Linear Algebra

Yuji Takai

2020-11-21

1 Introduction

Definition 1.1 (**Vector**)**.** Object that can be added together and multiplied by scalars to produce another object of the same kind

- Examples: geometric vectors, polynomials, audio signals, elements of $\mathbb{R}^n$

2 System of linear equations

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

- Every n-tuple $(x_1, \cdots, x_n) \in \mathbb{R}^n$ that satisfies the system of linear equations is a solution of the linear equation system
- Either *no, exactly one, or infinitely many* solutions can be obtained for a real-valued system of linear equations
- System of lilnear equation into matrix multiplication form

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

3 Matrices

3.1 Definitions

Definition 3.1 (**Matrix**). $m \times n$ tuple of elements a_{ij} ($m, n \in \mathbb{R}$, $i = 1, \cdots, m$ & $j = 1, \cdots, n$), which is ordered according to a rectangular scheme consisting of m rows and n columns

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}$$

- Represents systems of linear equations and linear functions (linear mappings)

Definition 3.2 (**Identity Matrix**). Matrix $\boldsymbol{I}_n \in \mathbb{R}^{n \times n}$ that contains 1 on the diagonal and 0 everywhere else

$$\boldsymbol{I}_n := \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \in \mathbb{R}^{n \times n}$$

Definition 3.3 (**Inverse**). For $\boldsymbol{A} \in \mathbb{R}^{n \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times n}$ that has the property $\boldsymbol{AB} = \boldsymbol{I}_n = \boldsymbol{BA}$ is the **inverse** of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$

- If $\boldsymbol{A}$ is **regular**/**invertible**/ **nonsingular**, $\boldsymbol{A}^{-1}$ exists
- Otherwise $\boldsymbol{A}$ is **singular**/**noninvertible**
- When $\boldsymbol{A}^{-1}$ exists, it is unique
- Can use the **determinant** to check whether a matrix is invertible
- If $\boldsymbol{A}$ is invertible, then so is $\boldsymbol{A}^T$, and $\boldsymbol{A}^{-T} := (\boldsymbol{A}^{-1})^{-T} = (\boldsymbol{A}^T)^{-1}$
- To compute the inverse, apply the **Gaussian elimination** on the augmented matrix $[\boldsymbol{A}|\boldsymbol{I}_n]$ to obtain the reduced-row echelon form $[\boldsymbol{I}_n|\boldsymbol{A}^{-1}]$

Definition 3.4 (Transpose). For $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times m}$ with $b_{ij} = a_{ji}$ is the **transpose** of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^T$

Definition 3.5 (Symmetric Matrix). $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ such that $\boldsymbol{A} = \boldsymbol{A}^T$

- Sum of symmetric matrices is always symmetric
- Product is always defined, but generally not symmetric

3.2 Operations on Matrix

Definition 3.6 (Matrix Addition). Element-wise sum for $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n}$

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

Definition 3.7 (Matrix Multiplication). Matrix multiplication of $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{B} \in \mathbb{R}^{n \times k}$ results in $\boldsymbol{C} \in \mathbb{R}^{\times k}$ with each element c_{ij} being **dot product** of the ith row of $\boldsymbol{A}$ and jth column of $\boldsymbol{B}$

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

- Matrices can only be multiplied if their "neighboring" dimensions match.
- Ex: an $n \times k$-matrix $\boldsymbol{A}$ can be multiplied with a $k \times m$-matrix $\boldsymbol{B}$, but only from the left side i.e. $\boldsymbol{AB} = \boldsymbol{C}$, and $\boldsymbol{BA}$ is undefined
- NOT defined as an element-wise operation on matrix elements i.e. $c_{ij} \neq a_{ij} b_{ij}$

Definition 3.8 (Hadamard product). Element-wise multiplication of equally-sized multi-dimensional arrays

Definition 3.9 (Scalar multiplication). If $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and $\lambda \in \mathbb{R}$, then $\lambda \boldsymbol{A} = \boldsymbol{K}$ where $K_{ij} = \lambda a_{ij}$

3.3 Properties of Matrix

- **Associativity**: $\forall \boldsymbol{A} \in \mathbb{R}^{m \times n},\ \boldsymbol{B} \in \mathbb{R}^{n \times p},\ \boldsymbol{C} \in \mathbb{R}^{p \times q}$: $(\boldsymbol{AB})\boldsymbol{C} = \boldsymbol{A}(\boldsymbol{BC})$
- **Distributivity**: $\forall \boldsymbol{A}, \boldsymbol{B} \in \mathbb{R}^{m \times n},\ \boldsymbol{C}, \boldsymbol{D} \in \mathbb{R}^{n \times p}$:
  - $(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$
  - $\boldsymbol{A}(\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{AC} + \boldsymbol{AD}$
- Multiplication with $\boldsymbol{I}$: $\forall \boldsymbol{A} \in \mathbb{R}^{m \times n}$: $\boldsymbol{I}_m \boldsymbol{A} = \boldsymbol{A}\boldsymbol{I}_n = \boldsymbol{A}$
- Inverse
  - $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I} = \boldsymbol{A}^{-1}\boldsymbol{A}$
  - $(\boldsymbol{AB})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$
  - $(\boldsymbol{A} + \boldsymbol{B})^{-1} \neq \boldsymbol{A}^{-1} + \boldsymbol{B}^{-1}$
- Transpose
  - $\left(\boldsymbol{A}^T\right)^T = \boldsymbol{A}$
  - $(\boldsymbol{A} + \boldsymbol{B})^T = \boldsymbol{A}^T + \boldsymbol{B}^T$
  - $(\boldsymbol{AB})^T = \boldsymbol{B}^T\boldsymbol{A}^T$
- Multiplication with scalar
  - **Associativity**
    - $(\lambda\psi)\boldsymbol{C} = \lambda(\psi\boldsymbol{C}), \quad \boldsymbol{C} \in \mathbb{R}^{m \times n}$
    - $\lambda(\boldsymbol{BC}) = (\lambda\boldsymbol{B})\boldsymbol{C} = \boldsymbol{B}(\lambda\boldsymbol{C}) = (\boldsymbol{BC})\lambda, \quad \boldsymbol{B} \in \mathbb{R}^{m \times n}, \boldsymbol{C} \in \mathbb{R}^{n \times k}$
    - $(\lambda\boldsymbol{C})^T = \boldsymbol{C}^T\lambda^T = \boldsymbol{C}^T\lambda = \lambda\boldsymbol{C}^T$ (since $\forall \lambda \in \mathbb{R} : \lambda = \lambda^T$)
  - **Distributivity**
    - $(\lambda + \psi)\boldsymbol{C} = \lambda\boldsymbol{C} + \psi\boldsymbol{C}, \quad \boldsymbol{C} \in \mathbb{R}^{m \times n}$
    - $\lambda(\boldsymbol{B} + \boldsymbol{C}) = \lambda\boldsymbol{B} + \lambda\boldsymbol{C}, \quad \boldsymbol{B}, \boldsymbol{C} \in \mathbb{R}^{m \times n}$

3.4 Row reduction

Definition 3.10 (**Elementary transformation**)**.** Transformation of the system of equation into a simpler form while keeping the solution set the same

- Exchange of 2 equations (rows in the matrix representing the system of equations)
- Multiplcation of an equation (row) with a constant $\lambda \in \mathbb{R}\backslash\{0\}$
- Addition of 2 equations (rows)

Definition 3.11 (**Pivot**)**.** Leading coefficient of a row and is always strictly to the right of the pivot of the row above it

- any equation system in row-echelon form always has a "staircase" structure

Definition 3.12 (**Row-echelon form**)**.** A matrix is in row-echelon form if:

- All rows that contain only zeros are at the bottom of the matrix i.e. 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 (**pivot**/ leading coefficient) is always strictly to the right of the pivot of the row above it

Definition 3.13 (**Basic variable**)**.** Variables corresponding to the pivots in the **row-echelon** form

Definition 3.14 (**Free variable**)**.** Variables corresponding to the non-pivots in the **row-echelon** form

Definition 3.15 (**Reduced row-echelon form**)**.** A matrix 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

Definition 3.16 (**Gaussian elimination**)**.** An algorithm that performs elementary transformations to bring a system of linear equations into reduced row-echelon form. Useful for:

- Computing determinants
- Checking whether a set of vectors is **linearly independent**
- Computing the inverse of a matrix
- Computing the rank of a matrix
- Determining a basis of a vector space

However, impractical for systems with millions of variables ($O(n^3)$ algorithm)

3.5 Algorithms for solving a system of linear equations

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

- If there is *no solution*, need to resort to approximate solutions like **linear regression**
- If $\boldsymbol{A}$ is a **square matrix** and **invertible**, can determine $\boldsymbol{A}^{-1}$ and solve the equation via $\boldsymbol{x} = \boldsymbol{A}^{-1}\boldsymbol{b}$
- If $\boldsymbol{A}$ has linearly independent columns:
  $$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \Leftrightarrow \boldsymbol{A}^T\boldsymbol{A}\boldsymbol{x} = \boldsymbol{A}^T\boldsymbol{b} \Leftrightarrow \boldsymbol{x} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\boldsymbol{b}$$
  - **Moore-Penrose pseudo-inverse** $(\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T$
  - Solution corresponds to the minimum norm least-squares solution
  - Requires many computations for the matrix-matrix product and computating inverse of $\boldsymbol{A}^T\boldsymbol{A}$
  - For reasons of numerical precision, not recommended to compute the inverse or pseudo-inverse
- Practical methods that indirectly solve systems of many linear equations
  - Stationary iterative methods: Richardson method, Jacobi method, Gauß-Seidel method, successive over-relaxation method
  - Krylov subspace methods: conjugate gradients, generalized minimal residual, biconjugate gradients
  - Key idea of iterative methods: set up an iteration of the form
    $$\boldsymbol{x}^{(k+1)} = \boldsymbol{C}\boldsymbol{x}^{(k)} + \boldsymbol{d}$$
    for suitable $\boldsymbol{C}$ and $\boldsymbol{d}$ that reduces the residual error $\|\boldsymbol{x}^{(k+1)} - \boldsymbol{x}_*\|$ in every iteration and converges to $\boldsymbol{x}_*$

4 Vector spaces

Definition 4.1 (**Group**)**.** For a set $\mathcal{G}$ and an operation $\otimes : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$, $G := (\mathcal{G}, \otimes)$ is called a **group** if the following hold:

1. **Closure** of $\mathcal{G}$ under $\otimes$: $\forall x, y \in \mathcal{G} : x \otimes y \in \mathcal{G}$
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 $y \otimes x = e$ (x^{-1}: inverse element of x)

Definition 4.2 (**Abelian group**)**.** Group $G = (\mathcal{G}, \otimes)$ with **commutative** property: $\forall x, y \in \mathcal{G}$: $x \otimes y = y \otimes x$

Definition 4.3 (**General Linear Group**)**.** The group with the set of regular/invertible matrices $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and matrix multiplication as operation, denoted by $GL(n, \mathbb{R})$

- NOT Abelian $\leftarrow$ since matrix multiplication is NOT commutative

Definition 4.4 (**Vector space**)**.** $V = (\mathcal{V}, +, \cdot)$ that is a set $\mathcal{V}$ with 2 operations

- $+$: $\mathcal{V} \times \mathcal{V} \to \mathcal{V}$ (**vector addition/inner operation**)
- $\cdot$: $\mathbb{R} \times \mathcal{V} \to \mathcal{V}$ (**multiplication by scalars/outer operation**)

where:

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

Definitions/Properties:

1. **Vectors**: the elements $\boldsymbol{x} \in \boldsymbol{V}$
2. **Scalars**: the elements $\lambda \in \mathbb{R}$
3. Neutral element of $(\mathcal{V}, +)$ = zero vector $\boldsymbol{0}$

Definition 4.5 (**Vector subspace**)**.** Let $V = (\mathcal{V}, +, \cdot)$ and $\mathcal{U} \subseteq \mathcal{V}$, $\mathcal{U} \neq \emptyset$. Then $U = (\mathcal{U}, +, \cdot)$ is **vector subspace**/**linear subspace** of V if U is a vector space with the vector space operations $+$ and $\cdot$ restricted to $\mathcal{U} \times \mathcal{U}$ and $\mathbb{R} \times \mathcal{U}$, denoted as $U \subseteq V$. To determine $U = (\mathcal{U}, +, \cdot)$ is a subspace of V:

1. $\mathcal{U} \neq \emptyset$, in particular: $\boldsymbol{0} \in \mathcal{U}$
2. Closure of U:
   - w.r.t the outer operation: $\forall \lambda \in \mathbb{R}, \forall \boldsymbol{x} \in \mathcal{U}$: $\lambda \boldsymbol{x} \in \mathcal{U}$
   - w.r.t the inner operation: $\forall \boldsymbol{x}, \boldsymbol{y} \in \mathcal{U}$: $\boldsymbol{x} + \boldsymbol{y} \in \mathcal{U}$

NOTE: every subspace $U \subseteq (\mathbb{R}^n, +, \cdot)$ is the solution space of a homogeneous system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$ for $\boldsymbol{x} \in \mathbb{R}^n$

5 Linear independence

Definition 5.1 (**Linear combination**)**.** For a vector space V and a finite number of vectors $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_n \in \boldsymbol{V}$, every $\boldsymbol{v} \in \boldsymbol{V}$ of the form

$$\boldsymbol{v} = \lambda_1 \boldsymbol{x}_1 + \cdots + \lambda_k \boldsymbol{x}_k = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i \in \boldsymbol{V}$$

with $\lambda_1, \cdots, \lambda_k \in \mathbb{R}$ is a **linear combination** of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$

Definition 5.2 (**Linear (In)dependence**)**.** For a vector space $\boldsymbol{V}$ with $k \in \mathbb{N}$ and $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \boldsymbol{V}$:

- If there is a non-trivial linear combination, such that $\mathbf{0} = \sum_{i=1}^{k} \lambda_i \boldsymbol{x}_i$, the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are **linearly dependent**
- If only the trivial solution exists (i.e. $\lambda_1 = \cdots = \lambda_k = 0$), the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are **linearly independent**

Properties of linear independence:

- k vectors are either linearly dependent or linearly independent = no third option
- If at least one of the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ is $\mathbf{0}$ or two vectors are identical, then they are linearly dependent
- The vectors $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k : x_i \neq \mathbf{0}, i = 1, \ldots, k\}$, $k \geq 2$ are linearly dependent iff (at least) one of them is a linear combination of the others
- Practical way of checking whether vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ are linearly independent is to use Gaussian elimination: write all vectors as columns of a matrix $\boldsymbol{A}$ and perform Gaussian elimination until the matrix is in row echelon form:
  - The **pivot** columns indicate the vectors, which are linearly independent of the vectors on the left
  - The **non-pivot** columns can be expressed as linear combinations of the pivot columns on their left
  - All column vectors are **linearly independent** iff *all* columns are pivot columns $\Leftrightarrow$ if there is at least one non-pivot column, the columns are linearly dependent

Vector space, basis and linear independence

- Vector space V with k linearly independent vectors $\boldsymbol{b}_1, ..., \boldsymbol{b}_k$ and m linear combinations

$$\boldsymbol{x}_1 = \sum_{i=1}^{k} \lambda_{i1} \boldsymbol{b}_i,$$
$$\vdots$$
$$\boldsymbol{x}_m = \sum_{i=1}^{k} \lambda_{im} \boldsymbol{b}_i$$

  - Defining $\boldsymbol{B} = [\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k]$:

$$\boldsymbol{x}_j = \boldsymbol{B}\boldsymbol{\lambda}_j, \quad \boldsymbol{\lambda}_j = \begin{bmatrix} \lambda_{1j} \\ \vdots \\ \lambda_{kj} \end{bmatrix}, \quad j = 1, \ldots, m$$

  -

-

6

7

8