
EECS 16A Designing Information Devices and Systems I

Spring 2021 Lecture Notes

Note 23

23.1 Introduction: Least Squares

In our previous exploration of linear algebra, we developed techniques such as Gaussian Elimination and matrix inversion to solve systems of linear equations exactly. Gaussian Elimination and inversion work best when we have perfect measurements of our system, i.e. when there are no errors or noise in the system. However, given a system which might be prone to noisy measurements, Gaussian Elimination might not be able to solve it; since the equations might end up being inconsistent and have no solution.

For example, in the GPS problem, we use cross-correlation to measure the distance between the satellites (beacons) and the receiver to generate the distances to feed into the trilateration algorithm. However, interference in the wireless signals or errors in correlation can easily lead to noisy measurements of the distance between the receiver and the transmitter.

We'd like to develop a method to solve a system of linear equations even when the system might be inconsistent due to noise. In particular, our goal is to understand how collecting more equations than unknowns can help us reduce the impact of noise and find a solution to the system of equations that might not be perfect, but is as close as possible.

This note develops the *Least Squares* technique for approximately solving systems of linear equations in the presence of noise. Our focus will be on *overdetermined* systems of equations (more equations than unknowns), so we can use the extra equations to reduce the impact of noise. Least squares is the fundamental idea behind *data fitting* and *machine learning*: In data fitting, we find lines or curves that best match the data. In machine learning, we use a best-fit curve to make predictions about new, unseen data.

23.2 Approximate Solutions to Linear Systems

So far, we have studied many problems that ultimately reduce to linear equations of the form

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

When given a system of linear equations, we have sought to assign values to variables that satisfy *all* of the equations *exactly*.

However, when there is noise, this may result in an *inconsistent* system, where no \vec{x} exists that satisfies all the equations exactly.

Fundamentally, least squares deals with choosing a solution \vec{x} that may not satisfy all equations, but minimizes the magnitude of the *error* in our solution. The error of the (approximate) solution \vec{x} is defined as:

$$\vec{e} = \vec{b} - A\vec{x}$$

Let us write out the columns of an $n \times m$ matrix A explicitly and rearrange the above expression, to obtain

$$\begin{bmatrix} | & | & \cdots & | \\ \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_m \\ | & | & \cdots & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix} + \vec{e} = \vec{b}.$$

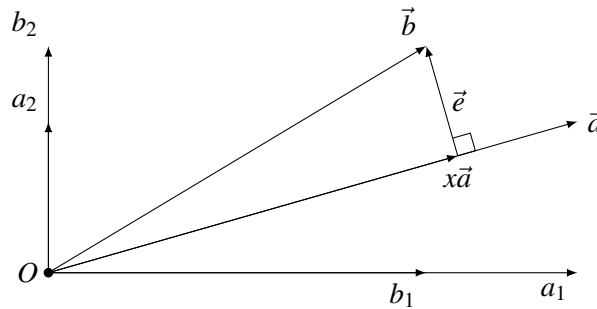
Expanding out the matrix multiplication, we obtain

$$(x_1\vec{a}_1 + x_2\vec{a}_2 + \dots + x_m\vec{a}_m) + \vec{e} = \vec{b},$$

In other words, we are trying to find the “closest” vector to \vec{b} in the span of the columns $\{\vec{a}_i\}$ of A .

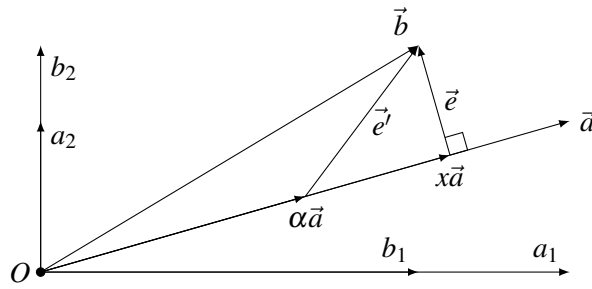
23.2.1 Least Squares: The 2D Case

How can we find such a vector? To gain some intuition, let’s simplify the problem slightly, working in $n = 2$ dimensional space, where A has $m = 1$ columns i.e, A is a 2×1 column vector \vec{a} . Then we would like to find a solution x that minimizes the error $\|\vec{e}\| = \|\vec{b} - x\vec{a}\|$ (note that x is a scalar solution, and \vec{b} and \vec{e} are therefore 2×1 column vectors:



Since x is a scalar, the vector $x\vec{a}$ will have the same direction as \vec{a} . To minimize the magnitude of the error vector, we therefore want to find the closest point on the line defined by the vector a , to the vector b . The diagram suggests that we should drop a perpendicular from \vec{b} onto \vec{a} to find the closest point (known as the *projection* of \vec{b} onto \vec{a}), making the error \vec{e} orthogonal to \vec{a} . We can make this intuition rigorous:

Consider a different solution $\alpha \in \mathbb{R}$ such that $\alpha\vec{a}$ lies in the span of \vec{a} (since α is a scalar, this will always be true). Since we assume that $\alpha \neq x$, we can assume without loss of generality that $\alpha < x$. Consider the error between $\alpha\vec{a}$ and \vec{b} : $\vec{e}' = \vec{b} - \alpha\vec{a}$. Plotting this on a diagram, we see that \vec{e}' forms the hypotenuse of a right triangle, where one leg is the vector \vec{e} :



Thus, by the Pythagorean Theorem, $\|\vec{e}'\|^2 = \|\vec{e}\|^2 + \|\vec{x}\vec{a} - \alpha\vec{a}\|^2$. Since $\alpha \neq x$, we know that the last term is strictly positive. Hence, $\|\vec{e}'\| > \|\vec{e}\|$. Therefore the error with $\alpha\vec{a}$ is greater than the error with $x\vec{a}$, and hence α is not an optimal solution. Since α was an arbitrary solution, this shows that x is indeed optimal.

We have shown that the projection of \vec{b} onto \vec{a} is the solution that minimizes the error \vec{e} . But what is this projection? Well, we know that $\vec{e} \perp x\vec{a}$. Using the properties of inner products, we get:

$$\begin{aligned}
 & \langle \vec{e}, x\vec{a} \rangle = 0 \\
 \implies & x \langle \vec{e}, \vec{a} \rangle = 0 \\
 \implies & \langle \vec{b} - x\vec{a}, \vec{a} \rangle = 0 \\
 \implies & \langle \vec{b}, \vec{a} \rangle - x \langle \vec{a}, \vec{a} \rangle = 0 \\
 \implies & x = \frac{\langle \vec{b}, \vec{a} \rangle}{\langle \vec{a}, \vec{a} \rangle}.
 \end{aligned}$$

This is known as the *projection* of \vec{b} onto the span of \vec{a}_1 . It is also commonly called an *orthogonal projection*. Since the orthogonal projection gives the smallest error solution to $A\vec{x} = \vec{b}$, we have essentially derived the Least Squares algorithm in two dimensions.

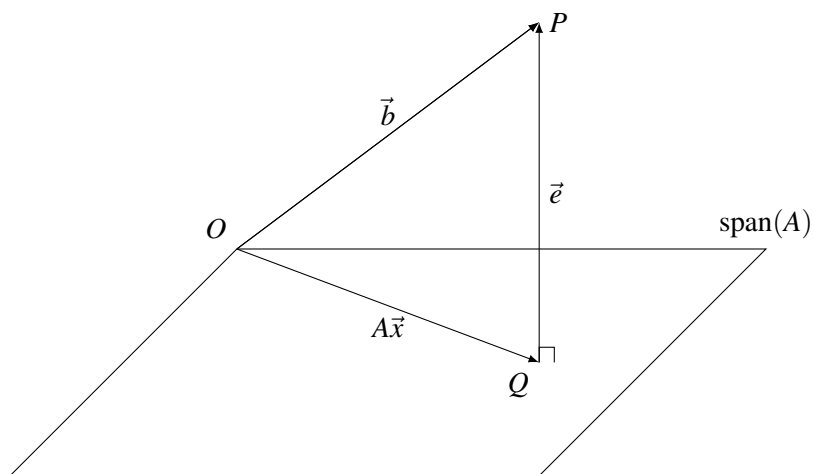
23.2.2 Least Squares: The General Case

Now, how do we generalize beyond two dimensions? Well, we argued in 2D that, in order to minimize the magnitude of the error, we should choose x_1 such that $\vec{e} \perp x_1$ - i.e \vec{e} is orthogonal to the subspace in which $x_1\vec{a}_1$ lies. The analogous claim in higher dimensions, given that we wish to choose \vec{x} to minimize the magnitude of the error

$$\vec{e} = \vec{b} - A\vec{x},$$

would be to choose \vec{x} that make \vec{e} orthogonal to the span of all possible $A\vec{x}$ - i.e \vec{e} that is orthogonal to the column space of A . Can we justify this claim?

We can generalize what we did earlier in the 2D case to higher dimensions to derive the general Least Squares algorithm. What are all the possible values that $A\vec{x}$ can take? This is the set of all linear combinations of the columns of A , i.e the columnspace of A . This is a hyperplane in n -dimensional space that passes through the origin. To help us visualize this, we will illustrate the span of the columns of A as a 2D plane embedded in 3D space:

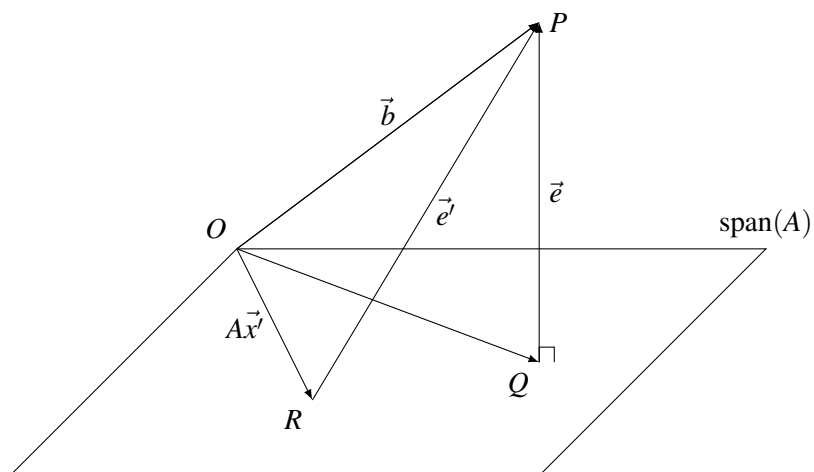


In the above diagram, the vector \vec{b} does not lie in the column space of A , which is why doing Gaussian Elimination on $A\vec{x} = \vec{b}$ would lead to an inconsistent set of equations that cannot be solved.

Our goal is to find that vector $A\vec{x}$ in $\text{span}(A)$ that is closest to \vec{b} (i.e. we want to minimize the magnitude of the error $\|\vec{e}\| = \|\vec{b} - A\vec{x}\|$). This will give us a valid solution \vec{x} (such that $A\vec{x}$ lies in the column space of A) and is as close as possible to one that generates the measurements \vec{b} .

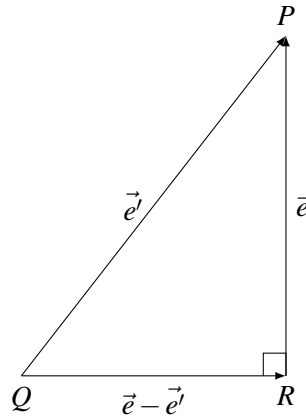
We drop a perpendicular from the tip of the vector \vec{b} (denoted by P) to the plane $\text{span}(A)$, and label the point of intersection with the column space of A point Q . Consider the vector $A\vec{x}$ that connects the origin O to point Q .

As before, our goal is to show that this choice of \vec{x} minimizes the magnitude of the error $\|\vec{e}\|$. To do so, by analogy with our 2D geometric proof, we will consider any other \vec{x}' , where $A\vec{x}'$ is the point R on the plane $\text{span}(A)$. Adding it to our figure, where $\vec{e}' = \vec{b} - A\vec{x}'$, we obtain



We'd like to show that $\|\vec{e}'\| \geq \|\vec{e}\|$. In the 2D case, we were able to do so by drawing a right triangle where \vec{e} was a leg and \vec{e}' the hypotenuse. Can we do something similar here?

From the figure, we see that the triangle formed by P , Q , and R is a triangle with the desired side lengths, as illustrated below:



But is it a *right* triangle? Observe that the bottom leg of the triangle is the vector $\vec{e} - \vec{e}' = A(\vec{x} - \vec{x}')$, which lies in the column space of A . Thus, by construction, it is orthogonal to \vec{e} , so this is indeed a right triangle!

Therefore, we can simply apply the Pythagorean theorem to conclude that $\|\vec{e}'\| \geq \|\vec{e}\|$, as desired. So as we have shown that no \vec{x}' can produce a smaller error than \vec{x} , we have proven that \vec{x} is the optimal solution!

23.2.2.1 Algebraic Proof

We can also show this using an algebraic proof. Assume that an \vec{x} exists such that $\vec{e} = \vec{b} - A\vec{x}$ is orthogonal to all elements in the column space of A . Now, imagine for the sake of *contradiction* that there existed an alternative \vec{x}' with corresponding error $\vec{e}' = \vec{b} - A\vec{x}'$, such that

$$\|\vec{e}'\| < \|\vec{e}\|.$$

In other words, we are imagining that there exists a strictly *better* solution to $A\vec{x} \approx \vec{b}$ than \vec{x} . From the equation relating the norms of the errors, we have that

$$\begin{aligned} & \|\vec{e}'\| < \|\vec{e}\| \\ \implies & \|\vec{e}'\|^2 < \|\vec{e}\|^2 \\ \implies & \langle \vec{e}', \vec{e}' \rangle < \langle \vec{e}, \vec{e} \rangle \end{aligned}$$

So far, all we've done is make simple manipulations to our inequality, squaring to write the norm of a vector in terms of inner products. Now, we will use a small trick. We need to use the fact that \vec{e} is orthogonal to any element in the span of A *somewhere* in our proof. In other words, we want an expression of the form $\langle \vec{e}, A\vec{v} \rangle$ (for some vector \vec{v}) to appear somewhere, so that we can set it to 0 and apply this fact.

Where could such an expression $A\vec{v}$ come from? One way is to recognize that

$$\begin{aligned}\vec{e}' &= \vec{b} - A\vec{x}' \\ &= \vec{b} - A\vec{x} + A\vec{x} - A\vec{x}' \\ &= \vec{e} + A(\vec{x} - \vec{x}'),\end{aligned}$$

adding and subtracting the same term $A\vec{x}$ in order to bring \vec{e} back into the equation. Making this substitution, we find that

$$\langle \vec{e} + A(\vec{x} - \vec{x}'), \vec{e} + A(\vec{x} - \vec{x}') \rangle < \langle \vec{e}, \vec{e} \rangle.$$

Now, we will apply the distributive property of inner products on the left-hand-side of the inequality, to obtain

$$\langle \vec{e}, \vec{e} \rangle + \langle \vec{e}, A(\vec{x} - \vec{x}') \rangle + \langle A(\vec{x} - \vec{x}'), \vec{e} \rangle + \langle A(\vec{x} - \vec{x}'), A(\vec{x} - \vec{x}') \rangle < \langle \vec{e}, \vec{e} \rangle.$$

Recall that inner products of real vectors are commutative, so

$$\langle \vec{e}, A(\vec{x} - \vec{x}') \rangle = \langle A(\vec{x} - \vec{x}'), \vec{e} \rangle.$$

Using the above identity to combine terms, and cancelling equal terms on both sides of the inequality, we find that

$$2\langle A(\vec{x} - \vec{x}'), \vec{e} \rangle + \langle A(\vec{x} - \vec{x}'), A(\vec{x} - \vec{x}') \rangle < 0.$$

And now we can use our orthogonality assumption! Since $A(\vec{x} - \vec{x}')$ lies in the span of A , we know that it is orthogonal to \vec{e} , so

$$\langle A(\vec{x} - \vec{x}'), \vec{e} \rangle = 0.$$

Substituting, we find that

$$\langle A(\vec{x} - \vec{x}'), A(\vec{x} - \vec{x}') \rangle < 0 \Rightarrow \|A(\vec{x} - \vec{x}')\| < 0.$$

However, we know that norms are always nonnegative, and so the left-hand-side of the above expression must be greater than or equal to 0. We have therefore reached a contradiction.

Thus, our original assumption that an \vec{x}' existed that was a better least squares solution than \vec{x} must have been false. So we have shown *if* an \vec{x} exists such that $\vec{b} - A\vec{x}$ is orthogonal to the span of A , that it must be the optimal solution to our least squares problem!

23.2.3 Least Squares

We know now that, *if* there exists an \vec{x} such that $\vec{e} = \vec{b} - A\vec{x}$ is orthogonal to *every vector* in the column space of A , that such a vector would be the optimum solution to our least squares problem and minimizes $\|\vec{e}\| = \|\vec{b} - A\vec{x}\|$.

Theorem 23.1: A vector \vec{e} is orthogonal to every vector in the column space of A if and only if it is orthogonal to each of the columns \vec{a}_i that form the basis of its column space.

Proof. Observe that if \vec{e} is orthogonal to every vector in the column space of A , then it is orthogonal to each of the \vec{a}_i , as each of the \vec{a}_i are in the column space of A .

Now, we will try to prove the converse. Consider an arbitrary vector $\vec{v} \in \text{span}(A)$. By definition, we know that there exist coefficients α_i such that we can express

$$\vec{v} = \alpha_1 \vec{a}_1 + \alpha_2 \vec{a}_2 + \dots + \alpha_m \vec{a}_m.$$

Now, consider our vector \vec{e} , that is orthogonal to each of the \vec{a}_i . In other words, we know that for any valid i ,

$$\langle \vec{e}, \vec{a}_i \rangle = 0.$$

We wish to show that \vec{e} is orthogonal to \vec{v} as well. To show this, it is natural to try evaluating $\langle \vec{e}, \vec{v} \rangle$, in the hope that this turns out to be 0. Doing so, substituting in our various definitions, and applying the distributive property of inner products, we see that

$$\begin{aligned} \langle \vec{e}, \vec{v} \rangle &= \langle \vec{e}, \alpha_1 \vec{a}_1 + \alpha_2 \vec{a}_2 + \dots + \alpha_m \vec{a}_m \rangle \\ &= \langle \vec{e}, \alpha_1 \vec{a}_1 \rangle + \langle \vec{e}, \alpha_2 \vec{a}_2 \rangle + \dots + \langle \vec{e}, \alpha_m \vec{a}_m \rangle \\ &= \alpha_1 \langle \vec{e}, \vec{a}_1 \rangle + \alpha_2 \langle \vec{e}, \vec{a}_2 \rangle + \dots + \alpha_m \langle \vec{e}, \vec{a}_m \rangle \\ &= \alpha_1 \cdot 0 + \alpha_2 \cdot 0 + \dots + \alpha_m \cdot 0 \\ &= 0, \end{aligned}$$

as desired! In other words, if our \vec{e} is orthogonal to all the basis vectors of a subspace, it is orthogonal to *every vector* in that subspace as well! \square

We know now that, for our error vector \vec{e} to be orthogonal to every vector in the column space of A , it suffices to say that it is orthogonal to each of the columns \vec{a}_i . Now, recall that the inner product of two vectors can be represented as a matrix multiplication, so

$$\langle \vec{e}, \vec{a}_i \rangle = 0 \iff \vec{a}_i^T \vec{e} = 0.$$

Since we have $\vec{a}_i^T \vec{e} = 0$ for all i , we can stack these scalar equations to form the single matrix equation

$$\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_m^T & - \end{bmatrix} \vec{e} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

But we have,

$$\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_m^T & - \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_m \\ | & | & & | \end{bmatrix}^T = A^T,$$

so we can simplify our equation to just

$$A^T \vec{e} = \vec{0}.$$

What do we do with this equation? We're interested in solving for \vec{x} , so it makes sense to substitute for

$\vec{e} = \vec{b} - A\vec{x}$, to obtain

$$\begin{aligned} A^T(\vec{b} - A\vec{x}) &= \vec{0} \\ \implies A^T A \vec{x} &= A^T \vec{b}. \end{aligned}$$

Now, observe that (unlike A) $A^T A$ is a square $m \times m$ matrix! Is it invertible? We will show below that it is, in the case where the columns of A are linearly independent. In this case $A^T A$ is a square matrix with only a trivial nullspace, we know that it has a unique inverse $(A^T A)^{-1}$. Pre-multiplying our earlier equation by this inverse (which we know exists), we obtain

$$\begin{aligned} A^T A \vec{x} &= A^T \vec{b} \\ \implies (A^T A)^{-1} A^T A \vec{x} &= (A^T A)^{-1} A^T \vec{b} \\ \implies \vec{x} &= (A^T A)^{-1} A^T \vec{b}. \end{aligned}$$

This is the general least squares solution for $A\vec{x} \approx \vec{b}$ when A has independent columns, and is our final result!

23.3 Is $A^T A$ invertible?

Let us figure out when exactly we can invert $A^T A$. First, we will prove a helper theorem.

Theorem 23.2: $\text{Null}(A^T A) = \text{Null}(A)$, even when A has a nontrivial nullspace.

Proof. To see this, consider an arbitrary $\vec{v} \in \text{Null}(A^T A)$. By definition, we have that

$$A^T A \vec{v} = \vec{0}.$$

Now, we will make the “magic” step¹ of pre-multiplying by \vec{v}^T , to obtain

$$\vec{v}^T A^T A \vec{v} = \vec{v}^T \vec{0} = 0.$$

Rearranging the left-hand-side of the above identity, we find that

$$(A\vec{v})^T (A\vec{v}) = 0 \implies \langle A\vec{v}, A\vec{v} \rangle = \|A\vec{v}\|^2 = 0.$$

Thus, it is clear that $A\vec{v} = \vec{0}$, so $\vec{v} \in \text{Null}(A)$. Thus, we have shown that $\text{Null}(A^T A) \subseteq \text{Null}(A)$. Now, consider an arbitrary vector $\vec{v}' \in \text{Null}(A)$. Pre-multiplying by A^T , we have that

$$\begin{aligned} A\vec{v}' &= \vec{0} \\ \implies A^T A \vec{v}' &= \vec{0}, \end{aligned}$$

so $\vec{v}' \in \text{Null}(A^T A)$.

¹As an aside, make sure to understand the intuition behind our “magic” step - we wanted to write the left-hand-side as an inner product, to obtain an equation involving the norm of $A\vec{v}$. This kind of pre-multiplication in order to get an inner product is fairly common in these sorts of proofs, and is a good thing to have in your “toolbox” of algebraic tricks.

Consequently, $\text{Null}(A) \subseteq \text{Null}(A^T A)$. Combining this with the earlier result, we have that

$$\text{Null}(A) = \text{Null}(A^T A),$$

as desired! □

We can now return to our least squares argument in the special case where A has linearly independent columns. Since all the \vec{a}_i are independent, it is a consequence that A has only a trivial null space. From our helper theorem, we can now see that $A^T A$ also only has a trivial nullspace!

23.4 Application of Least Squares

Gauss used this technique to predict where certain planets would be in their orbit. A scientist named Piazzi made 19 observations over the period of a month in regards to the orbit of Ceres. Gauss used some of these observations. He also knew the general shape of the orbit of planets due to Kepler's laws of planetary motion. Gauss set up equations like:

$$\alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y = 1$$

We can set up a matrix like:

$$\begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_n^2 & \dots & \dots & \dots & y_n \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \varepsilon \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix}$$

Here, the x 's and y 's are known: they are the coordinates of the measured positions of Ceres. The unknowns are $\alpha, \dots, \varepsilon$. We write the above equation with matrix/vector notation as follows

$$A\vec{v} = \vec{b}$$

where define

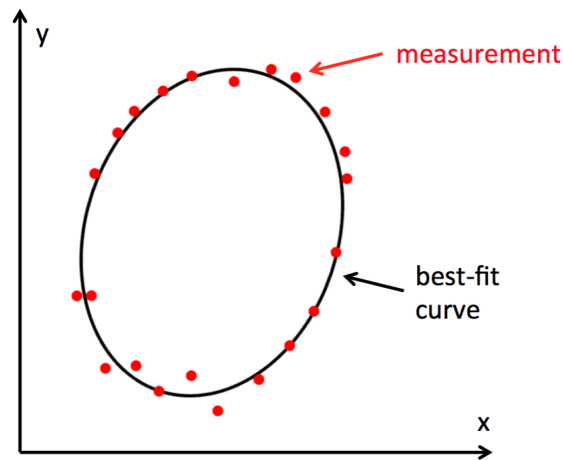
$$A = \begin{bmatrix} x_1^2 & x_1 y_1 & y_1^2 & x_1 & y_1 \\ \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ x_n^2 & \dots & \dots & \dots & y_n \end{bmatrix} \quad \vec{v} = \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \varepsilon \end{bmatrix} \quad \vec{b} = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ \vdots \\ 1 \end{bmatrix}$$

Now we can use the least squares formula to estimate the unknown coefficients in \vec{v} :

$$\vec{v} = (A^T A)^{-1} A^T \vec{b}$$

$$\begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \delta \\ \varepsilon \end{bmatrix} = (A^T A)^{-1} A^T \vec{b}$$

Once solved, we can now translate the coefficients, $\alpha, \dots, \varepsilon$, back into an ellipse using the original equation $\alpha x^2 + \beta xy + \gamma y^2 + \delta x + \varepsilon y = 1$. A possible result might look like this:



As we can see, the least squares method is useful for fitting noisy or approximate measurements to a curve, provided that we know the general shape of the curve. In addition, this example shows least squares is not limited to lines.

23.5 Practice Problems

These practice problems are also available in an interactive form on the course website.

1. True or False: Least squares is a method for solving an underdetermined system of linear equations.
2. Find the least squares solution to $\begin{bmatrix} 1 & 2 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \vec{x} = \begin{bmatrix} 2 \\ 0 \\ -8 \end{bmatrix}$.
3. True or False: Least squares always has a unique solution given by $\vec{x} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \vec{b}$.
4. True or False: We can use the least squares method to perform regression with a sine function, solving for $y = a \cdot \sin(bx)$. Make sure to also understand the reasoning why.
5. Let's say that we have a scenario where we have a set of data which includes information about a given set of patients. We have their height, weight, age, and white blood cell count. We are trying to create a predictor function for white blood cell count by solving an equation of the form $\mathbf{A}\vec{x} = \vec{b}$. What information should be in the \mathbf{A} matrix?
 - (a) The white blood cell counts.
 - (b) The height, weight, age, and white blood cell count of each patient.
 - (c) The unknown parameters $\alpha_1, \alpha_2, \alpha_3$.
 - (d) The height, weight, and age for each patient.

6. True or False: Let $\vec{x} = \text{proj}_{\text{Col}(\mathbf{A})} \vec{b}$ be the projection of \vec{b} onto the column space of a matrix \mathbf{A} . Then, $\mathbf{A}^T (\vec{b} - \vec{x}) = \vec{0}$.
7. True or False: The projection of a vector \vec{b} onto a set of vectors $\{\vec{a}_1, \vec{a}_2, \dots, \vec{a}_k\}$ is equal to $\frac{\langle \vec{b}, \vec{a}_1 \rangle}{\langle \vec{a}_1, \vec{a}_1 \rangle} \vec{a}_1 + \frac{\langle \vec{b}, \vec{a}_2 \rangle}{\langle \vec{a}_2, \vec{a}_2 \rangle} \vec{a}_2 + \dots + \frac{\langle \vec{b}, \vec{a}_k \rangle}{\langle \vec{a}_k, \vec{a}_k \rangle} \vec{a}_k$.
8. True or False: Given an arbitrary cost function, the error vector corresponding to the best approximation of a vector \vec{b} to the column space of \mathbf{A} is always orthogonal to $\text{Col}(\mathbf{A})$.
9. Find the best approximation \hat{x} to the system of equations $\begin{cases} a_1x = b_1 \\ a_2x = b_2 \end{cases}$ given the cost function $\text{cost}(x) = 2(b_1 - a_1\hat{x})^2 + (b_2 - a_2\hat{x})^2$.
- (a) $\frac{a_1b_1 + a_2b_2}{a_1^2 + a_2^2}$
- (b) $\frac{2a_1b_1 + a_2b_2}{a_1^2 + a_2^2}$
- (c) $\frac{2a_1b_1 + a_2b_2}{2a_1^2 + a_2^2}$
- (d) $\frac{a_1b_1 + 2a_2b_2}{a_1^2 + 2a_2^2}$