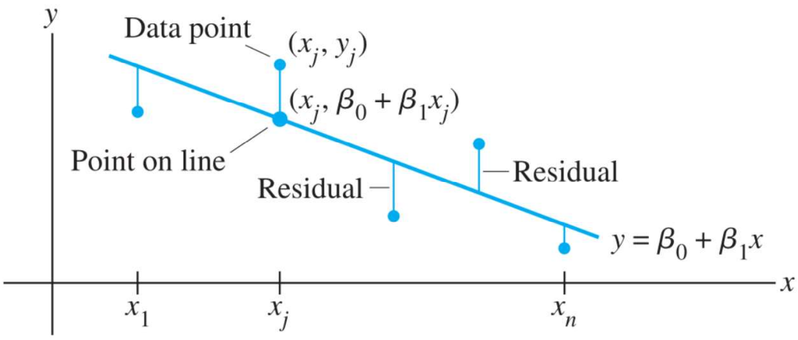
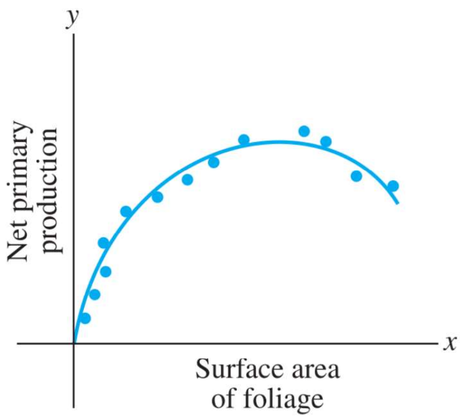
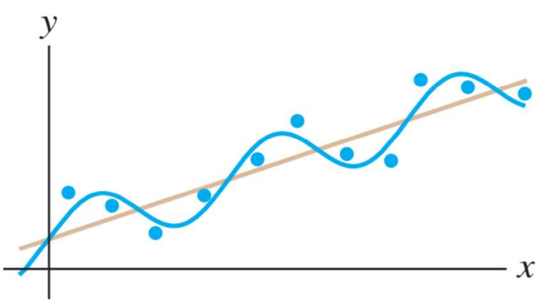
Leas-square problems and its applications

- In experimental science, to find the rules underlying the experimental datum, it is important to approximate them with well-known functions like polynomials or sinus-cosines.

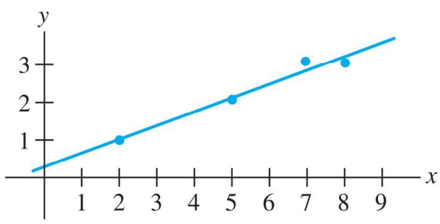






EXAMPLE 1 Find the equation $y = \beta_0 + \beta_1 x$ of the least-squares line that best fits the data points (2, 1), (5, 2), (7, 3), and (8, 3).

Solution From the graphics, it is clear that the points are not colinear, hence the set of equations



$$\begin{cases} \beta_0 + 2\beta_1 = 1 \\ \beta_0 + 5\beta_1 = 2 \\ \beta_0 + 7\beta_1 = 3 \\ \beta_0 + 8\beta_1 = 3 \end{cases}$$

has no solutions.

To best approximate the data, we would like the sum of squares

$$(*) \quad [(\beta_0 + 2\beta_1) - 1]^2 + [(\beta_0 + 5\beta_1) - 2]^2 + [(\beta_0 + 7\beta_1) - 3]^2 + [(\beta_0 + 8\beta_1) - 3]^2$$

to attain the minimum. In other words, we want to find $(\beta_1, \beta_2) \in \mathbb{R}^2$ such that the above sum of square attains the minimum.

Observe that (*) equals the square of the distance between

$$\begin{pmatrix} \beta_0 + 2\beta_1 \\ \beta_0 + 5\beta_1 \\ \beta_0 + 7\beta_1 \\ \beta_0 + 8\beta_1 \end{pmatrix} \in \mathbb{R}^4 \quad \text{and} \quad \underbrace{\begin{pmatrix} 1 \\ 2 \\ 3 \\ 3 \end{pmatrix}}_{\vec{y}} \in \mathbb{R}^4$$

$$= \underbrace{\begin{pmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}}_{\vec{\beta}}$$

In other worlds, we are look for $\vec{\beta} \in \mathbb{R}^2$ such that

$$\| \vec{y} - A\vec{\beta} \|^2 \quad \text{attain the minimum.}$$

Notice that

$$\{ A\vec{\beta} \mid \beta \in \mathbb{R}^2 \} = \operatorname{Span}\left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \\ 7 \\ 8 \end{bmatrix} \right\} = \operatorname{Col}(A)$$

is a subspace of $\mathbb{R}^4$. Hence the question is reduced to find the distance between $\vec{y}$ and the subspace $W = \operatorname{Col}(A)$.

The problem can be solved by projecting $\vec{y}$ onto W, i.e. β satisfies the equation

$$(**) \qquad A\vec{\beta} = \operatorname{Proj}_W \vec{y}.$$

The projection $\text{proj}_W(\vec{y})$ can be calculated directly if we know an orthogonal basis of W, which can be obtained by a Gram-Schmidt process applied to the column vectors of A.

Here we proceed in another way. Notice that

$$\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y}) \text{ must be orthogonal to } \text{col}(A).$$

In other words,

$$A^t(\vec{y} - \text{proj}_{\text{col}(A)}(\vec{y})) = 0$$

With (**), this implies that

$$A^t(\vec{y} - A\vec{\beta}) = 0, \quad \text{i.e.} \quad \boxed{A^tA\vec{\beta} = A^t\vec{y}.}$$

We can solve the last equation to find $\vec{\beta}$. Plug in A and $\vec{y}$, we get

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 5 \\ 1 & 7 \\ 1 & 8 \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 2 & 5 & 7 & 8 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 3 \end{bmatrix}.$$

$$\begin{pmatrix} 4 & 22 \\ 22 & 142 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 9 \\ 57 \end{pmatrix} \Rightarrow \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 2/7 \\ 5/14 \end{pmatrix}$$

So the least-square line is $y = \frac{2}{7} + \frac{5}{14}x$. $\boxed{\S}$

EXAMPLE 2 Suppose data points $(x_1, y_1), \ldots, (x_n, y_n)$ appear to lie along some sort of parabola instead of a straight line. For instance, if the x-coordinate denotes the production level for a company, and y denotes the average cost per unit of operating at a level of x units per day, then a typical average cost curve looks like a parabola that opens upward (Figure 3). In ecology, a parabolic curve that opens downward is used to model the net primary production of nutrients in a plant, as a function of the surface area of the foliage (Figure 4). Suppose we wish to approximate the data by an equation of the form

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 \tag{3}$$

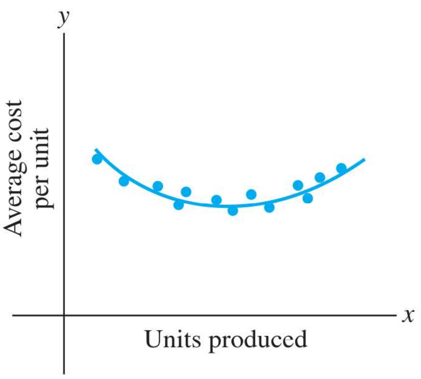
Describe the linear model that produces a "least-squares fit" of the data by equation (3).

Solution: As before, we want to find $(\beta_0, \beta_1, \beta_2) \in \mathbb{R}^3$ such that

$$\sum_{i=1}^{n} \left[(\beta_0 + \beta_1 x_i + \beta_2 x_i^2) - y_i \right]^2$$

attains the minimum. Let

$$\vec{x}_\beta = \begin{pmatrix} \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \\ \beta_0 + \beta_1 x_2 + \beta_2 x_2^2 \\ \vdots \\ \beta_0 + \beta_1 x_n + \beta_2 x_n^2 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$$



$$= \underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}}_{\vec{\beta}} \in \mathbb{R}^n .$$

then the question can be reformulated as finding the minimum of

$$\|\vec{y} - A\vec{\beta}\|^2, \quad \beta \in \mathbb{R}^3.$$

As $\{A\vec{\beta} \mid \beta \in \mathbb{R}^3\} = \mathrm{Col}(A)$ is a subspace of $\mathbb{R}^n$, the question can be solved by solving the equation

$$(*) \qquad A\vec{\beta} = \mathrm{Proj}_{\mathrm{Col}(A)}(\vec{y}).$$

As before, notice that $\vec{y} - \mathrm{Proj}_{\mathrm{Col}(A)}(\vec{y})$ is orthogonal to $\mathrm{Col}(A)$, this can be further transformed to

$$0 = A^t(\vec{y} - \mathrm{Proj}_{\mathrm{Col}(A)}(\vec{y})) \overset{(*)}{=} A^t(\vec{y} - A\vec{\beta}), \text{ i.e. } \boxed{A^tA\vec{\beta} = A^t\vec{y}}.$$

- All these questions lead naturally to the least-square problem (最小平方问题): The equation $A\mathbf{x} = \mathbf{b}$ is not necessarily solvable due to errors (误差) in the experimental datum, the best hope is then to find $\mathbf{x}_0$ such that $\|A\mathbf{x}_0 - \mathbf{b}\|$ attains the minimum.

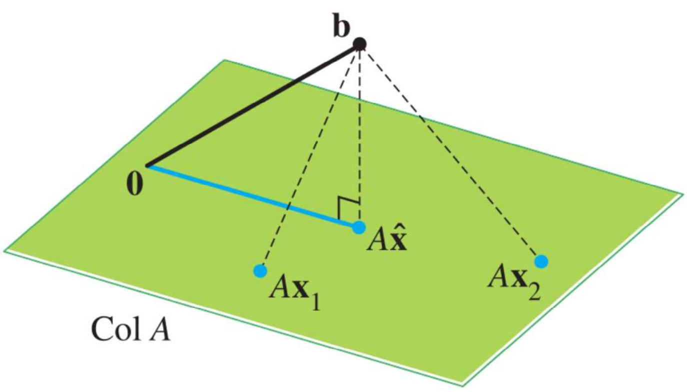
If A is $m \times n$ and $\mathbf{b}$ is in $\mathbb{R}^m$, a **least-squares solution** of $A\mathbf{x} = \mathbf{b}$ is an $\hat{\mathbf{x}}$ in $\mathbb{R}^n$ such that

$$\|\mathbf{b} - A\hat{\mathbf{x}}\| \le \|\mathbf{b} - A\mathbf{x}\|$$

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

Solution to the least-square problem

- Such a problem looks very much to find the distance between a point and a subspace in $\mathbb{R}^n$. Indeed, we can solve it by the best approximation theorem.
- Notice that the set $\{A\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n\} = \mathrm{Im}(A)$ coincides with the column space $\mathrm{Col}(A)$. Hence the least square problem can be reformulated as finding the distance between the point $\mathbf{b}$ and the subspace $\mathrm{Col}(A)$, i.e. to find the projection $\mathbf{b}_0$ of $\mathbf{b}$ in $\mathrm{Col}(A)$ and to find $\mathbf{x}_0$ such that $A\mathbf{x}_0 = \mathbf{b}_0$.



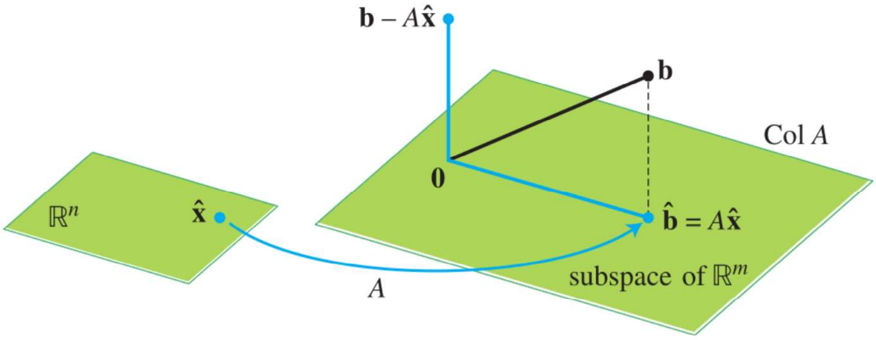


FIGURE 2 The least-squares solution $\hat{\mathbf{x}}$ is in $\mathbb{R}^n$.

- Theorem

The set of least-squares solutions of $A\mathbf{x} = \mathbf{b}$ coincides with the nonempty set of solutions of the normal equations $A^TA\mathbf{x} = A^T\mathbf{b}$.

Proof: We want to find $\vec{x}$ such that $\| A\vec{x} - \vec{b} \|$ attains the minimum. As we have explained before,

$$\{ A\vec{x} \mid \vec{x} \in \mathbb{R}^n \} = \mathrm{col}(A),$$

hence $\vec{x}$ will be solution to the equation

$$(*) \qquad A\vec{x} = \mathrm{proj}_{\mathrm{col}(A)}(\vec{b})$$

As before, $\vec{b} - \text{proj}_{\text{col}(A)}(\vec{b})$ will be orthogonal to $\text{col}(A)$, hence

$$A^t(\vec{b} - \text{proj}_{\text{col}(A)}(\vec{b})) = 0$$

$$(*) \quad \|$$

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

So $\vec{x}$ satisfies the equation

$$A^t A \vec{x} = A^t \vec{b}. \qquad \square$$

EXAMPLE 1 Find a least-squares solution of the inconsistent system $A\mathbf{x} = \mathbf{b}$ for

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

<u>Solution</u> According to the theorem, we need to solve the system

$$A^t A \vec{x} = A^t \vec{b}$$

We calculate

$$A^t A = \begin{pmatrix} 4 & 0 & 1 \\ 0 & 2 & 1 \end{pmatrix} \begin{pmatrix} 4 & 0 \\ 0 & 2 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 17 & 1 \\ 1 & 5 \end{pmatrix}.$$

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

So the equation becomes

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

$$\Rightarrow \quad \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 17 & 1 \\ 1 & 5 \end{bmatrix}^{-1} \begin{bmatrix} 19 \\ 11 \end{bmatrix} = \frac{1}{84} \begin{bmatrix} 5 & -1 \\ -1 & 17 \end{bmatrix} \begin{bmatrix} 19 \\ 11 \end{bmatrix}$$

$$= \frac{1}{84} \begin{bmatrix} 84 \\ 168 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

- Theorem

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

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

b. The columns of A are linearly independent.

c. The matrix A^TA is invertible.

When these statements are true, the least-squares solution $\hat{\mathbf{x}}$ is given by

$$\hat{\mathbf{x}} = (A^TA)^{-1}A^T\mathbf{b} \tag{4}$$

proof By the previous theorem, least-square solution to the equation $A\vec{x} = \vec{b}$ is the set of solutions to

$$A^t A \vec{x} = A^t \vec{b}. \qquad (\Rightarrow \text{ second assertion}).$$

The solution is unique if and only if $A^t A$ is invertible, hence the equivalence a) $\Leftrightarrow$ c)

For the matrix $A^t A$ to be invertible, i.e.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "A"] \arrow[rr, bend right, "A^t A \text{ invertible}"'] & \mathbb{R}^m \arrow[r, "A^t"] & \mathbb{R}^n \end{tikzcd} \quad \Leftrightarrow \quad \mathrm{Ker}(A^t A) = 0$$

as A is a square matrix.

claim: $\mathrm{Ker}(A^t A) = \mathrm{Ker}(A)$.

It is clear that $A\vec{x} = 0$ implies $A^t A \vec{x} = 0$, hence $\mathrm{Ker}(A) \subset \mathrm{Ker}(A^t A)$. Conversely, if $A^t A \vec{x} = 0$, then

$$0 = \vec{x}^{\,t} A^t A x = (A\vec{x})^t \cdot A\vec{x} = \| A\vec{x} \|^2$$

$$\Rightarrow \quad A\vec{x} = 0,$$

So $\mathrm{Ker}(A^t A) \subset \mathrm{Ker}(A)$. Combine them, get the equality.

Now $A^t A$ invertible $\Leftrightarrow \mathrm{Ker}(A^t A) = 0 \Leftrightarrow \mathrm{Ker}(A) = 0$

$\Leftrightarrow$ column vectors of A linearly independent, get b) $\Leftrightarrow$ c)

- Theorem

Given an $m \times n$ matrix A with linearly independent columns, let $A = QR$ be a QR factorization of A as in Theorem 12. Then, for each $\mathbf{b}$ in $\mathbb{R}^m$, the equation $A\mathbf{x} = \mathbf{b}$ has a unique least-squares solution, given by

$$\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b} \tag{6}$$

Proof By the previous theorem, the equation $A\vec{x} = \vec{b}$ has unique least square solution, as A has linearly independent column vectors. The solution is given by

$$\vec{x} = (A^tA)^{-1}A^t\vec{b} = ((QR)^tQR)^{-1}(QR)^t\vec{b}$$

$$= (R^t\underbrace{Q^tQ}_{I}R)^{-1}(R^tQ^t)\vec{b} = R^{-1}\underbrace{(R^t)^{-1}R^t}_{I}Q^t\vec{b} = R^{-1}Q^t\vec{b}. \quad \boxed{E}$$