf{R}^m \to \mathbf{R}^m$ **transform in neural network**

$$T_1(x) = Wx + b, \quad T_2(x) = \max(0, x) \Rightarrow (T_2 \circ T_1)(x) = \max(0, Wx + b)$$

Two operators cancel each other

scaling operators: $T_1, T_2 : \mathbf{R}^n \to \mathbf{R}^n$

$$\begin{aligned} T_1(x_1, x_2, \ldots, x_n) &= (a_1 x_1, a_2 x_2, \ldots, a_n x_n) \\ T_2(x_1, x_2, \ldots, x_n) &= (x_1/a_1, x_2/a_2, \ldots, x_n/a_n), \quad \forall a_k \neq 0 \\ (T_2 \circ T_1)(x) &= (T_1 \circ T_2)(x) = x \end{aligned}$$

shift operators: $T_1, T_2 : \mathbf{R}^n \to \mathbf{R}^n$

$$\begin{aligned} T_1(x_1, x_2, \ldots, x_n) &= (x_2, x_3, x_4, \ldots, x_n, x_1) \\ T_2(x_1, x_2, \ldots, x_n) &= (x_n, x_1, x_2, \ldots, x_{n-2}, x_{n-1}) \\ (T_2 \circ T_1)(x) &= T_2(x_2, x_3, \ldots, x_n, x_1) = x \\ (T_1 \circ T_2)(x) &= T_1(x_n, x_1, \ldots, x_{n-2}, x_{n-1}) = x \end{aligned}$$

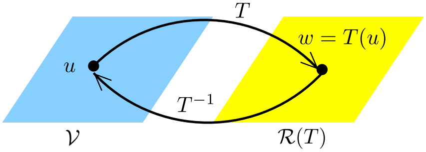
in these examples, T_2 brings the image under T_1 back to the original x !

Inverse of linear transformation

a linear transformation $T : \mathcal{V} \to \mathcal{W}$ is **invertible** if there is a transformation $S : \mathcal{W} \to \mathcal{V}$ satisfying

$$S \circ T = I_{\mathcal{V}} \quad \text{and} \quad T \circ S = I_{\mathcal{W}}$$

we call S the **inverse** of T and denote $S = T^{-1}$



$$T^{-1}(T(u)) = u \quad \forall u \in \mathcal{U}$$
$$T(T^{-1}(w)) = w \quad \forall w \in \mathcal{R}(T)$$

Facts:

- the inverse transformation $T^{-1} : \mathcal{R}(T) \to \mathcal{V}$ exists if and only if T is one-to-one
- $T^{-1} : \mathcal{R}(T) \to \mathcal{V}$ is also linear

✎

Inverse of matrix transformation

consider $T : \mathbf{R}^n \to \mathbf{R}^n$ where $T(x) = Ax$

- T is one-to-one if and only if A is invertible
- T^{-1} exists if and only if A is invertible

the inverse transformation must satisfy

$$T^{-1}(T(x)) = T^{-1}(Ax) = x, \quad \forall x \in \mathbf{R}^n$$

to find the description of T^{-1}

let $y = Ax$ and since A^{-1} exists, we can write $x = A^{-1}y$

$$T^{-1}(Ax) = T^{-1}(y) = A^{-1}y$$

this holds for all $y \in \mathbf{R}^n$ (since $y \in \mathcal{R}(A) = \mathbf{R}^n$)

conclusion: the inverse transformation is simply the matrix transformation given by A^{-1}

Inverse of difference operator

$$T : \mathbf{R}^n \to \mathbf{R}^n, \quad T(x) = \begin{bmatrix} x_1 \\ x_2 - x_1 \\ x_3 - x_2 \\ \vdots \\ x_n - x_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix} x \triangleq Ax$$

does T have an inverse ? if yes, what would it be ?

- please check ✎ that A is invertible and therefore T^{-1} exists
- $T^{-1}(x)$ is given

$$T^{-1}(x) = A^{-1}x = \begin{bmatrix} 1 & & & \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & 1 & 1 & 1 \end{bmatrix} x = \begin{bmatrix} x_1 \\ x_1 + x_2 \\ \vdots \\ x_1 + x_2 + \cdots + x_n \end{bmatrix}$$

T^{-1} is the cumulative sum operator ! (difference cancels with sum)

Inverse of transformation on $\mathbf{P}_n$

$T : \mathbf{P}_1 \to \mathbf{P}_1$, $T(p(x)) = p(x+c)$ where $c \in \mathbf{R}$ is given

- it can be verified ✎ that T is linear and one-to-one
- let $p(x) = a_0 + a_1 x$ be any polynomial in $\mathbf{P}_1$, T^{-1} must satisfy

$$T^{-1}(T(p(x)) = T^{-1}(a_0 + a_1(x+c)) = p(x) = a_0 + a_1 x, \quad \forall a_0, a_1 \in \mathbf{R}$$

- to find description of T^{-1}, let $q(x) = b_0 + b_1 x \triangleq a_0 + a_1(x+c)$ and we should write a_0, a_1 in terms of b_0, b_1

$$b_0 + b_1 x = a_0 + a_1 c + a_1 x \quad \Rightarrow \quad a_0 = b_0 - b_1 c, \;\; a_1 = b_1$$

- we can write $T^{-1}(b_0 + b_1 x) = b_0 - b_1 c + b_1 x = b_0 + b_1(x - c)$

it shows that $T^{-1}(q(x)) = q(x-c)$ (forward translation $x + c$ cancels with backward translation $x - c$)

Domain of T^{-1} may not be the whole co-domain of T

$T : \mathbf{R}^2 \to \mathbf{R}^{2\times 2}$ and given $a, c \neq 0$

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} ax_1 & 0 \\ 0 & cx_2 \end{bmatrix}$$

we can verify that ✎

- T is linear and one-to-one (hence, T^{-1} exists)
- $\mathcal{R}(T) = \mathrm{span}\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$ (not the whole $\mathbf{R}^{2\times 2}$)

$T^{-1} : \mathcal{R}(T) \to \mathbf{R}^2$ is defined from $\mathcal{R}(T)$ and must satisfy

$$T^{-1}\left(\begin{bmatrix} ax_1 & 0 \\ 0 & cx_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

it follows that $T^{-1}(Y) = (y_{11}/a, y_{22}/c)$ where $Y \in \mathcal{R}(T)$

Composition of one-to-one linear transformation

if $T_1 : \mathcal{U} \rightarrow \mathcal{V}$ and $T_2 : \mathcal{V} \rightarrow \mathcal{W}$ are one-to-one linear transformation, then

- $T_2 \circ T_1$ is one-to-one
- $(T_2 \circ T_1)^{-1} = T_1^{-1} \circ T_2^{-1}$

example: $T_1 : \mathbf{R}^n \rightarrow \mathbf{R}^n$, $\quad T_2 : \mathbf{R}^n \rightarrow \mathbf{R}^n$

$$\begin{aligned} T_1(x_1, x_2, \ldots, x_n) &= (a_1x_1, a_2x_2, \ldots, a_nx_n), \quad a_k \neq 0, k = 1, \ldots, n \\ T_2(x_1, x_2, \ldots, x_n) &= (x_2, x_3, \ldots, x_n, x_1) \end{aligned}$$

both T_1 and T_2 are invertible and the inverses are

$$\begin{aligned} T_1^{-1}(w_1, w_2, \ldots, w_n) &= ((1/a_1)w_1, (1/a_2)w_2, \ldots, (1/a_n)w_n) \\ T_2^{-1}(w_1, w_2, \ldots, w_n) &= (w_n, w_1, \ldots, w_{n-1}) \end{aligned}$$

from a direct calculation, the composition of T_1^{-1} with T_2^{-1} is

$$\begin{aligned}(T_1^{-1} \circ T_2^{-1})(w) &= T_1^{-1}(w_n, w_1, \ldots, w_{n-1}) \\ &= ((1/a_1)w_n, (1/a_2)w_1, \ldots, (1/a_n w_{n-1}))\end{aligned}$$

now consider the composition of T_2 with T_1

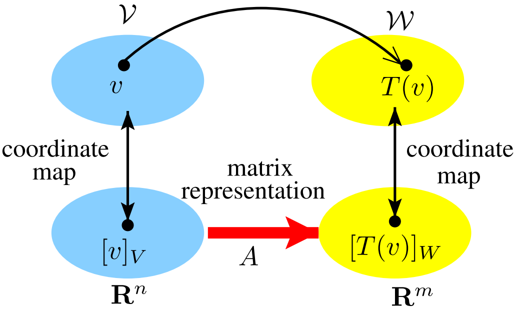
$$(T_2 \circ T_1)(x) = (a_2 x_2, \ldots, a_n x_n, a_1 x_1)$$

it is clear to see that

$$(T_2 \circ T_1) \circ (T_1^{-1} \circ T_2^{-1}) = I$$

Matrix representation for linear transformation

let $T : \mathcal{V} \to \mathcal{W}$ be a linear transformation



V is a basis for $\mathcal{V}$

$$\dim \mathcal{V} = n$$

W is a basis for $\mathcal{W}$

$$\dim \mathcal{W} = m$$

how to represent an image of T in terms of its coordinate vector ?

problem: find a matrix $A \in \mathbf{R}^{m\times n}$ that maps $[v]_V$ into $[T(v)]_W$

Key idea

the matrix A must satisfy

$$A[v]_V = [T(v)]_W, \quad \textbf{for all} \;\; v \in \mathcal{V}$$

hence, it suffices to hold *for all vector in a basis* for $\mathcal{V}$

suppose a basis for $\mathcal{V}$ is $V = \{v_1, v_2, \ldots, v_n\}$

$$A[v_1] = [T(v_1)], \quad A[v_2] = [T(v_2)], \quad \ldots, \quad A[v_n] = [T(v_n)]$$

(we have dropped the subscripts that refer to the choice of bases V, W

A is a matrix of size $m \times n$, so we can write A as

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \end{bmatrix}$$

where a_k's are the columns of A

the coordinate vectors of v_k's are simply the standard unit vectors

$$[v_1] = e_1, \quad [v_2] = e_2, \quad \ldots, \quad [v_n] = e_n$$

hence, we have

$$A[v_1] = a_1 = [T(v_1)], \quad A[v_2] = a_2 = [T(v_2)], \quad \cdots, \quad A[v_n] = a_n = [T(v_n)]$$

stack these vectors back in A

$$A = \begin{bmatrix} [T(v_1)] & [T(v_2)] & \cdots & [T(v_n)] \end{bmatrix}$$

- the columns of A are the coordinate maps of the images of the basis vectors in $\mathcal{V}$
- we call A the **matrix representation** for T relative to the bases V and W and denote it by
$$[T]_{W,V}$$
- a matrix representation *depends* on the **choice of bases** for $\mathcal{V}$ and $\mathcal{W}$

special case: $T : \mathbf{R}^n \to \mathbf{R}^m$, $T(x) = Bx$ we have $[T] = B$ relative to the *standard bases* for $\mathbf{R}^m$ and $\mathbf{R}^n$

Example 1

$T : \mathcal{V} \to \mathcal{W}$ where

$$\begin{array}{rclll} \mathcal{V} & = & \mathbf{P}_1 & \text{with a basis} & V = \{1, t\} \\ \mathcal{W} & = & \mathbf{P}_1 & \text{with a basis} & W = \{t-1, t\} \end{array}$$

define $T(p(t)) = p(t+1)$, find $[T]$ relative to V and W

solution.

find the mappings of vectors in V and their coordinates relative to W

$$\begin{array}{rclcl} T(v_1) = T(1) & = & 1 & = & -1 \cdot (t-1) + 1 \cdot t \\ T(v_2) = T(t) & = & t+1 & = & -1 \cdot (t-1) + 2 \cdot t \end{array}$$

hence $[T(v_1)]_W = (-1, 1)$ and $[T(v_2)]_W = (-1, 2)$

$$[T]_{WV} = \begin{bmatrix} [T(v_1)]_W & [T(v_2)]_W \end{bmatrix} = \begin{bmatrix} -1 & -1 \\ 1 & 2 \end{bmatrix}$$

Example 2

given a matrix representation for $T : \mathbf{P}_2 \to \mathbf{R}^2$

$$[T] = \begin{bmatrix} 5 & 2 & -1 \\ 3 & 0 & 4 \end{bmatrix}$$

relative to the bases $V = \{2 - t, t + 1, t^2 - 1\}$ and $W = \{(1, 0), (1, 1)\}$

find the image of $6t^2$ under T

solution. find the coordinate of $6t^2$ relative to V by writing

$$6t^2 = \alpha_1 \cdot (2 - t) + \alpha_2 \cdot (t + 1) + \alpha_3 \cdot (t^2 - 1)$$

solving for $\alpha_1, \alpha_2, \alpha_3$ gives

$$[6t^2]_V = \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix}$$

from the definition of $[T]$:

$$[T(6t^2)]_W = [T]_{WV}[6t^2]_V = \begin{bmatrix} 5 & 2 & -1 \\ 3 & 0 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ 2 \\ 6 \end{bmatrix} = \begin{bmatrix} 8 \\ 30 \end{bmatrix}$$

then we read from $[T(6t^2)]_W$ that

$$T(6t^2) = 8 \cdot (1,0) + 30 \cdot (1,1) = (38,30)$$

Matrix representation for linear operators

we say T is a **linear operator** if T is a linear transformation from $\mathcal{V}$ to $\mathcal{V}$

- typically we use the same basis for $\mathcal{V}$, says $V = \{v_1, v_2, \dots, v_n\}$
- a matrix representation for T relative to V is denoted by $[T]_V$ where

$$[T]_V = \begin{bmatrix} [T(v_1)] & [T(v_2)] & \dots & [T(v_n)] \end{bmatrix}$$

Theorem ✌

- T is one-to-one if and only if $[T]_V$ is invertible
- $[T^{-1}]_V = ([T]_V)^{-1}$

what is the matrix (relative to a basis) for the identity operator ?

Matrix representation for composite transformation

if $T_1 : \mathcal{U} \to \mathcal{V}$ and $T_2 : \mathcal{V} \to \mathcal{W}$ are linear transformations

and U, V, W are bases for $\mathcal{U}, \mathcal{V}, \mathcal{W}$ respectively

then

$$[T_2 \circ T_1]_{W,U} = [T_2]_{W,V} \cdot [T_1]_{V,U}$$

example: $T_1 : \mathcal{U} \to \mathcal{V}$, $T_2 : \mathcal{V} \to \mathcal{W}$

$$\mathcal{U} = \mathbf{P}_1, \quad \mathcal{V} = \mathbf{P}_2, \quad \mathcal{W} = \mathbf{P}_3$$

$$U = \{1, t\}, \quad V = \{1, t, t^2\}, \quad W = \{1, t, t^2, t^3\}$$

$$T_1(p(t)) = T_1(a_0 + a_1 t) = 2a_0 - 3a_1 t$$

$$T_2(p(t)) = 3tp(t)$$

find $[T_2 \circ T_1]$

solution. first find $[T_1]$ and $[T_2]$

$$\begin{array}{lllll} T_1(1) & = & 2 & = & 2\cdot 1 + 0\cdot t + 0\cdot t^2 \\ T_1(t) & = & -3t & = & 0\cdot 1 - 3\cdot t + 0\cdot t^2 \end{array} \implies [T_1] = \begin{bmatrix} 2 & 0 \\ 0 & -3 \\ 0 & 0 \end{bmatrix}$$

$$\begin{array}{lllll} T_2(1) & = & 3t & = & 0\cdot 1 + 3\cdot 1 + 0\cdot t^2 + 0\cdot t^3 \\ T_2(t) & = & 3t^2 & = & 0\cdot 1 + 0\cdot 1 + 3\cdot t^2 + 0\cdot t^3 \\ T_2(t^2) & = & 3t^3 & = & 0\cdot 1 + 0\cdot 1 + 0\cdot t^2 + 3\cdot t^3 \end{array} \implies [T_2] = \begin{bmatrix} 0 & 0 & 0 \\ 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

next find $[T_2 \circ T_1]$

$$\begin{array}{lllll} (T_2 \circ T_1)(1) & = & T_2(2) & = & 6t \\ (T_2 \circ T_1)(t) & = & T_2(-3t) & = & -9t^2 \end{array} \implies [T_2 \circ T_1] = \begin{bmatrix} 0 & 0 \\ 6 & 0 \\ 0 & -9 \\ 0 & 0 \end{bmatrix}$$

easy to verify that $[T_2 \circ T_1] = [T_2] \cdot [T_1]$

References

1. H.Anton and C. Rorres, *Elementary Linear Algebra*, John Wiley, 2011
2. W.K. Nicholson, *Linear Algebra with Applications*, McGraw-Hill, 2006
3. S. Boyd and L. Vandenberghe, *Introduction to Applied Linear Algebra: Vectors, Matrices, and Least squares*, Cambridge, 2018