Linear Algebra 18
Orthogonality

Wei-Shi Zheng,
wszheng@ieee.org, 2011

1 What Do You Learn from This Note

We still observe the unit vectors we have introduced in Chapter 1:

$$\vec{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \ \vec{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \ \vec{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{1}$$

Question: Have you tried to compute the inner product like $\vec{e}_1 \cdot \vec{e}_2$, $\vec{e}_1 \cdot \vec{e}_3$ and $\vec{e}_2 \cdot \vec{e}_3$?

You will actually find that $\vec{e}_1 \cdot \vec{e}_2 = 0$, $\vec{e}_1 \cdot \vec{e}_3 = 0$ and $\vec{e}_2 \cdot \vec{e}_3 = 0$. That is the three standard basis vectors are orthogonal, more precisely orthonormal. We have a special name for such a type of basis called ***orthogonal (orthonormal) basis***. We will comprehensively introduce this in this lecture note.

Basic Concept: orthogonal set (正交集), orthogonal basis (正交基), Orthogonal Matrix (正交矩阵), Orthogonal Projection (正交投影), Gram–Schmidt Process (格拉姆-施密特正交化)

2 Orthogonal Basis

2.1 Definition

Definition 1 (ORTHOGONAL SET (正交集))**.** Let $S = \{\vec{v}_1, \ldots, \vec{v}_r\} \subset \mathbb{R}^n - \{\vec{0}\}$. We say S is an orthogonal set iff for any $i, j = 1, \ldots, r$ and $i \neq j$, we have $\vec{v}_i \perp \vec{v}_j$. Furthermore, if $\vec{v}_1, \ldots, \vec{v}_r$ are all unit vectors then S is called an **orthonormal set**.

Example: Textbook P.384.

Definition 2 (ORTHOGONAL BASIS (正交基))**.** A basis $\mathcal{B}$ of $\mathbb{R}^n$ which is also orthogonal is called an orthogonal basis. Furthermore, $\mathcal{B}$ is called an orthonormal basis if it is orthonormal.

Example: The standard basis $\mathcal{E} = \{\vec{e}_1, \ldots, \vec{e}_n\}$ is an orthonormal basis of $\mathbb{R}^n$.

Example: Textbook P.389.

2.2 Some Properties

Connection between orthogonal set & linearly independent:

Theorem 3. Let $S = \{\vec{v}_1, \ldots, \vec{v}_r\}$ be an orthogonal set. Then S is linearly independent.

Proof. Let $c_1, \ldots, c_r$ be scalars. Suppose that $\vec{0} = c_1\vec{v}_1 + \cdots + c_r\vec{v}_r$. Then for each $i = 1, \ldots, r$, we have

$$0 = \vec{v}_i \cdot \vec{0} = \vec{v}_i \cdot (c_1\vec{v}_1 + \cdots + c_r\vec{v}_r) = c_1(\vec{v}_i \cdot \vec{v}_1) + \cdots + c_r(\vec{v}_i \cdot \vec{v}_1) = c_i(\vec{v}_i \cdot \vec{v}_i).$$

Since $\vec{v}_i \cdot \vec{v}_i \neq 0$, we must have $c_i = 0$ and the result follows. □

Orthogonal Matrix (正交矩阵):

Theorem 4. Let $U = (\vec{v}_1 \ \cdots \ \vec{v}_r) \in \mathbb{R}^{n\times r}$. Then
1. $\{\vec{v}_1, \ldots, \vec{v}_r\}$ is orthogonal iff U^TU is invertible and diagonal;
2. $\{\vec{v}_1, \ldots, \vec{v}_r\}$ is orthonormal iff $U^TU = I_r$.

Proof. 1.

$$\begin{aligned} & \{\vec{v}_1, \ldots, \vec{v}_r\} \text{ is orthogonal.} \\ \iff & \left[U^TU\right]_{ij} = \vec{v}_i^T\vec{v}_j = \vec{v}_i \cdot \vec{v}_j = \begin{cases} \|\vec{v}_i\| \neq 0 & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases} \\ \iff & U^TU \text{ is invertible and diagonal.} \end{aligned}$$

2. Similar to 1.. □

Definition 5 (ORTHOGONAL MATRIX). A matrix $U \in \mathbb{R}^n$ is said to be orthogonal iff $U^T U = I_n$ (or $U^T = U^{-1}$).

Example: For any $\theta \in \mathbb{R}$, $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is orthogonal.

Theorem 6.
1. I_n is orthogonal.
2. If $U \in \mathbb{R}^n$ is orthogonal then so is U^{-1}.
3. If U_1 and $U_2 \in \mathbb{R}^n$ are orthogonal then so is $U_1 U_2$.

(So the set of orthogonal matrices of size n is a group under multiplication.)

Proof. Easy, left as an exercise. □

Why is orthogonal basis useful?

Theorem 7. Let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ be an orthogonal basis of $\mathbb{R}^n$. Then for any $\vec{v} \in \mathbb{R}^n$ we have

$$[\vec{v}]_{\mathcal{B}} = \begin{pmatrix} \frac{(\vec{v} \cdot \vec{b}_1)}{(\vec{b}_1 \cdot \vec{b}_1)} & \cdots & \frac{(\vec{v} \cdot \vec{b}_n)}{(\vec{b}_n \cdot \vec{b}_n)} \end{pmatrix}^T .$$

So if $\mathcal{B}$ is orthonormal, then

$$[\vec{v}]_{\mathcal{B}} = \begin{pmatrix} (\vec{v} \cdot \vec{b}_1) & \cdots & (\vec{v} \cdot \vec{b}_n) \end{pmatrix}^T .$$

Proof. Suppose $[\vec{v}]_{\mathcal{B}} = (\vec{c}_1 \ \cdots \ \vec{c}_n)^T$. Then for $i = 1, \ldots, n$, we have

$$\vec{v} \cdot \vec{b}_i = (\vec{c}_1 \vec{b}_1 + \cdots + \vec{c}_n \vec{b}_n) \cdot \vec{b}_i = \vec{c}_1 (\vec{b}_1 \cdot \vec{b}_i) + \cdots + \vec{c}_n (\vec{b}_n \cdot \vec{b}_i) = \vec{c}_i (\vec{b}_i \cdot \vec{b}_i).$$

So $\vec{c}_i = \frac{(\vec{v} \cdot \vec{b}_i)}{(\vec{b}_i \cdot \vec{b}_i)}$. □

Example: Textbook P.385.

Theorem 8. Let $\mathcal{B} = \{\vec{b}_1, \ldots, \vec{b}_n\}$ be an orthonormal basis of $\mathbb{R}^n$. Then for any $\vec{u}$, $\vec{v} \in \mathbb{R}^n$,
1. $\vec{u} \cdot \vec{v} = [\vec{u}]_{\mathcal{B}} \cdot [\vec{v}]_{\mathcal{B}}$;
2. $\|\vec{v}\| = \|[\vec{v}]_{\mathcal{B}}\|$;
3. $d(\vec{u}, \vec{v}) = d([\vec{u}]_{\mathcal{B}}, [\vec{v}]_{\mathcal{B}})$.
This means the coordinate isomorphism with respect to $\mathcal{B}$ preserves dot product, norm and distance, in other word, the geometric structure of $\mathbb{R}^n$ is preserved.

Proof. 1.

$$\begin{aligned}
[\vec{u}]_{\mathcal{B}} \cdot [\vec{v}]_{\mathcal{B}} &= (\vec{u} \cdot \vec{b}_1)(\vec{v} \cdot \vec{b}_1) + \cdots + (\vec{u} \cdot \vec{b}_n)(\vec{v} \cdot \vec{b}_n) \\
&= \big(\vec{u} \cdot (\vec{v} \cdot \vec{b}_1)\vec{b}_1\big) + \cdots + \big(\vec{u} \cdot (\vec{v} \cdot \vec{b}_n)\vec{b}_n\big) \\
&= \vec{u} \cdot \big((\vec{v} \cdot \vec{b}_1)\vec{b}_1 + \cdots + (\vec{v} \cdot \vec{b}_n)\vec{b}_n\big) \\
&= \vec{u} \cdot \vec{v}
\end{aligned}$$

2. and 3. are derived from 1. directly. □

Theorem 9. $(\vec{u}_1 \ \cdots \ \vec{u}_n)$ *is orthogonal iff* $\{\vec{u}_1, \ldots, \vec{u}_n\}$ *is an orthonormal basis.*

Theorem 10.
Let $\mathcal{B}$ be an orthonormal basis. Then $\mathcal{B}'$ is an orthonormal basis iff $[\mathcal{B}']_{\mathcal{B}}$ is orthogonal.

Proof. Suppose that $\mathcal{B}' = \{b'_1, \ldots, b'_n\}$. Then

$$\begin{aligned}
\mathcal{B}' \text{ is orthonormal.} &\Leftrightarrow \{b'_1, \ldots, b'_n\} \text{ is orthonormal.} \\
&\Leftrightarrow \{[b'_1]_{\mathcal{B}}, \ldots, [b'_n]_{\mathcal{B}}\} \text{ is orthonormal.} \\
& \ [\text{BY THEOREM 8}] \\
&\Leftrightarrow ([b'_1]_{\mathcal{B}} \ \cdots \ [b'_n]_{\mathcal{B}}) \text{ is orthogonal.}
\end{aligned}$$

□

3 Orthogonal Projection (正交投影)

注：下面的要求大家完全掌握证明。

Theorem 11 (The Orthogonal Decomposition Theorem)**.** *Let W be s subspace of $\mathbb{R}^n$. Then each $\vec{y}$ in $\mathbb{R}^n$ can be written uniquely in the form*

$$\vec{y} = \hat{\vec{y}} + \vec{z},$$

where $\hat{\vec{y}}$ is in W and $\vec{z}$ is in $W^{\perp}$. In fact, if $\{\vec{u}_1, \cdots, \vec{u}_p\}$ is any orthogonal basis of W, then

$$\hat{\vec{y}} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{y} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p}\vec{u}_p$$

and

$$\vec{z} = \vec{y} - \hat{\vec{y}}.$$

Proof. Sketch of the proof:

1. Prove $\hat{\vec{y}} \in W$ (Linear combination of all basis in W)
2. Prove $\vec{z} \in W^{\perp}$ ($\vec{z} \perp \vec{u}_i$, for all $\vec{u}_i$)
3. Prove the uniqueness of the decomposition.

□

Definition 12 (orthogonal projection)**.** *We say the orthogonal projection of $\vec{y}$ onto a subspace W is*

$$proj_W \vec{y} = \frac{\vec{y} \cdot \vec{u}_1}{\vec{u}_1 \cdot \vec{u}_1}\vec{u}_1 + \cdots + \frac{\vec{y} \cdot \vec{u}_p}{\vec{u}_p \cdot \vec{u}_p}\vec{u}_p$$

where $\{\vec{u}_1, \cdots, \vec{u}_p\}$ is any orthogonal basis of W.

Theorem 13. *If $\{\vec{u}_1, \cdots, \vec{u}_p\}$ is any orthonormal basis of a subspace W of $\mathbb{R}^n$, then*

$$proj_W \vec{y} = (\vec{y} \cdot \vec{u}_1)\vec{u}_1 + \cdots + (\vec{y} \cdot \vec{u}_p)\vec{u}_p.$$

If $U = [\vec{u}_1, \cdots, \vec{u}_p]$ then

$$proj_W \vec{y} = UU^T \vec{y}.$$

If we replace the subspace W in the above theorem by a special subspace $L = span\{\vec{u}\}$, we can still have:

Theorem 14. *For any vector $\vec{y} \in \mathbb{R}^n$, we can have*

$$proj_L \vec{y} = \frac{\vec{y} \cdot \vec{u}}{\vec{u} \cdot \vec{u}},$$

and $\vec{z}$ is orthogonal to $\vec{u}$ in $\mathbb{R}^n$.

We call $proj_L \vec{y}$ the orthogonal projection of y onto L.

Theorem 15 (The Best Approximation Theorem)**.** *Let W be a subspace of $\mathbb{R}^n$, $\vec{y}$ be any vector in $\mathbb{R}^n$. Then $proj_W \vec{y}$ is the closet point in W to $\vec{y}$, in the sense that*

$$||\vec{y} - proj_W \vec{y}|| \leqslant ||\vec{y} - \vec{v}||$$

for all $\vec{v}$ in W distinct from $\hat{\vec{y}}$.

Proof. Since

$$\vec{y} - \vec{v} = \vec{y} - \text{proj}_W \vec{y} + \text{proj}_W \vec{y} - \vec{v},$$

and $\vec{y} - \text{proj}_W \vec{y} \in W^{\perp}$ and $\text{proj}_W \vec{y} - \vec{v} \in W$, so

$$||\vec{y} - \vec{v}||^2 = ||\vec{y} - \text{proj}_W \vec{y}||^2 + ||\text{proj}_W \vec{y} - \vec{v}||^2.$$

That is

$$||\vec{y} - \vec{v}||^2 \geqslant ||\vec{y} - \text{proj}_W \vec{y}||^2$$

□

4 Where to use the Best Approximation Theorem

Question: We just do nothing when the matrix equation $A\vec{x} = \vec{b}$ has no solution in real scenarios?

NO!!!NO!!!NO!!!! Let $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$. If there is no solution for the matrix equation $A\vec{x} = \vec{b}$, we are still required to find $\vec{x} \in \mathbb{R}^n$ such that $A\vec{x}$ is the best approximation of $\vec{b}$ in some sense. A feasible approach to this problem is to find $\vec{x}$ such that the distance between $\vec{b}$ and $A\vec{x}$ is as small as possible. We formulate this idea as follows:

Definition 16 (THE LEAST–SQUARES SOLUTION (VERSION 1))**.** Let $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$. Then a least–squares solution of $A\vec{x} = \vec{b}$ is an $\hat{\vec{x}} \in \mathbb{R}^n$ such that

$$\|\vec{b} - A\hat{\vec{x}}\| \leqslant \|\vec{b} - A\vec{x}\|$$

for all $\vec{x} \in \mathbb{R}^n$. The distance $d(\vec{b}, A\hat{\vec{x}}) = \|\vec{b} - A\hat{\vec{x}}\|$ is called the **least–squares error** of $A\vec{x} = \vec{b}$.

Define $W = \mathrm{Col}(A)$. Notice that $\vec{w} = A\vec{x}$ for some $\vec{x} \in \mathbb{R}^n$ iff $\vec{w} \in W$ and by THEOREM 15, the closest vector to $\vec{b}$ among all vectors in W is the orthogonal projection $\vec{b}_W$ of $\vec{b}$ onto W, which is unique. So DEFINITION 16 can be recast equivalently as

Definition 17 (THE LEAST–SQUARES SOLUTION (VERSION 2))**.** Let $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$. Then a least–squares solution of $A\vec{x} = \vec{b}$ is an $\hat{\vec{x}} \in \mathbb{R}^n$ such that

$$A\hat{\vec{x}} = \vec{b}_W,$$

where $W = \mathrm{Col}(A)$.

Question: How to compute the least-square solution?

The least–squares solution set of a matrix equation over $\mathbb{R}$ is identified completely as follows:

Theorem 18. Let $A \in \mathbb{R}^{m\times n}$ and $\vec{b} \in \mathbb{R}^m$. Then the least–squares solution set of the equation $A\vec{x} = \vec{b}$ is precisely the solution set of the equation

$$A^T A\vec{x} = A^T\vec{b}.$$

Proof. Define $W = \mathrm{Col}(A)$. Then

$$\begin{aligned} A\hat{\vec{x}} = \vec{b}_W. &\iff \vec{b} - A\hat{\vec{x}} \in W^\perp. \\ &\iff \vec{b} - A\hat{\vec{x}} \in \mathrm{Nul}(A^T) \quad [\textsc{since } W^\perp = \mathrm{Nul}(A^T)]. \\ &\iff A^T(\vec{b} - A\hat{\vec{x}}) = \vec{0}. \\ &\iff A^T A\hat{\vec{x}} = A^T\vec{b}. \end{aligned}$$

So the result follows. □

Examples: Textbook P.411, P.412, P.413.

Remark: If $\{\vec{a}_1, \dots, \vec{a}_n\}$ is an orthogonal set then

$$\vec{b}_W = \frac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}\vec{a}_1 + \cdots + \frac{\vec{b}\cdot\vec{a}_n}{\vec{a}_n\cdot\vec{a}_n}\vec{a}_n.$$

So $\hat{\vec{x}}$ can be written down directly as

$$\hat{\vec{x}} = \begin{pmatrix} \dfrac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1} & \cdots & \dfrac{\vec{b}\cdot\vec{a}_n}{\vec{a}_n\cdot\vec{a}_n} \end{pmatrix}^T.$$

Example: Textbook P.414.

Question: What is the case when A^TA is invertible?

If A^TA is invertible then the matrix equation $A\vec{x} = \vec{b}$ has a unique least–squares solution, namely $(A^TA)^{-1}A^T\vec{b}$. But this is not always the case. The next theorem gives a sufficient and necessary condition for A^TA being invertible.

Lemma 19. Let $A \in \mathbb{R}^{m\times n}$. Then
1. $\mathrm{Nul}(A) = \mathrm{Nul}(A^TA)$;
2. $\mathrm{rank}(A) = \mathrm{rank}(A^TA)$.

Proof. 1. Suppose that $\vec{x} \in \mathrm{Nul}(A)$, namely $A\vec{x} = \vec{0}$. Then $A^TAx = A^T\vec{0} = \vec{0}$. So $\vec{x} \in \mathrm{Nul}(A^TA)$, that is $\mathrm{Nul}(A) \subseteq \mathrm{Nul}(A^TA)$.

Conversely, suppose that $\vec{x} \in \mathrm{Nul}(A^TA)$, namely $A^TAx = \vec{0}$. Then

$$(A\vec{x}) \cdot (A\vec{x}) = (A\vec{x})^T(A\vec{x}) = ||A\vec{x}||^2 = 0,$$

which forces that $A\vec{x} = \vec{0}$. So $\vec{x} \in \mathrm{Nul}(A)$ and $\mathrm{Nul}(A^TA) \subseteq \mathrm{Nul}(A)$.
2. By the RANK & NULLITY THEOREM,

$$\mathrm{rank}(A) = n - \dim \mathrm{Nul}(A) = n - \dim \mathrm{Nul}(A^TA) = \mathrm{rank}(A^TA).$$

□

Theorem 20. Let $A = (\vec{a}_1 \ \cdots \ \vec{a}_n) \in \mathbb{R}^{m\times n}$. Then A^TA is invertible iff $\{\vec{a}_1, \dots, \vec{a}_n\}$ is linearly independent.

Proof. Notice that $A^TA \in \mathbb{R}^n$. Then

$$\begin{aligned} A^TA \text{ is invertible.} &\iff \mathrm{rank}(A^TA) = n. \\ &\iff \mathrm{rank}(A) = n \quad [\text{BY LEMMA } 4]. \\ &\iff \dim \mathrm{Col}(A) = \dim \mathrm{Span}\{\vec{a}_1, \dots, \vec{a}_n\} = n. \\ &\iff \{\vec{a}_1, \dots, \vec{a}_n\} \text{ is linearly independent.} \end{aligned}$$

□

5 Constructing Orthogonal Basis by Gram–Schmidt Process (格拉姆-施密特正交化)

Question: We have seen the usefulness of orthonormal/orthogonal basis. But how can we generate them?

注：这里告诉我们怎样从一组向量集中构造正交基

We shall show this by the means of THE GRAM–SCHMIDT PROCESS.

Theorem 21 (THE GRAM–SCHMIDT PROCESS)**.** Let $\{\vec{w}_1, \dots, \vec{w}_r\} \subseteq \mathbb{R}^n$ be a linearly independent set. Then we can construct an orthogonal set $\{\vec{v}_1, \dots, \vec{v}_r\} \subseteq \mathbb{R}^n$ such that

$$\text{Span}\{\vec{v}_1, \dots, \vec{v}_r\} = \text{Span}\{\vec{w}_1, \dots, \vec{w}_r\}.$$

Furthermore, by normalising $\{\vec{v}_1, \dots, \vec{v}_r\}$, we obtain an orthonormal set $\{\vec{u}_1, \dots, \vec{u}_r\} \subseteq \mathbb{R}^n$ such that

$$\text{Span}\{\vec{u}_1, \dots, \vec{u}_r\} = \text{Span}\{\vec{w}_1, \dots, \vec{w}_r\}.$$

Proof. We shall prove the first part of the theorem by induction on r(下面，我们用数学归纳法证明).

1. STEP 1: $r = 1$. Then we can take $\vec{v}_1 = \vec{w}_1$ and the result is trivial.
2. STEP 2: INDUCTIVE HYPOTHESIS. Assume the result holds for $r-1$.
3. STEP 3: INDUCTIVE STEP. Since $\{\vec{w}_1, \dots, \vec{w}_{r-1}\}$ is linearly independent, so by INDUCTIVE HYPOTHESIS, we can construct an orthogonal set $\{\vec{v}_1, \dots, \vec{v}_{r-1}\}$ such that

   $$\text{Span}\{\vec{v}_1, \dots, \vec{v}_{r-1}\} = \text{Span}\{\vec{w}_1, \dots, \vec{w}_{r-1}\}.$$

   Now, $\vec{v}_i \neq \vec{0}$, for all $i = 1, \dots, r-1$, so

   $$\vec{v}_r = \vec{w}_r - \frac{\vec{w}_r \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{w}_r \cdot \vec{v}_{r-1}}{\vec{v}_{r-1} \cdot \vec{v}_{r-1}}\vec{v}_{r-1}$$

   is well–defined. It is obvious that

   $$\text{Span}\{\vec{v}_1, \dots, \vec{v}_r\} = \text{Span}\{\vec{w}_1, \dots, \vec{w}_r\},$$

so $\vec{v}_r \neq \vec{0}$. Also for $i = 1, \ldots, r-1$, we have

$$\vec{v}_r \cdot \vec{v}_i = \vec{w}_r \cdot \vec{v}_i - \frac{\vec{w}_r \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 \cdot \vec{v}_i - \cdots - \frac{\vec{w}_r \cdot \vec{v}_{r-1}}{\vec{v}_{r-1} \cdot \vec{v}_{r-1}}\vec{v}_{r-1} \cdot \vec{v}_i = \vec{w}_r \cdot \vec{v}_i - \frac{\vec{w}_r \cdot \vec{v}_i}{\vec{v}_i \cdot \vec{v}_i}\vec{v}_i \cdot \vec{v}_i = 0.$$

Therefore $\{\vec{v}_1, \ldots, \vec{v}_r\}$ is an orthogonal set.

Finally, for $i = 1, \ldots, r$, define $\vec{u}_i = \frac{\vec{v}_i}{\|\vec{v}_i\|}$. Then $\{\vec{u}_1, \ldots, \vec{u}_r\}$ is orthonormal and

$$\mathrm{Span}\{\vec{u}_1, \ldots, \vec{u}_r\} = \mathrm{Span}\{\vec{w}_1, \ldots, \vec{w}_r\}.$$

□

The Gram–Schmidt Process:
According to the proof of the above theorem, we can write down $\vec{v}_1, \ldots, \vec{v}_r$ and $\vec{u}_1, \ldots, \vec{u}_r$ directly as follows:

$$\begin{aligned}
\vec{v}_1 &= \vec{w}_1, \\
\vec{v}_2 &= \vec{w}_2 - \frac{\vec{w}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1, \\
&\vdots \\
\vec{v}_i &= \vec{w}_i - \frac{\vec{w}_i \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{w}_i \cdot \vec{v}_{i-1}}{\vec{v}_{i-1} \cdot \vec{v}_{i-1}}\vec{v}_{i-1}, \\
&\vdots \\
\vec{v}_r &= \vec{w}_r - \frac{\vec{w}_r \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 - \cdots - \frac{\vec{w}_r \cdot \vec{v}_{r-1}}{\vec{v}_{r-1} \cdot \vec{v}_{r-1}}\vec{v}_{r-1},
\end{aligned}$$

$$\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}\ , \quad \vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|}\ , \quad \ldots\ , \quad \vec{u}_r = \frac{\vec{v}_r}{\|\vec{v}_r\|}.$$

This process of writing down $\vec{v}_1, \ldots, \vec{v}_r$ and $\vec{u}_1, \ldots, \vec{u}_r$ is called the GRAM–SCHMIDT PROCESS, which can be used to creating an orthonormal basis for any subspace of $\mathbb{R}^n$.

Examples: Textbook P.402, P.405.

QR Decomposition: Application of Gram–Schmidt:

An application of Gram–Schmidt is for decompose a matrix A whose columns are linearly independent into the following forms:

$$A = QR,$$

where Q is a matrix whose columns form an orthonormal basis for Col$A$ and R is an upper triangular invertible matrix with positive entries onits diagonal. More specifically, we have the following theorem:

Theorem 22 (QR factorization (QR分解)). Let $A = (\vec{w}_1 \ \cdots \ \vec{w}_r) \in \mathbb{R}^{m\times r}$ and $\operatorname{rank}(A) = r$. Then A can be factorised as $A = QR$, where $Q \in \mathbb{R}^{m\times r}$ such that the columns of Q form an orthonormal basis of Col(Q) and $R \in \mathbb{R}^{r\times r}$ is an upper triangular matrix such that diagonal entries are positive.

Proof. The rank of A is r indicates that $\{\vec{w}_1, \ldots, \vec{w}_r\}$ is linearly independent. So by Theorem 1, we can construct orthogonal set $\{\vec{v}_1, \ldots, \vec{v}_r\}$ and orthonormal set $\{\vec{u}_1, \ldots, \vec{u}_r\}$ such that for $i = 1, \ldots, r$,

$$\vec{w}_i = \frac{\vec{w}_i \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \cdots + \frac{\vec{w}_i \cdot \vec{v}_{i-1}}{\vec{v}_{i-1} \cdot \vec{v}_{i-1}}\vec{v}_{i-1} + \vec{v}_i, \quad \vec{v}_i = \|\vec{v}_i\|\vec{u}_i,$$

That is

$$(\vec{w}_1 \ \cdots \ \vec{w}_r) = (\vec{v}_1 \ \cdots \ \vec{v}_r) \begin{pmatrix} 1 & \frac{\vec{w}_2\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1} & \cdots & \frac{\vec{w}_r\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1} \\ 0 & 1 & \cdots & \frac{\vec{w}_r\cdot\vec{v}_2}{\vec{v}_2\cdot\vec{v}_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix}$$

and

$$(\vec{v}_1 \ \cdots \ \vec{v}_r) = (\vec{u}_1 \ \cdots \ \vec{u}_r)\operatorname{diag}(\|\vec{v}_1\|, \|\vec{v}_2\|, \ldots, \|\vec{v}_r\|).$$

So

$$\begin{aligned} (\vec{w}_1 \ \cdots \ \vec{w}_r) &= (\vec{u}_1 \ \cdots \ \vec{u}_r) \begin{pmatrix} \|\vec{v}_1\| & 0 & \cdots & 0 \\ 0 & \|\vec{v}_2\| & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\vec{v}_r\| \end{pmatrix} \begin{pmatrix} 1 & \frac{\vec{w}_2\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1} & \cdots & \frac{\vec{w}_r\cdot\vec{v}_1}{\vec{v}_1\cdot\vec{v}_1} \\ 0 & 1 & \cdots & \frac{\vec{w}_r\cdot\vec{v}_2}{\vec{v}_2\cdot\vec{v}_2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \\ &= (\vec{u}_1 \ \cdots \ \vec{u}_r) \begin{pmatrix} \|\vec{v}_1\| & \vec{w}_2 \cdot \vec{u}_1 & \cdots & \vec{w}_r \cdot \vec{u}_1 \\ 0 & \|\vec{v}_2\| & \cdots & \vec{w}_r \cdot \vec{u}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\vec{v}_r\| \end{pmatrix} \end{aligned}$$

Take $Q = (\vec{u}_1 \ \cdots \ \vec{u}_r)$ and $R = \begin{pmatrix} \|\vec{v}_1\| & \vec{w}_2 \cdot \vec{u}_1 & \cdots & \vec{w}_r \cdot \vec{u}_1 \\ 0 & \|\vec{v}_2\| & \cdots & \vec{w}_r \cdot \vec{u}_2 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \|\vec{v}_r\| \end{pmatrix}$ and the result follows. $\square$

Example: Textbook P.406.