EECS 16A Designing Information Devices and Systems I
Spring 2019 Discussion 12A

1. Mechanical Projection

In $\mathbb{R}^n$, the projection of vector $\vec{a}$ onto vector $\vec{b}$ is defined as:

$$\text{proj}_{\vec{b}}(\vec{a}) = \frac{\left\langle \vec{a}, \vec{b} \right\rangle}{\left\| \vec{b} \right\|^2} \vec{b} = \frac{\left\langle \vec{a}, \vec{b} \right\rangle}{\left\| \vec{b} \right\|} \hat{b}$$

where $\hat{b}$ is the normalized $\vec{b}$, i.e., a unit vector with the same direction as $\vec{b}$.

(a) Project $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ – that is, onto the x-axis. Graph these two vectors and the projection.

Answer: $\begin{bmatrix} 5 \\ 0 \end{bmatrix}$

(b) Project $\begin{bmatrix} 5 \\ 2 \end{bmatrix}$ onto $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ – that is, onto the y-axis. Graph these two vectors and the projection.

Answer: $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$

(c) Project $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$ onto $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$. Graph these two vectors and the projection.

Answer: $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$

(d) Project $\begin{bmatrix} 4 \\ -2 \end{bmatrix}$ onto $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. Graph these two vectors and the projection.

Answer: $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

(e) Project $\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$ onto the span of the vectors $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right\}$ – that is, onto the x-y plane in $\mathbb{R}^3$.

Answer: $\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$

(f) What is the geometric/physical interpretation of projection? Justify using the previous parts.

Answer:

Geometrically, the orthogonal projection yields the closest vector in the projection space, such that the difference between the original vector and the projected vector is orthogonal to the projected vector. That is, the projection is the best approximation for the original vector within the subspace contraint.

UCB EECS 16A, Spring 2019, Discussion 12A, All Rights Reserved. This may not be publicly shared without explicit permission.

(g) For part $(a)(b)$ and part $(c)(d)$, we looked at two different projections for one vector. For those cases, using only the projected vector $\text{proj}_{\vec{b}}(\vec{a})$ and the vector $\vec{b}$ we projected onto, do we have enough information to reconstruct the original vector $\vec{a}$?

Answer:

Yes in these cases. The vectors that we projected onto were linearly independent, so we have enough information to reconstruct.

(h) Given information about n projections of a vector in $\mathbb{R}^n$, when do we have enough information to reconstruct the original vector? **Answer:**

As stated above, we need those n projections to be on n linearly independent vectors. If these vectors that we are projecting onto are orthogonal, we don't even need to know the vectors; we only need the projections. If they are not orthogonal, we will need the vectors used for projection.

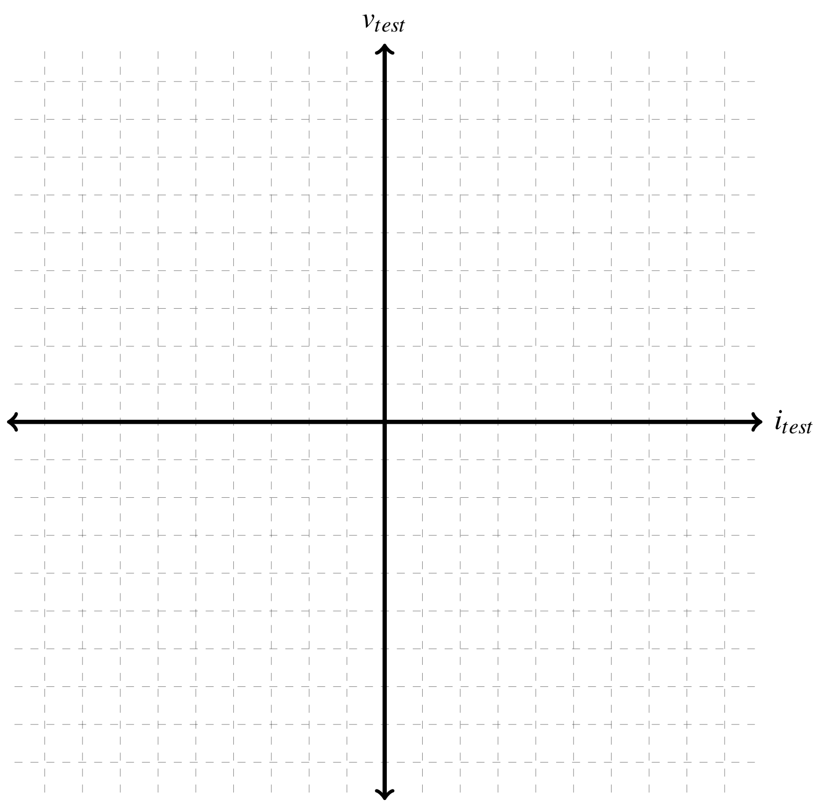
All Rights Reserved. This may not be publicly shared without explicit permission.

2. Ohm's Law With Noise

We are trying to measure the resistance of a black box. We apply various i_{test} currents and measure the ouput voltage v_{test}. Sometimes, we are quite fortunate to get nice numbers. Oftentimes, our measurement tools are a little bit noisy, and the values we get out of them are not accurate. However, if the noise is completely random, then the effect of it can be averaged out over many samples. So we repeat our test many times:

Test	i_{test} (mA)	v_{test} (V)
1	10	21
2	3	7
3	-1	-2
4	5	8
5	-8	-15
6	-5	-11

(a) Plot the measured voltage as a function of the current.



Answer:
Notice that these points *do not* lie on a line!

(b) Suppose we stack the currents and voltages to get $\vec{I} = \begin{bmatrix} 10 \\ 3 \\ -1 \\ 5 \\ -8 \\ -5 \end{bmatrix}$ and $\vec{V} = \begin{bmatrix} 21 \\ 7 \\ -2 \\ 8 \\ -15 \\ -11 \end{bmatrix}$. Is there a unique solution for R? What conditions must $\vec{I}$ and $\vec{V}$ satisfy in order for us to solve for R uniquely?

UCB EECS 16A, Spring 2019, Discussion 12A, All Rights Reserved. This may not be publicly shared without explicit permission.

Answer:

We cannot find the unique solution for R because $\vec{V}$ is not a scalar multiple of $\vec{I}$. In general, we need $\vec{V}$ to be a scalar multiple of $\vec{I}$ to be able to solve for R exactly (another linear algebraic way of saying this is that $\vec{V}$ is in the span of $\vec{I}$).

We know that the *physical* reason we are not able to solve for R is that we have imperfect observations of the voltage across the terminals, $\vec{V}$. Therefore, now that we know we cannot solve for R directly, a very pertinent goal would be to find a value of R that *approximates* the relationship between $\vec{I}$ and $\vec{V}$ as closely as possible.

Let's move on and see how we do this.

(c) Ideally, we would like to find R such that $\vec{V} = \vec{I}R$. If we cannot do this, we'd like to find a value of R that is the *best* solution possible, in the sense that $\vec{I}R$ is as "close" to $\vec{V}$ as possible. We are defining the sum of squared errors as a **cost function**. In this case the cost function for any value of R quantifies the difference between each component of $\vec{V}$ (i.e. v_j) and each component of $\vec{I}R$ (i.e. i_jR) and sum up the squares of these "differences" as follows:

$$\text{cost}(R) = \sum_{j=1}^{6} (v_j - i_jR)^2$$

Do you think this is a good cost function? Why or why not?

Answer:

For each point (i_j, v_j), we want $|v_j - i_jR|$ to be as small as possible. We can call this term the individual error term for this point.

One way of looking at the aggregate "error" in our fit is to add up the squares of the individual errors, so that all errors add up. This is precisely what we've done in the cost function. If we did not square the differences, then a positive difference and a negative difference would cancel each other out.

(d) Show that you can also express the above cost function in vector form, that is,

$$\text{cost}(R) = \left\langle (\vec{V} - \vec{I}R), (\vec{V} - \vec{I}R) \right\rangle$$

Hint: $\langle \vec{a}, \vec{b} \rangle = \vec{a}^T \vec{b} = \sum_i a_i b_i$

Answer:

Let's define the error vector as

$$\vec{e} = \vec{V} - \vec{I}R.$$

Then, we observe that $e_j = v_j - i_jR$.

Therefore,

$$\begin{aligned}
\text{cost}(R) &= \sum_{j=1}^{6} (v_j - i_jR)^2 \\
&= \sum_{j=1}^{6} e_j^2 \\
&= \|\vec{e}\|_2^2 \\
&= \langle \vec{e}, \vec{e} \rangle \\
&= \left\langle (\vec{V} - \vec{I}R), (\vec{V} - \vec{I}R) \right\rangle
\end{aligned}$$

(e) Find $\hat{R}$, which is defined as the optimal value of R that minimizes $\text{cost}(R)$.

Hint: Use calculus. The optimal $\hat{R}$ makes $\frac{d\text{cost}(\hat{R})}{dR} = 0$

Answer:

First, note that

$$\frac{d\text{cost}(R)}{dR} = -2\sum_{j=1}^{6} i_j(v_j - i_j R)$$

For $R = \hat{R}$, we will have $\frac{d\text{cost}(R)}{dR} = 0$. This means that

$$-2\sum_{j=1}^{6} i_j(v_j - i_j\hat{R}) = 0,$$

which will ultimately give us

$$\hat{R} = \frac{\sum_{j=1}^{6} i_j v_j}{\sum_{j=1}^{6} i_j^2} = \frac{\left\langle \vec{I}, \vec{V} \right\rangle}{\|\vec{I}\|^2}$$

In our particular example, $\left\langle \vec{I}, \vec{V} \right\rangle = 448$ and $\|\vec{I}\|^2 = 224$. Therefore, we will get $\hat{R} = 2\,\text{k}\Omega$.

Using the equation for least squares estimate with $A = \begin{bmatrix} 10 \\ 3 \\ -1 \\ 5 \\ -8 \\ -5 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 21 \\ 7 \\ -2 \\ 8 \\ -15 \\ -11 \end{bmatrix}$, we would have:

$$\hat{R} = (A^T A)^{-1} A^T \vec{b}$$
$$\hat{R} = \frac{\left\langle \vec{I}, \vec{V} \right\rangle}{\left\langle \vec{I}, \vec{I} \right\rangle}$$
$$\hat{R} = \frac{\left\langle \vec{I}, \vec{V} \right\rangle}{\|\vec{I}\|^2},$$

which gives us the same expression as before!

(f) On your original IV plot, also plot the line $v_{test} = \hat{R} i_{test}$. Can you visually see why this line "fits" the data well? How well would we have done if we had guessed $R = 3\,\text{k}\Omega$? What about $R = 1\,\text{k}\Omega$? Calculate the cost functions for each of these choices of R to validate your answer.

Answer:

When $\hat{R} = 2\,\text{k}\Omega$, we have

$$\begin{aligned} \text{cost}(2k) &= (21 - 2\cdot 10)^2 + (7 - 2\cdot 3)^2 + (-2 - 2\cdot(-1))^2 + \\ &\quad (8 - 2\cdot 5)^2 + (-15 - 2\cdot(-8))^2 + (-11 - 2\cdot(-5))^2 \\ &= 8. \end{aligned}$$

When $\hat{R} = 3\,\mathrm{k\Omega}$, we have

$$\begin{aligned}\text{cost}(3k) &= (21-3\cdot 10)^2 + (7-3\cdot 3)^2 + (-2-3\cdot(-1))^2 + \\ &\quad (8-3\cdot 5)^2 + (-15-3\cdot(-8))^2 + (-11-3\cdot(-5))^2 \\ &= 232.\end{aligned}$$

When $\hat{R} = 1\,\mathrm{k\Omega}$, we have

$$\begin{aligned}\text{cost}(1k) &= (21-1\cdot 10)^2 + (7-1\cdot 3)^2 + (-2-1\cdot(-1))^2 + \\ &\quad (8-1\cdot 5)^2 + (-15-1\cdot(-8))^2 + (-11-1\cdot(-5))^2 \\ &= 232.\end{aligned}$$

(g) Now, suppose that we add a new data point: $i_7 = 2\,\mathrm{mA}$, $v_7 = 4\,\mathrm{V}$. Will $\hat{R}$ increase, decrease, or remain the same? Why? What does that say about the line $v_{test} = \hat{R} i_{test}$?

Answer:

We can qualitatively see that $\hat{R}$ will remain 2. This is because we already obtained $\hat{R}$ to fit our previous data in the best way. Now, you should notice that this new piece of data (i_7, v_7) also lies exactly on the line $v_{test} = \hat{R} i_{test}$! Therefore, you have no reason to change $\hat{R}$. It is the best fit for the old data and will fit the new data anyway.

(h) Let's add another data point: $i_8 = 4\,\mathrm{mA}, v_8 = 11\,\mathrm{V}$. Will $\hat{R}$ increase, decrease, or remain the same? Why? What does that say about the line $v_{test} = \hat{R} i_{test}$?

Answer:

We can qualitatively see that $\hat{R}$ should be something greater than or equal to 2. This is because you have already obtained $\hat{R}$ to fit your previous data in the best way. Now, you notice that this new piece of data (i_8, v_8) also lies *above* the line $v_{test} = \hat{R} i_{test}$! Therefore, if you decreased $\hat{R}$, it would be a worse fit for the old data and the new data. You would increase $\hat{R}$ to find a better fit.

(i) Now your mischievous friend has hidden the black box. You want to predict what output voltage across the terminals if you applied 5.5 mA through the black box. What would your best guess be?

Answer:

Hopefully, by now, it makes sense to the class that you will estimate $\hat{V} = 5.5\,\mathrm{mA} \cdot \hat{R} = 5.5\,\mathrm{mA} \cdot 2\,\mathrm{k\Omega} = 11\,\mathrm{V}$. This is an example of estimation from machine learning! You have *learned* what is going on inside the black box, that is, $\hat{R}$, by making observations of $\vec{I}$ and $\vec{V}$. Now, you are using what you have learned, $\hat{R}$, to estimate $\hat{V}$ for new values of I.

UCB EECS 16A, Spring 2019, Discussion 12A, All Rights Reserved. This may not be publicly shared without explicit permission.