# Gram-Schmidt; Introduction to Determinant

## Lecture 7

Zhang Ce

Department of Electrical and Electronic Engineering
Southern University of Science and Technology

2022.11.15

# Table of Contents

# Table of Contents

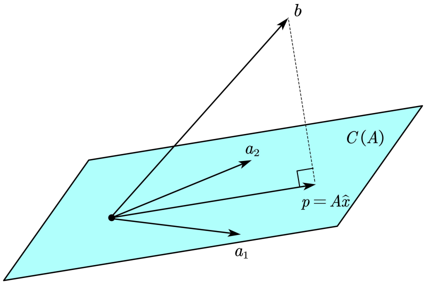
# Last Lecture, We Discuss...

Four parts in last lecture:

1. Orthogonal Vector and Subspaces
   inner product (dot product), definition of orthogonality
2. Projection onto Lines
   geometrical interpretation, projection matrix
3. Least Squares
   geometrical view and general projection matrices
4. Interesting Applications of Linear Algebra
   chemistry, circuit principles and computer vision

# Least-Squares

The core is: find the solution for $Ax = p$ (A is the matrix with $a_1, a_2$ in columns) where $p$ is the projection on $C(A)$, which can definitely gives a unique solution.




$$a_1{}^T (b - A\hat{x}) = a_2{}^T (b - A\hat{x}) = 0$$

The error $e = b - p = b - A\hat{x}$ is in the nullspace of $A$, while the projection $p = A\hat{x}$ is in the column space of $A$!

We can simplify that equation...

$$a_1{}^T(b - A\hat{x}) = a_2{}^T(b - A\hat{x}) = 0$$

$$\begin{bmatrix} a_1{}^T \\ a_2{}^T \end{bmatrix}(b - A\hat{x}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$A^T(b - A\hat{x}) = 0$$

$$A^T A\hat{x} = A^T b$$

Now, can you tell me why $A^T A\hat{x} = A^T b$ must be consistent? The reason behind: $Ax = p$ must have solutions.

Finally, we can get:

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

That is the least-squares solution, indicating how to combine the columns in $A$ can we get the nearest vector to $b$.

# Projection Matrices

The projection vector in the column space, which is also the nearest vector to $b$ in $C(A)$:

$$p = A\hat{x} = A\left(A^TA\right)^{-1}A^Tb = Pb$$

Now we can have projection matrices onto column space of $A$:

$$P = A\left(A^TA\right)^{-1}A^T$$

Can I take the inverse and simplify the expression further?

$$P = A\left(A^TA\right)^{-1}A^T = AA^{-1}(A^T)^{-1}A^T = I???$$

$A$ is not square matrix, $A^{-1}$ doesn't exist.

Recall that we have introduced projection matrices onto lines:

$$P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}}$$

It is just a special case (1-D case) for projection matrices.

Balance the chemical reaction equation:

$$P_4 + CuSO_4 + H_2O \to Cu_3P + H_2SO_4 + H_3PO_4$$

- For element $P$: $4x_1 = x_4 + x_6$.
- For element $Cu$: $x_2 = 3x_4$.
- For element $S$: $x_2 = x_5$.
- For element $O$: $4x_2 + x_3 = 4x_5 + 4x_6$.
- For element $H$: $2x_3 = 2x_5 + 3x_6$.

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

Guess the rank of the matrix, give explanations also.

$$\begin{bmatrix} 4 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -3 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 4 & 1 & 0 & -4 & -4 \\ 0 & 0 & 2 & 0 & -2 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 4 & 0 & 0 & -1 & 0 & -1 \\ 0 & 1 & 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & 12 & -4 & -4 \\ 0 & 0 & 0 & 3 & -1 & 0 \\ 0 & 0 & 0 & 0 & -2 & 5 \end{bmatrix}$$

Free variable $x_6$ set to 2, the solution is:

$$x = \begin{bmatrix} 11/12 & 5 & 8 & 5/3 & 5 & 2 \end{bmatrix}^T$$

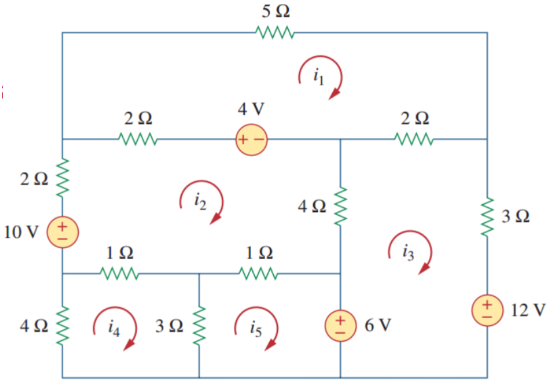
Multiply a constant 12, the solution becomes:

$$x = \begin{bmatrix} 11 & 60 & 96 & 20 & 60 & 24 \end{bmatrix}^T$$

So, the final chemical reaction equation is:

$$11 P_4 + 60 CuSO_4 + 96 H_2O = 20 Cu_3P + 60 H_2SO_4 + 24 H_3PO_4$$

Can you figure out the whole state of the following circuit?




Our method is: find the mesh current $i_1, i_2, i_3, i_4, i_5$.

The essence: Solve $Ax = b$ linear system!

$$\begin{bmatrix} 9 & -2 & -2 & 0 & 0 \\ -2 & 10 & -4 & -1 & -1 \\ -2 & -4 & 9 & 0 & 0 \\ 0 & -1 & 0 & 8 & -3 \\ 0 & -1 & 0 & -3 & 4 \end{bmatrix} \begin{bmatrix} i_1 \\ i_2 \\ i_3 \\ i_4 \\ i_5 \end{bmatrix} = \begin{bmatrix} 4 \\ 6 \\ -6 \\ 0 \\ -6 \end{bmatrix}$$

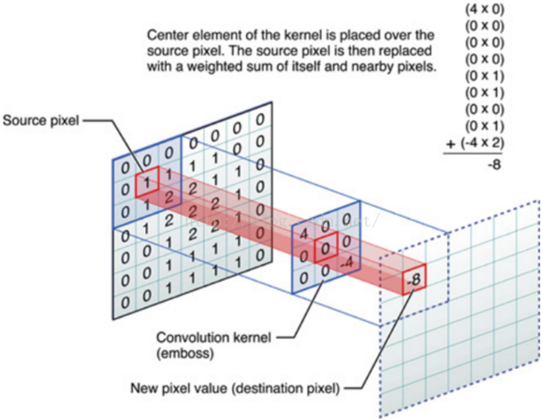
Every row is a KVL equation for a mesh.

If there is no voltage source, then the equation system becomes $Ax = 0$, the mesh currents are definitely all zero (that is to say the matrix is of full rank because all the meshes are independent, no repeat meshes occur)!

The result will be

$$i_1 = 0.283A, i_2 = 0.211A, i_3 = -0.488A, i_4 = -0.718A, i_5 = -1.986A$$

A new operation: convolution.




Now, let's see what will happen for different convolution kernel.

# Linear Algebra in Computer Vision

- Identity Filter:

$$* \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \longrightarrow$$

- Sharpness Filter (enhance contents):

$$* \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \longrightarrow$$

- Another Sharpness Filter (enhance edges):

$$* \begin{bmatrix} 1 & 1 & 1 \\ 1 & -7 & 1 \\ 1 & 1 & 1 \end{bmatrix} \longrightarrow$$

- Edge Detection Filter:

$$* \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \longrightarrow$$

# Linear Algebra in Computer Vision

- Average Box Filter:

$$* \ \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \longrightarrow$$

- Gauss Smoothing Filter:

$$* \ \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix} \longrightarrow$$

Gauss Smoothing Filter for many times:

*Original*    2 *times Smoothing*    5 *times Smoothing*    10 *times Smoothing*

Every time when you use the mobile phone, if you try to pull down the notification badge, Gauss Smoothing Filter is applied. That is the magic from Linear Algebra!

# Table of Contents

# We Will Learn...

Now, we step into a new era: we only consider square matrices now! Just as we do in Chapter 1.

- 3.4 Orthogonal Bases and Gram-Schmidt
- 4.1-4.2 Introduction and Properties of the determinants
- 4.3-4.4 Formulas of the Determinant and Applications
- 5.1 Introduction to Eigenvalues and Eigenvectors
- 5.2 Diagonalization of Matrix
- 5.5 Complex Matrices
- 5.6 Similarity transformation
- 6.1 Minima, Maxima and Saddle Points
- 6.2 Tests for Positive Definiteness
- 6.3 Singular Value Decomposition
- 6.4 Minimum Principles

More difficult! But if you are an expert in Chapter 2, those knowledge would not be very complex.

# Table of Contents

# Orthonormal Vectors

## Definition

The vectors $q_1, q_2, \cdots, q_n$ are orthonormal if

$$q_i{}^T q_j = \begin{cases} 0, \textit{if } i \neq j \\ 1, \textit{if } i = j \end{cases}$$

What are the properties of orthonormal vector set?

- Any 2 vectors are orthogonal.
- Every vector has a length of 1.

Well, give me some orthogonal vector set in $\mathbb{R}^2$.

(You may want to say the standard basis, that's true definitely, can you find some others?)

# Orthogonal Matrices

If you write the orthonormal vectors in the columns of a matrix (kind of familiar?) to form $Q$, any properties?

$Qx = 0$ has only the zero solution! That is because the column vectors are independent (orthogonal is the maximum independent).

How can I express the property of orthonormal vector set in matrix language (Think about $Q^TQ$)?

$$\begin{bmatrix} - & - & {q_1}^T & - & - \\ - & - & {q_2}^T & - & - \\ & & \vdots & & \\ & & \vdots & & \\ - & - & {q_m}^T & - & - \end{bmatrix}_{m\times n} \begin{bmatrix} | & | & & & | \\ | & | & & & | \\ q_1 & q_2 & \cdots & \cdots & q_m \\ | & | & & & | \\ | & | & & & | \end{bmatrix}_{n\times m} = I_{m\times m}$$

Are $m, n$ need to be equal? No, but if they are equal, orthogonal matrix.

For orthogonal matrices, $Q^T = Q^{-1}$. (familiar?)

# Orthogonal Matrices

Why we create $Q$ (not necessarily square)? How can it be related to the previous knowledge in this chapter?

We have learnt the matrix representation of projection onto $C(A)$:

$$P = A(A^T A)^{-1} A^T$$

If we write down the matrix representation of projection onto $C(Q)$:

$$P = Q(Q^T Q)^{-1} Q^T = QQ^T$$

Recall 2 properties for projection matrices. Check if $QQ^T$ satisfies them?

In Chapter 3.3 least-squares, if we change $A$ to $Q$, what will we get?

$$A^T A\hat{x} = A^T b \Rightarrow \hat{x} = Q^T b \Rightarrow \hat{x}_i = {q_i}^T b$$

Which shows us the projection of $b$ on the $i$-th basis of $C(Q)$ is just the inner product of $q_i$ and $b$, that is correct because $||q_i|| = 1$.

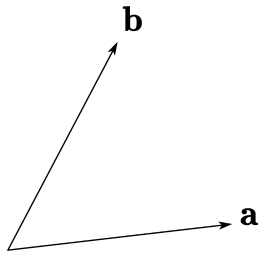
# Table of Contents

# Gram-Schmidt

We have known that orthogonal matrices have those "good" properties, a natural thought is: can I generalize and use the properties for all matrices?

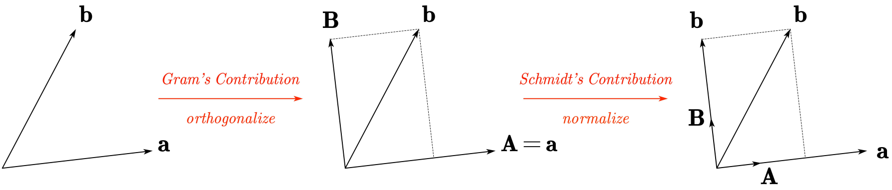
That is the same to ask: can I transform any linearly independent vector set to orthonormal vector set? Gram and Schmidt answer yes!

Everything is difficult if you don't choose an easy example to analyze... How about start with 2 non-orthogonal independent vectors in $\mathbb{R}^2$?



# Gram-Schmidt

*Geometrical Perspective*




*Algebraic Perspective*

$$\begin{bmatrix} | & | \\ | & | \\ \mathbf{a} & \mathbf{b} \\ | & | \\ | & | \end{bmatrix} \xrightarrow[\mathbf{B}=\mathbf{b}-\frac{\mathbf{b}^T\mathbf{a}}{\mathbf{a}^T\mathbf{a}}\mathbf{a}]{\mathbf{A}=\mathbf{a}} \begin{bmatrix} | & | \\ | & | \\ \mathbf{A} & \mathbf{B} \\ | & | \\ | & | \end{bmatrix} \xrightarrow[\mathbf{B}=\frac{\mathbf{B}}{||\mathbf{B}||}]{\mathbf{A}=\frac{\mathbf{A}}{||\mathbf{A}||}} \begin{bmatrix} | & | \\ | & | \\ \mathbf{A} & \mathbf{B} \\ | & | \\ | & | \end{bmatrix}$$

**Algorithm Summary:**

Gram's Part:

- Accept $\mathbf{a}$ to the orthogonal vector set.

$$\mathbf{A} = \mathbf{a}$$

- Subtract $\mathbf{A}$ component from $\mathbf{b}$ and add to the orthogonal vector set.

$$\mathbf{B} = \mathbf{b} - \frac{\mathbf{b}^T\mathbf{A}}{\mathbf{A}^T\mathbf{A}}\mathbf{A}$$

Schmidt's Part:

- Normalize the vectors in orthogonal vector set.

$$\mathbf{A} = \frac{\mathbf{A}}{||\mathbf{A}||}, \mathbf{B} = \frac{\mathbf{B}}{||\mathbf{B}||}$$

Generalize it to more complexed cases, let's have a try.

# Gram-Schmidt

## Example

Do Gram-Schmidt orthogonalization for the vectors

$$\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

**Solution:** Accept **a** to the orthogonal vector set:

$$a_1' = a_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Subtract $a_1'$ component from $a_2$ and add to the orthogonal vector set.

$$a_2' = a_2 - \frac{a_2^T a_1'}{a_1'^T a_1'} a_1' = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

# Gram-Schmidt

Subtract $a_1'$ and $a_2'$ component from $a_3$ and add to the orthogonal vector set.

$$a_3' = a_3 - \frac{a_3^T a_1'}{a_1'^T a_1'} a_1' - \frac{a_3^T a_2'}{a_2'^T a_2'} a_2'$$

$$= \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1/2}{3/2} \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix} = \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

Finally, normalization:

$$q_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, q_2 = \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -1 \\ 2 \end{bmatrix}, q_3 = \frac{1}{\sqrt{3}} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix}$$

# QR Decomposition

Experts in Linear Algebra will not stop here, they will go ahead to find the connection between $Q$ and $A$. That is the same with $LU$ decomposition, we find the matrix $L$ after we know how to find $U$. Now, we are going to find the connection $R$.

$$A = QR = \left(QQ^T\right) A \Rightarrow R = Q^T A$$

$$\begin{bmatrix} | & | & | \\ a & b & c \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} {q_1}^T a & {q_1}^T b & {q_1}^T c \\ {q_2}^T a & {q_2}^T b & {q_2}^T c \\ {q_3}^T a & {q_3}^T b & {q_3}^T c \end{bmatrix}$$

In the Gram-Schmidt process, we can guarantee that ${q_2}^T a = 0$, why?

$$\begin{bmatrix} | & | & | \\ a & b & c \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ q_1 & q_2 & q_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} {q_1}^T a & {q_1}^T b & {q_1}^T c \\ 0 & {q_2}^T b & {q_2}^T c \\ 0 & 0 & {q_3}^T c \end{bmatrix}$$

$R$ is upper triangular! QR decomposition complete. A little bit complex...

# Table of Contents

# Introduction

At first, I don't want to show you complex computations. A question is: determinant expresses what properties of matrix? Recall that every matrix can be recognized as a linear transformation.

Follow my opinion, consider the following matrix:

$$A = \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix}$$

Imaging the linear transformation process of this matrix, it is a linear transformation that stretches the 2-D plane on x and y directions! Actually, as you may know, $det(A) = 8$, which reflects the stretching extent of linear transformation. To say in another way, if we trace an square or any kind of shape during linear transformation, the area will be multiplied by 8 after linear transformation.

Video(3B1B): https://www.bilibili.com/video/BV1ys411472E?p=7

# Properties of Determinant

Now, let's consider the following properties for determinants geometrically.

- $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1, \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$
- $\begin{vmatrix} ta & b \\ tc & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}$
- Linearly dependent columns make the determinant 0
- $\det U = \begin{vmatrix} d_1 & * & * & \cdots & * \\ 0 & d_2 & * & \cdots & * \\ 0 & 0 & d_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & * \\ 0 & 0 & 0 & 0 & d_n \end{vmatrix} = d_1 d_2 d_3 \cdots d_n$
- $\det AB = (\det A)(\det B)$

We know all of them without any kinds of computations!

# Computation of Determinant

Important formula (don't ask why, just remember):

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \cdots \textit{too long}$$

And cofactor expansion (by row):

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$

You can also expand by column. Be careful with those signs.

We will get deeper in next lecture. Please understand the geometrical interpretation, that's enough for this lecture!

# Table of Contents

# Properties of Determinants

The order of these properties come from MIT 18.06 (Gilbert).

1. $\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1$
2. $\begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1$, exchange 2 rows reverse the sign
3. $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t \begin{vmatrix} a & b \\ c & d \end{vmatrix}, \begin{vmatrix} a+a' & b+b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}$
4. 2 equal rows $\rightarrow$ 0 determinant, easily proved by property 2
5. Subtract k times row m from row n will not change the determinant
6. Zero row $\rightarrow$ 0 determinant

The order of these properties come from MIT 18.06 (Gilbert).

7. $\det U = \begin{vmatrix} d_1 & * & * & \cdots & * \\ 0 & d_2 & * & \cdots & * \\ 0 & 0 & d_3 & \cdots & * \\ \vdots & \vdots & \vdots & \ddots & * \\ 0 & 0 & 0 & 0 & d_n \end{vmatrix} = d_1 d_2 d_3 \cdots d_n$
8. Zero det means singular, nonzero det means invertible
9. $\det AB = (\det A)(\det B)$
10. $\det A^T = \det A$

Permutations can be classified to odd and even! That is the same as multiplying a permutation matrix, and permutation matrices have -1 or 1 determinant. Odd row exchanges reverse the sign, while even row exchanges do not change the sign.