# Least squares

Simple example

| # ($n$) | Ohms ($y_n$) |
|---|---|
| 1 | 1068 |
| 2 | 988 |
| 3 | 1002 |
| 4 | 996 |

the measurement model and the associated error

$$
\begin{array}{ll}
y_1 = x + v_1 & e_1^2 = (y_1 - x)^2 \\
y_2 = x + v_2 & e_2^2 = (y_2 - x)^2 \\
y_3 = x + v_3 & e_3^2 = (y_3 - x)^2 \\
y_4 = x + v_4 & e_4^2 = (y_4 - x)^2
\end{array}
\tag{1}
$$

minimizing the error

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

$$
\begin{aligned}
\mathbf{e} = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix} &= \mathbf{y} - \mathbf{H}x \\
= \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} &- \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x
\end{aligned}
\tag{3}
$$

Squared error, what we want to minimize

$$
\begin{aligned}
\mathscr{L}_{\mathrm{LS}}(x) = e_1^2 + e_2^2 + e_3^2 + e_4^2 &= \mathbf{e}^T\mathbf{e} \\
&= (\mathbf{y} - \mathbf{H}x)^T(\mathbf{y} - \mathbf{H}x) \\
&= \mathbf{y}^T\mathbf{y} - x^T\mathbf{H}^T\mathbf{y} - \mathbf{y}^T\mathbf{H}x + x^T\mathbf{H}^T\mathbf{H}x
\end{aligned}
\tag{4}
$$

To get the smallest sol we do the derivative

$$
\begin{aligned}
\left.\frac{\partial \mathscr{L}}{\partial x}\right|_{x=\hat{x}} = -\mathbf{y}^T\mathbf{H} - \mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0 \\
-2\mathbf{y}^T\mathbf{H} + 2\hat{x}^T\mathbf{H}^T\mathbf{H} = 0
\end{aligned}
\tag{5}
$$

solving

$$
\hat{x}_{\mathrm{LS}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{y} \tag{6}
$$

X hat minimizes our squared errors. Note: this is only possible if $H$ is not singular / has an inverse. This can be stated as the dimentions sadifying $n \geq m$, aka more mesurments than states.

$$\mathbf{y} = \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \quad \mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{7}$$

$$\hat{x}_{\text{LS}} = \left( [11111] \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \right)^{-1} [11111] \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} = 1013.5 \tag{8}$$

# Weigted least squares

We migt want to weight the mesurments because we might have better sensores that we trust more.

| # | mulitmeter-A ($\sigma = 20$ Ohm) | mulitmeter-B ($\sigma = 2$Ohm) |
|---|---|---|
| 1 | 1068 | |
| 2 | 988 | |
| 3 | | 1002 |
| 4 | | 996 |

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix} \tag{9}$$
$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

Each time varying noiseterm $\mathbf{v_i}$ is an independent random variable across measurements and has an an different variance ( or standard deviation) accociated with it.

$$\mathrm{E}\left[v_i^2\right] = \sigma_i^2, \quad (i = 1, \ldots, m) \quad \mathbf{R} = \mathbb{E}\left[\mathbf{v}\mathbf{v}^T\right] = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_m^2 \end{bmatrix} \tag{10}$$

Each squared error term is now weighted by the inverse of the variance associated with the corresponding measurement.

$$\begin{aligned}\mathscr{L}_{\text{WLS}}(\mathbf{x}) &= \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e} \\ &= \frac{e_1^2}{\sigma_1^2} + \frac{e_2^2}{\sigma_2^2} + \ldots + \frac{e_m^2}{\sigma_m^2}\end{aligned} \quad \text{where} \quad \begin{bmatrix} e_1 \\ \vdots \\ e_m \end{bmatrix} = \mathbf{e} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} - \mathbf{H} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{11}$$

We approach it's minimization the same way as before,

$$\begin{aligned}\mathscr{L}_{\text{WLS}}(\mathbf{x}) &= \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e} \\ &= (\mathbf{y} - \mathbf{H}\mathbf{x})^T\mathbf{R}^{-1}(\mathbf{y} - \mathbf{H}\mathbf{x})\end{aligned} \tag{12}$$

Solving in the same mannor

$$\frac{\partial \mathscr{L}}{\partial \mathbf{x}}_{\mathbf{x}=\hat{\mathbf{x}}} = \mathbf{0} = -\mathbf{y}^T\mathbf{R}^{-\mathbf{I}}\mathbf{H} + \hat{\mathbf{x}}^T\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} \tag{13}$$

$$\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\hat{\mathbf{x}}_{\text{WLS}} = \mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \tag{14}$$

we get

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \tag{15}$$

An example

$$\mathbf{H} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{y} = \begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ & & & \sigma_4^2 \end{bmatrix} = \begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix} \tag{16}$$

$$\begin{aligned}\hat{x}_{\text{WLS}} &= \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \\ &= \left([1111]\begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}^{-1}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}\right)^{-1}[11111]\begin{bmatrix} 400 & & & \\ & 400 & & \\ & & 4 & \\ & & & 4 \end{bmatrix}^{-1}\begin{bmatrix} 1068 \\ 988 \\ 1002 \\ 996 \end{bmatrix} \\ &= \frac{1}{1/400 + 1/400 + 1/4 + 1/4}\left(\frac{1068}{400} + \frac{988}{400} + \frac{1002}{4} + \frac{996}{4}\right) \\ &= 999.3\end{aligned} \tag{17}$$

Summery

$$\begin{matrix} \mathscr{L}_{\text{LS}}(\mathbf{x}) = \mathbf{e}^T\mathbf{e} & \mathscr{L}_{\text{WLS}}(\mathbf{x}) = \mathbf{e}^T\mathbf{R}^{-1}\mathbf{e} \\ \hat{\mathbf{x}}_{\text{LS}} = \left(\mathbf{H}^T\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{y} & \hat{\mathbf{x}}_{\text{WLS}} = \left(\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}\right)^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y} \\ m \geq n & m \geq n \\ & \sigma_i^2 > 0 \end{matrix} \tag{18}$$

| Current (A) | Voltage (V) |
|---|---|
| 0.2 | 1.23 |
| 0.3 | 1.38 |
| 0.4 | 2.06 |
| 0.5 | 2.47 |
| 0.6 | 3.17 |

## Recursive Least Squares

Teqniue that can be used when you dont want to compute the whole badge at a time.

$$\begin{aligned}\mathscr{L}_{\mathrm{RLS}} &= \mathbb{E}\left[\left(x_k - \hat{x}_k\right)^2\right] \\ &= \sigma_k^2\end{aligned} \tag{19}$$

Generalizing this to the sates $n$

$$\begin{aligned}\mathscr{L}_{\mathrm{RIS}} &= \mathbb{E}\left[\left(x_{1k} - \hat{x}_{1k}\right)^2 + \ldots + \left(x_{nk} - \hat{x}_{nk}\right)^2\right] \\ &= \operatorname{Trace}\left(\mathbf{P}_k\right)\end{aligned} \tag{20}$$

$\mathbf{P}_k$ is the state covariance matrix. We can forumate a recursive deeition of this $\mathbf{P}_k$ as a funciton of $\mathbf{K}_k$, by using matrix calculus and derivatives

$$\mathbf{P}_k = \left(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k-1}\left(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k\right)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T \tag{21}$$

$$\mathbf{K}_k = \mathbf{P}_{k-1}\mathbf{H}_k^T\left(\mathbf{H}_k\mathbf{P}_{k-1}\mathbf{H}_k^T + \mathbf{R}_k\right)^{-1} \tag{22}$$

$$\begin{aligned}\mathbf{P}_k &= \mathbf{P}_{k-1} - \mathbf{K}_k\mathbf{H}_k\mathbf{P}_{k-1} \\ &= \left(\mathbf{1} - \mathbf{K}_k\mathbf{H}_k\right)\mathbf{P}_{k-1}\end{aligned} \tag{23}$$

The larger our gain matrix $\mathbf{H}$, the smaller our new estimator covariance will be. Intuitively, you can think of this gain matrix as balancing the information we get from our prior estimate and the information we receive from our new measurement.

### Algorithm

1. We initialize the algorithm with estimate of our unknown parameters and a corresponding covariance matrix. This initial guess could come from the first measurement we take and the covariance could come from technical specifications.

$$\begin{aligned}\hat{\mathbf{x}}_0 &= \mathbb{E}[\mathbf{x}] \\ \mathbf{P}_0 &= \mathbb{E}\left[\left(\mathbf{x} - \hat{\mathbf{x}}_0\right)\left(\mathbf{x} - \hat{\mathbf{x}}_0\right)^T\right]\end{aligned} \tag{24}$$

2. Set up our measurement model and pick values for our measurement covariance.

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x} + \mathbf{v}_k \tag{25}$$

3. Update the estimate and the covariance:

$$\begin{aligned}
\mathbf{K}_k &= \mathbf{P}_{k-1} \mathbf{H}_k^T \left( \mathbf{H}_k \mathbf{P}_{k-1} \mathbf{H}_k^T + \mathbf{R}_k \right)^{-1} \\
\hat{\mathbf{x}}_k &= \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \left( \mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1} \right) \\
\mathbf{P}_k &= \left( \mathbf{1} - \mathbf{K}_k \mathbf{H}_k \right) \mathbf{P}_{k-1}
\end{aligned} \tag{26}$$

## Least Squares, Method of Maximum Likelihood

Ganger alle gausian med hverandere

$$\begin{aligned} p(\mathbf{y}|x) &\propto \mathcal{N}\left(y_1; x, \sigma^2\right) \mathcal{N}\left(y_2; x, \sigma^2\right) \times \ldots \times \mathcal{N}\left(y_m; x, \sigma^2\right) \\ &= \frac{1}{\sqrt{(2\pi)^m \sigma^{2m}}} \exp\left(\frac{-\sum_{i=1}^{m}\left(y_i - x\right)^2}{2\sigma^2}\right) \end{aligned} \tag{27}$$

Prver vi  makisimere

$$\begin{aligned} \hat{x}_{\mathrm{MLE}} &= \underset{x}{\operatorname{argmax}} p(\mathbf{y}|x) \\ &= \underset{x}{\operatorname{argmax}} \log p(\mathbf{y}|x) \end{aligned} \tag{28}$$

Take the log and get something we are farmilliar

$$\log p(\mathbf{y}|x) = -\frac{1}{2R}\left(\left(y_1 - x\right)^2 + \ldots + \left(y_m - x\right)^2\right) + C \tag{29}$$

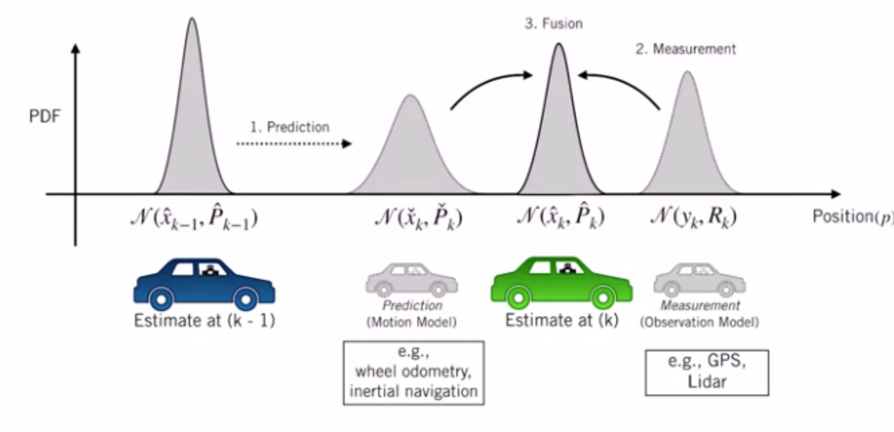
$$\hat{x}_{\mathrm{MLE}} = \underset{x}{\operatorname{argmin}} \frac{1}{2\sigma^2}\left(\left(y_1 - x\right)^2 + \ldots + \left(y_m - x\right)^2\right) \tag{30}$$

$$\hat{x}_{\mathrm{MLE}} = \underset{x}{\operatorname{argmin}} \frac{1}{2}\left(\frac{\left(y_1 - x\right)^2}{\sigma_1^2} + \ldots + \frac{\left(y_m - x\right)^2}{\sigma_m^2}\right) \tag{31}$$

$$\hat{x}_{\mathrm{MHE}} = \hat{x}_{\mathrm{LS}} = \underset{x}{\operatorname{argmin}} \mathscr{L}_{\mathrm{LS}}(x) = \underset{x}{\operatorname{argmax}} \mathscr{S}_{\mathrm{MHE}}(x) \tag{32}$$

Good, but sensetive for outliers

# Kalman



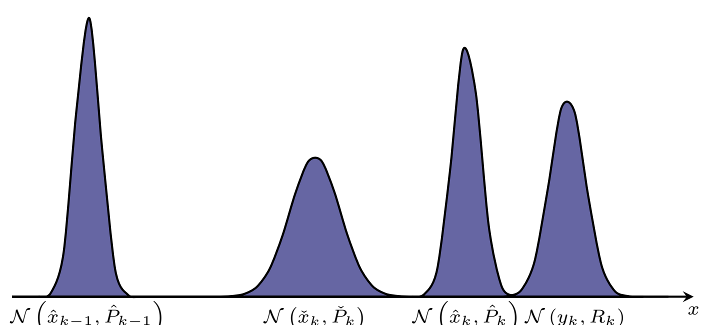




Figure 1: $y(x) = \frac{\sin(x)}{x}$