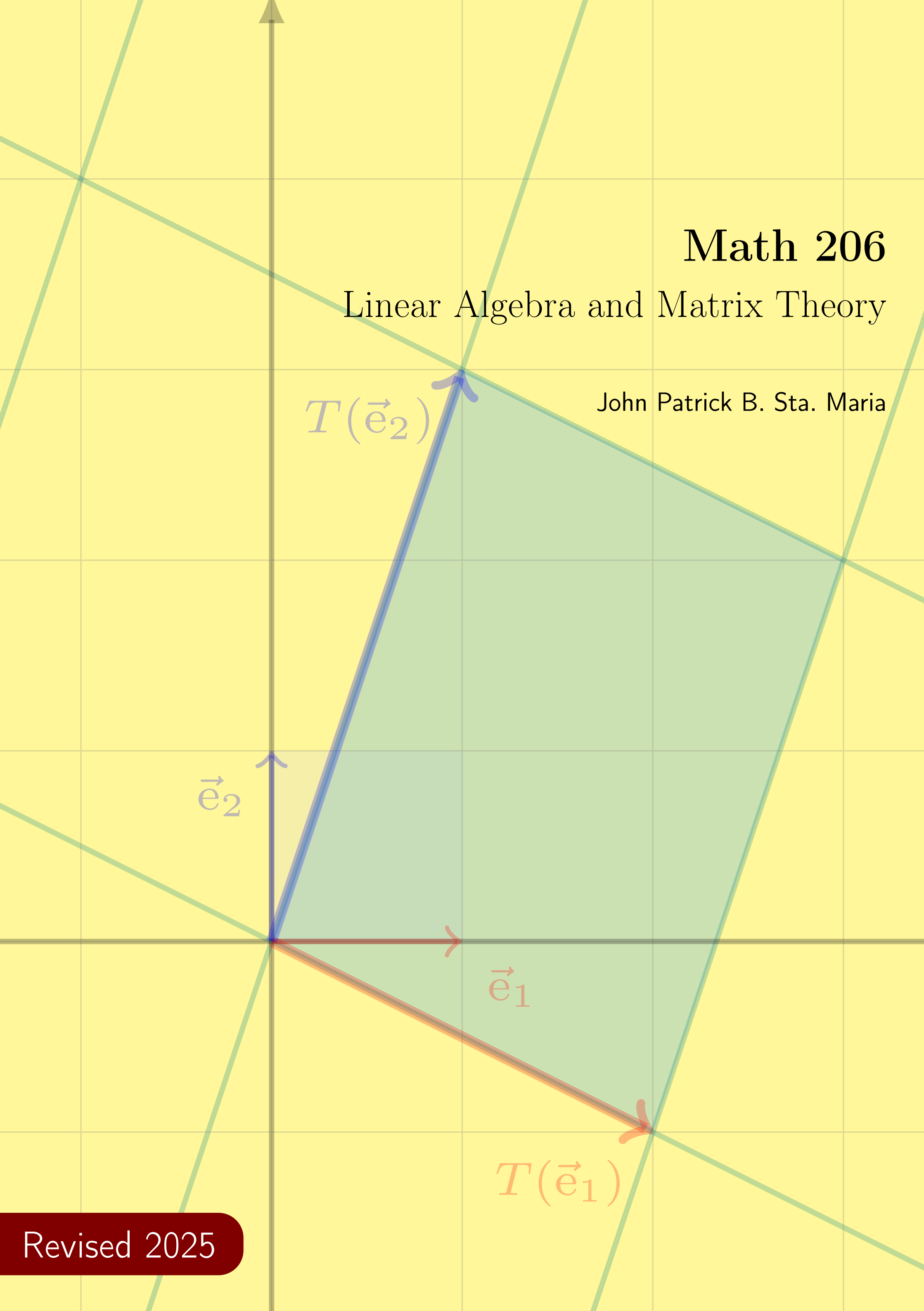
Math 206

Linear Algebra and Matrix Theory

John Patrick B. Sta. Maria



Revised 2025

Lesson 4.6: Least Squares Approximation

Learning Outcomes: At the end of the lesson, the students are able to

1. solve the minimum-distance problem using orthogonal projections;

2. compute least-squares solutions to linear systems;

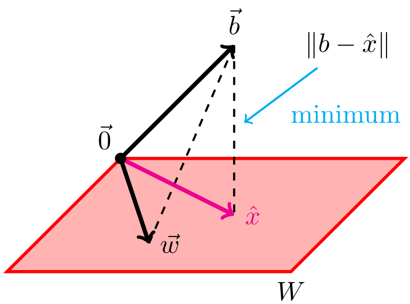
3. apply least-squares method to solve polynomial-fitting problems.

Best Approximation

Let W be a subspace of $\mathbb{R}^n$ and $\vec{b}$ be a given vector. In this lesson, we are interested to find the vector $\vec{w}$ in the subspace W which *best approximates* $\vec{b}$ in the sense that the distance $\|\vec{b}-\vec{w}\|$ is minimum. We state this problem formally as follows.

Minimum - Distance Problem in $\mathbb{R}^n$
Given a subspace W and a vector $\vec{b}$, find a vector $\hat{x} \in W$ such that $\|\vec{b}-\hat{x}\| \leq \|\vec{b}-\vec{w}\|$, for all $\vec{w} \in W$. In other formulation, we solve the minimization problem:

$$\begin{cases} \text{minimize:} & \|\vec{b}-\vec{w}\| \\ \text{subject to:} & \vec{w} \in W \end{cases}$$



Intuitively, if we think of W as a plane and $\vec{b}$ as a point in space, then the point in W nearest to $\vec{b}$ is the foot of an altitude to W. In linear algebra, this is what we call as the orthogonal projection onto W.

Theorem 284: The Best Approximation Theorem

If W is a subspace and $\vec{b}$ is a vector, then $\hat{x} = \operatorname{proj}_W \vec{b}$ is the unique solution to the *minimum-distance problem.* In other words, the vector in W of least distance from $\vec{b}$ is $\operatorname{proj}_W \vec{b}$.

Proof. Let $\vec{b} \in \mathbb{R}^n$ and let $\vec{w} \in W$. Note that we may write

$$\vec{b} - \vec{w} = (\vec{b} - \operatorname{proj}_W \vec{b}) + (\operatorname{proj}_W \vec{b} - \vec{w}).$$

Recall that $\vec{b} - \operatorname{proj}_W \vec{b} \in W^\perp$. Since $\operatorname{proj}_W \vec{b}$ and $\vec{w}$ are both elements of W, it follows that $\operatorname{proj}_W \vec{b} - \vec{w} \in W$. Thus, $\vec{b} - \operatorname{proj}_W \vec{b}$ and $\operatorname{proj}_W \vec{b} - \vec{w}$ are orthogonal. We may apply Theorem 30 to the above equation and get

$$\|\vec{b} - \vec{w}\|^2 = \|\vec{b} - \operatorname{proj}_W \vec{b}\|^2 + \|\operatorname{proj}_W \vec{b} - \vec{w}\|^2.$$

Since the second term of the right-hand side is non-negative, deleting it from the equation gives the inequality

$$\|\vec{b} - \vec{w}\|^2 \geq \|\vec{b} - \operatorname{proj}_W \vec{b}\|^2$$

implying that $\|\vec{b} - \operatorname{proj}_W \vec{b}\| \leq \|\vec{b} - \vec{w}\|$. Finally, we take $\hat{x} = \operatorname{proj}_W \vec{b}$.

The proof of the uniqueness of $\hat{x}$ is an easy exercise which is left to the students to verify. ▨

Remark. The quantity $d = \|\vec{b} - \operatorname{proj}_W \vec{b}\|$ is defined as the **distance** of $\vec{b}$ from W. By the projection theorem, we may also write

$$d = \|\operatorname{proj}_{W^\perp} \vec{b}\|.$$

Note that if $\vec{b} \in W$, then $\hat{x} = \vec{b}$.

Example 285. Find a formula for the distance in $\mathbb{R}^n$ of the point $(b_1, b_2, \ldots, b_n)$ and the hyperplane whose equation is $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$.

Solution. Let W denote the hyperplane $a_1x_1 + a_2x_2 + \cdots + a_nx_n = 0$ and let

$$\vec{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

We know that $W^\perp = \operatorname{span}\{\vec{a}\}$, where

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}.$$

Therefore, by the above remark, the distance d between $\vec{b}$ and W is

$$d = \| \operatorname{proj}_{W^\perp} \vec{x}\| = \| \operatorname{proj}_{\vec{a}} \vec{x}\| = \frac{|\vec{x} \cdot \vec{a}|}{\|a\|} = \frac{|a_1 b_1 + a_2 b_2 + \cdots + a_n b_n|}{\sqrt{a_1^2 + a_2^2 + \cdots + a_n^2}}.$$

Least-Squares Solution to Linear Systems

There may be instances in the applications where we are supposed to solve for a linear system $A\vec{x} = \vec{b}$, only to find out the this linear system is inconsistent. As an alternative, we instead find a vector $\hat{x}$ such that the distance $\|\vec{b} - A\hat{x}\|$ is minimum.

In this lesson, we shall be interested to find a vector $\hat{x}$ such that

$$\|\vec{b} - A\hat{x}\| \leq \|\vec{b} - A\vec{x}\|, \text{ for all } \vec{x} \in \mathbb{R}^n$$

Such a vector $\hat{x}$ is called a **least-squares solution** to the linear system $A\vec{x} = \vec{b}$. The norm $\|\vec{b} - A\hat{x}\|$ is called the **least-squares error**.

If we let

$$\vec{b} - A\hat{x} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix},$$

then the least-squares error is $\|\vec{b} - A\hat{x}\| = \sqrt{e_1^2 + e_2^2 + \cdots + e_n^2}$ and this expression is minimized if the sum of squares $e_1^2 + e_2^2 + \cdots + e_n^2$ is minimized, which explains the terminology.

How do we find a least-squares solution to $A\vec{x} = b$?

The key observation is the fact that

$$C(A) = \operatorname{Im} T_A = \{A\vec{x} \mid \vec{x} \in \mathbb{R}^n\}.$$

By Theorem 284, $\|\vec{b} - A\vec{x}\|$ is minimized, when

$$A\vec{x} = \operatorname{proj}_{C(A)} \vec{b} \tag{4.1}$$

The linear system (4.1) is consistent and its solutions are the least-squares solutions of $A\vec{x} = \vec{b}$. Hence, we are guaranteed that every linear system has a least-squares solution.

Note that we may rewrite linear system (4.1) as

$$\vec{b} - A\vec{x} = \vec{b} - \operatorname{proj}_{C(A)} \vec{b}.$$

Since $C(A)^{\perp} = N(A^t)$, the projection theorem implies that $\vec{b} - \operatorname{proj}_{C(A)} \vec{b} = \operatorname{proj}_{N(A^t)} \vec{b}$. It follows that

$$\vec{b} - A\vec{x} = \operatorname{proj}_{N(A^t)} \vec{b}.$$

Taking note that $\operatorname{proj}_{N(A^t)} \vec{b} \in N(A^t)$, if we multiply A^t from the left on both sides, we get

$$\begin{aligned} A^t(\vec{b} - A\vec{x}) &= A^t \operatorname{proj}_{N(A^t)} \vec{b} \\ \therefore\ A^t(\vec{b} - A\vec{x}) &= \vec{0}. \end{aligned}$$

Simplifying we get

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

This equation is called the **normal equation** and any of its solutions is a least-squares solution to $A\vec{x} = \vec{b}$. We have the following result.

Theorem 286: Least-Squares Solutions

The least-squares solutions to the linear system $A\vec{x} = \vec{b}$ are the solutions to the *normal equation*

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

Furthermore, if A has a full column rank, the normal equation has a unique solution

$$\hat{x} = (A^t A)^{-1} A^t\vec{b}.$$

Note that the second statement follows from Theorem 273, which implies that if A has a full column rank, then $A^t A$ is nonsingular.

We now illustrate this result through examples.

Example 287. Prove that the linear system

$$\begin{cases} x_1 + 2x_2 = 4 \\ 2x_1 - x_2 = 3 \\ x_1 + 3x_2 = 6 \end{cases}$$

is inconsistent. Find a least-squares solution and calculate the least-squares error.

Solution. We may write the linear system in matrix form $A\vec{x} = \vec{b}$, where

$$A = \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 3 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \qquad \text{and} \qquad \vec{b} = \begin{bmatrix} 4 \\ 3 \\ 6 \end{bmatrix}.$$

Applying the Gauss-Jordan method, we see that

$$[A \mid \vec{b}] = \left[\begin{array}{cc|c} 1 & 2 & 4 \\ 2 & -1 & 3 \\ 1 & 3 & 6 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{array}\right],$$

hence the linear system is inconsistent.

Notice that the columns of A are linearly independent, so A has a full column rank implying that A^tA is nonsingular. To find the least-squares solution, we compute

- $A^tA = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 3 \\ 3 & 14 \end{bmatrix}$;

- $(A^tA)^{-1} = \begin{bmatrix} 6 & 3 \\ 3 & 14 \end{bmatrix}^{-1} = \begin{bmatrix} \frac{14}{75} & -\frac{1}{25} \\ -\frac{1}{25} & \frac{2}{25} \end{bmatrix}$; and

- $A^t\vec{b} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 4 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 16 \\ 23 \end{bmatrix}$.

Therefore, the least-squares solution to the linear system is

$$\hat{x} = (A^tA)^{-1}A^t\vec{b} = \begin{bmatrix} \frac{14}{75} & -\frac{1}{25} \\ -\frac{1}{25} & \frac{2}{25} \end{bmatrix} \begin{bmatrix} 16 \\ 23 \end{bmatrix} = \begin{bmatrix} \frac{31}{15} \\ \frac{6}{5} \end{bmatrix}.$$

Finally, the least-square error is the norm of the vector

$$\vec{b} - A\hat{x} = \begin{bmatrix} 4 \\ 3 \\ 6 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 2 & -1 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} \frac{31}{15} \\ \frac{6}{5} \end{bmatrix} = \begin{bmatrix} -\frac{7}{15} \\ \frac{1}{15} \\ \frac{1}{3} \end{bmatrix},$$

which is

$$\|\vec{b} - A\hat{x}\| = \sqrt{\left(-\tfrac{7}{15}\right)^2 + \left(\tfrac{1}{15}\right)^2 + \left(\tfrac{1}{3}\right)^2} = \sqrt{\tfrac{1}{3}} \approx 0.5774$$

Theorem 286 guarantees that this is the smallest norm we can get.

Example 288. The linear system below is inconsistent.

$$\begin{cases} x_1 + 3x_2 = 4 \\ 2x_1 + 6x_2 = 5 \end{cases}$$

Find a least-squares solution and compute the least-squares error.

Solution. Define

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \text{and} \quad \vec{b} = \begin{bmatrix} 4 \\ 5 \end{bmatrix},$$

so that $A\vec{x} = \vec{b}$. In this case, A has rank 1, hence, A^tA is singular. To find a least-squares solution, we compute

- $A^tA = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 15 \\ 15 & 45 \end{bmatrix}.$
- $A^t\vec{b} = \begin{bmatrix} 1 & 2 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 14 \\ 42 \end{bmatrix}.$

Next, we solve the normal equation $A^tA\hat{x} = A^t\vec{b}$:

$$\left[\begin{array}{cc|c} 5 & 15 & 14 \\ 15 & 45 & 42 \end{array}\right] \xrightarrow{\text{RREF}} \left[\begin{array}{cc|c} 1 & 3 & \frac{14}{5} \\ 0 & 0 & 0 \end{array}\right].$$

The normal equation has infinitely many solutions of the form

$$\hat{x} = \begin{bmatrix} \frac{14}{5} - 3r \\ r \end{bmatrix} = \begin{bmatrix} \frac{14}{5} \\ 0 \end{bmatrix} + r \begin{bmatrix} -3 \\ 1 \end{bmatrix}, \text{ where } r \in \mathbb{R}.$$

This means that the original linear system has infinitely many least-squares solutions.

Note that the least-squares error vector is

$$\begin{aligned}\vec{b} - A\hat{x} &= \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} \frac{14}{5} - 3r \\ r \end{bmatrix} \\ &= \begin{bmatrix} 4 \\ 5 \end{bmatrix} - \begin{bmatrix} \frac{14}{5} \\ \frac{28}{5} \end{bmatrix} \\ &= \begin{bmatrix} \frac{6}{5} \\ -\frac{3}{5} \end{bmatrix},\end{aligned}$$

which is independent of the parameter r. This only means that although there are infinitely many least-squares solution, each one of them produces the same least-squares error which is

$$\|\vec{b} - A\hat{x}\| = \sqrt{\left(\tfrac{6}{5}\right)^2 + \left(-\tfrac{3}{5}\right)^2} = \frac{3\sqrt{5}}{5} \approx 1.3416.$$

Applications to Curve Fitting

Many problems in mathematical modeling involve finding a mathematical relationship between two variables x and y by "fitting" a curve $y = f(x)$ to a set of experimental data points

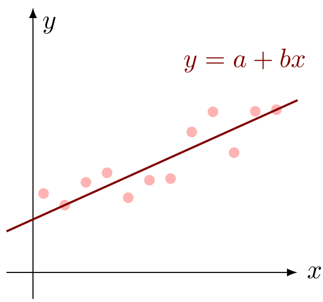
$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

To simplify the problem, we make certain assumptions about the function f and call the relationship $y = f(x)$ a **mathematical model** for the data.

The simplest type of curve-fitting problem is when f is assumed to be a *linear function* of the form

$$f(x) = a + bx,$$

where $a, b \in \mathbb{R}$. In this case, the line $y = a + bx$ is sometimes called the **best-fit line** or the **regression line**.



Suppose that the linear function $y = a + bx$ models the relationship between x and y and that the parameters a and b are to be determined by the data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

If we plug in these points to the equation $y = a + bx$, the unknown values of a and b may be obtained by solving the system of equations

$$\begin{aligned} a + bx_1 &= y_1 \\ a + bx_2 &= y_2 \\ &\vdots \\ a + bx_n &= y_n \end{aligned}$$

This is equivalent to the matrix equation $A\vec{p} = \vec{y}$, where

$$A = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \vec{p} = \begin{bmatrix} a \\ b \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

In most cases, the data points have errors since they are measured through experiments. Thus, the linear system $A\vec{p} = \vec{y}$ is highly likely to be inconsistent. In this case, we look for the least-squares solution to the linear system by solving the normal equation

$$A^t A\vec{p} = A^t\vec{y}.$$

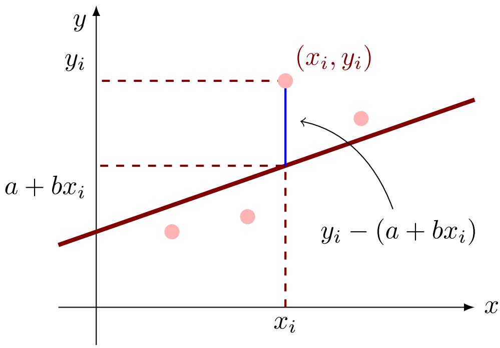
If the x-coordinates of the data points are not all the same, then A has full rank, so the unique least-squares solution is

$$\vec{p} = (A^t A)^{-1} A^t \vec{y}.$$

This least-squares solution minimizes the quantity

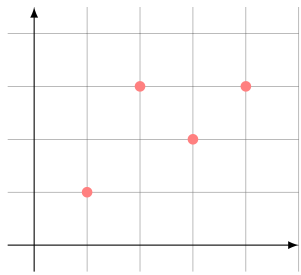
$$\|\vec{y} - A\vec{p}\|^2 = [y_1 - (a + bx_1)]^2 + [y_2 - (a + bx_2)]^2 + \cdots + [y_n - (a + bx_n)]^2,$$

which is the sum of squares of *residuals*. The calculated regression line $y = a + bx$ is also called the **line of best fit**.



Example 289. Find the least-squares line that best fits the points $(1,1), (2,3), (3,2)$, and $(4,3)$. Also, calculate the sum of the squares of the residuals.

Solution.



We set up the linear system $A\vec{p} = \vec{y}$, where

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}, \vec{p} = \begin{bmatrix} a \\ b \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} 1 \\ 3 \\ 2 \\ 3 \end{bmatrix}.$$

We compute the following:

- $A^tA = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 10 \\ 10 & 30 \end{bmatrix};$

- $(A^tA)^{-1} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{bmatrix};$

- $A^t\vec{y} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 25 \end{bmatrix}.$

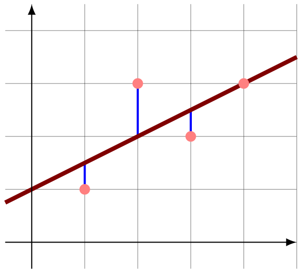
Therefore, the least-squares solution is

$$\begin{bmatrix} a \\ b \end{bmatrix} = \vec{p} = (A^t A)^{-1} A^t \vec{y} = \begin{bmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{5} \end{bmatrix} \begin{bmatrix} 9 \\ 25 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}.$$

Therefore, the line of best fit has equation

$$y = 1 + \tfrac{1}{2}x.$$

The graph of the line of best fit is shown below.



Now,

$$\vec{y} - A\vec{p} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ 1 \\ -\frac{1}{2} \\ 0 \end{bmatrix}.$$

Therefore, the sum of squares of the residuals is

$$\|\vec{y} - A\vec{p}\|^2 = \left(-\tfrac{1}{2}\right)^2 + 1^2 + \left(-\tfrac{1}{2}\right)^2 + 0^2 = \tfrac{3}{2} = 1.5.$$

Remark.

For a general linear regression problem on the set of data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n),$$

the normal equation $A^t A\vec{p} = A^t \vec{y}$ reduces to

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}.$$

From which it follows that

$$a = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2} \text{ and } b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}.$$

It is quite handy to memorize the formulas for a and b, but remember that we only need to remember the normal equation $A^t A\vec{p} = A^t \vec{y}$.

We next look at the case when the relationship of the variables x and y is a polynomial function of degree $n \geq 1$. As we shall see, the technique applied in the case when $m = 1$ naturally extends to a more general case.

Suppose that there is an evidence to believe that the variables x and y may be modeled by

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m,$$

where $a_i \in \mathbb{R}$, $i = 0, 1, \ldots, m$, and $a_m \neq 0$ and that this is the curve of that best fits the data points

$$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n).$$

If we want the curve to interpolate the points, then

$$\begin{aligned} y_1 &= a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_m x_1^m \\ y_2 &= a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_m x_2^m \\ &\vdots \\ y_n &= a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_m x_n^m \end{aligned}$$

which is equivalent to the linear system $A\vec{p} = \vec{y}$, where

$$A = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{bmatrix}, \vec{p} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{bmatrix}, \quad \text{and} \quad \vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}.$$

Observe that

- if $m = n - 1$ and the x_i's are distinct, the problem is equivalent to the polynomial interpolation problem studied in Lesson 2.1, hence, the problem has a unique solution.
- if $m < n-1$, due to errors in measurement, the linear system is usually inconsistent. Thus, we consider looking for the least-squares solution.
- if $m < n$ and at least $m + 1$ of the x-coordinates are distinct, then A has a full column rank. Thus, the normal equation

$$A^t A\vec{p} = A^t\vec{y}$$

has a unique solution

$$\vec{p} = (A^t A)^{-1} A^t\vec{y}.$$

Example 290. Find the best quadratic fit to the data:

x	-1	0	0	1	2
y	1	-1	1	3	4

Solution. We find the best quadratic fit $y = a_0 + a_1 x + a_2 x^2$, by solving $A\vec{p} = \vec{y}$, where

$$A = \begin{bmatrix} 1 & -1 & (-1)^2 \\ 1 & 0 & 0^2 \\ 1 & 0 & 0^2 \\ 1 & 1 & 1^2 \\ 1 & 2 & 2^2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{bmatrix}, \vec{p} = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}, \text{ and } \vec{y} = \begin{bmatrix} 1 \\ -1 \\ 1 \\ 3 \\ 4 \end{bmatrix}.$$

Here, $m = 2$ while $n = 5$, and 4 x-coordinates are distinct. Thus, there is a unique solution to the normal equation $A^t A\vec{p} = A^t\vec{y}$, which is

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \vec{p} = (A^t A)^{-1} A^t\vec{y} = \begin{bmatrix} 0.5806 \\ 0.6129 \\ 0.6452 \end{bmatrix}.$$

Note that the values are rounded off up to 4 decimal places. The quadratic function that fits the data is

$$y = 0.5806 + 0.6129x + 0.6452x^2.$$