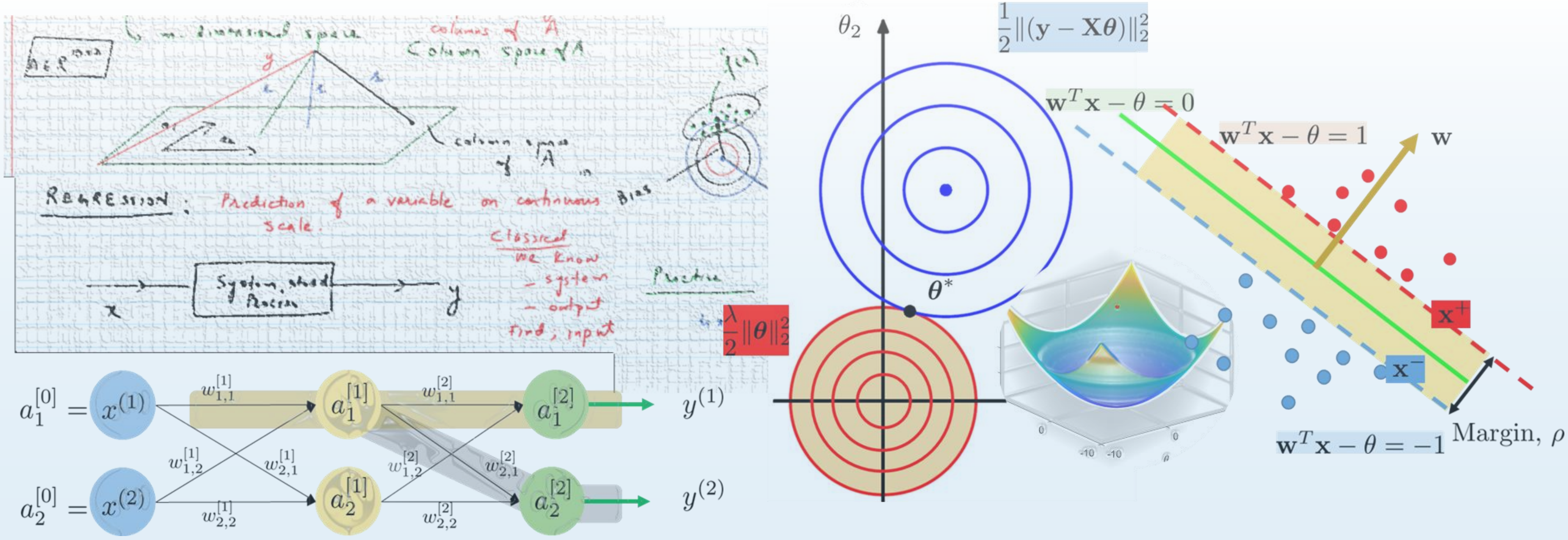


AI-501 Mathematics for AI

Matrices – Operations, Inverses and Linear Equations

Zubair Khalid

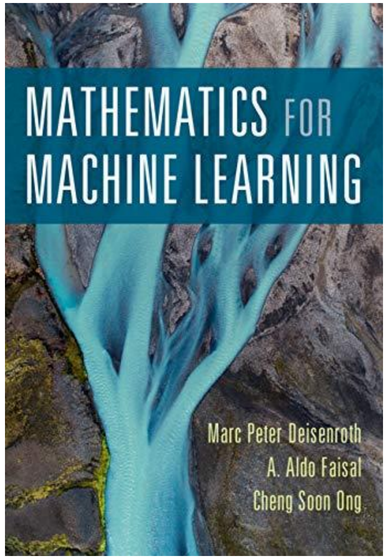
School of Science and Engineering

https://www.zubairkhalid.org/ai501_2024.html

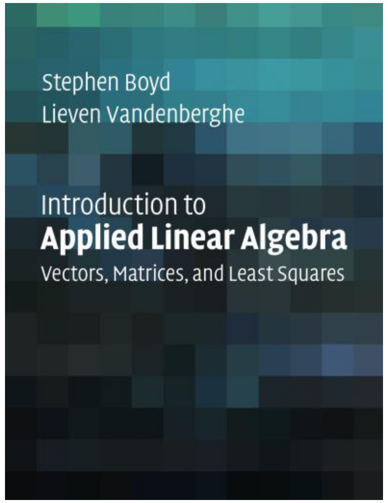
LUMS
A Not-for-Profit University

Outline

- *Matrices Notation*
- *Application Examples*
- *Operations on Matrices*
  - *Addition*
  - *Scaling*
  - *Transpose*
  - *Norm*



Chapter 2



Chapter 6

LUMS
A Not-for-Profit University

Matrices

Definition:

A matrix is a two dimensional (2D) vector or array of numbers.

Notation:

Usually denoted by a capital letter symbol; stack the list of numbers in 2D array.

For example, consider a matrix **A** of 6 real numbers represented as stack of 3 2-vectors using square or round parentheses:

$$A = \begin{bmatrix} -1.1 & 17.3 & 2.7 \\ 30.1 & 19.1 & 8.4 \end{bmatrix} \qquad A = \begin{pmatrix} -1.1 & 17.3 & 2.7 \\ 30.1 & 19.1 & 8.4 \end{pmatrix} \qquad 2 \times 3 \text{ matrix}$$

Size of a matrix: Number of rows (m) times number of columns (n); $m \times n$

We express matrix B of size $m \times n$ as $B \in \mathbf{R}^{m \times n}$ and call it $m \times n$-matrix.

Entry of a matrix: B_{ij} - entry in the matrix at i-th row and j-th column.

For example, $A_{21} = 30.1$.

Matrices

Square Matrix: $m = n$ **Tall Matrix:** $m > n$ **Wide Matrix:** $m < n$

Zero Matrix: A matrix with all elements equal to zero.

denoted by $\mathbf{0} \in \mathbf{R}^{m \times n}$.

Identity Matrix: A square matrix with diagonal elements equal to one and off diagonal elements equal to zero.

denoted by $I \equiv I_n \in \mathbf{R}^{n \times n}$ and is defined as

$$(I)_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases} = \delta_{i,j}$$

Diagonal Matrix:

Block Matrix:

$$A = \left[\begin{array}{cc} B & C \\ D & E \end{array} \right]$$

Triangular Matrix:

$$\begin{bmatrix} 1 & -1 & 0.7 \\ 0 & 1.2 & -1.1 \\ 0 & 0 & 3.2 \end{bmatrix} \qquad \begin{bmatrix} -0.6 & 0 \\ -0.3 & 3.5 \end{bmatrix}$$

Examples of Matrices - Applications

Image RGB

Each color represents a matrix.

Quantities

An mxn-matrix **A** can represent the amounts or quantities of n different resources or products held (or produced, or required) by an entity such as a company at m different locations or for m different customers.

For example, mxn-matrix represents the quantity of n products stocked in m number of warehouses.

Examples of Matrices - Applications

Time series grouped over time

- 12x20-matrix can represent the average monthly temperature, rainfall, pressure etc of 20 cities of Pakistan.
- 30x7-matrix can represent number of patients suffering from a disease over the next 30 days for 7 states/territories.
- Other examples include exchange rate, audio, and, in fact, any quantity that varies over time.

Operations on Matrices

Additivity and Scaling

Additivity of Matrices: The sum of two matrices A and B of the same dimensions is obtained by adding their corresponding entries:

$$(A+B)_{ij} = A_{ij} + B_{ij}.$$

Scaling of a Matrix by a Scalar: A matrix A can be scaled by a scalar α by multiplying each entry of the matrix by the scalar:

$$(\alpha A)_{ij} = \alpha A_{ij}.$$

Operations on Matrices

Additivity and Scaling

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix} + \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}$$

$$(-2)\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 6 & 0 \end{bmatrix} = \begin{bmatrix} -2 & -12 \\ -18 & -6 \\ -12 & 0 \end{bmatrix}$$

Operations on Matrices

Transpose and Concept of Symmetric Matrices

Transpose of a Matrix: The transpose of a matrix A, denoted as $A^\top$, is obtained by swapping its rows and columns. If A is an $m \times n$ matrix, then the transpose $A^\top$ is an $n \times m$ matrix, and its entries are given by:

$$(A^\top)_{ij} = A_{ji}.$$

Symmetric Matrices: A square matrix $A \in \mathbb{R}^{n \times n}$ is called symmetric if it is equal to its transpose:

$$A = A^\top.$$

This implies that the entries of A are symmetric across the main diagonal, i.e., $A_{ij} = A_{ji}$ for all i, j.

Skew-Symmetric Matrices: A square matrix $A \in \mathbb{R}^{n \times n}$ is called skew-symmetric if it is equal to the negative of its transpose:

$$A = -A^\top.$$

This implies that the diagonal entries of A are zero ($A_{ii} = 0$ for all i), and the off-diagonal entries satisfy $A_{ij} = -A_{ji}$ for all i, j.

Operations on Matrices

Transpose and Concept of Symmetric Matrices

$$\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix}^T = \begin{bmatrix} 0 & 7 & 3 \\ 4 & 0 & 1 \end{bmatrix}$$

Operations on Matrices

Transpose and Concept of Symmetric Matrices

- Any square matrix can be expressed as a sum of symmetric matrix and a skew symmeytric matrix.

This decomposition is given by:

$$A = S + K,$$

where:

$$S = \frac{A + A^\top}{2} \quad \text{(symmetric part)},$$

$$K = \frac{A - A^\top}{2} \quad \text{(skew-symmetric part)}.$$

Here, S is symmetric since $S^\top = S$, and K is skew-symmetric since $K^\top = -K$. This decomposition is useful for analyzing the properties of A, as it separates the matrix into its symmetric and skew-symmetric components.

Operations on Matrices

Matrix Norm

The Frobenius norm of a matrix $A \in \mathbb{R}^{m\times n}$, denoted as $\|A\|_F$, is a measure of the magnitude of the entries of the matrix, defined as

$$\|A\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|A_{ij}|^2}.$$

Alternatively, the Frobenius norm can be viewed as the 2-norm of the vector formed by stacking all the entries of A.

Use Cases

Measuring the Error of Matrix Approximations: To quantify the error between an original matrix A and its approximation $\hat{A}$. The error is measured as:

$$\|A - \hat{A}\|_F.$$

Regularization in Optimization Problems: In machine learning and optimization, the Frobenius norm can be used as a regularization term to prevent overfitting. Given a parameter matrix W, a regularization term such as $\|W\|_F^2$ is added to the objective function to penalize large weights.

Operations on Matrices

Trace of a Matrix

The trace of a square matrix $A \in \mathbb{R}^{n\times n}$, denoted by $\mathrm{tr}(A)$, is the sum of its diagonal elements:

$$\mathrm{tr}(A) = \sum_{i=1}^{n} A_{ii}.$$

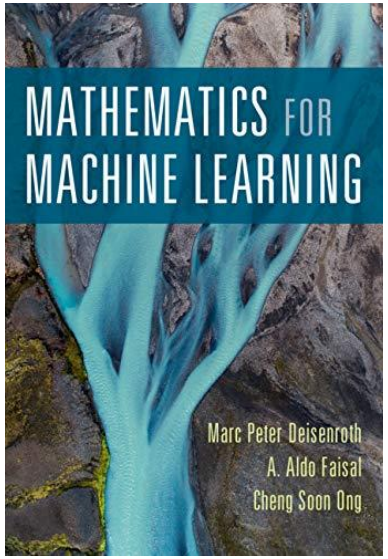
The trace is a linear operator and has several useful properties, such as being invariant under cyclic permutations of matrix products, i.e., $\mathrm{tr}(ABC) = \mathrm{tr}(CAB)$.

Use Cases in AI and Machine Learning

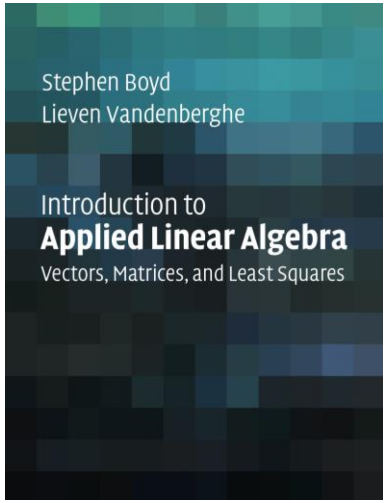
Dimensionality Reduction (PCA): In Principal Component Analysis (PCA), the trace is used to measure the total variance retained by a set of principal components. If Λ is a diagonal matrix containing the eigenvalues of the covariance matrix, the trace $\mathrm{tr}(\Lambda)$ represents the total variance, helping to select the components that capture the most variance. We will cover this in more detail later in the course.

Outline

- *Matrix-vector product*
- *Interpretations*
- *Application Examples*
- *Matrix-matrix product*



Chapter 2



Chapter 6 and 10

Matrix-Vector Multiplication

$A \in \mathbf{R}^{m\times n} \qquad x \in \mathbf{R}^{n\times 1}\ (\mathbf{R}^n)$

number of columns of A equals the size of x

$y = Ax \qquad y \in \mathbf{R}^{m\times 1}\ (\mathbf{R}^m)$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$y_i = \sum_{k=1}^{n} A_{ik}x_k = A_{i1}x_1 + \cdots + A_{in}x_n, \quad i = 1,\dots,m$$

Example

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

$$= \begin{bmatrix} (0)(2) + (2)(1) + (-1)(-1) \\ (-2)(2) + (1)(1) + (1)(-1) \end{bmatrix}$$

$$= \begin{bmatrix} 3 \\ -4 \end{bmatrix}$$

Matrix-Vector Multiplication

$$\left.\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \right\} \quad y = Ax$$

$$y_i = \sum_{k=1}^{n} A_{ik} x_k = A_{i1} x_1 + \cdots + A_{in} x_n, \quad i = 1, \dots, m$$

Interpretation In terms of Rows of Matrix

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrix-Vector Multiplication

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \Bigg\} \quad y = Ax$$

$$y_i = \sum_{k=1}^{n} A_{ik} x_k = A_{i1}x_1 + \cdots + A_{in}x_n, \quad i = 1, \ldots, m$$

Interpretation In terms of Columns of Matrix

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \ldots, & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad y = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

- This shows that $y = Ax$ is a linear combination of the columns of A; the coefficients in the linear combination are the elements of x.

Matrix-Vector Multiplication

Application Examples

- For example, 200x70-matrix represents the quantity of 70 products stocked in 200 warehouses.

Calculating Total Value of Stock in Each Warehouse
If the price of each product is represented by a vector $\mathbf{p} \in \mathbb{R}^{70}$, then the product

$$\mathbf{y} = A\mathbf{p}$$

will be a vector of length 200, where each element of $\mathbf{y}$ represents the total value of stock in the corresponding warehouse.

Estimating Total Weight of Stock in Each Warehouse
If the weight of each product is given by a vector $\mathbf{w} \in \mathbb{R}^{70}$, then

$$\mathbf{y} = A\mathbf{w}$$

will represent the total weight of stock in each warehouse, useful for logistics and transportation planning.

Matrix-Vector Multiplication

Application Examples

- For example, 200x70-matrix represents the quantity of 70 products stocked in 200 warehouses.

Calculating Stock Value per Warehouse with Seasonal or Discounted Pricing
If the prices of the products change due to seasonality or discounts, represented as a vector $\mathbf{p}_{\text{season}} \in \mathbb{R}^{70}$, then

$$\mathbf{y} = A\mathbf{p}_{\text{season}}$$

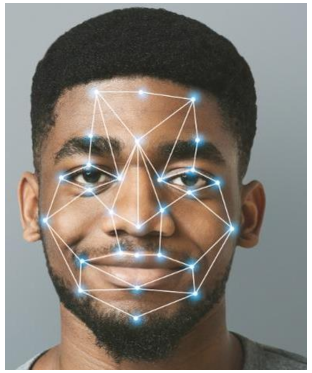
will provide the updated total value of stock in each warehouse.

Determining Warehouse Demand for Product Replenishment
If each product has a "replenishment coefficient" $\mathbf{r} \in \mathbb{R}^{70}$, indicating stock to be reordered based on warehouse quantities, then

$$\mathbf{y} = A\mathbf{r}$$

will yield the replenishment demand per warehouse.

Matrix-Vector Multiplication



Application Examples

Feature matrix and weight vector

Let $A \in \mathbb{R}^{d \times n}$ represent a matrix where each column corresponds to one of n images, and each column is a d-dimensional feature vector describing various characteristics of the images.

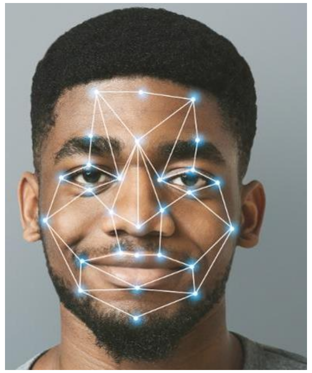
Transforming Image Features Using a Weight Vector for Classification

If there is a weight vector $\mathbf{w} \in \mathbb{R}^d$ that transforms or combines the features of each image for a classification task, then

$$\mathbf{y} = A^\top \mathbf{w}$$

will produce a vector $\mathbf{y} \in \mathbb{R}^n$, where each element corresponds to a transformed score for each image, which can be used for predicting class labels or performing further analysis.

Matrix-Vector Multiplication



Application Examples

Feature matrix and weight vector

Let $A \in \mathbb{R}^{d \times n}$ represent a matrix where each column corresponds to one of n images, and each column is a d-dimensional feature vector describing various characteristics of the images.

Calculating the Similarity of All Images to a Reference Image
If a reference image has a feature vector $\mathbf{q} \in \mathbb{R}^d$, then the product

$$\mathbf{y} = A^\top \mathbf{q}$$

will yield an n-dimensional vector $\mathbf{y}$, where each entry represents the similarity (e.g., dot product or cosine similarity) of the reference image to each of the n images. This can be useful in image retrieval or recommendation systems.

Matrix-Vector Multiplication

Application Examples

Expansion in a basis

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

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} a_1, & a_2, & \ldots, & a_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Matrix-Vector Multiplication

Application Examples

Linear dependence of columns

The columns of a general matrix $A \in \mathbb{R}^{m\times n}$ are linearly dependent if there exists a set of scalars $c_1, c_2, \dots, c_n$, not all zero, such that:

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0},$$

where $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n$ are the columns of A, and $\mathbf{0}$ is the zero vector in $\mathbb{R}^m$.

Equivalently, in matrix form:

$$A\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{0},$$

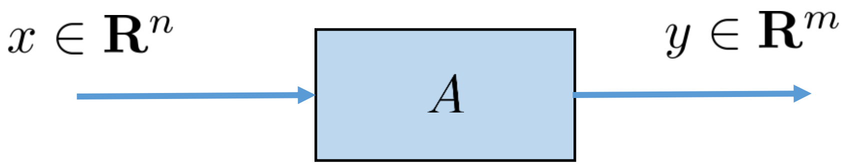
where $\begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix}^\top$ is a non-zero vector in $\mathbb{R}^n$.

Matrix-Vector Multiplication

Linear Transformation Interpretation:

Input-Output System Interpretation

$$y = Ax$$



Matrix-Vector Multiplication

Input-Output System Interpretation

$$x \in \mathbf{R}^n \longrightarrow \boxed{A} \longrightarrow y \in \mathbf{R}^m$$

Examples

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$$

Matrix-Vector Multiplication

Input-Output System Interpretation

$$x \in \mathbf{R}^n \longrightarrow \boxed{A} \longrightarrow y \in \mathbf{R}^m$$

Examples

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$\begin{bmatrix} 1 \\ -3 \\ 5 \end{bmatrix}$$

Matrix-Vector Multiplication

Input-Output System Interpretation

$$x \in \mathbf{R}^n \longrightarrow \boxed{A} \longrightarrow y \in \mathbf{R}^m$$

Examples

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Matrix-Vector Multiplication

Input-Output System Interpretation

$$x \in \mathbf{R}^n \longrightarrow \boxed{A} \longrightarrow y \in \mathbf{R}^m$$

Examples

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \quad \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

Generalization: Permuation matrix

- Permutaion matrix entries $P_{i,j} \in \{0, 1\}$
- one non-zero entry equal to one per row
- one non-zero entry equal to one per column

Matrix-Matrix Multiplication

$$A \in \mathbf{R}^{m\times n} \qquad B \in \mathbf{R}^{n\times p} \qquad C = AB$$

$$\text{no. of columns in A} = \text{no. of rows in B} = n$$

$$C \in \mathbf{R}^{m\times p}$$

$$\begin{bmatrix} C_{11} & C_{12} & \dots & C_{1p} \\ C_{21} & C_{22} & \dots & C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{m1} & C_{m2} & \dots & C_{mp} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \dots & B_{1p} \\ B_{21} & B_{22} & \dots & B_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ B_{n1} & B_{n2} & \dots & B_{np} \end{bmatrix}$$

$$[b_1, b_2, \dots, b_p]$$

$$C = AB \iff C_{ij} = \sum_{k=1}^{p} A_{ik}B_{kj} = A_{i1}B_{1j} + \cdots + A_{ip}B_{pj}, \qquad i = 1, \dots, m, \quad j = 1, \dots, n.$$

Matrix-Matrix Multiplication

Properties:

not *commutative:* $AB \neq BA$ in general

associative: $(AB)C = A(BC)$ so we write ABC

associative with scalar-matrix multiplication: $(\gamma A)B = \gamma(AB) = \gamma AB$

$$(AB)^T = B^T A^T$$

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} W & Y \\ X & Z \end{bmatrix} = \begin{bmatrix} AW + BX & AY + BZ \\ CW + DX & CY + DZ \end{bmatrix}$$

- Dimensions must be compatible.

Matrix-Matrix Multiplication

$$\begin{bmatrix} -1.5 & 3 & 2 \\ 1 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 \\ 0 & -2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 3.5 & -4.5 \\ -1 & 1 \end{bmatrix}$$

Outer product of Vectors:

$$\begin{bmatrix} a_1b_1 & a_1b_2 & \cdots & a_1b_n \\ a_2b_1 & a_2b_2 & \cdots & a_2b_n \\ \vdots & \vdots & & \vdots \\ a_mb_1 & a_mb_2 & \cdots & a_mb_n \end{bmatrix}$$

Gram Matrix:

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

Outline

- Systems of Linear Equations
  - Formulation
- Inverses
  - Left-inverse
  - Right-inverse
  - Inverse
  - Pseudo-inverse
  - Connection with the linear equations

Systems of Linear Equations

Formulation:

$$\boxed{\begin{array}{rcl} A_{11}x_1 + A_{12}x_2 + \cdots + A_{1n}x_n & = & b_1 \\ A_{21}x_1 + A_{22}x_2 + \cdots + A_{2n}x_n & = & b_2 \\ & \vdots & \\ A_{m1}x_1 + A_{m2}x_2 + \cdots + A_{mn}x_n & = & b_m \end{array}} \quad \left.\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \right\}$$

$$Ax = b$$

$$A \in \mathbf{R}^{m\times n} \qquad x \in \mathbf{R}^n \quad b \in \mathbf{R}^m$$

- one solution
- multiple solutions
- no solution

$x_1, x_2, \dots, x_n$ - variables or unknowns

$b_1, b_2, \dots, b_m$ - knowns, measurements, equation righ-hand side

A_{ij} - coefficient of the i-th equation associated with the j-th variable

Systems of Linear Equations

$$Ax = b$$

- $m < n$ under-determined
- $m = n$ square
- $m > n$ over-determined

Example 01

$x_1 + x_2 = 1, \quad x_1 = -1,$

$x_1 - x_2 = 0$

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

- no solution

Example 02

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

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

- multiple solutions

$$x = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad x = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Inverses

Left-Inverse:

X is a left inverse of A if

$$XA = I$$

A is left-invertible if it has at least one left inverse

Example:

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

Left inverses

$$\frac{1}{9}\begin{bmatrix} -11 & -10 & 16 \\ 7 & 8 & -11 \end{bmatrix} \qquad \begin{bmatrix} 0 & -1/2 & 3 \\ 0 & 1/2 & -2 \end{bmatrix}$$

Inverses

Left-Inverse:

Left-invertibility and column independence:

If A has a left inverse X then the columns of A are linearly independent.

Assume $Ax = 0$

$$X(Ax) = 0 \qquad (XA)x = Ix = x = 0$$

Connect with independence-dimension inequality:

When A is wide; $A \in \mathbf{R}^{m \times n}$ $\quad m < n$

- Columns are linearly dependent

A is not left invertible.

$A \in \mathbf{R}^{m \times n}$ can be left invertible

$m = n$ or $m > n$

Square or Tall

Inverses

Left Inverse: Connection with the Systems of Linear Equations

$$Ax = b$$

- $m = n$ square
- $m > n$ over-determined

- If A has a left inverse X, then we multiply with X the above system

$$X(Ax) = Xb \qquad\qquad x = Xb$$

- If solution exists for the system $Ax = b$.

  $x = Xb$ is **the** only solution of $Ax = b$.

- If there is no solution for the system $Ax = b$.

  $x = Xb$ does not **not** satisfy $Ax = b$.

If A has the left inverse X,
- there is **at most** one solution
- if exists, solution is $x = Xb$

In summary, a left inverse can be used to determine whether or not a solution of an over-determined set of linear equations exists, and when it does, find the unique solution.

Inverses

Right-Inverse:

X is a right inverse of A if

$$AX = I$$

A is right-invertible if it has at least one right inverse

Example:

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

Right inverses

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

Connection with the left Inverse:

If X is a right inverse of A, then X^T is the left inverse of A^T.

$$I = I^T = (AX)^T = X^T A^T \quad \Rightarrow \quad X^T A^T = I$$

Inverses

Right-Inverse:

Right-invertibility and row independence:

If A has a left inverse X then the columns of A are linearly independent.

$\downarrow$

If A has a right inverse X then the rows of A are linearly independent.

Connect with independence-dimension inequality:

When A is tall; $A \in \mathbf{R}^{m \times n}$ $\quad m > n$

- rows are linearly dependent

A is not right invertible.

$A \in \mathbf{R}^{m \times n}$ can be right invertible

$m = n$ or $m < n$

Square or Wide

Inverses

Right Inverse: Connection with the Systems of Linear Equations

$$Ax = b$$

- $m = n$ square
- $m < n$ under-determined

- If A has a right inverse X, then we substitute $x = Xb$ in the above system

$$A(Xb) = Ib = b \qquad \Rightarrow \qquad x = Xb \text{ solution of } Ax = b$$

- If solution exists for the system $Ax = b$.

  $x = Xb$ is **the** solution out of many solutions of $Ax = b$.

If A has the right inverse X,
- there is **at least** one solution
- one solution is $x = Xb$

- $m < n$ under-determined

- In summary, a right inverse can be used to find a solution of a square or underdetermined set of linear equations, for any vector b.

Inverses

Inverse:

If a matrix has both left and right inverses;

- they are unique and equal.

$$XA = I, \quad AY = I \qquad \Longrightarrow \qquad X = X(AY) = (XA)Y = Y$$

$X = Y$ is referred to as the inverse of the matrix A, denoted by A^{-1}.

Example:

$$A = \begin{bmatrix} -1 & 1 & -3 \\ 1 & -1 & 1 \\ 2 & 2 & 2 \end{bmatrix} \qquad A^{-1} = \frac{1}{4}\begin{bmatrix} 2 & 4 & 1 \\ 0 & -2 & 1 \\ -2 & -2 & 0 \end{bmatrix}$$

Inverses

Inverse: Connection with the Systems of Linear Equations

$$Ax = b$$

- If A is invertible, $Ax = b$ has the unique solution given by

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

Inverses

Inverse: Properties of Nonsingular or Invertible Matrix

Square matrix A is nonsingular if it is invertible.

Following statements are equivalent for a square matrix A.

1. A is left-invertible
2. the columns of A are linearly independent
3. A is right-invertible
4. the rows of A are linearly independent

$$\begin{tikzcd}
\text{left-invertible} \arrow[r, Rightarrow, "\text{(a)}"] & \text{linearly independent columns} \arrow[d, Rightarrow, "\text{(b)}"] \\
\text{linearly independent rows} \arrow[u, Rightarrow, "\text{(b')}"'] & \text{right-invertible} \arrow[l, Rightarrow, "\text{(a')}"']
\end{tikzcd}$$

Inverses

Inverse: Examples

The identity matrix I is invertible, with inverse $I^{-1} = I$, since $II = I$.

A diagonal matrix A is invertible if and only if its diagonal entries are nonzero. The inverse of an $n \times n$ diagonal matrix A with nonzero diagonal entries is

$$A^{-1} = \begin{bmatrix} \frac{1}{A_{11}} & 0 & \cdots & 0 \\ 0 & \frac{1}{A_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{A_{nn}} \end{bmatrix},$$

since

$$AA^{-1} = \begin{bmatrix} \frac{A_{11}}{A_{11}} & 0 & \cdots & 0 \\ 0 & \frac{A_{22}}{A_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{A_{nn}}{A_{nn}} \end{bmatrix} = I.$$

In compact notation, we have

$$\operatorname{diag}(A_{11}, \ldots, A_{nn})^{-1} = \operatorname{diag}(A_{11}^{-1}, \ldots, A_{nn}^{-1}).$$

Vandermonde Matrix

$$A = \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^{n-1} \\ 1 & t_2 & t_2^2 & \cdots & t_2^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_n & t_n^2 & \cdots & t_n^{n-1} \end{bmatrix}$$

Inverses

Inverse: Examples

the *Gram matrix* associated with a matrix

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

is the matrix of column inner products

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

the Gram matrix is nonsingular if only if A has linearly independent columns

$$\boxed{\begin{aligned} A^T Ax = 0 \quad & \Longrightarrow \quad x^T A^T Ax = (Ax)^T (Ax) = \|Ax\|^2 = 0 \\ & \Longrightarrow \quad Ax = 0 \\ & \Longrightarrow \quad x = 0 \end{aligned}}$$

Example

- $A \in \mathbf{R}^{m \times n}$ with orthonormal columns

$$A^T A = I$$

Inverses

Inverse: Examples

Orthonormal Matrix

- $A \in \mathbf{R}^{n \times n}$ with orthonormal columns

$$A^T A = I$$

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

- A^T is also orthonormal.

Orthogonal Matrix

- $A \in \mathbf{R}^{n \times n}$ with orthonormal columns

$$A^T A = I$$

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

- A^T is also orthogonal.

Matrix with orthonormal columns

- $A \in \mathbf{R}^{m \times n}$ with orthonormal columns

$$A^T A = I$$

Inner product $(Ax)^T(Ay) = x^T A^T A y = x^T y$

Norm $\|Ax\| = \left((Ax)^T(Ax)\right)^{1/2} = (x^T x)^{1/2} = \|x\|$

Distance $\|Ax - Ay\| = \|x - y\|$

Angle $\angle(Ax, Ay) = \angle(x, y)$

Linear transformation using 'matrix with orthonormal columns' preserves norm, distance, angle and inner product.

Inverses

Pseudo Inverse: Matrix with linearly independent columns

- suppose $A \in \mathbf{R}^{m \times n}$ has linearly independent columns
- this implies that A is tall or square ($m \geq n$)

the *pseudo-inverse* of A is defined as

$$A^{\dagger} = (A^T A)^{-1} A^T$$

(Left Pseudo-Inverse)

Equivalent Statements
- A is left-invertible
- the columns of A are linearly independent
- $A^T A$ is nonsingular

- A is left-invertible

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

Inverses

Pseudo Inverse: Matrix with linearly independent rows

- suppose $A \in \mathbf{R}^{m \times n}$ has linearly independent rows
- this implies that A is wide or square $(m \leq n)$

the *pseudo-inverse* of A is defined as

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

(Right Pseudo-Inverse)

Equivalent Statements

- A is right-invertible
- the rows of A are linearly independent
- AA^T is nonsingular

- A is right-invertible

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