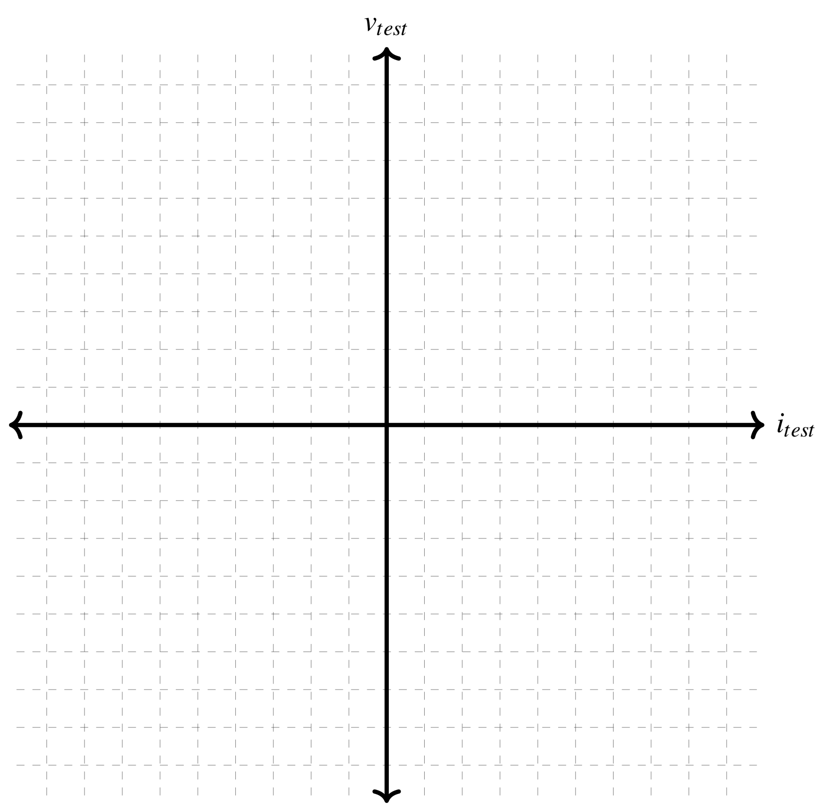
EECS 16A Designing Information Devices and Systems I
Summer 2020 Discussion 7A

1. Ohm's Law With Noise

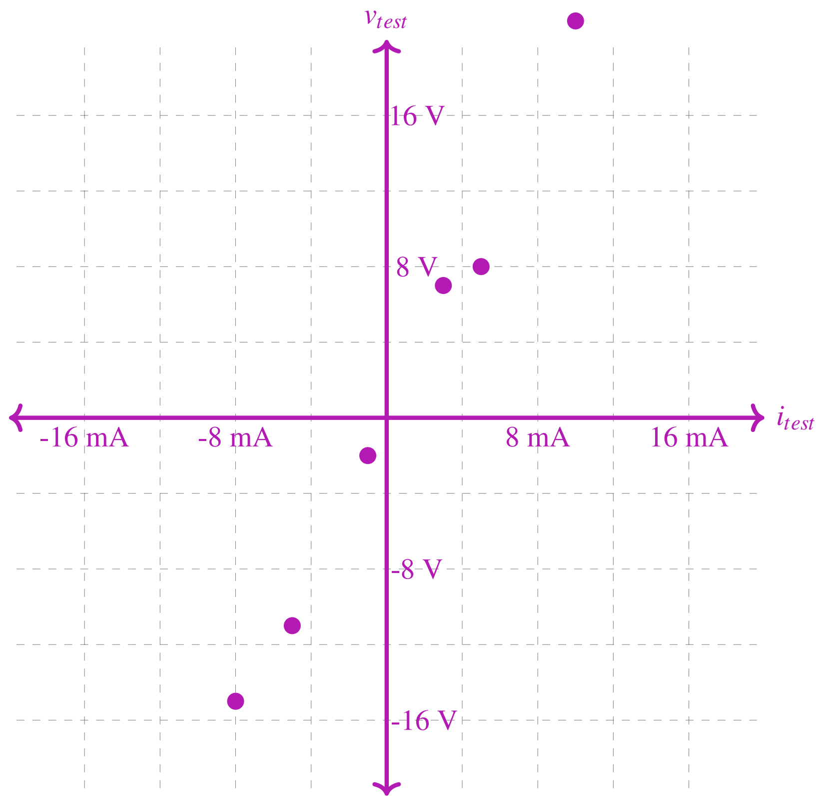
We are trying to measure the resistance of a black box. We apply various i_{test} currents and measure the ouput voltage v_{test}. Sometimes, we are quite fortunate to get nice numbers. Oftentimes, our measurement tools are a little bit noisy, and the values we get out of them are not accurate. However, 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?

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:

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

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

Answer:

For each point (i_j, v_j), we want $|v_j - i_j R|$ 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 $e_j = v_j - i_j R$.

Therefore,

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

(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 $\hat{R} = 2\,\text{k}\Omega$.

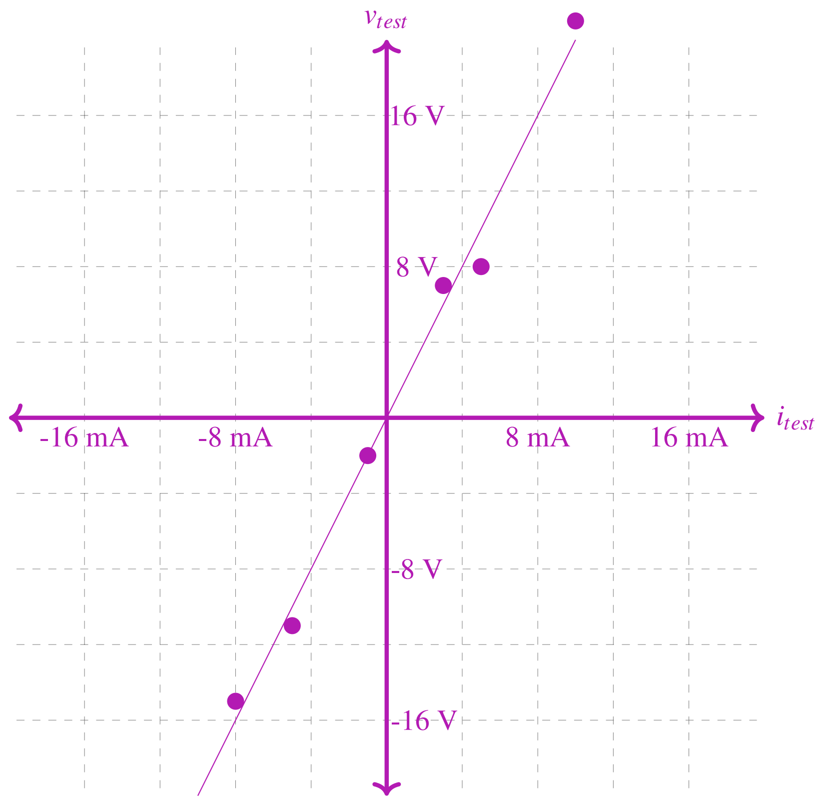
Using the equation for least squares estimate with $A = \begin{bmatrix} 10 \\ 3 \\ -1 \\ 5 \\ -8 \\ -5 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 21 \\ 7 \\ -2 \\ 8 \\ -15 \\ -11 \end{bmatrix}$, we would have:

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

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 $\hat{R} = 3\,\text{k}\Omega$, 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 $\hat{R} = 1\,\text{k}\Omega$, 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: $i_7 = 2\,\text{mA}$, $v_7 = 4\,\text{V}$. 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 kΩ. 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.

2. Orthonormal Matrices and Projections

An orthonormal matrix, $\mathbf{A}$, is a matrix whose columns, $\vec{a}_i$, are:

- Orthogonal (ie. $\langle \vec{a}_i, \vec{a}_j \rangle = 0$ when $i \neq j$)
- Normalized (ie. vectors with length equal to 1, $\|\vec{a}_i\| = 1$). This implies that $\|\vec{a}_i\|^2 = \langle \vec{a}_i, \vec{a}_i \rangle = 1$.

(a) Suppose that the matrix $\mathbf{A} \in \mathbb{R}^{N \times M}$ has linearly independent columns. The vector $\vec{y}$ in $\mathbb{R}^N$ is not in the subspace spanned by the columns of $\mathbf{A}$. What is the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$?

Answer: When finding a projection onto a subspace, we're trying to find the "closest" vector in that subspace. This can be found by first finding $\vec{x}$ that minimizes $\|\vec{y} - \mathbf{A}\vec{x}\|$. From least squares, we know that $\vec{\hat{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$. The projection of $\vec{y}$ onto the columns of $\mathbf{A}$ is then $\vec{\hat{y}} = \mathbf{A}\vec{\hat{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}$.

(b) Show if $\mathbf{A} \in \mathbb{R}^{N \times N}$ is an orthonormal matrix then the columns, $\vec{a}_i$, form a basis for $\mathbb{R}^N$.

Answer:

We want to show that the columns of $\mathbf{A}$ form a basis for $\mathbb{R}^N$. To show that the columns form a basis for $\mathbb{R}^N$ we need to show two things:

- The columns must form a set of N linearly independent vectors.
- Any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the vectors in the set.

We already know we have N vectors, so first we will show they are linearly independent. We shall do this by showing that $\mathbf{A}\vec{\beta} = \vec{0}$ implies that $\vec{\beta}$ can be only $\vec{0}$.

$$\mathbf{A}\vec{\beta} = \vec{0} \tag{1}$$
$$\beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N = \vec{0} \tag{2}$$

Then to exploit the properties of orthogonal vectors, we consider taking the inner product of each side of the above equation with $\vec{a}_i$.

$$\langle \vec{a}_i, \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \rangle = \langle \vec{a}_i, \vec{0} \rangle = 0 \tag{3}$$

Now we apply the distributive property of the inner product and the definition of orthonormal vectors,

$$\langle \vec{a}_i, \beta_1\vec{a}_1 \rangle + \ldots + \langle \vec{a}_i, \beta_i\vec{a}_i \rangle + \ldots + \langle \vec{a}_i, \beta_N\vec{a}_N \rangle = 0 \tag{4}$$
$$0 + \ldots + \beta_i\langle \vec{a}_i, \vec{a}_i \rangle + \ldots + 0 = 0 \tag{5}$$
$$0 + \ldots + \beta_i\vec{a}_i^T\vec{a}_i + \ldots + 0 = 0 \tag{6}$$

Because $\vec{a}_i^T\vec{a}_i = 1$, $\beta_i = 0$ for the equation to hold. Then, since this is true for all i from 1 to N, all the elements of the vector beta must be zero ($\vec{\beta} = \vec{0}$). Because $\vec{x} = \vec{0}$ implies $\vec{\beta} = \vec{0}$, the columns of $\mathbf{A}$ are linearly independent.

Now, we will show that any vector $\vec{x} \in \mathbb{R}^N$ can be represented as a linear combination of the columns of $\mathbf{A}$.

$$\vec{x} = \mathbf{A}\vec{\beta} = \beta_1\vec{a}_1 + \ldots + \beta_N\vec{a}_N \tag{7}$$

Because we know that the N columns of $\mathbf{A}$ are linearly independent, then there exists $\mathbf{A}^{-1}$. Applying the inverse to the equation above,

$$\mathbf{A}^{-1}\mathbf{A}\vec{\beta} = \mathbf{A}^{-1}\vec{x} \tag{8}$$

$$\vec{\beta} = \mathbf{A}^{-1}\vec{x}, \tag{9}$$

we find that there exists a unqiue β that allow us to represent any $\vec{x}$ as a linear combination of the columns of $\mathbf{A}$.

(c) When $\mathbf{A} \in \mathbb{R}^{N\times M}$ and $N \geq M$ (i.e. tall matrices), show that if the matrix is orthonormal, then $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

Answer: Want to show $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{M\times M}$.

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} \vec{a}_1^T\vec{a}_1 & \vec{a}_1^T\vec{a}_2 & \dots & \vec{a}_1^T\vec{a}_n \\ \vec{a}_2^T\vec{a}_1 & \vec{a}_2^T\vec{a}_2 & \dots & \vec{a}_2^T\vec{a}_n \\ \vdots & \vdots & & \vdots \end{bmatrix} = \mathbf{I}_{M\times M} \tag{10}$$

When $\vec{a}_i^T\vec{a}_i = \|\vec{a}_i\|^2 = 1$ and when $i \neq j$, $\vec{a}_i^T\vec{a}_j = 0$ because the column vectors are orthogonal.

(d) Again, suppose $\mathbf{A} \in \mathbb{R}^{N\times M}$ where $N \geq M$ is an orthonormal matrix. Show that the projection of $\vec{y}$ onto the subspace spanned by the columns of $\mathbf{A}$ is now $\mathbf{A}\mathbf{A}^T\vec{y}$.

Answer:

Starting with the result from part (a),

$$\mathbf{A}\hat{\vec{x}} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y}, \tag{11}$$

we can apply the result from part (c),

$$\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \mathbf{A}\mathbf{I}\mathbf{A}^T\vec{y} \tag{12}$$

$$= \mathbf{A}\mathbf{A}^T\vec{y} \tag{13}$$

(e) Given $\mathbf{A} \in \mathbb{R}^{N\times M} = \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and the columns of $\mathbf{A}$ are orthonormal, find the least squares solution to $\mathbf{A}\hat{\vec{x}} = \vec{y}$ where $\vec{y} = \begin{bmatrix} 5 & 12 & 7 & 8 \end{bmatrix}^T$.

Answer:

Method 1:

Since the columns of $\mathbf{A}$ are orthonormal, from part (d) we know that

$\hat{\vec{x}} = \mathbf{A}^T\vec{y} = \begin{bmatrix} \langle \vec{a}_1, \vec{y} \rangle \\ \langle \vec{a}_2, \vec{y} \rangle \\ \langle \vec{a}_3, \vec{y} \rangle \end{bmatrix}$.

Note that this is equivalent to projecting $\vec{y}$ onto each column of $\mathbf{A}$:

$\hat{x_1} = \frac{\langle \vec{a}_1, \vec{y} \rangle}{||\vec{a}_1||^2} = \langle \vec{a}_1, \vec{y} \rangle = 8$

$\hat{x_2} = \frac{\langle \vec{a}_2, \vec{y} \rangle}{||\vec{a}_2||^2} = \langle \vec{a}_2, \vec{y} \rangle = 7$

$\hat{x_3} = \frac{\langle \vec{a}_3, \vec{y} \rangle}{||\vec{a}_3||^2} = \langle \vec{a}_3, \vec{y} \rangle = \frac{17\sqrt{2}}{2}$

Method 2 (Alternatively you can use the least squares formula):

$$\hat{\vec{x}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\vec{y} = \left(\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 0 & \frac{\sqrt{2}}{2} \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} & 0 & 0 \end{bmatrix} \begin{bmatrix} 5 \\ 12 \\ 7 \\ 8 \end{bmatrix} = \begin{bmatrix} 8 \\ 7 \\ \frac{17\sqrt{2}}{2} \end{bmatrix}$$

