Chapter 7
ORTHOGONALITY

- Geometric concepts on *length* and *perpendicularity* of vectors
- Construct a basis of perpendicular eigenvectors for a given matrix or L.T.
- Conditions for the existence of such a basis

7.1 The Geometry of Vectors

- The **norm** (**length**) of any vector $\boldsymbol{v}$ in $\mathcal{R}^n$, denoted by $\|\boldsymbol{v}\|$, is defined by

$$\|\boldsymbol{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}\ .$$

- A vector $\boldsymbol{v}$ with $\|\boldsymbol{v}\|=1$ is called a **unit vector**.
- The **distance** between two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$ is defined as $\|\boldsymbol{u}-\boldsymbol{v}\|$.
- By the Pythagorean theorem, two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^2$ are perpendicular iff

$$\|\boldsymbol{v}-\boldsymbol{u}\|^2 = \|\boldsymbol{u}\|^2 + \|\boldsymbol{v}\|^2$$

$$(v_1-u_1)^2 + (v_2-u_2)^2 = u_1^2 + u_2^2 + v_1^2 + v_2^2$$

$$u_1v_1 + u_2v_2 = 0.$$

- Define the **dot product** of vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$ by

$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

  We say that $\boldsymbol{u}$ and $\boldsymbol{v}$ are **orthogonal** (**perpendicular**) if $\boldsymbol{u} \bullet \boldsymbol{v} = 0$.
  - ✓ The dot product of two vectors is a scalar.
  - ✓ $\boldsymbol{0}$ is orthogonal to every vector in $\mathcal{R}^n$.
- The dot product of $\boldsymbol{u}$ and $\boldsymbol{v}$ can be represented as a matrix product:

$$\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{u}^{\mathrm{T}} \boldsymbol{v}$$

If A is an $m \times n$ matrix, $\boldsymbol{u} \in \mathcal{R}^n$, and $\boldsymbol{v} \in \mathcal{R}^m$, then $A\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{u} \bullet A^{\mathrm{T}} \boldsymbol{v}$.

Theorem 7.1

For any vectors $\boldsymbol{u}$, $\boldsymbol{v}$, and $\boldsymbol{w}$ in $\mathcal{R}^n$ and every scalar c,

a) $\boldsymbol{u} \bullet \boldsymbol{u} = \|\boldsymbol{u}\|^2$.

b) $\boldsymbol{u} \bullet \boldsymbol{u} = 0$ iff $\boldsymbol{u} = \boldsymbol{0}$.

c) $\boldsymbol{u} \bullet \boldsymbol{v} = \boldsymbol{v} \bullet \boldsymbol{u}$.

d) $\boldsymbol{u} \bullet (\boldsymbol{v}+\boldsymbol{w}) = \boldsymbol{u} \bullet \boldsymbol{v} + \boldsymbol{u} \bullet \boldsymbol{w}$.

e) $(\boldsymbol{v}+\boldsymbol{w}) \bullet \boldsymbol{u} = \boldsymbol{v} \bullet \boldsymbol{u} + \boldsymbol{w} \bullet \boldsymbol{u}$.

f) $(c\boldsymbol{u}) \bullet \boldsymbol{v} = c(\boldsymbol{u} \bullet \boldsymbol{v}) = \boldsymbol{u} \bullet (c\boldsymbol{v})$.

g) $\|c\boldsymbol{u}\| = |c|\,\|\boldsymbol{u}\|$.

- By Theorem 7.1(g), any nonzero vector $\boldsymbol{v}$ can be **normalized**, that is, transformed into a unit vector by multiplying it by the scalar $\frac{1}{\|\boldsymbol{v}\|}$.
- Extend (d) and (e) of Theorem 7.1 to linear combinations,

$$\boldsymbol{u}\bullet\sum_{i=1}^{p} c_i\boldsymbol{v}_i = \sum_{i=1}^{p} c_i\boldsymbol{u}\bullet\boldsymbol{v}_i \text{ and}$$

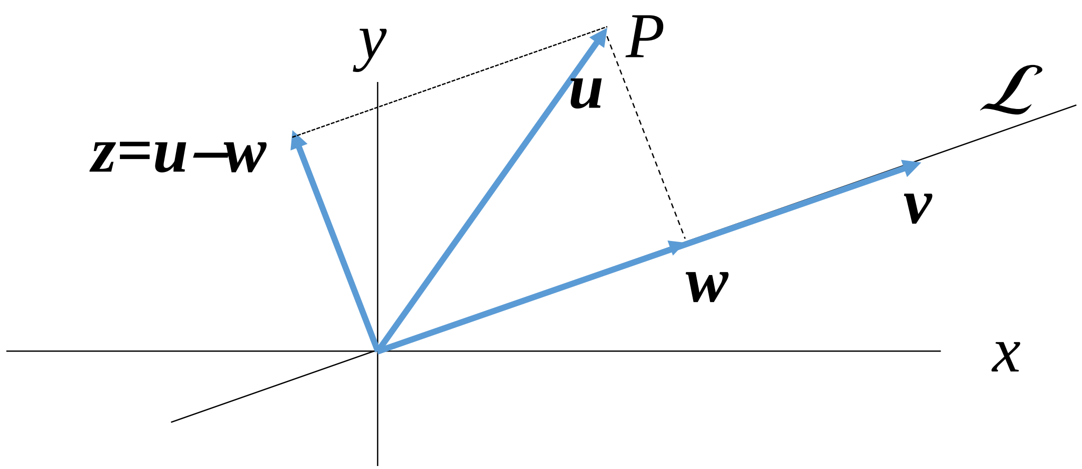
$$\left(\sum_{i=1}^{p} c_i\boldsymbol{v}_i\right)\bullet\boldsymbol{u} = \sum_{i=1}^{p} c_i\boldsymbol{v}_i\bullet\boldsymbol{u}$$

Theorem 7.2 **(Pythagorean Theorem in $\mathcal{R}^n$)**

Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vectors in $\mathcal{R}^n$. Then $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal iff

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

Orthogonal projection of a vector on a line



Find the distance from a point P to the line $\mathcal{L}$:

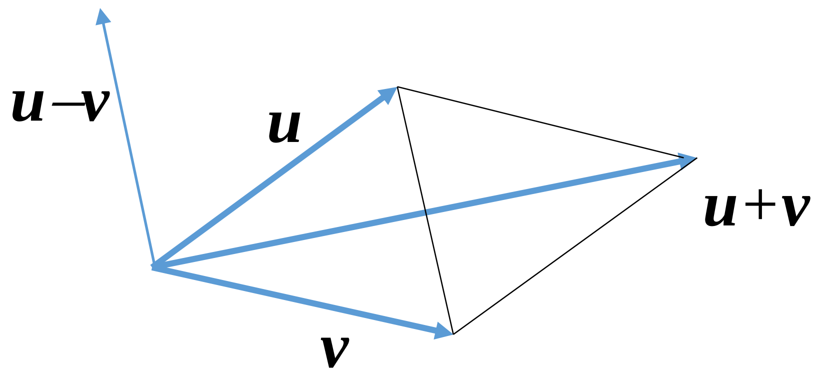
- The vector $\boldsymbol{w}$ is called the **orthogonal projection of *u* on $\mathcal{L}$**.
- To find $\boldsymbol{w}$ in terms of $\boldsymbol{u}$ and $\mathcal{L}$, let $\boldsymbol{v}$ be any nonzero vector along $\mathcal{L}$ and $\boldsymbol{z}=\boldsymbol{u}-\boldsymbol{w}$. Then $\boldsymbol{w}=c\boldsymbol{v}$ for some scalar c. Since $\boldsymbol{z}$ and $\boldsymbol{v}$ are orthogonal,

$$0=\boldsymbol{z}\bullet\boldsymbol{v}=(\boldsymbol{u}-\boldsymbol{w})\bullet v=(\boldsymbol{u}-c\boldsymbol{v})\bullet\boldsymbol{v}=\boldsymbol{u}\bullet\boldsymbol{v}-c\|\boldsymbol{v}\|^2.$$ So

✓ $c=\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}$ and thus $\boldsymbol{w}=\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}$

✓ $\|\boldsymbol{u}-\boldsymbol{w}\|=\left\|\boldsymbol{u}-\frac{\boldsymbol{u}\cdot\boldsymbol{v}}{\|\boldsymbol{v}\|^2}\boldsymbol{v}\right\|.$

The diagonals of a parallelogram are orthogonal iff the parallelogram is a rhombus.

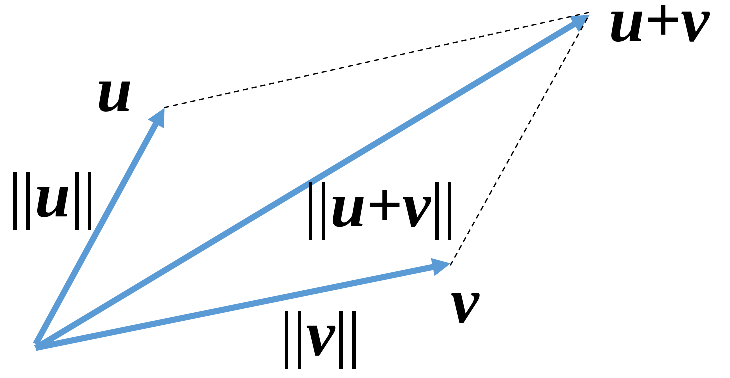


The diagonals of the rhombus are $\boldsymbol{u}+\boldsymbol{v}$ and $\boldsymbol{u}-\boldsymbol{v}$.
The dot product of them is

$$(\boldsymbol{u}+\boldsymbol{v})\bullet(\boldsymbol{u}-\boldsymbol{v})=||\boldsymbol{u}||^2-||\boldsymbol{v}||^2.$$

Thus the diagonals are orthogonal iff $||\boldsymbol{u}||^2=||\boldsymbol{v}||^2$.

The Cauchy-Schwarz and Triangle inequalities



Theorem 7.3 **(Cauchy-Schwarz Inequality)**
For any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$, we have

$$|\boldsymbol{u} \bullet \boldsymbol{v}| \leq \|\boldsymbol{u}\| \, \|\boldsymbol{v}\|.$$

Theorem 7.4 **(Triangle Inequality)**
For any vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$, we have

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

Computing average class size ?

7.2 Orthogonal vectors

A subset of $\mathcal{R}^n$ is an **orthogonal set** if every pair of distinct vectors in the set is orthogonal

- The subset is called **orthonormal set** if it is an orthogonal set consisting entirely of unit vectors.

Theorem 7.5

Any orthogonal set of nonzero vectors is linearly independent.

An orthogonal set that is also a basis for a subspace of $\mathcal{R}^n$ is called an **orthogonal basis** for the subspace.

- A basis that is also an orthonormal set is called an **orthonormal basis.**
- Multiplying vectors in an orthogonal basis by nonzero scalars produces a new orthogonal basis for the same subspace.

Let $\mathcal{S}=\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be an orthogonal basis for a subspace V of $\mathcal{R}^n$. Consider the problem that a vector $\boldsymbol{u}$ in V is to be represented as a linear combination of the vectors in $\mathcal{S}$:

Suppose that $\boldsymbol{u}=c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots+c_k\boldsymbol{v}_k$.

To obtain c_i, we use

$$\begin{aligned}\boldsymbol{u}\bullet\boldsymbol{v}_i&=(c_1\boldsymbol{v}_1+c_2\boldsymbol{v}_2+\cdots c_k\boldsymbol{v}_k)\bullet v_i\\&=c_i\|\boldsymbol{v}_i\|^2\end{aligned}$$

and hence $c_i = \frac{u\cdot v_i}{\|v_i\|^2}$. (It needn't solve a system of linearly equations.)

Representation of a Vector in terms of an Orthogonal or Orthonormal Basis

Let $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ be an orthogonal basis for a subspace V of $\mathcal{R}^n$ and let $\boldsymbol{u}$ be a vector in V. Then

$$\boldsymbol{u} = \frac{\boldsymbol{u}\cdot\boldsymbol{v}_1}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 + \frac{\boldsymbol{u}\cdot\boldsymbol{v}_2}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 + \cdots + \frac{\boldsymbol{u}\cdot\boldsymbol{v}_k}{\|\boldsymbol{v}_k\|^2}\boldsymbol{v}_k.$$

If, in addition, the orthogonal basis is an orthonormal basis for V, then

$$\boldsymbol{u} = (\boldsymbol{u}\cdot\boldsymbol{v}_1)\boldsymbol{v}_1 + (\boldsymbol{u}\cdot\boldsymbol{v}_2)\boldsymbol{v}_2 + \cdots + (\boldsymbol{u}\cdot\boldsymbol{v}_k)\boldsymbol{v}_k.$$

Two questions arise:

1. Does every subspace of $\mathcal{R}^n$ have an orthogonal basis?
2. If the answer is “YES”, how can it be found?

Gram-Schmidt (*orthogonalization*) *process* provides the following info:

Every subspace of $\mathcal{R}^n$ has an orthogonal and hence an orthonormal basis.

Theorem 7.6 (The Gram-Schmidt Process)

Let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_k\}$ be a basis for a subspace W of $\mathcal{R}^n$. Define

$$\boldsymbol{v}_1 = \boldsymbol{u}_1,$$

$$\boldsymbol{v}_2 = \boldsymbol{u}_2 - \frac{\boldsymbol{u}_2 \cdot \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1,$$

$$\boldsymbol{v}_3 = \boldsymbol{u}_3 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_3 \cdot \boldsymbol{v}_2}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2,$$

$$\vdots$$

$$\boldsymbol{v}_k = \boldsymbol{u}_k - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_1}{\|\boldsymbol{v}_1\|^2}\boldsymbol{v}_1 - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_2}{\|\boldsymbol{v}_2\|^2}\boldsymbol{v}_2 - \cdots - \frac{\boldsymbol{u}_k \cdot \boldsymbol{v}_{k-1}}{\|\boldsymbol{v}_{k-1}\|^2}\boldsymbol{v}_{k-1}.$$

Then $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i\}$ is an orthogonal set of nonzero vectors such that

$$\text{Span}\,\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_i\} = \text{Span}\,\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_i\}$$

for each i. So $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an orthogonal basis for W.

The QR Factorization of a Matrix

Suppose that $A=[\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_n]$ is an $m\times n$ matrix with linearly independent columns. Apply the Gram-Schmidt process to vectors $\boldsymbol{a}_i$ to obtain orthogonal vectors $\boldsymbol{v}_i$, $i=1,2,\ \ldots,\ n$. Then normalize these vectors to obtain orthonormal set $\{\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_n\}$.

We may write

$$\boldsymbol{a}_1 = r_{11}\boldsymbol{w}_1$$

$$\boldsymbol{a}_2 = r_{12}\,\boldsymbol{w}_1 + r_{22}\,\boldsymbol{w}_2$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.$$

$$\boldsymbol{a}_n = r_{1n}\,\boldsymbol{w}_1 + r_{2n}\,\boldsymbol{w}_2 + \ldots + r_{nn}\,\boldsymbol{w}_n$$

Define $Q=[\boldsymbol{w}_1\ \ \boldsymbol{w}_2\ \ \ldots\ \ \boldsymbol{w}_n\ ]$ and $R = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ 0 & r_{22} & \ldots & r_{2n} \\ \vdots & 0 & \ldots & r_{3n} \\ 0 & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & r_{nn} \end{bmatrix}$, with

$r_{ij}=\boldsymbol{a}_j \bullet \boldsymbol{w}_i$.

Then $A=Q[\boldsymbol{r}_1\ \ \boldsymbol{r}_2\ \ldots\ \boldsymbol{r}_n]=QR$, where $\boldsymbol{r}_{\mathrm{i}}=[r_{1i}\ \ r_{2i}\ \ r_{ii}\ \ 0\ \ldots.\ 0]^{\mathrm{T}}$.

The QR Factorization of a Matrix

Let A be an $m \times n$ matrix with linearly independent columns. There exists an $m \times n$ matrix Q whose columns form an orthonormal set in $\mathcal{R}^m$, and an $n \times n$ upper triangular matrix R such that $A=QR$. Furthermore, R can be chosen to have positive diagonal entries.

Suppose that A is an $m \times n$ matrix with linearly independent columns. Any factorization $A=QR$, where Q is an $m \times n$ matrix whose columns form an orthonormal set in $\mathcal{R}^m$ and R is an $n \times n$ upper triangular matrix, is called a ***QR* factorization of *A***.

For any QR factorization of A, the columns of Q forms an orthonormal basis for Col A.

Suppose that $A\boldsymbol{x}=\boldsymbol{b}$ is given where A is an $m\times n$ matrix with linearly independent columns. Let $A=QR$ be a QR factorization of A. Using the result $Q^TQ=I_n$, we have the equivalent systems

$$\begin{gathered}
A\boldsymbol{x}=\boldsymbol{b}\\
QR\boldsymbol{x}=\boldsymbol{b}\\
Q^TQR\boldsymbol{x}=Q^T\boldsymbol{b}\\
I_nR\boldsymbol{x}=Q^T\boldsymbol{b}\\
R\boldsymbol{x}=Q^T\boldsymbol{b}
\end{gathered}$$

$Q^TQ=Q^T[\boldsymbol{w}_1\ \ \boldsymbol{w}_2\ \ldots.\ \boldsymbol{w}_{\mathrm{n}}]$

$=[Q^T\boldsymbol{w}_1\ \ Q^T\boldsymbol{w}_2\ \ldots.\ Q^T\boldsymbol{w}_{\mathrm{n}}]$

$=[(\boldsymbol{w}_1^TQ)^T(\boldsymbol{w}_2^TQ)^T\ \ \ldots(\boldsymbol{w}_n^TQ)^T]$

$=[([\boldsymbol{w}_1^T\boldsymbol{w}_1\ \ \boldsymbol{w}_1^T\boldsymbol{w}_2\ \ \ \ldots.\ \ \boldsymbol{w}_1^T\boldsymbol{w}_n])^T\ \ ([\boldsymbol{w}_2^T\boldsymbol{w}_1\ \ \boldsymbol{w}_2^T\boldsymbol{w}_2\ \ \ \ldots.\ \ \boldsymbol{w}_2^T\boldsymbol{w}_n])^T$

$\ldots.\ ([\boldsymbol{w}_n^T\boldsymbol{w}_1\ \ \boldsymbol{w}_n^T\boldsymbol{w}_2\ \ \ \ldots.\ \ \boldsymbol{w}_n^T\boldsymbol{w}_n])^T]$

$=[\boldsymbol{e}_1\ \ \boldsymbol{e}_2\ \ldots.\ \boldsymbol{e}_{\mathrm{n}}]=I_n.$

7.3 Orthogonal Projections

Definition The **orthogonal complement** of a nonempty subset $\mathcal{S}$ of $\mathcal{R}^n$, denoted by $\mathcal{S}^\perp$, is the set of all vectors in $\mathcal{R}^n$ that are orthogonal to every vector in $\mathcal{S}$. That is,

$$\mathcal{S}^\perp = \{\boldsymbol{v} \in \mathcal{R}^n : \boldsymbol{v} \cdot \boldsymbol{u} = 0 \text{ for every } \boldsymbol{u} \text{ in } \mathcal{S}\}.$$

- If $\mathcal{S} = \mathcal{R}^n$, then $\mathcal{S}^\perp = \{\boldsymbol{0}\}$; and if $\mathcal{S} = \{\boldsymbol{0}\}$, then $\mathcal{S}^\perp = \mathcal{R}^n$.
- If $\mathcal{S}$ is any nonempty subset of $\mathcal{R}^n$, then $\boldsymbol{0}$ is in $\mathcal{S}^\perp$.
- Moreover, if $\boldsymbol{v}$ and $\boldsymbol{w}$ are in $\mathcal{S}^\perp$, then, for every vector $\boldsymbol{u} \in \mathcal{S}$, $(\boldsymbol{v}+\boldsymbol{w})\cdot\boldsymbol{u}=\boldsymbol{0}$ and therefore $(\boldsymbol{v}+\boldsymbol{w})$ is in $\mathcal{S}^\perp$.
  - ✓ $\mathcal{S}^\perp$ is closed under vector addition.
  - ✓ By similar arguments, $\mathcal{S}^\perp$ is closed under scalar multiplication.

  So, ..

The orthogonal complement of any nonempty subset of $\mathcal{R}^n$ is a subspace of $\mathcal{R}^n$.

For any nonempty subset $\mathcal{S}$ of $\mathcal{R}^n$, we have $\mathcal{S}^\perp = (\text{Span } \mathcal{S})^\perp$. In particular, the orthogonal complement of a basis for a subspace is the same as the orthogonal complement of the subspace.

For any matrix A, the orthogonal complement of the row space of A is the null space of A; that is,

$$(\text{Row } A)^\perp = \text{Null } A.$$

Applying the above to A^T, we have

$$(\text{Col A})^\perp = (\text{Row } A^T)^\perp = \text{Null } A^T.$$

Theorem 7.7 (Orthogonal Decomposition Theorem) Let W be a subspace of $\mathcal{R}^n$. Then, for any vector $\boldsymbol{u}$ in $\mathcal{R}^n$, there exist unique vectors $\boldsymbol{w}$ in W and $\boldsymbol{z}$ in $W^\perp$ such that $\boldsymbol{u}=\boldsymbol{w}+\boldsymbol{z}$. In addition, if $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_k\}$ is an orthonormal basis for W, then

$$\boldsymbol{w}=(\boldsymbol{u}\bullet\boldsymbol{v}_1)\boldsymbol{v}_1+(\boldsymbol{u}\bullet\boldsymbol{v}_2)\boldsymbol{v}_2+\ldots+(\boldsymbol{u}\bullet\boldsymbol{v}_k)\boldsymbol{v}_k.$$

Combining a basis for W with a basis for $W^\perp$, we have a basis for $\mathcal{R}^n$. Thus,

For any subspace W of $\mathcal{R}^n$,

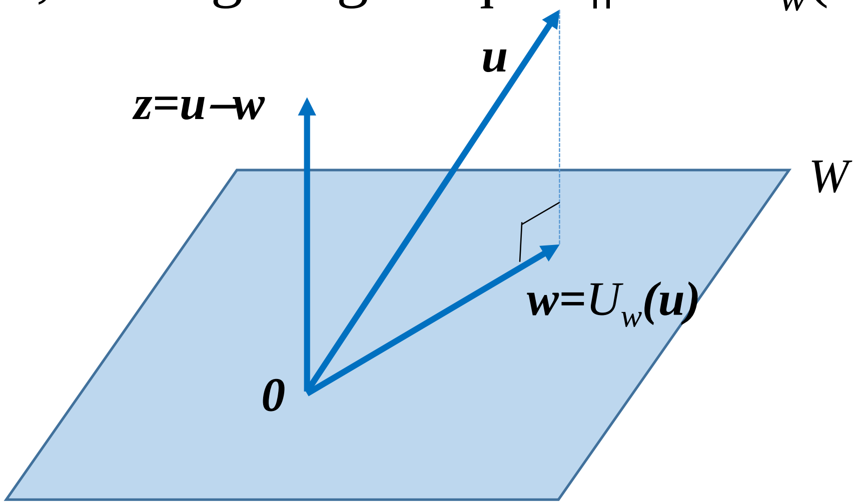
$$\dim W + \dim W^\perp = n.$$

Orthogonal Projections on Subspaces

Definitions Let W be a subspace of $\mathcal{R}^n$ and $\boldsymbol{u}$ be a vector in $\mathcal{R}^n$. The **orthogonal projection of $\boldsymbol{u}$ on $\boldsymbol{W}$** is the unique vector $\boldsymbol{w}$ in W such that $\boldsymbol{u}-\boldsymbol{w}$ is in $W^\perp$.

Furthermore, the function $U_w\colon \mathcal{R}^n \to \mathcal{R}^n$ such that $U_w(\boldsymbol{u})$ is the orthogonal projection of $\boldsymbol{u}$ on W for every $\boldsymbol{u}$ in $\mathcal{R}^n$ is called **the orthogonal projection operator** on W.

For any vector $\boldsymbol{u}$ in $\mathcal{R}^n$ but not in W, the vector $\boldsymbol{u}-U_w(\boldsymbol{u})$ is orthogonal to W, having length equal $\|\boldsymbol{u}-U_w(\boldsymbol{u})\|$.



The vector $\boldsymbol{w}$ is the orthogonal projection of $\boldsymbol{u}$ on W.

Any orthogonal projection U_w of $\mathcal{R}^n$ is *linear*, by showing that

- U_w preserves vector addition.

  Let $\boldsymbol{u_1}, \boldsymbol{u_2} \in \mathcal{R}^n$ and suppose that $U_w(\boldsymbol{u_1})=\boldsymbol{w_1}$ and $U_w(\boldsymbol{u_2})=\boldsymbol{w_2}$.

  Then there are unique vectors $\boldsymbol{z_1}, \boldsymbol{z}_2 \in W^{\perp}$ such that $\boldsymbol{u_1}=\boldsymbol{w_1}+\boldsymbol{z_1}$ and $\boldsymbol{u_2}=\boldsymbol{w_2}+\boldsymbol{z_2}$.

  Thus

  $$\boldsymbol{u_1}+\boldsymbol{u_2}=(\boldsymbol{w_1}+\boldsymbol{w_2})+(\boldsymbol{z_1}+\boldsymbol{z_2})$$

  Since $\boldsymbol{w_1}+\boldsymbol{w_2} \in W$ and $\boldsymbol{z_1}+\boldsymbol{z_2} \in W^{\perp}$,

  $$U_w(\boldsymbol{u_1}+\boldsymbol{u_2})= \boldsymbol{w_1}+\boldsymbol{w_2} =U_w(\boldsymbol{u_1})+U_w(\boldsymbol{u_2}) \ \#$$

- U_w preserves scalar multiplication

  Let c be a scalar.

  It follows that $c\boldsymbol{u_1}=c\boldsymbol{w_1}+c\boldsymbol{z_1}$ where $c\boldsymbol{w_1} \in W$ and $c\boldsymbol{z_1} \in W^{\perp}$.

  Then $U_w(c\boldsymbol{u_1})=c\boldsymbol{w_1}=cU_w(\boldsymbol{u_1})$ #

Linear transformation ⇒ Matrix transformation

The standard matrix of an orthogonal projection operator U_w on a subspace W of $\mathcal{R}^n$ is called the **orthogonal projection matrix** for W and is denoted P_w.

- The columns of P_w are the images of the standard vectors under U_w, that is, the orthogonal projections of the standard vectors. (See example 3 with $\boldsymbol{u}=\boldsymbol{e_i}$)

$$P_w = [U_w(\boldsymbol{e}_1) \quad U_w(\boldsymbol{e}_2) \quad \cdots \quad U_w(\boldsymbol{e}_n)]$$

- An alternative method for computing P_w, by Theorem 7.8.

Lemma Let C be a matrix whose columns are linearly independent. Then C^TC is invertible.

Proof:

Theorem 7.8

Let C be an $n\times k$ matrix whose columns form a basis for a subspace W of $\mathcal{R}^n$. Then

$$P_w = C(C^T C)^{-1} C^T.$$

Proof:

Let W be a subspace of $\mathcal{R}^n$, $\boldsymbol{w} = U_w(\boldsymbol{u})$ and $\boldsymbol{w}'$ be any vector in W.

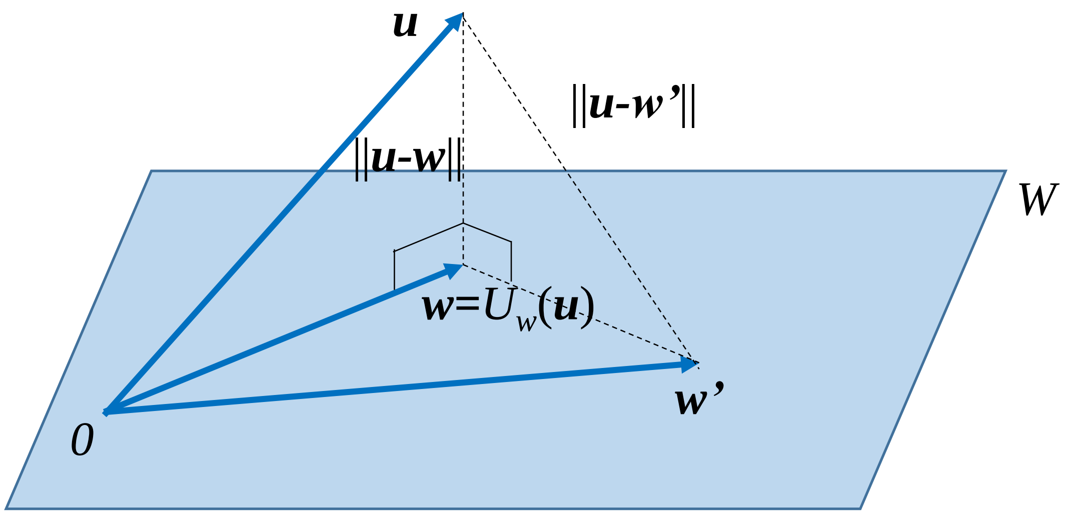
- $(\boldsymbol{u}-\boldsymbol{w})\bullet(\boldsymbol{w}-\boldsymbol{w}')=0$ since $(\boldsymbol{u}-\boldsymbol{w})\in W^\perp$ and $(\boldsymbol{w}-\boldsymbol{w}')\in W$.
- By the Pythagorean theory in R^n,

$$\|\boldsymbol{u}-\boldsymbol{w}'\|^2=\|(\boldsymbol{u}-\boldsymbol{w})+(\boldsymbol{w}-\boldsymbol{w}')\|^2=\|\boldsymbol{u}-\boldsymbol{w}\|^2+\|\boldsymbol{w}-\boldsymbol{w}'\|^2\geq\|\boldsymbol{u}-\boldsymbol{w}\|^2.$$

Closest Vector Property

Let W be a subspace of $\mathcal{R}^n$ and $\boldsymbol{u}$ be a vector in $\mathcal{R}^n$. Among all vectors in W, the vector closest to $\boldsymbol{u}$ is the orthogonal projection $U_w(\boldsymbol{u})$ of $\boldsymbol{u}$ on W.

Define the **distance from a vector $\boldsymbol{u}$ in $\mathcal{R}^n$ to a subspace W of** $\mathcal{R}^n$ to be $\|\boldsymbol{u} - U_w(\boldsymbol{u})\|$, the minimum distance between $\boldsymbol{u}$ and every vector in W.



7.4 Least-squares approximation and orthogonal projection matrices

- Deterministic vs Stochastic/Probabilistic!

- Given a set of data (x_1, y_1), (x_2, y_2), …., (x_n, y_n) obtained by empirical measurements, obtain their relationship by finding the line $y=a_0+a_1x$ that *best fits* the data.

The method of least squares:

Find a_0 and a_1 so that

$$E = [y_1 - (a_0 + a_1x_1)]^2 + [y_2 - (a_0 + a_1x_2)]^2 + \cdots + [y_n - (a_0 + a_1x_n)]^2$$

is minimized.

E is called the **error sum of squares**.

Let $\boldsymbol{v_1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$, $\boldsymbol{v_2} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$, $\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, and C=[$\boldsymbol{v_1}$ $\boldsymbol{v_2}$].

Then,

$$E = \|\boldsymbol{y} - (a_0\boldsymbol{v_1} + a_1\boldsymbol{v_2})\|^2$$

$\sqrt{E}$ is the distance between $\boldsymbol{y}$ and the vector $a_0\boldsymbol{v_1} + a_1\boldsymbol{v_2} \in$ W=Span{$\boldsymbol{v_1}$, $\boldsymbol{v_2}$}

The orthogonal projection of $\boldsymbol{y}$ on W is the vector in W that is nearest to $\boldsymbol{y}$.

$$a_0\boldsymbol{v_1} + a_1\boldsymbol{v_2} = C\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = P_w\boldsymbol{y}.$$

Suppose that $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ are linearly independent !
So $\mathcal{B}$={$\boldsymbol{v_1}$, $\boldsymbol{v_2}$} is a basis for W.

Apply Theorem 7.8 to obtain

$$C\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = C(C^T C)^{-1} C^T \boldsymbol{y}.$$

Multiplying both sides by C^T gives

$$C^T C \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = C^T C(C^T C)^{-1} C^T \boldsymbol{y} = C^T \boldsymbol{y}.$$

- The matrix equation $C^T C\boldsymbol{x} = C^T \boldsymbol{y}$, a system of linear equations, is called the **normal equations**.

By the Lemma preceding Theorem 7, $C^T C$ is invertible. The least-square line has the equation $y=a_0+a_1x$, where

$$\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = (C^T C)^{-1} C^T \boldsymbol{y}.$$

Find the best fit by a quadratic polynomial $y=a_0+a_1x+a_2x^2$:

The error sum of squares is

$$E = [y_1 - (a_0 + a_1x_1 + a_2x_1^2)]^2 + [y_2 - (a_0 + a_1x_2 + a_2x_2^2]^2 + \cdots + [y_n - (a_0 + a_1x_n + a_2\, x_n^2)]^2.$$

In this case,

$$\boldsymbol{v_1}= \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \boldsymbol{v_2} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \boldsymbol{v_3} = \begin{bmatrix} x_1^2 \\ x_2^2 \\ \vdots \\ x_n^2 \end{bmatrix}, \boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$, and $C=[\boldsymbol{v_1}\ \boldsymbol{v_2}\ \boldsymbol{v_3}]$.

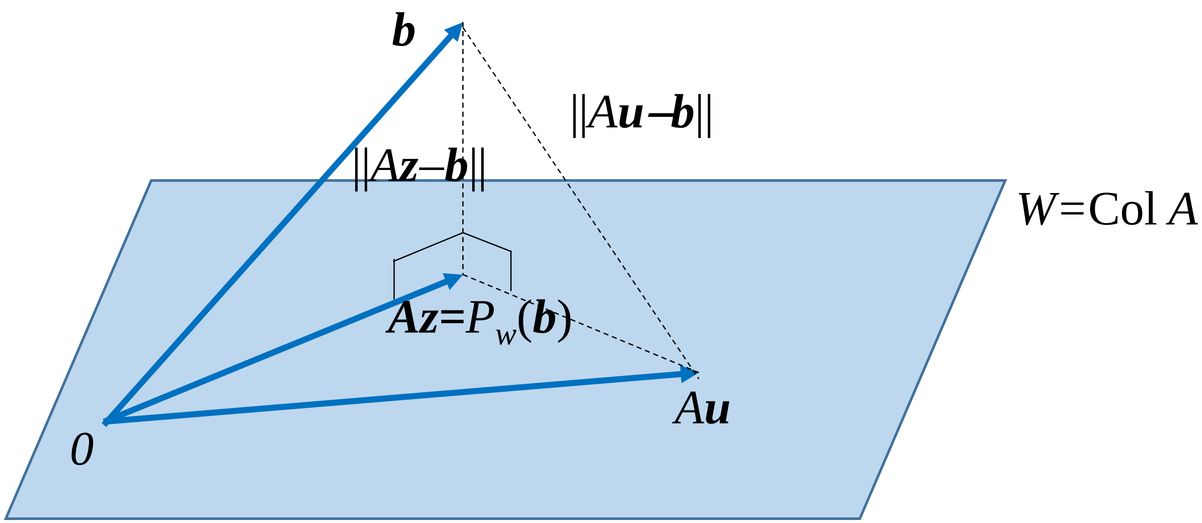
- Assume that $\boldsymbol{v_1}$, $\boldsymbol{v_2}$, and $\boldsymbol{v_3}$ are linearly independent !
- Obtain the normal equations

$$C^TC \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = C^T\boldsymbol{y}$$

The solution is

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = (C^TC)^{-\mathbf{1}}C^T\boldsymbol{y}.$$

Inconsistent systems of Linear Equations



- A system of linear equations $A\boldsymbol{x}=\boldsymbol{b}$ may be inconsistent.
- Obtain a vector $\boldsymbol{z}$ for which $||A\boldsymbol{z}-\boldsymbol{b}||$ is a minimum.
- The vector $\boldsymbol{z}$ minimizes $||A\boldsymbol{z}-\boldsymbol{b}||$ iff it is a solution of the system of linear equation $A\boldsymbol{x}=P_w\boldsymbol{b}$.

Solutions of Least Norm

Given a nonhomogeneous system of linear equations with an infinite set of solutions, select the solution of least norm.

A consistent system $A\boldsymbol{x}=\boldsymbol{c}$ of linear equations with $\boldsymbol{c}\neq\boldsymbol{0}$.
Let $\boldsymbol{v_0}$ be any solution of the system and $Z=$ Null A.

- A vector $\boldsymbol{v}$ is a solution of the system iff it is of the form $\boldsymbol{v}=\boldsymbol{v}_0+\boldsymbol{z}$, where $\boldsymbol{z}\in Z$.
- Select a vector $\boldsymbol{z}$ in Z so that $\|\boldsymbol{v}_0+\boldsymbol{z}\|$ is a minimum:

  Let $\boldsymbol{z}$ be the orthogonal projection of $-\boldsymbol{v}_0$ on Z; that is,

  $\boldsymbol{z} = P_Z(-\boldsymbol{v}_0)= -P_Z\,\boldsymbol{v}_0\,.$

Thus,

$$\boldsymbol{v}=\boldsymbol{v}_0-P_Z\,\boldsymbol{v}_0$$

is the unique solution of the system of least norm.

7.5 Orthogonal Matrices and Operators

Which linear operators T will satisfy $\|T(\boldsymbol{u})\|=\|\boldsymbol{u}\|$ for every $\boldsymbol{u}$ in $\mathcal{R}^n$?

Let Q be an $n\times n$ matrix and $\|Q\boldsymbol{u}\|=\|\boldsymbol{u}\|$ for every $\boldsymbol{u}$ in $\mathcal{R}^n$. $\Rightarrow$

- For the j-th column of Q, denoted by $\boldsymbol{q}_j$,

  $$\|\boldsymbol{q}_j\|=\|Q\boldsymbol{e}_j\|=\|\boldsymbol{e}_j\|=1. \qquad (1)$$

  Thus, the norm of every column of Q is 1.
- If $i\neq j$, we have

  $$\|\boldsymbol{q}_i + \boldsymbol{q}_j\|^2 =\|Q\boldsymbol{e}_i +Q\boldsymbol{e}_j\|^2 =\|Q(\boldsymbol{e}_i+\boldsymbol{e}_j)\|^2 =\|\boldsymbol{e}_i +\boldsymbol{e}_j\|^2=2=\|\boldsymbol{q}_i\|^2+\|\boldsymbol{q}_j\|^2. \qquad (2)$$

By Theorem 7.2, $\boldsymbol{q}_i$ and $\boldsymbol{q}_j$ are orthogonal. The columns of Q thus form an orthonormal basis for $\mathcal{R}^n$.

- ✓ An $n \times n$ matrix is an **orthogonal matrix** (or **orthogonal**) if its columns form an *orthonormal* basis for $\mathcal{R}^n$.
- ✓ A linear operator on $\mathcal{R}^n$ is called an **orthogonal operator** (or **orthogonal**) if its standard matrix is an orthogonal matrix.

Theorem 7.9

The following conditions about an $n \times n$ matrix Q are equivalent:

a) Q is orthogonal. $\leftarrow$ Unitary
b) $Q^T Q = I_n$.
c) Q is invertible and $Q^T = Q^{-1}$.
d) $Q\mathbf{u} \cdot Q\mathbf{v} = \mathbf{u} \cdot \mathbf{v}$ for any $\mathbf{u}$ and $\mathbf{v}$ in $\mathcal{R}^n$. (Q preserves dot products.)
e) $\|Q\mathbf{u}\| = \|\mathbf{u}\|$ for any $\mathbf{u}$ in $\mathcal{R}^n$. (Q preserves dot norms)

Proof ?

- An $n \times n$ matrix Q is orthogonal iff $Q^T = Q^{-1}$ ($Q^T Q = I_n$ or $QQ^T = I_n$).
- The condition $QQ^T = I_n$ is equivalent to the condition that the rows of Q form an orthonormal basis for $\mathcal{R}^n$.

Theorem 7.10

Let P and Q be $n \times n$ orthogonal matrices.

a) $\det Q = \pm 1$.

b) PQ is an orthogonal matrix.

c) Q^{-1} is an orthogonal matrix.

d) Q^T is an orthogonal matrix

Proof ?

A linear operator is orthogonal iff its standard matrix is orthogonal: $\Rightarrow$

If T is a linear operator on $\mathcal{R}^n$, then the following statements are equivalent:

a) T is an orthogonal operator.

b) $T(\boldsymbol{u}) \cdot T(\boldsymbol{v}) = \boldsymbol{u} \cdot \boldsymbol{v}$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$. (T preserves dot products.)

c) $\|T(\boldsymbol{u})\| = \|\boldsymbol{u}\|$ for all $\boldsymbol{u}$ in $\mathcal{R}^n$. (T preserves norms)

If T and U are orthogonal operators on $\mathcal{R}^n$, then TU and T^{-1} are orthogonal operators on $\mathcal{R}^n$.

Orthogonal Operator on the Euclidean Plane*

Theorem 7.11

Let T be an orthogonal linear operator on $\mathcal{R}^2$ with standard matrix Q.

a) If $\det Q = 1$, then T is a rotation.

b) If $\det Q = -1$, then T is a reflection.

Theorem 7.12

Let T and U be orthogonal linear operators on $\mathcal{R}^2$.

a) If both T and U are reflections, the TU is a rotation.

b) If one of T or U is a reflection and the other is a rotation, then TU is a reflection.

RIGID MOTIONS

A function F: $\mathcal{R}^n \rightarrow \mathcal{R}^n$ is called a **rigid motion** if

$$\|F(\boldsymbol{u})-F(\boldsymbol{v})\|=\|\boldsymbol{u}-\boldsymbol{v}\|$$

for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$.

- *Any orthogonal operator is a rigid motion.*
- *Any rigid motion that is also linear is an orthogonal operator.*
- A *translation* is one kind of rigid motion but usually not linear:
  For any $\boldsymbol{b}$ in $\mathcal{R}^n$, the function F_b: $\mathcal{R}^n \rightarrow \mathcal{R}^n$ defined by
  $$F_b(\boldsymbol{v})=\boldsymbol{v}+\boldsymbol{b}$$
  is called the **translation by *b***.
  - ✓ If $\boldsymbol{b}\neq\boldsymbol{0}$, F_b is not linear because $F_b(\boldsymbol{0})=\boldsymbol{b}\neq\boldsymbol{0}$.
  - ✓ F_b is a rigid motion.
- *The composition of two rigid motions* on $\mathcal{R}^n$ *is a rigid motion* on $\mathcal{R}^n$.
- Any rigid motion on $\mathcal{R}^n$ can be represented as the composition of an orthogonal operator followed by a translation.

Theorem 7.13

Let $T : \mathcal{R}^n \rightarrow \mathcal{R}^n$ be a rigid motion such that $T(\boldsymbol{0})=\boldsymbol{0}$.

a) $\|T(\boldsymbol{u})\|=\|\boldsymbol{u}\|$ for every $\boldsymbol{u}$ in $\mathcal{R}^n$.

b) $T(\boldsymbol{u})\cdot T(\boldsymbol{v})=\boldsymbol{u}\cdot\boldsymbol{v}$ for all $\boldsymbol{u}$ and $\boldsymbol{v}$ in $\mathcal{R}^n$.

c) T is linear.

d) T is an orthogonal operator.

Proof ?

Consider any rigid motion F on $\mathcal{R}^n$ and let T: $\mathcal{R}^n \rightarrow \mathcal{R}^n$ be defined by

$$T(\boldsymbol{v})=F(\boldsymbol{v})-F(\boldsymbol{0}).$$

- T is a rigid motion.
- $T(\boldsymbol{0})=\boldsymbol{0}$. Therefore T is an orthogonal operator by Theorem 7.13

$$F(\boldsymbol{v})=T(\boldsymbol{v})+F(\boldsymbol{0}) \text{ for any } \boldsymbol{v} \text{ in } \mathcal{R}^n.$$

Setting $\boldsymbol{b}=F(\boldsymbol{0})$, we have

$$F(\boldsymbol{v})=F_b T(\boldsymbol{v})$$

for any $\boldsymbol{v}$ in $\mathcal{R}^n$, and hence F is the composition $F=F_b T$.

Any rigid motion on $\mathcal{R}^n$ is the composition of an orthogonal operator followed by a translation. Hence any rigid motion on $\mathcal{R}^2$ is the composition of a rotation or a reflection, followed by a translation.

7.6 Symmetric Matrices

Consider that the columns of invertible matrix P form a basis for $\mathcal{R}^n$ consisting of eigenvectors of an $n \times n$ diagonalizable matrix A, and the diagonal entries of D are the corresponding eigenvalues. Then, $A=PDP^{-1}$.

Now suppose that the columns of P also form an orthonormal basis for $\mathcal{R}^n$. Then,

- $P^T = P^{-1}$, by Theorem 7.9
- $A^T = (PDP^{-1})^T = (PDP^T)^T = PD^TP^T = PDP^T = PDP^{-1} = A.$
  Thus $A^T=A$, a *symmetric* matrix.

(The above proves that "if there is an orthonormal basis for $\mathcal{R}^n$ consisting of eigenvectors of A, then A is symmetric.")

Theorem 7.14

If $\boldsymbol{u}$ and $\boldsymbol{v}$ are eigenvectors of a symmetric matrix that correspond to distinct eigenvalues, then $\boldsymbol{u}$ and $\boldsymbol{v}$ are orthogonal.

Proof ?

Theorem 7.15

An $n \times n$ matrix A is symmetric iff there is an orthonormal basis for $\mathcal{R}^n$ consisting of eigenvectors of A. In this case, there exists an orthogonal matrix P and a diagonal matrix D such that $P^T AP=D$.

Proof (? If time is available)

(Preliminary: "*Every eigenvalue of a symmetric matrix having real entries is real*".)

By Theorem 7.14, the vectors in any eigenspace of a symmetric $n \times n$ matrix A are orthogonal to the vectors in any other eigenspace of A.

- Combine all of the vectors from orthonormal bases for the distinct eigenspaces of A, we obtain an orthonormal basis for $\mathcal{R}^n$ consisting of eigenvectors of A

Quadratic Forms: (Self –study)

In the plane, the equations of all conic sections can be expressed by

$$ax^2+2bxy+cy^2+dx+ey+f=0.$$

The associated quadratic form of the above equation is

$$ax^2+2bxy+cy^2.$$

Assume $b \neq 0$. Let $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ and $\boldsymbol{v} = \begin{bmatrix} x \\ y \end{bmatrix}$. Then, the associated quadratic form can be written as $\boldsymbol{v}^{\mathrm{T}} A \boldsymbol{v}$.

……….

SPECTRAL DECOMPOSITION OF A MATRIX

Consider an $n\times n$ symmetric matrix A and an orthonormal basis $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ for $\mathcal{R}^n$ consisting of eigenvectors of A with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

Let $P= [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]$ and D denote the $n\times n$ diagonal matrix with diagonal entries $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively. Then,

$A=PDP^T$

$$=P[\lambda_1\boldsymbol{e}_1 \quad \lambda_2\boldsymbol{e}_2 \quad \cdots \quad \lambda_n\boldsymbol{e}_n]\,[\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n]^{\mathrm{T}}$$

$$=[P(\lambda_1\boldsymbol{e}_1) \quad P(\lambda_2\boldsymbol{e}_2) \quad \cdots \quad P(\lambda_n\boldsymbol{e}_n)]\begin{bmatrix}\boldsymbol{u}_1^T\\ \boldsymbol{u}_2^T\\ \vdots\\ \boldsymbol{u}_n^T\end{bmatrix}$$

$$=[\lambda_1\boldsymbol{u}_1 \quad \lambda_2\boldsymbol{u}_2 \quad \cdots \quad \lambda_n\boldsymbol{u}_n)]\begin{bmatrix}\boldsymbol{u}_1^T\\ \boldsymbol{u}_2^T\\ \vdots\\ \boldsymbol{u}_n^T\end{bmatrix} = \lambda_1\boldsymbol{u}_1\boldsymbol{u}_1^T + \lambda_2\boldsymbol{u}_2\boldsymbol{u}_2^T + \cdots + \lambda_1\boldsymbol{u}_n\boldsymbol{u}_n^T$$

- $P_i = \boldsymbol{u}_i \boldsymbol{u}_i^T$ is a matrix of rank 1.
- P_i is the orthogonal projection matrix for Span $\{\boldsymbol{u}_i\}$.

  Let W=Span $\{\boldsymbol{u}_i\}$ and C= $\boldsymbol{u}_i$. Applying Theorem 7.8, we have

$$P_w = \boldsymbol{u}_i(\boldsymbol{u}_i^T \boldsymbol{u}_i)^{-1}\boldsymbol{u}_i^T = \boldsymbol{u}_i(\boldsymbol{u}_i \cdot \boldsymbol{u}_i)^{-1}\boldsymbol{u}_i^T = \boldsymbol{u}_i(1)^{-1}\boldsymbol{u}_i^T = \boldsymbol{u}_i\boldsymbol{u}_i^T$$

- The representation

$$A = \lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_n P_n$$

  is called a **spectral decomposition** of A.
- P_i is symmetric and satisfies $P_i^2 = P_i$.

Theorem 7.16 (**Spectral Decomposition Theorem**)

Let A be an $n \times n$ symmetric matrix, and let $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n\}$ be an orthonormal basis for $\mathcal{R}^n$ consisting of eigenvectors of A with corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Then there exist symmetric matrices $P_1, P_2, \ldots, P_n$ such that the following results hold:

a) $A = \lambda_1 P_1 + \lambda_2 P_2 + \ldots + \lambda_n P_n$.

b) rank $P_i = 1$ for all i.

c) $P_i P_i = P_i$ for all i and $P_i P_j = O$ if $i \neq j$.

d) $P_i \boldsymbol{u}_i = \boldsymbol{u}_i$ for all i, and $P_i \boldsymbol{u}_j = \boldsymbol{0}$ if $i \neq j$.

(Spectral approximation ?)

7.7 Singular Value Decomposition (SVD)

Problems arise when

- A is not symmetric.
- A is not square, for which eigenvectors are not defined.

Theorem 7.17

Let A be an $m \times n$ matrix of rank k. Then there exist orthonormal bases $\mathcal{B}_1 = \{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ for $\mathcal{R}^n$ and $\mathcal{B}_2 = \{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_m\}$ for $\mathcal{R}^m$ and scalars

$$\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > 0$$

such that

$$A\boldsymbol{v}_i = \begin{cases} \sigma_i \boldsymbol{u}_i & \text{if } 1 \leq i \leq k \\ \boldsymbol{0} & \text{if } i > k \end{cases}$$

and

$$A^T\boldsymbol{u}_i = \begin{cases} \sigma_i \boldsymbol{v}_i & \text{if } 1 \leq i \leq k \\ \boldsymbol{0} & \text{if } i > k. \end{cases}$$

More on SVD

- If $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots., \boldsymbol{v}_n\}$ and $\{\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots., \boldsymbol{u}_m\}$ are any orthonormal bases for $\mathcal{R}^n$ and $\mathcal{R}^m$, respectively, that satisfy the two equations in Theorem 7.17,
  - ✓ then each $\boldsymbol{v}_i$ is an eigenvector of A^TA corresponding to the eigenvalue σ_i^2 if $i \leq k$ and to the eigenvalue 0 if $i > k$.

    ($\because$ $A^TA\,\boldsymbol{v}_i = A^T\sigma_i\,\boldsymbol{u}_i = \sigma_i\,A^T\boldsymbol{u}_i = \sigma_i\,(\sigma_i\boldsymbol{v}_i) = \sigma_i^2\,\boldsymbol{v}_i$ if $i \leq k$)
  - ✓ Furthermore, for $i=1,2, \ldots, k$, the vector $\boldsymbol{u}_i$ is an eigenvector of AA^T corresponding to the eigenvalue σ_i^2 and for $i > k$, the vector $\boldsymbol{u}_i$ is an eigenvector of AA^T corresponding to the eigenvalue 0.

$\Rightarrow$ σ_i's are the unique scalars satisfying the two equations in Theorem 7.17. They are called the **singular value** of matrix A.

- The orthonormal bases $\mathcal{B}_1$ and $\mathcal{B}_2$ in Theorem 7.17 are not unique.
- Consider a linear transformation T: $\mathcal{R}^n \rightarrow \mathcal{R}^m$, with standard matrix A. For any vector $\boldsymbol{v}_i \in \mathcal{B}_1$ and $\boldsymbol{u}_i \in \mathcal{B}_2$ such that $A\boldsymbol{v}_i = \sigma_i\boldsymbol{u}_i$, $\|T(c\boldsymbol{v}_i)\| = \sigma_i\|c\boldsymbol{v}_i\|$.

The Singular Value Decomposition of a Matrix

Follow the notations in Theorem 7.17. Define $n \times n$ orthogonal matrix V and $m \times m$ orthogonal matrix U, respectively, by

$$V=[\boldsymbol{v}_1 \ \ \boldsymbol{v}_2 \ \ \ldots. \ \ \boldsymbol{v}_n] \text{ and } U=[\boldsymbol{u}_1 \ \ \boldsymbol{u}_2 \ \ \ldots. \ \ \boldsymbol{u}_m].$$

Let Σ be defined by $\Sigma = \left[\begin{array}{cccc|cccc} \sigma_1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_k & 0 & 0 & \cdots & 0 \\ \hline 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{array}\right]$ (*)

Then,

$$AV = A[\boldsymbol{v}_1\ \ \boldsymbol{v}_2\ \ \ldots.\ \ \boldsymbol{v}_n] = [A\boldsymbol{v}_1\ \ A\boldsymbol{v}_2\ \ \ldots.\ A\boldsymbol{v}_n]$$
$$=[\sigma_1\boldsymbol{u}_1\quad \sigma_2\boldsymbol{u}_2\ \ \ldots\ \sigma_k\boldsymbol{u}_k\ \ \boldsymbol{0}\ \ \ldots\ \boldsymbol{0}]$$
$$=U\Sigma.$$

Since V is an orthogonal matrix,

$$A= U\Sigma V^{-1} = U\Sigma V^T.$$

Any factorization of an $m\times n$ matrix A into the product $A= U\Sigma V^T$, where U and V are orthogonal matrices and Σ is an $m\times n$ matrix of the above form, is called a **singular value decomposition** of A.

Theorem 7.18 (**Singular Value Decomposition**)

For any $m \times n$ matrix A of rank k. There exists $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_k > 0$, and an $m \times m$ orthogonal matrix U, and an $n \times n$ orthogonal matrix V such that

$$A = U\Sigma V^T$$

where Σ is the $m \times n$ matrix given in equation (*).

Further result: (proof ?)

If $A = U\Sigma V^T$ is any SVD of an $m \times n$ matrix A, then the nonzero diagonal entries of Σ are the singular values of A, and the columns of V and the columns of U, which form orthonormal bases for $\mathcal{R}^n$ and $\mathcal{R}^m$, respectively, satisfy the two equations in Theorem 7.17.

(? $A\boldsymbol{v}_i = U\Sigma V^T \boldsymbol{v}_i = U\Sigma\,\boldsymbol{e}_i = U\sigma_i \boldsymbol{e}_i = \sigma_i\, U\boldsymbol{e}_i = \sigma_i \boldsymbol{u}_i$ if $i \leq k$)

$\Rightarrow$ The columns of U and V in a singular decomposition of a matrix A are referred to as the *left* and *right* *singular vectors* of A, respectively.

Orthogonal Projections, Systems of Linear Equations, and the Pseudoinverse

Let A be an $m \times n$ matrix and $\boldsymbol{b}$ be in $\mathcal{R}^m$. Consider a system of linear equation $A\boldsymbol{x}=\boldsymbol{b}$.

- If it is consistent, a vector $\boldsymbol{u}$ in $\mathcal{R}^n$ is a solution iff $\|A\boldsymbol{u}-\boldsymbol{b}\|=0$.
- If it is inconsistent, $\|A\boldsymbol{u}-\boldsymbol{b}\|>0$ for every $\boldsymbol{u}$ in $\mathcal{R}^n$. A general objective is to find a vector $\boldsymbol{z}$ in $\mathcal{R}^n$ that minimizes the distance between $A\boldsymbol{u}$ and $\boldsymbol{b}$, that is, a vector $\boldsymbol{z}$ such that

$$\|A\boldsymbol{z}-\boldsymbol{b}\| \leq \|A\boldsymbol{u}-\boldsymbol{b}\| \text{ for all } \boldsymbol{u} \text{ in } \mathcal{R}^n.$$

  - ✓ This is the least-squares problem addressed previously where

$$\|A\boldsymbol{z}-\boldsymbol{b}\| \text{ is a minimum iff } A\boldsymbol{z}=P_W\boldsymbol{b},$$

where W=Col A and P_W is the orthogonal projection matrix for W.

Theorem 7.19

Let A be an $m \times n$ matrix of rank k having a singular value decomposition $A = U\Sigma V^T$, and let $W = \text{Col } A$. Let D be the $m \times m$ diagonal matrix whose first k diagonal entries are 1s and whose other entries are 0s. Then

$$P_W = UDU^T.$$

Proof?

Let $P = UDU^T$.

- Since $P^2 = P^T = P$, P is an orthogonal projection matrix for some subspace of $\mathcal{R}^m$.

  (For $\boldsymbol{u}, \boldsymbol{v} \in \mathcal{R}^m$, $P\boldsymbol{v} \bullet (\boldsymbol{u} - P\boldsymbol{u}) = \boldsymbol{v} \bullet P^T(\boldsymbol{u} - P\boldsymbol{u}) = \boldsymbol{v} \bullet P(\boldsymbol{u} - P\boldsymbol{u}) = \boldsymbol{v} \bullet (P\boldsymbol{u} - P^2\boldsymbol{u}) = 0$)
- Show that this subspace is W.

  …..

Modify the $m \times n$ matrix Σ to obtain a new $n \times m$ matrix $\Sigma^{\dagger}$ defined by

$$\Sigma^{\dagger} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \frac{1}{\sigma_k} & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{bmatrix} \qquad (**)$$

Thus $\Sigma\Sigma^{\dagger}$ =D, and hence

$$A(V\Sigma^{\dagger}U^T) = U\Sigma V^T(V\Sigma^{\dagger}U^T) = UDU^T = P.$$

$\rightarrow$ Col $P \subset$ Col A $(=W)$

$\rightarrow$ Col P=Col A $=W$ because dim(Col P)$=k$.

Therefore $P=P_W$.

Let A be an $m \times n$ matrix with a singular value decomposition $A = U\Sigma V^T$, $\boldsymbol{b}$ be a vector in $\mathcal{R}^m$, and $\boldsymbol{z} = V\Sigma^{\dagger} U^T\boldsymbol{b}$, where $\Sigma^{\dagger}$ is as in (**). Then the following statement are true:

a) If the system $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, then $\boldsymbol{z}$ is the unique solution of least norm.

b) If the system $A\boldsymbol{x} = \boldsymbol{b}$ is inconsistent, the $\boldsymbol{z}$ is the unique vector of least norm such that

$$\|A\boldsymbol{z} - \boldsymbol{b}\| \leq \|A\boldsymbol{u} - \boldsymbol{b}\|$$

for all $\boldsymbol{u}$ in $\mathcal{R}^n$.

- Although a singular value decomposition of a matrix $A = U\Sigma V^T$ is not unique, the matrix $V\Sigma^{\dagger}U^T$ is unique.
- For a given matrix $A=U\Sigma V^T$, the matrix $V\Sigma^{\dagger}U^T$ is called the **pseudoinverse**, or **Moore-Penrose generalized inverse**, of A and is denoted $A^{\dagger}$.
  - ✓ The pseudoinverse of Σ is $\Sigma^{\dagger}$.
  - ✓ The terminology pseudoinverse is due to that fact that if A is invertible, then $A^{-1}=A^{\dagger}$.

Applications of the Pseudoinverse

For any $m \times n$ matrix A and any vector $\boldsymbol{b}$ in $\mathcal{R}^m$, the following statements are true:

1. The orthogonal projection matrix for Col A is $AA^\dagger$.
2. The unique vector of least norm that minimizes $\|A\boldsymbol{u} - \boldsymbol{b}\|$ for $\boldsymbol{u}$ in $\mathcal{R}^n$ is $A^\dagger\boldsymbol{b}$.

Therefore, if $A\boldsymbol{x} = \boldsymbol{b}$ is consistent, $A^\dagger\boldsymbol{b}$ is the unique solution of least norm.

7.8 Principal Component Analysis (PCA)