Key Question: How do we "learn" models from data, and make predictions?

Agenda

- Quick Review from Last Time
- Multilateration
- Projections
- Least Squares

Summary (from yesterday)

Help!

1. Identify which satellites are in the transmitted signal
2. Find the time delay/shift for each satellite signal
3. Utilize the shifts to find the distance to each satellite
4. Use trilateration to find unknown coordinates.

Yesterday, we found that in a 2D plane, using information from three satellites, our system of equations for trilateration was:

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2-\|\vec{s}_2\|^2+d_2^2-d_1^2$$

$$2(\vec{s}_1-\vec{s}_3)^T\vec{x} = \|\vec{s}_1\|^2-\|\vec{s}_3\|^2+d_3^2-d_1^2$$

However, here we assumed that the distances d_1, d_2, and d_3 were given. In reality, $d = \tau c$ where c is a constant representing the velocity of the transmitted signal, and τ is the time delay calculated after conducting cross-correlation.

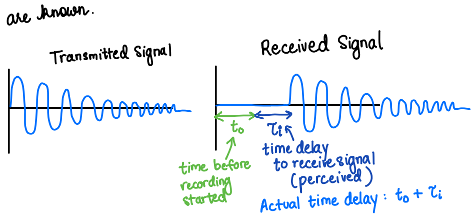
Given that $d_1=\tau_1 c$, $d_2=\tau_2 c$, and $d_3=\tau_3 c$, we can rewrite the above equations as:

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2-\|\vec{s}_2\|^2+\tau_2^2c^2-\tau_1^2c^2 = \|\vec{s}_1\|^2-\|\vec{s}_2\|^2+c^2(\tau_2^2-\tau_1^2)$$

$$2(\vec{s}_1-\vec{s}_3)^T\vec{x} = \|\vec{s}_1\|^2-\|\vec{s}_3\|^2+\tau_3^2c^2-\tau_1^2c^2 = \|\vec{s}_1\|^2-\|\vec{s}_3\|^2+c^2(\tau_3^2-\tau_1^2)$$

We designed this process in the last two lectures to help us identify our location on the surface of the Earth, but the way we found our time delays depended on the clock in our receiver and the satellites being perfectly in sync. We assumed that the receiver started recording at the same time ($t=0$) that the satellites were broadcasting. But how do we know to start recording exactly at that time $t=0$? We don't have an atomic clock on our cell phone receiver, unlike the precise atomic clocks on satellites.

What we do know is the relative delays between the signals transmitted by the satellites, because we know the satellites are in sync. This means that, in the above equations, if we assume, without loss of generality, that $\vec{s}_1$ is the satellite signal we received first, then the exact time delay τ_1 in reference to our receiver clock is unknown, but we do know that τ_2 and τ_3 arrived some units after the first signal. This means that $\Delta\tau_2=\tau_2-\tau_1$ and $\Delta\tau_3=\tau_3-\tau_1$ are known.



$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\tau_2^2-\tau_1^2)$$

$$2(\vec{s}_1-\vec{s}_3)^T\vec{x} = \|\vec{s}_1\|^2 - \|\vec{s}_3\|^2 + c^2(\tau_3^2-\tau_1^2)$$

Let's rewrite these equations with the change in our knowns and unknowns.

expanding a difference of two squares.

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\tau_2-\tau_1)(\tau_2+\tau_1)$$

trying to replace with $\Delta\tau_2 = \tau_2 - \tau_1$

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\tau_2-\tau_1)(\tau_2-\tau_1+2\tau_1)$$

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\Delta\tau_2)(\Delta\tau_2+2\underline{\tau_1})$$

unknown

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} - 2c^2\Delta\tau_2\tau_1 = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\Delta\tau_2)^2$$

$$2(\vec{s}_1-\vec{s}_3)^T\vec{x} - 2c^2\Delta\tau_3\tau_1 = \|\vec{s}_1\|^2 - \|\vec{s}_3\|^2 + c^2(\Delta\tau_3)^2 \quad \leftarrow \text{similarly, we find this equation.}$$

Now, we have three unknowns – x_1, x_2, and τ_1 – but only two equations!

What do we do? Pick another satellite to write an equation for!

Multilateration

$$2(\vec{s}_1-\vec{s}_2)^T\vec{x} - 2c^2\Delta\tau_2\tau_1 = \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\Delta\tau_2)^2$$

$$2(\vec{s}_1-\vec{s}_3)^T\vec{x} - 2c^2\Delta\tau_3\tau_1 = \|\vec{s}_1\|^2 - \|\vec{s}_3\|^2 + c^2(\Delta\tau_3)^2$$

$$2(\vec{s}_1-\vec{s}_4)^T\vec{x} - 2c^2\Delta\tau_4\tau_1 = \|\vec{s}_1\|^2 - \|\vec{s}_4\|^2 + c^2(\Delta\tau_4)^2 \quad \leftarrow \text{Our newest equation!}$$

Now, we can set up a matrix-vector equation.

$$2\begin{bmatrix} s_{11}-s_{21} & s_{12}-s_{22} & -c^2\Delta\tau_2 \\ s_{11}-s_{31} & s_{12}-s_{32} & -c^2\Delta\tau_3 \\ s_{11}-s_{41} & s_{12}-s_{42} & -c^2\Delta\tau_4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \tau_1 \end{bmatrix} = \begin{bmatrix} \|\vec{s}_1\|^2 - \|\vec{s}_2\|^2 + c^2(\Delta\tau_2)^2 \\ \|\vec{s}_1\|^2 - \|\vec{s}_3\|^2 + c^2(\Delta\tau_3)^2 \\ \|\vec{s}_1\|^2 - \|\vec{s}_4\|^2 + c^2(\Delta\tau_4)^2 \end{bmatrix}$$

Solve this using Gaussian Elimination!

Thus, to find our position in d-dimensional space when our receiver clock is not synchronized with the precise atomic clocks of the satellite, we need at least d+2 satellites.
With synchronized clocks, we needed d+1 satellites.

We can generate even more equations with information from satellites.

This means we will have more equations than unknowns!

$$\begin{bmatrix} \\ A \\ \\ \end{bmatrix}\begin{bmatrix} \vec{x} \end{bmatrix} = \begin{bmatrix} \\ \vec{b} \\ \\ \end{bmatrix}$$

Overdetermined System

When is there a solution to $A\vec{x} = \vec{b}$?

- When $\vec{b} \in$ span{columns of A}

$$a_{11}x_1 + a_{12}x_2 = b_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2$$
$$a_{31}x_1 + a_{32}x_2 = b_3$$

What if the equations are inconsistent due to noise (no $\vec{x}$ exists that satisfies all the equations exactly)?

- We have to find the closest approximate solution instead!

$$a_{11}x_1 + a_{12}x_2 = b_1 + e_1$$
$$a_{21}x_1 + a_{22}x_2 = b_2 + e_2$$
$$a_{31}x_1 + a_{32}x_2 = b_3 + e_3$$

So far...

We have:

- Measurements: $\vec{b}$
- Model: $A\vec{x} = \vec{b}$

But we also have a problem...

$A\vec{x} = \vec{b}$ has no solution!

What do we do?

We have to find an $\hat{x}$ such that $A\hat{x}$ is as close to $\vec{b}$ as we can get.

In other words, we can't find a perfect $\vec{x}$ to solve $A\vec{x} = \vec{b}$, so instead, we are trying to find $\hat{x}$ such that the length of error $\vec{e} = \vec{b} - A\hat{x}$ is minimized.

Let's consider a scalar problem in two dimensions first.

Our unknown value is x, and we are given two equations:

$$a_1 x = b_1$$
$$a_2 x = b_2$$

Our matrix-vector equation is thus:

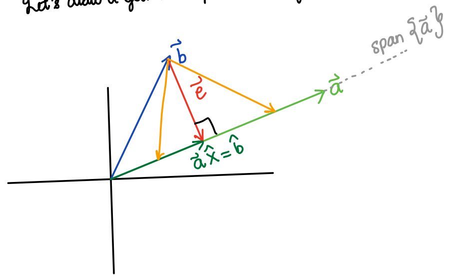
$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} x = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

We have more equations than unknowns, so this is an overdetermined system.

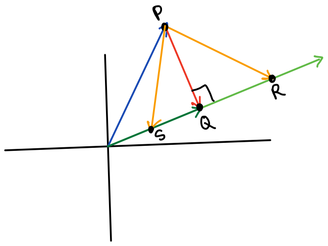
Our goal is to find $\hat{x}$ that has the smallest error:

$$\|\vec{e}\| = \|\vec{a}x - \vec{b}\| = \|\vec{b} - \vec{a}x\|$$

Let's draw a geometric representation of this:



For which $\hat{x}$ is $\vec{b}-\vec{a}\hat{x}$ the shortest?



Proof: Let's use the Pythagorean theorem.

$$\underset{\text{hypotenuse}}{(PR)^2} = (PQ)^2 + \underbrace{(QR)^2}_{\text{must be } >0 \text{ since it's a length.}}$$

$$\therefore \ (PR)^2 > (PQ)^2$$
$$PR > PQ$$

We can test this for any other potential $\vec{e}$, such as PS, but we find that dropping a perpendicular from $\vec{b}$ onto $\vec{a}$, such that $\vec{e}$ is orthogonal to $\vec{a}$, provides us with the shortest error vector.

Projection

The shortest distance between a point and a line is the orthogonal projection.

To find $\hat{x}$ in our 2D problem above, we need to find the orthogonal projection.

We know, from the Pythagorean theorem proof above, that $\vec{e} \perp \vec{a}$, and thus $\vec{e} \perp \hat{b}$ since we know $\hat{b}$ will be in the span of A.

Thus, $\vec{e}$ is orthogonal to both $\vec{a}$ and $\hat{b}$.

$$\vec{e} \perp \vec{a}, \ \vec{e} \perp \hat{b}$$
$$<\vec{e}, \vec{a}> = 0$$
$$\vec{e} = \vec{b} - A\hat{x} = \vec{b} - \hat{b}$$
$$<\vec{b} - \hat{b}, \vec{a}> = 0$$
$$<\vec{b}, \vec{a}> - <\hat{b}, \vec{a}> = 0$$
$$<\vec{b}, \vec{a}> = <\hat{b}, \vec{a}>$$
$$<\vec{b}, \vec{a}> = <\vec{a}\hat{x}, \vec{a}>$$
$$<\vec{b}, \vec{a}> = \hat{x} <\vec{a}, \vec{a}>$$
$$<\vec{b}, \vec{a}> = \hat{x} \|\vec{a}\|^2$$
$$\hat{x} = \frac{<\vec{b}, \vec{a}>}{\|\vec{a}\|^2}$$

Least Squares Solution!

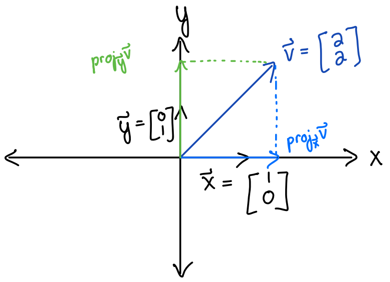
Thus,

$$\boxed{\hat{b} = \vec{a}\hat{x} = \frac{<\vec{b}, \vec{a}>}{\|\vec{a}\|^2}\vec{a}}$$

Given vectors $\vec{a}, \vec{b}$, we say that the orthogonal projection of $\vec{b}$ onto $\vec{a}$ is:

$$\text{proj}_{\vec{a}}(\vec{b}) = \frac{\vec{a}^T\vec{b}}{\|\vec{a}\|^2}\vec{a}$$

Example: Given a vector $\vec{v} = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$, what is the projection of $\vec{v}$ onto the x-axis? What is the projection of $\vec{v}$ onto the y-axis?



$$proj_{\vec{x}}(\vec{v}) = \frac{\vec{x}^T\vec{v}}{\|\vec{x}\|^2}\vec{x} = \frac{\begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix}}{1^2 + 0^2} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \frac{2}{1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$$

$$proj_{\vec{y}}(\vec{v}) = \frac{\vec{y}^T\vec{v}}{\|\vec{y}\|^2}\vec{y} = \frac{\begin{bmatrix} 0 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ 2 \end{bmatrix}}{0^2 + 1^2} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{2}{1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 2 \end{bmatrix}$$

The projection of a vector is like its shadow onto another vector, or the component of a vector in a particular direction.

Now, let's consider a problem in higher dimensions.

Our unknown values are x_1 and x_2, and we are given three equations, such that our matrix-vector equation is:

$$\underbrace{\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}_{\vec{x}} = \underbrace{\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}}_{\vec{b}}, \quad \text{columns } \vec{a}_1, \vec{a}_2$$

We have more equations than unknowns, so this is an overdetermined system.

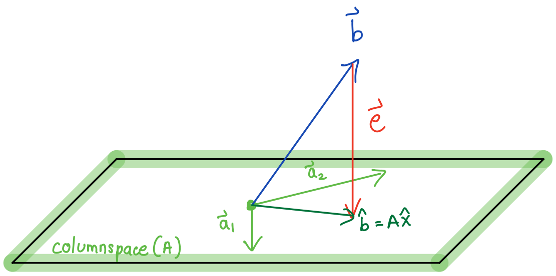
We are assuming there is noise in this system, making it inconsistent since $\vec{b}$ is not in the columnspace of A.

We are looking for $\hat{x}$ that is a solution with the smallest possible error.

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

Our solution will be an orthogonal projection onto the columnspace of A.

Let's draw a geometric representation of this:



We are trying to find:

$$\underset{\hat{x}}{\operatorname{argmin}} \|\vec{e}\| = \|A\hat{x} - \vec{b}\|$$

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

$\vec{e}$ must be orthogonal to the columnspace of A.

$\vec{e} \perp \vec{a}_i \quad \underset{\text{for all } i}{\underset{\uparrow}{\forall i}}$, so $\langle \vec{a}_i, \vec{e} \rangle = 0$

$$\langle \vec{a}_i, \vec{b} - \hat{b} \rangle = 0$$

$$\vec{a}_i^T (\vec{b} - \hat{b}) = 0$$

$$\begin{bmatrix} - & \vec{a}_1^T & - \\ - & \vec{a}_2^T & - \\ & \vdots & \\ - & \vec{a}_n^T & - \end{bmatrix} \begin{bmatrix} | \\ \vec{b} - \hat{b} \\ | \end{bmatrix} = \vec{0}$$

$$A^T(\vec{b} - A\hat{x}) = \vec{0}$$

$$A^T\vec{b} - A^TA\hat{x} = \vec{0}$$

$$A^T\vec{b} = \underbrace{A^TA}_{\text{if } A \text{ is full rank, then } A^TA \text{ is invertible.}}\hat{x}$$

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

$$\boxed{\hat{b} = A(A^TA)^{-1}A^T\vec{b}}$$

Least Squares

An important method used to approximate the solutions of overdetermined systems.

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

Does this match our 2D least squares solution?

$$(A^TA)^{-1}A^T\vec{b} = \left([a_1 \; a_2] \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \right)^{-1} \vec{a}^T\vec{b}$$

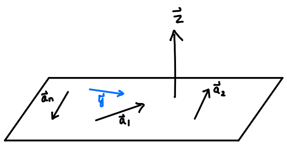
$$= (a_1^2 + a_2^2)^{-1} \vec{a}^T \vec{b}$$

$$= \frac{\vec{a}^T\vec{b}}{\|\vec{a}\|^2}$$ ✓ yes!

How do we know for a fact that $\vec{e}$ is orthogonal to the columnspace of A if it is orthogonal to the columns of A?

Theorem: Consider matrix A and $\vec{y} \in$ colspace(A). If $\exists \vec{z}$ such that $\langle \vec{z}, \vec{a}_i \rangle = 0$, then $\langle \vec{z}, \vec{y} \rangle = 0$

Proof:



We know: A, $\vec{y} \in$ colspace(A), so $\vec{y} = c_1\vec{a}_1 + c_2\vec{a}_2 + \dots + c_n\vec{a}_n$

$$\left.\begin{aligned} &\langle \vec{z}, \vec{a}_1 \rangle = 0 \\ &\langle \vec{z}, \vec{a}_2 \rangle = 0 \\ &\quad \vdots \\ &\langle \vec{z}, \vec{a}_n \rangle = 0 \end{aligned}\right\} \vec{z} \text{ is orthogonal to all vectors in colspace(A)}$$

We want to show: $\langle \vec{z}, \vec{y} \rangle = 0$

$$\begin{aligned} \langle \vec{z}, \vec{y} \rangle &= \langle \vec{z}, c_1\vec{a}_1 + c_2\vec{a}_2 + \dots + c_n\vec{a}_n \rangle \\ &= \langle \vec{z}, c_1\vec{a}_1 \rangle + \langle \vec{z}, c_2\vec{a}_2 \rangle + \dots + \langle \vec{z}, c_n\vec{a}_n \rangle \\ &= c_1\langle \vec{z}, \vec{a}_1 \rangle + c_2\langle \vec{z}, \vec{a}_2 \rangle + \dots + c_n\langle \vec{z}, \vec{a}_n \rangle \\ &= c_1(0) + c_2(0) + \dots + c_n(0) \\ &= 0 \checkmark \end{aligned}$$

Thus, we showed that if our error vector $\vec{e}$ is orthogonal the columns of A, it is also orthogonal to the columnspace of A where $A\hat{x} = \hat{b}$ lies.

So, when we have an overdetermined system, we now have a closed form solution using least squares!

When might we have issues with this solution?

If A^TA is not invertible!

When is A^TA invertible? When it is square, has linearly independent columns, and a trivial nullspace.

A is not invertible, since it is not square, but when A has linearly independent columns, it has a trivial nullspace.

We can show that Null(A^TA) = Null(A):

Proof: We need to show both directions – that Null(A^TA) $\subseteq$ Null(A) and Null(A) $\subseteq$ Null(A^TA).

Let $\vec{u} \in$ Null(A^TA).

$A^TA\vec{u} = \vec{0}$

Multiplying both sides by $\vec{u}^T$:

$\vec{u}^TA^TA\vec{u} = \vec{u}^T\vec{0}$

$(A\vec{u})^TA\vec{u} = 0$

$\|A\vec{u}\|^2 = 0$

For the norm of a vector to be 0, the vector must be the zero vector.

Thus, $A\vec{u} = 0$, so $\vec{u} \in$ Null(A).

$\therefore$ Null(A^TA) $\subseteq$ Null(A).

Now, let $\vec{v} \in$ Null(A)

$A\vec{v} = \vec{0}$

Multiply both sides by A^T.

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

$A^TA\vec{v} = \vec{0}$, so $\vec{v} \in$ Null(A^TA).

$\therefore$ Null(A) $\subseteq$ Null(A^TA)

Since both directions are true, we have found that Null(A^TA) = Null(A).

This means that when A has linearly independent columns, A^TA will be invertible.

Next time: examples and applications of least squares.