Least Squares

Justin Wyss-Gallifent

July 21, 2021

3.1 Introduction

Let's go back to the matrix equation

$$A\bar{x} = \bar{b}$$

We know that a unique solution exists if A is invertible and if A is not invertible then there are either no solutions or infinitely many solutions.

Specifically the question that we'd like to address here is what can we do if there are no solutions at all? One answer might be to just stop, however maybe we could ask the question - what's the nearest solution we could find?

In other words if we can't find $\bar{x}$ so that $A\bar{x} = \bar{b}$, can we find some $\bar{x}$ so that $A\bar{x}$ is as close as possible to $\bar{b}$?

More rigorously can we find some $\hat{x}$ such that:

$$\text{For all } \bar{x} \text{ we have } ||A\hat{x} - \bar{b}|| \leq ||A\bar{x} - \bar{b}||$$

3.2 Reminder - Solutions and Column Space

First let's recall:

Definition 3.2.0.1. Given an $n \times m$ matrix A the *column space* of A is the subspace of $\mathbb{R}^n$ given by:

$$\text{Col}(A) = \text{span}\{\text{Columns of } A\}$$

Fact 3.2.0.1. The column space of A is exactly the vectors $\bar{b}$ such that $A\bar{x} = \bar{b}$ has at least one solution.

To reinforce this, a simple example will do:

Example 3.1. The equation

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \bar{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

has a solution if and only if there are x_1, x_2 with

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \bar{x} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \\ 0 & 3 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 1x_1 + 2x_2 \\ 0x_1 + 3x_2 \\ -1x_1 + 0x_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

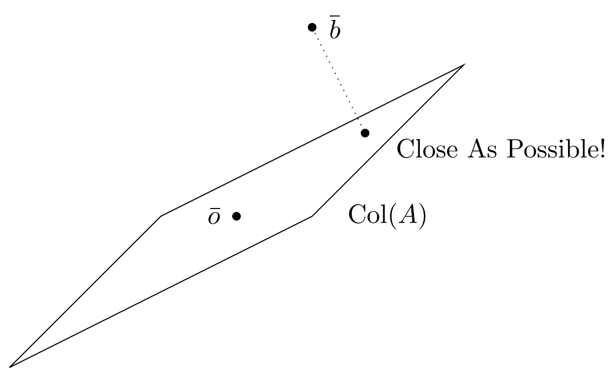
$$x_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix}$$

$$\begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \in \text{col}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} \right\}$$

3.3 The Intuition and Theory

So the situation we're in is that $\bar{b}$ is not in Col(A) and we wish to find $\hat{x}$ so that $A\hat{x}$, which is in Col(A), is as close as possible to $\bar{b}$.

Here's a picture of the situation:



This picture suggests that we can obtain a solution by projecting $\bar{b}$ onto Col(A) to get $\text{Pr}_{\text{Col}(A)}\bar{b}$ and then finding $\hat{x}$ to solve the equation:

$$A\hat{x} = \text{Pr}_{\text{Col}(A)}\bar{b}$$

Assuming this is correct, the problem with this approach, practically, is that calculating $\mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$ requires having an orthogonal basis for Col(A) and this is procedurally intense especially when A is large. So what we'll do is find a sneaky way to find $\hat{x}$ a different way. Just to be clear, we will solve this equation, but we won't solve it by finding $\mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$.

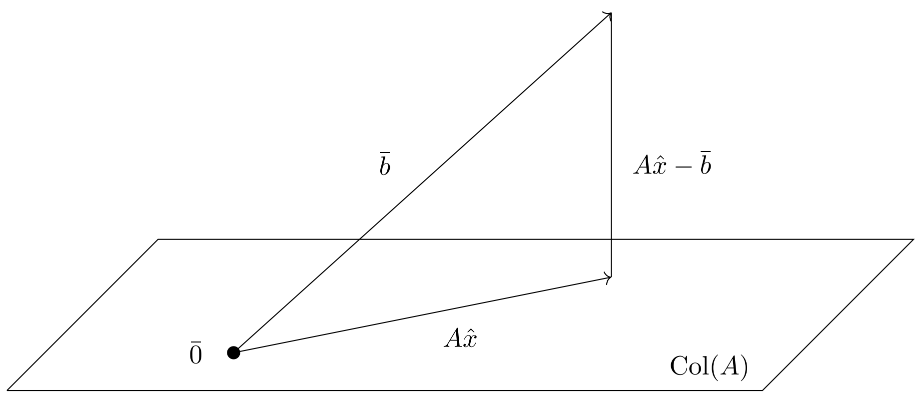
3.4 Theory: Least-Squares Solution

The approach is based on the following two things:

(a) There is a sneaky way of finding $\hat{x}$ such that $A\hat{x} = \mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$.

(b) The $\hat{x}$ we find actually does satisfy $||A\hat{x} - \bar{b}|| \leq ||A\bar{x} - \bar{b}||$ for all $\bar{x}$.

First, let's see how we can do part (a). Finding $\hat{x}$ such that $A\hat{x} = \mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$ means the vector $A\hat{x}-\bar{b}$ is perpendicular to Col(A) as illustrated by this picture:



From here note that $A\hat{x} - \bar{b}$ being perpendicular to Col(A) is equivalent to

$$
\begin{aligned}
A\bar{x} \cdot (A\hat{x} - \bar{b}) &= 0 && \text{for all } \bar{x} \\
(A\bar{x})^T(A\hat{x} - \bar{b}) &= 0 && \text{for all } \bar{x} \\
\bar{x}^T A^T (A\hat{x} - \bar{b}) &= 0 && \text{for all } \bar{x} \\
A^T (A\hat{x} - \bar{b}) &= \bar{0} \\
A^T A\hat{x} &= A^T\bar{b}
\end{aligned}
$$

(Note that $\bar{x}^T\bar{y} = 0$ for all $\bar{x}$ means $\bar{y} = \bar{0}$ because no nonzero vector can be perpendicular to every vector.)

So we simply solve this final equation instead. Notice that this final equation must have a solution since we're effectively solving $A\hat{x} = \mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$. It may have more than one, though, and it will have exactly one precisely when the columns of A are linearly independent (because this is always the case for systems of equations that have solutions) which is precisely when A^TA is invertible (because we're solving $A^TA\hat{x} = A^T\bar{b}$ to get the job done.)

Second, how about part (b). We want to make sure that finding $\hat{x}$ such that $A\hat{x} = \text{Pr}_{\text{Col}(A)}\bar{b}$ actually satisfies $||A\hat{x} - \bar{b}|| \leq ||A\bar{x} - \bar{b}||$ for all $\bar{x}$.

Well as we've seen this $\hat{x}$ is such that $A\hat{x} - \bar{b}$ is perpendicular to $\text{Col}(A)$.

For any $\bar{x}$, since $A\bar{x}$ and $A\hat{x}$ are both in $\text{Col}(A)$, so is $A\bar{x} - A\hat{x}$, so then $A\bar{x} - A\hat{x}$ is perpendicular to $A\hat{x} - \bar{b}$.

From here observe that:

$$(A\hat{x} - \bar{b}) + (A\bar{x} - A\hat{x}) = A\bar{x} - \bar{b}$$

and since the two on the left are perpendicular by the Pythagorean Theorem we have:

$$||A\hat{x} - \bar{b}||^2 + ||A\bar{x} - A\hat{x}||^2 = ||A\bar{x} - \bar{b}||^2$$

and therefore

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

3.5 Practical: Least-Squares Solution

Definition 3.5.0.1. Given the matrix equation

$$A\bar{x} = \bar{b}$$

a *least-squares solution* is a solution $\hat{x}$ satisfying

$$||A\hat{x} - \bar{b}|| \leq ||A\bar{x} - \bar{b}|| \text{ for all } \bar{x}$$

Such an $\hat{x}$ will also satisfy both

$$A\hat{x} = \text{Pr}_{\text{Col}(A)}\bar{b}$$

and

$$A^TA\hat{x} = A^T\bar{b}$$

This latter equation is typically the one used in practice.

Note that there may be either one or infinitely many least-squares solutions.

If the columns of A are linearly independent then there is exactly one solution and this solution is

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

If the columns of A are not linearly independent then there are infinitely many least-squares solutions.

In most situations we will encounter there is just one least-squares solution. From a real-world standpoint this is because we typically use least-squares for overdetermined systems (more equations than unknowns) which yields a matrix equation in which the matrix has more rows than columns. This typically results in columns which are linearly independent.

Definition 3.5.0.2. The *least-squares error* is the difference

$$||A\hat{x} - \bar{b}||$$

which measures how far our $A\hat{x}$ is from the desired $\bar{b}$.

Theorem 3.5.0.1. If the columns of A are linearly independent and if $A\bar{x} = \bar{b}$ has a solution then the least-square solution is the actual solution.

Proof. If $A\bar{x} = \bar{b}$ has a solution then $\bar{b} \in \text{Col}(A)$ and so $\text{Pr}_{\text{Col}(A)}\bar{b} = \bar{b}$ and so $A\hat{x} = \text{Pr}_{\text{Col}(A)}\bar{b}$ is equivalent to $A\hat{x} = \bar{b}$. Thus the solution to $A^T A\hat{x} = A^T\bar{b}$ satisfies $A\hat{x} = \bar{b}$. □

Corollary 3.5.0.1. If A is invertible then this is even more obvious.

Proof. If A is invertible then

$$\left(A^T A\right)^{-1} A^T\bar{b} = A^{-1}\left(A^T\right)^{-1} A^T\bar{b} = A^{-1}\bar{b}$$

in which case the solution to the least-squares equation reduces to the solution to the original equation. □

Example 3.2. Consider the system of equations

$$\begin{aligned} x + 2y &= 6 \\ x + y &= 4 \\ x - y &= 1 \end{aligned}$$

I've chosen this so it's clear that there is no solution. The first two equations have solution $x = 2$, $y = 2$ but this fails in the third, so there is no solution to all three.

Rephrased as a matrix equation:

$$\begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \bar{x} = \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix}$$

We instead solve the least-squares equation:

$$
\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \hat{x} &= \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix}^T \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix} \\
\begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \hat{x} &= \begin{bmatrix} 1 & 1 & 1 \\ 2 & 1 & -1 \end{bmatrix} \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix} \\
\begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix} \hat{x} &= \begin{bmatrix} 11 \\ 15 \end{bmatrix} \\
\hat{x} &= \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 11 \\ 15 \end{bmatrix} \\
\hat{x} &= \begin{bmatrix} 18/7 \\ 23/14 \end{bmatrix}
\end{aligned}
$$

The least-squares error is given by:

$$
\begin{aligned}
||A\hat{x} - \bar{b}|| &= \left|\left| \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 18/7 \\ 23/14 \end{bmatrix} - \begin{bmatrix} 6 \\ 4 \\ 1 \end{bmatrix} \right|\right| \\
&= \left|\left| \begin{bmatrix} -1/7 \\ 3/14 \\ -1/14 \end{bmatrix} \right|\right| \\
&= \frac{1}{\sqrt{14}} \\
&\approx 0.2673
\end{aligned}
$$

This is as small as $||A\bar{x} - \bar{b}||$ can be for any $\bar{x}$. You can test this to convince yourself by plugging in some other $\bar{x}$, maybe some close to $\hat{x}$ and some not. Note that $\hat{x} \approx \begin{bmatrix} 2.57143 \\ 1.64286 \end{bmatrix}$.

$$
\begin{aligned}
\left|\left| A \begin{bmatrix} 2.6 \\ 1.6 \end{bmatrix} - \bar{b} \right|\right| &\approx 0.28284 > 0.26726 \\
\left|\left| A \begin{bmatrix} 2.55 \\ 1.62 \end{bmatrix} - \bar{b} \right|\right| &\approx 0.27911 > 0.26726 \\
\left|\left| A \begin{bmatrix} 3 \\ 2 \end{bmatrix} - \bar{b} \right|\right| &\approx 1.41421 > 0.26726
\end{aligned}
$$

Example 3.3. Consider the system

$$
\begin{aligned}
x + 2y &= 3 \\
x + 2y &= 5
\end{aligned}
$$

Clearly this has no solutions. As a matrix equation we have

$$
\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 3 \\ 5 \end{bmatrix}
$$

Notice that the columns of A are not linearly independent. The least-squares approach gives us:

$$
\begin{aligned}
\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \widehat{\begin{bmatrix} x \\ y \end{bmatrix}} &= \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 3 \\ 5 \end{bmatrix} \\
\begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \widehat{\begin{bmatrix} x \\ y \end{bmatrix}} &= \begin{bmatrix} 1 & 1 \\ 2 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 5 \end{bmatrix} \\
\begin{bmatrix} 2 & 4 \\ 4 & 8 \end{bmatrix} \widehat{\begin{bmatrix} x \\ y \end{bmatrix}} &= \begin{bmatrix} 8 \\ 16 \end{bmatrix}
\end{aligned}
$$

We see that there are infinitely many solutions of the form

$$
\begin{bmatrix} 4 - 2\alpha \\ \alpha \end{bmatrix} \text{ for } \alpha \in \mathbb{R}
$$

Each of these is then a least-squares solution.

3.6 Picture of a Simple Case

In closing, a really simple example can help nail down what we've done.

Consider the matrix equation

$$
\begin{bmatrix} 2 \\ 1 \end{bmatrix} [x_1] = \begin{bmatrix} 2 \\ 2 \end{bmatrix}
$$

Obviously there is no solution. Graphically $\text{Col}(A)$ is the set of multiples of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and there is no solution since $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ is not a multiple of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

When we solve the least-squares problem as follows:

$$
\begin{aligned}
\begin{bmatrix} 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \end{bmatrix} \widehat{[x_1]} &= \begin{bmatrix} 2 \\ 1 \end{bmatrix}^T \begin{bmatrix} 2 \\ 2 \end{bmatrix} \\
[5]\,\widehat{[x_1]} &= [6] \\
\widehat{[x_1]} &= [6/5]
\end{aligned}
$$

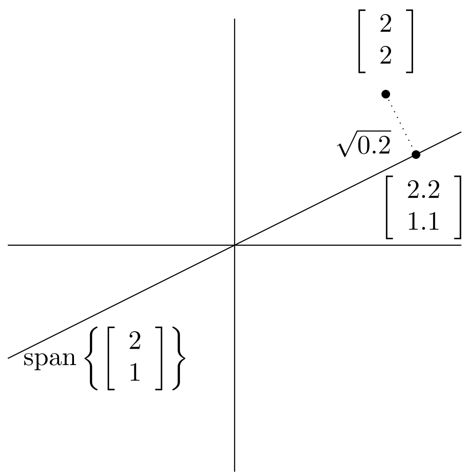
So that

$$A\widehat{[x_1]} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} [6/5] = \begin{bmatrix} 2.4 \\ 1.2 \end{bmatrix}$$

which is in $\mathrm{Col}(A)$ and is as close as possible to $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$, with that distance being the least-squares error:

$$\left|\left| \begin{bmatrix} 2.4 \\ 1.2 \end{bmatrix} - \begin{bmatrix} 2 \\ 1 \end{bmatrix} \right|\right| = \left|\left| \begin{bmatrix} 0.4 \\ 0.2 \end{bmatrix} \right|\right| = \sqrt{0.4^2 + 0.2^2} = \sqrt{0.2}$$

as shown here:



3.7 Matlab

The transpose of a matrix can be done either with the `transpose` command or an apostrophe:

```
>> A = [1 2;1 1;1 -1];
>> transpose(A)
ans =
     1     1     1
     2     1    -1
>> A'
ans =
     1     1     1
     2     1    -1
```

Practically speaking if the columns of A are linearly independent then the least-squares solution can be easily found in Matlab:

```
>> A = [1 2;1 1;1 -1];
>> b = [6;4;1];
>> inv(A'*A)*A'*b
ans =
    2.5714
    1.6429
```

The `norm` command is useful for the least-squares error:

```
>> A = [1 2;1 1;1 -1];
>> b = [6;4;1];
>> x=inv(A'*A)*A'*b;
>> norm(A*x-b)
ans =
    0.2673
```

3.8 Exercises

Exercise 3.1. Find the least-squares solution and least-squares error for the matrix equation

$$\begin{bmatrix} 1 & 2 & 3 \\ -1 & 0 & 2 \\ 5 & 1 & 1 \\ 2 & 2 & 0 \end{bmatrix} \bar{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}$$

Exercise 3.2. Find the vector in col$A$ closest to $\bar{b}$ where:

$$A = \begin{bmatrix} 1 & 2 \\ 0 & -3 \\ 2 & 6 \end{bmatrix} \text{ and } \bar{b} = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

Exercise 3.3. Using least squares, find the vector in

$$\text{span}\left\{ \begin{bmatrix} 1 \\ 2 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ 3 \\ 0 \\ 3 \end{bmatrix} \right\} \text{ closest to } \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Exercise 3.4. Assuming the dimensions all work out, is it possible for the matrix equation $A^T A\bar{x} = A^T\bar{b}$ to have no solutions? Explain.

Exercise 3.5. Could a system of equations with more variables than equations have a unique least squares solution? Explain.

Exercise 3.6. Using least squares, find the vector in

$$\operatorname{span}\left\{\begin{bmatrix}1\\2\end{bmatrix}\right\} \text{ closest to } \begin{bmatrix}5\\5\end{bmatrix}$$

Exercise 3.7. Use the method of least-squares to find the point on the line $y = 3x$ closest to $(2, 3)$.

Exercise 3.8. Suppose A is invertible so that $A\bar{x} = \bar{b}$ actually has a single solution but you use the method of least-squares anyway. Show that the solution you get via least-squares is the actual solution. Hint: Manipulate the least-squares formula.

Exercise 3.9. Show that the following does not have a unique least-squares solution by attempting to find such a solution and explaining where the process fails:

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

Exercise 3.10. Consider the following matrix equation $A\bar{x} = \bar{b}$ with:

$$\begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 0 & 3 \end{bmatrix} \bar{x} = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}$$

(a) Find the least-squares solution the long way:

  (i) Find an orthogonal basis for $\mathrm{Col}(A)$. You can do this by calling the columns $\bar{c}_1$ and $\bar{c}_2$ and then using $\{\bar{c}_1, \bar{c}_2 - \mathrm{Pr}_{\bar{c}_1}\bar{c}_2\}$ as a basis.

  (ii) Find $\mathrm{Pr}_{\mathrm{Col}(A)}\bar{b}$. This equals the sum of the projections of $\bar{b}$ onto each of the basis vectors. Call this $\hat{b}$.

  (iii) Solve $A\bar{x} = \hat{b}$.

(b) Find the least-squares solution using the easy method and verify that they're the same.

Exercise 3.11. Repeat the previous question with:

$$\begin{bmatrix} 2 & -2 \\ 1 & 4 \\ 1 & 2 \end{bmatrix} \bar{x} = \begin{bmatrix} -1 \\ 3 \\ 1 \end{bmatrix}$$

Exercise 3.12. Explain why a least-squares problem always has a solution. Your answer should touch on the issue of the column space and what is really going on under the hood.

Exercise 3.13. Assuming it exists we know that the least-squares solution is given by:

$$\hat{x} = \left(A^T A\right)^{-1} A^T \bar{b}$$

What is mathematically wrong with the following attempt to simplify this? Specifically, which equals signs are not valid and why?

$$\hat{x} = \left(A^T A\right)^{-1} A^T \bar{b} = A^{-1} \left(A^T\right)^{-1} A^T \bar{b} = A^{-1} \bar{b}$$

Exercise 3.14. Suppose we're solving for the least squares solution to $A\bar{x} = \bar{b}$. Why will switching the order of the columns in A have no effect on the solution?