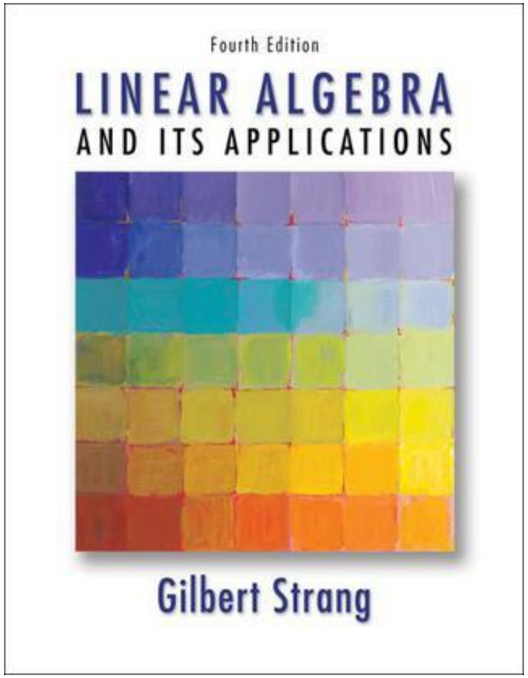
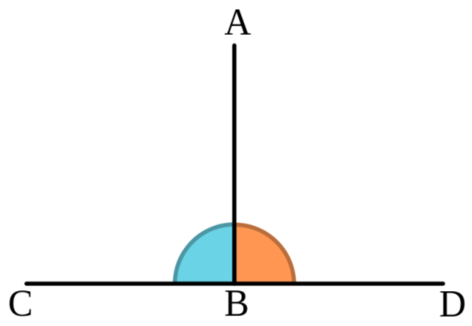
Linear Algebra



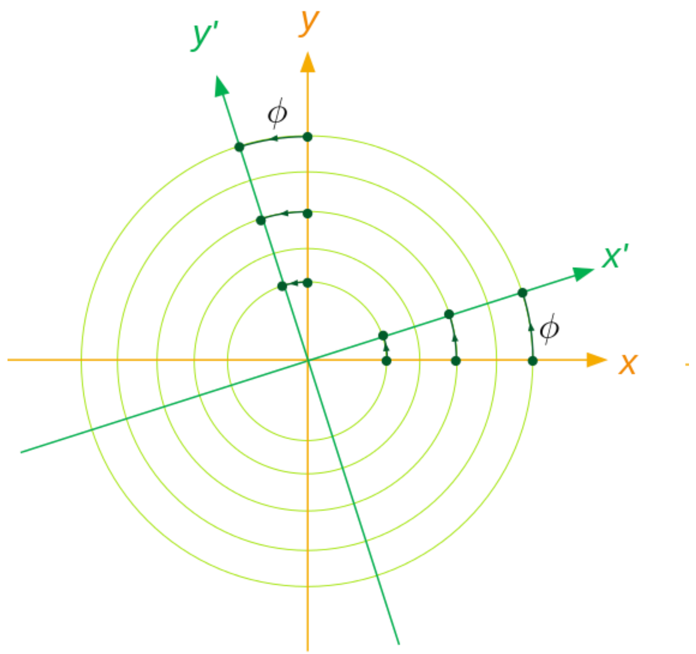
Instructor: Jing YAO

3 Orthogonality
(正交性)



3.3

PROJECTIONS AND LEAST SQUARES（投影与最小二乘）

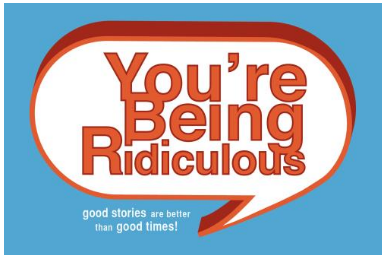


- Least squares
- Projection Matrix
- The matrix $A^{T}A$
- Fitting of Data (数据拟合)

Introduction

Solve *Ax* = *b*

when there is no solution?



IT HAPPENS ALL THE TIME!

(e.g., it often happens when m>n)

In spite of their unsolvability, inconsistent equations arise all the time in practice.

They have to be solved!

Elimination is going to fail.

Throw away some equations? This is hard to justify if all *m* equations come from the same source.

Better way:

choose the **x** *that minimizes an average error E in the m equations.*

We start with an example.

Example 1. The system of linear equations with *only one* unknown x

$$\begin{cases} 2x = b_1, \\ 3x = b_2, \\ 4x = b_3. \end{cases}$$

has solutions if and only if (b_1, b_2, b_3) is proportional to (2, 3, 4).

If it has no solution, we wish to find a value $\hat{x}$ *which minimizes the difference*

$$E^2 = (2x - b_1)^2 + (3x - b_2)^2 + (4x - b_3)^2.$$

Differentiating it with respect to x, we have

$$\frac{\mathrm{d}E^2}{\mathrm{d}x} = 4(2x - b_1) + 6(3x - b_2) + 8(4x - b_3)$$

equating 0 leads to

$$\hat{x} = \frac{2b_1 + 3b_2 + 4b_3}{2^2 + 3^2 + 4^2} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}},$$

where $\boldsymbol{a} = (2,3,4)^{\mathrm{T}}$ and $\boldsymbol{b} = (b_1, b_2, b_3)^{\mathrm{T}}$.

The general case is the same. We "solve" $\boldsymbol{a}x = \boldsymbol{b}$ by minimizing

$$E^2 = \left\| \boldsymbol{a}x - \boldsymbol{b} \right\|^2 = (a_1x - b_1)^2 + (a_2x - b_2)^2 + \cdots + (a_mx - b_m)^2.$$

The derivative of E^2 is zero at the point $\hat{x}$, if

$$(a_1x - b_1)a_1 + (a_2x - b_2)a_2 + \cdots + (a_mx - b_m)a_m = 0.$$

We are *minimizing the distance* from $\boldsymbol{b}$ to the line through $\boldsymbol{a}$, and calculus gives the same answer,

$$\hat{x} = \frac{a_1b_1 + a_2b_2 + \cdots + a_mb_m}{a_1^2 + a_2^2 + \cdots + a_m^2} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}},$$

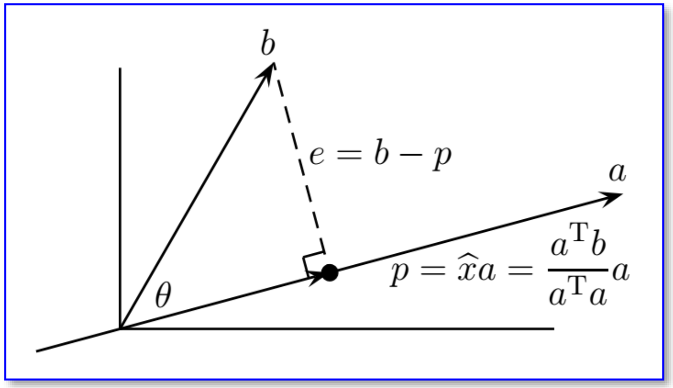
where $\boldsymbol{a} = (a_1, a_2, \ldots, a_m)^{\mathrm{T}}$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)^{\mathrm{T}}$.

Lemma

The least-squares solution to a problem $\boldsymbol{a}x = \boldsymbol{b}$ in one unknown is

$$\hat{x} = \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}.$$

(*The error vector* $\boldsymbol{e}$ *connecting* $\boldsymbol{b}$ *to* $\boldsymbol{p}$ *must be perpendicular to* $\boldsymbol{a}$)



I. Best Approximation – Least Squares Solution

Our next task is to establish a similar formula in the general case (Let $\boldsymbol{A}$ be an $(m \times n)$-matrix).

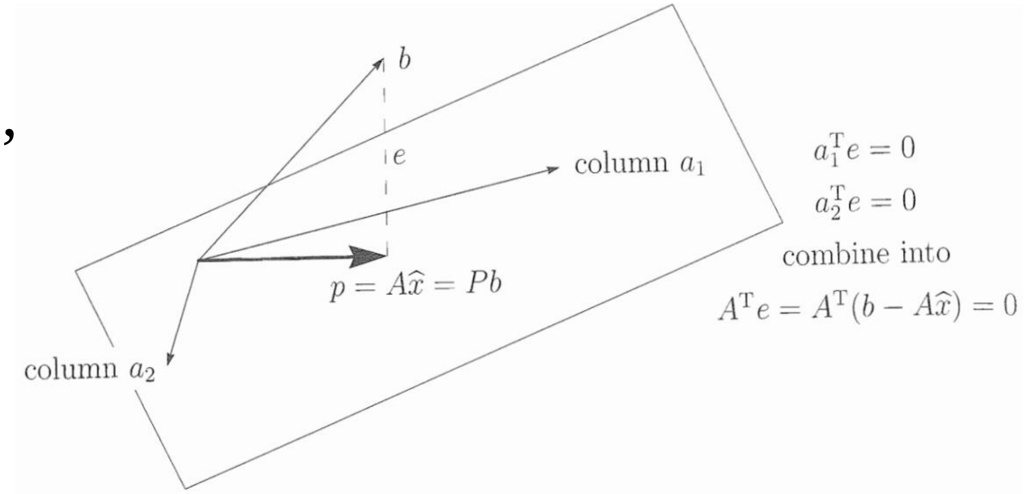
Suppose that $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent. We wish to find $\hat{\boldsymbol{x}}$ such that $||\boldsymbol{A}\hat{\boldsymbol{x}} - \boldsymbol{b}||$ is as small as possible.

The column space C($\boldsymbol{A}$) consists of all vectors of the form $\boldsymbol{Ax}$ with $\boldsymbol{x} \in \mathbf{R}^n$, i.e., $C(\boldsymbol{A}) = \{\boldsymbol{Ax} \mid \boldsymbol{x} \in \mathbf{R}^n\}$.

Since $\boldsymbol{Ax} = \boldsymbol{b}$ has no solution, we have $\boldsymbol{b} \neq \boldsymbol{Ax}$ for any $\boldsymbol{x} \in \mathbf{R}^n$,

that is, the vector $\boldsymbol{b}$ is not in the column space C($\boldsymbol{A}$).

Among the vectors of C($\boldsymbol{A}$), the vector $\boldsymbol{b}$ is nearest to <u>the projection of $\boldsymbol{b}$</u> in C($\boldsymbol{A}$).



Projection onto the column space of a 3 by 2 matrix

Namely, $\boldsymbol{A}\widehat{\boldsymbol{x}}$ should be the projection of $\boldsymbol{b}$ in C($\boldsymbol{A}$).

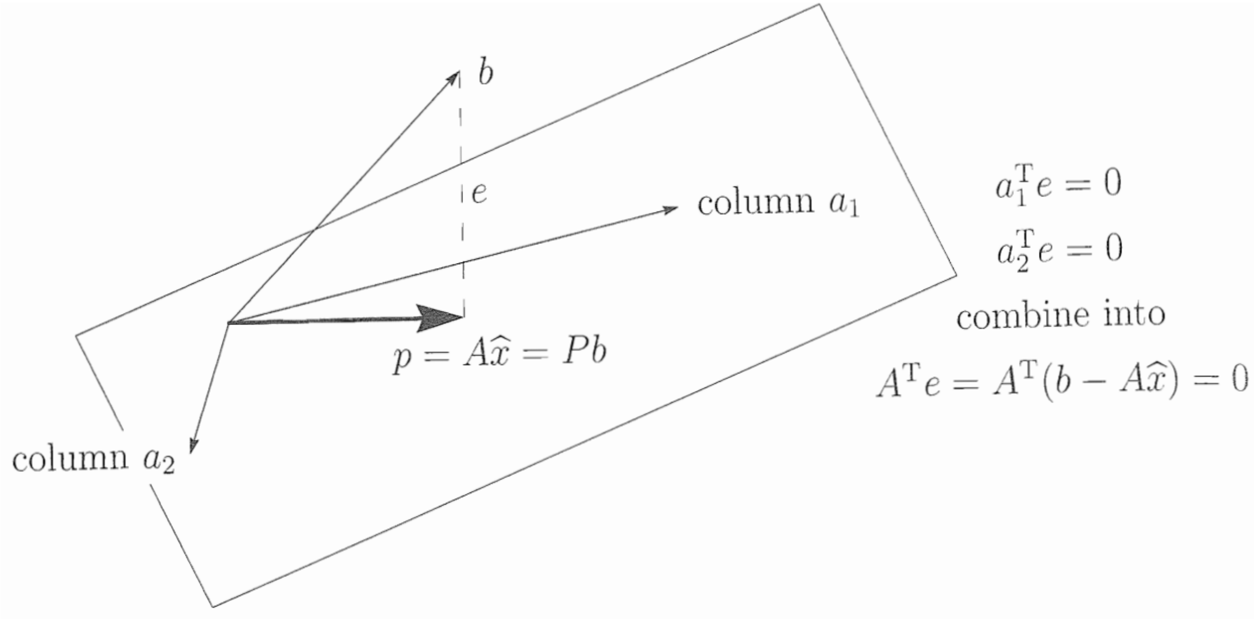
In other words, the difference $\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}}$ is perpendicular to C($\boldsymbol{A}$), i.e.,

$$\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}} \perp \mathrm{C}(A).$$

Equivalently, $\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}}$ is perpendicular to all columns of $\boldsymbol{A}$.

Since all vectors perpendicular to the column space lie in the *left nullspace*, thus $\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}}$ lies in the left nullspace of $\boldsymbol{A}$, so

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}}) = \boldsymbol{0}.$$



Projection onto the column space of a 3 by 2 matrix

Theorem 1. *If a system $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent (has no solution), its least-squares solution minimizes $||\boldsymbol{Ax} - \boldsymbol{b}||^2$:*

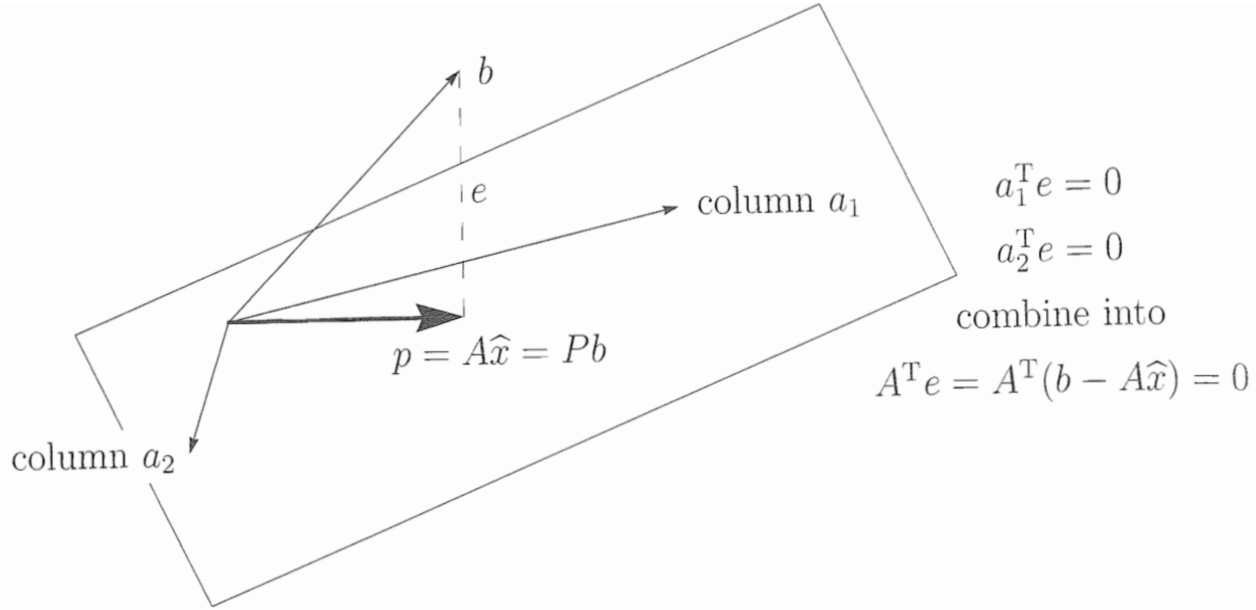
$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \qquad \textbf{(Normal equations)}$$

Moreover, if $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible, then

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \qquad \textbf{(Best estimate)}$$

The projection of $\boldsymbol{b}$ onto the column space is the nearest point $\boldsymbol{A}\hat{\boldsymbol{x}}$:

$$\boldsymbol{p} = \boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}. \qquad \textbf{(Projection)}$$



Projection onto the column space of a 3 by 2 matrix

Example 2. The following system of linear equations $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ has no solution

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 3 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix}.$$

We wish to find an approximation of $\boldsymbol{x}$, denoted by $\widehat{\boldsymbol{x}}$, such that $\boldsymbol{A}\widehat{\boldsymbol{x}}$ is very close to $\boldsymbol{b}$, that is, such that the 'error' $||\boldsymbol{b} - \boldsymbol{A}\widehat{\boldsymbol{x}}||$ is as small as possible.

There is a formula for finding such an approximation $\widehat{\boldsymbol{x}}$:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

Let $\widehat{\boldsymbol{x}} = (x_0, y_0, z_0)^{\mathrm{T}}$. Then

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 3 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 3 \\ 3 & 6 & 9 \\ 3 & 9 & 18 \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 6 \\ 15 \end{bmatrix},$$

Using the typical method, we can find solutions for (x_0, y_0, z_0), *one of* which is $(-4, 3, 0)$.

Remark 1. Suppose $\boldsymbol{b}$ is actually in the column space of $\boldsymbol{A}$—it is a combination $\boldsymbol{b} = \boldsymbol{Ax}$ of the columns. Then the projection of $\boldsymbol{b}$ is still $\boldsymbol{b}$:

b in column space $$\boldsymbol{p} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{Ax} = \boldsymbol{b}.$$

The closest point $\boldsymbol{p}$ is just $\boldsymbol{b}$ itself—which is obvious.

Remark 2. At the other extreme, suppose $\boldsymbol{b}$ is *perpendicular* to every column, so $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{0}$. In this case $\boldsymbol{b}$ projects to the zero vector:

b in left nullspace $$\boldsymbol{p} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{0} = \boldsymbol{0}.$$

Remark 3. When $\boldsymbol{A}$ is square and invertible, the column space is the whole space. Every vector projects to itself, $\boldsymbol{p}$ equals $\boldsymbol{b}$, and $\hat{\boldsymbol{x}} = \boldsymbol{x}$:

If A is invertible $$\boldsymbol{p} = \boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{A}\boldsymbol{A}^{-1}(\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \boldsymbol{b}.$$

When $\boldsymbol{A}$ is rectangular that is not possible.

Remark 4. Suppose $\boldsymbol{A}$ has only one column, containing $\boldsymbol{a}$. Then the matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is the number $\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}$ and $\hat{\boldsymbol{x}}$ is $\dfrac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{b}}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}}$. We return to the earlier formula.

II. Projection Matrices and the Matrix $A^{\mathrm{T}}A$

For a matrix $\boldsymbol{A}$ of size $m \times n$, the column space $\mathrm{C}(\boldsymbol{A})$ is a subspace of $\mathbf{R}^m$. Is there a projection from $\mathbf{R}^m$ to $\mathrm{C}(\boldsymbol{A})$?

As mentioned above, $\boldsymbol{A}\hat{\boldsymbol{x}}$ is the vector in $\mathrm{C}(\boldsymbol{A})$ that is the projection of $\boldsymbol{b}$ in $\mathrm{C}(\boldsymbol{A})$, and is nearest to $\boldsymbol{b}$, which is

$$\boldsymbol{A}\hat{\boldsymbol{x}} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

So the matrix $\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}$ is a transformation: $\boldsymbol{b} \mapsto \boldsymbol{A}\hat{\boldsymbol{x}}$.

Corollary. *Let* $\boldsymbol{A}$ *be a matrix of size* $m \times n$ *such that* $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ *is invertible. Then the projection of* $\mathbf{R}^m$ *to* $\mathrm{C}(\boldsymbol{A})$ *has matrix*

$$\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}.$$

It is easy to see that the matrix $\boldsymbol{P} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}$ is symmetric, and

$$\boldsymbol{P}^2 = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{A}\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\mathrm{T}} = \boldsymbol{P}.$$

Theorem 2. *Any symmetric matrix* $\boldsymbol{P}$ *with* $\boldsymbol{P}^2 = \boldsymbol{P}$ *represents a projection*. (任何对称的幂等矩阵都对应一个投影变换)

Proof. Let $\boldsymbol{P}$ have size $m \times m$.

We claim that $\boldsymbol{P}$ projects $\mathbf{R}^m$ to the column space $\mathrm{C}(\boldsymbol{P})$.

Let $\boldsymbol{v} \in \mathbf{R}^m$. Consider the difference $\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}$. A vector of $\mathrm{C}(\boldsymbol{P})$ has the form $\boldsymbol{P}\boldsymbol{x}$, and

$$(\boldsymbol{P}\boldsymbol{x})^{\mathrm{T}}(\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{v} - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}^{\mathrm{T}}\boldsymbol{P}\boldsymbol{v} = \boldsymbol{x}^{\mathrm{T}}(\boldsymbol{P} - \boldsymbol{P}^2)\boldsymbol{v} = 0.$$

Thus $\boldsymbol{v} - \boldsymbol{P}\boldsymbol{v}$ is perpendicular to $\boldsymbol{P}\boldsymbol{x}$, and so $\boldsymbol{P}\boldsymbol{v}$ is the projection of $\boldsymbol{v}$ in $\mathrm{C}(\boldsymbol{P})$.

Remark. In this case, $\boldsymbol{I} - \boldsymbol{P}$ also represents a projection.

There is a simple way to decide *whether $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible*.

Theorem 3. *The matrices $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ and $\boldsymbol{A}$ have the same nullspace.*
In particular, if $\boldsymbol{A}$ has full column rank, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible.

Proof. If $\boldsymbol{Ax} = \boldsymbol{0}$, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{0}$, and $\mathrm{N}(\boldsymbol{A}) \subseteq \mathrm{N}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})$.

Conversely, suppose that $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{0}$. Then

$$||\boldsymbol{Ax}||^2 = (\boldsymbol{Ax})^{\mathrm{T}}(\boldsymbol{Ax}) = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{Ax} = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{0} = 0,$$

which implies that $\boldsymbol{Ax} = \boldsymbol{0}$.

Thus $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) \subseteq \mathrm{N}(\boldsymbol{A})$, and so $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \mathrm{N}(\boldsymbol{A})$.

Let $\boldsymbol{A}$ be of size $m \times n$. Then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is of size $n \times n$.

If $\boldsymbol{A}$ has column rank n, then $\mathrm{N}(\boldsymbol{A}) = \{\boldsymbol{0}\}$, and so $\mathrm{N}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}) = \{\boldsymbol{0}\}$.
Therefore, $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is invertible.

Note: *$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is square and symmetric.*

Moreover, if $\boldsymbol{A}$ has independent columns, then $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$ is square, symmetric, and invertible.

(如果$\boldsymbol{A}$的列线性无关, 则$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$是方阵、对称而且可逆)

For Example,

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix}, \text{then}$$

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 5 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 8 \\ 8 & 30 \end{bmatrix} \text{ is invertible.}$$

$$\boldsymbol{B} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix}, \text{then}$$

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} = \begin{bmatrix} 1 & 1 & 1 \\ 3 & 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 3 & 9 \\ 9 & 27 \end{bmatrix} \text{ is } \boldsymbol{not} \text{ invertible.}$$

III. Least Squares Fitting of Data (数据的最小二乘拟合)

Look at a line

$$C + Dt = b.$$

If there is no experimental error, then two measurements of b will determine the line.

But if there is error, by a series of experiments, we obtain a system with two unknowns C, D:

$$\begin{cases} C + Dt_1 = b_1, \\ C + Dt_2 = b_2, \\ \quad \cdots \\ C + Dt_m = b_m. \end{cases}$$

In matrix form $\boldsymbol{Ax} = \boldsymbol{b}$, where

$$\boldsymbol{A} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_m \end{bmatrix}, \qquad \boldsymbol{x} = \begin{bmatrix} C \\ D \end{bmatrix}, \qquad \boldsymbol{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}.$$

If the system $\boldsymbol{Ax} = \boldsymbol{b}$ is inconsistent, then we wish to find $\widehat{\boldsymbol{x}} = \left(\hat{C}, \widehat{D}\right)^{\mathrm{T}}$ to minimize the squared error E^2:

$$E^2 = \left|\left|\boldsymbol{b} - \boldsymbol{Ax}\right|\right|^2 = (b_1 - C - Dt_1)^2 + \cdots + (b_m - C - Dt_m)^2.$$

By the theorem obtained above, we have

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

which is

$$\begin{bmatrix} m & t_1 + \cdots + t_m \\ t_1 + \cdots + t_m & t_1^2 + \cdots + t_m^2 \end{bmatrix}\begin{bmatrix} \hat{C} \\ \widehat{D} \end{bmatrix} = \begin{bmatrix} b_1 + \cdots + b_m \\ t_1 b_1 + \ \ldots + t_m b_m \end{bmatrix}.$$

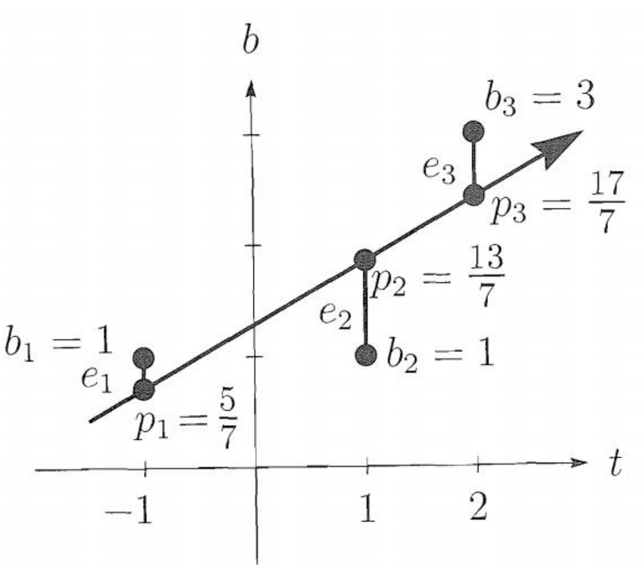
Theorem 4. *Let the measurements $b_1, b_2, \ldots, b_m$ be given at distinct points $t_1, t_2, \ldots, t_m$. Then the line*

$$\hat{C} + \widehat{D}t = b$$

with $\widehat{\boldsymbol{x}} = \left(\hat{C}, \widehat{D}\right)^{\mathrm{T}}$ which minimizes E^2 comes from least squares:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

Example 3 Three measurements b_1, b_2, b_3 are marked on the figure:



$b = 1$ at $t = -1$; $b = 1$ at $t = 1$; $b = 3$ at $t = 2$.

$$\begin{bmatrix} 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}.$$

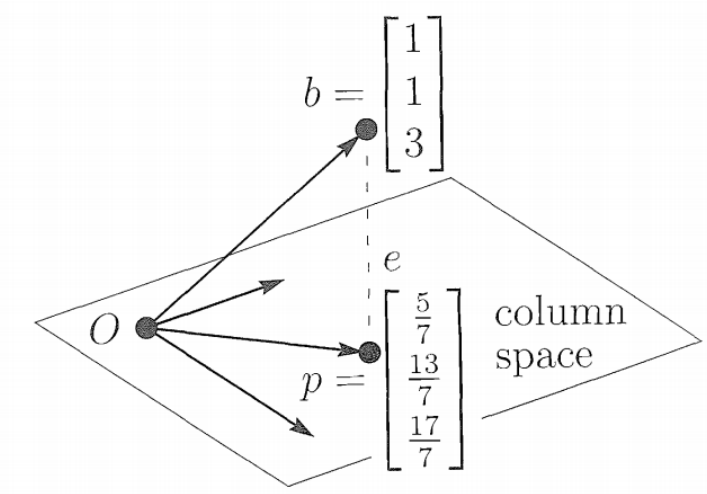
$\boldsymbol{Ax} = \boldsymbol{b}$ can't be solved because the points are not on a line.

Therefore they are solved by least squares:

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\widehat{\boldsymbol{x}} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

is $\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} C \\ D \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \end{bmatrix} \Rightarrow \hat{C} = \frac{9}{7},\ \hat{D} = \frac{4}{7}.$

The best line is $\frac{9}{7} + \frac{4}{7}t$.



Remark. The mathematics of least squares is not limited to fitting the data by straight lines– *nonlinear least squares*.

Key words:

Least Squares

The matrix $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$; *Projection matrix*

Fitting of Data

Homework

See Blackboard