Module 1 | Lesson 1 (Part 1)

The Squared Error Criterion and the Method of Least Squares

Estimating Resistance

Measurement	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Let x be the resistance. Assume it is a **constant**, but **unknown**.

We make measurements, y, of the resistance.

We model our measurements as corrupted by noise v.

$$\boxed{y = x + v}$$

Estimating Resistance

Measurement	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Measurement Model

'Actual' resistance *Measurement noise*

$$y_1 = x + v_1$$

$$y_2 = x + v_2$$

$$y_3 = x + v_3$$

$$y_4 = x + v_4$$

Estimating Resistance

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

Measurement Model

$$y_1 = x + v_1$$

$$y_2 = x + v_2$$

$$y_3 = x + v_3$$

$$y_4 = x + v_4$$

Squared Error

$$e_1^2 = (y_1 - x)^2$$

$$e_2^2 = (y_2 - x)^2$$

$$e_3^2 = (y_3 - x)^2$$

$$e_4^2 = (y_4 - x)^2$$

The squared error *criterion*:

$$\hat{x}_{\text{LS}} = \text{argmin}_{\,x} \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\text{LS}}(x)$$

squared error cost func.
or loss func.

The 'best' estimate of resistance is the one that minimizes the sum of squared errors

Minimizing the Squared Error Criterion

$$\boxed{\hat{x}_{\mathrm{LS}} = \operatorname{argmin}_x \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\mathrm{LS}}(x)}$$

Let's re-write our criterion using vectors:

This matrix is called the 'Jacobian'

$$\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} = \mathbf{y} - \mathbf{H}x$$

$$= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x$$

Minimizing the Squared Error Criterion

$$\boxed{\hat{x}_{\text{LS}} = \text{argmin}_{\,x} \left(e_1^2 + e_2^2 + e_3^2 + e_4^2\right) = \mathscr{L}_{\text{LS}}(x)}$$

Now, we can express our criterion as follows,

$$\begin{aligned}
\mathscr{L}_{\text{LS}}(x) = e_1^2 + e_2^2 + e_3^2 + e_4^2 &= \mathbf{e}^T\mathbf{e} \\
&= (\mathbf{y} - \mathbf{H}x)^T(\mathbf{y} - \mathbf{H}x) \\
&= \mathbf{y}^T\mathbf{y} - x^T\mathbf{H}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}x + x^T\mathbf{H}^T\mathbf{H}x
\end{aligned}$$

Minimizing the Squared Error Criterion

$$\boxed{\mathscr{L}(x) = \mathbf{e}^T\mathbf{e} = \mathbf{y}^T\mathbf{y} - x^T\mathbf{H}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}x + x^T\mathbf{H}^T\mathbf{H}x}$$

To minimize this, we can compute the partial derivative with respect to our parameter, set to 0, and solve for an extremum:

$$\left.\frac{\partial \mathscr{L}}{\partial x}\right|_{x=\hat{x}} = -\mathbf{y}^T\mathbf{H} - \mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0$$

$$-2\mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0$$

Re-arranging, we arrive at:

$$\hat{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$$

x-hat minimizes our squared error criterion!

Minimizing the Squared Error Criterion

Careful! We will only be able to solve for $\hat{x}$ if $(\mathbf{H}^T\mathbf{H})^{-1}$ exists.

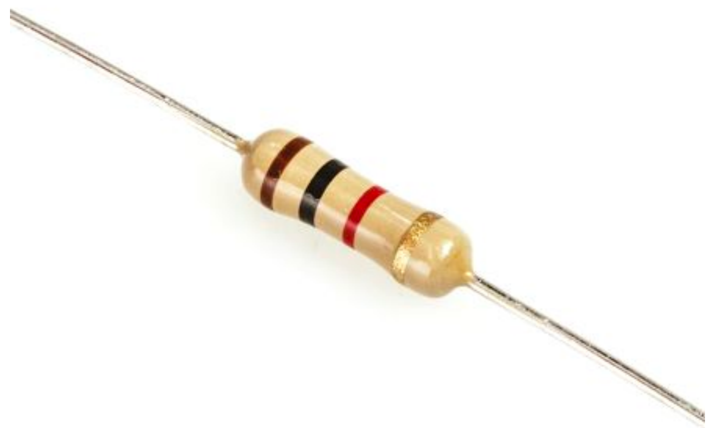
If we have m measurements, and n unknown parameters, then:

$$\mathbf{H} \in \mathbb{R}^{m \times n} \qquad \mathbf{H}^T\mathbf{H} \in \mathbb{R}^{n \times n}$$

This means that $(\mathbf{H}^T\mathbf{H})^{-1}$ exists only if there are at least as many measurements as there are unknown parameters:

$$m \geq n$$

Minimizing the Squared Error Criterion



Returning to our problem, we see that:

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996

$$\mathbf{y} = \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \qquad \mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\hat{x}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$$

$$= \left([1111] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} [1111] \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} = \frac{1}{4}(1068 + 988 + 1002 + 996) = 1013.5 \quad \text{Ohms}$$

The least squares solution is just the mean of our measurements!

Method of Least Squares | Assumptions

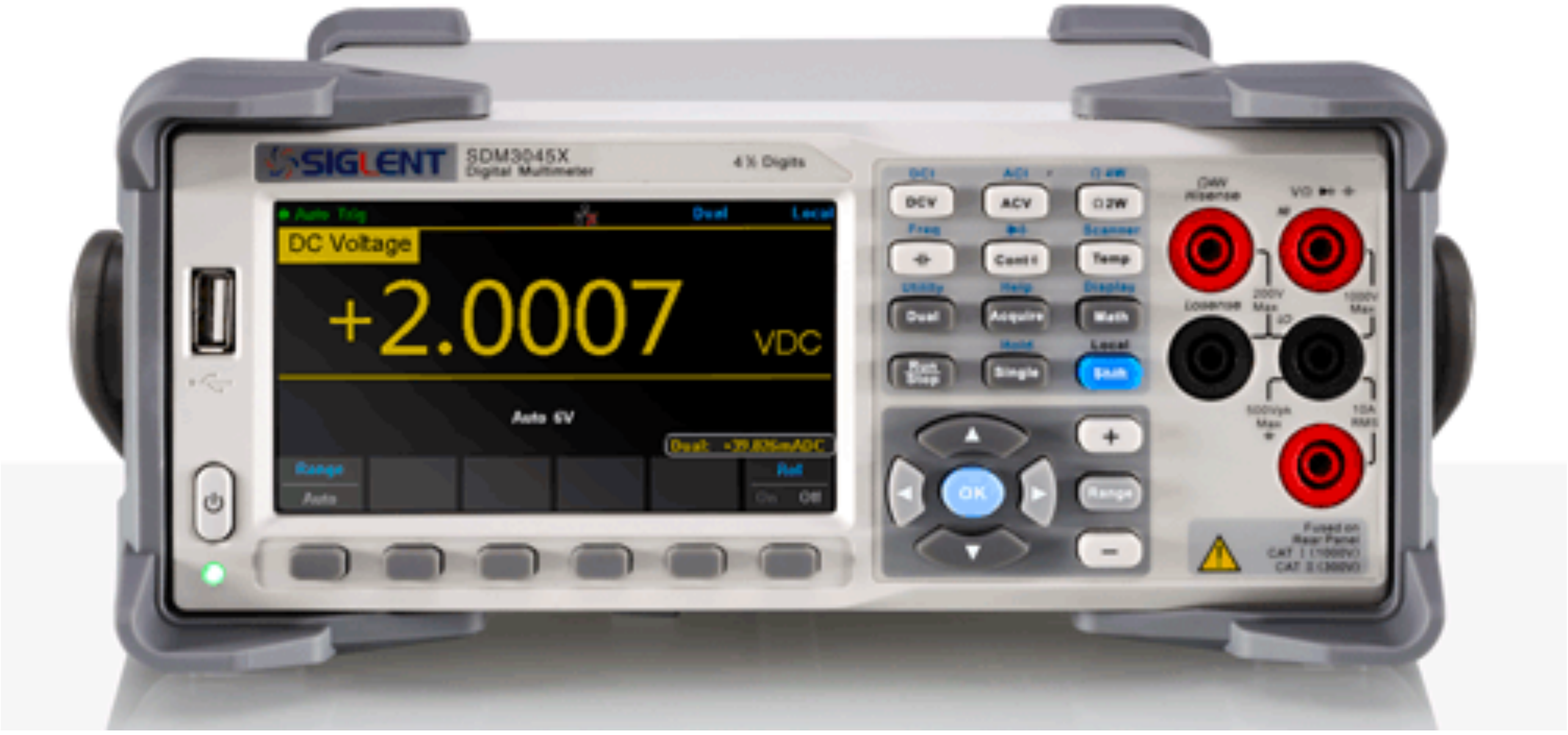
- Our measurement model, $y = x + v$, is **linear.**
- Measurements are **equally weighted.**
  (we do not suspect that some have more noise than others).

Module 1 | Lesson 1 (Part 2)

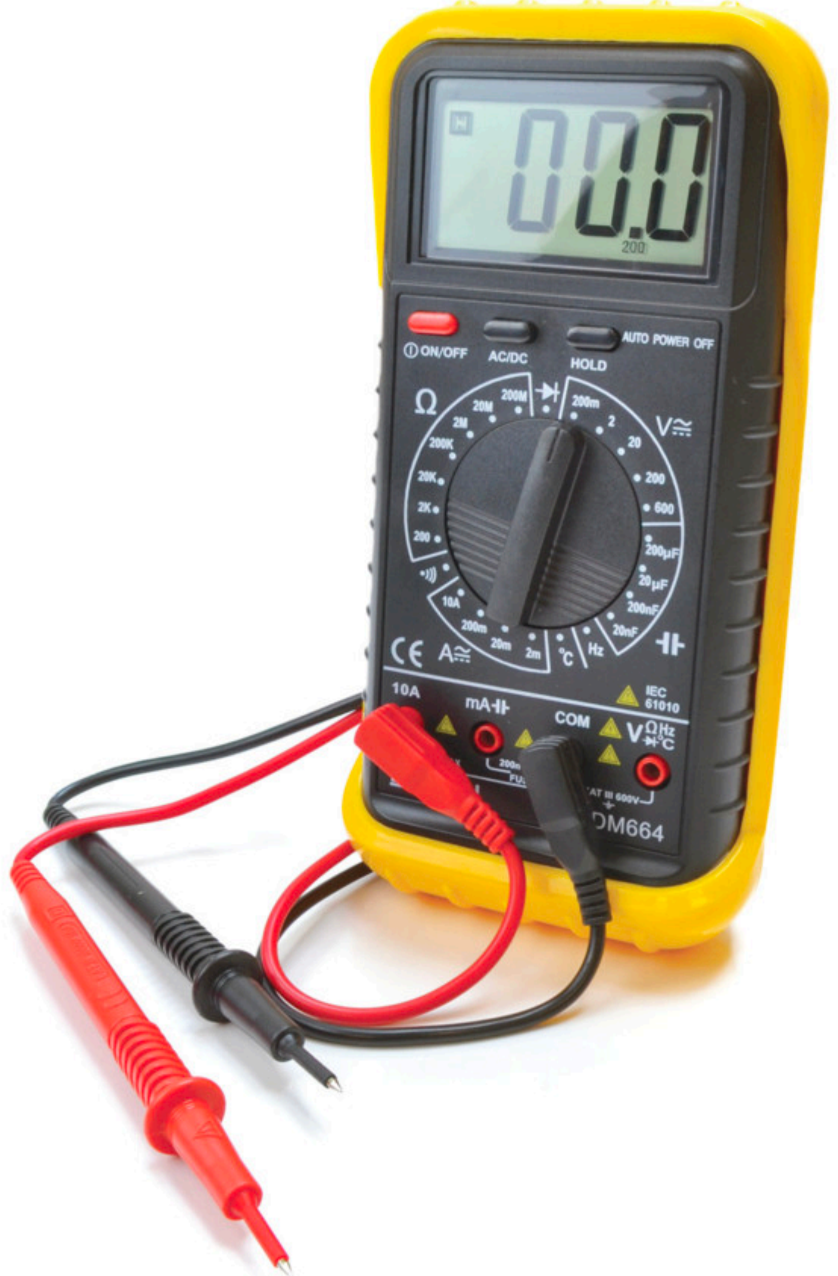
Weighted Least Squares

Method of *Weighted* Least Squares

- Suppose we take measurements with multiple multimeters, some of which are better than others



vs.



Method of *Weighted* Least Squares

If we assume each noise term is independent, but of different variance,

$$\mathbb{E}[v_i^2] = \sigma_i^2, \quad (i = 1, \ldots, m) \qquad \mathbf{R} = \mathbb{E}[\mathbf{v}\mathbf{v}^T] = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix}$$

Then we can define a *weighted* least squares criterion as:

$$\begin{aligned} \mathscr{L}_{\mathrm{WLS}}(\mathbf{x}) &= \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e} \\ &= \frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} + \ldots + \frac{e_m^2}{\sigma_m^2} \end{aligned}$$

where

$$\begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} = \mathbf{e} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} - \mathbf{H} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

The higher the expected noise, the lower the weight we place on the measurement.

Minimizing the *Weighted* Least Squares Criterion

Expanding our new criterion,

$$\begin{aligned}\mathscr{L}_{\mathrm{WLS}}(\mathbf{x}) &= \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e} \\ &= (\mathbf{y}-\mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})\end{aligned}$$

We can minimize it as before, but accounting for the new weighting term:

$$\hat{\mathbf{x}} = \operatorname{argmin}_{\mathbf{x}} \mathscr{L}(\mathbf{x}) \longrightarrow \left.\frac{\partial \mathscr{L}}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}} = \mathbf{0} = -\mathbf{y}^T\mathbf{R}^{-1}\mathbf{H} + \hat{\mathbf{x}}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}$$

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\hat{\mathbf{x}}_{\mathrm{WLS}} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$$

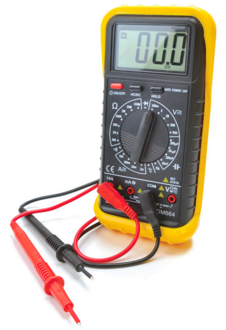
The weighted normal equations!

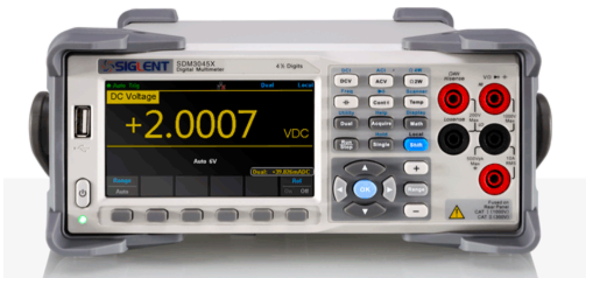
Method of *Weighted* Least Squares

The weighted normal equations

$$\boxed{\hat{\mathbf{x}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}}$$

Let's adapt our example:





	Resistance Measurements (Ohms)	
#	Multimeter A ($\sigma = 20$ Ohms)	Multimeter B ($\sigma = 2$ Ohms)
1	1068	
2	988	
3		1002
4		996

Method of *Weighted* Least Squares

Once we define the relevant quantities, we plug-and-chug to get:

$$\mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \mathbf{y} = \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \qquad \mathbf{R} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ & & & \sigma_4^2 \end{bmatrix} = \begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}$$

$$\begin{aligned}
\hat{x}_{\mathrm{WLS}} &= (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \\
&= \left([1111] \begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} [1111] \begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \\
&= \frac{1}{1/400 + 1/400 + 1/4 + 1/4} \left(\frac{1068}{400} + \frac{988}{400} + \frac{1002}{4} + \frac{996}{4} \right) \\
&= 999.3 \text{ Ohms}
\end{aligned}$$

Ordinary versus Weighted Least Squares

	Least Squares	Weighted Least Squares
Loss / Criterion	$\mathscr{L}_{\mathrm{LS}}(\mathbf{x}) = \mathbf{e}^T\mathbf{e}$	$\mathscr{L}_{\mathrm{WLS}}(\mathbf{x}) = \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e}$
Solution	$\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y}$	$\hat{\mathbf{x}}_{\mathrm{WLS}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$
Limitations	$m \geq n$	$m \geq n$ $\sigma_i^2 > 0$

Accurate noise modelling is crucial to utilize various sensors effectively!