Linear Algebra

Review

Ngoc Hoang Luong

Faculty of Computer Science
University of Information Technology (UIT)
Vietnam National University - Ho Chi Minh City (VNU-HCM)

Maths for Computer Science, Fall 2022

Machine Learning and Linear Algebra

[1]

[1] https://xkcd.com/1838/

References

The contents of this document are taken mainly from the following sources:

- Gilbert Strang. Linear Algebra and Learning from Data. `https://math.mit.edu/~gs/learningfromdata/`
- Gilbert Strang. Introduction to Linear Algebra. `http://math.mit.edu/~gs/linearalgebra/`
- Gilbert Strang. Linear Algebra for Everyone. `http://math.mit.edu/~gs/everyone/`

Table of Contents

Table of Contents

Vectors

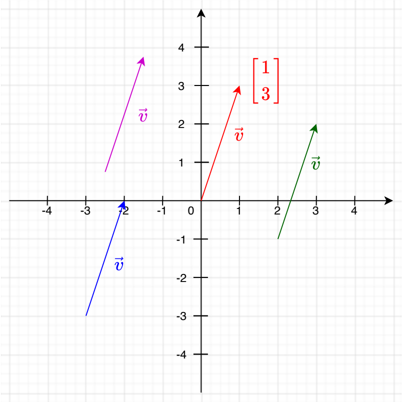
- Vectors are arrays of numerical values.
- Each numerical value is referred to as *coordinate*, *component*, *entry*, or *dimension*.
- The number of components is the vector *dimensionality*.
- e.g., a vector representation of a person: 25 years old (Age), making 30 dollars an hour (Salary), having 6 years of experience (Experience): $[25, 30, 6]$.
- Vectors are special objects that can be added together and multiplied by scalars to produce another object of the same kind.

Geometric Vectors

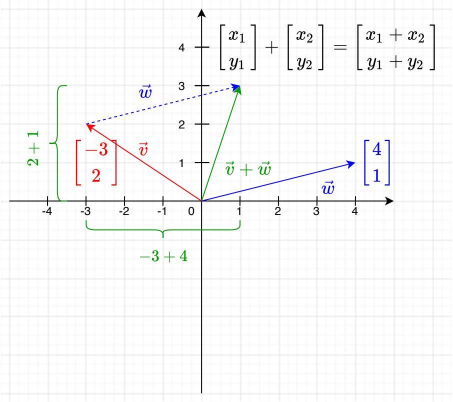
- **Geometric vectors** are often visualized as a quantity that has a **magnitude** as well as a **direction**.
- e.g., the velocity of a person moving at 1 meter/second in the eastern direction and 3 meters/second in the northern direction can be described as a directed line from the origin to $(1, 3)$.
- The **tail** of the vector is at the origin. The **head** is at $(1, 3)$.
- Geometric vectors can have arbitrary tails.
- Two geometric vectors can be added, such that $\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{z}$ is another geometric vector.
- Multiplication by a scalar $\lambda \boldsymbol{x}, \lambda \in \mathbb{R}$, is also a geometric vector.

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Vectors



$\vec{v}$
$\begin{bmatrix} 1 \\ 3 \end{bmatrix}$
$\vec{v}$
$\vec{v}$
$\vec{v}$



$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} x_1 + x_2 \\ y_1 + y_2 \end{bmatrix}$
$\vec{w}$
$2+1$
$\begin{bmatrix} -3 \\ 2 \end{bmatrix}$
$\vec{v}$
$\vec{v}+\vec{w}$
$\begin{bmatrix} 4 \\ 1 \end{bmatrix}$
$\vec{w}$
$-3+4$

Vectors

- Polynomials are vectors. Adding two polynomials results in another polynomial. Multiplied by a scalar, the result is also a polynomial.
- Audio signals are also vectors. Addition of two audio signals and scalar multiplication result in new audio signals.
- Elements of $\mathbb{R}^n$ (tuples of n real numbers) are vectors. For example,

$$\boldsymbol{a} = \begin{bmatrix} 6 \\ 14 \\ -3 \end{bmatrix} \in \mathbb{R}^3$$

  is a triplet of numbers. Adding two vectors $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^n$ component-wise results in another vectors $\boldsymbol{a} + \boldsymbol{b} = \boldsymbol{c} \in \mathbb{R}^n$. Multiplying $\boldsymbol{a} \in \mathbb{R}^n$ by $\lambda \in \mathbb{R}$ results in a scaled vector $\lambda \boldsymbol{a} \in \mathbb{R}^n$.

Basic Operations with Vectors

- Vector of the same dimensionality can be added or subtracted.
- Consider two d-dimensional vectors:

$$\boldsymbol{x}+\boldsymbol{y}=\begin{bmatrix}x_1\\ \dots\\ x_d\end{bmatrix}+\begin{bmatrix}y_1\\ \dots\\ y_d\end{bmatrix}=\begin{bmatrix}x_1+y_1\\ \dots\\ x_d+y_d\end{bmatrix}\quad \boldsymbol{x}-\boldsymbol{y}=\begin{bmatrix}x_1\\ \dots\\ x_d\end{bmatrix}-\begin{bmatrix}y_1\\ \dots\\ y_d\end{bmatrix}=\begin{bmatrix}x_1-y_1\\ \dots\\ x_d-y_d\end{bmatrix}$$

- Vector addition is commutative: $\boldsymbol{x}+\boldsymbol{y}=\boldsymbol{y}+\boldsymbol{x}$.

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Basic Operations with Vectors

- A vector $\boldsymbol{x} \in \mathbb{R}^d$ can be scaled by a factor $a \in \mathbb{R}$ as follows

$$\boldsymbol{v} = a\boldsymbol{x} = a \begin{bmatrix} x_1 \\ \dots \\ x_d \end{bmatrix} = \begin{bmatrix} ax_1 \\ \dots \\ ax_d \end{bmatrix}$$

- Scalar multiplication operation scales the "length" of the vector, but does not change the "direction" (i.e., relative values of different components)

Basic Operations with Vectors

- The **dot product** between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is the sum of the element-wise multiplication of their individual components.

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{d} x_i y_i$$

- The dot product is commutative:

$$\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{d} x_i y_i = \sum_{i=1}^{d} y_i x_i = \boldsymbol{y} \cdot \boldsymbol{x}$$

- The dot product is distributive:

$$\boldsymbol{x} \cdot (\boldsymbol{y} + \boldsymbol{z}) = \boldsymbol{x} \cdot \boldsymbol{y} + \boldsymbol{x} \cdot \boldsymbol{z}$$

Basic Operations with Vectors

- The dot product of a vector with itself produces the squared Euclidean norm. The norm defines the vector length and is denoted by $\|\cdot\|$:

$$\|x\|^2 = \boldsymbol{x} \cdot \boldsymbol{x} = \sum_{i=1}^{d} x_i^2$$

- The Euclidean norm of $x \in \mathbb{R}^d$ is defined as

$$\|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2} = \sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}$$

  and computes the Euclidean distance of $\boldsymbol{x}$ from the origin.
- The Euclidean norm is also known as the L_2-norm.

Basic Operations with Vectors

- A generalization of the Euclidean norm is the L_p-norm, denoted by $\|\cdot\|_p$:

$$\|x\|_p = \left(\sum_{i=1}^{d} |x_i|^p\right)^{(1/p)}$$

  where p is a positive value.
- When $p = 1$, we have the Manhattan norm, or the L_1-norm.

Basic Operations with Vectors

- Vectors can be **normalized** to unit length by dividing them with their norm:

$$\boldsymbol{x}' = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} = \frac{\boldsymbol{x}}{\sqrt{\boldsymbol{x} \cdot \boldsymbol{x}}}$$

- The resulting vector is a **unit vector**.
- The squared Euclidean distance $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ can be shown to be the dot product of $\boldsymbol{x} - \boldsymbol{y}$ with itself:

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = (\boldsymbol{x} - \boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y}) = \sum_{i=1}^{d} (x_i - y_i)^2$$

Basic Operations with Vectors

- **Cauchy-Schwarz Inequality**: the dot product between a pair of vectors is bounded above by the product of their lengths.

$$\left|\sum_{i=1}^{d} x_i y_i\right| = |\boldsymbol{x} \cdot \boldsymbol{y}| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$$

- **Triangle Inequality**: Consider the triangle formed by the origin, $\boldsymbol{x}$, and $\boldsymbol{y}$, the side length $\|\boldsymbol{x} - \boldsymbol{y}\|$ is no greater than the sum $\|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ of the other two sides.

UIT
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Basic Operations with Vectors

- Consider the triangle created by the origin, $\boldsymbol{x}$, and $\boldsymbol{y}$. Find the angle θ between $\boldsymbol{x}$ and $\boldsymbol{y}$.
- The side lengths of this triangle are: $a = \|\boldsymbol{x}\|$, $b = \|\boldsymbol{y}\|$, and $c = \|\boldsymbol{x} - \boldsymbol{y}\|$. Using the cosine law, we have:

$$\begin{aligned}
\cos(\theta) = \frac{a^2 + b^2 - c^2}{2ab} &= \frac{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \\
&= \frac{\|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - (\boldsymbol{x} - \boldsymbol{y}) \cdot (\boldsymbol{x} - \boldsymbol{y})}{2\|\boldsymbol{x}\|\|\boldsymbol{y}\|} \\
&= \frac{\boldsymbol{x} \cdot \boldsymbol{y}}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}
\end{aligned}$$

- Two vectors are **orthogonal** if their dot product is 0.
- The vector $\mathbf{0}$ is considered orthogonal to every vector.

Matrices

Definition

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

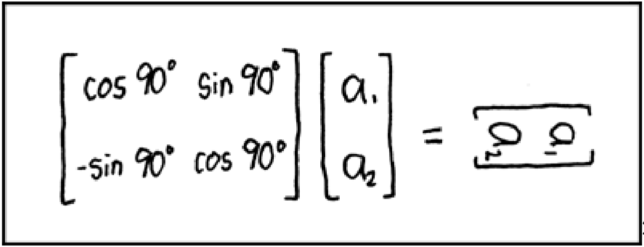
$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}, \quad a_{ij} \in \mathbb{R}$$

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

Matrices

- A matrix has the same number of rows as columns is a **square** matrix. Otherwise, it is a **rectangular** matrix.
- A matrix having more rows than columns is referred to as *tall*, while a matrix having more columns than rows is referred to as *wide* or *fat*.
- A scalar can be considered as a 1×1 "matrix".
- A d-dimensional vector can be considered a $1 \times d$ matrix when it is treated as a **row vector**.
- A d-dimensional vector can be considered a $d \times 1$ matrix when it is treated as a **column vector**.
- By defaults, vectors are assumed to be column vectors.

Matrix-Vector Multiplication



[2]

[2] https://xkcd.com/184/

Matrix-Vector Multiplication Ax

- Multiply A times $\boldsymbol{x}$ using rows of A.

$$\begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ 2x_1 + 4x_2 \\ 3x_1 + 7x_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{a}_1^* \boldsymbol{x} \\ \boldsymbol{a}_2^* \boldsymbol{x} \\ \boldsymbol{a}_3^* \boldsymbol{x} \end{bmatrix}$$

  $A\boldsymbol{x} =$ dot products of rows of A with $\boldsymbol{x}$.

- Multiply A times $\boldsymbol{x}$ using columns of A.

$$\begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix} = x_1 \boldsymbol{a}_1 + x_2 \boldsymbol{a}_2$$

  $A\boldsymbol{x} =$ combination of columns of $\boldsymbol{a}_1,\ \boldsymbol{a}_2$ (of A) scaled by scalars $x_1,\ x_2$ respectively.

Linear Combinations of Columns

Ax

Ax is a linear combination of the columns of A.

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

$$A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n$$

Column space of $A = \mathbf{C}(A) =$ all vectors $A\boldsymbol{x}$

$=$ all linear combinations of the columns

Column Space of A

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$$

- Each $A\boldsymbol{x}$ is a vector in the $\mathbb{R}^3$ space.
- All combinations $A\boldsymbol{x} = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2$ produce what part of $\mathbb{R}^3$?
- Answer: a **plane**, containing:
  - the line of all vectors $x_1\boldsymbol{a}_1$,
  - the line of all vectors $x_2\boldsymbol{a}_2$,
  - the sum of any vector on one line + any vector on the other line, filling out an **infinite plane** containing the two lines, but not the whole $\mathbb{R}^3$.

Definition

The combinations of the columns fill out the column space of A.

Column Space of A

$$Ax = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$$

- **C**(A) is plane.
- The plane includes $(0,0)$, produced when $x_1 = x_2 = 0$.
- The plane includes $(5,6,10) = \boldsymbol{a}_1 + \boldsymbol{a}_2$ and $(-1,-2,-4) = \boldsymbol{a}_1 - \boldsymbol{a}_2$. Every combination $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2$ is in **C**(A).
- The probability the plane does not include a random point **rand**(3,1)? Which points are in the plane?

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

$\boldsymbol{b}$ is in **C**(A) exactly when $A\boldsymbol{x} = \boldsymbol{b}$ has a solution $\boldsymbol{x}$.
$\boldsymbol{x}$ shows how to express $\boldsymbol{b}$ as a combination of the columns of A.

Column Space of A

- $\boldsymbol{b} = (1,1,1)$ is not in $\mathbf{C}(A)$ because

$$A\boldsymbol{x} = \begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \quad \text{is unsolvable.}$$

- What is the column space of A_2?

$$\begin{bmatrix} 2 & 3 & 5 \\ 2 & 4 & 6 \\ 3 & 7 & 10 \end{bmatrix}$$

  - $\boldsymbol{a}_3 = \boldsymbol{a}_1 + \boldsymbol{a}_2$, is already in $\mathbf{C}(A)$, the plane of $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$.
  - Including this **dependent** column does not go beyond $\mathbf{C}(A)$.
  - $\mathbf{C}(A_2)$=$\mathbf{C}(A)$.

- What is the column space of A_3?

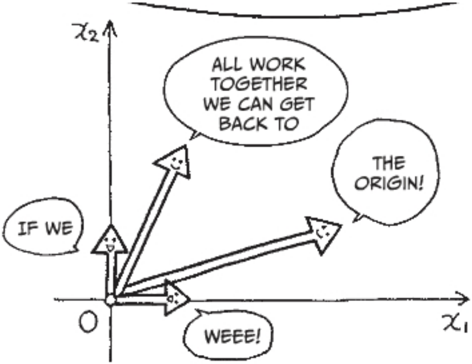
$$\begin{bmatrix} 2 & 3 & 1 \\ 2 & 4 & 1 \\ 3 & 7 & 1 \end{bmatrix}$$

  - $\boldsymbol{a}_3 = (1,1,1)$ is not in the plane $\mathbf{C}(A)$.
  - Visualize the xy-plane and a third vector (x_3, y_3, z_3) out of the plane (meaning that $z_3 \neq 0$).
  - $\mathbf{C}(A_3)$=$\mathbb{R}^3$.

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Column Spaces of $\mathbb{R}^3$

- Subspaces of $\mathbb{R}^3$:
  - The zero vector $(0, 0, 0)$.
  - A line of all vectors $x_1\boldsymbol{a}_1$.
  - A plane of all vectors $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2$.
  - The whole $\mathbb{R}^3$ with all vectors $x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + x_3\boldsymbol{a}_3$.
- Vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ need to be **independent**. The only combination that gives the zero vector is $0\boldsymbol{a}_1 + 0\boldsymbol{a}_2 + 0\boldsymbol{a}_3$.
- The zero vector is in every subspace.



LINEAR DEPENDENCE [3]

[3] https://mathsci2.appstate.edu/ sjg/class/2240/hf14.html

Independent Columns, Basis, and Ranks of A

Definition

A **basis** for a subspace is a full set of independent vectors: All vectors in the space are combinations of the basis vector.

Create a matrix C whose columns come directly from A:

- If column 1 of A is not all zero, put it into C.
- If column 2 of A is not a multiple of column 1, put it into C.
- If column 3 of A is not a combination of columns 1 and 2, put it into C. *Continue*.
- At the end, C will have r columns ($r \leq n$). They are independent columns, and they are a "basis" for the column space $\mathbf{C}(A)$.

Independent Columns, Basis, and Ranks of A

If $A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix}$ then $C = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix}$ $\quad$ $n = 3$ columns in A; $r = 2$ columns in C

If $A = \begin{bmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{bmatrix}$ then $C = A$ $\quad$ $n = 3$ columns in A; $r = 3$ columns in C

If $A = \begin{bmatrix} 1 & 2 & 5 \\ 1 & 2 & 5 \\ 1 & 2 & 5 \end{bmatrix}$ then $C = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\quad$ $n = 3$ columns in A; $r = 1$ columns in C

- The number r counts independent columns.
- It is the "dimension" of the column space of A and C (same space).

Definition

The **rank** of a matrix is the **dimension** of its column space.

Rank Factorization $A = CR$

- The matrix C connects to A by a third matrix R: $A = CR$.
- $A \in \mathbb{R}^{m\times n},\ C \in \mathbb{R}^{m\times r},\ R \in \mathbb{R}^{r\times n}$

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} = CR$$

- C multiplies the first column of R produces column 1 of A.
- C multiplies the second column of R produces column 2 of A.
- C multiplies the third column of R produces column 3 of A.
- Combinations of the columns of C produce the columns of A $\longrightarrow$ Put the right numbers in R.

Definition

$R =$**rref**(A) = row-reduced echelon form of A.

Rank Factorization $A = CR$

$$A = \begin{bmatrix} 1 & 3 & 8 \\ 1 & 2 & 6 \\ 0 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 3 \\ 1 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 2 \end{bmatrix} = CR$$

- The matrix R has $r = 2$ rows $\boldsymbol{r}_1^*,\ \boldsymbol{r}_2^*$.
- Multiply row 1 of C with R, we get $\boldsymbol{r}_1^* + 3\boldsymbol{r}_2^* \rightarrow$ row 1 of A.
- Multiply row 2 of C with R, we get $\boldsymbol{r}_1^* + 2\boldsymbol{r}_2^* \rightarrow$ row 2 of A.
- Multiply row 3 of C with R, we get $0\boldsymbol{r}_1^* + 1\boldsymbol{r}_2^* \rightarrow$ row 3 of A.
- R has independent rows: No row is a combination of the other rows. Hint: Look at the zeros and ones in R - the identity matrix I in R.
- The rows of R are a **basis for the row space** of A.
- Notation: The row space of matrix $A = \mathbf{C}(A^\top)$.

Rank Factorization $A = CR$

1. The r columns of C are independent (by their construction).
2. Every column of A is a combination of those r columns of C (because $A = CR$).
3. The r rows of R are independent (they contain the matrix I_r).
4. Every row of A is a combination of those r rows of R (because $A = CR$).

Key facts:

- The r columns of C is a **basis** for **C**(A): dimension r.
- The r rows of R is a **basis** for **C**$(A^\top)$: dimension r.

Notice

The number of independent columns = The number of independent rows.
The column space and row space of A both have dimension r.
The column rank of A = The row rank of A.

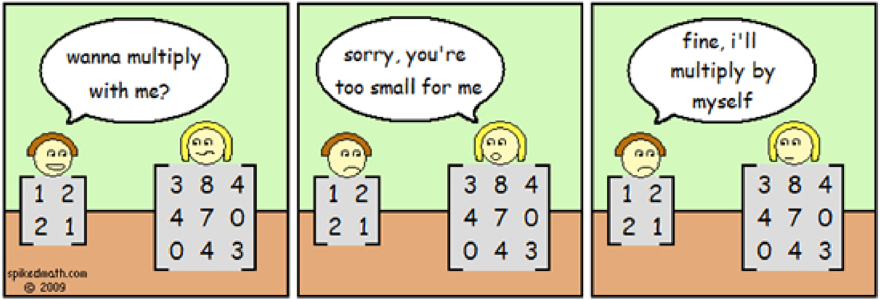
Q&A

Question: If an $n \times n$ matrix A has n independent columns, then $C = ?,\ R = ?$

Answer: $C = A,\ R = I$.

Table of Contents

UIT
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN



[4]

[4] https://mathsci2.appstate.edu/ sjg/class/2240/hf14.html

Compute AB by Inner Products

- **Inner products** (rows times columns) produce each of the numbers in $AB = C$:

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ a_{21} & a_{22} & a_{23} \\ \cdot & \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot & b_{13} \\ \cdot & \cdot & b_{23} \\ \cdot & \cdot & b_{33} \end{bmatrix} = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & c_{23} \\ \cdot & \cdot & \cdot \end{bmatrix}$$

- $c_{ij} =$ (row i of A)$\cdot$(column j of B)

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} = \boldsymbol{a}_i^* \boldsymbol{b}_j$$

Rank-1 Matrix

- **Outer products** (columns times rows) produce **rank one matrices**.

$$\boldsymbol{u}\boldsymbol{v}^\top = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 3 & 4 & 6 \end{bmatrix} = \begin{bmatrix} 6 & 8 & 12 \\ 6 & 8 & 12 \\ 3 & 4 & 6 \end{bmatrix}$$

- An $m \times 1$ matrix (a column $\boldsymbol{u}$) times a $1 \times p$ matrix (a row $\boldsymbol{v}^\top$) gives an $m \times p$ matrix.
- All columns of $\boldsymbol{u}\boldsymbol{v}^\top$ are multiples of $\boldsymbol{u}$.
- All rows of $\boldsymbol{u}\boldsymbol{v}^\top$ are multiples of $\boldsymbol{v}^\top$.
- The column space of $\boldsymbol{u}\boldsymbol{v}^\top$ is the line through $\boldsymbol{u}$.
- The row space of $\boldsymbol{u}\boldsymbol{v}^\top$ is the line through $\boldsymbol{v}$.
- All non-zero matrices $\boldsymbol{u}\boldsymbol{v}^\top$ have rank one.

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

$AB =$ Sum of Rank-1 Matrices

- The product AB is the sum of columns $\boldsymbol{a}_k$ times rows $\boldsymbol{b}_k^*$.

$$AB = \begin{bmatrix} | & & | \\ \boldsymbol{a}_1 & \dots & \boldsymbol{a}_n \\ | & & | \end{bmatrix} \begin{bmatrix} - & \boldsymbol{b}_1^* & - \\ & \vdots & \\ - & \boldsymbol{b}_n^* & - \end{bmatrix} = \boldsymbol{a}_1\boldsymbol{b}_1^* + \boldsymbol{a}_2\boldsymbol{b}_2^* + \cdots + \boldsymbol{a}_n\boldsymbol{b}_n^*$$

- Example:

$$\begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 4 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \begin{bmatrix} 2 & 4 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 12 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 17 \end{bmatrix}$$

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Insight from Column times Row

- Looking for the important part of a matrix A.
- Factor A into CR and look at the pieces $\boldsymbol{c}_k \boldsymbol{r}_k^*$ of $A = CR$.
- Factoring A into CR is the reverse of multiplying $CR = A$.
- The inside information about A is not visible until A is factored.

Important Factorizations

1. $A = LU$: elimination
2. $A = QR$: orthogonalization
3. $S = Q\Lambda Q^\top$: eigenvalues and orthonormal eigenvectors
4. $A = X\Lambda X^{-1}$: diagonalization
5. $A = U\Sigma V^\top$: Singular Value Decomposition (SVD)

Inverse Matrices

- The square matrix A is invertible if there exists a matrix A^{-1} that

$$A^{-1}A = I \text{ and } AA^{-1} = I$$

- The matrix A cannot have two different inverses. Suppose $BA = I$ and also $AC = I$. Then $B = C$.

$$B(AC) = (BA)C \text{ gives } BI = IC \text{ or } B = C.$$

- If A is invertible, the one and only solution to $A\boldsymbol{x} = \boldsymbol{b}$ is $\boldsymbol{x} = A^{-1}\boldsymbol{b}$.
- If $A\boldsymbol{x} = \boldsymbol{0}$ for a nonzero vector $\boldsymbol{x}$, then A has no inverse.
- If A and B are invertible then so is AB. The inverse of AB is

$$(AB)^{-1} = B^{-1}A^{-1}$$

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

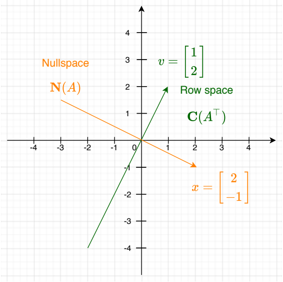
Table of Contents

Example 1

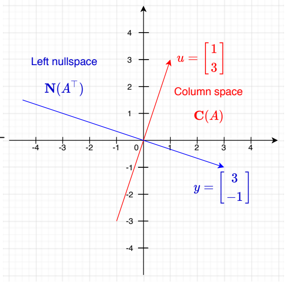
$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^\top$$

- Column space $\mathbf{C}(A)$ is the line through $\boldsymbol{u} = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$.
- Row space $\mathbf{C}(A^\top)$ is the line through $\boldsymbol{v} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$.
- Nullspace $\mathbf{N}(A)$ is the line through $\boldsymbol{x} = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$. $A\boldsymbol{x} = \mathbf{0}$.
- Left nullspace $\mathbf{N}(A^\top)$ is the line through $\boldsymbol{y} = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$. $A^\top\boldsymbol{y} = \mathbf{0}$.

Example 1



$$A = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} = \boldsymbol{u}\boldsymbol{v}^\top$$



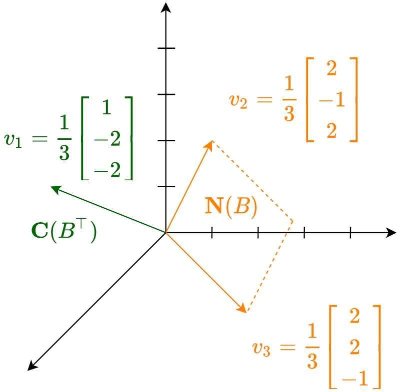
Definition

The column space **C**(A) contains all combinations of the columns of A.
The row space **C**$(A^\top)$ contains all combinations of the columns of $A^\top$.
The nullspace **N**(A) contains all solutions $\boldsymbol{x}$ to $A\boldsymbol{x} = \mathbf{0}$.
The left nullspace **N**(A) contains all solutions $\boldsymbol{y}$ to $A^\top\boldsymbol{y} = \mathbf{0}$.

Example 2

$$B = \begin{bmatrix} 1 & -2 & -2 \\ 3 & -6 & -6 \end{bmatrix}$$

- The row space $\mathbf{C}(B^\top)$ is the infinite line through $\boldsymbol{v}_1 = \frac{1}{3}(1, -2, -2)$.
- $B\boldsymbol{x} = \mathbf{0}$ has solutions $\boldsymbol{x}_1 = (2, 1, 0)$ and $\boldsymbol{x}_2 = (2, 0, 1)$.
- $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are in the same plane with $v_2 = \frac{1}{3}(2, -1, 2)$ and $v_3 = \frac{1}{3}(2, 2, -1)$.
- The nullspace $\mathbf{N}(B)$ has an **orthonormal basis** $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$, is the infinite plane of $\boldsymbol{v}_2$ and $\boldsymbol{v}_3$.
- $\boldsymbol{v}_1, \boldsymbol{v}_2, \boldsymbol{v}_3$: an orthonormal basis for $\mathbb{R}^3$.



Subspaces of A

If $A\boldsymbol{x} = \mathbf{0}$ then $\begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

- $\boldsymbol{x}$ is orthogonal to every row of A.
- Every $\boldsymbol{x}$ in the nullspace of A is orthogonal to the row space of A.
- Every $\boldsymbol{y}$ in the nullspace of $A^\top$ is orthogonal to the column space of A.

$$\begin{array}{lcccc} & \mathbf{N}(A) \perp & \mathbf{C}(A^\top) & \mathbf{N}(A^\top) \perp & \mathbf{C}(A) \\ \text{Dimensions} & n-r & r & m-r & r \end{array}$$

- Two orthogonal subspaces. The dimensions add to n and to m.

Table of Contents

$Ax = b$ by Elimination

The usual order:

- Column 1.
  - Row 1 is the first pivot row.
  - Multiply row 1 by numbers $l_{21}, l_{31}, \ldots, l_{n1}$ and subtract from rows $2, 3, \ldots, n$ of A respectively.

$$\text{Multipliers } l_{21} = \frac{a_{21}}{a_{11}} \quad l_{31} = \frac{a_{31}}{a_{11}} \quad \ldots \quad l_{n1} = \frac{a_{n1}}{a_{11}}$$

$$[A \mid \boldsymbol{b}] = \left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 4 & 5 & -3 & 6 & 9 \\ -2 & 5 & -2 & 6 & 4 \\ 4 & 11 & -4 & 8 & 2 \end{array}\right] \to \left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 0 & 3 & -1 & 2 & -1 \\ 0 & 6 & -3 & 8 & 9 \\ 0 & 9 & -2 & 4 & -8 \end{array}\right]$$

$Ax = b$ by Elimination

The usual order:

- Column 2.
  - The new row 2 is the second pivot row.
  - Multiply row 2 by numbers $l_{32}, l_{42}, \ldots, l_{n2}$ and subtract from rows $3, 4, \ldots, n$ of A respectively.

$$\text{Multipliers } l_{32} = \frac{a_{32}}{a_{22}} \quad l_{42} = \frac{a_{42}}{a_{22}} \quad \ldots \quad l_{n2} = \frac{a_{n2}}{a_{22}}$$

$$\left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 0 & 3 & -1 & 2 & -1 \\ 0 & 6 & -3 & 8 & 9 \\ 0 & 9 & -2 & 4 & -8 \end{array}\right] \to \left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 0 & 3 & -1 & 2 & -1 \\ 0 & 0 & -1 & 4 & 11 \\ 0 & 0 & 1 & -2 & -5 \end{array}\right]$$

$Ax = \boldsymbol{b}$ by Elimination

The usual order:

- Column 3.
  - The new row 3 is the third pivot row.
  - Multiply row 3 by numbers $l_{43}, l_{53}, \dots, l_{n3}$ and subtract from rows $4, 5, \dots, n$ of A respectively.

$$\text{Multipliers } l_{43} = \frac{a_{43}}{a_{33}} \quad l_{53} = \frac{a_{53}}{a_{33}} \quad \dots \quad l_{n3} = \frac{a_{n3}}{a_{33}}$$

$$\left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 0 & 3 & -1 & 2 & -1 \\ 0 & 0 & -1 & 4 & 11 \\ 0 & 0 & 1 & -2 & -5 \end{array}\right] \to \left[\begin{array}{cccc|c} 2 & 1 & -1 & 2 & 5 \\ 0 & 3 & -1 & 2 & -1 \\ 0 & 0 & -1 & 4 & 11 \\ 0 & 0 & 0 & 2 & 6 \end{array}\right] = [U \mid \boldsymbol{c}]$$

- Columns 3 to n: Eliminating on A until obtaining the upper triangular U: n pivots on its diagonal.

$$\begin{aligned} 2x_1 + \; x_2 - x_3 + 2x_4 &= 5 \\ 3x_2 - x_3 + 2x_4 &= -1 \\ - x_3 + 4x_4 &= 11 \\ 2x_4 &= 6 \end{aligned}$$

By back substitution, we get

$$x_4 = 3, \quad x_3 = 1, \quad x_2 = -2, \quad x_1 = 1$$

Lower Triangular L and Upper Triangular U

- Elimination on $A\boldsymbol{x} = \boldsymbol{b}$ produces the upper triangular matrix

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

- and the lower triangular matrix

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 2 & 1 & 0 \\ 2 & 3 & -1 & 1 \end{bmatrix}$$

- Elimination factors A into a lower triangular L times an upper triangular U.

$$A = LU$$

The Factorization $A = LU$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} \begin{bmatrix} \text{pivot row 1} \\ \text{pivot row 2} \\ \text{pivot row 3} \\ \text{pivot row 4} \end{bmatrix} = \begin{bmatrix} 2 & 1 & -1 & 2 \\ 4 & 5 & -3 & 6 \\ -2 & 5 & -2 & 6 \\ 4 & 11 & -4 & 8 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ l_{21} \\ l_{31} \\ l_{41} \end{bmatrix} \begin{bmatrix} \text{pivot row 1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & x & x & x \\ 0 & x & x & x \\ 0 & x & x & x \end{bmatrix}$$

$$\boxed{l_{ij} = \frac{a_{ij}}{a_{jj}}}$$

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

$$= \begin{bmatrix} 2 & 1 & -1 & 2 \\ 4 & 2 & -2 & 4 \\ -2 & -1 & 1 & -2 \\ 4 & 2 & -2 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 3 & -1 & 2 \\ 0 & 6 & -3 & 8 \\ 0 & 9 & -2 & 4 \end{bmatrix}$$

The first step reduces the 4×4 problem to a 3×3 problem by removing $l_1 u_1^*$.

The Factorization $A = LU$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} \begin{bmatrix} \text{pivot row 1} \\ \text{pivot row 2} \\ \text{pivot row 3} \\ \text{pivot row 4} \end{bmatrix} = \boldsymbol{l}_1 \boldsymbol{u}_1^* + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 3 & -1 & 2 \\ 0 & 6 & -3 & 8 \\ 0 & 9 & -2 & 4 \end{bmatrix}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \begin{bmatrix} 0 \\ 1 \\ l_{32} \\ l_{42} \end{bmatrix} \begin{bmatrix} \text{pivot row 2} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{bmatrix} \qquad \boxed{l_{ij} = \frac{a_{ij}}{a_{jj}}}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 0 & 3 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & x & x \\ 0 & 0 & x & x \end{bmatrix}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 3 & -1 & 2 \\ 0 & 6 & -2 & 4 \\ 0 & 9 & -3 & 6 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

The second step reduces the 3×3 problem to a 2×2 problem by removing $l_2 u_2^*$.

The Factorization $A = LU$

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} \begin{bmatrix} \text{pivot row 1} \\ \text{pivot row 2} \\ \text{pivot row 3} \\ \text{pivot row 4} \end{bmatrix} = \boldsymbol{l}_1 \boldsymbol{u}_1^* + \boldsymbol{l}_2 \boldsymbol{u}_2^* + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 1 & -2 \end{bmatrix}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \boldsymbol{l}_2 \boldsymbol{u}_2^* + \begin{bmatrix} 0 \\ 0 \\ 1 \\ l_{43} \end{bmatrix} \begin{bmatrix} \text{pivot row 3} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x \end{bmatrix}$$

$$\boxed{l_{ij} = \frac{a_{ij}}{a_{jj}}}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \boldsymbol{l}_2 \boldsymbol{u}_2^* + \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & -1 & 4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x \end{bmatrix}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \boldsymbol{l}_2 \boldsymbol{u}_2^* + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 4 \\ 0 & 0 & 1 & -4 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

$$= \boldsymbol{l}_1 \boldsymbol{u}_1^* + \boldsymbol{l}_2 \boldsymbol{u}_2^* + \boldsymbol{l}_3 \boldsymbol{u}_3^* + \boldsymbol{l}_4 \boldsymbol{u}_4^*$$

The third step reduces the 2×2 problem to a single number by removing $l_3 u_3^*$.

Elimination and $A = LU$

- Start from $\begin{bmatrix} A & \boldsymbol{b} \end{bmatrix} = \begin{bmatrix} LU & \boldsymbol{b} \end{bmatrix}$.
- Elimination produces $\begin{bmatrix} U & L^{-1}\boldsymbol{b} \end{bmatrix} = \begin{bmatrix} U & \boldsymbol{c} \end{bmatrix}$.
- Elimination on $A\boldsymbol{x} = \boldsymbol{b}$ produces the equation $U\boldsymbol{x} = \boldsymbol{c}$ that are ready for back substitution.
- $A = LU = \sum \boldsymbol{l}_i \boldsymbol{u}_i^* =$ **sum of rank one matrices**.

Table of Contents

Orthogonality

- Orthogonal $\sim$ perpendicular.
- Orthogonal vectors $\boldsymbol{x}$ and $\boldsymbol{y}$:

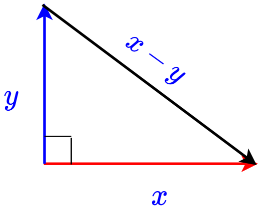
$$\boldsymbol{x}^\top \boldsymbol{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 0$$

Law of Cosines: θ is the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$:

$$\|\boldsymbol{x} - \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 - 2\|\boldsymbol{x}\|\|\boldsymbol{y}\| \cos\theta$$

Orthogonal vectors have $\cos\theta = 0$.
Pythagoras Law:

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{y}\|^2 &= \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 \\ (\boldsymbol{x} - \boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}) &= \boldsymbol{x}^\top \boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y} \\ \boldsymbol{x}^\top \boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y} - \boldsymbol{x}^\top \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{x} &= \boldsymbol{x}^\top \boldsymbol{x} + \boldsymbol{y}^\top \boldsymbol{y} \\ \boldsymbol{x}^\top \boldsymbol{y} &= 0 \end{aligned}$$



Orthogonal Basis

- Orthogonal basis for a subspace: Every pair of basis vectors has $\boldsymbol{v}_i^\top \boldsymbol{v}_j = 0$
- Orthonormal basis: Orthogonal basis of unit vectors: Every $\boldsymbol{v}_i^\top \boldsymbol{v}_i = 1$ (length 1).
- From orthogonal to orthonormal, divide every basis vector $\boldsymbol{v}_i$ by its length $\|\boldsymbol{v}_i\|$.
- The standard basis is orthogonal (and orthonormal) in $\mathbb{R}^n$:

$$\text{Standard basis } \boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k} \text{ in } \mathbb{R}^3 \quad \boldsymbol{i} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{j} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{k} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

- Every subspace of $\mathbb{R}^n$ has an orthogonal basis.

Orthogonal Subspaces

- Subspace **S** is orthogonal to subspace **T**: Every vector in **S** is orthogonal to every vector in **T**.

Orthogonal Subspaces

- The row space $\mathbf{C}(A^\top)$ is orthogonal to the nullspace $\mathbf{N}(A)$.

$$A\boldsymbol{x} = \begin{bmatrix} \text{row } 1 \\ \vdots \\ \text{row } m \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{x} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

- The column space $\mathbf{C}(A)$ is orthogonal to the left nullspace $\mathbf{N}(A^\top)$.

$$A^\top\boldsymbol{y} = \begin{bmatrix} (\text{column } 1)^\top \\ \vdots \\ (\text{column } m)^\top \end{bmatrix} \begin{bmatrix} \\ \boldsymbol{y} \\ \\ \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$$

Orthogonal Subspaces

- Every vector $\boldsymbol{v}$ in $\mathbb{R}^n$ has a row space component $\boldsymbol{v}_{row}$ and a nullspace component $\boldsymbol{v}_{null}$: $\boldsymbol{v} = \boldsymbol{v}_{row} + \boldsymbol{v}_{null}$

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

- The row space $\mathbf{C}(A^\top)$ is the plane of all vectors $\beta_1 \boldsymbol{a}_1^* + \beta_2 \boldsymbol{a}_2^*$.
- The nullspace $\mathbf{N}(A)$ is the line through $\boldsymbol{u} = (0, 0, 1)$: all vectors $\beta_3 \boldsymbol{u}$

$$\boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} \in \mathbb{R}^3 \quad \boldsymbol{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \underbrace{\beta_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta_2 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}}_{\boldsymbol{v}_{row}} + \underbrace{\beta_3 \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}}_{\boldsymbol{v}_{null}}$$

- Dimensions: $\textbf{dim } \mathbf{C}(A^\top) + \textbf{dim } \mathbf{N}(A) = r + (n - r) = n$.
- A row space basis (r vectors) and a nullspace basis ($n - r$ vectors) produces a basis for the whole $\mathbb{R}^n$ (n vectors).

Fundamental Theorem in Linear Algebra

The row space and nullspace of A are orthogonal complements in $\mathbb{R}^n$.

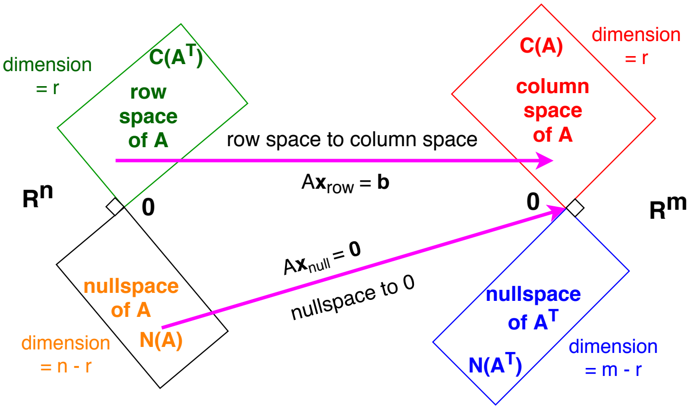
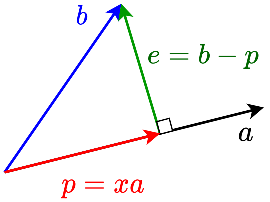


Figure: Two pairs of orthogonal subspaces.



- $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$
- $\boldsymbol{p} = x\boldsymbol{a}$
- Because $\boldsymbol{e}$ is orthogonal to $\boldsymbol{a}$:

$$\begin{aligned} \boldsymbol{a}^\top \boldsymbol{e} &= 0 \\ \boldsymbol{a}^\top(\boldsymbol{b} - \boldsymbol{p}) &= 0 \\ \boldsymbol{a}^\top(\boldsymbol{b} - x\boldsymbol{a}) &= 0 \\ x\boldsymbol{a}^\top \boldsymbol{a} &= \boldsymbol{a}^\top \boldsymbol{b} \\ x &= \frac{\boldsymbol{a}^\top \boldsymbol{b}}{\boldsymbol{a}^\top \boldsymbol{a}} \end{aligned}$$

- Therefore, $\boldsymbol{p} = \boldsymbol{a}x = \boldsymbol{a}\frac{\boldsymbol{a}^\top \boldsymbol{b}}{\boldsymbol{a}^\top \boldsymbol{a}}$
- There is a projection matrix P that $\boldsymbol{p} = P\boldsymbol{b}$.

$$\boxed{P = \frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top \boldsymbol{a}}}$$

Projection onto a Line

$$\boxed{P = \frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}}}$$

- Column space of A: matrix-vector multiplication $A\boldsymbol{x} \in \mathbf{C}(A)$.
- $\boldsymbol{p} = P\boldsymbol{b}$. What is the column space $\mathbf{C}(P)$?
- $\mathbf{C}(P)$ is the line through $\boldsymbol{a}$.
- Is P symmetric?

$$P^\top = \left(\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}}\right)^\top = \frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}} = P. \quad \text{Yes.}$$

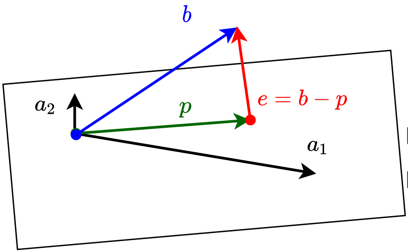
- What if we project $\boldsymbol{b}$ twice?

$$P^2 = \left(\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}}\right)\left(\frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}}\right) = \frac{\boldsymbol{a}\boldsymbol{a}^\top}{\boldsymbol{a}^\top\boldsymbol{a}} = P$$

Projection onto a Subspace

- Why bother with projection?
- Because $A\boldsymbol{x} = \boldsymbol{b}$ may have no solution ($m \gg n$). $\boldsymbol{b}$ might not in the column space $\mathbf{C}(A)$.
- Solve $A\hat{\boldsymbol{x}} = \boldsymbol{p}$ instead, where $\boldsymbol{p}$ is the projection of $\boldsymbol{b}$ onto the column space $\mathbf{C}(A)$.

Projection onto a Subspace



- Choose two independent vectors $\boldsymbol{a}_1,\ \boldsymbol{a}_2$ in the plane to form a basis.

$$A = \begin{bmatrix} | & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 \\ | & | \end{bmatrix}$$

- Plane of $\boldsymbol{a}_1,\ \boldsymbol{a}_2$ = Column space of A.
- $\boldsymbol{p}$ is a linear combination of $\boldsymbol{a}_1,\ \boldsymbol{a}_2$.

$$\begin{aligned} p &= \hat{x}_1\boldsymbol{a}_1 + \hat{x}_2\boldsymbol{a}_2 \\ &= A\hat{\boldsymbol{x}} \end{aligned}$$

- Find $\hat{\boldsymbol{x}}$.

Projection onto a Subspace

- $\boldsymbol{p} = A\hat{\boldsymbol{x}}$. Find $\hat{\boldsymbol{x}}$.
- $\boldsymbol{e} = \boldsymbol{b} - \boldsymbol{p}$ is perpendicular to the plane.

$$\begin{bmatrix} & \boldsymbol{a}_1^\top & \\ & \boldsymbol{a}_2^\top & \end{bmatrix} \begin{bmatrix} | \\ \boldsymbol{e} \\ | \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{aligned} A^\top \boldsymbol{e} &= \boldsymbol{0} \\ A^\top (b - A\hat{\boldsymbol{x}}) &= \boldsymbol{0} \\ A^\top A\hat{\boldsymbol{x}} &= A^\top \boldsymbol{b} \\ \hat{\boldsymbol{x}} &= (A^\top A)^{-1} A^\top \boldsymbol{b} \end{aligned}$$

- We have $\boldsymbol{p} = A\hat{\boldsymbol{x}} = A(A^\top A)^{-1} A^\top \boldsymbol{b}$.
- The projection matrix P:

$$\boxed{P = A(A^\top A)^{-1} A^\top}$$

Projection onto a Subspace

$$\boxed{P = A(A^\top A)^{-1}A^\top}$$

- Is P symmetric?

$$\begin{aligned} P^\top = (A(A^\top A)^{-1}A^\top)^\top &= A((A^\top A)^{-1})^\top A^\top \\ &= A((A^\top A)^\top)^{-1}A^\top \\ &= A(A^\top A)^{-1}A^\top = P \end{aligned}$$

  Yes.
- Is $P^2 = P$?

$$\begin{aligned} P^2 &= A(A^\top A)^{-1}A^\top A(A^\top A)^{-1}A^\top \\ &= A(A^\top A)^{-1}(A^\top A)(A^\top A)^{-1}A^\top \\ &= A(A^\top A)^{-1}A^\top = P \end{aligned}$$

  Yes.

UIT
TRƯỜNG ĐẠI HỌC
CÔNG NGHỆ THÔNG TIN

Q with Orthonormal Columns

$$Q_1 = \frac{1}{3}\begin{bmatrix} 2 \\ 2 \\ -1 \end{bmatrix} \qquad Q_1^\top Q_1 = \begin{bmatrix} 1 \end{bmatrix}$$

$$Q_2 = \frac{1}{3}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix} \qquad Q_2^\top Q_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$Q_3 = \frac{1}{3}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix} \qquad Q_3^\top Q_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

- Columns of Q's are orthonormal.
- Each one of those matrices has $Q^\top Q = I$.
- $Q^\top$ is a left inverse of Q.
- $Q_3 Q_3^\top = I$. $Q_3^\top$ is also a right inverse.

Orthogonal Projection

- All the matrices $P = QQ^\top$ have $P^T = P$.

$$P^\top = (QQ^\top)^\top = QQ^\top = P$$

- All the matrices $P = QQ^\top$ have $P^2 = P$.

$$P^2 = (QQ^\top)(QQ^\top) = Q(Q^\top Q)Q^\top = QQ^\top = P$$

- P is a projection matrix.

Orthogonal Projection

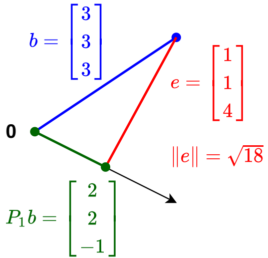
If $P^2 = P = P^\top$ then $P\boldsymbol{b}$ is the orthogonal projection of $\boldsymbol{b}$ onto the column space of P.

TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

Orthogonal Projection

- Project $\boldsymbol{b} = (3,3,3)$ on the Q_1 line. $P_1 = Q_1 Q_1^\top$

$$P_1\boldsymbol{b} = \frac{1}{9}\begin{bmatrix}2\\2\\-1\end{bmatrix}\begin{bmatrix}2 & 2 & -1\end{bmatrix}\begin{bmatrix}3\\3\\3\end{bmatrix} = \frac{1}{9}\begin{bmatrix}2\\2\\-1\end{bmatrix}9 = \begin{bmatrix}2\\2\\-1\end{bmatrix}$$

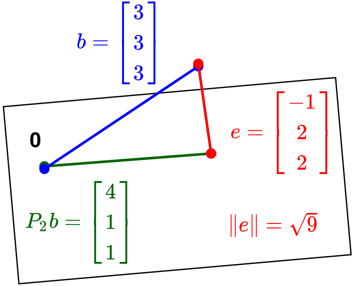


- P_1 splits $\boldsymbol{b}$ in 2 perpendicular parts: projection $P_1\boldsymbol{b}$ and error $\boldsymbol{e} = \boldsymbol{b} - P_1\boldsymbol{b}$

Orthogonal Projection

- Project $\boldsymbol{b} = (3,3,3)$ on the Q_2 plane. $P_2 = Q_2 Q_2^\top$

$$P_2\boldsymbol{b} = \frac{1}{9}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \end{bmatrix}\begin{bmatrix} 3 \\ 3 \\ 3 \end{bmatrix} = \frac{1}{9}\begin{bmatrix} 2 & 2 \\ 2 & -1 \\ -1 & 2 \end{bmatrix}\begin{bmatrix} 9 \\ 9 \end{bmatrix} = \begin{bmatrix} 4 \\ 1 \\ 1 \end{bmatrix}$$



- P_2 projects $\boldsymbol{b}$ on the column space of Q_2.
- The error vector $\boldsymbol{b} - P_2\boldsymbol{b}$ is shorter than $\boldsymbol{b} - P_1\boldsymbol{b}$.

$$Q_3 = \frac{1}{3}\begin{bmatrix} 2 & 2 & -1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix}$$

- What is $P_3\boldsymbol{b} = Q_3Q_3^\top\boldsymbol{b}$?
- Project $\boldsymbol{b}$ onto the whole space $\mathbb{R}^3$.
- $P_3 = Q_3Q_3^\top = I$. Thus, $P_3\boldsymbol{b} = \boldsymbol{b}$. Vector $\boldsymbol{b}$ is in $\mathbb{R}^3$ already.
- The error $\boldsymbol{e}$ is zero!!!

Orthogonalization

- Determine if a list of vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$ is linearly independent.

Gram-Smidth algorithm

given vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_k$
for $i = 1, \ldots, k$

1. Orthogonalization. $\tilde{\boldsymbol{q}}_i = \boldsymbol{a}_i - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{a}_i)\boldsymbol{q}_1 - \ldots - (\boldsymbol{q}_{i-1}^\mathsf{T}\boldsymbol{a}_i)\boldsymbol{q}_{i-1}$
2. Test for linear dependence. If $\tilde{\boldsymbol{q}}_i = 0$, quit.
3. Normalization. $\boldsymbol{q}_i = \tilde{\boldsymbol{q}}_i / \|\tilde{\boldsymbol{q}}_i\|$

- If the vectors are linearly independent, the Gram-Smidth algorithm produces an orthonormal collection of vectors $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$.
- If the vectors $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_{j-1}$ are linearly independent, but $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_j$ are linearly dependent, the algorithm detects this and terminates.

Orthogonalization: Example

$$\boldsymbol{a}_1 = (-1, 1, -1, 1), \quad \boldsymbol{a}_2 = (-1, 3, -1, 3), \quad \boldsymbol{a}_3 = (1, 3, 5, 7)$$

Applying the Gram-Smidth algorithm gives the following results.

- $i = 1$:

$$\tilde{\boldsymbol{q}}_1 = \boldsymbol{a}_1$$
$$\boldsymbol{q}_1 = \frac{1}{\|\tilde{\boldsymbol{q}}_1\|}\tilde{\boldsymbol{q}}_1 = (-1/2, 1/2, -1/2, 1/2)$$

- $i = 2$:

$$\begin{aligned}\tilde{\boldsymbol{q}}_2 &= \boldsymbol{a}_2 - (\boldsymbol{q}_1^\mathsf{T}\boldsymbol{a}_2)\boldsymbol{q}_1 \\ &= (-1, 3, -1, 3) - 4(-1/2, 1/2, -1/2, 1/2) = (1, 1, 1, 1)\end{aligned}$$
$$\boldsymbol{q}_2 = \frac{1}{\|\tilde{\boldsymbol{q}}_2\|}\tilde{\boldsymbol{q}}_2 = (1/2, 1/2, 1/2, 1/2)$$

Orthogonalization: Example

- $i = 3$:

$$\begin{aligned}
\tilde{\boldsymbol{q}}_3 &= \boldsymbol{a}_3 - (\boldsymbol{q}_1^\top \boldsymbol{a}_3)\boldsymbol{q}_1 - (\boldsymbol{q}_2^\top \boldsymbol{a}_3)\boldsymbol{q}_2 \\
&= \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \end{bmatrix} - 2 \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} - 8 \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -2 \\ -2 \\ 2 \\ 2 \end{bmatrix} \\
\boldsymbol{q}_3 &= \frac{1}{\|\tilde{\boldsymbol{q}}_3\|}\tilde{\boldsymbol{q}}_3 = (-1/2, -1/2, 1/2, 1/2)
\end{aligned}$$

- The completion of the Gram-Smidth algorithm without early termination indicates that the vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$ are linearly independent.

QR factorization: $A = QR$

$$A = QR$$

$$\begin{bmatrix} | & | & & | \\ \boldsymbol{a}_1 & \boldsymbol{a}_2 & \dots & \boldsymbol{a}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \boldsymbol{q}_1 & \boldsymbol{q}_2 & \dots & \boldsymbol{q}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & r_{nn} \end{bmatrix}$$

$$r_{kk} = \|\tilde{\boldsymbol{q}}_k\|$$

$$r_{k-1,k} = \boldsymbol{q}_{k-1}^{\mathsf{T}} \boldsymbol{a}_k$$

QR factorization: $A = QR$

$$
\begin{aligned}
\hat{\boldsymbol{x}} &= (A^\mathsf{T} A)^{-1} A^\mathsf{T} \boldsymbol{b} \\
&= ((QR)^\mathsf{T} (QR))^{-1} (QR)^\mathsf{T} \boldsymbol{b} \\
&= (R^\mathsf{T} Q^\mathsf{T} QR)^{-1} R^\mathsf{T} Q^\mathsf{T} \boldsymbol{b} \\
&= (R^\mathsf{T} R)^{-1} R^\mathsf{T} Q^\mathsf{T} \boldsymbol{b} \qquad (\texttt{because } Q^\mathsf{T} Q = I) \\
&= R^{-1} R^{-\mathsf{T}} R^\mathsf{T} Q^\mathsf{T} \boldsymbol{b} \\
&= R^{-1} Q^\mathsf{T} \boldsymbol{b}
\end{aligned}
$$

Solving for $\hat{\boldsymbol{x}}$ by solving $R\hat{\boldsymbol{x}} = Q^\mathsf{T} \boldsymbol{b}$ with back-substitution.