Lecture 3: Linear Algebra

CSE4130: 기초머신러닝

Junsuk Choe (최준석)

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

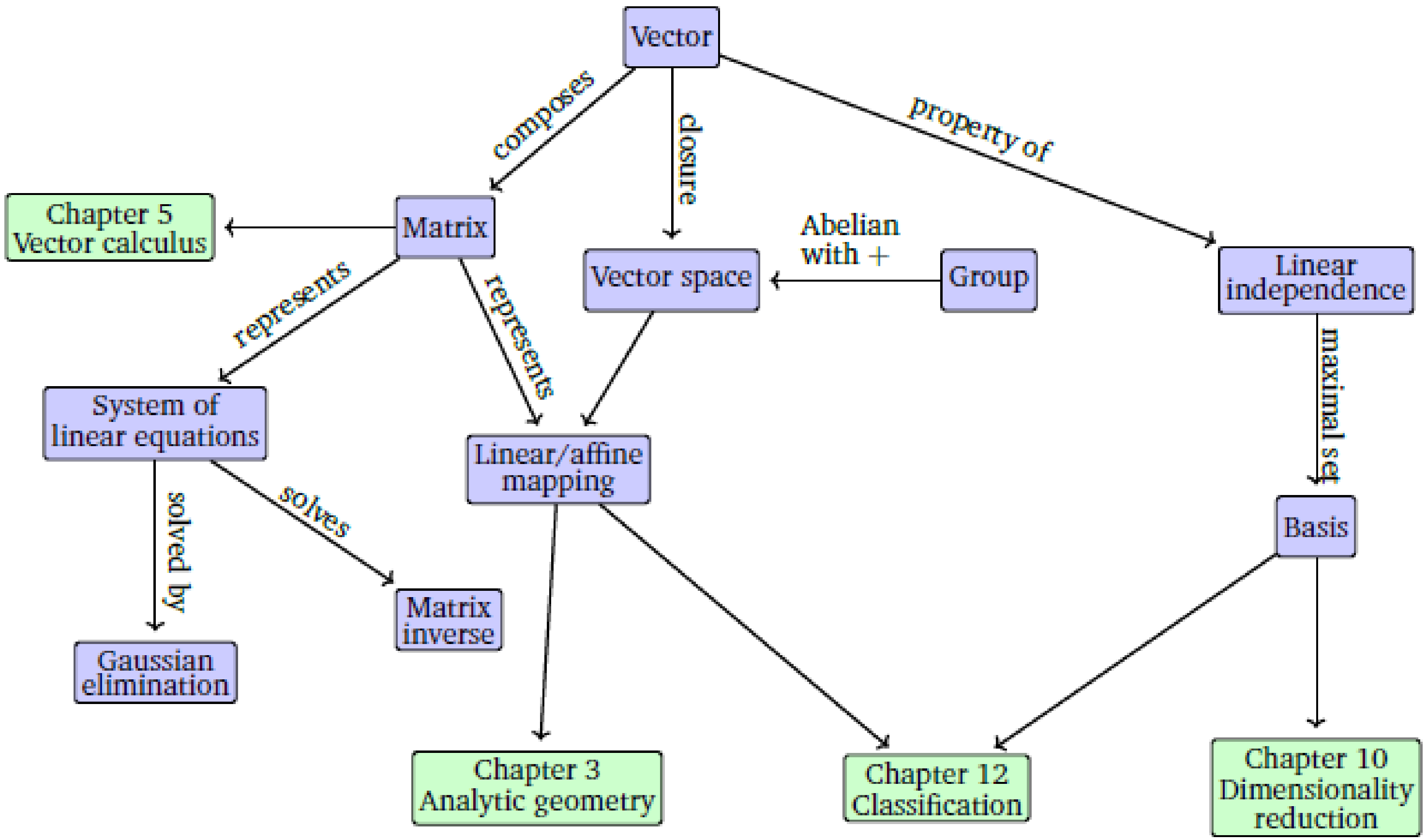
(5) Linear Independence

(6) Basis and Rank

(7) Linear Mappings

(8) Affine Spaces

Roadmap



Vector
composes
closure
property of
Chapter 5
Vector calculus
Matrix
Abelian
with +
Vector space
Group
Linear
independence
represents
represents
maximal set
System of
linear equations
Linear/affine
mapping
solves
Basis
solved by
Matrix
inverse
Gaussian
elimination
Chapter 3
Analytic geometry
Chapter 12
Classification
Chapter 10
Dimensionality
reduction

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

(5) Linear Independence

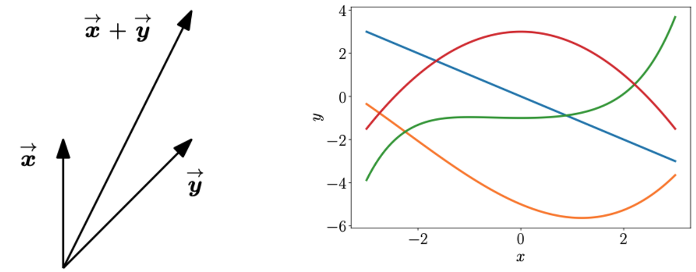
(6) Basis and Rank

(7) Linear Mappings

(8) Affine Spaces

Linear Algebra

- Algebra: a set of objects and a set of rules or operations to manipulate those objects
- Linear algebra
  - Object: vectors $\boldsymbol{v}$
  - Operations: their additions ($\boldsymbol{v} + \boldsymbol{w}$) and scalar multiplication ($k\boldsymbol{v}$)
- Examples
  - Geometric vectors
    - High school physics
  - Polynomials
  - Audio signals
  - Elements of $\mathbb{R}^n$



(a) Geometric vectors.
(b) Polynomials.

System of Linear Equations

- For unknown variables $(x_1, \cdots, x_n) \in \mathbb{R}^n$,

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- Three cases of solutions

- No solution

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

- Unique solution

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

- Infinitely many solutions

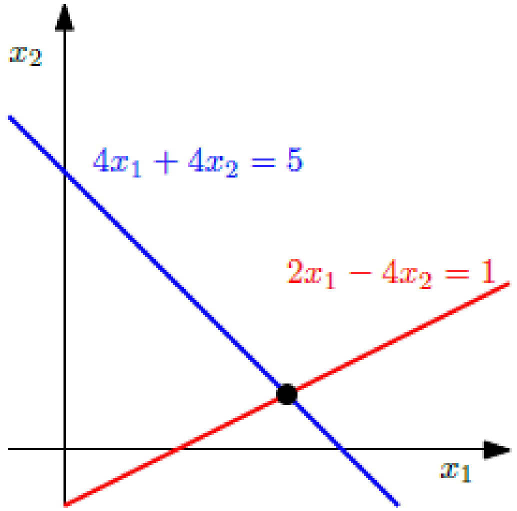
$$\begin{aligned} x_1 + x_2 + x_3 &= 3 \\ x_1 - x_2 + 2x_3 &= 2 \\ 2x_1 \phantom{{}-x_2} + 3x_3 &= 5 \end{aligned}$$

- Question. Under what conditions, one of the above three cases occur?

Geometric Interpretation

- In a system of linear equations with two variables x_1, x_2, each linear equation defines a line on the x_1x_2-plane.
- Since a solution to a system of linear equations must satisfy all equations simultaneously, the solution set is the intersection of these lines.
- This intersection set can be a line (if the linear equations describe the same line), a point, or empty (when the lines are parallel).

$$
\begin{aligned}
4x_1 + 4x_2 &= 5 \\
2x_1 - 4x_2 &= 1
\end{aligned}
$$



Matrix Representation

- A collection of linear equations

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &= b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

- Matrix representations:

$$\begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \end{pmatrix} x_1 + \cdots + \begin{pmatrix} a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} x_n = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \iff \underbrace{\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}}_{\boldsymbol{A}} \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\boldsymbol{x}} = \underbrace{\begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}}_{\boldsymbol{b}}$$

- Understanding $\boldsymbol{A}$ is the key to answering various questions about this linear system $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

Roadmap

Matrix: Definition

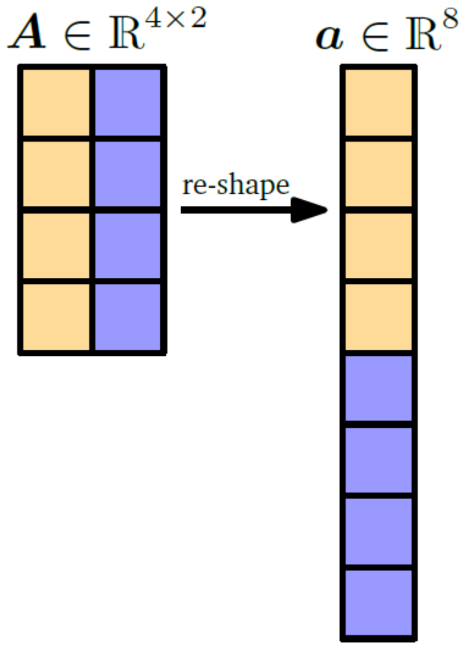
- **Definition 2.1** (Matrix). With $m, n \in \mathbb{N}$ a real-valued (m, n) matrix $\boldsymbol{A}$ is an $m \cdot n$-tuple of elements $a_{ij}, i = 1, ..., m, j = 1, ..., n$, which is ordered according to a rectangular scheme consisting of m rows and n columns:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, a_{ij} \in \mathbb{R}.$$

- By convention $(1, n)$-matrices are called *rows* and $(m, 1)$-matrices are called *columns*. These special matrices are also called *row/column vectors*.

Matrix: Definition

- $\mathbb{R}^{m\times n}$ is the set of all real-valued (m, n)-matrices. $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ can be equivalently represented as $\mathbf{a} \in \mathbb{R}^{mn}$ by stacking all n columns of the matrix into a long vector.



Matrix: Addition and Multiplication

- For two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{m \times n}$,

$$\boldsymbol{A} + \boldsymbol{B} := \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix} \in \mathbb{R}^{m \times n}$$

- For two matrices $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\boldsymbol{B} \in \mathbb{R}^{n \times k}$, the elements c_{ij} of the product $\boldsymbol{C} = \boldsymbol{A}\boldsymbol{B} \in \mathbb{R}^{m \times k}$ is:

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, k.$$

- Example. $\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$ and $\boldsymbol{B} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$, compute $\boldsymbol{A}\boldsymbol{B}$ and $\boldsymbol{B}\boldsymbol{A}$.

Identity Matrix and Matrix Properties

- A square matrix[1] $\boldsymbol{I}_n$ with $I_{ii} = 1$ and $I_{ij=0}$ for $i \neq j$, where n is the number of rows and columns. For example,

$$\boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{I}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

- Associativity: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times p}$, $\boldsymbol{C} \in \mathbb{R}^{p \times q}$, $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- Distributivity: For $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$, and $\boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n \times p}$,
  (i) $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$ and (ii) $\boldsymbol{A}(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD}$
- Multiplication with the identity matrix: For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{I}_m\boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$

[1] $\#$ of rows $= \#$ of cols

Inverse and Transpose

- For a square matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B}$ is the inverse of A, denoted by $\boldsymbol{A}^{-1}$, if
$$\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}.$$
- Called regular/invertible/nonsingular, if it exists.
- If it exists, it is unique.
- How to compute? For 2×2 matrix,
$$\boldsymbol{A}^{-1} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

- For a matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ with $b_{ij} = a_{ji}$ is the transpose of $\boldsymbol{A}$, which we denote by $\boldsymbol{A}^{\mathsf{T}}$.
- Example. For $\boldsymbol{A} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$,
$$\boldsymbol{A}^{\mathsf{T}} = \begin{pmatrix} 0 & 1 & 0 \\ 2 & -1 & 1 \end{pmatrix}$$
- If $\boldsymbol{A} = \boldsymbol{A}^{\mathsf{T}}$, $\boldsymbol{A}$ is called symmetric.

Inverse and Transpose: More Properties

- $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$
- $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
- $(\boldsymbol{A}+\boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$
- $\boldsymbol{A}^{-T} := (\boldsymbol{A}^{-1})^{\mathsf{T}} = (\boldsymbol{A}^{\mathsf{T}})^{-1}$

- $(\boldsymbol{A}^{\mathsf{T}})^{\mathsf{T}} = \boldsymbol{A}$
- $(\boldsymbol{A}+\boldsymbol{B})^{\mathsf{T}} = \boldsymbol{A}^{\mathsf{T}} + \boldsymbol{B}^{\mathsf{T}}$
- $(\boldsymbol{A}\boldsymbol{B})^{\mathsf{T}} = \boldsymbol{B}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}$
- If $\boldsymbol{A}$ is invertible, so is $\boldsymbol{A}^{\mathsf{T}}$.

Scalar Multiplication

- Multiplication by a scalar $\lambda \in \mathbb{R}$ to $\boldsymbol{A} \in \mathbb{R}^{m \times n}$
- Example. For $\boldsymbol{A} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$, $3 \times \boldsymbol{A} = \begin{pmatrix} 0 & 6 \\ 3 & -3 \\ 0 & 3 \end{pmatrix}$
- Associativity
  - $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C})$
  - $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda$
  - $(\lambda\boldsymbol{C})^{\mathsf{T}} = \boldsymbol{C}^{\mathsf{T}}\lambda^{\mathsf{T}} = \boldsymbol{C}^{\mathsf{T}}\lambda = \lambda\boldsymbol{C}^{\mathsf{T}}$
- Distributivity
  - $(\lambda + \psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}$
  - $\lambda(\boldsymbol{B} + \boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}$

Roadmap

Example

$$\begin{aligned}-3x \quad\quad\;\; + 2z &= -1\\ x - 2y + 2z &= -5/3\\ -x - 4y + 6z &= -13/3\end{aligned}$$

- ρ_i: i-th equation
- Express the equation as its augmented matrix.

$$\left(\begin{array}{ccc|c}-3 & 0 & 2 & -1\\ 1 & -2 & 2 & -5/3\\ -1 & -4 & 6 & -13/3\end{array}\right) \xrightarrow[-(1/3)\rho_1+\rho_3]{(1/3)\rho_1+\rho_2} \left(\begin{array}{ccc|c}-3 & 0 & 2 & -1\\ 0 & -2 & 8/3 & -2\\ 0 & -4 & 16/3 & -4\end{array}\right)$$

$$\xrightarrow{-2\rho_2+\rho_3} \left(\begin{array}{ccc|c}-3 & 0 & 2 & -1\\ 0 & -2 & 8/3 & -2\\ 0 & 0 & 0 & 0\end{array}\right)$$

The two nonzero rows give $-3x + 2z = -1$ and $-2y + (8/3)z = -2$.

[1]Examples from this slide to the next several slides come from Jim Hefferson's Linear Algebra book.

- Parametrizing $-3x + 2z = -1$ and $-2y + (8/3)z = -2$ gives:

$$x = (1/3) + (2/3)z$$
$$y = 1 + (4/3)z$$
$$z = z$$

$$\{\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} 2/3 \\ 4/3 \\ 1 \end{pmatrix} z \mid z \in \mathbb{R}\}$$

This helps us understand the set of solutions, e.g., each value of z gives a different solution.

z	0	1	2	$-1/2$
solution $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$	$\begin{pmatrix} 1/3 \\ 1 \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 7/3 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 5/3 \\ 11/3 \\ 2 \end{pmatrix}$	$\begin{pmatrix} 0 \\ 1/3 \\ -1/2 \end{pmatrix}$

Form of solution sets

- The system $\begin{aligned} x + 2y - z \phantom{{}+w} &= 2 \\ 2x - y - 2z + w &= 5 \end{aligned}$ reduces in this way.

$$\left(\begin{array}{cccc|c} 1 & 2 & -1 & 0 & 2 \\ 2 & -1 & -2 & 1 & 5 \end{array}\right) \xrightarrow{-2\rho_1+\rho_2} \left(\begin{array}{cccc|c} 1 & 2 & -1 & 0 & 2 \\ 0 & -5 & 0 & 1 & 1 \end{array}\right)$$

- It has solutions of this form.

$$\begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 12/5 \\ -1/5 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} z + \begin{pmatrix} -2/5 \\ 1/5 \\ 0 \\ 1 \end{pmatrix} w \qquad \text{for } z, w \in \Re$$

- Note that taking $z = w = 0$ shows that the first vector is a particular solution of the system.

General $=$ Particular $+$ Homogeneous

- General approach
  1. Find a particular solution to $\boldsymbol{Ax} = \boldsymbol{b}$
  2. Find all solutions to the homogeneous equation $\boldsymbol{Ax} = \mathbf{0}$
     - $\mathbf{0}$ is a trivial solution
  3. Combine the solutions from steps 1. and 2. to the general solution
- Questions: A formal algorithm that performs the above?
  - Gauss-Jordan method: convert into a "beautiful" form (formally reduced row-echelon form)
  - Elementary transformations: (i) row swapping (ii) multiply by a constant (iii) row addition
- Such a form allows an algorithmic way of solving linear equations

Row-Echelon Form

Definition 2.6 A matrix is in row-echelon form if:

- All rows that contain only zeros are at the bottom of the matrix; correspondingly, all rows that contain at least one nonzero element are on top of rows that contain only zeros.
- Looking at nonzero rows only, the first nonzero number from the left (also called the *pivot* or the *leading coefficient*) is always strictly to the right of the pivot of the row above it.
- The variables corresponding to the pivots in the row-echelon form are called *basic variables* and the other variables are *free variables*.

Row-Echelon Form

- An equation system is in *reduced row-echelon form* if:
  - It is in row-echelon form.
  - Every pivot is 1.
  - The pivot is the only nonzero entry in its column.
- **Gaussian elimination** is an algorithm that performs elementary transformations to bring a system of linear equations into reduced row-echelon form.

Example: Unique Solution

- Start as usual by getting echelon form.

$$\begin{array}{rcrcrcr} x & + & y & - & z & = & 2 \\ 2x & - & y & & & = & -1 \\ x & - & 2y & + & 2z & = & -1 \end{array} \xrightarrow[-1\rho_1+\rho_3]{-2\rho_1+\rho_2} \begin{array}{rcrcrcr} x & + & y & - & z & = & 2 \\ & & -3y & + & 2z & = & -5 \\ & & -3y & + & 3z & = & -3 \end{array} \xrightarrow{-1\rho_2+\rho_3} \begin{array}{rcrcrcr} x & + & y & - & z & = & 2 \\ & & -3y & + & 2z & = & -5 \\ & & & & z & = & 2 \end{array}$$

- Make all the leading entries one.

$$\xrightarrow{(-1/3)\rho_2} \begin{array}{rcrcrcr} x & + & y & - & z & = & 2 \\ & & y & - & (2/3)z & = & 5/3 \\ & & & & z & = & 2 \end{array}$$

- Finish by using the leading entries to eliminate upwards, until we can read off the solution.

$$\begin{array}{rcrcrcr} x & + & y & - & z & = & 2 \\ & & y & - & (2/3)z & = & 5/3 \\ & & & & z & = & 2 \end{array} \xrightarrow[(2/3)\rho_3+\rho_2]{\rho_3+\rho_1} \begin{array}{rcrcrcr} x & + & y & & & = & 4 \\ & & y & & & = & 3 \\ & & & & z & = & 2 \end{array} \xrightarrow{-\rho_2+\rho_1} \begin{array}{rcrcrcr} x & & & & & = & 1 \\ & & y & & & = & 3 \\ & & & & z & = & 2 \end{array}$$

Example: Infinite Number of Solutions

$$\begin{aligned} x - y \qquad - 2w &= 2 \\ x + y + 3z + \ w &= 1 \\ -y + \ z - \ w &= 0 \end{aligned}$$

- Start by getting echelon form and turn the leading entries to 1's.

$$\xrightarrow{-1\rho_1+\rho_2} \left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 2 \\ 0 & 2 & 3 & 3 & -1 \\ 0 & -1 & 1 & -1 & 0 \end{array}\right)$$

$$\xrightarrow{(1/2)\rho_2+\rho_3} \left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 2 \\ 0 & 2 & 3 & 3 & -1 \\ 0 & 0 & 5/2 & 1/2 & -1/2 \end{array}\right)$$

$$\xrightarrow[(2/5)\rho_3]{(1/2)\rho_2} \left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 2 \\ 0 & 1 & 3/2 & 3/2 & -1/2 \\ 0 & 0 & 1 & 1/5 & -1/5 \end{array}\right)$$

- Eliminate upwards.

$$\xrightarrow{-(3/2)\rho_3+\rho_2} \left(\begin{array}{cccc|c} 1 & -1 & 0 & -2 & 2 \\ 0 & 1 & 0 & 6/5 & -1/5 \\ 0 & 0 & 1 & 1/5 & -1/5 \end{array}\right)$$

$$\xrightarrow{\rho_2+\rho_1} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -4/5 & 9/5 \\ 0 & 1 & 0 & 6/5 & -1/5 \\ 0 & 0 & 1 & 1/5 & -1/5 \end{array}\right)$$

- The parameterized solution set is:

$$\{\begin{pmatrix} 9/5 \\ -1/5 \\ -1/5 \\ 0 \end{pmatrix} + \begin{pmatrix} 4/5 \\ -6/5 \\ -1/5 \\ 1 \end{pmatrix} w \mid w \in \mathbb{R}\}$$

Calculating the Inverse

- To compute the inverse $\boldsymbol{A}^{-1}$ of $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, we need to find a matrix $\boldsymbol{X}$ that satisfies $\boldsymbol{AX} = \boldsymbol{I}_n$. Then, $\boldsymbol{X} = \boldsymbol{A}^{-1}$.
- We can write this down as a set of simultaneous linear equations $\boldsymbol{AX} = \boldsymbol{I}_n$, where we solve for $\boldsymbol{X} = [x_1 | \cdots | x_n]$.
- We use the augmented matrix notation for a compact representation of this set of systems of linear equations and obtain $[\boldsymbol{A} | \boldsymbol{I}_n] \rightarrow \cdots \rightarrow [\boldsymbol{I}_n | \boldsymbol{A}^{-1}]$.
- This means that if we bring the augmented equation system into reduced row-echelon form, we can read out the inverse on the right-hand side of the equation system.

Example: Calculating the Inverse

Determine the inverse of

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

Cases of Solution Sets

		number of solutions of the homogeneous system	
		one	*infinitely many*
particular solution exists?	*yes*	unique solution	infinitely many solutions
	no	no solutions	no solutions

Algorithms for Solving System of Linear Equations

1. Pseudo-inverse

   $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \iff \boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^{\mathsf{T}}\boldsymbol{b} \iff \boldsymbol{x} = \left(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathsf{T}}\boldsymbol{b}$$

   - $\left(\boldsymbol{A}^{\mathsf{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathsf{T}}$: *Moore-Penrose pseudo-inverse*
   - many computations: matrix product, inverse, etc
2. Gaussian elimination
   - intuitive and constructive way
   - cubic complexity (in terms of # of simultaneous equations)
3. Iterative methods
   - practical ways to solve indirectly

   (a) stationary iterative methods: Richardson method, Jacobi method, Gaus-Seidel method, successive over-relaxation method

   (b) Krylov subspace methods: conjugate gradients, generalized minimal residual, biconjugate gradients

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

(5) Linear Independence

(6) Basis and Rank

(7) Linear Mappings

(8) Affine Spaces

Group

- A set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \mapsto \mathcal{G}$. $G := (\mathcal{G}, \otimes)$ is called a group, if:
  1. Closure. $\forall x, y \in \mathcal{G}$, $x \otimes y \in \mathcal{G}$
  2. Associativity. $\forall x, y, z \in \mathcal{G}$, $(x \otimes y) \otimes z = x \otimes (y \otimes z)$
  3. Neutral element. $\exists e \in \mathcal{G}$, $\forall x \in \mathcal{G}$, $x \otimes e = x$ and $e \otimes x = x$
  4. Inverse element. $\forall x \in \mathcal{G}$, $\exists y \in \mathcal{G}$, $x \otimes y = e$ and $y \otimes x = e$. We often use $x^{-1} = y$.

- $G = (\mathcal{G}, \otimes)$ is an Abelian group, if the following is additionally met:
  - Communicativity. $\forall x, y \in \mathcal{G}$, $x \otimes y = y \otimes x$

Examples

- $(\mathbb{Z}, +)$ is an Abelian group
- $(\mathbb{N} \cup \{0\}, +)$ is not a group (because inverses are missing)
- $(\mathbb{Z}, \cdot)$ is not a group
- $(\mathbb{R}, \cdot)$ is not a group (because of no inverse for 0)
- $(\mathbb{R}^n, +)$, $(\mathbb{Z}^n, +)$ are Abelian, if $+$ is defined componentwise
- $(\mathbb{R}^{m \times n}, +)$ is Abelian (with componentwise $+$)
- $(\mathbb{R}^{n \times n}, \cdot)$
  - Closure and associativity follow directly
  - Neutral element: $\boldsymbol{I}_n$
  - The inverse $\boldsymbol{A}^{-1}$ may exist or not. So, generally, it is not a group. However, the set of invertible matrices in $\mathbb{R}^{n \times n}$ with matrix multiplication is a group, called general linear group.

General Linear Group

- The set of regular (invertible) matrices $A \in \mathbb{R}^{n \times n}$ is a group with respect to matrix multiplication and is called $GL(n, \mathbb{R})$. However, general linear group since matrix multiplication is not commutative, the group is not Abelian.

Vector Spaces

Definition. A real-valued vector space $V = (\mathcal{V}, +, \cdot)$ is a set $\mathcal{V}$ with two operations

(*a*) $+ : \mathcal{V} \times \mathcal{V} \mapsto \mathcal{V}$ (vector addition)

(*b*) $\cdot : \mathbb{R} \times \mathcal{V} \mapsto \mathcal{V}$ (scalar multiplication),

where

1. $(\mathcal{V}, +)$ is an Abelian group
2. Distributivity.
   - $\forall \lambda \in \mathbb{R}, \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V},\ \lambda \cdot (\boldsymbol{x} + \boldsymbol{y}) = \lambda \cdot \boldsymbol{x} + \lambda \boldsymbol{y}$
   - $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \mathcal{V},\ (\lambda + \psi) \cdot \boldsymbol{x} = \lambda \cdot \boldsymbol{x} + \psi \cdot \boldsymbol{x}$
3. Associativity. $\forall \lambda, \psi \in \mathbb{R}, \boldsymbol{x} \in \boldsymbol{\mathcal{V}},\ \lambda \cdot (\psi \cdot \boldsymbol{x}) = (\lambda \psi) \cdot \boldsymbol{x}$
4. Neutral element. $\forall \boldsymbol{x} \in \mathcal{V},\ 1 \cdot \boldsymbol{x} = \boldsymbol{x}$

Example

- $\mathcal{V} = \mathbb{R}^n$ with
  - Vector addition: $\boldsymbol{x} + \boldsymbol{y} = (x_1 + y_1, \ldots, x_n + y_n)$
  - Scalar multiplication: $\lambda \boldsymbol{x} = (\lambda x_1, \ldots, \lambda x_n)$

- $\mathcal{V} = \mathbb{R}^{m \times n}$ with
  - Vector addition: $\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$
  - Scalar multiplication: $\lambda \boldsymbol{A} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}$

Vector Subspaces

Definition. Consider a vector space $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{U} \subset \mathcal{V}$. Then, $U = (\mathcal{U}, +, \cdot)$ is called vector subspace (simply linear subspace or subspace) of V if U is a vector space with two operations '$+$' and '$\cdot$' restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$.

Examples

- For every vector space V, V and $\{\mathbf{0}\}$ are the trivial subspaces.
- The solution set of $\boldsymbol{Ax} = \mathbf{0}$ is the subspace of $\mathbb{R}^n$.
- The solution of $\boldsymbol{Ax} = \boldsymbol{b}$ ($\boldsymbol{b} \neq \mathbf{0}$) is not a subspace of $\mathbb{R}^n$.
- The intersection of arbitrarily many subspaces is a subspace itself.

Roadmap

Linear Independence

- Definition. For a vector space V and vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in V$, every $\mathbf{v} \in V$ of the form $\mathbf{v} = \lambda_1 \mathbf{x}_1 + \cdots + \lambda_k \mathbf{x}_k$ with $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ is a linear combination of the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \in V$.
- Definition. If there is a non-trivial linear combination such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \mathbf{x}_i$ with at least one $\lambda_i \neq 0$, the vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly dependent. If only the trivial solution exists, i.e., $\lambda_1 = \ldots = \lambda_k = 0$, $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly independent.
- Meaning. A set of linearly independent vectors consists of vectors that have no redundancy.
- Useful fact. The vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ are linearly dependent, iff (at least) one of them is a linear combination of the others.
  - $x - 2y = 2$ and $2x - 4y = 4$ are linearly dependent.

Checking Linear Independence

- Gauss elimination to get the row echelon form
- All column vectors are linearly independent iff all columns are pivot columns (why?).
- Example.

$$\boldsymbol{x}_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \\ 4 \end{pmatrix}, \quad \boldsymbol{x}_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \end{pmatrix}, \quad \boldsymbol{x}_3 = \begin{pmatrix} -1 \\ -2 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & 1 & -2 \\ -3 & 0 & 1 \\ 4 & 2 & 1 \end{pmatrix} \quad \rightsquigarrow \cdots \rightsquigarrow \quad \begin{pmatrix} 1 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

- Every column is a pivot column. Thus, $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3$ are linearly independent.

Linear Combinations of Linearly Independent Vectors

- Vector space V with k linearly independent vectors $\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_k$
- m linear combinations $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$. (Q) Are they linearly independent?

$$\begin{aligned} \boldsymbol{x}_1 &= \lambda_{11}\boldsymbol{b}_1 + \lambda_{21}\boldsymbol{b}_2 + \cdots + \lambda_{k1}\boldsymbol{b}_k \\ &\vdots \\ \boldsymbol{x}_m &= \lambda_{1m}\boldsymbol{b}_1 + \lambda_{2m}\boldsymbol{b}_2 + \cdots + \lambda_{km}\boldsymbol{b}_k \end{aligned}$$

$$\boldsymbol{x}_j = \overbrace{\begin{pmatrix}\boldsymbol{b}_1, & \cdots, & \boldsymbol{b}_k\end{pmatrix}}^{\boldsymbol{B}} \overbrace{\begin{pmatrix}\lambda_{1j} \\ \vdots \\ \lambda_{kj}\end{pmatrix}}^{\boldsymbol{\lambda}_j}, \quad \boldsymbol{x}_j = \boldsymbol{B}\boldsymbol{\lambda}_j$$

- $\sum_{j=1}^{m} \psi_j \boldsymbol{x}_j = \sum_{j=1}^{m} \psi_j \boldsymbol{B}\boldsymbol{\lambda}_j = \boldsymbol{B}\sum_{j=1}^{m} \psi_j \boldsymbol{\lambda}_j$
- $\{\boldsymbol{x}\}$ linearly independent $\iff$ $\{\boldsymbol{\lambda}\}$ linearly independent

Example

$$
\begin{aligned}
\boldsymbol{x}_1 &= \boldsymbol{b}_1 - 2\boldsymbol{b}_2 + \boldsymbol{b}_3 - \boldsymbol{b}_4 \\
\boldsymbol{x}_2 &= -4\boldsymbol{b}_1 - 2\boldsymbol{b}_2 + 4\boldsymbol{b}_4 \\
\boldsymbol{x}_3 &= 2\boldsymbol{b}_1 + 3\boldsymbol{b}_2 - \boldsymbol{b}_3 - 3\boldsymbol{b}_4 \\
\boldsymbol{x}_4 &= 17\boldsymbol{b}_1 - 10\boldsymbol{b}_2 + 11\boldsymbol{b}_3 + \boldsymbol{b}_4
\end{aligned}
$$

Example

$$\begin{aligned}
\boldsymbol{x}_1 &= \boldsymbol{b}_1 - 2\boldsymbol{b}_2 + \boldsymbol{b}_3 - \boldsymbol{b}_4 \\
\boldsymbol{x}_2 &= -4\boldsymbol{b}_1 - 2\boldsymbol{b}_2 + 4\boldsymbol{b}_4 \\
\boldsymbol{x}_3 &= 2\boldsymbol{b}_1 + 3\boldsymbol{b}_2 - \boldsymbol{b}_3 - 3\boldsymbol{b}_4 \\
\boldsymbol{x}_4 &= 17\boldsymbol{b}_1 - 10\boldsymbol{b}_2 + 11\boldsymbol{b}_3 + \boldsymbol{b}_4
\end{aligned}$$

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\lambda}_1 & \boldsymbol{\lambda}_2 & \boldsymbol{\lambda}_3 & \boldsymbol{\lambda}_4 \end{pmatrix} = \begin{pmatrix} 1 & -4 & 2 & 17 \\ -2 & -2 & 3 & -10 \\ 1 & 0 & -1 & 11 \\ -1 & -4 & -3 & 1 \end{pmatrix} \rightsquigarrow \cdots \rightsquigarrow \begin{pmatrix} 1 & 0 & 0 & -7 \\ 0 & 1 & 0 & -15 \\ 0 & 0 & 1 & -18 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- The last column is not a pivot column. Thus, $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_3$ are linearly dependent.

Roadmap

Generating Set and Basis

- Definition. A vector space $V = (\mathcal{V}, +, \cdot)$ and a set of vectors $\mathcal{A} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k\} \subset \mathcal{V}$.
    - If every $v \in \mathcal{V}$ can be expressed as a linear combination of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$, $\mathcal{A}$ is called a generating set of V.
    - The set of all linear combinations of $\mathcal{A}$ is called the span of $\mathcal{A}$.
    - If $\mathcal{A}$ spans the vector space V, we use $V = \text{span}[\mathcal{A}]$ or $V = \text{span}[\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k]$
- Definition. The minimal generating set $\mathcal{B}$ of V is called basis of V. We call each element of $\mathcal{B}$ basis vector. The number of basis vectors is called dimension of V.
- Properties
    - $\mathcal{B}$ is a maximally[2] linearly independent set of vectors in V.
    - Every vector $x \in V$ is a linear combination of $\mathcal{B}$, which is unique.

[2] Adding any other vector to this set will make it linearly dependent.

Examples

- Different bases $\mathbb{R}^3$

$$\mathcal{B}_1 = \{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \}, \mathcal{B}_2 = \{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \},$$

$$\mathcal{B}_3 = \{ \begin{pmatrix} 0.5 \\ 0.8 \\ 0.4 \end{pmatrix}, \begin{pmatrix} 1.8 \\ 0.3 \\ 0.3 \end{pmatrix}, \begin{pmatrix} -2.2 \\ -1.3 \\ 3.5 \end{pmatrix} \}$$

- Linearly independent, but not maximal. Thus, not a basis.

$$\mathcal{A} = \{ \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, \begin{pmatrix} 2 \\ -1 \\ 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \\ -4 \end{pmatrix} \}$$

Determining a Basis

- Want to find a basis of a subspace $U = \text{span}[\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_m]$
    1. Construct a matrix $\boldsymbol{A} = \begin{pmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_m \end{pmatrix}$
    2. Find the row-echelon form of $\boldsymbol{A}$.
    3. Collect the pivot columns.
- Logic: Collect $\boldsymbol{x}_i$ so that we have only trivial solution. Pivot columns tell us which set of vectors is linearly independent.
- See example 2.17 (pp. 35)

Example 2.17 (pp. 35)

Rank (1)

- Definition. The rank of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ denoted by $\mathrm{rk}(\boldsymbol{A})$ is $\#$ of linearly independent columns
  - Same as the number of linearly independent rows
- $$\boldsymbol{A} = \begin{pmatrix} 1 & 2 & 1 \\ -2 & -3 & 1 \\ 3 & 5 & 0 \end{pmatrix} \rightsquigarrow \cdots \rightsquigarrow \begin{pmatrix} 1 & 2 & 1 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \end{pmatrix}$$

  Thus, $\mathrm{rk}(\boldsymbol{A}) = 2$.
- $\mathrm{rk}(\boldsymbol{A}) = \mathrm{rk}(\boldsymbol{A}^{\mathsf{T}})$

Rank (2)

- The columns (resp. rows) of $\boldsymbol{A}$ span a subspace U (resp. W) with $\dim(U) = \text{rk}(\boldsymbol{A})$ (resp. $\dim(W) = \text{rk}(\boldsymbol{A})$), and a basis of U (resp. W) can be found by Gauss elimination of $\boldsymbol{A}$ (resp. $\boldsymbol{A}^{\mathsf{T}}$).

- For all $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\text{rk}(\boldsymbol{A}) = n$, iff $\boldsymbol{A}$ is regular (invertible).

- The linear system $\boldsymbol{Ax} = \boldsymbol{b}$ is solvable, iff $\text{rk}(\boldsymbol{A}) = \text{rk}(\boldsymbol{A}|\boldsymbol{b})$.

- For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, the subspace of solutions for $\boldsymbol{Ax} = \mathbf{0}$ possesses dimension $n - \text{rk}(\boldsymbol{A})$.

- $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ has full rank if its rank equals the largest possible rank for a matrix of the same dimensions. The rank of the full-rank matrix $\boldsymbol{A}$ is $\min(\#\text{ of cols}, \#\text{ of rows})$.

Example 2.18 (Rank)

Roadmap

Linear Mapping (1)

- Interest: A mapping that preserves the structure of the vector space
- Definition. For vector spaces V, W, a mapping $\Phi : V \mapsto W$ is called a linear mapping (or homomorphism/linear transformation), if, for all $\boldsymbol{x}, \boldsymbol{y} \in V$ and all $\lambda \in \mathbb{R}$,
  - $\Phi(\boldsymbol{x} + \boldsymbol{y}) = \Phi(\boldsymbol{x}) + \Phi(\boldsymbol{y})$
  - $\Phi(\lambda \boldsymbol{x}) = \lambda \Phi(\boldsymbol{x})$
- Definition. A mapping $\Phi : \mathcal{V} \mapsto \mathcal{W}$ is called
  - Injective (단사), if $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{V}$, $\Phi(\boldsymbol{x}) = \Phi(\boldsymbol{y}) \implies \boldsymbol{x} = \boldsymbol{y}$
  - Surjective (전사), if $\Phi(\mathcal{V}) = \mathcal{W}$
  - Bijective (전단사), if it is injenctive and surjective.

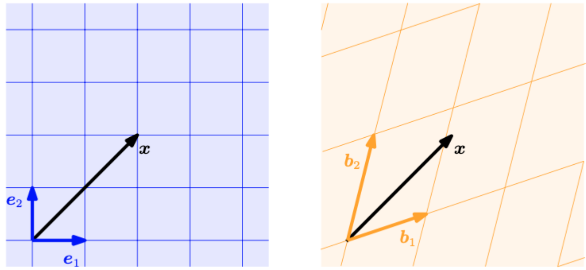
Exampe 2.19 (Homomorphism)

Linear Mapping (2)

- For bjective mapping, there exists an inverse mapping Φ^{-1}.
- Isomorphism if Ψ is linear and bijective.
- Theorem. Vector spaces V and W are isomorphic, iff $\dim(V) = \dim(W)$.
  - Vector spaces of the same dimension are kind of the same thing.
- Other properties
  - For two linear mappings Φ and Ψ, $\Phi \circ \Psi$ is also a linear mapping.
  - If Φ is an isomorphism, so is Φ^{-1}.
  - For two linear mappings Φ and Ψ, $\Phi + \Psi$ and $\lambda\Psi$ for $\lambda \in \mathbb{R}$ are linear.

Coordinates

- A basis defines a coordinate system.



- Consider an ordered basis $B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$ of vector space V. Then, for any $\boldsymbol{x} \in V$, there exists a unique linear combination

  $$\boldsymbol{x} = \alpha_1 \boldsymbol{b}_1 + \ldots + \alpha_n \boldsymbol{b}_n.$$

- We call $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix}$ the coordinate of $\boldsymbol{x}$ with respect to $B = (\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n)$.

- Basis change $\Longrightarrow$ Coordinate change

Basis Change

- Consider a vector space V and two coordinate systems defined by $B = (\boldsymbol{b}_1, \dots, \boldsymbol{b}_n)$ and $B' = (\boldsymbol{b}'_1, \dots, \boldsymbol{b}'_n)$.
- **Question.** For $(x_1, \dots, x_n)_B \to (y_1, \dots, y_n)_{B'}$, what is $(y_1, \dots, y_n)_{B'}$?
- **Theorem.** $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}'_1 & \dots & \boldsymbol{b}'_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_1 & \dots & \boldsymbol{b}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$
- Regard $\boldsymbol{A}_\Phi = \begin{pmatrix} \boldsymbol{b}'_1 & \dots & \boldsymbol{b}'_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_1 & \dots & \boldsymbol{b}_n \end{pmatrix}$ as a linear map

Example

- $B = ((1,0),(0,1))$ and $B' = ((2,1),(1,2))$
- $(4,2)_B \to (x,y)_{B'}$?
- Using $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}'_1 & \dots & \boldsymbol{b}'_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_1 & \dots & \boldsymbol{b}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

Example

- $B = ((1,0),(0,1))$ and $B' = ((2,1),(1,2))$
- $(4,2)_B \to (x,y)_{B'}$?
- Using $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \boldsymbol{b}'_1 & \dots & \boldsymbol{b}'_n \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{b}_1 & \dots & \boldsymbol{b}_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$,

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 4 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \end{pmatrix}$$

Transformation Matrix

- Two vector spaces
  - V with basis $B = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n)$ and W with basis $C = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_m)$
- What is the coordinate in C-system for each basis $\boldsymbol{b}_j$? For $j = 1, \ldots, n$,

$$\Phi(\boldsymbol{b}_j) = \alpha_{1j}\boldsymbol{c}_1 + \cdots + \alpha_{mj}\boldsymbol{c}_m \iff \Phi(\boldsymbol{b}_j) = \begin{pmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{pmatrix} \begin{pmatrix} \alpha_{1j} \\ \vdots \\ \alpha_{mj} \end{pmatrix}$$

$$\implies \begin{pmatrix} \Phi(\boldsymbol{b}_1) & \cdots & \Phi(\boldsymbol{b}_n) \end{pmatrix} = \begin{pmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{pmatrix} \overbrace{\begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1n} \\ \vdots & & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mn} \end{pmatrix}}^{\boldsymbol{A}_\Phi}$$

- $\hat{y} = \boldsymbol{A}_\Phi \hat{x}$, where $\hat{x}$ is the vector w.r.t B and $\hat{y}$ is the vector w.r.t. C

Example 2.21 (Transformation Matrix)

Basis Change: General Case (1)

- For linear mapping $\Phi : V \mapsto W$, consider bases B, B' of V and C, C' of W

  $B = \begin{pmatrix} \boldsymbol{b}_1 & \cdots & \boldsymbol{b}_n \end{pmatrix}, \ B' = \begin{pmatrix} \boldsymbol{b}'_1 & \cdots & \boldsymbol{b}'_n \end{pmatrix} \quad C = \begin{pmatrix} \boldsymbol{c}_1 & \cdots & \boldsymbol{c}_m \end{pmatrix}, \ C' = \begin{pmatrix} \boldsymbol{c}'_1 & \cdots & \boldsymbol{c}'_m \end{pmatrix}.$
- (inter) transformation matrices $\boldsymbol{A}_\Phi$ from B to C and $\boldsymbol{A}'_\Phi$ from B' to C'
- (intra) transformation matrices $\boldsymbol{S}$ from B' to B and $\boldsymbol{T}$ from C' to C
- Theorem. $\boldsymbol{A}'_\Phi = \boldsymbol{T}^{-1}\boldsymbol{A}_\Phi \boldsymbol{S}$

$$\begin{tikzcd}
V \arrow[r, "\Phi"] & W \\
B \arrow[r, "\Phi_{CB}", "\boldsymbol{A}_\Phi"'] & C \\
\tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\boldsymbol{S}"'] \arrow[r, "\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C} \arrow[u, "\boldsymbol{T}", "\Xi_{C\tilde{C}}"']
\end{tikzcd}
\qquad
\begin{tikzcd}
V \arrow[r, "\Phi"] & W \\
B \arrow[r, "\Phi_{CB}", "\boldsymbol{A}_\Phi"'] & C \arrow[d, "\boldsymbol{T}^{-1}"', "\Xi_{\tilde{C}C} = \Xi^{-1}_{C\tilde{C}}"] \\
\tilde{B} \arrow[u, "\Psi_{B\tilde{B}}", "\boldsymbol{S}"'] \arrow[r, "\tilde{\boldsymbol{A}}_\Phi", "\Phi_{\tilde{C}\tilde{B}}"'] & \tilde{C}
\end{tikzcd}$$

Basis Change: General Case (2)

- **Definition 2.21** (Equivalence). Two matrices $\boldsymbol{A}, \boldsymbol{A}' \in \mathbb{R}^{m \times n}$ are *equivalent* if there exist regular matrices $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{T} \in \mathbb{R} m \times m$, such that $\boldsymbol{A}' = \boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{S}$.
- **Definition 2.22** (Similarity). Two matrices $\boldsymbol{A}, \boldsymbol{A}' \in \mathbb{R}^{n \times n}$ are *similar* if there exist a regular matrix $\boldsymbol{S} \in \mathbb{R}^{n \times n}$ with $\boldsymbol{A}' = \boldsymbol{S}^{-1}\boldsymbol{A}\boldsymbol{S}$.
- Remark. Similar matrices are always equivalent. However, equivalent matrices are not necessarily similar.

Example 2.24 (Basis Change)

Image and Kernel

- Consider a linear mapping $\Phi : V \mapsto W$. The kernel (or null space) is the set of vectors in V that maps to $\mathbf{0} \in W$ (i.e., neutral element).

  Definition. $\ker(\Phi) := \Phi^{-1}(\mathbf{0}_W) = \{\boldsymbol{v} \in V : \Phi(\boldsymbol{v}) = \mathbf{0}_W\}$

- Image/range: set of vectors $w \in W$ that can be reached by Φ from any vector in V
- V: domain, W: codomain

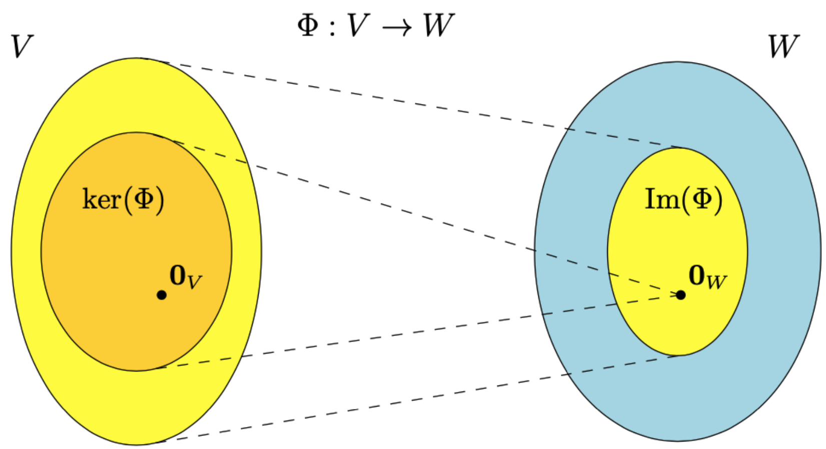


Image and Kernel: Properties

- $\mathbf{0}_V \in \ker(\Phi)$ (because $\Phi(\mathbf{0}_V) = \mathbf{0}_W$)
- Both $\mathsf{Im}(\Phi)$ and $\ker(\Phi)$ are subspaces of W and V, respectively.
- Φ is one-to-one (injective) $\iff \ker(\Phi) = \{0\}$ (i.e., only $\mathbf{0}$ is mapped to $\mathbf{0}$)
- Since Φ is a linear mapping, there exists $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ such that $\Phi : \boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$. Then, $\mathsf{Im}(\Phi) =$ column space of $\boldsymbol{A}$ which is the span of column vectors of $\boldsymbol{A}$.
- $\mathsf{rk}(\boldsymbol{A}) = \dim(\mathsf{Im}(\Phi))$
- $\ker(\Phi)$ is the solution set of the homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \mathbf{0}$

Example 2.25 (Image and Kernel of a Linear Mapping)

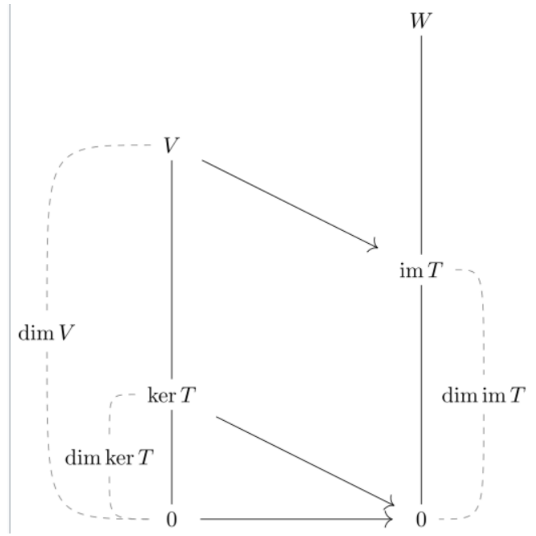
Rank-Nullity Theorem

Theorem.

$$\dim(\ker(\Phi)) + \dim(\mathrm{Im}(\Phi)) = \dim(V)$$

- If $\dim(\mathrm{Im}(\Phi)) < \dim(V)$, the kernel contains more than just $\mathbf{0}$.
- If $\dim(\mathrm{Im}(\Phi)) < \dim(V)$, $\boldsymbol{A}_\Phi \boldsymbol{x} = \mathbf{0}$ has infinitely many solutions.
- If $\dim(V) = \dim(W)$ (e.g., $V = W = \mathbb{R}^n$), the followings are equivalent: Φ is
  - (1) injective, (2) surjective, (3) bijective,
  - In this case, Φ defines $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, where $\boldsymbol{A}$ is regular.
- Simplified version. For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$,

$$\mathrm{rk}(\boldsymbol{A}) + \mathrm{nullity}(\boldsymbol{A}) = n$$



[2] Nullity: the dimension of null space (kernel)

Roadmap

(1) Systems of Linear Equations

(2) Matrices

(3) Solving Systems of Linear Equations

(4) Vector Spaces

(5) Linear Independence

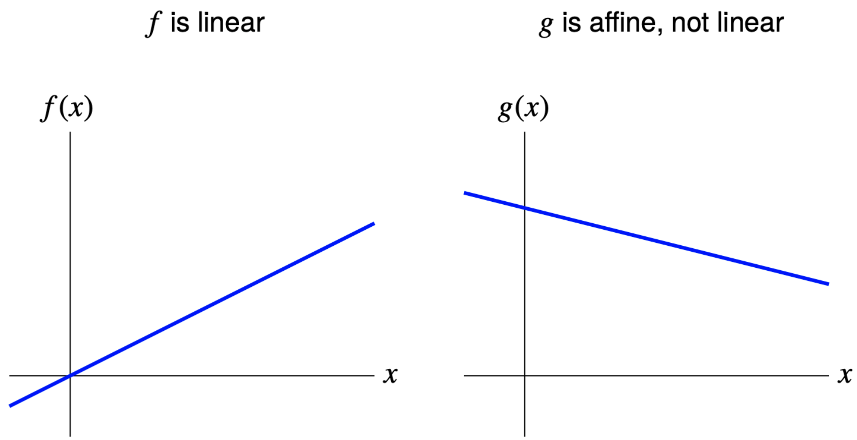
(6) Basis and Rank

(7) Linear Mappings

(8) Affine Spaces

Linear vs. Affine Function

- linear function: $f(x) = ax$
- affine function: $f(x) = ax + b$
- sometimes people refer to affine functions as linear



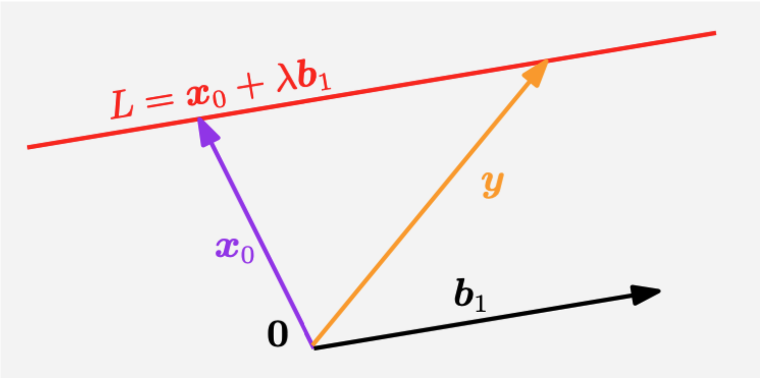
Affine Subspace

- Spaces that are offset from the origin. Not a vector space.
- Definition. Consider a vector space V, $\boldsymbol{x}_0 \in V$ and a subspace $U \subset V$. Then, the subset $L = \boldsymbol{x_0} + U := \{\boldsymbol{x_0} + \boldsymbol{u} : \boldsymbol{u} \in U\}$ is called affine subspace or linear manifold of V.
- U is called direction or direction space, and $\boldsymbol{x}_0$ is support point.
- An affine subspace is not a vector subspace of V for $\boldsymbol{x}_0 \notin U$.
- ***Parametric equation.*** A k-dimensional affine space $L = \boldsymbol{x_0} + U$. If $(\boldsymbol{b}_1, \dots, \boldsymbol{b}_k)$ is an ordered basis of U, any element $\boldsymbol{x} \in L$ can be uniquely described as

  $$\boldsymbol{x} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \cdots + \lambda_k \boldsymbol{b}_k, \quad \lambda_1, \dots, \lambda_k \in \mathbb{R}$$

Example

- In $\mathbb{R}^2$, one-dimensional affine subspace: line. $\boldsymbol{y} = \boldsymbol{x}_0 + \lambda \boldsymbol{b}_1$. $U = \text{span}[\boldsymbol{b}_1]$
- In $\mathbb{R}^3$, two-dimensional affine subspace: plane. $\boldsymbol{y} = \boldsymbol{x}_0 + \lambda_1 \boldsymbol{b}_1 + \lambda_2 \boldsymbol{b}_2$. $U = \text{span}[\boldsymbol{b}_1, \boldsymbol{b}_2]$
- In $\mathbb{R}^n$, $(n-1)$-dimensional affine subspace: hyperplane. $\boldsymbol{y} = \boldsymbol{x}_0 + \sum_{k=1}^{n-1} \lambda_i \boldsymbol{b}_i$. $U = \text{span}[\boldsymbol{b}_1, \ldots, \boldsymbol{b}_n]$



- For a linear mapping $\Phi : V \mapsto W$ and a vector $\boldsymbol{a} \in W$, the mapping $\phi : V \mapsto W$ with $\phi(\boldsymbol{x}) = \boldsymbol{a} + \Phi(\boldsymbol{x})$ is an affine mapping from V to W. The vector $\boldsymbol{a}$ is called the translation vector.

Questions?

References

[1] This lecture slide is mainly based upon https://yung-web.github.io/home/courses/mathml.html (made by Prof. Yung Yi, KAIST EE)

[2] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. Mathematics for machine learning. Cambridge University Press, 2020.