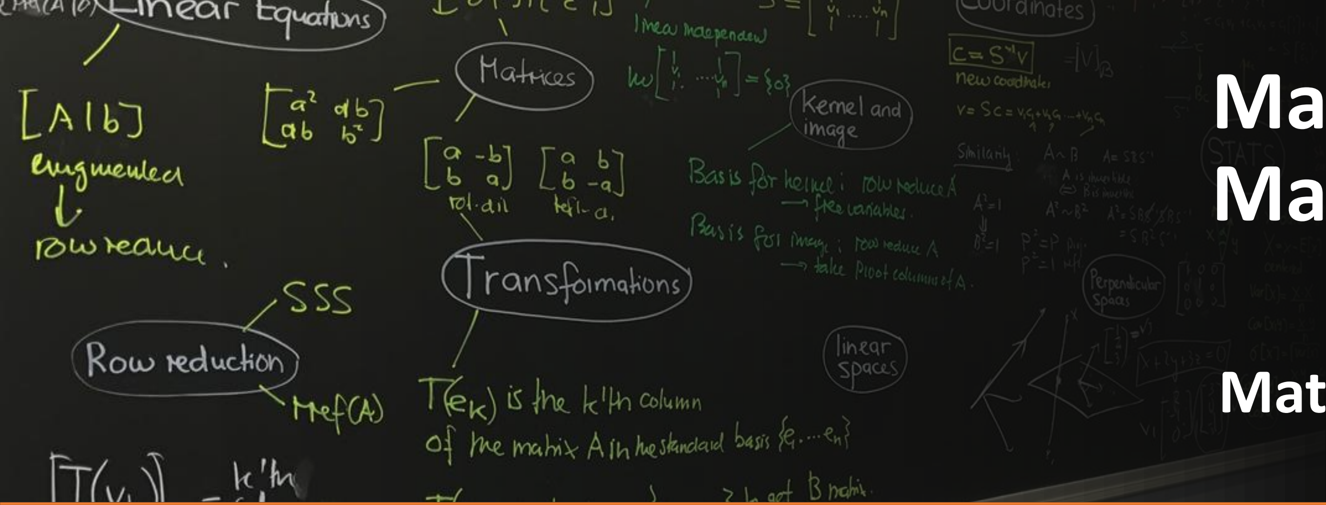


Mathematical Foundations for Machine Learning and Data Science

Matrices – Notation, Application Examples and Basic Operations

Dr. Zubair Khalid

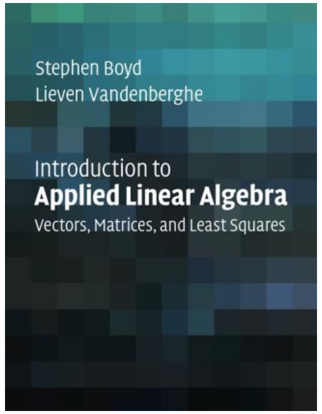
Department of Electrical Engineering
School of Science and Engineering
Lahore University of Management Sciences

https://www.zubairkhalid.org/ee212_2021.html

LUMS
A Not-for-Profit University

Outline

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



Chapter 6

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:

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

Triangular Matrix:

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

$$U_{i,j} = \begin{cases} a_{i,j} & \text{for } i \leq j \\ 0 & \text{for } i > j. \end{cases}$$

A

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 the number of expected COVID-19 in Pakistan cases 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

$A \in R^{m \times n}, \quad B \in R^{m \times n}$

$A + B = C \in R^{m \times n}$

$C_{ij} = A_{ij} + B_{ij} \qquad i = 1, \dots, m, \quad j = 1, \dots, n$

$$\underbrace{\begin{bmatrix} 0 & 4 \\ 7 & 0 \\ 3 & 1 \end{bmatrix}}_{A} + \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 0 & 4 \end{bmatrix}}_{B} = \underbrace{\begin{bmatrix} 1 & 6 \\ 9 & 3 \\ 3 & 5 \end{bmatrix}}_{C}$$

* $A \in R^{m \times n}, \quad \alpha \in R$

$\alpha A = B$

$B_{ij} = \alpha A_{ij}$

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

Operations on Matrices

Transpose and Concept of Symmetric Matrices

* $A \in R^{m \times n}$
* $A^T \in R^{n \times m}$

$$(A^T)_{ij} = A_{ji}$$

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

Symmetric Matrix

$A \in R^{n \times n}$ is symmetric if $A = A^T$

Skew-Symmetric

$$A = -A^T$$

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.

$B \in R^{n \times n}$

$B = A + C$

$A = A^T$

$C = -C^T$

* $A = \frac{1}{2}(B + B^T)$

* $C = \frac{1}{2}(B - B^T)$

$\frac{1}{2}(2B) = B = A + C$

Operations on Matrices

Matrix Norm

$A \in \mathbb{R}^{m \times n}$

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

FROBENIUS Norm of a matrix

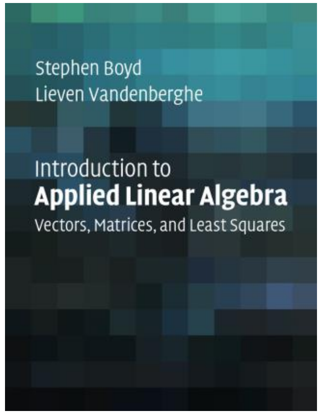
Operations on Matrices

Trace of a Matrix

- $A \in R^{n \times n}$
- $tr(A) = \sum_{i=1}^{n} a_{ii}$

Outline

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



Chapters 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

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

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

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

$= \langle$ i-th row of A, x $\rangle$

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

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

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

Interpretation In terms of Rows of Matrix

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

$$r_i \in R^n$$

$$y_i = r_i^T x$$

$$= \langle r_i, x \rangle$$

Matrix-Vector Multiplication

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

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

Interpretation In terms of Columns of Matrix

$$y \quad = \quad A \quad x$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = [a_1,\ a_2,\ \dots,\ a_n] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$a_i \in R^m$

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

$$\boxed{y = Ax}$$

- 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.

A Not-for-Profit University

Matrix-Vector Multiplication

Application Examples

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

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mn} \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{70} \end{bmatrix}$$

$x_i \rightarrow$ value of i-th product

$m = 200$

$n = 70$

We want to determine value of all products in each warehouse

$$y = Ax \in R^{200}$$

Matrix-Vector Multiplication

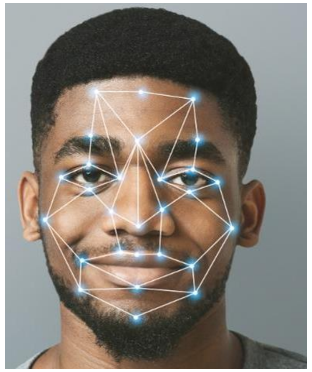
Application Examples

Feature matrix and weight vector

70 features

$$A = \begin{matrix} \text{Photo1} \\ \text{Photo2} \\ \vdots \\ 1000 \end{matrix} \begin{bmatrix} & & \\ & A_{21} & \\ & & \end{bmatrix}$$

$A \in R^{1000 \times 70}$



$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_{70} \end{bmatrix}$$

$$\boxed{y = Aw} \in R^{1000}$$

Matrix-Vector Multiplication

Application Examples

Expansion in a basis

R^n, $a_1, a_2, \ldots, a_n \in R^n$

$b \in R^n$ $\quad b = \beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_n a_n$

$$b = A \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \Rightarrow \boxed{b = A\beta}$$

Linear dependence of columns

$Ax = 0$ for some $x \neq 0$ (Linear dependence b/w columns)

$Ax = 0 \Rightarrow x = 0$ (Linear independence)

LUMS
A Not-for-Profit University

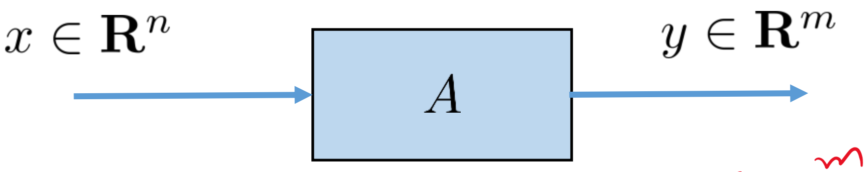
Matrix-Vector Multiplication

Linear Transformation Interpretation:

Input-Output System Interpretation

$$y = Ax$$

$f: R^n \to R^m$



$\boxed{y = Ax}$

$u \in R^n$ $\quad v \in R^m$

$A(\alpha x) = \alpha Ax$

$= \alpha y$

$x + u$ $\quad y + v$

* Scaling

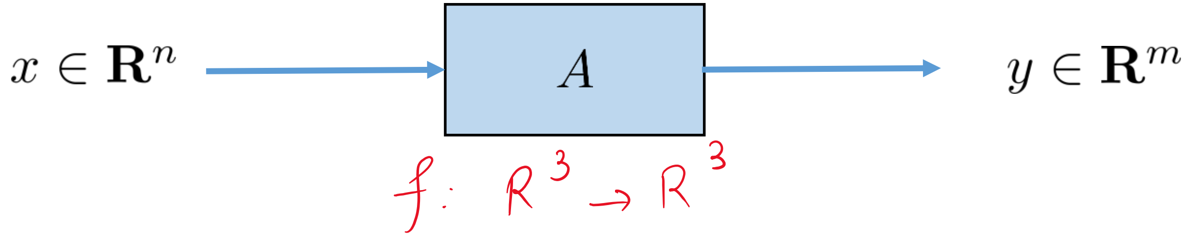
$f: R^n \to R^m$

$\boxed{f(\alpha x + \beta u) = \alpha f(x) + \beta f(u)}$

* Additivity

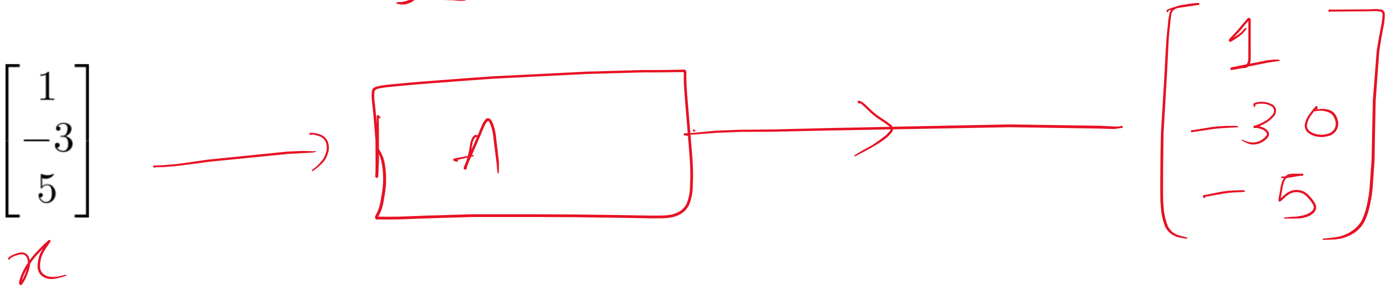
Matrix-Vector Multiplication

Input-Output System Interpretation



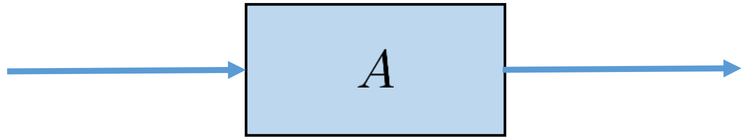
Examples

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



Matrix-Vector Multiplication

Input-Output System Interpretation



$f: R^3 \rightarrow R^2$

Examples

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ x_3 \end{bmatrix} \in R^2$$

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

$$y = \begin{bmatrix} -2 \\ 5 \end{bmatrix} \in R^2$$

LUMS
A Not-for-Profit University

Matrix-Vector Multiplication

Input-Output System Interpretation

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

$f: R^4 \to R^4$

Reverser

Examples

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

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

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

no. of columns in A = no. of rows in B = n

$C \in \mathbf{R}^{m\times p}$ $\quad (m \times n) \quad (n \times p)$

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

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

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

$C_{ij} = \langle$ i-th row of A, j-th column of B $\rangle$

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

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

$A \in R^{2 \times 3}$ $B \in R^{3 \times 2}$ C

Outer product of Vectors:

$a \in R^n$, $b \in R^n$

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

Gram Matrix:

$A \in R^{m \times n}$,

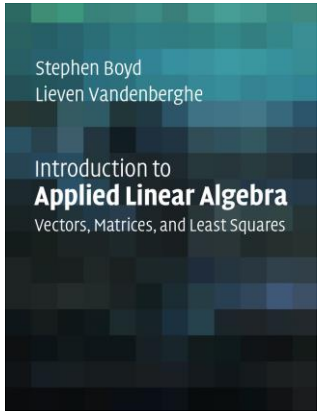
$$G = A^T A = \begin{bmatrix} a_1^T a_1 & a_1^T a_2 & \cdots & a_1^T a_n \\ a_2^T a_1 & a_2^T a_2 & \cdots & a_2^T a_n \\ \vdots & \vdots & \ddots & \vdots \\ a_n^T a_1 & a_n^T a_2 & \cdots & a_n^T a_n \end{bmatrix}$$

$(n \times m)\ (m \times n$

$$G_{ij} = a_i^T a_j$$

Outline

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



Chapters 8 and 11

Systems of Linear Equations

Formulation:

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

$$\begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

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

$$Ax = b$$

- one solution
- multiple solutions
- no solution

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

$b_1, b_2, \ldots, b_m$ - 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 $Ax = 0$

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

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.

↓

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 manu 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

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

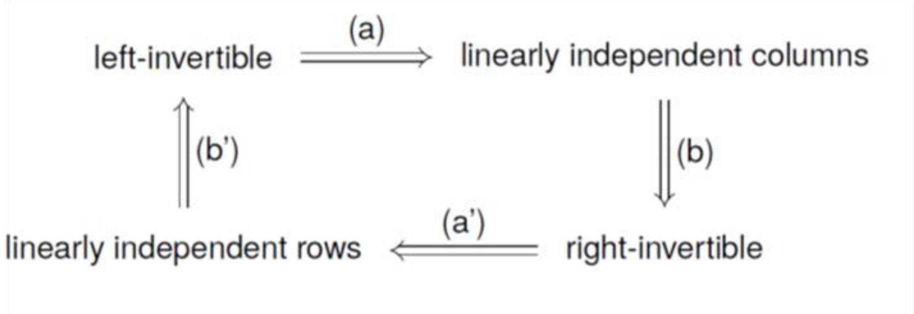
Inverses

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



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

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

since

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

In compact notation, we have

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

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

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

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

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

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

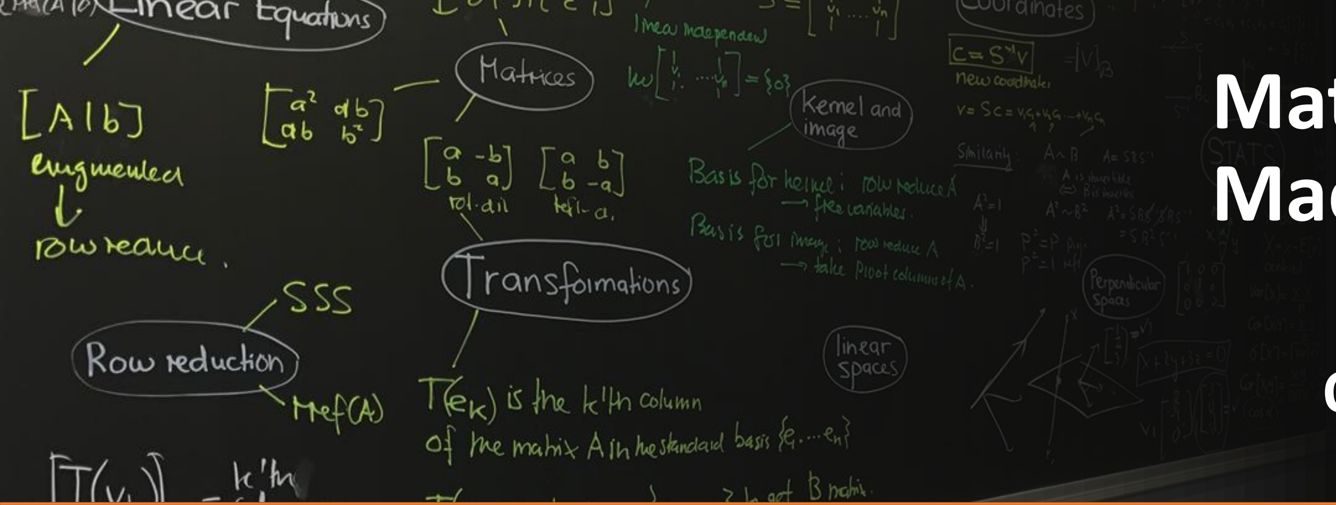
(Right Pseudo-Inverse)

Equivalent Statements

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

- A is right-invertible

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



Mathematical Foundations for Machine Learning and Data Science

QR Factorization and Solution of Linear Equations

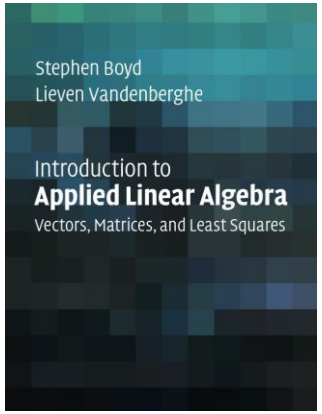
Dr. Zubair Khalid

Department of Electrical Engineering
School of Science and Engineering
Lahore University of Management Sciences

https://www.zubairkhalid.org/ee212_2020.html

Outline

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



Chapters 10.4 and 11.3

Triangular Matrix

- Square matrix $A \in \mathbf{R}^n$ is lower triangular if

$$A_{ij} = 0, \quad j > i$$

$$A = \begin{bmatrix} A_{11} & 0 & \cdots & 0 & 0 \\ A_{21} & A_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_{n-1,1} & A_{n-1,2} & \cdots & A_{n-1,n-1} & 0 \\ A_{n1} & A_{n2} & \cdots & A_{n,n-1} & A_{nn} \end{bmatrix}$$

- Square matrix $A \in \mathbf{R}^n$ is upper triangular if $A_{ij} = 0$ for $j < i$

- Triangular matrix A with nonzero diagonal elements is nonsingular.

$$Ax = 0 \quad \Longrightarrow \quad x = 0$$

Triangular Matrix

Linear Equations with Lower Triangular Matrix

$Ax = b$ $\qquad$ $A \in \mathbf{R}^n$ is lower triangular

$$\begin{bmatrix} A_{11} & 0 & \cdots & 0 & 0 \\ A_{21} & A_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & 0 & 0 \\ A_{n-1,1} & A_{n-1,2} & \cdots & A_{n-1,n-1} & 0 \\ A_{n1} & A_{n2} & \cdots & A_{n,n-1} & A_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

Using forward substitution:

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

Triangular Matrix

Linear Equations with Upper Triangular Matrix

$Ax = b$ $\qquad$ $A \in \mathbf{R}^n$ is upper triangular

Using back substitution:

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

QR Factorization

if $A \in \mathbf{R}^{m \times n}$ has linearly independent columns then it can be factored as

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}$$

- vectors $q_1, \ldots, q_n$ are orthonormal m-vectors:

$$\|q_i\| = 1, \qquad q_i^T q_j = 0 \quad \text{if } i \neq j$$

- diagonal elements R_{ii} are nonzero

QR Factorization

if $A \in \mathbf{R}^{m \times n}$ has linearly independent columns then it can be factored as

$$A = QR$$

Q-factor

- Q is $m \times n$ with orthonormal columns ($Q^T Q = I$)
- if A is square ($m = n$), then Q is orthogonal ($Q^T Q = QQ^T = I$)

R-factor

- R is $n \times n$, upper triangular, with nonzero diagonal elements
- R is nonsingular (diagonal elements are nonzero)

QR Factorization

How to compute?

Gram–Schmidt algorithm

Gram–Schmidt QR algorithm computes Q and R column by column

- after k steps we have a partial QR factorization

$$\begin{bmatrix} a_1 & a_2 & \cdots & a_k \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & \cdots & q_k \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1k} \\ 0 & R_{22} & \cdots & R_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{kk} \end{bmatrix}$$

- columns $q_1, \ldots, q_k$ are orthonormal
- diagonal elements $R_{11}, R_{22}, \ldots, R_{kk}$ are positive

Orthogonalization

$$\tilde{q}_i = a_i - \sum_{j=1}^{i-1} (q_j^T a_i)\, q_j$$

Normalization

$$q_i = \frac{\tilde{q}_i}{\|\tilde{q}_i\|}$$

QR Factorization

How to compute?

Gram–Schmidt algorithm

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1n} \\ 0 & R_{22} & \cdots & R_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & R_{nn} \end{bmatrix}$$

- column k of the equation $A = QR$ reads

$$a_k = R_{1k}q_1 + R_{2k}q_2 + \cdots + R_{k-1,k}q_{k-1} + R_{kk}q_k$$

$$R_{1k} = q_1^T a_k, \qquad R_{2k} = q_2^T a_k, \qquad \ldots, \qquad R_{k-1,k} = q_{k-1}^T a_k$$

QR Factorization

Example

$$
\begin{bmatrix} a_1 & a_2 & a_3 \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix}
$$

$$
= \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}
$$

First column of Q and R

$$
\tilde{q}_1 = a_1 = \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \qquad R_{11} = \|\tilde{q}_1\| = 2, \qquad q_1 = \frac{1}{R_{11}}\tilde{q}_1 = \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix}
$$

QR Factorization

Example

Second column of Q and R

- compute $R_{12} = q_1^T a_2 = 4$
- compute

$$\tilde{q}_2 = a_2 - R_{12} q_1 = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix} - 4 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

- normalize to get

$$R_{22} = \|\tilde{q}_2\| = 2, \qquad q_2 = \frac{1}{R_{22}} \tilde{q}_2 = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

QR Factorization

Example

Third column of Q and R

- compute $R_{13} = q_1^T a_3 = 2$ and $R_{23} = q_2^T a_3 = 8$
- compute

$$\tilde{q}_3 = a_3 - R_{13}q_1 - R_{23}q_2 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - 2\begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} - 8\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix}$$

- normalize to get

$$R_{33} = \|\tilde{q}_3\| = 4, \qquad q_3 = \frac{1}{R_{33}}\tilde{q}_3 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

QR Factorization

Example

Final result

$$
\begin{bmatrix} -1 & -1 & 1 \\ 1 & 3 & 3 \\ -1 & -1 & 5 \\ 1 & 3 & 7 \end{bmatrix} = \begin{bmatrix} q_1 & q_2 & q_3 \end{bmatrix} \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ 0 & R_{22} & R_{23} \\ 0 & 0 & R_{33} \end{bmatrix}
$$

$$
= \begin{bmatrix} -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 \\ -1/2 & 1/2 & 1/2 \\ 1/2 & 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 2 & 4 & 2 \\ 0 & 2 & 8 \\ 0 & 0 & 4 \end{bmatrix}
$$

QR Factorization

Solving Linear Equations

$$Ax = b$$

$$x = A^{-1}b = R^{-1}Q^Tb$$

QR factorization of nonsingular matrix

every nonsingular $A \in \mathbf{R}^{n \times n}$ has a QR factorization

$$A = QR$$

- $Q \in \mathbf{R}^{n \times n}$ is orthogonal ($Q^TQ = QQ^T = I$)
- $R \in \mathbf{R}^{n \times n}$ is upper triangular with positive diagonal elements

Algorithm: to solve $Ax = b$ with nonsingular $A \in \mathbf{R}^{n \times n}$,

1. factor A as $A = QR$
2. compute $y = Q^Tb$
3. solve $Rx = y$ by back substitution

QR Factorization

Solving Linear Equations – Pseudo Inverse

$$Ax = b$$

A is left-invertible
Columns are linear independent

$$A = QR$$

$$A^T A = (QR)^T (QR) = R^T Q^T QR = R^T R,$$

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

A is right-invertible
rows are linear independent

$$A^T = QR$$

$$AA^T = (QR)^T (QR) = R^T Q^T QR = R^T R$$

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