Northeastern University, Department of Mathematics

MATH G5110: Applied Linear Algebra and Matrix Analysis. (Fall 2020)

• Instructor: **He Wang** Email: **he.wang@northeastern.edu**

§6 Inner product spaces

Contents

1. Inner Product Spaces

Recall that for vectors $\vec{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$ in $\mathbb{R}^n$, the **dot product** of $\vec{u}$ and $\vec{v}$ is

$$\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i v_i = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Similarly the dot product of $\vec{u}$ and $\vec{v}$ in $\mathbb{C}^n$ is $\vec{u} \cdot \vec{v} = \sum_{i=1}^{n} u_i \overline{v_i}$

Theorem 1 (Properties of the dot Product)**.** *For vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and a scalar $c \in \mathbb{R}$, the following hold:*
(1.) $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$.
(2.) $(\vec{u} + \vec{v}) \cdot \vec{w} = \vec{u} \cdot \vec{w} + \vec{v} \cdot \vec{w}$.
(3.) $(c\vec{u}) \cdot \vec{v} = c(\vec{u} \cdot \vec{v}) = \vec{u} \cdot (c\vec{v})$.
(4.) $\vec{u} \cdot \vec{u} \geq 0$, *and* $\vec{u} \cdot \vec{u} = 0$ *if and only if* $\vec{u} = \vec{0}$.

More generally, we can define inner product on a general vector space V over $\mathbb{R}$ as

Definition 2 (Inner Product)**.** Let V be a real vector space. An **inner product** on V is a binary function

$$\langle -, - \rangle : V \times V \to \mathbb{R}$$

such that for vectors $\vec{u}, \vec{v}, \vec{w} \in V$ and a scalar $c \in \mathbb{R}$, the following hold:

(1.) $\langle \vec{u}, \vec{v} \rangle = \langle \vec{v}, \vec{u} \rangle$.
(2.) $\langle \vec{u} + \vec{v}, \vec{w} \rangle = \langle \vec{u}, \vec{w} \rangle + \langle \vec{v}, \vec{w} \rangle$.
(3.) $\langle c\vec{u}, \vec{v} \rangle = c\langle \vec{u}, \vec{v} \rangle$.
(4.) $\langle \vec{u}, \vec{u} \rangle \geq 0$
(5.) $\langle \vec{u}, \vec{u} \rangle = 0$ if and only if $\vec{u} = \vec{0}$.

We call V an **inner product space**.

Remark: For complex number field $\mathbb{C}$, item (1) is conjugate symmetry $\langle \vec{u}, \vec{v} \rangle = \overline{\langle \vec{v}, \vec{u} \rangle}$.

Over $\mathbb{R}$, by symmetry, $\langle \vec{u}, c\vec{v} \rangle = c\langle \vec{u}, \vec{v} \rangle$.

Over $\mathbb{C}$, by conjugate symmetry, $\langle \vec{u}, c\vec{v} \rangle = \overline{\langle c\vec{v}, \vec{u} \rangle} = \overline{c\langle \vec{v}, \vec{u} \rangle} = \bar{c}\overline{\langle \vec{v}, \vec{u} \rangle} = \bar{c}\langle \vec{u}, \vec{v} \rangle$.

Example 3. (Weighted dot products) Let $c_1, ..., c_n$ be positive numbers. The weighted inner product on $\mathbb{F}^n$ is

$$\langle \vec{v}, \vec{w} \rangle := \sum_{i=1}^{n} c_i v_i \overline{w_i}$$

Check that it satisfies all axioms.

On $\mathbb{R}^n$, we don't have to put conjugate.

Example 4. Let $P_n(\mathbb{F})$ be the vector space of polynomials of degree at most n with coefficient in $\mathbb{F}$.

An inner product on $P_n(\mathbb{R})$ can be defined as

$$\langle p, q \rangle = \int_0^1 p(t)q(t)\ dt$$

or as $\langle p, q \rangle = \displaystyle\int_0^1 p(t)\overline{q(t)}\ dt$ on $P_n(\mathbb{C})$.

In a inner product space, we have geometry and more tools to work with. Most properties of inner product space are similar as dot products.

Definition 5. Two vectors $\vec{u}$ and $\vec{v}$ are called **orthogonal** if $\langle \vec{v}, \vec{u} \rangle = 0$.

2. **Norms**

Definition 6 (Norm of a Vector)**.** Let V be a inner product space over $\mathbb{F}$. The **length** or **norm** of a vector $\vec{v} \in V$, denoted by $||\vec{v}||$, is defined as

$$||\vec{v}|| = \sqrt{\langle \vec{v}, \vec{v} \rangle}$$

where $v_1, \ldots, v_n$ are the coordinates of $\vec{v}$.

$||\vec{v}|| = 0$ if and only if $\vec{v} = 0$.

A vector $\vec{u}$ is called an **unit vector** if $||\vec{u}|| = 1$.

If a vector $\vec{w}$ is not an unit vector, we can find a unit vector on the same direction defined by

$$\frac{\vec{w}}{||\vec{w}||}$$

and called the **normalization** of $\vec{w}$.

Proposition 7. *For any vector $\vec{v} \in V$ and any scalar $c \in \mathbb{F}$ one obtains*

$$||c \cdot \vec{v}|| = |c| \cdot ||\vec{v}||.$$

Proof. The proof is the same for $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$.

$$||c\vec{v}||^2 = \langle c\vec{v}, c\vec{v}\rangle = c\bar{c}\langle \vec{v}, \vec{v}\rangle = |c|^2||\vec{v}||.$$

□

Theorem 8 (Pythagorean Theorem)**.** *If two vectors $\vec{u}, \vec{v} \in V$ are orthogonal, then they satisfy the* ***Pythagorean Relation***

$$||\vec{u} + \vec{v}||^2 = ||\vec{u}||^2 + ||\vec{v}||^2.$$

Proof. Proof: Two vectors $\vec{u}, \vec{v} \in V$ are orthogonal if and only if $\langle \vec{u}, \vec{v}\rangle = 0$.

$$||\vec{u} + \vec{v}||^2 = \langle \vec{u} + \vec{v}, \vec{u} + \vec{v}\rangle = \langle \vec{u}, \vec{u}\rangle + \langle \vec{u}, \vec{v}\rangle + \langle \vec{v}, \vec{u}\rangle + \langle \vec{v}, \vec{v}\rangle = ||\vec{u}||^2 + ||\vec{v}||^2$$

□

Definition 9. Let $L = \text{Span}\{\vec{w}\}$ be the subspace in V spanned by $\vec{w} \in V$. For a given vector $\vec{y} \in V$, the vector

$$\text{proj}_L(\vec{y}) := \left(\frac{\langle \vec{y}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}\right) \vec{w}$$

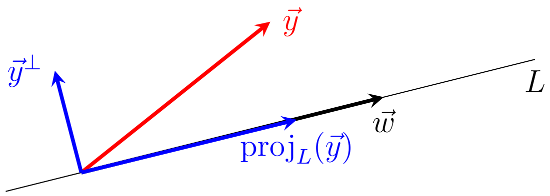
is called the **orthogonal projection of $\vec{y}$ onto** L (or onto $\vec{w}$) and

$$\vec{y}^{\perp} := \vec{y} - \text{proj}_L(\vec{y})$$

the component of $\vec{y}$ orthogonal to L (or $\vec{w}$).

Proposition 10. *Let $\vec{w}$ be a nonzero vector in V. Any vector $\vec{y} \in V$ can be uniquely written as the sum of a scalar product of $\vec{w} \in V$ and a vector orthogonal to $\vec{w}$.*

Proof. $\vec{y} = \text{proj}_L(\vec{y}) + \vec{y}^{\perp}$. Suppose there is another decomposition $\vec{y} = \vec{a} + \vec{b}$ such that $\vec{a} = c\vec{w}$ and $\vec{b}$ is orthogonal to $\vec{w}$. Then $\langle \vec{y}, \vec{w}\rangle = \langle \vec{a} + \vec{b}, \vec{w}\rangle = \langle c\vec{w}, \vec{w}\rangle + \langle \vec{b}, \vec{w}\rangle = c\langle \vec{w}, \vec{w}\rangle$. Hence $c = \dfrac{\langle \vec{y}, \vec{w}\rangle}{\langle \vec{w}, \vec{w}\rangle}$ □



Theorem 11 (**Cauchy-Schwarz inequality**).

$$|\langle \vec{x}, \vec{y} \rangle| \leq ||\vec{x}|| \cdot ||\vec{y}||$$

The equality holds if and only if $\vec{y} = c\vec{x}$.

Proof. $\vec{y} = \text{proj}_{\vec{x}}(\vec{y}) + \vec{y}^{\perp}$.
$||\vec{y}||^2 = \langle \vec{y}, \vec{y} \rangle = ||\text{proj}_{\vec{x}}(\vec{y})||^2 + ||\vec{y}^{\perp}||^2 = \dfrac{\langle \vec{y}, \vec{x} \rangle^2}{\langle \vec{x}, \vec{x} \rangle} + ||\vec{y}^{\perp}||^2 \geq \dfrac{\langle \vec{y}, \vec{x} \rangle^2}{\langle \vec{x}, \vec{x} \rangle}$.
The equality holds if and only if $\vec{y}^{\perp} = \vec{0}$. □

In particular, $||\vec{y}|| \geq ||\text{proj}_{\vec{x}}(\vec{y})||$.

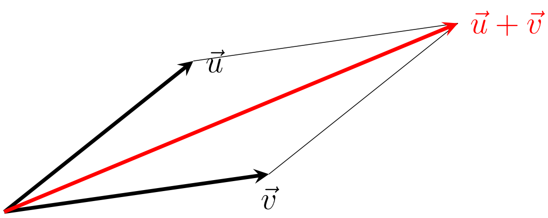
Proposition 12 (Triangle Inequality). *Two vectors* $\vec{u}, \vec{v} \in V$ *satisfy*

$$||\vec{u} + \vec{v}|| \leq ||\vec{u}|| + ||\vec{v}||.$$

Proof.

$$||\vec{u} + \vec{v}||^2 = \langle \vec{u} + \vec{v}, \vec{u} + \vec{v} \rangle = \langle \vec{u}, \vec{u} \rangle + \langle \vec{u}, \vec{v} \rangle + \langle \vec{v}, \vec{u} \rangle + \langle \vec{v}, \vec{v} \rangle = ||\vec{u}||^2 + ||\vec{v}||^2 + 2Re\langle \vec{u}, \vec{v} \rangle$$
$$\leq ||\vec{u}||^2 + ||\vec{v}||^2 + 2|\langle \vec{u}, \vec{v} \rangle| \leq ||\vec{u}||^2 + ||\vec{v}||^2 + 2||\vec{u}|| \cdot ||\vec{v}|| = (||\vec{u}|| + ||\vec{v}||)^2$$

□



Definition 13. (Angles Between Vectors) The **angle between two nonzero vectors** $\vec{u}, \vec{v} \in V$ is the the angle $0 \leq \theta \leq \pi$ satisfying

$$\langle \vec{u}, \vec{v} \rangle = ||\vec{u}|| \cdot ||\vec{v}|| \cdot \cos\theta.$$

Or we can write

$$\theta = \arccos \frac{\langle \vec{u}, \vec{v} \rangle}{||\vec{u}|| \cdot ||\vec{v}||}.$$

In particular, when $\langle \vec{u}, \vec{v} \rangle = 0$, the angle $\theta = \frac{\pi}{2}$.

A vector space V with norm is called a **normed vector space**. In fact, not every norm is defined by inner product. More generally, one can define normed space by axioms:

Definition 14. A **norm** on V is a map from V to $\mathbb{F}$ such that

(1) $||\vec{x}|| \geq 0$ for all $\vec{x} \in V$. $||\vec{x}|| = 0$ if and only if $\vec{x} = \vec{0}$.
(2) $||c\vec{x}|| = |c| \cdot ||\vec{x}||$ for all $\vec{x} \in V$ and $c \in \mathbb{F}$.
(3) The triangle inequality $||\vec{u} + \vec{v}|| \leq ||\vec{u}|| + ||\vec{v}||$ holds for all vectors in V.

Definition 15 (Distance Between Vectors)**.** The **distance** $\text{dist}(\vec{u}, \vec{v})$ between vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ is defined as

$$\text{dist}(\vec{u}, \vec{v}) = ||\vec{u} - \vec{v}||.$$

Note that $\text{dist}(\vec{u}, \vec{v}) \geq 0$ for any pair of vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, and $\text{dist}(\vec{u}, \vec{v}) = 0$ if and only if $\vec{u} = \vec{v}$. Also, $\text{dist}(\vec{u}, \vec{v}) = \text{dist}(\vec{v}, \vec{u})$.

Example 16. (l^p spaces) Let $1 \leq p < \infty$, it is natural to define l^p norms on $\mathbb{F}^n$

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

When $p = 2$, it is the norm induced by the dot product.

Example 17. (l^∞ spaces) It is natural to define l^∞ norms on $\mathbb{F}^n$

$$||\vec{x}||_\infty = \max_{1 \leq i \leq n} \{|x_i|\}$$

Example 18. (Norms on $\mathbb{F}^{m \times n}$ induced by norms on $\mathbb{F}^n$) Normed matrix vector spaces $\mathbb{F}^{m \times n}$. Using norms on $\mathbb{F}^n$, one can define norms on matrix vector spaces

$$||A|| = \sup\{||A\vec{x}|| \mid \vec{x} \in \mathbb{F}^n \text{ with } ||\vec{x}|| = 1\} = \sup\{\frac{||A\vec{x}||}{||\vec{x}||} \mid \vec{x} \neq \vec{0} \in \mathbb{F}^n\}$$

Example 19. Infinity norm on $\mathbb{F}^{m \times n}$.

$$||A||_\infty = \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$

3. **Orthogonal Projections and Orthonormal Bases**

Definition 20 (Orthogonal Set)**.** A set $\{\vec{u}_1, \ldots, \vec{u}_p\}$ of vectors in a inner vector space V is called **orthogonal** if $\langle \vec{u}_i, \vec{u}_j \rangle = 0$ for any choice of indices $i \neq j$.

Proposition 21.
- *Orthogonal vectors are linear independent.*
- *Orthonormal vectors $\{\vec{u}_1, \ldots, \vec{u}_n\}$ in $\mathbb{R}^n$ form a basis of $\mathbb{R}^n$.*

Proof. Suppose $\{\vec{u}_1, \ldots, \vec{u}_p\}$ is orthogonal and $c_1\vec{u}_1 + \cdots + c_p\vec{u}_p = \vec{0}$. Then $\langle c_i\vec{u}_i, \vec{u}_i \rangle = c_i\langle \vec{u}_i, \vec{u}_i \rangle = 0$ for each $i = 1, 2, ..., p$. So, $c_i = 0$.

□

Definition 22. • An **orthogonal basis** for a subspace W of an inner product space V is any basis for W which is also an orthogonal set.

• If each vector is a **unit** vector in an orthogonal basis, then it is an **orthonormal basis**.

Theorem 23 (Coordinates with respect to an orthogonal basis)**.** *Let $\mathscr{B} = \{\vec{u}_1, \ldots, \vec{u}_p\}$ be an* ***orthogonal*** *basis for a subspace W of an inner product space V, and let $\vec{y}$ be any vector in W. Then*

$$\vec{y} = \left(\frac{\langle \vec{y}, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle}\right) \vec{u}_1 + \cdots + \left(\frac{\langle \vec{y}, \vec{u}_p \rangle}{\langle \vec{u}_p, \vec{u}_p \rangle}\right) \vec{u}_p$$

If $W = \mathbb{R}^n$, then the $\mathscr{B}$-coordinates of $\vec{y}$ are given by:

$$[\vec{y}]_{\mathscr{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \quad \textit{with} \quad c_i = \frac{\langle \vec{y}, \vec{u}_i \rangle}{\langle \vec{u}_i, \vec{u}_i \rangle} = \frac{\langle \vec{y}, \vec{u}_i \rangle}{||\vec{u}_i||^2}$$

In particular, let $\mathscr{B} = \{\vec{u}_1, \ldots, \vec{u}_p\}$ be an ***orthonormal*** *basis for a subspace W of $\mathbb{R}^n$, and let $\vec{y}$ be any vector in W. Then*

$$\vec{y} = \langle \vec{y}, \vec{u}_1 \rangle \vec{u}_1 + \cdots + \langle \vec{y}, \vec{u}_p \rangle \vec{u}_p$$

Proof. Suppose $\vec{y} = c_1\vec{u}_1 + \cdots + c_p\vec{u}_p$. □

Let V be an inner product space.

We have defined orthogonal projection onto a vector in V. We can define the orthogonal projection on to an subspace W of V.

More generally, given a subspace W of V and a vector $\vec{y} \in V$, we can ask if/how one can find a decomposition of $\vec{y}$ as

$$\vec{y} = \text{proj}_W(\vec{y}) + \vec{y}^{\perp}$$

with $\text{proj}_W(\vec{y}) \in W$ (**the orthogonal projection of $\vec{y}$ on to W**) and $\vec{y}^{\perp}$ is the **component of $\vec{y}$ perpendicular to W**.

Theorem 24 (Orthogonal Decomposition)**.** *Let W be any subspace of V and let $\vec{y} \in V$ be any vector. Then there exists a unique decomposition*

$$\vec{y} = \text{proj}_W(\vec{y}) + \vec{y}^{\perp}$$

with $\text{proj}_W(\vec{y}) \in W$ and $\vec{y}^{\perp}$ is perpendicular to W.

Theorem 25 (Orthogonal Decomposition)**.** *If $\{\vec{u}_1, \ldots, \vec{u}_p\}$ is an* orthogonal basis *for W, then*

$$\text{proj}_W(\vec{y}) = \left(\frac{\langle \vec{y}, \vec{u}_1 \rangle}{\langle \vec{u}_1, \vec{u}_1 \rangle}\right) \vec{u}_1 + \cdots + \left(\frac{\langle \vec{y}, \vec{u}_p \rangle}{\langle \vec{u}_p, \vec{u}_p \rangle}\right) \vec{u}_p$$

and $\vec{y}^{\perp} = \vec{y} - \text{proj}_W(\vec{y})$.

Definition 26 (Orthogonal Complements)**.** Given a nonempty **subset** (finite or infinite) W of V, we define its **orthogonal complement** $W^\perp$ (pronounced "W perp") as the set of all vectors $\vec{v} \in V$ such that

$$\langle \vec{v}, \vec{w} \rangle = 0, \qquad \text{for all } \vec{w} \in W.$$

Expressed in set notation:

$$W^\perp = \{\vec{v} \in V \mid \langle \vec{v}, \vec{w} \rangle = 0 \text{ for all } \vec{w} \in W\}$$

Theorem 27. *Let S be a subset of V. Let $W = \text{Span}(S)$, then*

1. $S^\perp = W^\perp$
2. *If $W = \text{Span}(\mathscr{B})$, then $W^\perp = \mathscr{B}^\perp$*
3. *$W^\perp$ is a subspace of V (even when S is not).*
4. $(W^\perp)^\perp = W$.
5. $\dim W + \dim W^\perp = \dim V$.

Theorem 28. *Let W be a subspace of V, then*

$$V = W \oplus W^\perp$$

Let A be an $m \times n$ matrix.

The **row space** of A is Row(A), spanned by the row vectors of A.

The **column space** of A is Col(A), so Col(A) = im(A).

The kernel of A is also called the **null space** of A, denoted Nul(A).

Theorem 29. *Let A be an $m \times n$ matrix, then*

$$(\text{Row } A)^\perp = \ker(A) \quad \textit{and} \quad (\text{im } A)^\perp = \ker A^T.$$

More over,

$$\mathbb{F}^m = \ker A^T \oplus \text{im } A$$

4. Gram-Schmidt process and QR-factorization

The **Gram-Schmidt process** is an algorithm that produces an orthogonal (or orthonormal) basis for any subspace W of V by starting with any basis for W.

Theorem 30 (Gram-Schmidt (Orthogonalize)). *Let W be a subspace of V and let $\vec{b}_1, \cdots, \vec{b}_p$ be a basis for W. Define vectors $\vec{v}_1, \ldots, \vec{v}_p$ as*

$$\begin{aligned}
\vec{v}_1 &= \vec{b}_1 \\
\vec{v}_2 &= \vec{b}_2 - \frac{\langle \vec{b}_2, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 \\
\vec{v}_3 &= \vec{b}_3 - \frac{\langle \vec{b}_3, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{b}_3, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}\vec{v}_2 \\
&\vdots \\
\vec{v}_p &= \vec{b}_p - \frac{\langle \vec{b}_p, \vec{v}_1 \rangle}{\langle \vec{v}_1, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{b}_p, \vec{v}_2 \rangle}{\langle \vec{v}_2, \vec{v}_2 \rangle}\vec{v}_2 - \cdots - \frac{\langle \vec{b}_p, \vec{v}_{p-1} \rangle}{\langle \vec{v}_{p-1}, \vec{v}_{p-1} \rangle}\vec{v}_{p-1}
\end{aligned}$$

Then $\{\vec{v}_1, \ldots, \vec{v}_p\}$ is an orthogonal basis for W and

$$\text{Span}\{\vec{b}_1, \cdots, \vec{b}_k\} = \text{Span}\{\vec{v}_1, \cdots, \vec{v}_k\}$$

for and $k = 1, \ldots, p$.

Theorem 31 (Gram-Schmidt (Normalize)). *If $\{\vec{v}_1, \ldots, \vec{v}_p\}$ is an orthogonal basis for W, then $\{\vec{u}_1, \ldots, \vec{u}_p\}$ is an orthonormal basis for W, where, $\vec{u}_i = \dfrac{\vec{v}_i}{||\vec{v}_i||}$ for $i = 1, \ldots, p$.*

$$\text{Basis} \xrightarrow{\text{orthogonalize}} \text{Orthogonal basis} \xrightarrow{\text{normalize}} \text{Orthonormal basis.}$$

- Note that the formula for computing $\vec{v}_i$ for any $i = 2, 3, \ldots, p$ can be written as

$$\begin{aligned}
\vec{v}_i &= \vec{b}_i - \text{proj}_{\vec{v}_1}(\vec{b}_i) - \text{proj}_{\vec{v}_2}(\vec{b}_i) - \cdots - \text{proj}_{\vec{v}_{i-1}}(\vec{b}_i) \\
&= \vec{b}_i - \text{proj}_{\text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}\}}(\vec{b}_i).
\end{aligned}$$

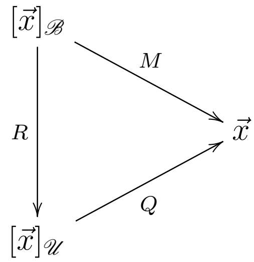
  So, $\vec{v}_i = \vec{b}_i^{\perp}$ respect to $\text{Span}\{\vec{v}_1, \ldots, \vec{v}_{i-1}\}$.
- This formula is *inductive* in that the computation of $\vec{v}_i$ relies on the vectors $\vec{v}_1, \ldots, \vec{v}_{i-1}$.

QR-Factorization.

QR**-Factorization** is the matrix version of Gram-Schmidt process for a subspace W of $\mathbb{F}^n$:

$$\text{Basis } \mathscr{B} = \{\vec{b}_1, \ldots, \vec{b}_p\} \xrightarrow{\text{orthogonalize}} \text{Orthogonal basis } \mathscr{V} = \{\vec{v}_1, \ldots, \vec{v}_p\}$$
$$\xrightarrow{\text{normalize}} \text{Orthonormal basis } \mathscr{U} = \{\vec{u}_1, \ldots, \vec{u}_p\}.$$

Given a vector in W, let's compare their coordinates:



Each matrix defines an isomorphism. So, $M = QR$.

Here $M = [\vec{b}_1 \ \ldots \ \vec{b}_p]$ and $Q = [\vec{u}_1, \ldots, \vec{u}_p]$.

Theorem 32. *Given a $n \times p$ matrix $M = [\vec{b}_1 \ \ldots \ \vec{b}_p]$ with independent columns. There is a unique decomposition*

$$M = QR$$

where, $Q = [\vec{u}_1, \ldots, \vec{u}_p]$ has orthonormal columns and R is an $p \times p$ upper triangular matrix with

$$r_{ii} = ||\vec{v}_i|| \text{ for } i = 1, \ldots, p \text{ and } r_{ij} = \langle \vec{u}_i, \vec{b}_j \rangle \text{ for } i < j.$$

Proof. Proof(for $p = 3$): From Gram-Schmidt process, write $\vec{b}_i$ as linear combinations of $\vec{u}_i$.

$$\begin{aligned}
\vec{b}_1 &= \vec{v}_1 = ||\vec{v}_1||\vec{u}_1 \\
\vec{b}_2 &= \vec{v}_2 + \frac{\langle \vec{b}_2, \vec{v}_1 \rangle}{||\vec{v}_1||^2}\vec{v}_1 = ||\vec{v}_2||\vec{u}_2 + \langle \vec{b}_2, \vec{u}_1 \rangle \vec{u}_1 \\
\vec{b}_3 &= \vec{v}_3 + \frac{\langle \vec{b}_3, \vec{v}_1 \rangle}{||\vec{v}_1||^2}\vec{v}_1 + \frac{\langle \vec{b}_3, \vec{v}_2 \rangle}{||\vec{v}_2||^2}\vec{v}_2 = ||\vec{v}_3||\vec{u}_3 + \langle \vec{b}_3, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{b}_3, \vec{u}_2 \rangle \vec{u}_2
\end{aligned}$$

So,

$$[\vec{b}_1 \ \vec{b}_2 \ \vec{b}_3] = [\vec{u}_1 \ \vec{u}_2 \ \vec{u}_3] \begin{bmatrix} ||\vec{v}_1|| & \langle \vec{u}_1, \vec{b}_2 \rangle & \langle \vec{u}_1, \vec{b}_3 \rangle \\ 0 & ||\vec{v}_2|| & \langle \vec{u}_2, \vec{b}_3 \rangle \\ 0 & 0 & ||\vec{v}_3|| \end{bmatrix}$$

□

5. **Orthogonal Transformations and Orthogonal Matrices**

Let V be a inner product space.

Definition 33. A linear transformation $T : V \to V$ is called **orthogonal** if

$$||T(\vec{x})|| = ||\vec{x}|| \text{ for all } \vec{x} \in V$$

that is, T preserves the length of vectors.

Example 34. Whether or not the following transformations are orthogonal.

(1.) Rotations $S : \mathbb{R}^2 \to \mathbb{R}^2$ are orthogonal transformations.

The matrix of rotation $S = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is orthogonal.

(2.) Reflections $R : \mathbb{R}^2 \to \mathbb{R}^2$ are orthogonal transformations.

The matrix of reflection matrix $R = \begin{bmatrix} a & b \\ b & -a \end{bmatrix}$ with $a^2 + b^2 = 1$ is orthogonal.

(3.) Orthogonal projections $P : \mathbb{R}^2 \to \mathbb{R}^2$ are NOT orthogonal transformations.

The matrix of an orthogonal transformation $T : \mathbb{F}^n \to \mathbb{F}^n$ is called an **orthogonal matrix**.

Theorem 35. *Let U be an $n \times n$ orthogonal matrix and let $\vec{x}$ and $\vec{y}$ be any vectors in $\mathbb{F}^n$. Then*
(1) $||U\vec{x}|| = ||\vec{x}||$.
(2) $\langle U\vec{x}, U\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$.
(3) $\langle U\vec{x}, U\vec{y}\rangle = 0$ *if and only if* $\langle \vec{x}, \vec{y}\rangle = 0$.

Proof. The transformation $T(\vec{x}) = U\vec{x}$ is orthogonal. So, we have 1.
For 2. $||U(\vec{x}+\vec{y})||^2 = ||\vec{x}+\vec{y}||^2 = (\vec{x}+\vec{y}) \cdot (\vec{x}+\vec{y}) = ||\vec{x}||^2 + 2\vec{x}\cdot\vec{y} + ||\vec{y}||^2$
$||U(\vec{x}+\vec{y})||^2 = ||U(\vec{x}) + U(\vec{y})||^2 = ||U(\vec{x})||^2 + ||U(\vec{y})||^2 + 2\langle U\vec{x}, U\vec{y}\rangle$.
Compare two formulas, we have $\langle U\vec{x}, U\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$. □

Proposition 36. *U is an orthogonal matrix if and only if $\langle U\vec{x}, U\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$ for any $\vec{x}$ and $\vec{y}$ in $\mathbb{R}^n$.*

The above theorem says that orthogonal transformations **preserve inner products**, hence also **preserve angles** and orthogonality.

Using the geometric meaning of the orthogonal transformation, we have

Theorem 37. *1. If A is orthogonal, then A is invertible and A^{-1} is orthogonal.*
2. If A and B are orthogonal, then AB is orthogonal.

Theorem 38. *The $n \times n$ matrix U is orthogonal if and only if $\{\vec{u}_1, \ldots, \vec{u}_n\}$ is an orthonormal set.*

Proof. Proof. "⇒" Suppose U is an orthogonal matrix. We prove that $\{\vec{u}_1, \dots, \vec{u}_n\}$ is an orthonormal set.
Use the property that U is orthogonal if and only if $\langle U\vec{x}, U\vec{y}\rangle = \langle \vec{x}, \vec{y}\rangle$. Apply the formula to standard vectors $\vec{x} = \vec{e}_i$ and $\vec{y} = \vec{e}_j$.

$$U\vec{e}_i = [u_1\ u_2\ \cdots\ u_n]\begin{bmatrix}0\\ \vdots\\ 1\\ \vdots\\ 0\end{bmatrix} = \vec{u}_i$$

Hence $\langle \vec{u}_i, \vec{u}_j\rangle = \langle U\vec{e}_i, U\vec{e}_j\rangle = \langle \vec{e}_i, \vec{e}_j\rangle = \begin{cases} 0 & \text{when } i \neq j \\ 1 & \text{when } i = j \end{cases}$

So, $\{\vec{u}_1, \dots, \vec{u}_n\}$ is orthonormal.
"⇐" Suppose $\{\vec{u}_1, \dots, \vec{u}_n\}$ is an orthonormal set. We show that U is an orthogonal matrix.

For any $\vec{x} \in \mathbb{R}^n$, $U\vec{x} = [u_1\ u_2\ \cdots\ u_n]\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} = x_1\vec{u}_1 + x_2\vec{u}_2 + \cdots + x_n\vec{u}_n$

$||U\vec{x}||^2 = \langle (x_1\vec{u}_1 + x_2\vec{u}_2 + \cdots + x_n\vec{u}_n), (x_1\vec{u}_1 + x_2\vec{u}_2 + \cdots + x_n\vec{u}_n)\rangle = x_1^2 + \cdots + x_n^2 = ||\vec{x}||^2$
So, $||U\vec{x}|| = ||\vec{x}||$ and hence U is an orthogonal matrix. □

Recall the transpose of a matrix: Given an $m \times n$ matrix A, we define the ***transpose matrix*** A^T as the $n \times m$ matrix whose (i, j)-th entry is the (j, i)-th entry of A. The dot product can be written as matrix product

$$\vec{v} \cdot \vec{w} = \vec{v}^T\vec{w}$$

Theorem 39. *The $n \times n$ matrix A is orthogonal if and only if $A^TA = I_n$; if and only if $A^{-1} = A^T$.*

Proof. Proof. A is orthogonal if and only if $\{\vec{a}_1, \dots, \vec{a}_n\}$ is orthonormal, i.e., $\vec{a}_i \cdot \vec{a}_j = 1$ if $i \neq j$ and $||\vec{a}_i|| = 1$.
On the other side, (write for the case $n = 3$)

$$A^TA = \begin{bmatrix}\vec{a}_1^T\\ \vec{a}_2^T\\ \vec{a}_3^T\end{bmatrix}\begin{bmatrix}\vec{a}_1 & \vec{a}_2 & \vec{a}_3\end{bmatrix} = \begin{bmatrix}\vec{a}_1^T\vec{a}_1 & \vec{a}_1^T\vec{a}_2 & \vec{a}_1^T\vec{a}_3\\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \vec{a}_2^T\vec{a}_3\\ \vec{a}_3^T\vec{a}_1 & \vec{a}_3^T\vec{a}_2 & \vec{a}_3^T\vec{a}_3\end{bmatrix} = \begin{bmatrix}\vec{a}_1\cdot\vec{a}_1 & \vec{a}_1\cdot\vec{a}_2 & \vec{a}_1\cdot\vec{a}_3\\ \vec{a}_2\cdot\vec{a}_1 & \vec{a}_2\cdot\vec{a}_2 & \vec{a}_2\cdot\vec{a}_3\\ \vec{a}_3\cdot\vec{a}_1 & \vec{a}_3\cdot\vec{a}_2 & \vec{a}_3\cdot\vec{a}_3\end{bmatrix} = I_3$$

□

Theorem 40. *Let W be any subspace of $\mathbb{R}^n$ with an orthonormal basis $\{\vec{u}_1, \dots, \vec{u}_p\}$. Let $U = [\vec{u}_1\ \vec{u}_2 \cdots \vec{u}_p]$. For any $\vec{y} \in \mathbb{R}^n$,*

$$\text{proj}_W(\vec{y}) = UU^T\vec{y}.$$

That is, the ***matrix of the projection*** *onto W is*

$$P = UU^T$$

Remark: 1. $p < n$ since W is a subspace of $\mathbb{R}^n$. When $p = n$, then $P = I_n$.

2. We always have $U^TU = I$ for orthonormal basis $\{\vec{u}_1, \dots, \vec{u}_p\}$.

The theorem comes from the following formula from §5.1. The idea is to translate dot product to matrix product.

$$
\begin{aligned}
\text{proj}_W(\vec{y}) &= (\vec{y}\cdot\vec{u}_1)\vec{u}_1 + (\vec{y}\cdot\vec{u}_2)\vec{u}_2 + \cdots + (\vec{y}\cdot\vec{u}_p)\vec{u}_p \\
&= [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p] \begin{bmatrix} \vec{y}\cdot\vec{u}_1 \\ \vec{y}\cdot\vec{u}_2 \\ \vdots \\ \vec{y}\cdot\vec{u}_p \end{bmatrix} \\
&= [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p] \begin{bmatrix} \vec{u}_1\cdot\vec{y} \\ \vec{u}_2\cdot\vec{y} \\ \vdots \\ \vec{u}_p\cdot\vec{y} \end{bmatrix} \\
&= [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p] \begin{bmatrix} \vec{u}_1^T\vec{y} \\ \vec{u}_2^T\vec{y} \\ \vdots \\ \vec{u}_p^T\vec{y} \end{bmatrix} \\
&= [\vec{u}_1\ \vec{u}_2\ \cdots\ \vec{u}_p] \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_p^T \end{bmatrix} \vec{y} \\
&= UU^T\vec{y}
\end{aligned}
$$

6. The adjoint of a linear operator