
1 Least-Squares Approximation

Often an exact solution to a problem in applied mathematics is difficult or impossible to obtain. However, it is usually just as useful to find an approximation to a solution. In particular, finding “linear approximations” is a powerful technique in applied mathematics. One basic case is the situation where a system of linear equations has no solution, and it is desirable to find a “best approximation” to a solution to the system.

We begin by defining the “best approximation” in a natural way, and showing that computing the best approximation reduces to solving a related system of linear equations called the *normal equations*. Next, we demonstrate a common application where a collection of data points is approximated by a line (or a curve). We conclude this section by showing that *QR*-factorization provides us with a more efficient way to solve the normal equations and compute the best approximation.

Best Approximate Solutions

Exploration 1 Let

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. A quick examination of the last two rows should convince you that this equation has no solutions. In other words, $\mathbf{b}$ is not in the span of the columns of A .

If $\mathbf{z}$ were an exact solution to $A\mathbf{x} = \mathbf{b}$, then $\mathbf{b} - A\mathbf{z}$ would be $\mathbf{0}$. Since the equation does not have a solution, we will attempt to find the next best thing to a solution by finding $\mathbf{z}$ such that $\|\mathbf{b} - A\mathbf{z}\|$ is as small as possible. The quantity $\|\mathbf{b} - A\mathbf{z}\|$ is called the error.

The following GeoGebra interactive will help you understand the geometry behind finding $\mathbf{z}$.

RIGHT-CLICK and DRAG to rotate the image for a better view.

Access GeoGebra interactives through the online version of this text at

<https://ximera.osu.edu/linearalgebradv3/LinearAlgebraInteractiveIntro>.

Record your best guess for $\mathbf{z}$ – you will have a chance to check your answer in Example ??.

What did you discover about the geometry of minimizing $\|\mathbf{b} - A\mathbf{z}\|$? Select all that apply.

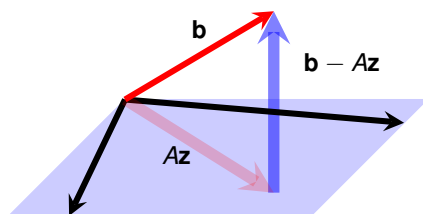
Select All Correct Answers:

1. $\mathbf{z}$ is orthogonal to the plane spanned by the columns of A .
2. $\|\mathbf{b} - A\mathbf{z}\|$ is orthogonal to $\text{col}(A)$.
3. $\mathbf{b} - A\mathbf{z}$ is orthogonal to $\text{col}(A)$.
4. $A\mathbf{z}$ is orthogonal to $\text{col}(A)$.
5. $A\mathbf{z}$ is an orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$.

Our geometric observations will help us develop a method for finding $\mathbf{z}$.

Suppose A is an $m \times n$ matrix, and $\mathbf{b}$ is a column vector in $\mathbb{R}^m$. Consider the matrix equation $A\mathbf{x} = \mathbf{b}$. If this equation does not have a solution, we can attempt to find a best approximation by finding $\mathbf{z}$ which minimizes the error, $\|\mathbf{b} - A\mathbf{z}\|$. The expression $\|\mathbf{b} - A\mathbf{z}\|$ is also sometimes called the *residual*. The error (or the residual) is given in terms of a vector norm. Recall that our definition of the norm involves the sum of squares of the vector components. When we minimize the norm, we minimize the sum of squares. This is why the method we are describing is often referred to as *least squares*. We will explore this idea further later in this section.

In the case when $\text{col}(A)$ is a subspace of $\mathbb{R}^3$, we can see geometrically that $\mathbf{z}$ is the best approximation if and only if $A\mathbf{z}$ is an orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$, and the error is the magnitude of $\mathbf{b} - A\mathbf{z}$, as shown below.



What we observed above, holds in general. We will use this fact to find $\mathbf{z}$.

Every vector in $\text{col}(A)$ can be written in the form $A\mathbf{x}$ for some $\mathbf{x}$ in $\mathbb{R}^n$. Our goal is to find $\mathbf{z}$ such that $A\mathbf{z}$ is the orthogonal projection of $\mathbf{b}$ onto $\text{col}(A)$. By Corollary ??, every vector $A\mathbf{x}$ in $\text{col}(A)$ is orthogonal to $\mathbf{b} - A\mathbf{z}$.

Therefore, we have

$$\begin{aligned} A^T(\mathbf{b} - A\mathbf{z}) &= \mathbf{0} \\ A^T\mathbf{b} - A^TA\mathbf{z} &= \mathbf{0} \\ A^TA\mathbf{z} &= A^T\mathbf{b} \end{aligned} \tag{1}$$

Since $\mathbf{b} - A\mathbf{z}$ is normal to the subspace $\text{col}(A)$, we call the system of linear equations in (??) the *normal equations* for $\mathbf{z}$. If A^TA is invertible, then we can write

$$\mathbf{z} = (A^TA)^{-1}A^T\mathbf{b} \tag{2}$$

We will return to the question of invertibility of A^TA in Theorem ??. For now, let's revisit the problem posed in Exploration ??.

Example 1. We now return to the matrix equation $A\mathbf{x} = \mathbf{b}$ of Exploration ?? to find $\mathbf{z}$ that best approximates a solution.

Explanation. Recall that

$$A = \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$$

In this case, $(A^T A)^{-1}$ exists. Applying equation (??), we compute

$$\begin{aligned} \mathbf{z} &= \left(\begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 3 & 1 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 3 & 1 & 1 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 11 & 7 \\ 7 & 9 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 10 \end{bmatrix} = \begin{bmatrix} 0.18 & -0.14 \\ -0.14 & 0.22 \end{bmatrix} \begin{bmatrix} 10 \\ 10 \end{bmatrix} \\ &= \begin{bmatrix} 0.4 \\ 0.8 \end{bmatrix} \end{aligned}$$

Compare this answer to your guess in Exploration ???. If your guess was correct, nice job! If your guess was different, try setting $\mathbf{z}$ to the correct answer and use the GeoGebra interactive in Exploration ?? to examine the geometry of the problem.

We now come back to the question of when $A^T A$ is invertible.

Theorem 1. If columns of matrix A are linearly independent, then $A^T A$ is invertible.

Proof Let A be a matrix with linearly independent columns. We will show that $(A^T A) \mathbf{x} = \mathbf{0}$ has only the trivial solution. For $\mathbf{x}$, a solution of $A^T A \mathbf{x} = \mathbf{0}$, we have

$$\begin{aligned} \|\mathbf{Ax}\|^2 &= (\mathbf{Ax}) \cdot (\mathbf{Ax}) \\ &= (\mathbf{Ax})^T \mathbf{Ax} \\ &= \mathbf{x}^T A^T \mathbf{Ax} \\ &= \mathbf{x}^T \cdot \mathbf{0} = 0 \end{aligned}$$

Therefore $\mathbf{Ax} = \mathbf{0}$. By linear independence of the columns of A we conclude that $\mathbf{x} = \mathbf{0}$. ■

We summarize our findings in the following theorem.

Theorem 2. Let A be an $m \times n$ matrix, let $\mathbf{b}$ be a column vector in $\mathbb{R}^m$. Consider the matrix equation

$$\mathbf{Ax} = \mathbf{b}$$

1. Any solution $\mathbf{z}$ to the normal equations

$$(A^T A) \mathbf{z} = A^T \mathbf{b}$$

is a best approximation to a solution to $\mathbf{Ax} = \mathbf{b}$ in the sense that $\|\mathbf{b} - \mathbf{Az}\|$ is minimized.

2. If the columns of A are linearly independent, then $A^T A$ is invertible and $\mathbf{z}$ is given uniquely by

$$\mathbf{z} = (A^T A)^{-1} A^T \mathbf{b}$$

Example 2. The system of linear equations

$$\begin{aligned} 3x - y &= 4 \\ x + 2y &= 0 \\ 2x + y &= 1 \end{aligned}$$

has no solution. Find the vector $\mathbf{z}$ that best approximates a solution.

Explanation. We have

$$A = \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix}, \quad \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

The normal equations are $(A^T A) \mathbf{z} = A^T \mathbf{b}$. We compute

$$\begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 2 \\ 2 & 1 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 3 & 1 & 2 \\ -1 & 2 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ 0 \\ 1 \end{bmatrix}.$$

$$\begin{bmatrix} 14 & 1 \\ 1 & 6 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 14 \\ -3 \end{bmatrix}$$

We observe that $A^T A$ is invertible. Multiplying on the left by $(A^T A)^{-1}$ yields

$$\mathbf{z} = \frac{1}{83} \begin{bmatrix} 87 \\ -56 \end{bmatrix}$$

With this vector $\mathbf{z}$, the left sides of the equations become

$$\begin{array}{rclcl} 3(87/83) & - & (-56/83) & \approx & 3.82 \\ (87/83) & + & 2(-56/83) & \approx & -0.30 \\ 2(87/83) & + & (-56/83) & \approx & 1.42 \end{array}$$

This is as close as possible to a solution.

Example 3. The average number g of goals per game scored by a hockey player seems to be related linearly to two factors: the number x_1 of years of experience and the number x_2 of goals in the preceding 10 games. The data on the following page were collected on four players. Find the linear function $g = a_0 + a_1 x_1 + a_2 x_2$ that best fits the data.

g	x_1	x_2
0.8	5	3
0.8	3	4
0.6	1	5
0.4	2	1

Explanation. If the relationship is given by $g = a_0 + a_1 x_1 + a_2 x_2$, then the data can be described as follows:

$$\begin{bmatrix} 1 & 5 & 3 \\ 1 & 3 & 4 \\ 1 & 1 & 5 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix}$$

Using Theorem ??, we get

$$\mathbf{z} = \frac{1}{42} \underbrace{\begin{bmatrix} 119 & -17 & -19 \\ -17 & 5 & 1 \\ -19 & 1 & 5 \end{bmatrix}}_{(A^T A)^{-1}} \underbrace{\begin{bmatrix} 1 & 1 & 1 & 1 \\ 5 & 3 & 1 & 2 \\ 3 & 4 & 5 & 1 \end{bmatrix}}_{A^T} \begin{bmatrix} 0.8 \\ 0.8 \\ 0.6 \\ 0.4 \end{bmatrix} = \begin{bmatrix} 0.14 \\ 0.09 \\ 0.08 \end{bmatrix}$$

Hence the best-fitting function is $g = 0.14 + 0.09x_1 + 0.08x_2$.

Application of Least Squares to Curve Fitting

In [Curve Fitting](#), we discussed how to fit a function to a set of data points, so that the graph of the function passes through each of the points. We also discussed why doing so is sometimes impossible (two points lie one above the other), and may not even be desirable (overfitting). In this section we will learn how to approximate a collection of data points with a line (or a curve) that fits the “trend” of the points. We will start with data that fit a linear pattern.

Exploration 2 Consider the points $(1, 1)$, $(2, 3)$ and $(4, 4)$. These points do not lie on a straight line, but they have a general upward linear trend. (Typically there would be many more points to consider, but we will limit our exploration to what we can do by hand.) Our goal is to find a line that fits these points as closely as possible.

We are looking for a function f of the form $f(x) = ax + b$ such that the following infeasible system is satisfied as closely as possible

$$\begin{aligned}a(1) + b &= 1 \\a(2) + b &= 3 \\a(4) + b &= 4\end{aligned}$$

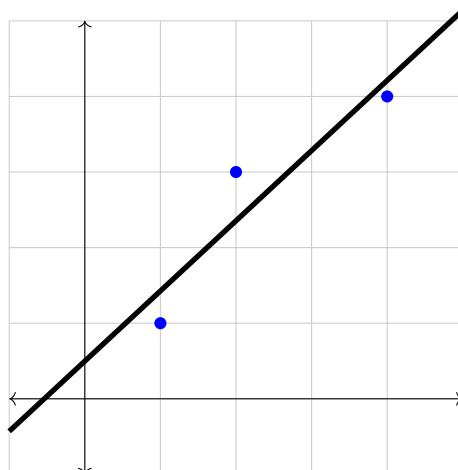
From the first part of this section we know how to find a best approximation. By Theorem ??, we have

$$\begin{aligned}\mathbf{z} = \begin{bmatrix} a \\ b \end{bmatrix} &= (A^T A)^{-1} A^T \mathbf{b} \\&= \left(\begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 4 & 1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} \\&= \begin{bmatrix} 13/14 \\ 1/2 \end{bmatrix}\end{aligned}$$

According to our computations, the line that best fits the data is given by

$$f(x) = \frac{13}{14}x + \frac{1}{2}$$

Let's take a look.



We found this fit by minimizing $\|\mathbf{b} - \mathbf{Az}\|$. We will now investigate the meaning of this expression in relation to the line and the data points.

$$\mathbf{b} - \mathbf{Az} = \begin{bmatrix} 1 \\ 3 \\ 4 \end{bmatrix} - \begin{bmatrix} (13/14)(1) + 1/2 \\ (13/14)(2) + 1/2 \\ (13/14)(4) + 1/2 \end{bmatrix} \approx \begin{bmatrix} -0.43 \\ 0.64 \\ -0.21 \end{bmatrix} \quad (3)$$

Observe that each entry of $\mathbf{b} - \mathbf{Az}$ is the signed vertical distance between a particular point and the line.

Instead of computing the error, $\|\mathbf{b} - \mathbf{Az}\|$, we will compute $\|\mathbf{b} - \mathbf{Az}\|^2$ to avoid the square root.

$$\|\mathbf{b} - \mathbf{Az}\|^2 = (-0.43)^2 + 0.64^2 + (-0.21)^2 \approx 0.64 \quad (4)$$

Minimizing $\|\mathbf{b} - \mathbf{Az}\|$ also minimizes $\|\mathbf{b} - \mathbf{Az}\|^2$. Therefore, what we have minimized is the sum of squares of the vertical distances between the data points and the line. The following GeoGebra interactive will help you explore this idea.

Access GeoGebra interactives through the online version of this text at

<https://ximera.osu.edu/linearalgebradv3/LinearAlgebraInteractiveIntro>.

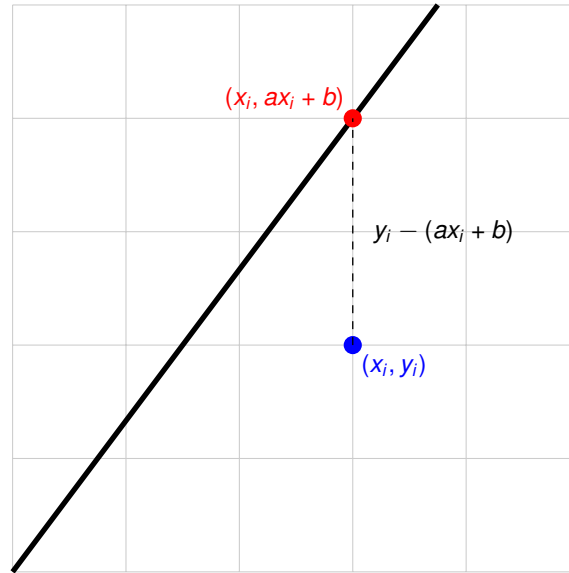
In Exploration ?? we discovered that $\|\mathbf{b} - \mathbf{Az}\|^2$ is the sum of squares of vertical distances between the given data points and the proposed line. By minimizing $\|\mathbf{b} - \mathbf{Az}\|$, we minimize the sum of squares of vertical distances. This observation holds in general. Given a collection of points $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, finding a linear function of the form $f(x) = ax + b$ that best fits the points we would find a best solution to the system

$$\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

by minimizing

$$\left\| \begin{bmatrix} y_1 - (ax_1 + b) \\ y_2 - (ax_2 + b) \\ \vdots \\ y_n - (ax_n + b) \end{bmatrix} \right\|^2 = (y_1 - (ax_1 + b))^2 + (y_2 - (ax_2 + b))^2 + \cdots + (y_n - (ax_n + b))^2$$

A geometric interpretation of $y_i - (ax_i + b)$ is shown below.



The line we obtain in this fashion is called a *line of best fit* or a *trendline*, and the method we used is referred to as the *method of least squares*.

We can apply the method of least squares to find best fitting non-linear functions.

Example 4. Find the least squares approximating quadratic polynomial of the form $f(x) = ax^2 + bx + c$ for the following points.

$$(-3, 3), (-1, 1), (0, 1), (1, 2), (3, 4)$$

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{aligned} a(-3)^2 + b(-3) + c &= 3 \\ a(-1)^2 + b(-1) + c &= 1 \\ a(0)^2 + b(0) + c &= 1 \\ a(1)^2 + b(1) + c &= 2 \\ a(3)^2 + b(3) + c &= 4 \end{aligned}$$

This corresponds to the matrix equation

$$\begin{bmatrix} 9 & -3 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 1 \\ 9 & 3 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 1 \\ 2 \\ 4 \end{bmatrix}$$

Multiplying on the left by A^T gives us the normal equations.

$$A^T A \mathbf{z} = A^T \mathbf{b}$$

$$\begin{bmatrix} 164 & 0 & 20 \\ 0 & 20 & 0 \\ 20 & 0 & 5 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 66 \\ 11 \\ 4 \end{bmatrix}$$

It turns out that $A^T A$ is invertible, so it is easy to solve for $\mathbf{z}$. You can use technology to accomplish this.

We arrive at the solution

$$\mathbf{z} = \begin{bmatrix} 0.26 \\ 0.20 \\ 1.15 \end{bmatrix}$$

Therefore, the quadratic function of best fit is given by $f(x) = 0.26x^2 + 0.2x + 1.15$. You can see the graph and the points shown below.

For more information about doing least squares with Octave, please see [Octave for Chapter 9](#).

Example 5. Given the data points $(-1, 0)$, $(0, 1)$, and $(1, 4)$, find the least squares approximating function of the form $f(x) = ax + b2^x$.

Explanation. We are looking for an approximate solution to the system of equations

$$\begin{aligned} a(-1) + b(2^{-1}) &= 0 \\ a(0) + b(2^0) &= 1 \\ a(1) + b(2^1) &= 4 \end{aligned}$$

This corresponds to the matrix equation

$$\begin{bmatrix} -1 & 1/2 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}$$

Using the normal equations, we obtain

$$A^T A \mathbf{z} = A^T \mathbf{b}$$

$$\frac{1}{4} \begin{bmatrix} 8 & 6 \\ 6 & 21 \end{bmatrix} \mathbf{z} = \begin{bmatrix} 4 \\ 9 \end{bmatrix}$$

Solving for $\mathbf{z}$ yields

$$\mathbf{z} = \begin{bmatrix} 10/11 \\ 16/11 \end{bmatrix}$$

Therefore, the function of best fit (of the given form) is given by

$$f(x) = \frac{10}{11}x + \frac{16}{11}(2^x)$$

Practice Problems

Problem 3 Find the best approximation to a solution to the system of equations.

$$\begin{array}{rrcrcl} 3x & + & y & + & z & = & 6 \\ 2x & + & 3y & - & z & = & 1 \\ 2x & - & y & + & z & = & 0 \\ 3x & - & 3y & + & 3z & = & 8 \end{array}$$

Enter answers in fraction form below.

$$x = \dots\dots, \quad y = \dots\dots, \quad z = \dots\dots$$

Problem 4 Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.

$$(2, 4), (4, 3), (7, 2), (8, 1)$$

Enter your answers in fraction form.

$$f(x) = \dots\dots x + \dots\dots$$

Problem 5 Find a linear function of best fit for the given set of data points. Examine how well your line fits the points by typing the equation of the line into the Desmos window.

$$(-2, 3), (-1, 1), (0, 0), (1, -2), (2, -4)$$

$$f(x) = \dots\dots x + \dots\dots$$

Problem 6 Use Octave to find the least squares approximating quadratic function for the following data points.

$$(-2, 1), (0, 0), (3, 2), (4, 3)$$

Round your answers to three decimal places.

$$f(x) = \dots\dots x^2 + \dots\dots x + \dots\dots$$

For more information about doing least squares with Octave, please see [Octave for Chapter 9](#).

Text Source

A portion of this section has been adapted from W. Keith Nicholson, *Linear Algebra with Applications*, Lyryx 2021-A, Open Edition, pp. 308-319.