§6.5 Least-Square Problems

In real application, $A\vec{x}=\vec{b}$ has no solution, we want to find $\hat{\vec{x}}$ such that $A\hat{\vec{x}}$ as close as possible to $\vec{b}$.

general least-squares problem: To find an $\hat{\vec{x}}$ that makes $\|b-A\hat{\vec{x}}\|$ as small as possible.

Def: If A is $m\times n$ and $\vec{b}$ is in $\mathbb{R}^m$, a least-squares solution of $A\vec{x}=\vec{b}$ is an $\hat{\vec{x}}$ in $\mathbb{R}^n$ such that

$$\|b-A\hat{\vec{x}}\| \leq \|b-A\vec{x}\|$$

for all $\vec{x}$ in $\mathbb{R}^n$.

* Solution of the General Least-Squares Problem

Thm: The set of least-squares solutions of $A\vec{x}=\vec{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\vec{x}=A^T\vec{b}$. — Normal equation for $Ax=b$

Proof: Let $\hat{\vec{b}}=\text{proj}_{\text{Col}A}\vec{b}$.

Since $\hat{\vec{b}}\in\text{Col}A$, $A\vec{x}=\hat{\vec{b}}$ is consistent. There exists an $\hat{\vec{x}}\in\mathbb{R}^n$ such that

$$A\hat{\vec{x}}=\hat{\vec{b}}.$$

By the Best Approximation Theorem,

$$\|\vec{b}-A\hat{\vec{x}}\| = \|\vec{b}-\hat{\vec{b}}\| \leq \|b-A\vec{x}\| \text{ for any } \vec{x}\in\mathbb{R}^n.$$

Now we are going to prove $\hat{\vec{x}}$ satisfy

$$A^TA\vec{x}=A^T\vec{b}.$$

In fact, $\vec{b}-\hat{\vec{b}}$ is orthogonal to $\text{Col}A$.

Let $A = [\vec{a}_1, \cdots, \vec{a}_n]$. Then $a_j^T(\vec{b} - \hat{\vec{b}}) = 0$, $j = 1, \cdots, n$.

Hence $[\vec{a}_1, \cdots, \vec{a}_n]^T (\vec{b} - \hat{\vec{b}}) = 0$

That is $A^T(\vec{b} - \hat{\vec{b}}) = 0$

$$A^T(\vec{b} - A\hat{\vec{x}}) = 0$$

i.e. $A^T A \hat{\vec{x}} = A^T \vec{b}$.

Conversely, suppose $\hat{\vec{x}}$ satisfies $A^T A \hat{\vec{x}} = A^T \vec{b}$. Then

$$A^T(A\hat{\vec{x}} - \vec{b}) = 0$$

That is, $A\hat{\vec{x}} - \vec{b}$ is orthogonal to $\vec{a}_j$, $j = 1, \cdots, n$.

Hence $A\hat{\vec{x}} - \vec{b}$ is orthogonal to all vectors in $\mathrm{Col}A$.

Therefore $\vec{b} = A\hat{\vec{x}} + b - A\hat{\vec{x}}$ with $A\hat{\vec{x}} \in \mathrm{Col}A$ and $b - A\hat{\vec{x}} \in (\mathrm{Col}A)^\perp$.

By the uniqueness of orthogonal decomposition, $A\hat{x}$ must be the orthogonal projection of $\vec{b}$ onto $\mathrm{Col}A$. That is, $A\hat{x} = \hat{b}$ and $\hat{x}$ is a least-squares solution.

Example: Find a least-squares solution of the inconsistent system $A\vec{x} = \vec{b}$ for

$$A = \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix}$$

Solution: $$A^T A = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}$$

$$A^T \vec{b} = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \\ 11 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

The normal equation is

$$\begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}^{-1}\begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$= \frac{1}{84}\begin{pmatrix} 5 & -1 \\ -1 & 17 \end{pmatrix}\begin{pmatrix} 19 \\ 11 \end{pmatrix}$$

$$= \frac{1}{84}\begin{pmatrix} 84 \\ 168 \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Example: Find a least-square solution of $A\vec{x} = \vec{b}$ for

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix}$$

Solution: $$A^TA = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$$

$$= \begin{pmatrix} 6 & 2 & 2 & 2 \\ 2 & 2 & 0 & 0 \\ 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 2 \end{pmatrix}$$

$$A^T\vec{b} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} -3 \\ -1 \\ 0 \\ 2 \\ 5 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ -4 \\ 2 \\ 6 \end{pmatrix}$$

The augmented matrix for $A^TA\vec{x} = A^T\vec{b}$ is

$$\begin{pmatrix} 6 & 2 & 2 & 2 & 4 \\ 2 & 2 & 0 & 0 & -4 \\ 2 & 0 & 2 & 0 & 2 \\ 2 & 0 & 0 & 2 & 6 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 & 1 & 3 \\ 0 & 1 & 0 & -1 & -5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\begin{cases} x_1 + x_4 = 3 \\ x_2 - x_4 = -5 \\ x_3 - x_4 = -2 \end{cases} \Rightarrow \begin{cases} x_1 = 3 - x_4 \\ x_2 = -5 + x_4 \\ x_3 = -2 + x_4 \end{cases}$$

So the general least-squares solution of $A\vec{x} = \vec{b}$ has the form

$$\hat{x} = \begin{pmatrix} 3 - x_4 \\ -5 + x_4 \\ -2 + x_4 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -5 \\ -2 \\ 0 \end{pmatrix} + x_4 \begin{pmatrix} -1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Thm: Let A be an $m \times n$ matrix. The following statements are logically equivalent:

a) The equation $A\vec{x} = \vec{b}$ has a unique least-squares solution for each $\vec{b}$ in $\mathbb{R}^m$.

b) The columns of A are linearly independent.

c) The matrix A^TA is invertible.

Alternative Calculations of Least-Squares Solutions

1) Find a least-squares solution of $A\vec{x} = \vec{b}$ when the columns of A are orthogonal.

Example: Find a least-squares solution of $A\vec{x}=\vec{b}$ for

$$A = \begin{pmatrix} 1 & -6 \\ 1 & -2 \\ 1 & 1 \\ 1 & 7 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} -1 \\ 2 \\ 1 \\ 6 \end{pmatrix}$$

Solution: Because the columns $\vec{a}_1$ and $\vec{a}_2$ of A are orthogonal, the orthogonal projection of $\vec{b}$ onto $\mathrm{Col}A$ is given by

$$\hat{b} = \frac{\vec{b}\cdot\vec{a}_1}{\vec{a}_1\cdot\vec{a}_1}\vec{a}_1 + \frac{\vec{b}\cdot\vec{a}_2}{\vec{a}_2\cdot\vec{a}_2}\vec{a}_2$$

$$= \frac{8}{4}\vec{a}_1 + \frac{45}{90}\vec{a}_2 = 2\vec{a}_1 + \frac{1}{2}\vec{a}_2$$

$$= [\vec{a}_1, \vec{a}_2]\begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix}$$

Hence $\hat{x} = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix}$.

2) If the columns of A are linearly independent, the least-squares solution can often be computed more reliable through a QR factorization of A.

Thm: Given an $m\times n$ matrix A with linearly independent columns, let $A=QR$ be a QR factorization of A. Then, for each $\vec{b}$ in $\mathbb{R}^m$, the equation $A\vec{x}=\vec{b}$ has a unique least-squares solution, given by

$$\hat{x} = R^{-1}Q^T\vec{b}$$

Proof: Let $\hat{\vec{x}} = R^{-1}Q^T\vec{b}$. Then

$$A\hat{\vec{x}} = QR\hat{\vec{x}} = QRR^{-1}Q^T\vec{b} = QQ^T\vec{b}.$$

Note that the columns of Q form an orthonormal basis for Col$A$. Hence $QQ^T\vec{b}$ is the orthogonal projection $\hat{\vec{b}}$ of $\vec{b}$ onto Col$A$.

So $\hat{\vec{x}}$ is a least-square solution.

The uniqueness follows from the fact that the columns of A are linearly independent.

Ex: Find the least-squares solution of $A\vec{x} = \vec{b}$ for

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}, \quad \vec{b} = \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix}.$$

using QR factorization.

Solution: First, we find the orthonormal basis for Col$A$ using Gram-Schmit process. Let $A = [\vec{a}_1, \vec{a}_2, \vec{a}_3]$

$$\vec{v}_1 = \vec{a}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \quad \vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|} = \frac{1}{\sqrt{1^2+1^2+1^2+1^2}}\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\vec{v}_2 = \vec{a}_2 - (\vec{a}_2 \cdot \vec{u}_1)\vec{u}_1$$

$$= \begin{pmatrix} 3 \\ 1 \\ 1 \\ 3 \end{pmatrix} - \left(\frac{1}{2}\cdot 3 + 1\cdot\frac{1}{2} + 1\cdot\frac{1}{2} + 3\cdot\frac{1}{2}\right)\begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 3 \\ 1 \\ 1 \\ 3 \end{pmatrix} - 4 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

$$\vec{u}_2 = \frac{\vec{v}_2}{\|\vec{v}_2\|} = \frac{1}{\sqrt{1^2+(-1)^2+(-1)^2+1^2}} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix}$$

$$= \frac{1}{2} \begin{pmatrix} 1 \\ -1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$\vec{v}_3 = \vec{a}_3 - (\vec{a}_3 \cdot \vec{u}_1)\vec{u}_1 - (\vec{a}_3 \cdot \vec{u}_2)\vec{u}_2$$

$$= \begin{pmatrix} 5 \\ 0 \\ 2 \\ 3 \end{pmatrix} - \left(5 \cdot \tfrac{1}{2} + 0 \cdot \tfrac{1}{2} + 2 \cdot \tfrac{1}{2} + 3 \cdot \tfrac{1}{2}\right) \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - \left(5 \cdot \tfrac{1}{2} - 0 \cdot \tfrac{1}{2} - 2 \cdot \tfrac{1}{2} + 3 \cdot \tfrac{1}{2}\right) \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 5 \\ 0 \\ 2 \\ 3 \end{pmatrix} - 5 \cdot \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} - 3 \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

$$= \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix}$$

$$\vec{u}_3 = \frac{1}{\sqrt{1^2+(-1)^2+1^2+(-1)^2}} \begin{pmatrix} 1 \\ -1 \\ 1 \\ -1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}$$

$$Q = [\vec{u}_1, \vec{u}_2, \vec{u}_3] = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$$

$$R = Q^T Q R = Q^T A = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 1 & 3 & 5 \\ 1 & 1 & 0 \\ 1 & 1 & 2 \\ 1 & 3 & 3 \end{pmatrix}$$

$$= \begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix}$$

Then $Q^T b = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 5 \\ 7 \\ -3 \end{pmatrix} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix}$

The least-squares solution $\hat{x}$ satisfies $R\vec{x} = Q^T \vec{b}$. That is,

$$\begin{pmatrix} 2 & 4 & 5 \\ 0 & 2 & 3 \\ 0 & 0 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 6 \\ -6 \\ 4 \end{pmatrix}$$

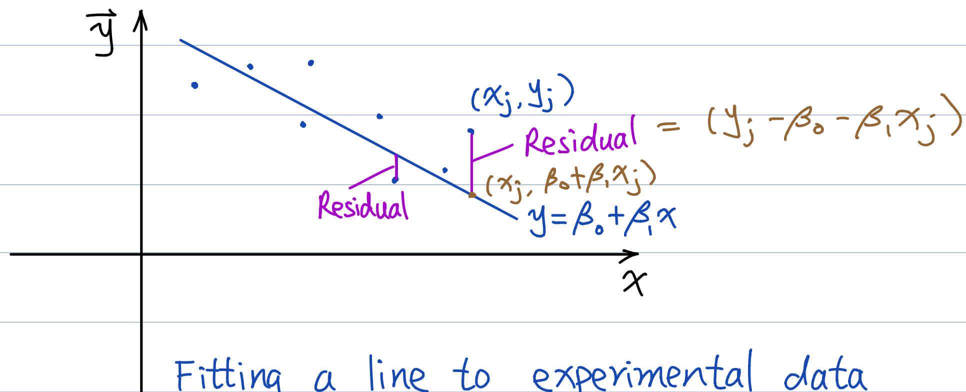
$$\begin{cases} 2x_1 + 4x_2 + 5x_3 = 6 \\ \qquad 2x_2 + 3x_3 = -6 \\ \qquad\qquad 2x_3 = 4 \end{cases} \Rightarrow \begin{cases} x_1 = 10 \\ x_2 = -6 \\ x_3 = 2 \end{cases}$$

Exercise: What can you say about the least-squares solution of $A\vec{x} = \vec{b}$ when $\vec{b}$ is orthogonal to the columns of A?

Solution: If $\vec{b}$ is orthogonal to the columns of A, then the projection of $\vec{b}$ onto the column space of A is $\vec{0}$. In this case, a least-squares solution $\hat{x}$ of $A\vec{x} = \vec{b}$ satisfies $A\hat{\vec{x}} = \vec{0}$.

§6.6 Applications to Linear Models

* Least - Squares Lines



Fitting a line to experimental data

Def: The least-squares Line is the line $y = \beta_0 + \beta_1 x$ that minimizes the sum of the squares of the residuals.

That is, $\min_{\vec{\beta}} \|X\vec{\beta} - \vec{y}\|^2$

where $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \cdots \\ 1 & x_n \end{pmatrix}$, $\vec{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$, $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$.

Remark: Computing the least-squares solution of $X\beta = \vec{y}$ is equivalent to finding the β that determines the least-squares line.

Ex: Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points $(2,1)$, $(5,2)$, $(7,3)$ and $(8,3)$.

Solution:

$$X = \begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix}, \qquad \vec{y} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}$$

The normal equation is $X^T X \vec{\beta} = X^T \vec{y}$

where

$$X^T X = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix} = \begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}$$

$$X^T \vec{y} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix} = \begin{pmatrix} 9 \\ 57 \end{pmatrix}$$

Hence $\vec{\beta} = (X^T X)^{-1} X^T \vec{y} = \begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix}^{-1} \begin{pmatrix} 9 \\ 57 \end{pmatrix}$

$$= \frac{1}{84} \begin{pmatrix} 142 & -22 \\ -22 & 4 \end{pmatrix} \begin{pmatrix} 9 \\ 57 \end{pmatrix}$$

$$= \frac{1}{84} \begin{pmatrix} 24 \\ 30 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{2}{7} \\ \frac{5}{14} \end{pmatrix}$$

Thus the least-squares line has the equation

$$y = \frac{2}{7} + \frac{5}{14} x$$