Topic 2 - Inner Product Space

AMA3724 Further Mathematical Methods(2024/25 Semester 1)
Lecturer: Jianbo Cui

- Inner Product and norm in $\mathbb{R}^n$
- Orthogonality
- Orthogonal subspace
- Projection
- Orthonormal set
- Gram-Schmidt Algorithm
- Orthonormal basis
- Least Squares Solution

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Inner product and norm in $\mathbb{R}^n$

Inner product

Given two vectors $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ in $\mathbb{R}^n$. The **inner product (scalar product)** of $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x} = \begin{bmatrix} y_1 & y_2 & \cdots & y_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Norm and unit vector

1. The **norm** of a vector $\mathbf{x}$ is defined by

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}.$$

2. A vector $\mathbf{u}$ is called a **unit vector** if $\|\mathbf{u}\| = 1$.

Inner product and norm in $\mathbb{R}^n$

Example For $\mathbf{x} = \begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix}3\\-1\\2\\1\\-1\end{bmatrix}$,

$$\begin{aligned}
\langle \mathbf{x}, \mathbf{y}\rangle &= \mathbf{y}^T\mathbf{x} = \begin{bmatrix}3 & -1 & 2 & 1 & -1\end{bmatrix}\begin{bmatrix}1\\2\\3\\4\\5\end{bmatrix}\\
&= 1\cdot 3 + 2\cdot(-1) + 3\cdot 2 + 4\cdot 1 + 5\cdot(-1)\\
&= 6\\
\|\mathbf{x}\| &= \left(1^2+2^2+3^2+4^2+5^2\right)^{\frac{1}{2}} = \sqrt{55}\\
\|\mathbf{y}\| &= \left(3^2+(-1)^2+2^2+1^2+(-1)^2\right)^{\frac{1}{2}} = 4
\end{aligned}$$

Let $\mathbf{z} = \dfrac{1}{\sqrt{55}}\mathbf{x}$. Then

$$\|\mathbf{z}\| = 1.$$

[1]:
```
x = sp.Matrix([[1,2,3,4,5]]).T
y = sp.Matrix([[3,-1,2,1,-1]]).T
```

[2]:
```
#Inner product of x and y
In [2]: y.T@x
```

[2]: $\begin{bmatrix}6\end{bmatrix}$

[3]:
```
# norm of x and y
x.norm()
```

[3]: $\sqrt{55}$

[4]:
```
y.norm()
```

[4]: 4

[5]:
```
z = x/x.norm(); z
```

[5]: $\begin{bmatrix}\frac{\sqrt{55}}{55}\\ \frac{2\sqrt{55}}{55}\\ \frac{3\sqrt{55}}{55}\\ \frac{4\sqrt{55}}{55}\\ \frac{\sqrt{55}}{11}\end{bmatrix}$

[6]:
```
z.norm()
```

[6]: 1

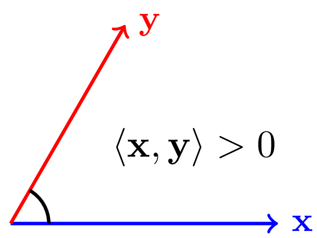
Inner product and norm in $\mathbb{R}^n$

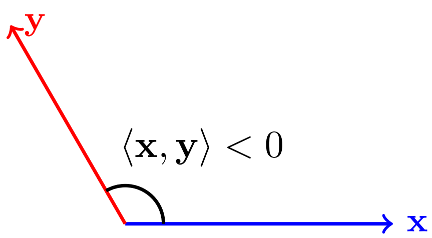
Properties of inner product

For any $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{w}$ in $\mathbb{R}^n$ and any constant c,

1. $\langle \mathbf{x}, \mathbf{y}\rangle = \langle \mathbf{y}, \mathbf{x}\rangle$;
2. $\langle \mathbf{x}+\mathbf{y}, \mathbf{w}\rangle = \langle \mathbf{x}, \mathbf{w}\rangle + \langle \mathbf{y}, \mathbf{w}\rangle$;
3. $\langle c\mathbf{x}, \mathbf{y}\rangle = c\langle \mathbf{x}, \mathbf{y}\rangle$;
4. $\langle \mathbf{x}, \mathbf{x}\rangle \geq 0$.

Furthermore, $\langle \mathbf{x}, \mathbf{x}\rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.



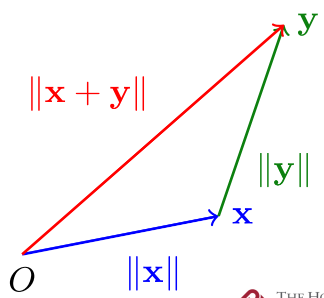


Properties of norm

For any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$ and any constant c,

1. $\|c\mathbf{x}\| = |c|\cdot\|\mathbf{x}\|$;
2. $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$;
3. $\|\mathbf{x}\| \geq 0$.

Furthermore, $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

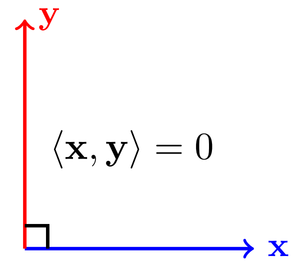


Orthogonality

Orthogonality

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be **orthogonal** if

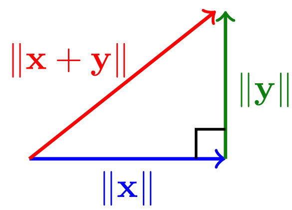
$$\langle \mathbf{x}, \mathbf{y} \rangle = 0.$$



Pythagorean Theorem

Two vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal if and only if

$$\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$



Proof. For any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

$$\|\mathbf{x}+\mathbf{y}\|^2 = \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2.$$

Thus, $\|\mathbf{x}+\mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2$ if and only if $\langle \mathbf{x}, \mathbf{y} \rangle = 0$.
So the theorem holds.

Orthogonality

Example

1. Let $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}$.

   Then $\langle \mathbf{x}, \mathbf{y} \rangle = 0$ and thus $\mathbf{x}$ and $\mathbf{y}$ are orthogonal.
   Notice also that

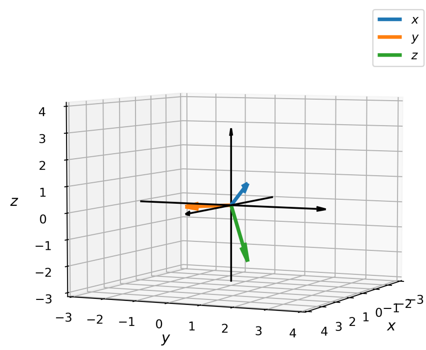
$$\|\mathbf{x}+\mathbf{y}\|^2 = \left\| \begin{bmatrix} 2 \\ 0 \\ 0 \\ 2 \end{bmatrix} \right\|^2 = 8 = \left\| \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right\|^2 + \left\| \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} \right\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2.$$

2. Let $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, and $\mathbf{z} = \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$.

   Then

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle = \langle \mathbf{y}, \mathbf{z} \rangle = 0.$$

   Thus, $\mathbf{x}$, $\mathbf{y}$, and $\mathbf{z}$ are mutually orthogonal.



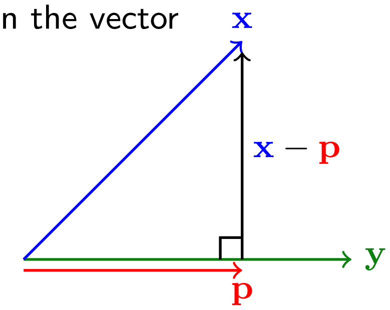
Orthogonality

Vector projection

Suppose $\mathbf{x}$ and $\mathbf{y}$ are two nonzero vectors in $\mathbb{R}^n$, then the vector

$$\mathbf{p} = \operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$$

is the **vector projection** of $\mathbf{x}$ onto $\mathbf{y}$.



Vector projection vs Orthogonality

Let $\mathbf{p} = \operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$. The vectors $\mathbf{x} - \mathbf{p}$ and $\mathbf{y}$ are orthogonal, i.e.,

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{y} \rangle = 0.$$

Proof. Let $\mathbf{p} = \operatorname{proj}_{\mathbf{y}} \mathbf{x}$. Then

$$\langle \mathbf{x} - \mathbf{p}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle - \langle \mathbf{p}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle - \left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \right) \langle \mathbf{y}, \mathbf{y} \rangle = 0.$$

The Hong Kong Polytechnic University 香港理工大學

Orthogonality

Example 2.1 Let $\mathbf{x} = \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix}$. Find the orthogonal projection of $\mathbf{x}$ onto $\mathbf{y}$. Then write $\mathbf{x}$ as the sum of two orthogonal vectors, one is a multiple of $\mathbf{y}$ and one is orthogonal to $\mathbf{y}$.

Solution. Notice that

$$\langle \mathbf{x}, \mathbf{y} \rangle = 35 \quad \text{and} \quad \langle \mathbf{y}, \mathbf{y} \rangle = 21.$$

Then the orthogonal projection of $\mathbf{x}$ onto $\mathbf{y}$ is

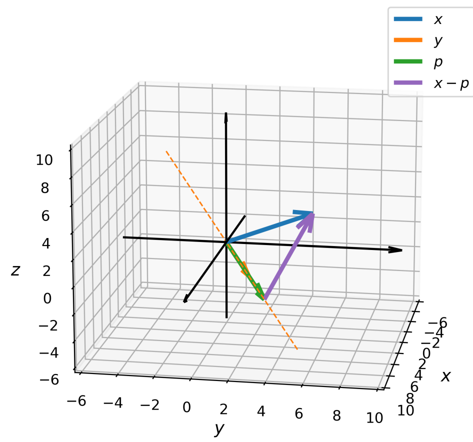
$$\mathbf{p} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y} = \frac{35}{21} \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix} = \begin{bmatrix} 20/3 \\ 10/3 \\ -5/3 \end{bmatrix}.$$

Also the component of $\mathbf{x}$ orthogonal to $\mathbf{y}$ is

$$\mathbf{x} - \mathbf{p} = \begin{bmatrix} 7 \\ 6 \\ 5 \end{bmatrix} - \begin{bmatrix} 20/3 \\ 10/3 \\ -5/3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 8/3 \\ 20/3 \end{bmatrix}.$$

Then $\mathbf{p}$ is a multiple of $\mathbf{u}$ while $\mathbf{x} - \mathbf{p}$ is orthogonal to $\mathbf{y}$ and

$$\mathbf{x} = \mathbf{p} + (\mathbf{x} - \mathbf{p}).$$



Orthogonality

Cauchy-Schwarz inequality

For any $\mathbf{x}$ and $\mathbf{y}$ in $\mathbb{R}^n$,

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

Proof. The result is clear if $\mathbf{x}$ or $\mathbf{y}$ is zero. Suppose $\mathbf{x}$ and $\mathbf{y}$ are nonzero. Let $\mathbf{p} = \operatorname{proj}_{\mathbf{y}} \mathbf{x} = \dfrac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$. Then $\mathbf{p}$ and $\mathbf{x} - \mathbf{p}$ are orthogonal and hence by Pythagorean Theorem,

$$\|\mathbf{p}\|^2 + \|\mathbf{x} - \mathbf{p}\|^2 = \|\mathbf{x}\|^2 \quad \Longrightarrow \quad \|\mathbf{p}\|^2 = \|\mathbf{x}\|^2 - \|\mathbf{x} - \mathbf{p}\|^2 \leq \|\mathbf{x}\|^2.$$

Then

$$\left(\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \right)^2 \langle \mathbf{y}, \mathbf{y} \rangle = \|\mathbf{p}\|^2 \leq \|\mathbf{x}\|^2 \quad \Longrightarrow \quad (\langle \mathbf{x}, \mathbf{y} \rangle)^2 \leq \|\mathbf{x}\|^2 \|\mathbf{y}\|^2.$$

Therefore, $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|$.

Orthogonal subspace

Orthogonal subspace

Two subspaces U and W of a vector space V are orthogonal if

$$\langle \mathbf{u}, \mathbf{w} \rangle = 0 \quad \text{for all} \quad \mathbf{u} \in U \text{ and } \mathbf{w} \in W.$$

Example Let

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{u}_2 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}, \quad \mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \quad \mathbf{w}_2 = \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \\ 1 \\ -1 \\ -1 \\ 1 \end{bmatrix}.$$

Set $U = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$ and $W = \operatorname{Span}\{\mathbf{w}_1, \mathbf{w}_2\}$. Then U and W are orthogonal.

Proof. For any $\mathbf{u} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 \in U$ and $\mathbf{w} = d_1\mathbf{w}_1 + d_2\mathbf{w}_2 \in V$,

$$\begin{aligned} \langle \mathbf{u}, \mathbf{w} \rangle &= \langle c_1\mathbf{u}_1 + c_2\mathbf{u}_2, d_1\mathbf{w}_1 + d_2\mathbf{w}_2 \rangle \\ &= c_1d_1\langle \mathbf{u}_1, \mathbf{w}_1 \rangle + c_1d_2\langle \mathbf{u}_1, \mathbf{w}_2 \rangle + c_2d_1\langle \mathbf{u}_2, \mathbf{w}_1 \rangle + c_2d_2\langle \mathbf{u}_2, \mathbf{w}_2 \rangle = 0. \end{aligned}$$

Orthogonal complement

The orthogonal complement, denoted $S^\perp$, of a subspace S in a vector space V contains all vectors that are orthogonal to S, that is

$$S^\perp = \{\mathbf{w} \in V : \langle \mathbf{w}, \mathbf{v} \rangle = \mathbf{0} \quad \text{for all } \mathbf{v} \in S\}.$$

The orthogonal complement is also a subspace.

Example (cont.) Let $U = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2\}$. Set $A = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix}$. Then $U = \operatorname{Col} A$.

$$\begin{aligned}
\mathbf{y} \in U^\perp \quad &\Longleftrightarrow \quad \langle \mathbf{y}, \mathbf{u} \rangle = 0 \quad \text{for all } \mathbf{u} \in U \\
&\Longleftrightarrow \quad \langle \mathbf{y}, \mathbf{u}_1 \rangle = \langle \mathbf{y}, \mathbf{u}_2 \rangle = 0 \\
&\Longleftrightarrow \quad \begin{bmatrix}0\\0\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1^T\mathbf{y}\\ \mathbf{u}_2^T\mathbf{y}\end{bmatrix} = \begin{bmatrix}\mathbf{u}_1^T\\ \mathbf{u}_2^T\end{bmatrix}\mathbf{y} = A^T\mathbf{y} \\
&\Longleftrightarrow \quad \mathbf{y} \in \operatorname{Nul} A^T.
\end{aligned}$$

Thus, $U^\perp = \operatorname{Nul} A^T = \operatorname{Span}\left\{\begin{bmatrix}-1\\0\\1\\0\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\0\\1\\0\\0\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\0\\0\\1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\0\\0\\0\\1\\0\\0\end{bmatrix}, \begin{bmatrix}-1\\0\\0\\0\\0\\0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\-1\\0\\1\\0\\0\\0\\1\end{bmatrix}\right\}$.

Orthogonal complement of column space

Given an $m \times n$ matrix A.

- $\operatorname{Nul} A$ is the orthogonal complement of $\operatorname{Col} A^T$ in $\mathbb{R}^n$.
- $\operatorname{Nul} A^T$ is the orthogonal complement of $\operatorname{Col} A$ in $\mathbb{R}^m$.

Proof. For any $\mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned}
\mathbf{y} \in (\operatorname{Col} A^T)^\perp \quad &\Longleftrightarrow \quad \langle \mathbf{y}, \mathbf{b} \rangle = 0 \quad \text{for all } \mathbf{b} \in \operatorname{Col} A^T \\
&\Longleftrightarrow \quad \langle \mathbf{y}, A^T\mathbf{x} \rangle = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^m \\
&\Longleftrightarrow \quad (A^T\mathbf{x})^T\mathbf{y} = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^m \\
&\Longleftrightarrow \quad \mathbf{x}^T A\mathbf{y} = 0 \quad \text{for all } \mathbf{x} \in \mathbb{R}^m \\
&\Longleftrightarrow \quad A\mathbf{y} = \mathbf{0} \\
&\Longleftrightarrow \quad \mathbf{y} \in \operatorname{Nul} A.
\end{aligned}$$

Therefore, $\operatorname{Nul} A = (\operatorname{Col} A^T)^\perp$.

Finally, by replacing A by A^T, we have

$$\operatorname{Nul} A^T = (\operatorname{Col} A)^\perp.$$

Projection

Given $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. Recall that $\mathbf{p} = \operatorname{proj}_{\mathbf{y}} \mathbf{x} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y}$ is the vector projection of $\mathbf{x}$ onto a one-dimensional subspace $\operatorname{Span}\{\mathbf{y}\}$. Let

$$P = \frac{1}{\mathbf{y}^T \mathbf{y}} \mathbf{y}\mathbf{y}^T,$$

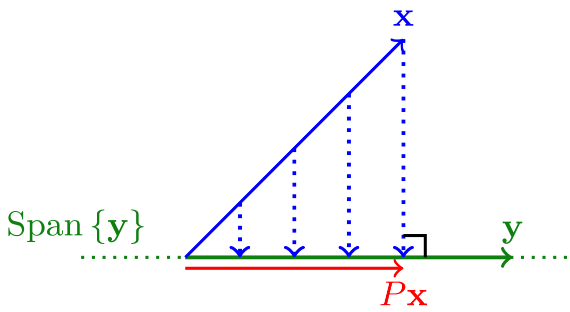
which is an $n \times n$ matrix. Then

$$P\mathbf{x} = \left(\frac{1}{\mathbf{y}^T\mathbf{y}} \mathbf{y}\mathbf{y}^T \right) \mathbf{x} = \frac{1}{\mathbf{y}^T\mathbf{y}} \mathbf{y}(\mathbf{y}^T\mathbf{x}) = \frac{\mathbf{y}^T\mathbf{x}}{\mathbf{y}^T\mathbf{y}} \mathbf{y} = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{y}, \mathbf{y} \rangle} \mathbf{y} = \mathbf{p}.$$

We call the P the projection matrix onto $\operatorname{Span}\{\mathbf{y}\}$.

Note that

- P is symmetric, i.e., $P^T = P$.
- $P^2 = P$.
- $P\mathbf{x} \in \operatorname{Span}\{\mathbf{y}\}$.
- $\mathbf{x} - P\mathbf{x} \in (\operatorname{Span}\{\mathbf{y}\})^{\perp}$.



Proof.

$$P^2 = \left(\frac{1}{\mathbf{y}^T\mathbf{y}} \mathbf{y}\mathbf{y}^T \right) \left(\frac{1}{\mathbf{y}^T\mathbf{y}} \mathbf{y}\mathbf{y}^T \right) = \frac{1}{(\mathbf{y}^T\mathbf{y})^2} \mathbf{y}\mathbf{y}^T\mathbf{y}\mathbf{y}^T = \frac{\mathbf{y}^T\mathbf{y}}{(\mathbf{y}^T\mathbf{y})^2} \mathbf{y}\mathbf{y}^T = \frac{1}{\mathbf{y}^T\mathbf{y}} \mathbf{y}\mathbf{y}^T = P.$$

Projection

Given a set of linearly independent vectors $\mathbf{a}_1, \ldots, \mathbf{a}_p \in \mathbb{R}^n$ and set the $n \times p$ matrix

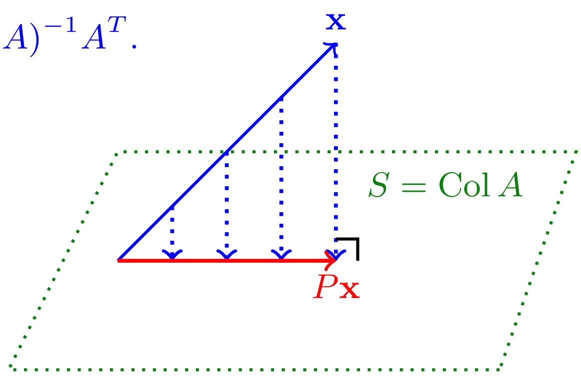
$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_p \end{bmatrix}.$$

Also let $S = \operatorname{Col} A = \operatorname{Span}\{\mathbf{a}_1, \ldots, \mathbf{a}_p\}$. Define the $n \times n$ projection matrix

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

For any $\mathbf{x} \in \mathbb{R}^n$,

- P is symmetric, i.e., $P^T = P$.
- $P^2 = P$.
- $P\mathbf{x} \in S = \operatorname{Col} A$.
- $\mathbf{x} - P\mathbf{x} \in S^{\perp} = (\operatorname{Col} A)^{\perp}$.



Proof. For any $\mathbf{x} \in \mathbb{R}^n$,

$$P\mathbf{x} = A(A^TA)^{-1}A^T\mathbf{x} = A\left((A^TA)^{-1}A^T\mathbf{x}\right) \in \operatorname{Col} A.$$

Recall that $(\operatorname{Col} A)^{\perp} = \operatorname{Nul} A^T$. Then

$$A^T(\mathbf{x} - P\mathbf{x}) = A^T\mathbf{x} - A^TA(A^TA)^{-1}A^T\mathbf{x} = A^T\mathbf{x} - A^T\mathbf{x} = \mathbf{0}.$$

So $\mathbf{x} - P\mathbf{x} \in \operatorname{Nul} A^T = (\operatorname{Col} A)^{\perp}$.

Exercise: Show that A^TA is invertible if and only if A has linearly independent columns.

Projection

Null space and column space of A^TA

For any $m \times n$ matrix A,

$$\mathrm{Nul}\, A^TA = \mathrm{Nul}\, A \quad \text{and} \quad \mathrm{Col}\, A^TA = \mathrm{Col}\, A^T.$$

Proof. For any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x} \in \mathrm{Nul}\, A \implies A\mathbf{x} = \mathbf{0} \implies A^TA\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in \mathrm{Nul}\, A^TA.$$

Therefore, $\mathrm{Nul}\, A \subseteq \mathrm{Nul}\, A^TA$. Now

$$\begin{aligned}\mathbf{x} \in \mathrm{Nul}\, A^TA &\implies A^TA\mathbf{x} = \mathbf{0} \implies \mathbf{x}^TA^TA\mathbf{x} = 0 \\ &\implies \|A\mathbf{x}\|^2 = 0 \implies A\mathbf{x} = \mathbf{0} \implies \mathbf{x} \in \mathrm{Nul}\, A.\end{aligned}$$

Thus, $\mathrm{Nul}\, A^TA \subseteq \mathrm{Nul}\, A$, and hence $\mathrm{Nul}\, A = \mathrm{Nul}\, A^TA$. Finally,

$$\mathrm{Col}\, A^T = (\mathrm{Nul}\, A)^\perp = (\mathrm{Nul}\, A^TA)^\perp = \mathrm{Col}\,(A^TA)^T = \mathrm{Col}\, A^TA.$$

Orthonormal set

Orthogonal and orthonormal sets

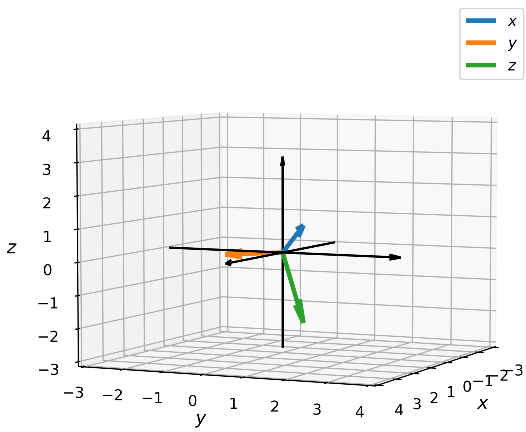
1. A set of vectors $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ is said to be an **orthogonal set** if these vectors are mutually orthogonal to each other, i.e.,

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0 \quad \text{whenever} \quad i \neq j.$$

2. A set of vectors $\{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is said to be an **orthonormal set** if these vectors are **unit vectors** and they form an orthogonal set.

1. $\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$ forms an orthogonal set in $\mathbb{R}^3$ only.

2. $\left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}, \begin{bmatrix} \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ -\frac{2}{\sqrt{6}} \end{bmatrix} \right\}$ forms an orthonormal set in $\mathbb{R}^3$.



[1]:
```
x1 = sp.Matrix([[1,1,1]]).T
x2 = sp.Matrix([[1,-1,0]]).T
x3 = sp.Matrix([[1,1,-2]]).T
display(x1,x2,x3)
```

$$\begin{bmatrix}1\\1\\1\end{bmatrix}$$
$$\begin{bmatrix}1\\-1\\0\end{bmatrix}$$
$$\begin{bmatrix}1\\1\\-2\end{bmatrix}$$

[2]:
```
# Inner products of all pairs
display(x1.T@x2, x1.T@x3, x2.T@x3)
```

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

[3]:
```
# the norm of vectors
x1.norm(), x2.norm(), x3.norm()
```

[3]: (sqrt(3), sqrt(2), sqrt(6))

[4]:
```
# combine vectors into a matrix
In [4]: X = sp.BlockMatrix([x1,x2,x3]).
↪as_explicit(); X
```

[4]: $$\begin{bmatrix}1 & 1 & 1\\1 & -1 & 1\\1 & 0 & -2\end{bmatrix}$$

[5]:
```
X.T@X
```

[5]: $$\begin{bmatrix}3 & 0 & 0\\0 & 2 & 0\\0 & 0 & 6\end{bmatrix}$$

[6]:
```
u1 = sp.Matrix([[1/sp.sqrt(3), 1/sp.
↪sqrt(3), 1/sp.sqrt(3)]]).T
u2 = sp.Matrix([[1/sp.sqrt(2), -1/sp.
↪sqrt(2), 0]]).T
u3 = sp.Matrix([[1/sp.sqrt(6), 1/sp.
↪sqrt(6), -2/sp.sqrt(6)]]).T
display(u1,u2,u3)
```

$$\begin{bmatrix}\frac{\sqrt{3}}{3}\\\frac{\sqrt{3}}{3}\\\frac{\sqrt{3}}{3}\end{bmatrix}$$
$$\begin{bmatrix}\frac{\sqrt{2}}{2}\\-\frac{\sqrt{2}}{2}\\0\end{bmatrix}$$
$$\begin{bmatrix}\frac{\sqrt{6}}{6}\\\frac{\sqrt{6}}{6}\\-\frac{\sqrt{6}}{3}\end{bmatrix}$$

[7]:
```
# Inner products of all pairs
display(u1.T@u2, u1.T@u3, u2.T@u3)
```

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

[8]:
```
# the norm of vectors
u1.norm(), u2.norm(), u3.norm()
```

[8]: (1, 1, 1)

[9]:
```
# combine vectors into a matrix
U = sp.BlockMatrix([u1,u2,u3]).
↪as_explicit(); U
```

[9]: $$\begin{bmatrix}\frac{\sqrt{3}}{3} & \frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6}\\\frac{\sqrt{3}}{3} & -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{6}\\\frac{\sqrt{3}}{3} & 0 & -\frac{\sqrt{6}}{3}\end{bmatrix}$$

[10]:
```
U.T@U
```

[10]: $$\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$$

Orthonormal set

Let $X = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}$. Then

$$X^T X = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \mathbf{x}_3^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{x}_2 & \mathbf{x}_1^T\mathbf{x}_3 \\ \mathbf{x}_2^T\mathbf{x}_1 & \mathbf{x}_2^T\mathbf{x}_2 & \mathbf{x}_2^T\mathbf{x}_3 \\ \mathbf{x}_3^T\mathbf{x}_1 & \mathbf{x}_3^T\mathbf{x}_2 & \mathbf{x}_3^T\mathbf{x}_3 \end{bmatrix}$$

$$= \begin{bmatrix} \langle \mathbf{x}_1, \mathbf{x}_1 \rangle & \langle \mathbf{x}_2, \mathbf{x}_1 \rangle & \langle \mathbf{x}_3, \mathbf{x}_1 \rangle \\ \langle \mathbf{x}_1, \mathbf{x}_2 \rangle & \langle \mathbf{x}_2, \mathbf{x}_2 \rangle & \langle \mathbf{x}_3, \mathbf{x}_2 \rangle \\ \langle \mathbf{x}_1, \mathbf{x}_3 \rangle & \langle \mathbf{x}_2, \mathbf{x}_3 \rangle & \langle \mathbf{x}_3, \mathbf{x}_3 \rangle \end{bmatrix}.$$

Therefore, $X^T X = I_3$ implies

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \begin{cases} 1 & i = j, \\ 0 & i \neq j. \end{cases}$$

Thus, $\{\mathbf{x}_1 \mathbf{x}_2, \mathbf{x}_3\}$ forms an orthonormal set.

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Orthonormal set

Orthogonal set $\Rightarrow$ Orthonormal set

Suppose $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ is an orthogonal set. Let

$$\mathbf{u}_j = \frac{1}{\|\mathbf{x}_j\|}\mathbf{x}_j \quad \text{for } j = 1, \ldots, k.$$

Then the set $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ forms an orthonormal set and

$$\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}.$$

Let $\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$.

- Notice that $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ is NOT an orthogonal set.
- Can we construct an orthogonal/orthonormal set $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ from $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$ and

$$\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}?$$

THE HONG KONG POLYTECHNIC UNIVERSITY 香港理工大學

Gram-Schmidt Algorithm

Given a set of linearly independent vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$. Let

$$
\begin{aligned}
\mathbf{y}_1 &= \mathbf{x}_1 & \mathbf{u}_1 &= \frac{1}{\|\mathbf{y}_1\|}\mathbf{y}_1 \\
\mathbf{y}_2 &= \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 & \mathbf{u}_2 &= \frac{1}{\|\mathbf{y}_2\|}\mathbf{y}_2 \\
\mathbf{y}_3 &= \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{x}_3, \mathbf{u}_2 \rangle \mathbf{u}_2 & \mathbf{u}_3 &= \frac{1}{\|\mathbf{y}_3\|}\mathbf{y}_3 \\
&\vdots \\
\mathbf{y}_k &= \mathbf{x}_k - \langle \mathbf{x}_k, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{x}_k, \mathbf{u}_2 \rangle \mathbf{u}_2 - \cdots - \langle \mathbf{x}_k, \mathbf{u}_{k-1} \rangle \mathbf{u}_{k-1} & \mathbf{u}_k &= \frac{1}{\|\mathbf{y}_k\|}\mathbf{y}_k.
\end{aligned}
$$

Then

- $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\}$ forms an orthogonal set, and
- $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$ forms an orthonormal set.

Furthermore,

$$\operatorname{Span}\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} = \operatorname{Span}\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_k\} = \operatorname{Span}\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}.$$

Example 2.2 Let $\mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$, $\mathbf{x}_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix}$, and $\mathbf{x}_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$.

Define

$$\mathbf{y}_1 = \mathbf{x}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \qquad \mathbf{u}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

$$\mathbf{y}_2 = \mathbf{x}_2 - \langle \mathbf{x}_2, \mathbf{u}_1 \rangle \mathbf{u}_1 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix} - \frac{3}{2}\cdot\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix} \qquad \mathbf{u}_2 = \frac{2}{\sqrt{3}}\begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}$$

$$
\begin{aligned}
\mathbf{y}_3 &= \mathbf{x}_3 - \langle \mathbf{x}_3, \mathbf{u}_1 \rangle \mathbf{u}_1 - \langle \mathbf{x}_3, \mathbf{u}_2 \rangle \mathbf{u}_2 \\
&= \begin{bmatrix}0\\0\\1\\1\end{bmatrix} - 1\cdot\frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \frac{1}{\sqrt{3}}\cdot\frac{2}{\sqrt{3}}\begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix} = \begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix} \qquad \mathbf{u}_3 = \frac{\sqrt{3}}{\sqrt{2}}\begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}.
\end{aligned}
$$

Now

$$\mathbf{y}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \quad \mathbf{y}_2 = \begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}, \quad \text{and} \quad \mathbf{y}_3 = \begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}.$$

Then $\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\}$ forms an orthogonal set and

$$\text{Span}\{\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}.$$

Also

$$\mathbf{u}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \ \mathbf{u}_2 = \frac{2}{\sqrt{3}}\begin{bmatrix}-3/4\\1/4\\1/4\\1/4\end{bmatrix}, \quad \text{and} \quad \mathbf{u}_3 = \frac{\sqrt{3}}{\sqrt{2}}\begin{bmatrix}0\\-2/3\\1/3\\1/3\end{bmatrix}.$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ forms an orthonormal set and

$$\text{Span}\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \text{Span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}.$$

Let $A = \begin{bmatrix}\mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3\end{bmatrix} = \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\1&1&1\end{bmatrix}$. Then

$$A = \begin{bmatrix}1&0&0\\1&1&0\\1&1&1\\1&1&1\end{bmatrix} = \underbrace{\begin{bmatrix}1/2 & -3/2\sqrt{3} & 0\\1/2 & 1/2\sqrt{3} & -2/\sqrt{6}\\1/2 & 1/2\sqrt{3} & 1/\sqrt{6}\\1/2 & 1/2\sqrt{3} & 1/\sqrt{6}\end{bmatrix}}_{Q}\underbrace{\begin{bmatrix}2 & 3/2 & 1\\0 & \sqrt{3}/2 & 1/\sqrt{3}\\0 & 0 & \sqrt{2}/\sqrt{3}\end{bmatrix}}_{R} = QR.$$

Here $Q = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \mathbf{u}_3\end{bmatrix}$ has orthonormal columns and $R = [r_{ij}]$ is an invertible upper triangular matrix with

$$r_{ii} = \|\mathbf{y}_i\| \quad \text{and} \quad r_{ij} = \langle \mathbf{u}_i, \mathbf{x}_j\rangle \quad \text{for } i \le j.$$

QR Factorization

For any $m \times n$ matrix A, A has a factorization

$$A = QR,$$

where Q is an $m \times n$ matrix with orthonormal columns (i.e., $Q^TQ = I_n$) and R is an $n \times n$ upper triangular matrix. If the columns of A are linearly independent, then R is invertible.

Gram-Schmidt Algorithm

[1]:
```
import sympy as sp
import numpy as np
```

[2]:
```
A = sp.Matrix([[1,0,0], [1,1,0], [1,1,1], [1,1,1]]); A
```

[2]:
$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

[3]:
```
# QR decomposition @ SymPy
Q,R = A.QRdecomposition()
display(Q,R)
```

$$\begin{bmatrix} \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \\ \frac{1}{2} & \frac{\sqrt{3}}{6} & -\frac{\sqrt{6}}{3} \\ \frac{1}{2} & \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{6} \\ \frac{1}{2} & \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{6} \end{bmatrix}$$

$$\begin{bmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{3} \\ 0 & 0 & \frac{\sqrt{6}}{3} \end{bmatrix}$$

[4]:
```
Q@R
```

[4]:
$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Gram-Schmidt Algorithm

[5]:
```
Q.T@Q
```

[5]:
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

[6]:
```
A = np.array([[1,0,0], [1,1,0], [1,1,1], [1,1,1]]); A
```

[6]:
```
array([[1, 0, 0],
       [1, 1, 0],
       [1, 1, 1],
       [1, 1, 1]])
```

[7]:
```
# QR decomposition @ NumPy
Q,R = np.linalg.qr(A)
display(Q,R)
```

```
array([[-0.5       ,  0.8660254 ,  0.        ],
       [-0.5       , -0.28867513,  0.81649658],
       [-0.5       , -0.28867513, -0.40824829],
       [-0.5       , -0.28867513, -0.40824829]])

array([[-2.        , -1.5       , -1.        ],
       [ 0.        , -0.8660254 , -0.57735027],
       [ 0.        ,  0.        , -0.81649658]])
```

Gram-Schmidt Algorithm

[8]:
```
Q@R
```

[8]:
```
array([[ 1.00000000e+00,  8.69063787e-17, -1.34358683e-16],
       [ 1.00000000e+00,  1.00000000e+00,  1.61842956e-16],
       [ 1.00000000e+00,  1.00000000e+00,  1.00000000e+00],
       [ 1.00000000e+00,  1.00000000e+00,  1.00000000e+00]])
```

[9]:
```
Q.T@Q
```

[9]:
```
array([[1.00000000e+00, 0.00000000e+00, 0.00000000e+00],
       [0.00000000e+00, 1.00000000e+00, 6.34802232e-18],
       [0.00000000e+00, 6.34802232e-18, 1.00000000e+00]])
```

[10]:
```
# QR decomposition (complete form) @ NumPy
Q,R = np.linalg.qr(A,mode='complete')
display(Q,R)
```

```
array([[-5.00000000e-01,  8.66025404e-01,  0.00000000e+00, 2.14690125e-18],
       [-5.00000000e-01, -2.88675135e-01,  8.16496581e-01, -8.68448956e-17],
       [-5.00000000e-01, -2.88675135e-01, -4.08248290e-01, -7.07106781e-01],
       [-5.00000000e-01, -2.88675135e-01, -4.08248290e-01, 7.07106781e-01]])

array([[-2.        , -1.5       , -1.        ],
       [ 0.        , -0.8660254 , -0.57735027],
       [ 0.        ,  0.        , -0.81649658],
       [ 0.        ,  0.        ,  0.        ]])
```


Orthonormal basis

Orthonormal basis

Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for a finite dimensional vector space V. Suppose

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{u}_j \quad \text{or equivalently,} \quad [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Then

$$c_j = \langle \mathbf{v}, \mathbf{u}_j \rangle, \quad \|\mathbf{v}\|^2 = \sum_{j=1}^{n} c_j^2, \quad \text{and} \quad [\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B}}^T \mathbf{v},$$

where $P_{\mathcal{B}} = P_{\mathcal{E}\leftarrow\mathcal{B}} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Furthermore, if $\mathbf{w} = \sum_{j=1}^{n} d_j \mathbf{u}_j$, then

$$d_j = \langle \mathbf{w}, \mathbf{u}_j \rangle \quad \text{and} \quad \langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} c_j d_j.$$

Orthonormal basis

Orthogonal set $\Longrightarrow$ Linearly independent set

If $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ is an orthogonal set, then $\mathcal{B}$ is linearly independent.

Proof. Exercise.

Orthonormal basis

Let $\mathcal{B} = \{\mathbf{u}_1, \ldots, \mathbf{u}_n\}$ be an orthonormal basis for a finite dimensional vector space V. Suppose

$$\mathbf{v} = \sum_{j=1}^{n} c_j \mathbf{u}_j \quad \text{or equivalently,} \quad [\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix}.$$

Then

$$c_j = \langle \mathbf{v}, \mathbf{u}_j \rangle, \quad \|\mathbf{v}\|^2 = \sum_{j=1}^{n} c_j^2, \quad \text{and} \quad [\mathbf{v}]_{\mathcal{B}} = P_{\mathcal{B}}^T \mathbf{v},$$

where $P_{\mathcal{B}} = P_{\mathcal{E} \leftarrow \mathcal{B}} = \begin{bmatrix} \mathbf{u}_1 & \cdots & \mathbf{u}_n \end{bmatrix}$. Furthermore, if $\mathbf{w} = \sum_{j=1}^{n} d_j \mathbf{u}_j$, then

$$d_j = \langle \mathbf{w}, \mathbf{u}_j \rangle \quad \text{and} \quad \langle \mathbf{v}, \mathbf{w} \rangle = \sum_{j=1}^{n} c_j d_j.$$

Orthonormal basis

Proof. For any $j = 1, \ldots, n$,

$$\langle \mathbf{v}, \mathbf{u}_j \rangle = \left\langle \sum_{k=1}^{n} c_k \mathbf{u}_k, \mathbf{u}_j \right\rangle = \sum_{k=1}^{n} c_k \langle \mathbf{u}_k, \mathbf{u}_j \rangle = c_j \langle \mathbf{u}_j, \mathbf{u}_j \rangle = c_j.$$

Thus,

$$[\mathbf{v}]_{\mathcal{B}} = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \langle \mathbf{v}, \mathbf{u}_1 \rangle \\ \vdots \\ \langle \mathbf{v}, \mathbf{u}_n \rangle \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \mathbf{v} \\ \vdots \\ \mathbf{u}_n^T \mathbf{v} \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \mathbf{v} = P_{\mathcal{B}}^T \mathbf{v}.$$

Similarly, if $\mathbf{w} = \sum_{j=1}^{n} d_j \mathbf{u}_j$, then $d_j = \langle \mathbf{w}, \mathbf{u}_j \rangle$. Now

$$\langle \mathbf{v}, \mathbf{w} \rangle = \left\langle \sum_{j=1}^{n} c_j \mathbf{u}_j, \sum_{k=1}^{n} d_k \mathbf{u}_k \right\rangle = \sum_{j=1}^{n} \sum_{k=1}^{n} c_j d_k \langle \mathbf{u}_j, \mathbf{u}_k \rangle = \sum_{j=1}^{n} c_j d_j \langle \mathbf{u}_j, \mathbf{u}_j \rangle = \sum_{j=1}^{n} c_j d_j.$$

Finally, $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \sum_{j=1}^{n} c_j^2$.

Orthonormal basis

Example 2.3 Let $\mathbf{u}_1 = \begin{bmatrix} \frac{1}{3\sqrt{2}} \\ \frac{1}{3\sqrt{2}} \\ \frac{-4}{3\sqrt{2}} \end{bmatrix}$, $\mathbf{u}_2 = \begin{bmatrix} \frac{2}{3} \\ \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$, $\mathbf{u}_3 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{-1}{\sqrt{2}} \\ 0 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, and $\mathbf{y} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$. Show that $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ forms an orthonormal basis for $\mathbb{R}^3$. Find the coordinate vectors $[\mathbf{x}]_{\mathcal{B}}$ and $[\mathbf{y}]_{\mathcal{B}}$.

Solution. Note that

$$\|\mathbf{u}_1\| = \|\mathbf{u}_2\| = \|\mathbf{u}_3\| = 1 \quad \text{and} \quad \langle \mathbf{u}_1, \mathbf{u}_2 \rangle = \langle \mathbf{u}_1, \mathbf{u}_3 \rangle = \langle \mathbf{u}_2, \mathbf{u}_3 \rangle = 0.$$

So $\mathcal{B} = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ forms an orthonormal basis for $\mathbb{R}^3$. Now let $P_{\mathcal{B}} = \begin{bmatrix} \frac{1}{3\sqrt{2}} & \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{1}{3\sqrt{2}} & \frac{2}{3} & \frac{-1}{\sqrt{2}} \\ \frac{-4}{3\sqrt{2}} & \frac{1}{3} & 0 \end{bmatrix}$. Then

$$[\mathbf{x}]_{\mathcal{B}} = P_{\mathcal{B}}^T \mathbf{x} = \begin{bmatrix} \frac{1}{3\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{-4}{3\sqrt{2}} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -\frac{2}{3\sqrt{2}} \\ \frac{5}{3} \\ 0 \end{bmatrix}.$$

$$[\mathbf{y}]_{\mathcal{B}} = P_{\mathcal{B}}^T \mathbf{y} = \begin{bmatrix} \frac{1}{3\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{-4}{3\sqrt{2}} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} -\frac{9}{3\sqrt{2}} \\ 3 \\ -\frac{1}{\sqrt{2}} \end{bmatrix}.$$

Orthonormal basis

[1]:
```python
import sympy as sp
from sympy import sqrt
```

[2]:
```python
PB = sp.Matrix([[1/sqrt(2)/3,'2/3',1/sqrt(2)], [1/sqrt(2)/3,'2/3',-1/sqrt(2)], [-4/sqrt(2)/3,'1/3',0]]); PB
```

[2]: $\begin{bmatrix} \frac{\sqrt{2}}{6} & \frac{2}{3} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{6} & \frac{2}{3} & -\frac{\sqrt{2}}{2} \\ -\frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \end{bmatrix}$

[3]:
```python
# Compute all inner products <ui,uj>
PB.T@PB
```

[3]: $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

[4]:
```python
x = sp.Matrix([[1,1,1]]).T
y = sp.Matrix([[1,2,3]]).T
display(x,y)
```

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

$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

[5]:
```python
# Coordinate vectors
PB.T@x
```

[5]: $\begin{bmatrix} -\frac{\sqrt{2}}{3} \\ \frac{5}{3} \\ 0 \end{bmatrix}$

[6]:
```python
PB.T@y
```

[6]: $\begin{bmatrix} -\frac{3\sqrt{2}}{2} \\ 3 \\ -\frac{\sqrt{2}}{2} \end{bmatrix}$

Orthonormal basis

Example 2.4 The two vectors $\mathbf{u}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\1\\-1\end{bmatrix}$ and $\mathbf{u}_2 = \frac{1}{6}\begin{bmatrix}1\\1\\3\\5\end{bmatrix}$ form an orthonormal set in $\mathbb{R}^4$. Extend this set to an orthornormal basis for $\mathbb{R}^4$.

Solution. Consider the matrix

$$A = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2\end{bmatrix} = \begin{bmatrix}\frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{3}{6}\\ -\frac{1}{2} & \frac{5}{6}\end{bmatrix}.$$

Then direct computations yield that $\mathbf{x}_1 = \begin{bmatrix}-1\\1\\0\\0\end{bmatrix}$ and $\mathbf{x}_2 = \begin{bmatrix}4\\0\\-3\\1\end{bmatrix}$ form a basis for $\operatorname{Nul} A^T = (\operatorname{Col} A)^\perp$. Thus, $\{\mathbf{x}_1, \mathbf{x}_2\}$ is orthogonal to $\{\mathbf{u}_1, \mathbf{u}_2\}$. Now apply Gram-Schmidt Algorithm on $\{\mathbf{x}_1, \mathbf{x}_2\}$, we obtain

$$\mathbf{u}_3 = \frac{1}{\sqrt{2}}\begin{bmatrix}-1\\1\\0\\0\end{bmatrix} \quad \text{and} \quad \mathbf{u}_4 = \frac{1}{3\sqrt{2}}\begin{bmatrix}2\\2\\-3\\1\end{bmatrix}.$$

Then $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4\}$ forms an orthornormal basis for $\mathbb{R}^4$.

Orthonormal basis

[1]:
```
A = sp.Matrix([['1/2','1/2','1/2','-1/
↪2'], ['1/6','1/6','3/6','5/6']]).T; A
```

[1]:
$$\begin{bmatrix}\frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{6}\\ \frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{5}{6}\end{bmatrix}$$

[2]:
```
# Nullspace of A.T
A.T.nullspace(); B
```

[2]:
```
[Matrix([
 [-1],
 [ 1],
 [ 0],
 [ 0]]),
 Matrix([
 [ 4],
 [ 0],
 [-3],
 [ 1]])]
```

[3]:
```
# Matrix with basis vectors for null␣
↪space as columns
B = sp.BlockMatrix(B).as_explicit(); B
```

[3]:
$$\begin{bmatrix}-1 & 4\\ 1 & 0\\ 0 & -3\\ 0 & 1\end{bmatrix}$$

[4]:
```
# QR Factorization
Q,R = B.QRdecomposition()
display(Q,R)
```

$$\begin{bmatrix}-\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{3}\\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{3}\\ 0 & -\frac{\sqrt{2}}{2}\\ 0 & \frac{\sqrt{2}}{6}\end{bmatrix}$$

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

Orthonormal basis

Orthogonal matrices

An $n \times n$ square matrix U is said to be **orthogonal** if

$$U^T U = U U^T = I_n.$$

That is, U^T is the inverse of U.

Example The following matrices are orthogonal matrices.

$$\begin{bmatrix} \sqrt{3}/2 & 1/2 \\ -1/2 & \sqrt{3}/2 \end{bmatrix}, \quad \begin{bmatrix} \frac{1}{3\sqrt{2}} & \frac{2}{3} & \frac{1}{\sqrt{2}} \\ \frac{1}{3\sqrt{2}} & \frac{2}{3} & \frac{-1}{\sqrt{2}} \\ \frac{-4}{3\sqrt{2}} & \frac{1}{3} & 0 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{12} & 1/2 \\ 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{12} & 1/2 \\ 0 & 2/\sqrt{6} & -1/\sqrt{12} & 1/2 \\ 0 & 0 & 3/\sqrt{12} & 1/2 \end{bmatrix}.$$

Some properties for orthogonal matrices

For an $n \times n$ orthogonal matrix U,

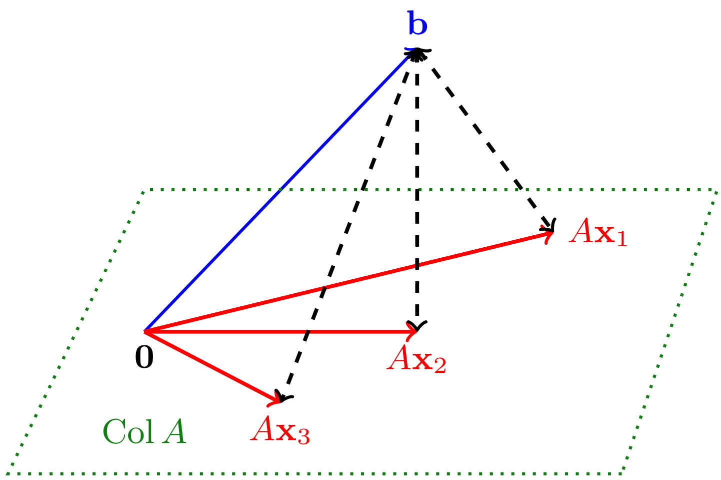
1. U^T is the inverse of U;
2. $\operatorname{Col} U = \mathbb{R}^n$ and $\operatorname{Row} U = \mathbb{R}^{1\times n}$;
3. Column vectors of U form an orthonormal basis in $\mathbb{R}^n$;
4. Row vectors of U form an orthonormal basis in $\mathbb{R}^{1\times n}$;
5. $\det(U) = \pm 1$.

Least Squares Solution

Suppose A is $m \times n$ and $A\mathbf{x} = \mathbf{b}$ is **inconsistent**.

- Find the 'nearest' solution of $A\mathbf{x} = \mathbf{b}$.
- Find a solution $\mathbf{x}$ such that

  $\|\mathbf{b} - A\mathbf{x}\|$ is minimum.



Consider the normal equation

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

For any $\mathbf{b} \in \mathbb{R}^m$,

$$A^T\mathbf{b} \in \operatorname{Col} A^T = \operatorname{Col} A^T A \quad \Longrightarrow \quad \text{there exists } \mathbf{x} \in \mathbb{R}^n \text{ such that } A^T A\mathbf{x} = A^T\mathbf{b}.$$

Therefore, the normal equation $A^T A\mathbf{x} = A^T\mathbf{b}$ is always consistent!

Least Squares Solution

Suppose $\hat{\mathbf{x}}$ is a solution to the equation $A^TA\mathbf{x} = A^T\mathbf{b}$. Then

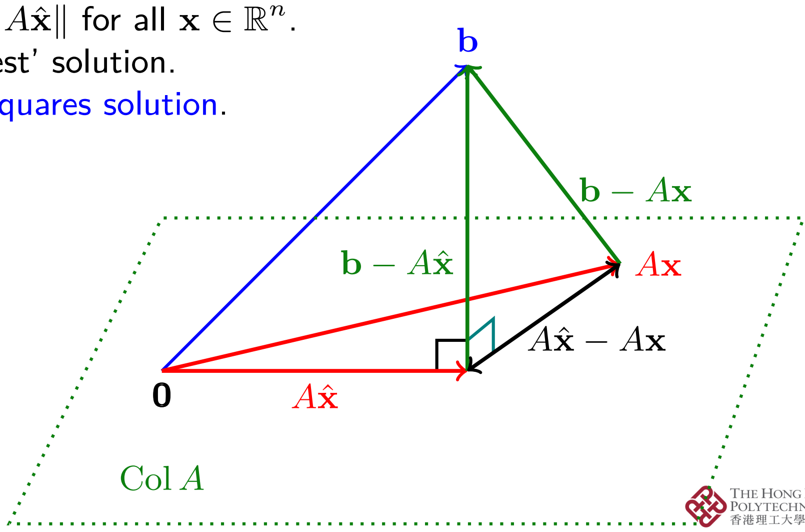
$$A^T(\mathbf{b} - A\hat{\mathbf{x}}) = A^T\mathbf{b} - A^TA\hat{\mathbf{x}} = A^T\mathbf{b} - A^T\mathbf{b} = \mathbf{0},$$

and so $\mathbf{b} - A\hat{\mathbf{x}} \in \mathrm{Nul}\,A^T = (\mathrm{Col}\,A)^\perp$.

On the other hand, any vector $\mathbf{x} \in \mathbb{R}^n$, $A(\hat{\mathbf{x}} - \mathbf{x}) \in \mathrm{Col}\,A$. Then

$$\begin{aligned} \|\mathbf{b} - A\mathbf{x}\|^2 &= \|(\mathbf{b} - A\hat{\mathbf{x}}) + (A\hat{\mathbf{x}} - A\mathbf{x})\|^2 \\ &= \|\mathbf{b} - A\hat{\mathbf{x}}\|^2 + \|A\hat{\mathbf{x}} - A\mathbf{x}\|^2 \geq \|\mathbf{b} - A\hat{\mathbf{x}}\|^2. \end{aligned}$$

Thus, $\|\mathbf{b} - A\mathbf{x}\| \geq \|\mathbf{b} - A\hat{\mathbf{x}}\|$ for all $\mathbf{x} \in \mathbb{R}^n$.
Therefore, $\hat{\mathbf{x}}$ is the 'nearest' solution.
This is so called a least squares solution.

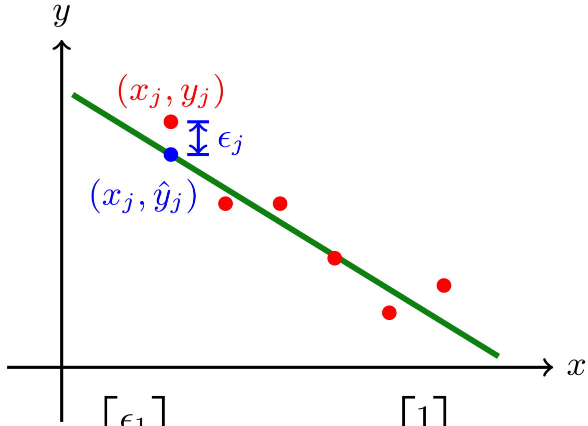


Least Squares Solution

- Suppose A^TA is invertible. Then the normal equation has the unique solution
$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b}.$$
Recall that
$$A\hat{\mathbf{x}} = A(A^TA)^{-1}A^T\mathbf{b}$$
is the projection of $\mathbf{b}$ onto the column space of A.
- Suppose A^TA is not invertible. Then the normal equation has infinite many solutions.
- If A has a QR-factorization, i.e.,$A = QR$, then
$$R^TQ^TQR\mathbf{x} = R^TQ^T\mathbf{b} \quad \Longrightarrow \quad Q^TQR\mathbf{x} = Q^T\mathbf{b} \quad \Longrightarrow \quad R\mathbf{x} = Q^T\mathbf{b}.$$
- If A has orthogonal columns, then $A^TA = I_n$ and
$$\hat{\mathbf{x}} = A^T\mathbf{b}.$$

Linear regression

Given a collection of data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.



Suppose

$$y_j = \alpha + \beta x_j + \epsilon_j \quad \text{for } j = 1, \ldots, n.$$

Set

$$\hat{y}_j = \alpha + \beta x_j \quad \text{for } j = 1, \ldots, n.$$

Define

$$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \hat{\mathbf{y}} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}, \quad \text{and} \quad \mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Then

$$\hat{\mathbf{y}} = \alpha\mathbf{e} + \beta\mathbf{x} = \underbrace{\begin{bmatrix} \mathbf{e} & \mathbf{x} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\mathbf{u}} = \underbrace{\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \alpha \\ \beta \end{bmatrix}}_{\mathbf{u}} = A\mathbf{u} \quad \text{and} \quad \boldsymbol{\epsilon} = \mathbf{y} - \hat{\mathbf{y}}.$$

Linear regression

Here we would like to

$$\text{minimize } \|\boldsymbol{\epsilon}\| = \|\mathbf{y} - \hat{\mathbf{y}}\| = \|\mathbf{y} - A\mathbf{u}\|.$$

Then the least squares solution is

$$\hat{\mathbf{u}} = (A^T A)^{-1} A^T \mathbf{y}.$$

Let

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad \bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j,$$

$$S_{xy} = \sum_{j=1}^{n}(x_j - \bar{x})(y_j - \bar{y}) = \left(\sum_{j=1}^{n} x_j y_j\right) - n\bar{x}\bar{y}$$

$$S_{xx} = \sum_{j=1}^{n}(x_j - \bar{x})^2 = \left(\sum_{j=1}^{n} x_j^2\right) - n\bar{x}^2.$$

Then

$$\mathbf{e}^T\mathbf{e} = n, \quad \mathbf{e}^T\mathbf{x} = n\bar{x}, \quad \mathbf{e}^T\mathbf{y} = n\bar{y}, \quad \mathbf{x}^T\mathbf{x} = S_{xx} + n\bar{x}^2, \quad \mathbf{x}^T\mathbf{y} = S_{xy} + n\bar{x}\bar{y}.$$

Linear regression

It follows that

$$
\begin{aligned}
A^T\mathbf{y} &= \begin{bmatrix}\mathbf{e}^T\\ \mathbf{x}^T\end{bmatrix}\mathbf{y} = \begin{bmatrix}\mathbf{e}^T\mathbf{y}\\ \mathbf{x}^T\mathbf{y}\end{bmatrix} = \begin{bmatrix} n\bar{y}\\ S_{xy}+n\bar{x}\bar{y}\end{bmatrix},\\
(A^TA)^{-1} &= \left(\begin{bmatrix}\mathbf{e}^T\\ \mathbf{x}^T\end{bmatrix}\begin{bmatrix}\mathbf{e} & \mathbf{x}\end{bmatrix}\right)^{-1} = \begin{bmatrix}\mathbf{e}^T\mathbf{e} & \mathbf{e}^T\mathbf{x}\\ \mathbf{x}^T\mathbf{e} & \mathbf{x}^T\mathbf{x}\end{bmatrix}^{-1}\\
&= \begin{bmatrix} n & n\bar{x}\\ n\bar{x} & S_{xx}+n\bar{x}^2\end{bmatrix}^{-1} = \frac{1}{nS_{xx}}\begin{bmatrix} S_{xx}+n\bar{x}^2 & -n\bar{x}\\ -n\bar{x} & n\end{bmatrix}\\
\hat{\mathbf{u}} = (A^TA)^{-1}A^T\mathbf{y} &= \frac{1}{nS_{xx}}\begin{bmatrix} S_{xx}+n\bar{x}^2 & -n\bar{x}\\ -n\bar{x} & n\end{bmatrix}\begin{bmatrix} n\bar{y}\\ S_{xy}+n\bar{x}\bar{y}\end{bmatrix}\\
&= \frac{1}{nS_{xx}}\begin{bmatrix} n\bar{y}S_{xx}-n\bar{x}S_{xy}\\ nS_{xy}\end{bmatrix} = \frac{1}{S_{xx}}\begin{bmatrix}\bar{y}S_{xx}-\bar{x}S_{xy}\\ S_{xy}\end{bmatrix}.
\end{aligned}
$$

That is,

$$\hat{\beta} = \frac{S_{xy}}{S_{xx}} \quad \text{and} \quad \hat{\alpha} = \frac{\bar{y}S_{xx}-\bar{x}S_{xy}}{S_{xx}} = \bar{y}-\bar{x}\cdot\frac{S_{xy}}{S_{xx}} = \bar{y}-\hat{\beta}\bar{x}.$$

The equation $y = \hat{\alpha}+\hat{\beta}x+\epsilon$ is known as the simple linear regression model.
Notice that

$$\bar{y} = \hat{\alpha}+\hat{\beta}\bar{x}.$$

Inner product space

Inner product space

An inner product on a vector space V is an operation $\langle\cdot,\cdot\rangle$ on V that satisfies the following conditions.
For any $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{w}$ in V and any constant c,

1. $\langle\mathbf{x},\mathbf{y}\rangle = \overline{\langle\mathbf{y},\mathbf{x}\rangle}$;
2. $\langle\mathbf{x}+\mathbf{y},\mathbf{w}\rangle = \langle\mathbf{x},\mathbf{w}\rangle+\langle\mathbf{y},\mathbf{w}\rangle$;
3. $\langle c\mathbf{x},\mathbf{y}\rangle = c\langle\mathbf{x},\mathbf{y}\rangle$;
4. $\langle\mathbf{x},\mathbf{x}\rangle \geq 0$.

A vector space with an inner product is called an inner product space.

Examples

1. $V=\mathbb{R}^n$ with $\langle\mathbf{x},\mathbf{y}\rangle = \mathbf{y}^T\mathbf{x} = \displaystyle\sum_{j=1}^n x_jy_j$.
2. $V=\mathbb{R}^n$ with $\langle\mathbf{x},\mathbf{y}\rangle = \displaystyle\sum_{j=1}^n w_jx_jy_j$,
   where $w_1,\ldots,w_n$ are positive real numbers.

Inner product space

3. $V = \mathbb{C}^n$ with $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^* \mathbf{x} = \sum_{j=1}^{n} x_j \bar{y}_j$.

4. $V = \mathbb{R}^{m \times n}$ with $\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$.

5. $V = \mathbb{C}^{m \times n}$ with $\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \bar{b}_{ij}$.

6. $V = \mathbb{P}_n$ with $\langle p, q \rangle = \sum_{i=1}^{n} p(x_i) q(x_i)$,

   where $x_1, \ldots, x_n$ are distinct real numbers.

7. $V = C[a, b]$ with $\langle f, g \rangle = \int_a^b f(x) g(x)\, dx$.

8. $V = C[a, b]$ with $\langle f, g \rangle = \int_a^b w(x) f(x) g(x)\, dx$,

   where $w(x)$ is a positive continuous function on $[a, b]$.