

6.4.3 Two Methods to Compute the Projection

We have now presented *two* methods for computing the orthogonal projection of a vector: this [theorem in Section 6.3](#) involves row reduction, and the [projection formula](#) requires an orthogonal basis. Here are some guidelines for which to use in a given situation.

1. If you already have an orthogonal basis, it is almost always easier to use the [projection formula](#). This often happens in the sciences.
2. If you are going to have to compute the projections of many vectors onto the same subspace, it is worth your time to run Gram–Schmidt to produce an orthogonal basis, so that you can use the [projection formula](#).
3. If you only have to project one or a few vectors onto a subspace, it is faster to use the [theorem in Section 6.3](#). This is the method we will follow in [Section 6.5](#).

6.5 The Method of Least Squares

Objectives

1. Learn examples of best-fit problems.
2. Learn to turn a best-fit problem into a least-squares problem.
3. *Recipe*: find a least-squares solution (two ways).
4. *Picture*: geometry of a least-squares solution.
5. *Vocabulary words*: **least-squares solution**.

In this section, we answer the following important question:

Suppose that $Ax = b$ does not have a solution. What is the best approximate solution?

For our purposes, the best approximate solution is called the *least-squares solution*. We will present two methods for finding least-squares solutions, and we will give several applications to best-fit problems.

6.5.1 Least-Squares Solutions

We begin by clarifying exactly what we will mean by a “best approximate solution” to an inconsistent matrix equation $Ax = b$.

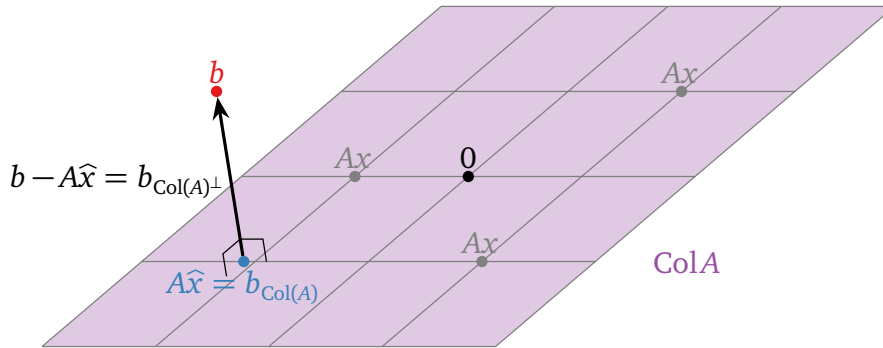
Definition. Let A be an $m \times n$ matrix and let b be a vector in \mathbf{R}^m . A **least-squares solution** of the matrix equation $Ax = b$ is a vector \hat{x} in \mathbf{R}^n such that

$$\text{dist}(b, A\hat{x}) \leq \text{dist}(b, Ax)$$

for all other vectors x in \mathbf{R}^n .

Recall that $\text{dist}(v, w) = \|v - w\|$ is the [distance](#) between the vectors v and w . The term “least squares” comes from the fact that $\text{dist}(b, Ax) = \|b - Ax\|$ is the square root of the sum of the squares of the entries of the vector $b - Ax$. So a least-squares solution minimizes the sum of the squares of the differences between the entries of $A\hat{x}$ and b . In other words, a least-squares solution solves the equation $Ax = b$ as closely as possible, in the sense that the sum of the squares of the difference $b - Ax$ is minimized.

Least Squares: Picture Suppose that the equation $Ax = b$ is inconsistent. Recall from this [note in Section 2.3](#) that the column space of A is the set of all other vectors c such that $Ax = c$ is consistent. In other words, $\text{Col}(A)$ is the set of all vectors of the form Ax . Hence, the [closest vector](#) of the form Ax to b is the orthogonal projection of b onto $\text{Col}(A)$. This is denoted $b_{\text{Col}(A)}$, following this [notation in Section 6.3](#).



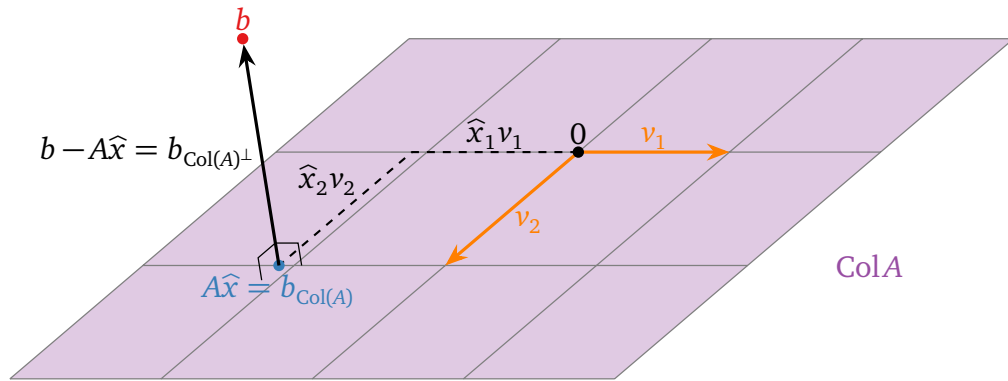
A least-squares solution of $Ax = b$ is a solution \hat{x} of the consistent equation $Ax = b_{\text{Col}(A)}$

Note. If $Ax = b$ is consistent, then $b_{\text{Col}(A)} = b$, so that a least-squares solution is the same as a usual solution.

Where is \hat{x} in this picture? If v_1, v_2, \dots, v_n are the columns of A , then

$$A\hat{x} = A \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{pmatrix} = \hat{x}_1 v_1 + \hat{x}_2 v_2 + \cdots + \hat{x}_n v_n.$$

Hence the entries of \hat{x} are the “coordinates” of $b_{\text{Col}(A)}$ with respect to the spanning set $\{v_1, v_2, \dots, v_m\}$ of $\text{Col}(A)$. (They are honest \mathcal{B} -coordinates if the columns of A are linearly independent.)



[Use this link to view the online demo](#)

The violet plane is $\text{Col}(A)$. The closest that Ax can get to b is the closest vector on $\text{Col}(A)$ to b , which is the orthogonal projection $b_{\text{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A , and the coefficients of \hat{x} are the lengths of the green lines. Click and drag b to move it.

We learned to solve this kind of orthogonal projection problem in [Section 6.3](#).

Theorem. Let A be an $m \times n$ matrix and let b be a vector in \mathbf{R}^m . The least-squares solutions of $Ax = b$ are the solutions of the matrix equation

$$A^T A x = A^T b$$

Proof. By this [theorem in Section 6.3](#), if \hat{x} is a solution of the matrix equation $A^T A x = A^T b$, then $A\hat{x}$ is equal to $b_{\text{Col}(A)}$. We argued above that a least-squares solution of $Ax = b$ is a solution of $Ax = b_{\text{Col}(A)}$. \square

In particular, finding a least-squares solution means solving a consistent system of linear equations. We can translate the above theorem into a recipe:

Recipe 1: Compute a least-squares solution. Let A be an $m \times n$ matrix and let b be a vector in \mathbf{R}^n . Here is a method for computing a least-squares solution of $Ax = b$:

1. Compute the matrix $A^T A$ and the vector $A^T b$.
2. Form the augmented matrix for the matrix equation $A^T A x = A^T b$, and row reduce.
3. This equation is always consistent, and any solution \hat{x} is a least-squares solution.

To reiterate: once you have found a least-squares solution \hat{x} of $Ax = b$, then $b_{\text{Col}(A)}$ is equal to $A\hat{x}$.

Example. Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \quad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

What quantity is being minimized?

Solution. We have

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

and

$$A^T b = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 5 & 3 & 0 \\ 3 & 3 & 6 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & -3 \\ 0 & 1 & 5 \end{array} \right).$$

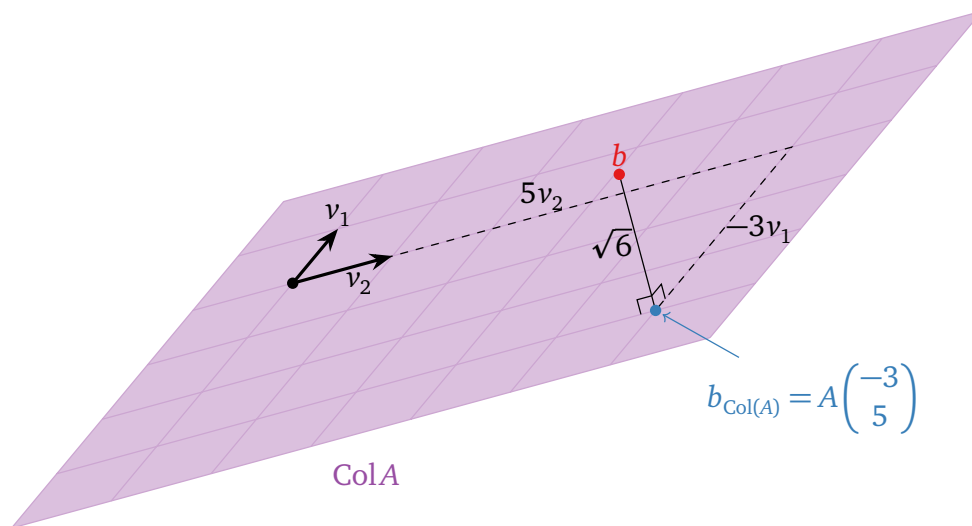
Therefore, the only least-squares solution is $\hat{x} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$.

This solution minimizes the distance from $A\hat{x}$ to b , i.e., the sum of the squares of the entries of $b - A\hat{x} = b - b_{\text{Col}(A)} = b_{\text{Col}(A)^\perp}$. In this case, we have

$$\|b - A\hat{x}\| = \left\| \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 5 \\ 2 \\ -1 \end{pmatrix} \right\| = \left\| \begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix} \right\| = \sqrt{1^2 + (-2)^2 + 1^2} = \sqrt{6}.$$

Therefore, $b_{\text{Col}(A)} = A\hat{x}$ is $\sqrt{6}$ units from b .

In the following picture, v_1, v_2 are the columns of A :



[Use this link to view the online demo](#)

The violet plane is $\text{Col}(A)$. The closest that Ax can get to b is the closest vector on $\text{Col}(A)$ to b , which is the orthogonal projection $b_{\text{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A , and the coefficients of \hat{x} are the \mathcal{B} -coordinates of $b_{\text{Col}(A)}$, where $\mathcal{B} = \{v_1, v_2\}$.

Example. Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 2 & 0 \\ -1 & 1 \\ 0 & 2 \end{pmatrix} \quad b = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Solution. We have

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

and

$$A^T b = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} = \begin{pmatrix} 2 \\ -2 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{cc|c} 5 & -1 & 2 \\ -1 & 5 & -2 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{cc|c} 1 & 0 & 1/3 \\ 0 & 1 & -1/3 \end{array} \right).$$

Therefore, the only least-squares solution is $\hat{x} = \frac{1}{3} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$.

[Use this link to view the online demo](#)

The red plane is $\text{Col}(A)$. The closest that Ax can get to b is the closest vector on $\text{Col}(A)$ to b , which is the orthogonal projection $b_{\text{Col}(A)}$ (in blue). The vectors v_1, v_2 are the columns of A , and the coefficients of \hat{x} are the \mathcal{B} -coordinates of $b_{\text{Col}(A)}$, where $\mathcal{B} = \{v_1, v_2\}$.

The reader may have noticed that we have been careful to say “the least-squares solutions” in the plural, and “a least-squares solution” using the indefinite article. This is because a least-squares solution need not be unique: indeed, if the columns of A are linearly dependent, then $Ax = b_{\text{Col}(A)}$ has infinitely many solutions. The following theorem, which gives equivalent criteria for uniqueness, is an analogue of this [corollary in Section 6.3](#).

Theorem. Let A be an $m \times n$ matrix and let b be a vector in \mathbf{R}^m . The following are equivalent:

1. $Ax = b$ has a unique least-squares solution.
2. The columns of A are linearly independent.
3. $A^T A$ is invertible.

In this case, the least-squares solution is

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

Proof. The set of least-squares solutions of $Ax = b$ is the solution set of the consistent equation $A^T A x = A^T b$, which is a translate of the solution set of the homogeneous equation $A^T A x = 0$. Since $A^T A$ is a square matrix, the equivalence of 1 and 3 follows from the [invertible matrix theorem in Section 5.1](#). The set of least squares-solutions is also the solution set of the consistent equation $Ax = b_{\text{Col}(A)}$, which has a unique solution if and only if the columns of A are linearly independent by this [important note in Section 2.5](#). \square

Example (Infinitely many least-squares solutions). Find the least-squares solutions of $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & -1 \\ 1 & 2 & -3 \end{pmatrix} \quad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

Solution. We have

$$A^T A = \begin{pmatrix} 3 & 3 & -3 \\ 3 & 5 & -7 \\ -3 & -7 & 11 \end{pmatrix} \quad A^T b = \begin{pmatrix} 6 \\ 0 \\ 6 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccc|c} 3 & 3 & -3 & 6 \\ 3 & 5 & -7 & 0 \\ -3 & -7 & 11 & 6 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 1 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array} \right).$$

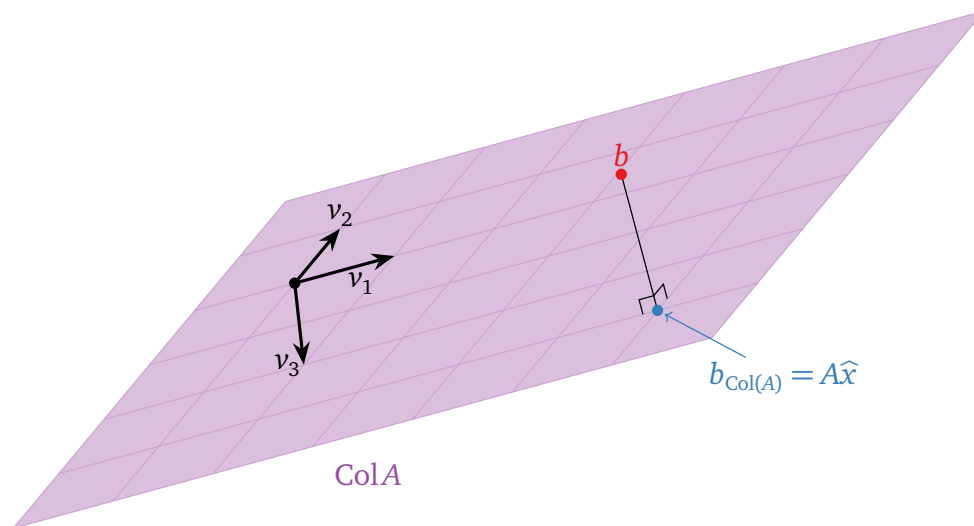
The free variable is x_3 , so the solution set is

$$\begin{cases} x_1 = -x_3 + 5 \\ x_2 = 2x_3 - 3 \\ x_3 = x_3 \end{cases} \xrightarrow[\text{vector form}]{\text{parametric}} \hat{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix} + \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix}.$$

For example, taking $x_3 = 0$ and $x_3 = 1$ gives the least-squares solutions

$$\hat{x} = \begin{pmatrix} 5 \\ -3 \\ 0 \end{pmatrix} \quad \text{and} \quad \hat{x} = \begin{pmatrix} 4 \\ -1 \\ 1 \end{pmatrix}.$$

Geometrically, we see that the columns v_1, v_2, v_3 of A are coplanar:



Therefore, there are many ways of writing $b_{\text{Col}(A)}$ as a linear combination of v_1, v_2, v_3 .

[Use this link to view the online demo](#)

The three columns of A are coplanar, so there are many least-squares solutions. (The demo picks one solution when you move b .)

As usual, calculations involving projections become easier in the presence of an orthogonal set. Indeed, if A is an $m \times n$ matrix with *orthogonal* columns u_1, u_2, \dots, u_m , then we can use the [projection formula in Section 6.4](#) to write

$$b_{\text{Col}(A)} = \frac{b \cdot u_1}{u_1 \cdot u_1} u_1 + \frac{b \cdot u_2}{u_2 \cdot u_2} u_2 + \cdots + \frac{b \cdot u_m}{u_m \cdot u_m} u_m = A \begin{pmatrix} (b \cdot u_1)/(u_1 \cdot u_1) \\ (b \cdot u_2)/(u_2 \cdot u_2) \\ \vdots \\ (b \cdot u_m)/(u_m \cdot u_m) \end{pmatrix}.$$

Note that the least-squares solution is unique in this case, since [an orthogonal set is linearly independent](#).

Recipe 2: Compute a least-squares solution. Let A be an $m \times n$ matrix with *orthogonal* columns u_1, u_2, \dots, u_m , and let b be a vector in \mathbf{R}^n . Then the least-squares solution of $Ax = b$ is the vector

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \dots, \frac{b \cdot u_m}{u_m \cdot u_m} \right).$$

This formula is particularly useful in the sciences, as matrices with orthogonal columns often arise in nature.

Example. Find the least-squares solution of $Ax = b$ where:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

Solution. Let u_1, u_2, u_3 be the columns of A . These form an orthogonal set, so

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \frac{b \cdot u_3}{u_3 \cdot u_3} \right) = \left(\frac{-3}{2}, \frac{-3}{2}, \frac{8}{4} \right) = \left(-\frac{3}{2}, -\frac{3}{2}, 2 \right).$$

Compare this [example in Section 6.4](#).

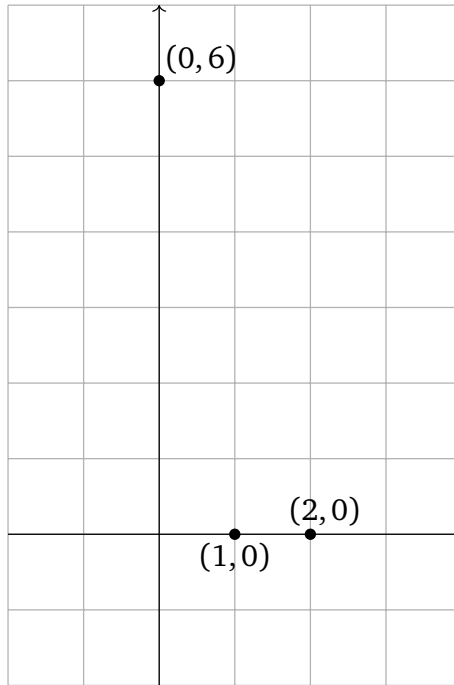
6.5.2 Best-Fit Problems

In this subsection we give an application of the method of least squares to data modeling. We begin with a basic example.

Example (Best-fit line). Suppose that we have measured three data points

$$(0, 6), \quad (1, 0), \quad (2, 0),$$

and that our model for these data asserts that the points should lie on a line. Of course, these three points do not actually lie on a single line, but this could be due to errors in our measurement. How do we predict which line they are supposed to lie on?



The general equation for a (non-vertical) line is

$$y = Mx + B.$$

If our three data points were to lie on this line, then the following equations would be satisfied:

$$\begin{aligned} 6 &= M \cdot 0 + B \\ 0 &= M \cdot 1 + B \\ 0 &= M \cdot 2 + B. \end{aligned} \tag{6.5.1}$$

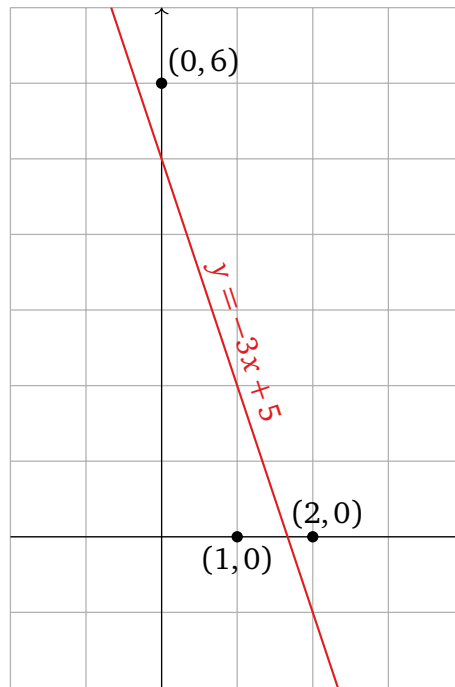
In order to find the best-fit line, we try to solve the above equations in the unknowns M and B . As the three points do not actually lie on a line, there is no actual solution, so instead we compute a least-squares solution.

Putting our linear equations into matrix form, we are trying to solve $Ax = b$ for

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \\ 2 & 1 \end{pmatrix} \quad x = \begin{pmatrix} M \\ B \end{pmatrix} \quad b = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix}.$$

We solved this least-squares problem in this [example](#): the only least-squares solution to $Ax = b$ is $\hat{x} = \begin{pmatrix} M \\ B \end{pmatrix} = \begin{pmatrix} -3 \\ 5 \end{pmatrix}$, so the best-fit line is

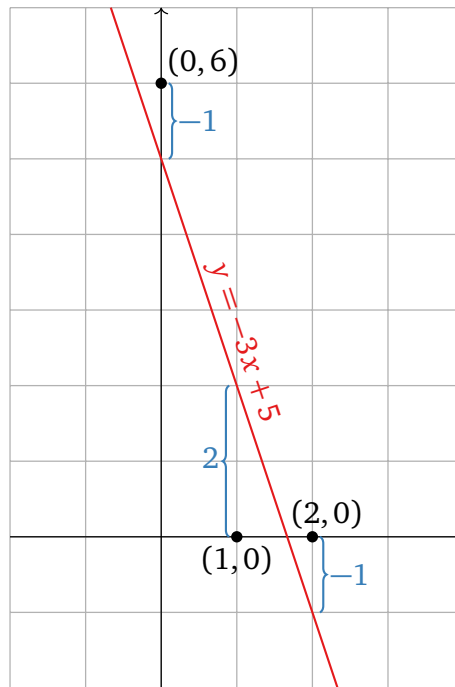
$$y = -3x + 5.$$



What exactly is the line $y = f(x) = -3x + 5$ minimizing? The least-squares solution \hat{x} minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the left-hand side of (6.5.1), and

$$A \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -3(0) + 5 \\ -3(1) + 5 \\ -3(2) + 5 \end{pmatrix} = \begin{pmatrix} f(0) \\ f(1) \\ f(2) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the y -coordinates of the graph of the line at the values of x we specified in our data points, and b is the vector whose entries are the y -coordinates of those data points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points:



$$b - A\hat{x} = \begin{pmatrix} 6 \\ 0 \\ 0 \end{pmatrix} - A \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix}$$

The best-fit line minimizes the sum of the squares of these vertical distances.

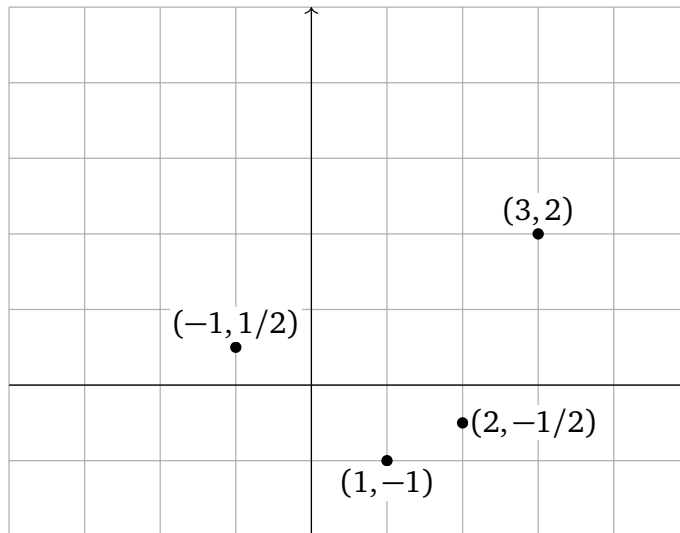
Interactive: Best-fit line.

[Use this link to view the online demo](#)

The best-fit line minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit line changes.

Example (Best-fit parabola). Find the parabola that best approximates the data points

$$(-1, 1/2), \quad (1, -1), \quad (2, -1/2), \quad (3, 2).$$



What quantity is being minimized?

Solution. The general equation for a parabola is

$$y = Bx^2 + Cx + D.$$

If the four points were to lie on this parabola, then the following equations would be satisfied:

$$\begin{aligned} \frac{1}{2} &= B(-1)^2 + C(-1) + D \\ -1 &= B(1)^2 + C(1) + D \\ -\frac{1}{2} &= B(2)^2 + C(2) + D \\ 2 &= B(3)^2 + C(3) + D. \end{aligned} \tag{6.5.2}$$

We treat this as a system of equations in the unknowns B, C, D . In matrix form, we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \end{pmatrix} \quad b = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix}.$$

We find a least-squares solution by multiplying both sides by the transpose:

$$A^T A = \begin{pmatrix} 99 & 35 & 15 \\ 35 & 15 & 5 \\ 15 & 5 & 4 \end{pmatrix} \quad A^T b = \begin{pmatrix} 31/2 \\ 7/2 \\ 1 \end{pmatrix},$$

then forming an augmented matrix and row reducing:

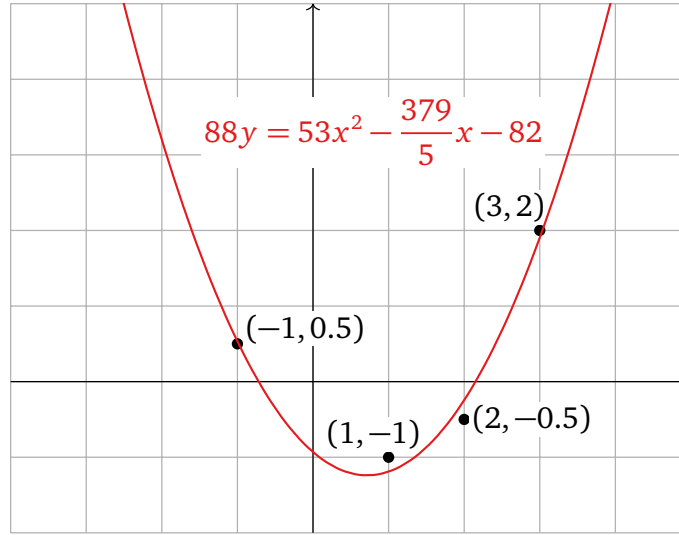
$$\left(\begin{array}{ccc|c} 99 & 35 & 15 & 31/2 \\ 35 & 15 & 5 & 7/2 \\ 15 & 5 & 4 & 1 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 53/88 \\ 0 & 1 & 0 & -379/440 \\ 0 & 0 & 1 & -41/44 \end{array} \right) \Rightarrow \hat{x} = \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix}.$$

The best-fit parabola is

$$y = \frac{53}{88}x^2 - \frac{379}{440}x - \frac{41}{44}.$$

Multiplying through by 88, we can write this as

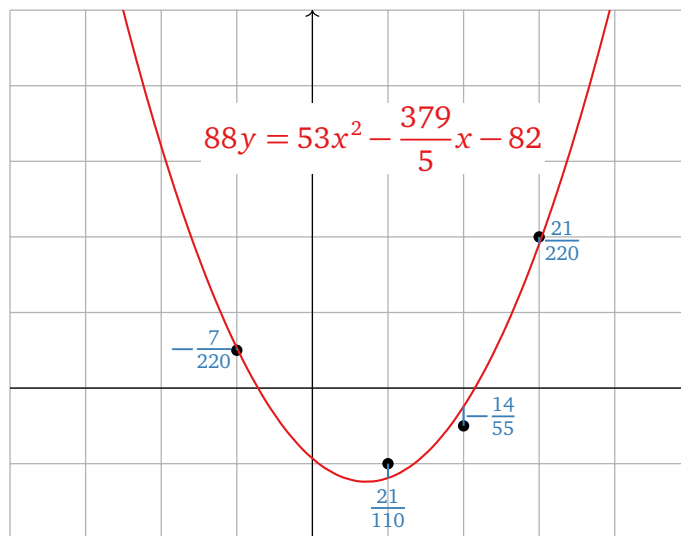
$$88y = 53x^2 - \frac{379}{5}x - 82.$$



Now we consider what exactly the parabola $y = f(x)$ is minimizing. The least-squares solution \hat{x} minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the left-hand side of (6.5.2), and

$$A\hat{x} = \begin{pmatrix} \frac{53}{88}(-1)^2 - \frac{379}{440}(-1) - \frac{41}{44} \\ \frac{53}{88}(1)^2 - \frac{379}{440}(1) - \frac{41}{44} \\ \frac{53}{88}(2)^2 - \frac{379}{440}(2) - \frac{41}{44} \\ \frac{53}{88}(3)^2 - \frac{379}{440}(3) - \frac{41}{44} \end{pmatrix} = \begin{pmatrix} f(-1) \\ f(1) \\ f(2) \\ f(3) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the y -coordinates of the graph of the parabola at the values of x we specified in our data points, and b is the vector whose entries are the y -coordinates of those data points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points:



$$b - A\hat{x} = \begin{pmatrix} 1/2 \\ -1 \\ -1/2 \\ 2 \end{pmatrix} - A \begin{pmatrix} 53/88 \\ -379/440 \\ -41/44 \end{pmatrix} = \begin{pmatrix} -7/220 \\ 21/110 \\ -14/55 \\ 21/220 \end{pmatrix}$$

The best-fit parabola minimizes the sum of the squares of these vertical distances.

[Use this link to view the online demo](#)

The best-fit parabola minimizes the sum of the squares of the vertical distances (vertical). Click and drag the points to see how the best-fit parabola changes.

Example (Best-fit linear function). Find the linear function $f(x, y)$ that best approximates the following data:

x	y	$f(x, y)$
1	0	0
0	1	1
-1	0	3
0	-1	4

What quantity is being minimized?

Solution. The general equation for a linear function in two variables is

$$f(x, y) = Bx + Cy + D.$$

We want to solve the following system of equations in the unknowns B, C, D :

$$\begin{aligned} B(1) + C(0) + D &= 0 \\ B(0) + C(1) + D &= 1 \\ B(-1) + C(0) + D &= 3 \\ B(0) + C(-1) + D &= 4. \end{aligned} \tag{6.5.3}$$

In matrix form, we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & -1 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ 1 \\ 3 \\ 4 \end{pmatrix}.$$

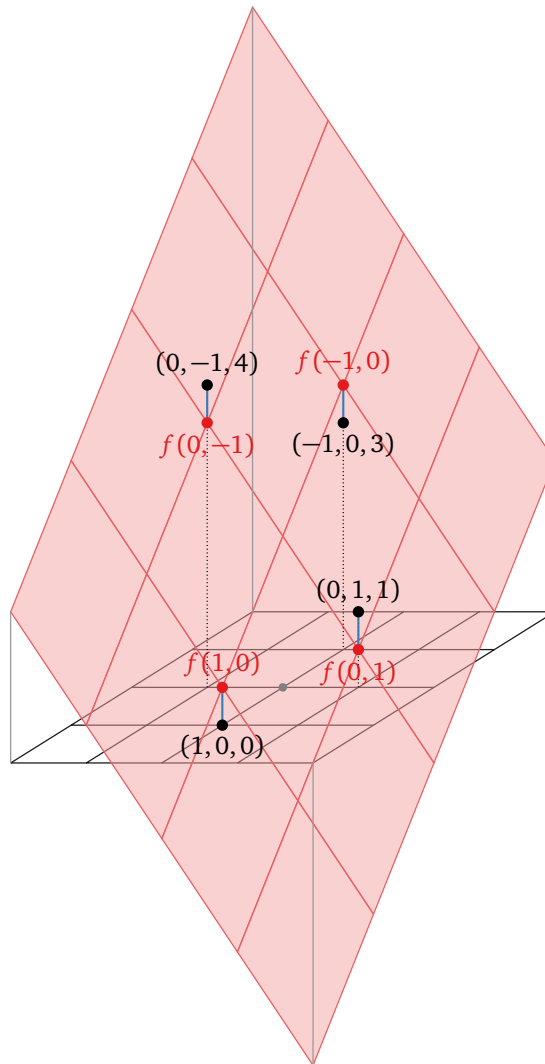
We observe that the columns u_1, u_2, u_3 of A are *orthogonal*, so we can use [recipe 2](#):

$$\hat{x} = \left(\frac{b \cdot u_1}{u_1 \cdot u_1}, \frac{b \cdot u_2}{u_2 \cdot u_2}, \frac{b \cdot u_3}{u_3 \cdot u_3} \right) = \left(\frac{-3}{2}, \frac{-3}{2}, \frac{8}{4} \right) = \left(-\frac{3}{2}, -\frac{3}{2}, 2 \right).$$

Therefore, the best-fit linear equation is

$$f(x, y) = -\frac{3}{2}x - \frac{3}{2}y + 2.$$

Here is a picture of the graph of $f(x, y)$:



Now we consider what quantity is being minimized by the function $f(x, y)$. The least-squares solution \hat{x} minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$. The vector b is the right-hand side of (6.5.3), and

$$A\hat{x} = \begin{pmatrix} -\frac{3}{2}(1) - \frac{3}{2}(0) + 2 \\ -\frac{3}{2}(0) - \frac{3}{2}(1) + 2 \\ -\frac{3}{2}(-1) - \frac{3}{2}(0) + 2 \\ -\frac{3}{2}(0) - \frac{3}{2}(-1) + 2 \end{pmatrix} = \begin{pmatrix} f(1, 0) \\ f(0, 1) \\ f(-1, 0) \\ f(0, -1) \end{pmatrix}.$$

In other words, $A\hat{x}$ is the vector whose entries are the values of f evaluated on the points (x, y) we specified in our data table, and b is the vector whose entries are the desired values of f evaluated at those points. The difference $b - A\hat{x}$ is the vertical distance of the graph from the data points, as indicated in the above picture. The best-fit linear function minimizes the sum of these vertical distances.

[Use this link to view the online demo](#)

The best-fit linear function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit linear function changes.

All of the above examples have the following form: some number of data points (x, y) are specified, and we want to find a function

$$y = B_1g_1(x) + B_2g_2(x) + \cdots + B_mg_m(x)$$

that best approximates these points, where g_1, g_2, \dots, g_m are fixed functions of x . Indeed, in the best-fit line example we had $g_1(x) = x$ and $g_2(x) = 1$; in the best-fit parabola example we had $g_1(x) = x^2$, $g_2(x) = x$, and $g_3(x) = 1$; and in the best-fit linear function example we had $g_1(x_1, x_2) = x_1$, $g_2(x_1, x_2) = x_2$, and $g_3(x_1, x_2) = 1$ (in this example we take x to be a vector with two entries). We evaluate the above equation on the given data points to obtain a system of linear equations in the unknowns B_1, B_2, \dots, B_m —once we evaluate the g_i , they just become numbers, so it does not matter what they are—and we find the least-squares solution. The resulting best-fit function minimizes the sum of the squares of the vertical distances from the graph of $y = f(x)$ to our original data points.

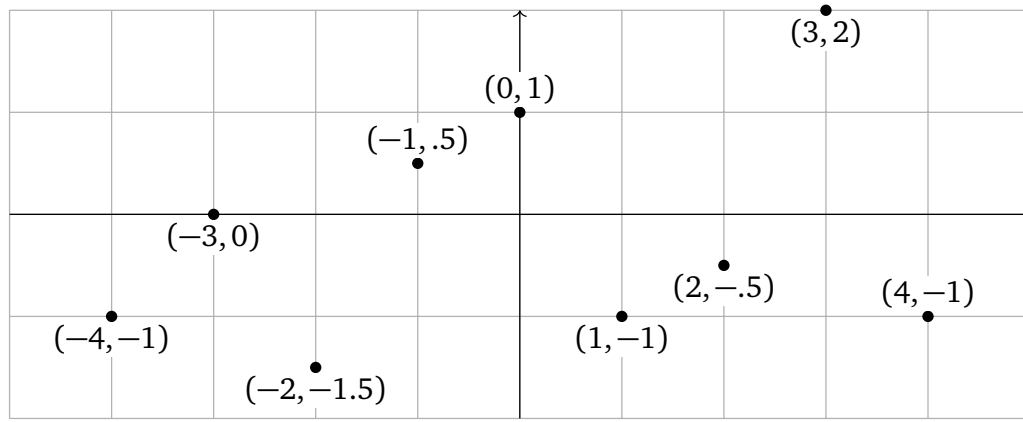
To emphasize that the nature of the functions g_i really is irrelevant, consider the following example.

Example (Best-fit trigonometric function). What is the best-fit function of the form

$$y = B + C \cos(x) + D \sin(x) + E \cos(2x) + F \sin(2x) + G \cos(3x) + H \sin(3x)$$

passing through the points

$$\begin{pmatrix} -4 \\ -1 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \end{pmatrix}, \begin{pmatrix} -2 \\ -1.5 \end{pmatrix}, \begin{pmatrix} -1 \\ .5 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 2 \\ -1.5 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ -1 \end{pmatrix}?$$



Solution. We want to solve the system of equations

$$\begin{aligned}
 -1 &= B + C \cos(-4) + D \sin(-4) + E \cos(-8) + F \sin(-8) + G \cos(-12) + H \sin(-12) \\
 0 &= B + C \cos(-3) + D \sin(-3) + E \cos(-6) + F \sin(-6) + G \cos(-9) + H \sin(-9) \\
 -1.5 &= B + C \cos(-2) + D \sin(-2) + E \cos(-4) + F \sin(-4) + G \cos(-6) + H \sin(-6) \\
 0.5 &= B + C \cos(-1) + D \sin(-1) + E \cos(-2) + F \sin(-2) + G \cos(-3) + H \sin(-3) \\
 1 &= B + C \cos(0) + D \sin(0) + E \cos(0) + F \sin(0) + G \cos(0) + H \sin(0) \\
 -1 &= B + C \cos(1) + D \sin(1) + E \cos(2) + F \sin(2) + G \cos(3) + H \sin(3) \\
 -0.5 &= B + C \cos(2) + D \sin(2) + E \cos(4) + F \sin(4) + G \cos(6) + H \sin(6) \\
 2 &= B + C \cos(3) + D \sin(3) + E \cos(6) + F \sin(6) + G \cos(9) + H \sin(9) \\
 -1 &= B + C \cos(4) + D \sin(4) + E \cos(8) + F \sin(8) + G \cos(12) + H \sin(12).
 \end{aligned}$$

All of the terms in these equations are *numbers*, except for the unknowns B, C, D, E, F, G, H :

$$\begin{aligned}
 -1 &= B - 0.6536C + 0.7568D - 0.1455E - 0.9894F + 0.8439G + 0.5366H \\
 0 &= B - 0.9900C - 0.1411D + 0.9602E + 0.2794F - 0.9111G - 0.4121H \\
 -1.5 &= B - 0.4161C - 0.9093D - 0.6536E + 0.7568F + 0.9602G + 0.2794H \\
 0.5 &= B + 0.5403C - 0.8415D - 0.4161E - 0.9093F - 0.9900G - 0.1411H \\
 1 &= B + C + D + E + F + G + H \\
 -1 &= B + 0.5403C + 0.8415D - 0.4161E + 0.9093F - 0.9900G + 0.1411H \\
 -0.5 &= B - 0.4161C + 0.9093D - 0.6536E - 0.7568F + 0.9602G - 0.2794H \\
 2 &= B - 0.9900C + 0.1411D + 0.9602E - 0.2794F - 0.9111G + 0.4121H \\
 -1 &= B - 0.6536C - 0.7568D - 0.1455E + 0.9894F + 0.8439G - 0.5366H.
 \end{aligned}$$

Hence we want to solve the least-squares problem

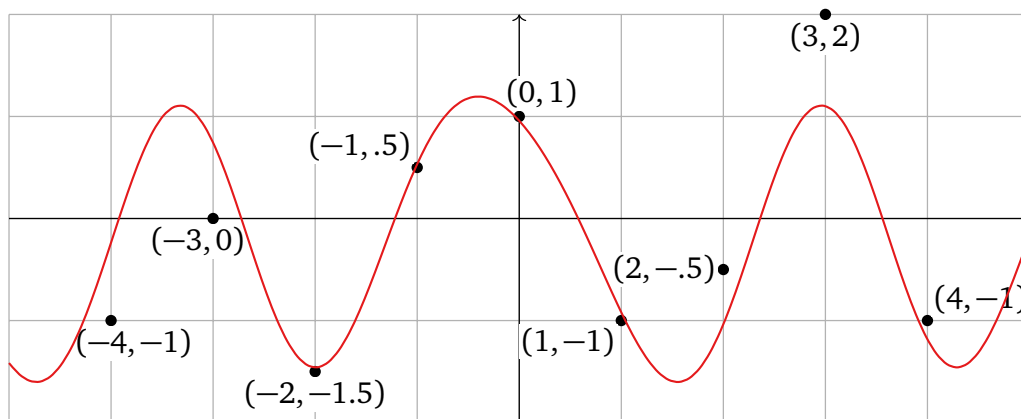
$$\begin{pmatrix} 1 & -0.6536 & 0.7568 & -0.1455 & -0.9894 & 0.8439 & 0.5366 \\ 1 & -0.9900 & -0.1411 & 0.9602 & 0.2794 & -0.9111 & -0.4121 \\ 1 & -0.4161 & -0.9093 & -0.6536 & 0.7568 & 0.9602 & 0.2794 \\ 1 & 0.5403 & -0.8415 & -0.4161 & -0.9093 & -0.9900 & -0.1411 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0.5403 & 0.8415 & -0.4161 & 0.9093 & -0.9900 & 0.1411 \\ 1 & -0.4161 & 0.9093 & -0.6536 & -0.7568 & 0.9602 & -0.2794 \\ 1 & -0.9900 & 0.1411 & 0.9602 & -0.2794 & -0.9111 & 0.4121 \\ 1 & -0.6536 & -0.7568 & -0.1455 & 0.9894 & 0.8439 & -0.5366 \end{pmatrix} \begin{pmatrix} B \\ C \\ D \\ E \\ F \\ G \\ H \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ -1.5 \\ 0.5 \\ 1 \\ -1 \\ -0.5 \\ 2 \\ -1 \end{pmatrix}.$$

We find the least-squares solution with the aid of a computer:

$$\hat{x} \approx \begin{pmatrix} -0.1435 \\ 0.2611 \\ -0.2337 \\ 1.116 \\ -0.5997 \\ -0.2767 \\ 0.1076 \end{pmatrix}.$$

Therefore, the best-fit function is

$$y \approx -0.1435 + 0.2611 \cos(x) - 0.2337 \sin(x) + 1.116 \cos(2x) - 0.5997 \sin(2x) \\ - 0.2767 \cos(3x) + 0.1076 \sin(3x).$$



$$y \approx -0.14 + 0.26 \cos(x) - 0.23 \sin(x) + 1.11 \cos(2x) - 0.60 \sin(2x) - 0.28 \cos(3x) + 0.11 \sin(3x)$$

As in the previous examples, the best-fit function minimizes the sum of the squares of the vertical distances from the graph of $y = f(x)$ to the data points.

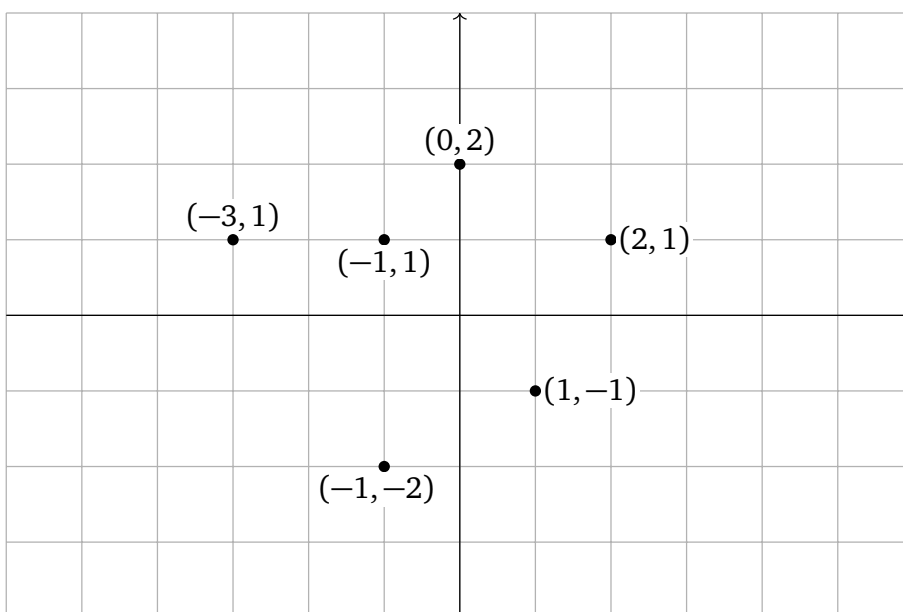
[Use this link to view the online demo](#)

The best-fit function minimizes the sum of the squares of the vertical distances (violet). Click and drag the points to see how the best-fit function changes.

The next example has a somewhat different flavor from the previous ones.

Example (Best-fit ellipse). Find the best-fit ellipse through the points

$$(0, 2), (2, 1), (1, -1), (-1, -2), (-3, 1), (-1, 1).$$



What quantity is being minimized?

Solution. The general equation for an ellipse (actually, for a nondegenerate conic section) is

$$x^2 + By^2 + Cxy + Dx + Ey + F = 0.$$

This is an *implicit equation*: the ellipse is the set of all solutions of the equation, just like the unit circle is the set of solutions of $x^2 + y^2 = 1$. To say that our data points lie on the ellipse means that the above equation is satisfied for the given

values of x and y :

$$\begin{aligned}
 (0)^2 + B(2)^2 + C(0)(2) + D(0) + E(2) + F &= 0 \\
 (2)^2 + B(1)^2 + C(2)(1) + D(2) + E(1) + F &= 0 \\
 (1)^2 + B(-1)^2 + C(1)(-1) + D(1) + E(-1) + F &= 0 \\
 (-1)^2 + B(-2)^2 + C(-1)(-2) + D(-1) + E(-2) + F &= 0 \\
 (-3)^2 + B(1)^2 + C(-3)(1) + D(-3) + E(1) + F &= 0 \\
 (-1)^2 + B(-1)^2 + C(-1)(-1) + D(-1) + E(-1) + F &= 0.
 \end{aligned} \tag{6.5.4}$$

To put this in matrix form, we move the constant terms to the right-hand side of the equals sign; then we can write this as $Ax = b$ for

$$A = \begin{pmatrix} 4 & 0 & 0 & 2 & 1 \\ 1 & 2 & 2 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 \\ 4 & 2 & -1 & -2 & 1 \\ 1 & -3 & -3 & 1 & 1 \\ 1 & 1 & -1 & -1 & 1 \end{pmatrix} \quad x = \begin{pmatrix} B \\ C \\ D \\ E \\ F \end{pmatrix} \quad b = \begin{pmatrix} 0 \\ -4 \\ -1 \\ -1 \\ -9 \\ -1 \end{pmatrix}.$$

We compute

$$A^T A = \begin{pmatrix} 36 & 7 & -5 & 0 & 12 \\ 7 & 19 & 9 & -5 & 1 \\ -5 & 9 & 16 & 1 & -2 \\ 0 & -5 & 1 & 12 & 0 \\ 12 & 1 & -2 & 0 & 6 \end{pmatrix} \quad A^T b = \begin{pmatrix} -19 \\ 17 \\ 20 \\ -9 \\ -16 \end{pmatrix}.$$

We form an augmented matrix and row reduce:

$$\left(\begin{array}{ccccc|c} 36 & 7 & -5 & 0 & 12 & -19 \\ 7 & 19 & 9 & -5 & 1 & 17 \\ -5 & 9 & 16 & 1 & -2 & 20 \\ 0 & -5 & 1 & 12 & 0 & -9 \\ 12 & 1 & -2 & 0 & 6 & -16 \end{array} \right) \xrightarrow{\text{RREF}} \left(\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 0 & 405/266 \\ 0 & 1 & 0 & 0 & 0 & -89/133 \\ 0 & 0 & 1 & 0 & 0 & 201/133 \\ 0 & 0 & 0 & 1 & 0 & -123/266 \\ 0 & 0 & 0 & 0 & 1 & -687/133 \end{array} \right).$$

The least-squares solution is

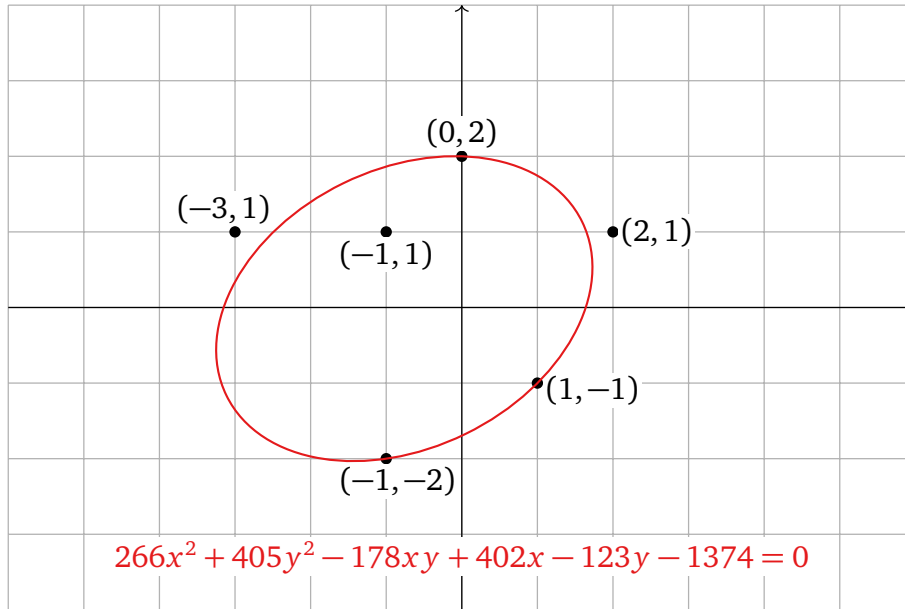
$$\hat{x} = \begin{pmatrix} 405/266 \\ -89/133 \\ 201/133 \\ -123/266 \\ -687/133 \end{pmatrix},$$

so the best-fit ellipse is

$$x^2 + \frac{405}{266}y^2 - \frac{89}{133}xy + \frac{201}{133}x - \frac{123}{266}y - \frac{687}{133} = 0.$$

Multiplying through by 266, we can write this as

$$266x^2 + 405y^2 - 178xy + 402x - 123y - 1374 = 0.$$



Now we consider the question of what quantity is minimized by this ellipse. The least-squares solution \hat{x} minimizes the sum of the squares of the entries of the vector $b - A\hat{x}$, or equivalently, of $A\hat{x} - b$. The vector $-b$ contains the constant terms of the left-hand sides of (6.5.4), and

$$A\hat{x} = \begin{pmatrix} \frac{405}{266}(2)^2 - \frac{89}{133}(0)(2) + \frac{201}{133}(0) - \frac{123}{266}(2) - \frac{687}{133} \\ \frac{405}{266}(1)^2 - \frac{89}{133}(2)(1) + \frac{201}{133}(2) - \frac{123}{266}(1) - \frac{687}{133} \\ \frac{405}{266}(-1)^2 - \frac{89}{133}(1)(-1) + \frac{201}{133}(1) - \frac{123}{266}(-1) - \frac{687}{133} \\ \frac{405}{266}(-2)^2 - \frac{89}{133}(-1)(-2) + \frac{201}{133}(-1) - \frac{123}{266}(-2) - \frac{687}{133} \\ \frac{405}{266}(1)^2 - \frac{89}{133}(-3)(1) + \frac{201}{133}(-3) - \frac{123}{266}(1) - \frac{687}{133} \\ \frac{405}{266}(-1)^2 - \frac{89}{133}(-1)(-1) + \frac{201}{133}(-1) - \frac{123}{266}(-1) - \frac{687}{133} \end{pmatrix}$$

contains the rest of the terms on the left-hand side of (6.5.4). Therefore, the entries of $A\hat{x} - b$ are the quantities obtained by evaluating the function

$$f(x, y) = x^2 + \frac{405}{266}y^2 - \frac{89}{133}xy + \frac{201}{133}x - \frac{123}{266}y - \frac{687}{133}$$

on the given data points.

If our data points actually lay on the ellipse defined by $f(x, y) = 0$, then evaluating $f(x, y)$ on our data points would always yield zero, so $A\hat{x} - b$ would be the zero vector. This is not the case; instead, $A\hat{x} - b$ contains the *actual* values of $f(x, y)$ when evaluated on our data points. The quantity being minimized is the sum of the squares of these values:

minimized =

$$f(0, 2)^2 + f(2, 1)^2 + f(1, -1)^2 + f(-1, -2)^2 + f(-3, 1)^2 + f(-1, 1)^2.$$

One way to visualize this is as follows. We can put this best-fit problem into the framework of this [example](#) by asking to find an equation of the form

$$f(x, y) = x^2 + By^2 + Cxy + Dx + Ey + F$$

which best approximates the data table

x	y	$f(x, y)$
0	2	0
2	1	0
1	-1	0
-1	-2	0
-3	1	0
-1	-1	0.

The resulting function minimizes the sum of the squares of the vertical distances from these data points $(0, 2, 0)$, $(2, 1, 0)$, \dots , which lie on the xy -plane, to the graph of $f(x, y)$.

[Use this link to view the online demo](#)

The best-fit ellipse minimizes the sum of the squares of the vertical distances (violet) from the points $(x, y, 0)$ to the graph of $f(x, y)$ on the left. The ellipse itself is the zero set of $f(x, y)$, on the right. Click and drag the points on the right to see how the best-fit ellipse changes. Can you arrange the points so that the best-fit conic section is actually a hyperbola?

Note. Gauss invented the method of least squares to find a best-fit ellipse: he correctly predicted the (elliptical) orbit of the asteroid Ceres as it passed behind the sun in 1801.