Algebraic Basis in Control Theory
控制理论中的代数基础

哈尔滨工业大学（深圳）

李衍杰

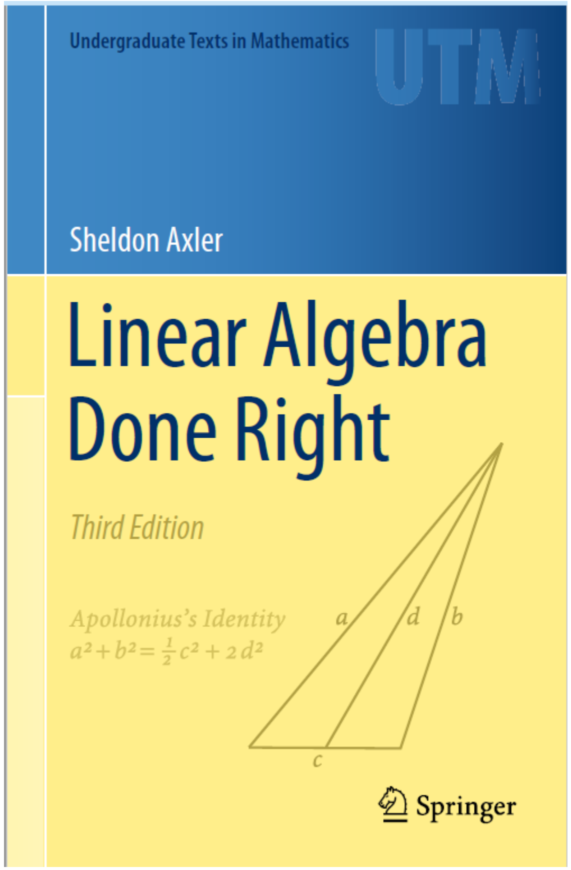
Contact Information

- 李衍杰
  - Room 1011, Building G
  - Phone: 86530531
  - Email: autolyj@hit.edu.cn
- Teaching Assistant
  - 马帅康，陈鹏彬，鄢柯剑，李心怡
  - Room 415， Building A
  - QQ群：747589654

群名称:23春-控制理论中的代数基础

群　号:747589654

Reference Book



没有一种数学思想如当初刚被发现时那样发表出来。一旦问题解决了，思考的程序便颠倒过来，把火热的思考变成冰冷的美丽。

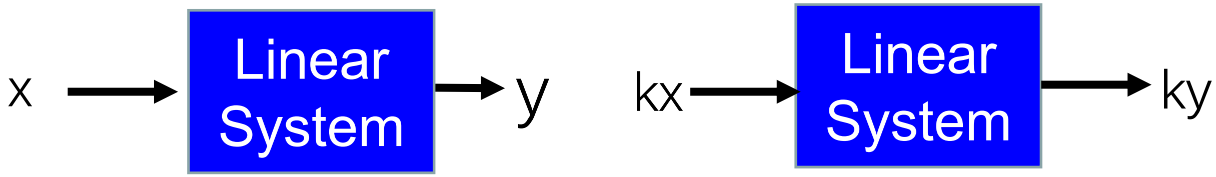
——————————弗赖登塔尔

课程目的

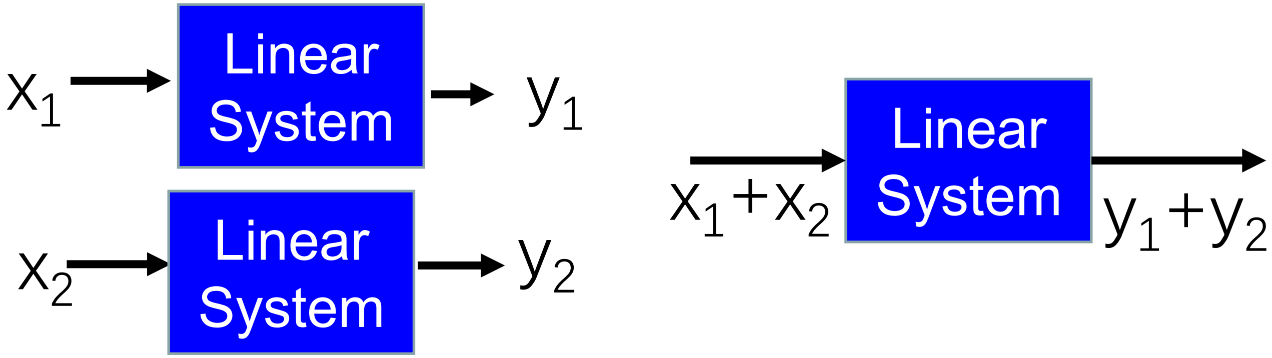
线性代数变为了计算科学，离散的概念需要具体的含义

Linear System

Persevering Multiplication



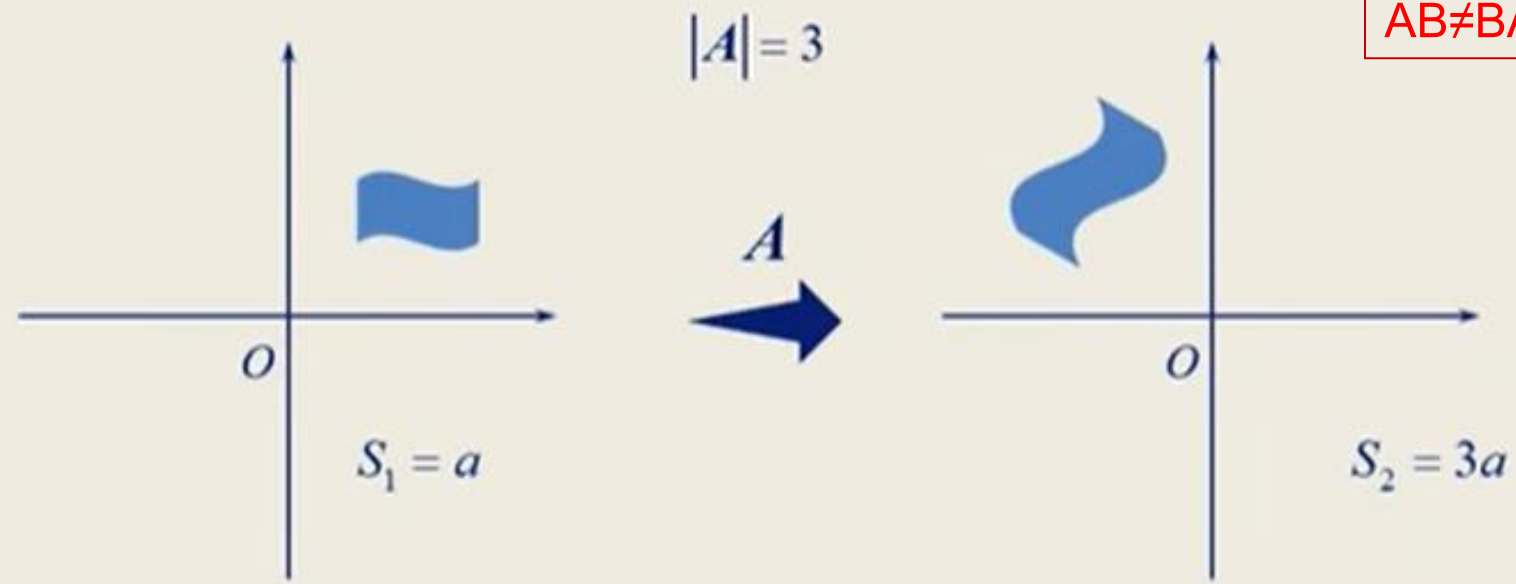
Persevering Addition



Hung-yi Lee

矩阵的行列式

AB≠BA？？



$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$

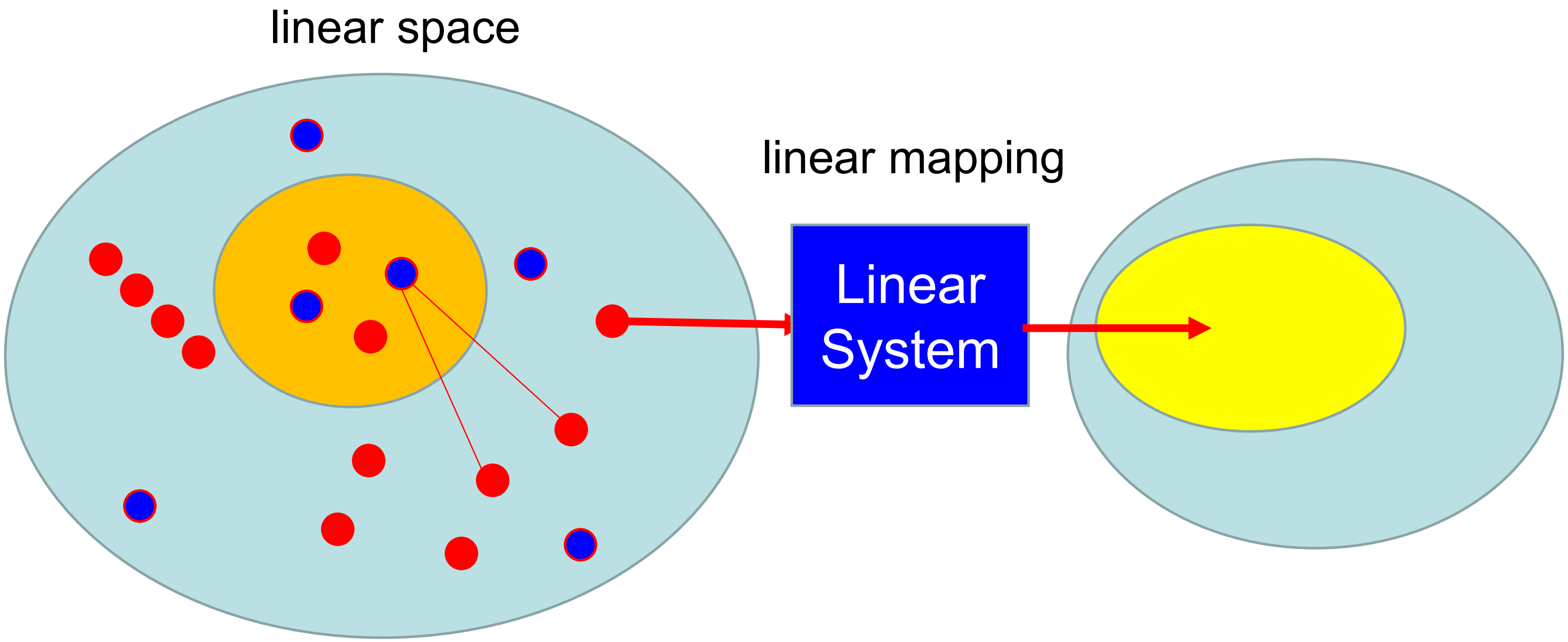
$|A| = 6$

如何理解 $|AB| = |A||B|$

AB 表示依次施加变换 B 和变换 A,

先后放大 $|B|$倍和 $|A|$倍,

总共放大 $|A||B|$倍.



An intuitive map for Linear Algebra

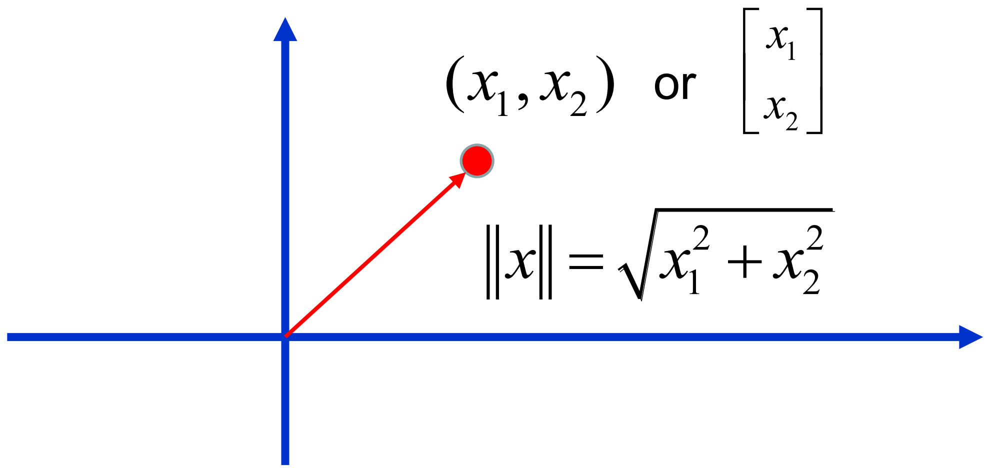
In making the definition of a vector space, we generalized the linear structure (addition and scalar multiplication) of R^2 and R^3.

We ignored other important features, such as the notions of length and angle.

Inner Product Space

Euclidean distance

R^2 space



R^n space

Euclidean distance

We choose a point ***0*** as origin in real ***n*** -dimensional Euclidean space: the *length* of any vector ***x*** in space, denoted as $\|x\|$, is defined as its *distance* to the origin.

Denote the Cartesian coordinates of ***x*** as $x_1, \ldots, x_n$

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Norm & Distance

- **Norm**: Norm of vector v is the length of v
  - Denoted $||v||$

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

- **Distance**: The distance between two vectors u and v is defined by $||v - u||$

$$v = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad u = \begin{bmatrix} 2 \\ -3 \\ 0 \end{bmatrix} \quad v - u = \begin{bmatrix} -1 \\ 5 \\ 3 \end{bmatrix} \quad ||v - u|| = \sqrt{(-1)^2 + 5^2 + 3^2}$$

$$= \sqrt{35}$$

$$||v|| = \sqrt{1^2 + 2^2 + 3^2} \quad = \sqrt{14}$$

dot product
scalar product
Inner product

6.2 **Definition** ***dot product*** （Scalar product）

For $x, y \in \mathbf{R}^n$, the ***dot product*** of x and y, denoted $x \cdot y$, is defined by

$$x \cdot y = x_1 y_1 + \cdots + x_n y_n,$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$.

Standard inner product

real vector space

- $x \cdot x \geq 0$ for all $x \in \mathbf{R}^n$;
- $x \cdot x = 0$ if and only if $x = 0$;
- for $y \in \mathbf{R}^n$ fixed, the map from $\mathbf{R}^n$ to $\mathbf{R}$ that sends $x \in \mathbf{R}^n$ to $x \cdot y$ is linear;
- $x \cdot y = y \cdot x$ for all $x, y \in \mathbf{R}^n$.

Definition. A Euclidean structure in a linear space $\boldsymbol{X}$ over $\boxed{\text{the reals}}$ is furnished by a real-valued function of two vector arguments called a *scalar product* and denoted as $(\boldsymbol{x}, \boldsymbol{y})$, which has the following properties:

(i) $(\boldsymbol{x}, \boldsymbol{y})$ is a bilinear function; that is, it is a linear function of each argument when the other is kept fixed.

(ii) It is symmetric:

$$(\boldsymbol{x}, \boldsymbol{y}) = (\boldsymbol{y}, \boldsymbol{x}).$$

(iii) It is positive:

$$(\boldsymbol{x}, \boldsymbol{x}) > \boldsymbol{0} \quad \text{except for } \boldsymbol{x} = \boldsymbol{0}.$$

Examples

Standard inner product $(x, y) = x^T y$

1. Consider another inner product on R^n

$< x, y > = x^T M y$ Where ***M*** is symmetric positive-define matrix

$$M = \begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}$$

$$\langle x, y \rangle = x_1 y_1 + \frac{1}{2} x_1 y_2 + \frac{1}{2} x_2 y_1 + x_2 y_2$$

2. The vector space of real functions whose domain is an closed interval [a, b] with inner product

$$< f, g >= \int_a^b f(x)g(x)dx$$

Consider a vector space of polynomials with degree less than 2 and its domain is [0,1] we have

$$< s - \frac{\sqrt{3}}{3}, s + \frac{\sqrt{3}}{3} >= \int_0^1 s^2 - 1/3 ds = 0$$

3. Consider matrix linear space $R^{m \times n}$

Given A and B in $\mathbb{R}^{m \times n}$, we can define an inner product by

$$\langle A, B \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

complex vector spaces

Recall that if $\lambda = a + bi$, where $a, b \in \mathbf{R}$, then

- the absolute value of λ, denoted $|\lambda|$, is defined by $|\lambda| = \sqrt{a^2 + b^2}$;
- the complex conjugate of λ, denoted $\bar{\lambda}$, is defined by $\bar{\lambda} = a - bi$;
- $|\lambda|^2 = \lambda\bar{\lambda}$.

For $z = (z_1, \ldots, z_n) \in \mathbf{C}^n$, we define the norm of z by

$$\|z\| = \sqrt{|z_1|^2 + \cdots + |z_n|^2}.$$

$$\|z\|^2 = z_1\overline{z_1} + \cdots + z_n\overline{z_n}$$

the inner product of

$w = (w_1, \ldots, w_n) \in \mathbf{C}^n$ with z should equal

① $w \cdot z = w_1\overline{z_1} + \cdots + w_n\overline{z_n}$

$$w \cdot z = wz^H \qquad \begin{aligned} w &= (w_1, \ldots, w_n) \\ z &= (z_1, \ldots, z_n) \end{aligned}$$

② $w \cdot z = \overline{w_1}z_1 + \cdots + \overline{w_n}z_n$

$$w \cdot z = \bar{w}z^T \qquad \begin{aligned} w &= (w_1, \ldots, w_n) \\ z &= (z_1, \ldots, z_n) \end{aligned}$$

$$\boxed{w \cdot z = \overline{z \cdot w}}$$

Fields of Complex number

An ***inner product*** on V is a function that takes each ordered pair (u, v) of elements of V to a number $\langle u, v \rangle \in \mathbf{F}$ and has the following properties:

positivity

$$\langle v, v \rangle \geq 0 \text{ for all } v \in V;$$

definiteness

$$\langle v, v \rangle = 0 \text{ if and only if } v = 0;$$

additivity in first slot

$$\langle u + v, w \rangle = \langle u, w \rangle + \langle v, w \rangle \text{ for all } u, v, w \in V;$$

homogeneity in first slot

$$\langle \lambda u, v \rangle = \lambda \langle u, v \rangle \text{ for all } \lambda \in \mathbf{F} \text{ and all } u, v \in V;$$

conjugate symmetry

$$\langle u, v \rangle = \overline{\langle v, u \rangle} \text{ for all } u, v \in V.$$

6.4 **Example** ***inner products***

(a) The ***Euclidean inner product*** on $\mathbf{F}^n$ is defined by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n) \rangle = w_1 \overline{z_1} + \cdots + w_n \overline{z_n}.$$

(b) If $c_1, \dots, c_n$ are positive numbers, then an inner product can be defined on $\mathbf{F}^n$ by

$$\langle (w_1, \dots, w_n), (z_1, \dots, z_n) \rangle = c_1 w_1 \overline{z_1} + \cdots + c_n w_n \overline{z_n}.$$

(c) An inner product can be defined on the vector space of continuous real-valued functions on the interval $[-1, 1]$ by

$$\langle f, g \rangle = \int_{-1}^{1} f(x) g(x) \, dx.$$

(d) An inner product can be defined on $\mathcal{P}(\mathbf{R})$ by

$$\langle p, q \rangle = \int_{0}^{\infty} p(x) q(x) e^{-x} \, dx.$$

6.7 Basic properties of an inner product

(a) For each fixed $u \in V$, the function that takes v to $\langle v, u \rangle$ is a linear map from V to $\mathbf{F}$.

(b) $\langle 0, u \rangle = 0$ for every $u \in V$.

(c) $\langle u, 0 \rangle = 0$ for every $u \in V$.

(d) $\langle u, v + w \rangle = \langle u, v \rangle + \langle u, w \rangle$ for all $u, v, w \in V$.

(e) $\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle$ for all $\lambda \in \mathbf{F}$ and $u, v \in V$.

Proof

(a) Part (a) follows from the conditions of additivity in the first slot and homogeneity in the first slot in the definition of an inner product.

(b) Part (b) follows from part (a) and the result that every linear map takes 0 to 0.

(c) Part (c) follows from part (a) and the conjugate symmetry property in the definition of an inner product.

(d) Suppose $u, v, w \in V$. Then

$$\begin{aligned} \langle u, v + w \rangle &= \overline{\langle v + w, u \rangle} \\ &= \overline{\langle v, u \rangle + \langle w, u \rangle} \\ &= \overline{\langle v, u \rangle} + \overline{\langle w, u \rangle} \\ &= \langle u, v \rangle + \langle u, w \rangle. \end{aligned}$$

(e) Suppose $\lambda \in \mathbf{F}$ and $u, v \in V$. Then

$$\begin{aligned} \langle u, \lambda v \rangle &= \overline{\langle \lambda v, u \rangle} \\ &= \overline{\lambda \langle v, u \rangle} \\ &= \bar{\lambda} \overline{\langle v, u \rangle} \\ &= \bar{\lambda} \langle u, v \rangle, \end{aligned}$$

as desired. ■

6.8 **Definition** ***norm***, $\|v\|$ (Euclidean length)

For $v \in V$, the ***norm*** of v, denoted $\|v\|$, is defined by

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

Example

$$\|f\| = \sqrt{\int_{-1}^{1} \left(f(x)\right)^2 dx.} \qquad \|A\| = \sqrt{\langle A, A \rangle} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2$$

Definition, The distance of two vectors ***x*** and ***y*** in a linear space with Euclidean norm is defined as ||***x*** - ***y***||.

6.10 Basic properties of the norm

Suppose $v \in V$.

(a) $\|v\| = 0$ if and only if $v = 0$.

(b) $\|\lambda v\| = |\lambda| \, \|v\|$ for all $\lambda \in \mathbf{F}$.

Orthogonal and Orthonormal

Definition. Two vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ are called *orthogonal* (*perpendicular*), denoted as $\boldsymbol{x} \perp \boldsymbol{y}$, if

$$(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{0}. \qquad \textbf{(13)}$$

6.12 Orthogonality and 0

(a) 0 is orthogonal to every vector in V.

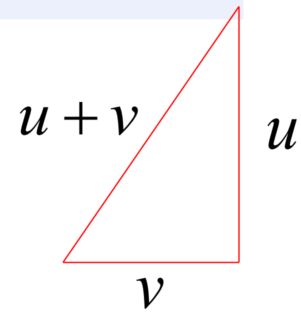
(b) 0 is the only vector in V that is orthogonal to itself.

Equalities and Inequalities

6.13 Pythagorean Theorem （勾股定理）

Suppose u and v are orthogonal vectors in V. Then

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2.$$



Proof We have

$$\begin{aligned}\|u+v\|^2 &= \langle u+v, u+v\rangle \\ &= \langle u,u\rangle + \langle u,v\rangle + \langle v,u\rangle + \langle v,v\rangle \\ &= \|u\|^2 + \|v\|^2,\end{aligned}$$

as desired.

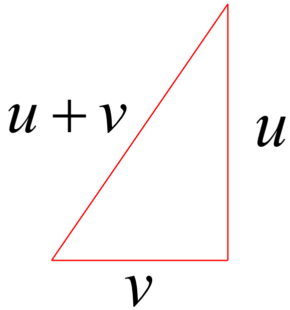
Converse proposition？

6.15 Cauchy–Schwarz Inequality

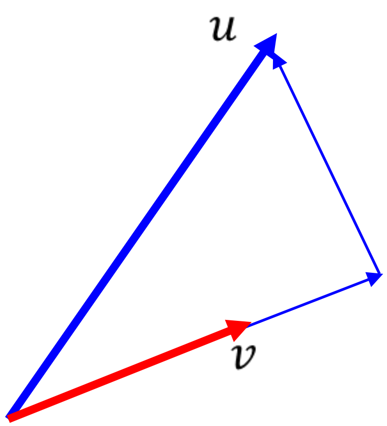
Suppose $u, v \in V$. Then

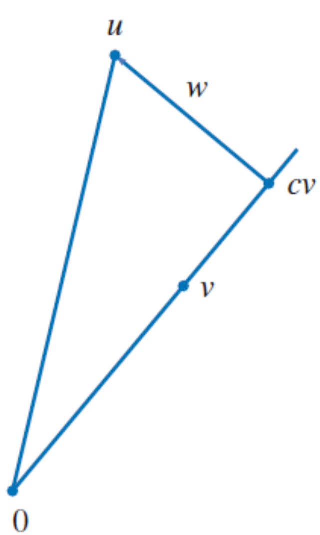
$$|\langle u, v\rangle| \leq \|u\|\,\|v\|.$$

This inequality is an equality if and only if one of u, v is a scalar multiple of the other.



$$u = cv + (u - cv)$$





Thus we need to choose c so that v is orthogonal to $(u - cv)$. In other words, we want

$$0 = \langle u - cv, v\rangle = \langle u, v\rangle - c\|v\|^2.$$

The equation above shows that we should choose c to be $\langle u, v\rangle / \|v\|^2$. Making this choice of c, we can write

$$u = \frac{\langle u, v\rangle}{\|v\|^2} v + \left(u - \frac{\langle u, v\rangle}{\|v\|^2} v\right).$$

6.14 An orthogonal decomposition

Suppose $u, v \in V$, with $v \neq 0$. Set $c = \dfrac{\langle u, v\rangle}{\|v\|^2}$ and $w = u - \dfrac{\langle u, v\rangle}{\|v\|^2}v$. Then

$$\langle w, v\rangle = 0 \quad \text{and} \quad u = cv + w.$$

Proof If $v = 0$, then both sides of the desired inequality equal 0. Thus we can assume that $v \neq 0$. Consider the orthogonal decomposition

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

given by 6.14, where w is orthogonal to v. By the Pythagorean Theorem,

$$\|u\|^2 = \left\| \frac{\langle u, v \rangle}{\|v\|^2} v \right\|^2 + \|w\|^2$$

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

6.16

$$\geq \frac{|\langle u, v \rangle|^2}{\|v\|^2}.$$

Multiplying both sides of this inequality by $\|v\|^2$ and then taking square roots gives the desired inequality.

对于实数域线性空间的另一种证明方法：

Theorem 1 (**Schwarz Inequality**). For all x, y,

$$|(x,y)| \le \|x\|\|y\|. \qquad \textbf{(9)}$$

Proof. Consider the function $q(t)$ of the real variable t defined by

$$q(t) = \|x + ty\|^2. \qquad \textbf{(10)}$$

Using the definition of norm and properties of inner product, we can write

$$q(t) = \|x\|^2 + 2t(x,y) + t^2\|y\|^2. \qquad \textbf{(10)'}$$

Assume that $y \neq 0$ and set $t = -(x,y)/\|y\|^2$ in **(10)'**.

Since **(10)** shows that $q(t) \geq 0$ for all t, we get that

$$\|x\|^2 - \frac{(x,y)^2}{\|y\|^2} \geq 0$$

This proves **(9).** For $y=0$, **(9)** is trivially true.

6.17 **Example** ***examples of the Cauchy–Schwarz Inequality***

(a) If $x_1, \ldots, x_n, y_1, \ldots, y_n \in \mathbf{R}$, then Cauchy(1821)

$$|x_1 y_1 + \cdots + x_n y_n|^2 \leq (x_1{}^2 + \cdots + x_n{}^2)(y_1{}^2 + \cdots + y_n{}^2).$$

(b) If f, g are continuous real-valued functions on $[-1, 1]$, then

$$\left|\int_{-1}^{1} f(x)g(x)\,dx\right|^2 \leq \left(\int_{-1}^{1} (f(x))^2\,dx\right)\left(\int_{-1}^{1} (g(x))^2\,dx\right).$$

Schwartz(1886)

Applications

Let a, b, c be positive real numbers, we have

$$\frac{2}{a+b}+\frac{2}{b+c}+\frac{2}{c+a} \geq \frac{9}{a+b+c}$$

Proof:

$$(a+b+b+c+c+a) \times [\frac{1}{a+b}+\frac{1}{b+c}+\frac{1}{c+a}]$$

$$= \left[\left(\sqrt{a+b}\right)^2+\left(\sqrt{b+c}\right)^2+(\sqrt{a+c})^2\right]\left[\left(\sqrt{\frac{1}{a+b}}\right)^2+\left(\sqrt{\frac{1}{b+c}}\right)^2+\left(\sqrt{\frac{1}{a+c}}\right)^2\right]$$

$$\geq \left(\sqrt{a+b}\sqrt{\frac{1}{a+b}}+\sqrt{b+c}\sqrt{\frac{1}{b+c}}+\sqrt{a+c}\sqrt{\frac{1}{a+c}}\right)^2$$

$$= 9$$

$$\boxed{\begin{aligned}&\frac{1}{x}+\frac{1}{y}+\frac{1}{z} > \frac{9}{x+y+z} \leftrightarrow \frac{x+y+z}{x}+\frac{x+y+z}{y}+\frac{x+y+z}{z} > 9\\ &\leftrightarrow \frac{y}{x}+\frac{z}{x}+\frac{x}{y}+\frac{z}{y}+\frac{x}{z}+\frac{y}{z} > 6 \leftrightarrow (\frac{y}{x}+\frac{x}{y})+(\frac{z}{x}+\frac{x}{z})+(\frac{y}{z}+\frac{z}{y}) > 6\end{aligned}}$$

Applications

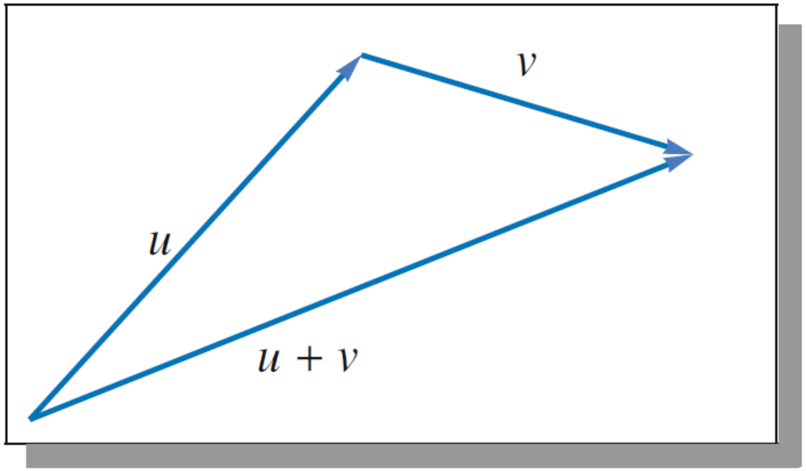
log-sum-exp: $f(x) = \log \sum_{k=1}^{n} \exp x_k$ is convex

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z} \mathbf{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \qquad (z_k = \exp x_k)$$

to show $\nabla^2 f(x) \succeq 0$, we must verify that $v^T \nabla^2 f(x) v \geq 0$ for all v:

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

since $(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$ (from Cauchy-Schwarz inequality)



6.18 Triangle Inequality

Suppose $u, v \in V$. Then

$$\|u + v\| \leq \|u\| + \|v\|.$$

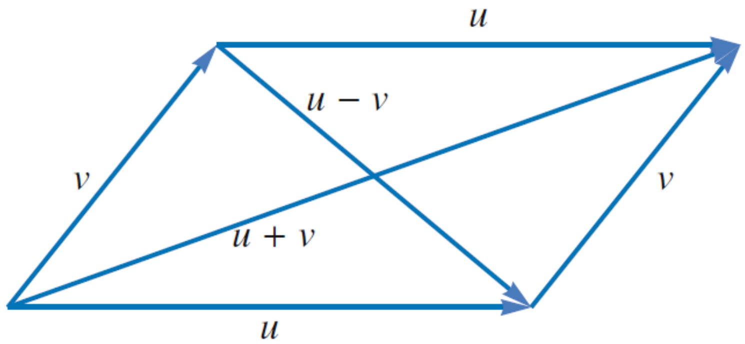
This inequality is an equality if and only if one of u, v is a nonnegative multiple of the other.

Proof We have

$$
\begin{aligned}
\|u+v\|^2 &= \langle u+v, u+v\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \langle v,u\rangle \\
&= \langle u,u\rangle + \langle v,v\rangle + \langle u,v\rangle + \overline{\langle u,v\rangle} \\
&= \|u\|^2 + \|v\|^2 + 2\operatorname{Re}\langle u,v\rangle \\
&\le \|u\|^2 + \|v\|^2 + 2|\langle u,v\rangle| && \textbf{6.19} \\
&\le \|u\|^2 + \|v\|^2 + 2\|u\|\,\|v\| && \textbf{6.20} \\
&= (\|u\| + \|v\|)^2,
\end{aligned}
$$

where 6.20 follows from the Cauchy–Schwarz Inequality (6.15). Taking square roots of both sides of the inequality above gives the desired inequality.

Parallelogram Equality



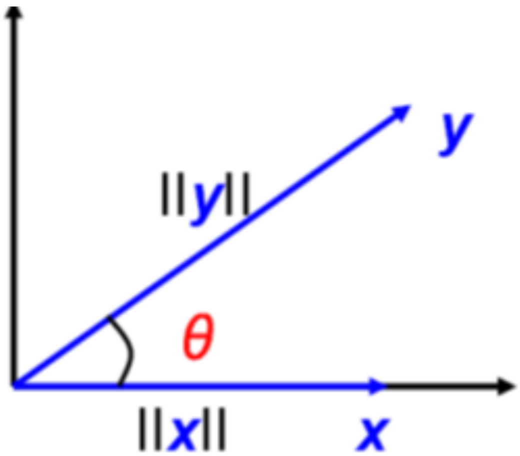
Suppose $u, v \in V$. Then

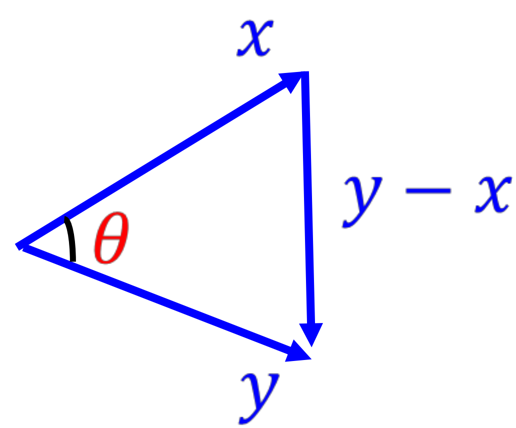
$$\|u+v\|^2 + \|u-v\|^2 = 2(\|u\|^2 + \|v\|^2).$$

Proof:

$$\begin{aligned}
\|u+v\|^2 + \|u-v\|^2 &= \langle u+v, u+v\rangle + \langle u-v, u-v\rangle \\
&= \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \langle v, u\rangle \\
&\qquad + \|u\|^2 + \|v\|^2 - \langle u, v\rangle - \langle v, u\rangle \\
&= 2(\|u\|^2 + \|v\|^2),
\end{aligned}$$

Geometric property (R^2)





$\boldsymbol{x}$ = (||$\boldsymbol{x}$||,$\boldsymbol{0}$)

$\boldsymbol{y}$ = (||$\boldsymbol{y}$||cos θ , ||$\boldsymbol{y}$||sin θ).

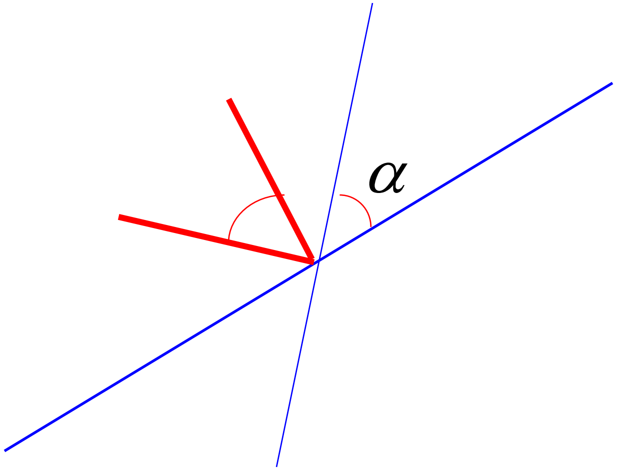
$\|y-x\|^2 = \|y\|^2 + \|x\|^2$-2 $\|y\|\ \|x\|\cos\theta$

<x,y>= ||$\boldsymbol{x}$|| ||$\boldsymbol{y}$||cos θ

where θ the angle between $\boldsymbol{x}$ and $\boldsymbol{y}$.

Definition: The angle between two vectors ***x*** and ***y*** in a linear space with Euclidean norm is defined as

$$\boldsymbol{\theta} = arccos\frac{(\boldsymbol{x}, \boldsymbol{y})}{\|\boldsymbol{x}\|\|\boldsymbol{y}\|}$$



$$l_1:\ a_1x + b_1y + c_1 = 0$$
$$l_2:\ a_2x + b_2y + c_2 = 0$$

$\longrightarrow$

$$\cos\alpha = \frac{|a_1a_2 + b_1b_2|}{\sqrt{a_1^2 + b_1^2}\sqrt{a_2^2 + b_2^2}}$$

Theorem 2.

$$\|x\| = \max\, (x,y), \|y\| = 1.$$

Proof. By Theorem 1, we have

$$|(x,y)| \le \|x\|\|y\| \le \|x\|$$

for any y with $\|y\| = 1$.

Let $y = x/\|x\|$, then $(x,y) = \|x\|$.

$$(x,y) = x^T y$$

$$(x,y) = x^T M y$$

$$M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\boxed{\begin{array}{ll} \max & a_1 y_1 + a_2 y_2 \\ s.t. & y_1^2 + y_2^2 = 1 \\ & \\ \max & 2a_1 y_1 + a_1 y_2 + a_2 y_1 + 2a_2 y_2 \\ s.t. & y_1^2 + y_1 y_2 + y_2^2 = 1/2 \end{array}}$$

Additional content

General norm

Definition

A vector space V is said to be a **normed linear space** if, to each vector $\mathbf{v} \in V$, there is associated a real number $\|\mathbf{v}\|$, called the **norm** of $\mathbf{v}$, satisfying

1. $\|\mathbf{v}\| \geq 0$ with equality if and only if $\mathbf{v} = 0$.
2. $\|\alpha\mathbf{v}\| = |\alpha|\|\mathbf{v}\|$ for any scalar α.
3. $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ for all $\mathbf{v}, \mathbf{w} \in V$.

Norms on R^n

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{1/p}$$

$$\|v\|_1 = |v_1| + |v_2| + \cdots + |v_n|$$

$$\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

$$\|v\|_3 = \sqrt[3]{|v_1|^3 + |v_2|^3 + \cdots + |v_n|^3}$$

$$\|v\|_4 = \sqrt[4]{v_1^4 + v_2^4 + \cdots + v_n^4}$$

$$\|v\|_\infty = \max_i |v_i|$$

The norm is not linear. To inject linearity into the discussion

Only norm defined by inner product have the Pythagorean theorem

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$$

are orthogonal; however,

$$\|\mathbf{x}_1\|_\infty^2 + \|\mathbf{x}_2\|_\infty^2 = 4 + 16 = 20$$

while

$$\|\mathbf{x}_1 + \mathbf{x}_2\|_\infty^2 = 16$$

If, however, $\|\cdot\|_2$ is used, then

$$\|\mathbf{x}_1\|_2^2 + \|\mathbf{x}_2\|_2^2 = 5 + 20 = 25 = \|\mathbf{x}_1 + \mathbf{x}_2\|_2^2$$

Cauchy-Schwarz Inequality for general norm

$$|<x,y>| \leq ||x||\,||y||_*$$

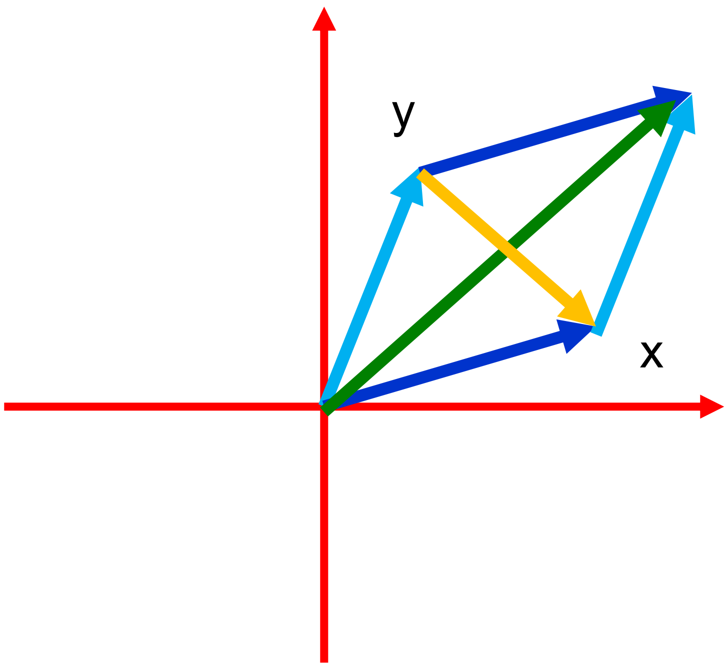
$$||y||_* = \sup_{||x|| \leq 1} <x,y>$$

设 $\|\cdot\|$ 是线性空间 V 上的一个范数. 如果它满足平行四边形公式, 即对任意两个向量 x 和 y,

$$\|x+y\|^2+\|x-y\|^2=2\left(\|x\|^2+\|y\|^2\right),$$

则存在唯一的 V 上的一个内积 $\langle\cdot\rangle$, 使得对任意向量 x

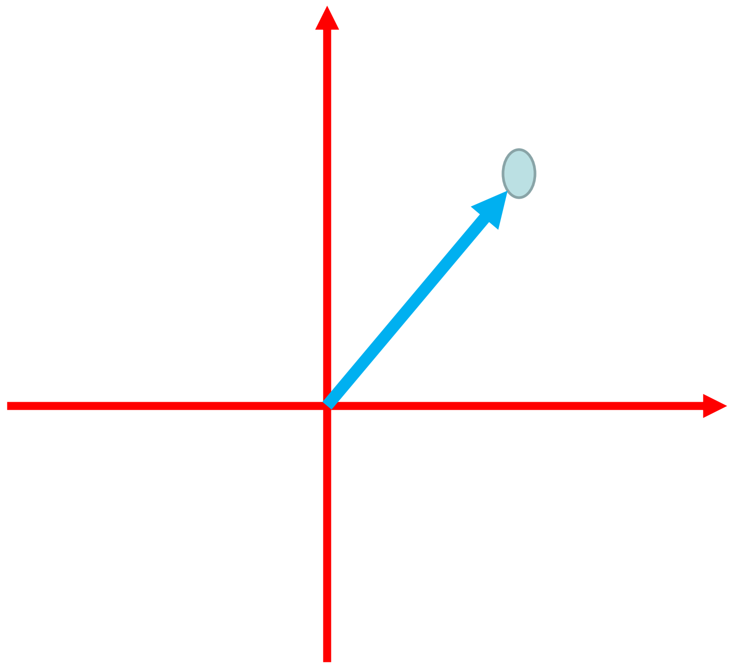
$$\|x\|=\sqrt{\langle x,x\rangle}.$$



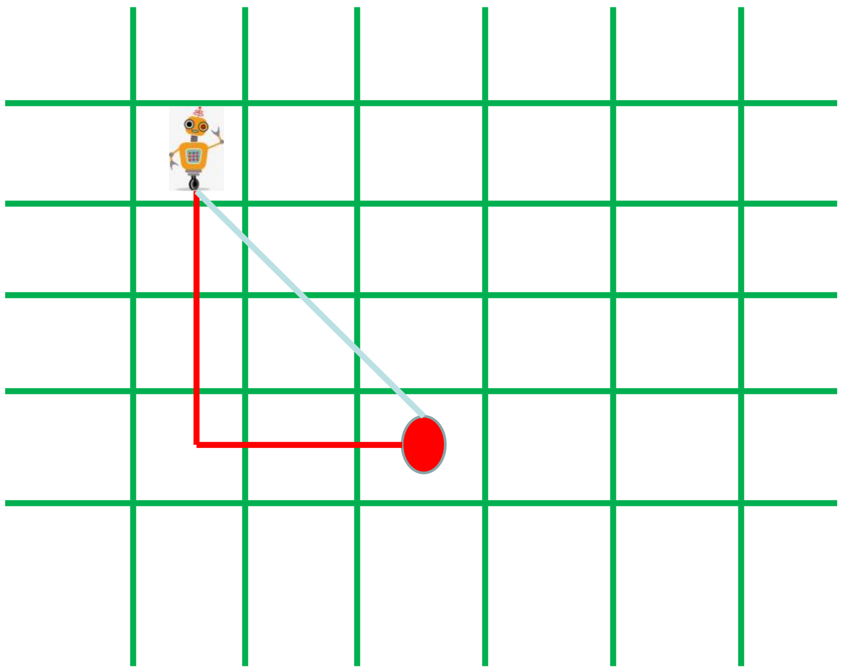
问题：既然已经有了长度，为什么还要考虑不是长度的范数?

引入范数使得我们可以考虑两个向量的距离．而距离好处有两个，分别是定性的和定量的．定性的方面是可以考虑极限，进而引入连续，导数与积分等数学分析的方法．定量的方面是可以为最优逼近问题提供逼近性能指标．定性的角度来说，各种范数的作用是等价的(范数的等价性)；但就提供符合实际需求的逼近性能指标来说，不同的范数起着不同的作用，给出的“最优解”也不同.

Different norms



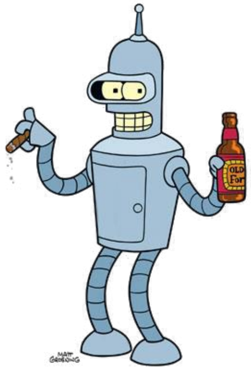
$\|x\|_2$



$\|x\|_1$

L1 Norm Regularization and Sparsity

over-fitting



把打扒捕拉 被劈盐派喔

把打扒捕拉 被劈盐派喔

揪

https://blog.mlreview.com/l1-norm-regularization-and-sparsity-explained-for-dummies-5b0e4be3938a

Why does a sparse solution avoid over-fitting

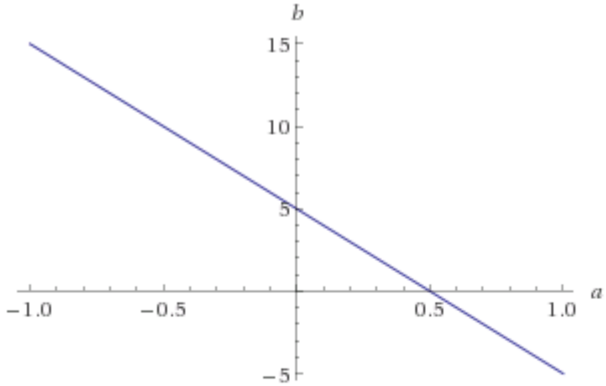
$$Ax = b$$

$$\begin{array}{ccccc} A & & x & & b \\ \begin{bmatrix} \cdots & \cdots & \cdots \end{bmatrix} & \times & \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix} & = & \begin{bmatrix} \vdots \end{bmatrix} \end{array}$$

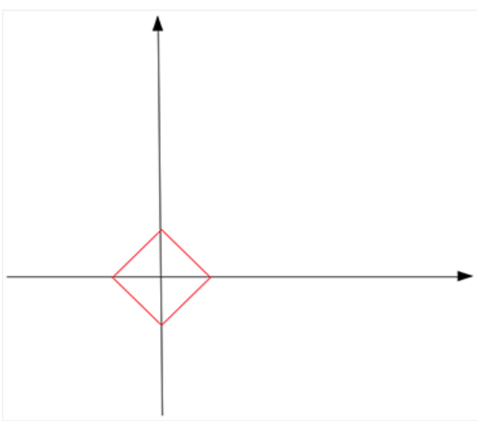
A sparse solution may make the vector x smaller (sparse)

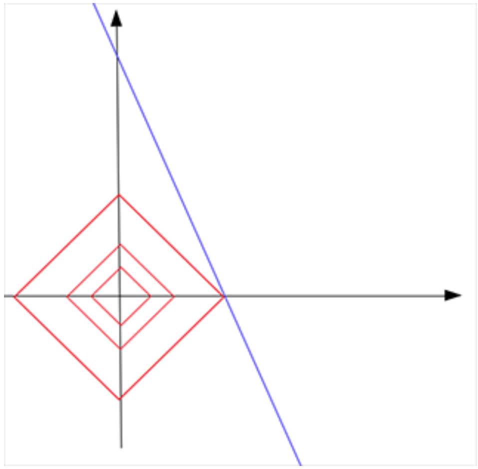
Why does a small L1 norm give a sparse solution

y = a * x + b $\qquad$ [10, 5] $\qquad$ $\begin{bmatrix}10 & 1\end{bmatrix} \times \begin{pmatrix}a \\ b\end{pmatrix} = \begin{pmatrix}5\end{pmatrix}$ $\qquad$ b = 5 – 10 * a



(a from −1 to 1)

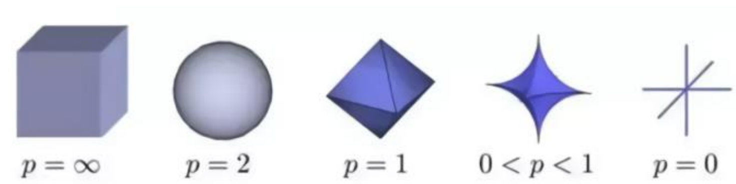


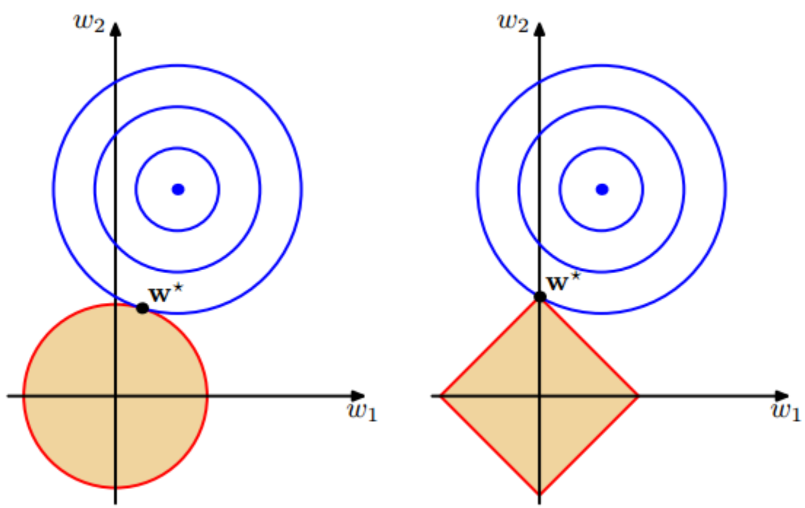


1-norm regularization and sparse

$$\min_{w} f(w) \quad s.t. \quad ||w||_2 \leq 1$$

$$\min_{w} f(w) \quad s.t. ||w||_1 \leq 1$$





When there is a zombie outbreak, which one should be the weapon of choice?

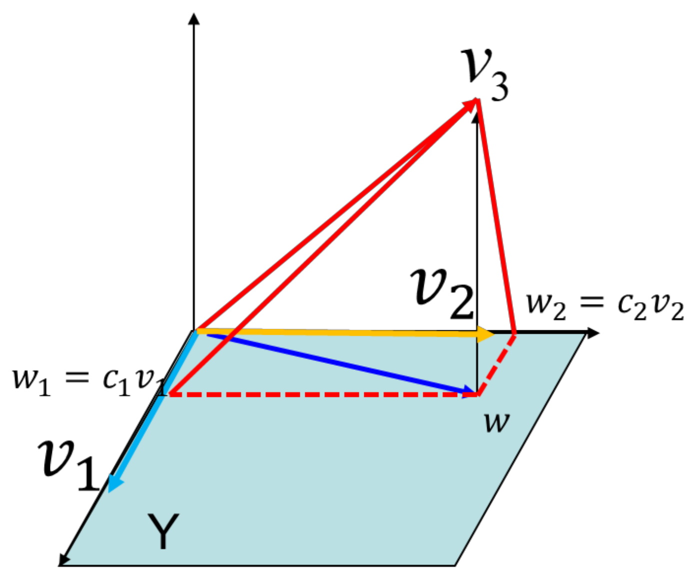
Orthogonal Projection

6.14 An orthogonal decomposition

Suppose $u, v \in V$, with $v \neq 0$. Set $c = \dfrac{\langle u, v\rangle}{\|v\|^2}$ and $w = u - \dfrac{\langle u, v\rangle}{\|v\|^2}v$. Then

$$\langle w, v\rangle = 0 \quad \text{and} \quad u = cv + w.$$

Orthogonal Projection



Orthogonal Set

- A set of vectors is called an orthogonal set if every pair of distinct vectors in the set is orthogonal. Standard inner product

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix} \right\}$$ An orthogonal set?

By definition, a set with only one vector is an orthogonal set.

Is orthogonal set independent?

Independent?

- Any orthogonal set of nonzero vectors is linearly independent.

Let $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be an orthogonal set $\mathbf{v}_i \neq \mathbf{0}$ for $i = 1, 2, \ldots, k$.

Assume $c_1, c_2, \ldots, c_k$ make $c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$

$$(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_i\mathbf{v}_i + \cdots + c_k\mathbf{v}_k) \cdot \mathbf{v}_i$$

$$= \quad c_1\mathbf{v}_1 \cdot \mathbf{v}_i + c_2\mathbf{v}_2 \cdot \mathbf{v}_i + \cdots + c_i\mathbf{v}_i \cdot \mathbf{v}_i + \cdots + c_k\mathbf{v}_k \cdot \mathbf{v}_i$$

$$= \quad c_i(\mathbf{v}_i \cdot \mathbf{v}_i) \quad = \quad c_i\underbrace{||\mathbf{v}_i||^2}_{\neq 0} \Rightarrow c_i = 0$$

$$\Rightarrow c_1 = c_2 = \cdots = c_k = 0$$

Problem

- Example: $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an basis for $\mathcal{R}^3$

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$$

$$\text{Let } \mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3.$$

$$c_1 = ?, c_2 = ?, c_3 = ?$$

Orthogonal basis?

Orthogonal Basis

- Let $S = \{v_1, v_2, \cdots, v_k\}$ be an orthogonal basis for a subspace W, and let u be a vector in W.

$$u = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

$$\frac{u \cdot v_1}{\|v_1\|^2} \quad \frac{u \cdot v_2}{\|v_2\|^2} \quad \frac{u \cdot v_k}{\|v_k\|^2}$$

Proof

To find c_i

$$u \cdot v_i = (c_1 v_1 + c_2 v_2 + \cdots + c_i v_i + \cdots + c_k v_k) \cdot v_i$$

$$= c_1 v_1 \cdot v_i + c_2 v_2 \cdot v_i + \cdots + c_i v_i \cdot v_i + \cdots + c_k v_k \cdot v_i$$

$$= c_i (v_i \cdot v_i) = c_i \|v_i\|^2 \Rightarrow c_i = \frac{u \cdot v_i}{\|v_i\|^2}$$

Example

- Example: $\mathcal{S} = \{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is an orthogonal basis for $\mathcal{R}^3$

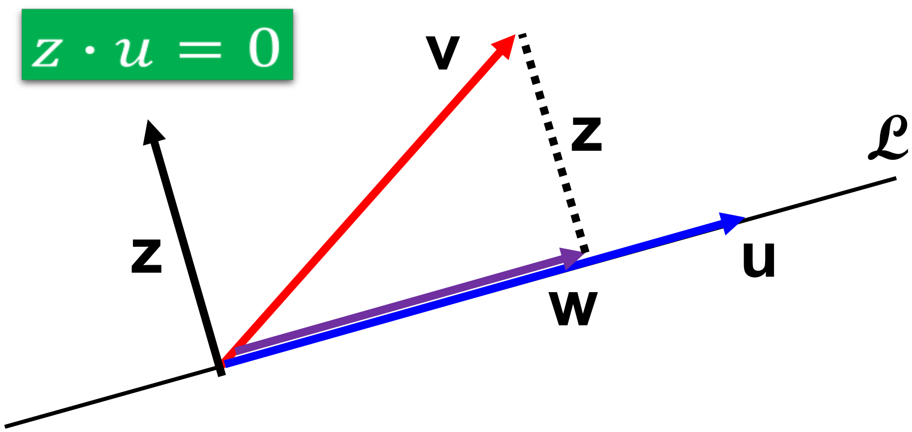
$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$$

$$\text{Let } \mathbf{u} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} \text{ and } \mathbf{u} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + c_3\mathbf{v}_3.$$

$$c_1 = \frac{\mathbf{u} \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2} \qquad c_2 = \frac{\mathbf{u} \cdot \mathbf{v}_2}{||\mathbf{v}_2||^2} \qquad c_3 = \frac{\mathbf{u} \cdot \mathbf{v}_3}{||\mathbf{v}_3||^2}$$

Orthogonal Projection on a line

- Orthogonal projection of a vector on a line



v: any vector

u: any nonzero vector on $\mathcal{L}$

w: orthogonal projection of **v** onto $\mathcal{L}$, **w** = c**u**

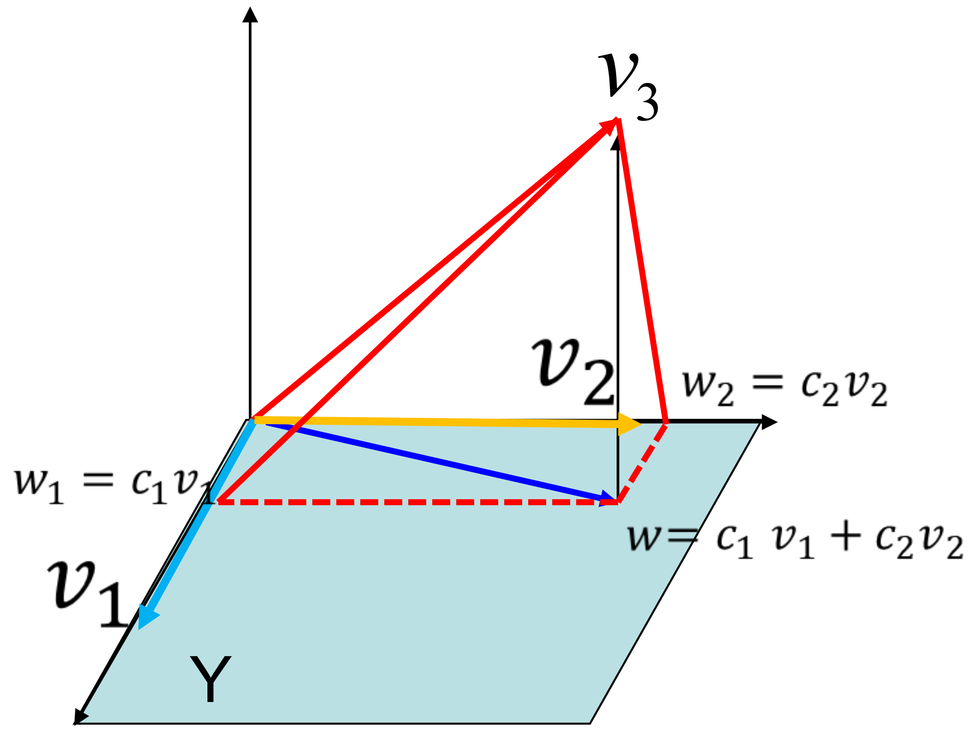
z: **v** – **w**

$$(v - w) \cdot u = (v - cu) \cdot u = v \cdot u - cu \cdot u = v \cdot u - c\|u\|^2$$

=0

$$c = \frac{v \cdot u}{\|u\|^2} \qquad w = cu = \frac{v \cdot u}{\|u\|^2} u$$

Distance from tip of **v** to $\mathcal{L}$: $\|z\| = \|v - w\| = \left\| v - \frac{v \cdot u}{\|u\|^2} u \right\|$



$$(v_3 - w_{1,}, v_1) = (v_3 - c_1 v_1, v_1) = (v_3, v_1) - c_1(v_1, v_1) = 0$$

$$w_1 = \frac{(v_3, v_1)}{\|v_1\|^2} v_1 \qquad w_2 = \frac{(v_3, v_2)}{\|v_2\|^2} v_2$$

$$w = w_1 + w_2 = \frac{(v_3, v_1)}{\|v_1\|^2} v_1 + \frac{(v_3, v_2)}{\|v_2\|^2} v_2$$

Orthogonal Projection

- Let $S = \{v_1, v_2, \cdots, v_k\}$ be an orthogonal basis for a subspace W, and let u be a vector in W.

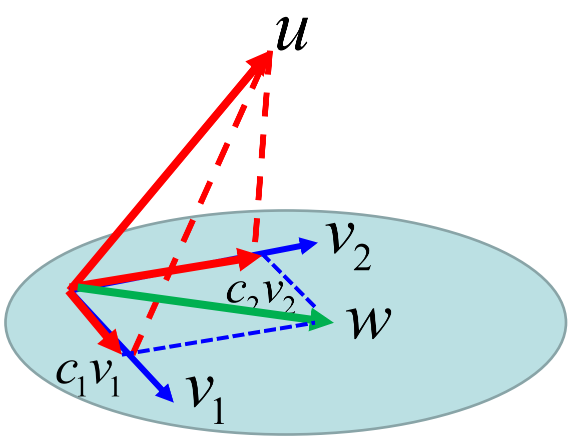
$$u = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

$$\frac{u \cdot v_1}{\|v_1\|^2} \quad \frac{u \cdot v_2}{\|v_2\|^2} \qquad \qquad \frac{u \cdot v_k}{\|v_k\|^2}$$

- Let u be **any vector** , and w is the orthogonal projection of u on W.

$$w = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

$$\frac{u \cdot v_1}{\|v_1\|^2} \quad \frac{u \cdot v_2}{\|v_2\|^2} \quad \frac{u \cdot v_k}{\|v_k\|^2}$$



“才能最好于孤独中培养，品格最好在世界的波涛汹涌中形成”

——歌德

Orthogonal Basis

Let $\{u_1, u_2, \cdots, u_k\}$ be a basis of a subspace W. How to transform $\{u_1, u_2, \cdots, u_k\}$ into an orthogonal basis

$$
\begin{aligned}
\mathbf{v}_1 &= \mathbf{u}_1, \\
\mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2}\mathbf{v}_1, \\
\mathbf{v}_3 &= \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2}\mathbf{v}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{v}_2}{||\mathbf{v}_2||^2}\mathbf{v}_2,
\end{aligned}
$$

Gram-Schmidt Process

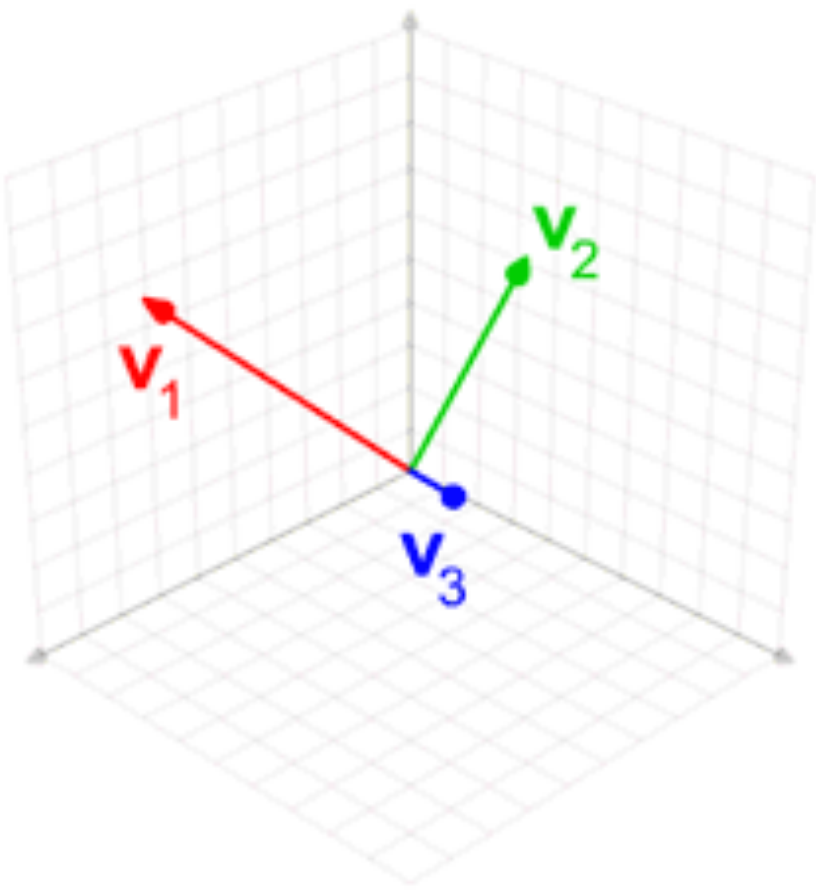
$$
\mathbf{v}_k = \mathbf{u}_k - \frac{\mathbf{u}_k \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2}\mathbf{v}_1 - \frac{\mathbf{u}_k \cdot \mathbf{v}_2}{||\mathbf{v}_2||^2}\mathbf{v}_2 - \cdots - \frac{\mathbf{u}_k \cdot \mathbf{v}_{k-1}}{||\mathbf{v}_{k-1}||^2}\mathbf{v}_{k-1}
$$

Then $\{v_1, v_2, \cdots, v_k\}$ is an orthogonal basis for W

$$
\text{Span}\ \{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_i\} = \text{Span}\ \{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_i\}
$$

orthogonal

Visualization



https://www.youtube.com/watch?v=Ys28-Yq21B8

Example

$\mathcal{S}$ = {$\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$} is a basis for subspace W

$$u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} \qquad u_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

(L.I. vectors)

Then $\mathcal{S}'$ = {$\mathbf{v}_1$, $\mathbf{v}_2$, $\mathbf{v}_3$} is an orthogonal basis for W .

$\mathcal{S}''$ = {$\mathbf{v}_1$, $\mathbf{v}_2$, 4$\mathbf{v}_3$}??? is also an orthogonal basis.

$$\mathbf{v}_1 = \mathbf{u}_1$$

$$\mathbf{v}_2 = \mathbf{u}_2 - \frac{\mathbf{u}_2 \cdot \mathbf{v}_1}{||\mathbf{v}_1||^2}\mathbf{v}_1 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 1 \end{bmatrix} - \frac{4}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix},$$

$$\mathbf{v}_3 = \mathbf{u}_3 - \frac{\mathbf{u}_3 \cdot \mathbf{u}_1}{||\mathbf{v}_1||^2}\mathbf{v}_1 - \frac{\mathbf{u}_3 \cdot \mathbf{u}_2}{||\mathbf{v}_2||^2}\mathbf{v}_2 = \begin{bmatrix} 1 \\ 1 \\ 2 \\ 1 \end{bmatrix} - \frac{5}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{(-1)}{2}\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix} = \frac{1}{4}\begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}.$$

Orthonormal Set

- A set of vectors is called an orthonormal set if it is an orthogonal set, and the norm of all the vectors is 1

$$\mathcal{S} = \left\{ \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix} \right\}$$

$$\frac{1}{\sqrt{14}}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \quad \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} \quad \frac{1}{\sqrt{42}}\begin{bmatrix} 5 \\ -4 \\ 1 \end{bmatrix}$$

Is orthonormal set independent?

Yes

A vector that has norm equal to 1 is called a unit vector.

6.23 **Definition** *orthonormal*

- A list of vectors is called ***orthonormal*** if each vector in the list has norm 1 and is orthogonal to all the other vectors in the list.
- In other words, a list $e_1, \ldots, e_m$ of vectors in V is orthonormal if

$$\langle e_j, e_k \rangle = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k. \end{cases}$$

6.24 **Example** *orthonormal lists*

(a) The standard basis in $\mathbf{F}^n$ is an orthonormal list.

(b) $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right)$ is an orthonormal list in $\mathbf{F}^3$.

(c) $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right), \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right), \left(\frac{1}{\sqrt{6}}, \frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}\right)$ is an orthonormal list in $\mathbf{F}^3$.

6.25 The norm of an orthonormal linear combination

If $e_1, \ldots, e_m$ is an orthonormal list of vectors in V, then

$$\|a_1 e_1 + \cdots + a_m e_m\|^2 = |a_1|^2 + \cdots + |a_m|^2$$

for all $a_1, \ldots, a_m \in \mathbf{F}$.

6.26 An orthonormal list is linearly independent

Every orthonormal list of vectors is linearly independent.

6.27 Definition ***orthonormal basis***

An ***orthonormal basis*** of V is an orthonormal list of vectors in V that is also a basis of V.

6.28 An orthonormal list of the right length is an orthonormal basis

Every orthonormal list of vectors in V with length $\dim V$ is an orthonormal basis of V.

6.29 Example Show that

$$\left(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}\right), \left(\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}\right), \left(\tfrac{1}{2}, -\tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right), \left(-\tfrac{1}{2}, \tfrac{1}{2}, -\tfrac{1}{2}, \tfrac{1}{2}\right)$$

is an orthonormal basis of $\mathbf{F}^4$.

Example

Suppose n is a positive integer. Prove that

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \dots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \dots, \frac{\sin nx}{\sqrt{\pi}}$$

is an orthonormal list of vectors in $C[-\pi, \pi]$, the vector space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x)\, dx.$$

Proof:

$$\int (\sin jt)^2\, dt = \frac{2jt - \sin 2jt}{4j}$$

$$\int (\cos jt)^2\, dt = \frac{2jt + \sin 2jt}{4j}.$$

$$\int (\sin jt)(\sin kt)\, dt = \frac{j\sin(j-k)t + k\sin(j-k)t - j\sin(j+k)t + k\sin(j+k)t}{2(j-k)(j+k)}$$

$$\int (\sin jt)(\cos kt)\, dt = \frac{j\cos(j-k)t + k\cos(j-k)t + j\cos(j+k)t - k\cos(j+k)t}{2(k-j)(j+k)}$$

$$\int (\cos jt)(\cos kt)\, dt = \frac{j\sin(j-k)t + k\sin(j-k)t + j\sin(j+k)t - k\sin(j+k)t}{2(j-k)(j+k)}$$

$$\int (\sin jt)(\cos jt)\, dt = -\frac{(\cos jt)^2}{2j}.$$

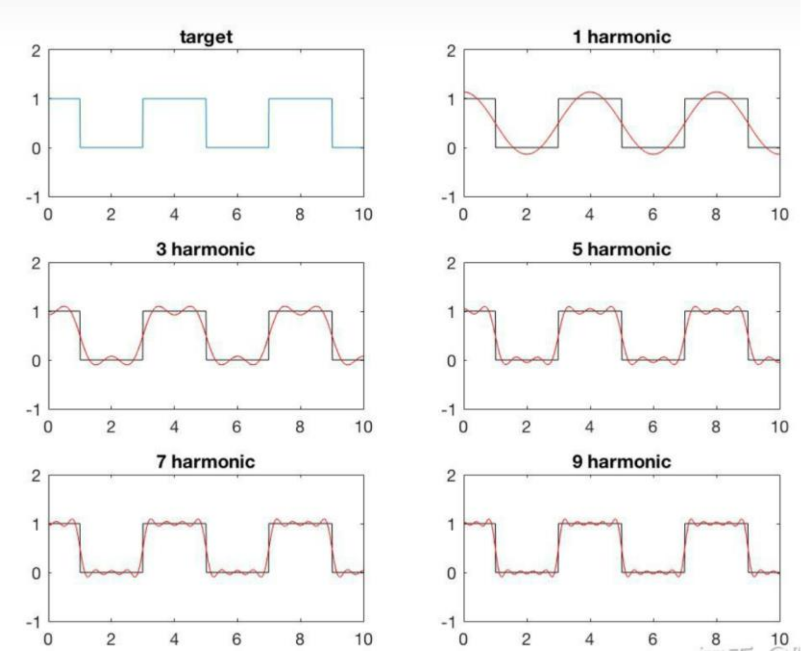
Example

Fourier Series

$$< f, g >= \frac{1}{l}\int_{-l}^{l} f(x)g(x)dx$$

$[-l, l]$ Periodic function

$$f(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty}\left[a_k \cos(k\omega t) + b_k \sin(k\omega t)\right]$$



$$\omega = \frac{2\pi}{2l}$$

$$\begin{cases} a_k = \frac{1}{l}\int_{-l}^{l} f(x)\cos\frac{n\pi x}{l}dx(n = 0, 1, 2, \dots) \\ b_k = \frac{1}{l}\int_{-l}^{l} f(x)\sin\frac{n\pi x}{l}dx(n = 1, 2, \dots) \end{cases}$$

In general, given a basis $e_1, \ldots, e_n$ of V and a vector $v \in V$, we know that there is some choice of scalars $a_1, \ldots, a_n \in \mathbf{F}$ such that

$$v = a_1 e_1 + \cdots + a_n e_n.$$

Coordinate $(a_1, a_2, \ldots \ldots . a_n)$

6.30 Writing a vector as linear combination of orthonormal basis

Suppose $e_1, \ldots, e_n$ is an orthonormal basis of V and $v \in V$. Then

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n$$

and

$$\|v\|^2 = |\langle v, e_1 \rangle|^2 + \cdots + |\langle v, e_n \rangle|^2.$$

6.31 Gram–Schmidt Procedure

Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors in V. Let $e_1 = v_1/\|v_1\|$. For $j = 2, \ldots, m$, define e_j inductively by

$$e_j = \frac{v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}}{\|v_j - \langle v_j, e_1\rangle e_1 - \cdots - \langle v_j, e_{j-1}\rangle e_{j-1}\|}.$$

Then $e_1, \ldots, e_m$ is an orthonormal list of vectors in V such that

$$\operatorname{span}(v_1, \ldots, v_j) = \operatorname{span}(e_1, \ldots, e_j)$$

for $j = 1, \ldots, m$.

6.33 Example Find an orthonormal basis of $\mathcal{P}_2(\mathbf{R})$, where the inner product is given by $\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)\,dx$.

Solution We will apply the Gram–Schmidt Procedure (6.31) to the basis $1, x, x^2$.

To get started, with this inner product we have

$$\|1\|^2 = \int_{-1}^{1} 1^2\,dx = 2.$$

Thus $\|1\| = \sqrt{2}$, and hence $e_1 = \sqrt{\frac{1}{2}}$.

Now the numerator in the expression for e_2 is

$$x - \langle x, e_1\rangle e_1 = x - \left(\int_{-1}^{1} x\sqrt{\tfrac{1}{2}}\,dx\right)\sqrt{\tfrac{1}{2}} = x.$$

We have

$$\|x\|^2 = \int_{-1}^{1} x^2\,dx = \tfrac{2}{3}.$$

Thus $\|x\| = \sqrt{\frac{2}{3}}$, and hence $e_2 = \sqrt{\frac{3}{2}}x$.

Now the numerator in the expression for e_3 is

$$\begin{aligned}
x^2 &- \langle x^2, e_1\rangle e_1 - \langle x^2, e_2\rangle e_2 \\
&= x^2 - \Big(\int_{-1}^{1} x^2\sqrt{\tfrac{1}{2}}\,dx\Big)\sqrt{\tfrac{1}{2}} - \Big(\int_{-1}^{1} x^2\sqrt{\tfrac{3}{2}}x\,dx\Big)\sqrt{\tfrac{3}{2}}x \\
&= x^2 - \tfrac{1}{3}.
\end{aligned}$$

We have

$$\|x^2 - \tfrac{1}{3}\|^2 = \int_{-1}^{1} \big(x^4 - \tfrac{2}{3}x^2 + \tfrac{1}{9}\big)\,dx = \tfrac{8}{45}.$$

Thus $\|x^2 - \frac{1}{3}\| = \sqrt{\frac{8}{45}}$, and hence $e_3 = \sqrt{\frac{45}{8}}\big(x^2 - \frac{1}{3}\big)$.

Thus

$$\sqrt{\tfrac{1}{2}},\ \sqrt{\tfrac{3}{2}}x,\ \sqrt{\tfrac{45}{8}}\big(x^2 - \tfrac{1}{3}\big)$$

Theorem 4 (Gram-Schmidt). Given an arbitrary basis $y^{(1)},\dots, y^{(n)}$ in a finite-dimensional linear space equipped with a Euclidean structure, there is a related basis $x^{(1)},\dots, x^{(n)}$ with the following properties:

(i) $x^{(1)},\dots, x^{(n)}$ is an orthonormal basis.

(ii) $x^{(k)}$ is a linear combination of $y^{(1)},\dots,y^{(k)}$, for all k.

Proof. We proceed recursively; suppose $x^{(1)},\dots,x^{(k-1)}$ have already been constructed. We set

$$x^{(k)} = c\left(y^{(k)} - \sum_{1}^{k-1} c_j x^{(j)} \right).$$

$$\left(x^{(k)}, x^{(l)}\right) = c\left(\left(y^{(k)}, x^{(l)}\right) - \sum_{1}^{k-1} c_j \left(x^{(j)}, x^{(l)}\right)\right)$$

$$= c\left(\left(y^{(k)}, x^{(l)}\right) - c_l\right) = 0, \quad \text{for } l = 1, 2, \ldots, k-1.$$

Since $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k-1)}$ are already orthonormal, it is easy to see that $\boldsymbol{x}^{(k)}$ defined above is orthogonal to them if we choose

$$c_l = \left(y^{(k)}, x^{(l)}\right), \quad l = 1, \ldots, k-1,$$

Finally we choose $\boldsymbol{c}$ so that $\|\boldsymbol{x}^k\| = \boldsymbol{1}$.

R^n *space*

$$x^{(k)} = c\left(y^{(k)} - \sum_{1}^{k-1} c_j x^{(j)} \right).$$

$y^{(1)} = a_{11} x^{(1)}$ $y^{(2)} = a_{21} x^{(1)} + a_{22} x^{(2)}$

$y^{(k)} = a_{k1} x^{(1)} + \cdots + a_{kk} x^{(k)}$

$$Y = [y^{(1)} \quad \cdots \quad y^{(k)}]$$

$$= [x^{(1)} \quad \cdots \quad x^{(k)}] \begin{bmatrix} a_{11} & a_{12} \ldots. & a_{1k} \\ 0 & & \\ \vdots & a_{22} \ldots. & a_{2k} \\ & \vdots & \vdots \\ 0 & 0 \ldots.. & a_{kk} \end{bmatrix}$$

$Y = QR$, Q is an orthogonal matrix

Let ***y*** be any other vector in ***X***; it can be expressed as

$$y = \sum b_k x^{(k)}.$$

$$x = \sum a_j x^{(j)} \qquad \textbf{(16)}$$

get

$$(x, y) = \sum \sum a_j b_k \left(x^{(j)}, y^{(k)}\right)$$

$$= \sum a_j b_j. \qquad \textbf{(17)}$$

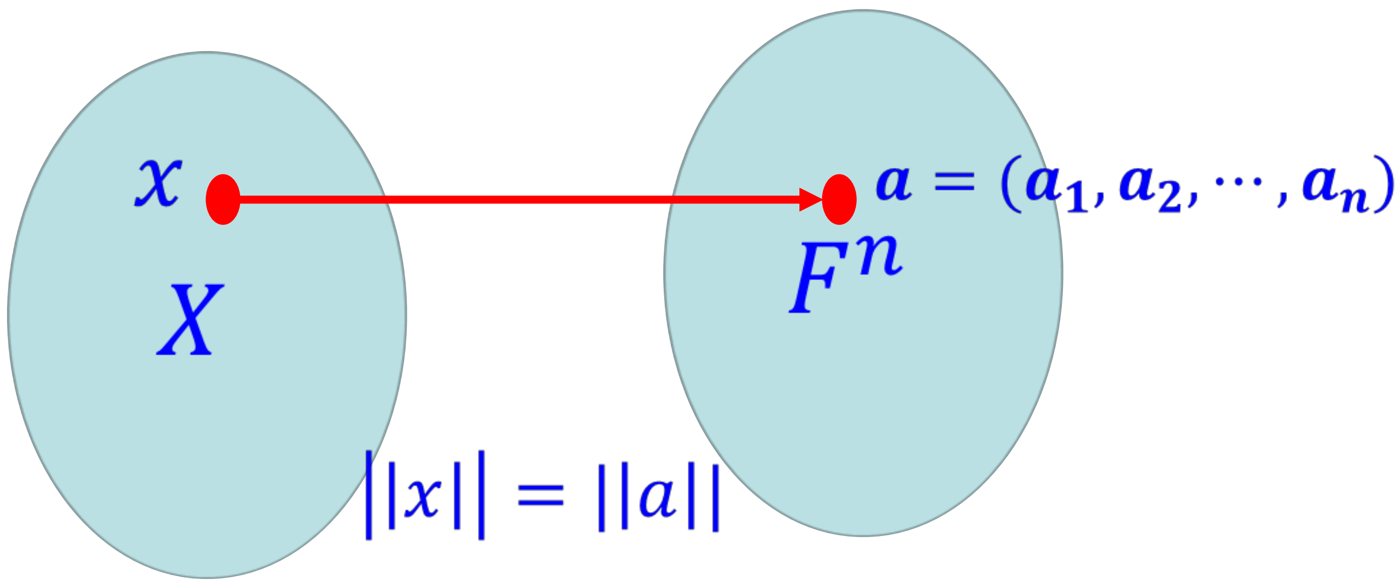
In particular, for ***y*** = ***x*** we get

$$\|x\|^2 = \sum a_j^2. \qquad \textbf{(17)'}$$

Equation **(17)** shows that the mapping defined by **(16)**,

$$\boldsymbol{x} \rightarrow (\boldsymbol{a_1},\ldots,\boldsymbol{a_n}),$$

carries the space $\boldsymbol{X}$ with a Euclidean structure into $\boldsymbol{F^n}$, and carries the scalar product of $\boldsymbol{X}$ into the standard scalar product **(2)** of $\boldsymbol{F^n}$.



$$||x|| = ||a||$$

$$< x, y > = < a, b >$$

6.39 Definition ***linear functional***

A ***linear functional*** on V is a linear map from V to $\mathbf{F}$. In other words, a linear functional is an element of $\mathcal{L}(V, \mathbf{F})$.

6.40 Example The function $\varphi : \mathbf{F}^3 \to \mathbf{F}$ defined by

$$\varphi(z_1, z_2, z_3) = 2z_1 - 5z_2 + z_3$$

is a linear functional on $\mathbf{F}^3$. We could write this linear functional in the form

$$\varphi(z) = \langle z, u \rangle$$

for every $z \in \mathbf{F}^3$, where $u = (2, -5, 1)$.

6.42 Riesz Representation Theorem

Suppose V is finite-dimensional and φ is a linear functional on V. Then there is a unique vector $u \in V$ such that

$$\varphi(v) = \langle v, u \rangle$$

for every $v \in V$.

Proof First we show there exists a vector $u \in V$ such that $\varphi(v) = \langle v, u \rangle$ for every $v \in V$. Let $e_1, \ldots, e_n$ be an orthonormal basis of V. Then

$$\begin{aligned} \varphi(v) &= \varphi(\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n) \\ &= \langle v, e_1 \rangle \varphi(e_1) + \cdots + \langle v, e_n \rangle \varphi(e_n) \\ &= \langle v, \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n \rangle \end{aligned}$$

for every $v \in V$, where the first equality comes from 6.30. Thus setting

6.43 $$u = \overline{\varphi(e_1)} e_1 + \cdots + \overline{\varphi(e_n)} e_n,$$

we have $\varphi(v) = \langle v, u \rangle$ for every $v \in V$, as desired.

Now we prove that only one vector $u \in V$ has the desired behavior. Suppose $u_1, u_2 \in V$ are such that

$$\varphi(v) = \langle v, u_1 \rangle = \langle v, u_2 \rangle$$

for every $v \in V$. Then

$$0 = \langle v, u_1 \rangle - \langle v, u_2 \rangle = \langle v, u_1 - u_2 \rangle$$

for every $v \in V$. Taking $v = u_1 - u_2$ shows that $u_1 - u_2 = 0$. In other words, $u_1 = u_2$, completing the proof of the uniqueness part of the result. ■

6.41 **Example** The function $\varphi\colon \mathcal{P}_2(\mathbf{R}) \to \mathbf{R}$ defined by

$$\varphi(p) = \int_{-1}^{1} p(t)\bigl(\cos(\pi t)\bigr)\,dt$$

is a linear functional on $\mathcal{P}_2(\mathbf{R})$ (here the inner product on $\mathcal{P}_2(\mathbf{R})$ is multiplication followed by integration on $[-1, 1]$; see 6.33). It is not obvious that there exists $u \in \mathcal{P}_2(\mathbf{R})$ such that

$$\varphi(p) = \langle p, u\rangle$$

for every $p \in \mathcal{P}_2(\mathbf{R})$ [we cannot take $u(t) = \cos(\pi t)$ because that function is not an element of $\mathcal{P}_2(\mathbf{R})$].

6.44 Example Find $u \in \mathcal{P}_2(\mathbf{R})$ such that

$$\int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\,dt = \int_{-1}^{1} p(t)u(t)\,dt$$

for every $p \in \mathcal{P}_2(\mathbf{R})$.

Solution Let $\varphi(p) = \int_{-1}^{1} p(t)\big(\cos(\pi t)\big)\,dt$. Applying formula 6.43 from the proof above, and using the orthonormal basis from Example 6.33, we have

$$u(x) = \Big(\int_{-1}^{1} \sqrt{\tfrac{1}{2}}\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{1}{2}} + \Big(\int_{-1}^{1} \sqrt{\tfrac{3}{2}}t\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{3}{2}}x$$
$$+ \Big(\int_{-1}^{1} \sqrt{\tfrac{45}{8}}\big(t^2 - \tfrac{1}{3}\big)\big(\cos(\pi t)\big)\,dt\Big)\sqrt{\tfrac{45}{8}}\big(x^2 - \tfrac{1}{3}\big).$$

A bit of calculus shows that

$$u(x) = -\frac{45}{2\pi^2}\big(x^2 - \tfrac{1}{3}\big).$$

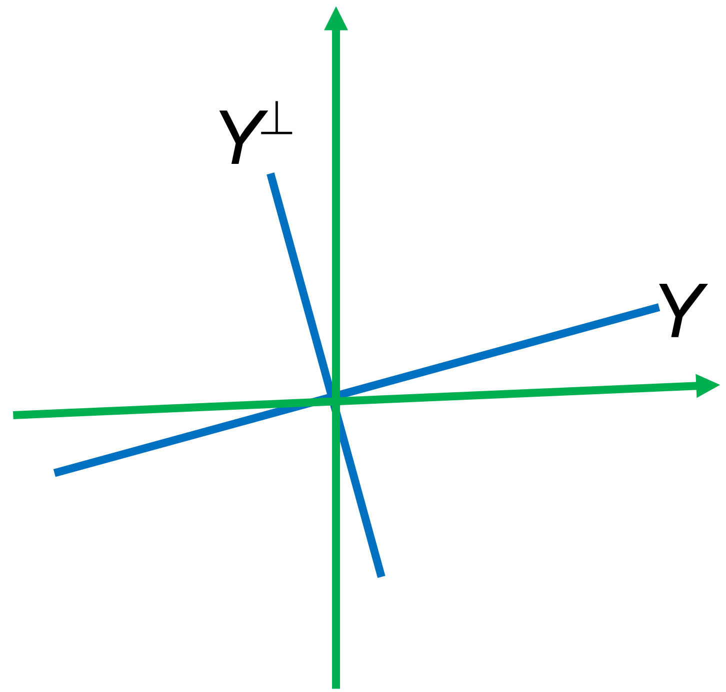
Orthogonal complement

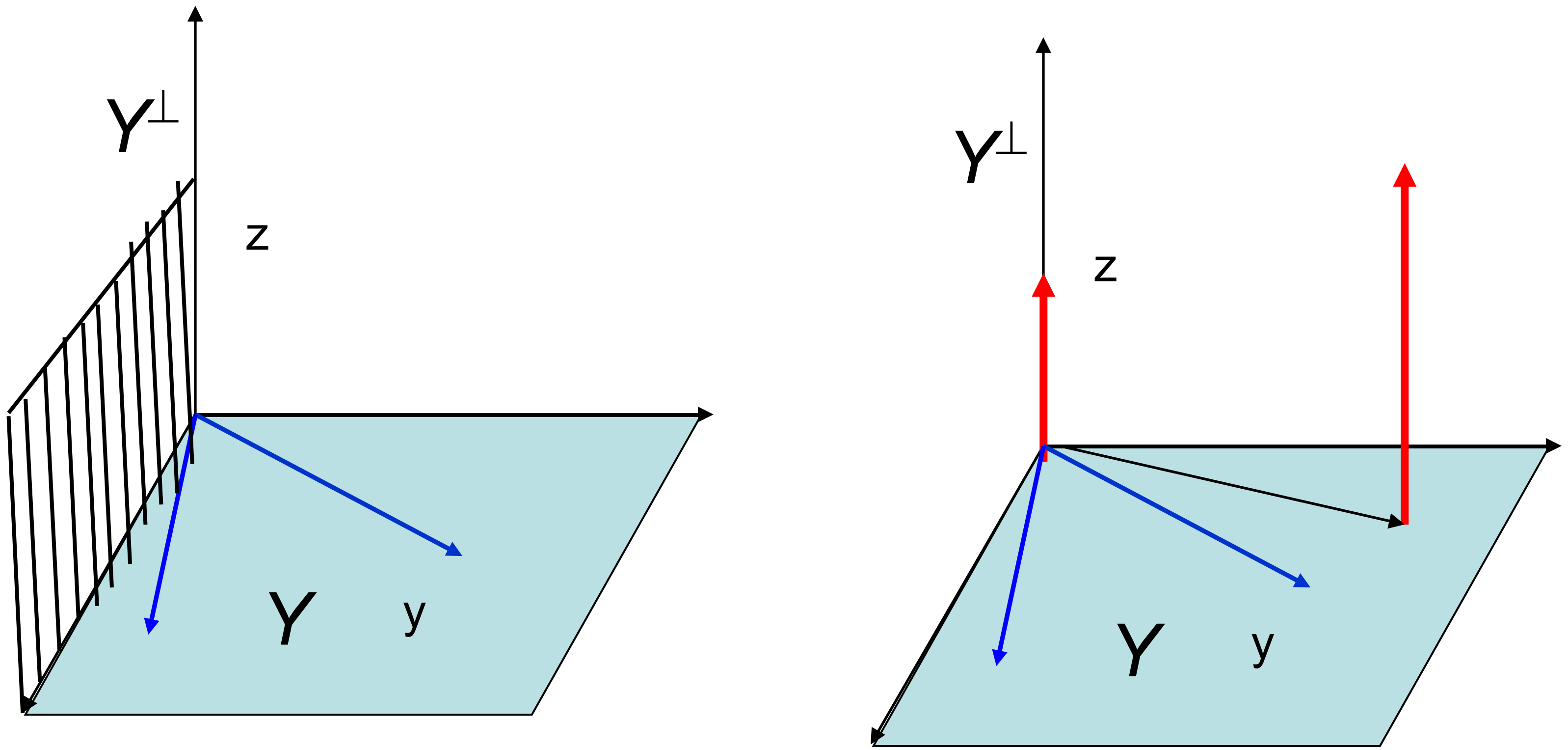
6.45 Definition ***orthogonal complement,*** $U^\perp$

If U is a subset of V, then the ***orthogonal complement*** of U, denoted $U^\perp$, is the set of all vectors in V that are orthogonal to every vector in U:

$$U^\perp = \{v \in V : \langle v, u \rangle = 0 \text{ for every } u \in U\}.$$

$$Y^{\perp} = \{z: (y, z) = 0, \forall y \in Y\}$$




$Y^{\perp}$
z
Y
y
$Y^{\perp}$
z
Y
y

6.46 Basic properties of orthogonal complement

(a) If U is a subset of V, then $U^{\perp}$ is a subspace of V.

(b) $\{0\}^{\perp} = V$.

(c) $V^{\perp} = \{0\}$.

(d) If U is a subset of V, then $U \cap U^{\perp} \subset \{0\}$.

(e) If U and W are subsets of V and $U \subset W$, then $W^{\perp} \subset U^{\perp}$.

Proof

(a) Suppose U is a subset of V. Then $\langle 0, u\rangle = 0$ for every $u \in U$; thus $0 \in U^{\perp}$.

Suppose $v, w \in U^{\perp}$. If $u \in U$, then

$$\langle v + w, u\rangle = \langle v, u\rangle + \langle w, u\rangle = 0 + 0 = 0.$$

Thus $v + w \in U^{\perp}$. In other words, $U^{\perp}$ is closed under addition.

Similarly, suppose $\lambda \in \mathbf{F}$ and $v \in U^{\perp}$. If $u \in U$, then

$$\langle \lambda v, u\rangle = \lambda\langle v, u\rangle = \lambda \cdot 0 = 0.$$

Thus $\lambda v \in U^{\perp}$. In other words, $U^{\perp}$ is closed under scalar multiplication. Thus $U^{\perp}$ is a subspace of V.

(b) Suppose $v \in V$. Then $\langle v, 0 \rangle = 0$, which implies that $v \in \{0\}^{\perp}$. Thus $\{0\}^{\perp} = V$.

(c) Suppose $v \in V^{\perp}$. Then $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus $V^{\perp} = \{0\}$.

(d) Suppose U is a subset of V and $v \in U \cap U^{\perp}$. Then $\langle v, v \rangle = 0$, which implies that $v = 0$. Thus $U \cap U^{\perp} \subset \{0\}$.

(e) Suppose U and W are subsets of V and $U \subset W$. Suppose $v \in W^{\perp}$. Then $\langle v, u \rangle = 0$ for every $u \in W$, which implies that $\langle v, u \rangle = 0$ for every $u \in U$. Hence $v \in U^{\perp}$. Thus $W^{\perp} \subset U^{\perp}$. ■

Examples

For Y = Span{$\mathbf{u}_1$, $\mathbf{u}_2$}, where $\mathbf{u}_1 = [\,1\ \ 1\ \ -1\ \ 4\,]^T$ and $\mathbf{u}_2 = [\,1\ -1\ \ 1\ \ 2\,]^T$

Please find the $Y^{\perp}$.

i.e., $\mathbf{v} = [\,x_1\ \ x_2\ \ x_3\ \ x_4\,]^T$ satisfies

$$\begin{array}{l} x_1 + x_2 - x_3 + 4x_4 = 0 \\ x_1 - x_2 + x_3 + 2x_4 = 0. \end{array} \Leftrightarrow \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} -3x_4 \\ x_3 - x_4 \\ x_3 \\ x_4 \end{bmatrix} = x_3 \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} -3 \\ -1 \\ 0 \\ 1 \end{bmatrix}$$

$\Leftrightarrow \mathcal{B} = \left\{ \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -3 \\ -1 \\ 0 \\ 1 \end{bmatrix} \right\}$ is a basis for $Y^{\perp}$.

Examples

- For any matrix A

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

$\mathbf{v} \in (\text{Row } A)^{\perp} \Leftrightarrow$ For all $\mathbf{w} \in$ Span{rows of A}, $(\mathbf{w}, \mathbf{v})=0$

$\Leftrightarrow A\mathbf{v} = \mathbf{0}$.

$$\boxed{(Col\ A)^{\perp} = Null\ A^{T}}$$

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

6.47 Direct sum of a subspace and its orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$V = U \oplus U^{\perp}.$$

Proof First we will show that

6.48 $$V = U + U^{\perp}.$$

To do this, suppose $v \in V$. Let $e_1, \ldots, e_m$ be an orthonormal basis of U. Obviously

6.49 $$v = \underbrace{\langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m}_{u} + \underbrace{v - \langle v, e_1 \rangle e_1 - \cdots - \langle v, e_m \rangle e_m}_{w}.$$

Let u and w be defined as in the equation above. Clearly $u \in U$. Because $e_1, \ldots, e_m$ is an orthonormal list, for each $j = 1, \ldots, m$ we have

$$\begin{aligned} \langle w, e_j \rangle &= \langle v, e_j \rangle - \langle v, e_j \rangle \\ &= 0. \end{aligned}$$

Thus w is orthogonal to every vector in $\operatorname{span}(e_1, \ldots, e_m)$. In other words, $w \in U^{\perp}$. Thus we have written $v = u + w$, where $u \in U$ and $w \in U^{\perp}$, completing the proof of 6.48.

From 6.46(d), we know that $U \cap U^{\perp} = \{0\}$. Along with 6.48, this implies that $V = U \oplus U^{\perp}$ (see 1.45). ■

6.50 Dimension of the orthogonal complement

Suppose V is finite-dimensional and U is a subspace of V. Then

$$\dim U^{\perp} = \dim V - \dim U.$$

6.51 The orthogonal complement of the orthogonal complement

Suppose U is a finite-dimensional subspace of V. Then

$$U = (U^{\perp})^{\perp}.$$

Proof First we will show that

6.52 $$U \subset (U^\perp)^\perp.$$

To do this, suppose $u \in U$. Then $\langle u, v\rangle = 0$ for every $v \in U^\perp$ (by the definition of $U^\perp$). Because u is orthogonal to every vector in $U^\perp$, we have $u \in (U^\perp)^\perp$, completing the proof of 6.52.

To prove the inclusion in the other direction, suppose $v \in (U^\perp)^\perp$. By 6.47, we can write $v = u + w$, where $u \in U$ and $w \in U^\perp$. We have $v - u = w \in U^\perp$. Because $v \in (U^\perp)^\perp$ and $u \in (U^\perp)^\perp$ (from 6.52), we have $v - u \in (U^\perp)^\perp$. Thus $v - u \in U^\perp \cap (U^\perp)^\perp$, which implies that $v - u$ is orthogonal to itself, which implies that $v - u = 0$, which implies that $v = u$, which implies that $v \in U$. Thus $(U^\perp)^\perp \subset U$, which along with 6.52 completes the proof. ■

6.53 **Definition** ***orthogonal projection,*** P_U

Suppose U is a finite-dimensional subspace of V. The ***orthogonal projection*** of V onto U is the operator $P_U \in \mathcal{L}(V)$ defined as follows: For $v \in V$, write $v = u + w$, where $u \in U$ and $w \in U^{\perp}$. Then $P_U v = u$.

6.54 **Example** Suppose $x \in V$ with $x \neq 0$ and $U = \operatorname{span}(x)$. Show that

$$P_U v = \frac{\langle v, x \rangle}{\|x\|^2} x$$

for every $v \in V$.

Solution Suppose $v \in V$. Then

$$v = \frac{\langle v, x \rangle}{\|x\|^2} x + \left(v - \frac{\langle v, x \rangle}{\|x\|^2} x \right),$$

where the first term on the right is in $\operatorname{span}(x)$ (and thus in U) and the second term on the right is orthogonal to x (and thus is in $U^{\perp}$). Thus $P_U v$ equals the first term on the right, as desired.

6.55 Properties of the orthogonal projection P_U

Suppose U is a finite-dimensional subspace of V and $v \in V$. Then

(a) $P_U \in \mathcal{L}(V)$;

(b) $P_U u = u$ for every $u \in U$;

(c) $P_U w = 0$ for every $w \in U^\perp$;

(d) range $P_U = U$;

(e) null $P_U = U^\perp$;

(f) $v - P_U v \in U^\perp$;

(g) $P_U{}^2 = P_U$;

(h) $\|P_U v\| \le \|v\|$;

(i) for every orthonormal basis $e_1, \ldots, e_m$ of U,

$$P_U v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_m \rangle e_m.$$

Proof

(a) To show that P_U is a linear map on V, suppose $v_1, v_2 \in V$. Write

$$v_1 = u_1 + w_1 \quad \text{and} \quad v_2 = u_2 + w_2$$

with $u_1, u_2 \in U$ and $w_1, w_2 \in U^\perp$. Thus $P_U v_1 = u_1$ and $P_U v_2 = u_2$. Now

$$v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2),$$

where $u_1 + u_2 \in U$ and $w_1 + w_2 \in U^\perp$. Thus

$$P_U(v_1 + v_2) = u_1 + u_2 = P_U v_1 + P_U v_2.$$

Similarly, suppose $\lambda \in \mathbf{F}$. The equation $v = u + w$ with $u \in U$ and $w \in U^\perp$ implies that $\lambda v = \lambda u + \lambda w$ with $\lambda u \in U$ and $\lambda w \in U^\perp$. Thus $P_U(\lambda v) = \lambda u = \lambda P_U v$.

Hence P_U is a linear map from V to V.

(b) Suppose $u \in U$. We can write $u = u + 0$, where $u \in U$ and $0 \in U^{\perp}$. Thus $P_U u = u$.

(c) Suppose $w \in U^{\perp}$. We can write $w = 0 + w$, where $0 \in U$ and $w \in U^{\perp}$. Thus $P_U w = 0$.

(d) The definition of P_U implies that range $P_U \subset U$. Part (b) implies that $U \subset$ range P_U. Thus range $P_U = U$.

(e) Part (c) implies that $U^{\perp} \subset$ null P_U. To prove the inclusion in the other direction, note that if $v \in$ null P_U then the decomposition given by 6.47 must be $v = 0 + v$, where $0 \in U$ and $v \in U^{\perp}$. Thus null $P_U \subset U^{\perp}$.

(f) If $v = u + w$ with $u \in U$ and $w \in U^{\perp}$, then

$$v - P_U v = v - u = w \in U^{\perp}.$$

(g) If $v = u + w$ with $u \in U$ and $w \in U^{\perp}$, then

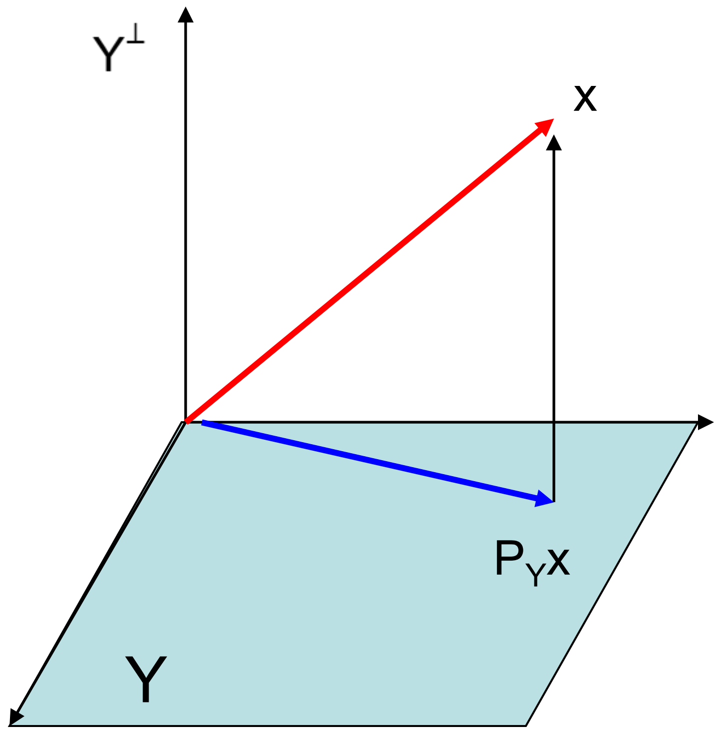
$$(P_U{}^2)v = P_U(P_U v) = P_U u = u = P_U v.$$

(h) If $v = u + w$ with $u \in U$ and $w \in U^{\perp}$, then

$$\|P_U v\|^2 = \|u\|^2 \leq \|u\|^2 + \|w\|^2 = \|v\|^2,$$

where the last equality comes from the Pythagorean Theorem.

(i) The formula for $P_U v$ follows from equation 6.49 in the proof of 6.47. ■



$Y^{\perp}$
x
$P_Y x$
Y

6.56 Minimizing the distance to a subspace

Suppose U is a finite-dimensional subspace of V, $v \in V$, and $u \in U$. Then

$$\|v - P_U v\| \leq \|v - u\|.$$

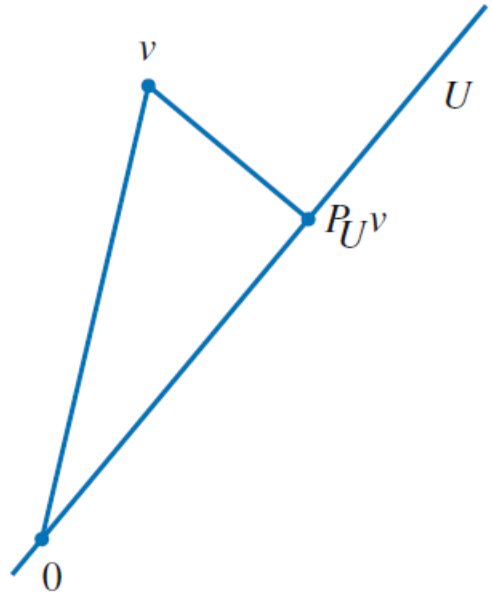
Furthermore, the inequality above is an equality if and only if $u = P_U v$.

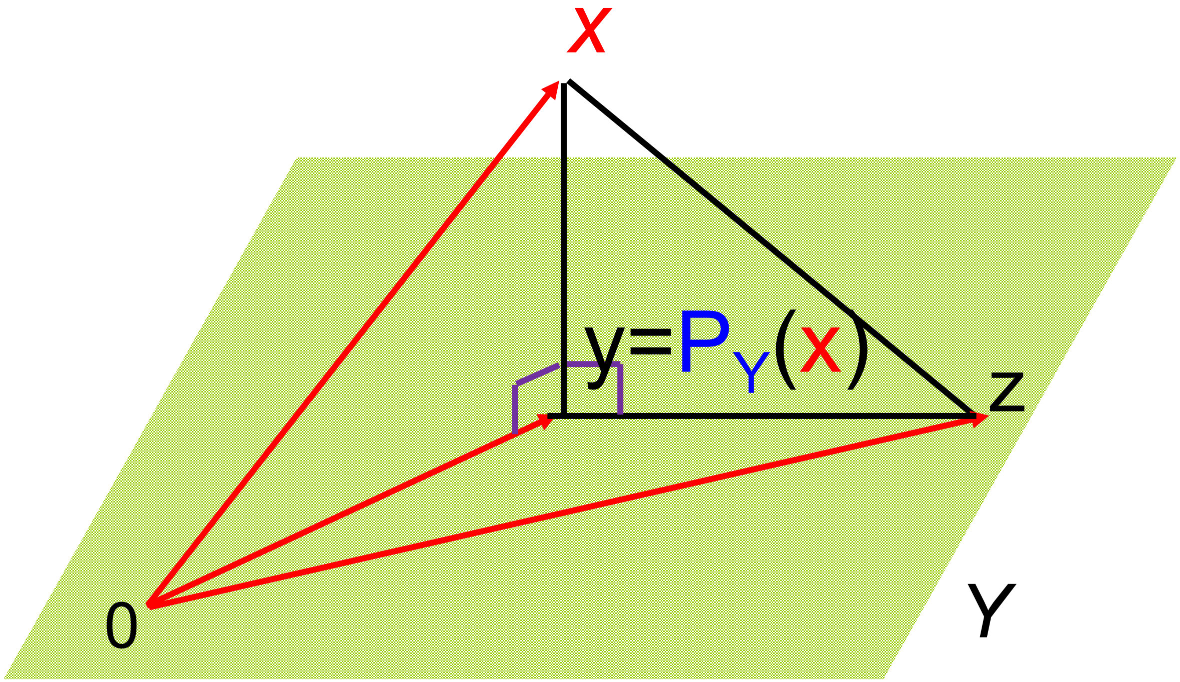
Proof We have

6.57
$$\begin{aligned}\|v - P_U v\|^2 &\leq \|v - P_U v\|^2 + \|P_U v - u\|^2 \\ &= \|(v - P_U v) + (P_U v - u)\|^2 \\ &= \|v - u\|^2,\end{aligned}$$

where the first line above holds because $0 \leq \|P_U v - u\|^2$, the second line above comes from the Pythagorean Theorem [which applies because $v - P_U v \in U^{\perp}$ by 6.55(f), and $P_U v - u \in U$], and the third line above holds by simple algebra. Taking square roots gives the desired inequality.

Our inequality above is an equality if and only if 6.57 is an equality, which happens if and only if $\|P_U v - u\| = 0$, which happens if and only if $u = P_U v$. ■


v
U
$P_U v$
0


x
y=PY(x)
z
0
Y

Example

Example Find a polynomial u with real coefficients and degree at most 5 that approximates $\sin x$ as well as possible on the interval $[-\pi, \pi]$, in the sense that

$$\int_{-\pi}^{\pi} |\sin x - u(x)|^2 \, dx$$

is as small as possible. Compare this result to the Taylor series approximation.

Solution Let $C_{\mathbf{R}}[-\pi, \pi]$ denote the real inner product space of continuous real-valued functions on $[-\pi, \pi]$ with inner product

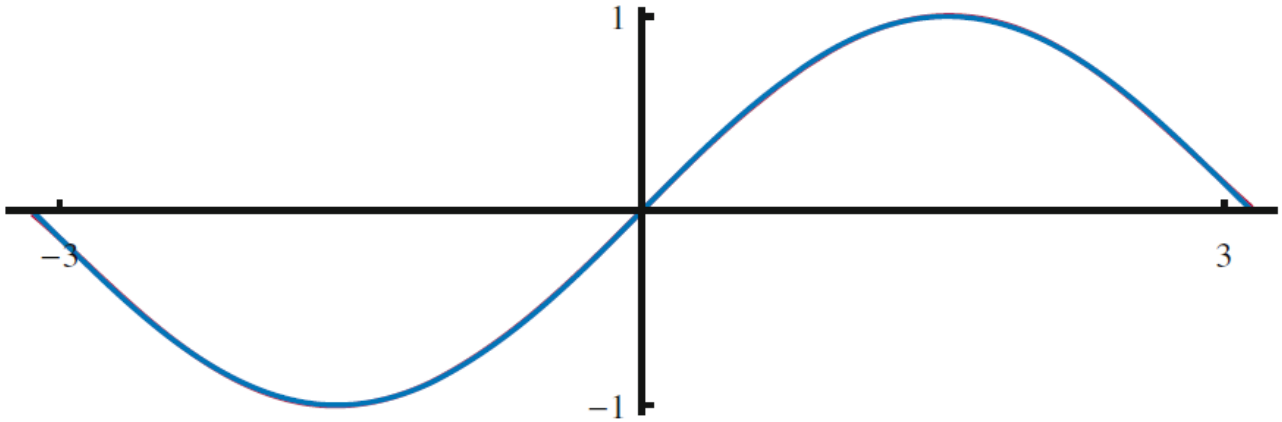
$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x)g(x)\,dx.$$

Let $v \in C_{\mathbf{R}}[-\pi, \pi]$ be the function defined by $v(x) = \sin x$. Let U denote the subspace of $C_{\mathbf{R}}[-\pi, \pi]$ consisting of the polynomials with real coefficients and degree at most 5. Our problem can now be reformulated as follows:

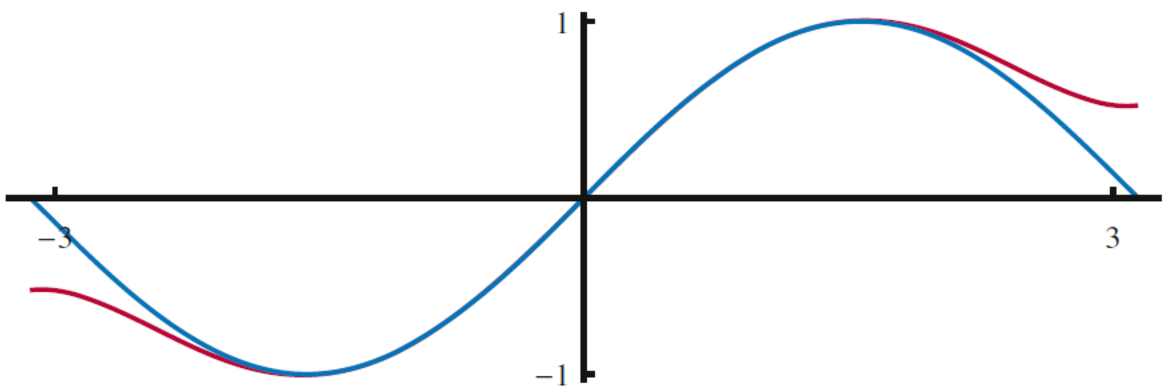
Find $u \in U$ such that $\|v - u\|$ is as small as possible.

$$1, x, x^2, x^3, x^4, x^5 \longrightarrow e_1, e_2, e_3, e_4, e_5, e_6$$

$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5$$



$$u(x) = 0.987862x - 0.155271x^3 + 0.00564312x^5$$



$$x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

Orthogonal Projection

- Let $S = \{v_1, v_2, \cdots, v_k\}$ be an orthogonal basis for a subspace W. Let u be any vector, and w is the orthogonal projection of u on W.

$$w = c_1 v_1 + c_2 v_2 + \cdots + c_k v_k$$

$$\frac{u \cdot v_1}{\|v_1\|^2} \quad \frac{u \cdot v_2}{\|v_2\|^2} \quad \frac{u \cdot v_k}{\|v_k\|^2}$$

$$C = [v_1 \quad \cdots \quad v_k] \qquad P_W = C(C^T C)^{-1} C^T \qquad w = P_W u$$

Orthogonal Projection Matrix

- Let C be an n x k matrix whose columns form a basis for a subspace W

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

$$P_W = C(C^TC)^{-1}C^T$$

n x n

Proof: Let $\mathbf{u} \in \mathcal{R}^n$ and $\mathbf{w} = P_W(\mathbf{u})$.

Since $W = \text{Col } C$, $\mathbf{w} = C\mathbf{b}$ for some $\mathbf{b} \in \mathcal{R}^k$

and $\mathbf{u} - \mathbf{w} \in W^{\perp}$

$\Rightarrow \mathbf{0} = C^T(\mathbf{u} - \mathbf{w}) = C^T\mathbf{u} - C^T\mathbf{w} = C^T\mathbf{u} - C^TC\mathbf{b}$.

$\Rightarrow C^T\mathbf{u} = C^TC\mathbf{b}$.

$\Rightarrow \mathbf{b} = (C^TC)^{-1}C^T\mathbf{u}$ and $\mathbf{w} = C(C^TC)^{-1}C^T\mathbf{u}$ as C^TC is invertible.

Least square problem $\min_{\{b\}} \|Cb - u\|_2$

Let C be a matrix with linearly independent columns. Then $C^T C$ is invertible.

Proof: We want to prove that C^TC has independent columns.

Suppose $C^TC\mathbf{b} = \mathbf{0}$ for some $\mathbf{b}$.

$\Rightarrow \mathbf{b}^TC^TC\mathbf{b} = (C\mathbf{b})^TC\mathbf{b} = (C\mathbf{b}) \bullet (C\mathbf{b}) = \|C\mathbf{b}\|^2 = 0.$

$\Rightarrow C\mathbf{b} = \mathbf{0} \Rightarrow \mathbf{b} = \mathbf{0}$ since C has L.I. columns.

Thus C^TC is invertible.

Orthogonal Projection Matrix

- Example: Let *W* be the 2-dimensional subspace of $\mathcal{R}^3$ with equation $x_1 - x_2 + 2x_3 = 0$.

$$P_W = C(C^T C)^{-1} C^T$$

W has a basis $\left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \right\}$ $\quad C = \begin{bmatrix} 1 & -2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$

$$P_W = \frac{1}{6} \begin{bmatrix} 5 & 1 & -2 \\ 1 & 5 & 2 \\ -2 & 2 & 2 \end{bmatrix} \qquad P_W \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 2 \end{bmatrix}$$

Least square problems

A least squares problem can generally be formulated as an overdetermined linear system of equations. Recall that an overdetermined system is one involving more equations than unknowns. Such systems are usually inconsistent. Thus, given an $m \times n$ system $A\mathbf{x} = \mathbf{b}$ with $m > n$, we cannot expect in general to find a vector $\mathbf{x} \in \mathbb{R}^n$ for which $A\mathbf{x}$ equals $\mathbf{b}$. Instead, we can look for a vector $\mathbf{x}$ for which $A\mathbf{x}$ is "closest" to $\mathbf{b}$. As you might expect, orthogonality plays an important role in finding such an $\mathbf{x}$.

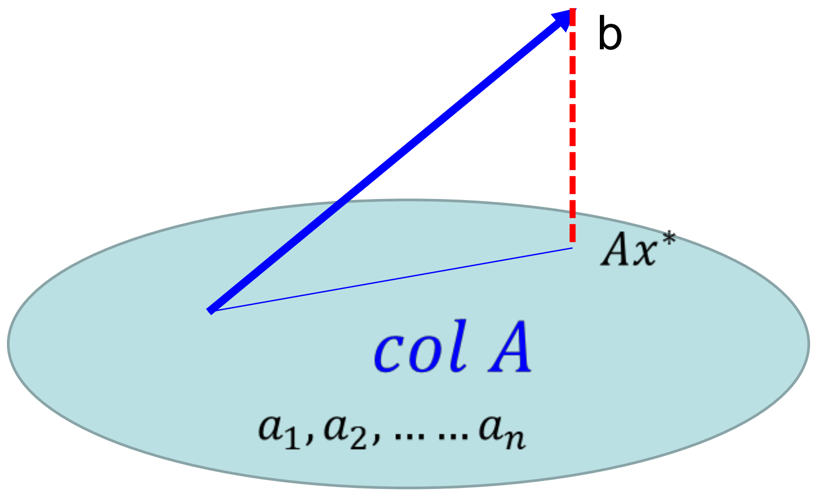
Least square problems

If we are given a system of equations $A\mathbf{x} = \mathbf{b}$, where A is an $m \times n$ matrix with $m > n$ and $\mathbf{b} \in \mathbb{R}^m$, then, for each $\mathbf{x} \in \mathbb{R}^n$, we can form a *residual*

$$r(\mathbf{x}) = \mathbf{b} - A\mathbf{x}$$

The distance between $\mathbf{b}$ and $A\mathbf{x}$ is given by

$$\|\mathbf{b} - A\mathbf{x}\| = \|r(\mathbf{x})\|$$



$A = [a_1, a_2, \dots \dots a_n]$ $\quad P_{col(A)} b = A(A^T A)^{-1} A^T b$

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

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

$\{y | A^T y = 0, y \in R^m\}$ $\quad A^T(b - Ax^*) = 0$

Thank you