

main

nybo

October 16, 2019

Contents

1	Least squares	2
2	Weigted least squares	4
3	Recursive Least Squares	6
	3.0.1 Algorithm	6
4	Least Squares, Method of Maximum Likelihood	7
5	Kalman	8
	5.1 Linear dynamical system	8

1 Least squares

Simple example

# (n)	Ohms (y _n)
1	1068
2	988
3	1002
4	996

the measurement model and the associated error

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

minimizing the error

$$\hat{x}_{\text{LS}} = \operatorname{argmin}_x (e_1^2 + e_2^2 + e_3^2 + e_4^2) = \mathcal{L}_{\text{LS}}(x)$$

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

Squared error, what we want to minimize

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

To get the smallest sol we do the derivative

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

solving

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

\hat{x} minimizes our squared errors. Note: this is only possible if H is not singular / has an inverse. This can be stated as the dimensions satisfying $n \geq m$, aka more measurements than states.

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

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

2 Weighted least squares

We might want to weight the measurements because we might have better sensors that we trust more.

#	multimeter-A ($\sigma = 20 \text{ Ohm}$)	multimeter-B ($\sigma = 20 \text{ Ohm}$)
1	1068	
2	988	
3		1002
4		996

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \mathbf{H} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}$$

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

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

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

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

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

We approach its minimization the same way as before,

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

Solving in the same manner

$$\frac{\partial \mathcal{L}}{\partial \mathbf{x}} \bigg|_{\mathbf{x}=\hat{\mathbf{x}}} = \mathbf{0} = -\mathbf{y}^T \mathbf{R}^{-1} \mathbf{H} + \hat{\mathbf{x}}^T \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$$

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

we get

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

An example

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

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

Summery

$$\begin{array}{ll}
\mathcal{L}_{\text{LS}}(\mathbf{x}) = \mathbf{e}^T \mathbf{e} & \mathcal{L}_{\text{WLS}}(\mathbf{x}) = \mathbf{e}^T \mathbf{R}^{-1} \mathbf{e} \\
\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y} & \hat{\mathbf{x}}_{\text{WLS}} = (\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{y} \\
m \geq n & m \geq n \\
& \sigma_i^2 > 0
\end{array}$$

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

3 Recursive Least Squares

Technique that can be used when you don't want to compute the whole batch at a time.

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

Generalizing this to the states n

$$\begin{aligned}\mathcal{L}_{\text{RLS}} &= \mathbb{E} \left[(x_{1k} - \hat{x}_{1k})^2 + \dots + (x_{nk} - \hat{x}_{nk})^2 \right] \\ &= \text{Trace}(\mathbf{P}_k)\end{aligned}$$

\mathbf{P}_k is the state covariance matrix. We can formulate a recursive definition of this \mathbf{P}_k as a function of \mathbf{K}_k , by using matrix calculus and derivatives

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

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

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

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.

3.0.1 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.

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

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

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

3. Update the estimate and the covariance:

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

4 Least Squares, Method of Maximum Likelihood

Ganger alle gaussian med hverandere

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

Prøver vi å maksimere

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

Take the log and get something we are familiar

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

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

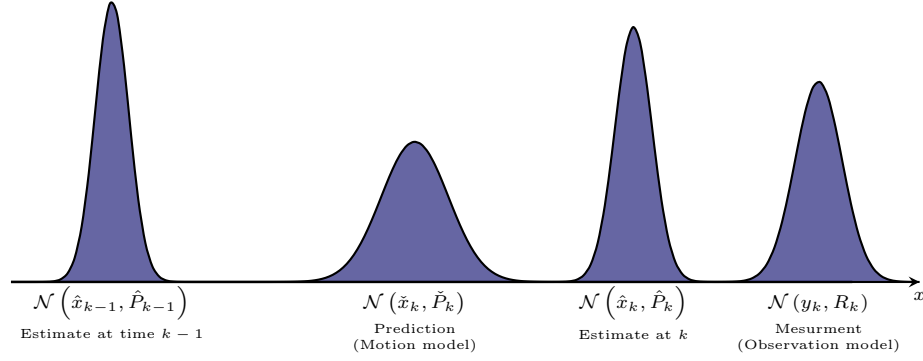
$$\hat{x}_{\text{MLE}} = \underset{x}{\operatorname{argmin}} \frac{1}{2} \left(\frac{(y_1 - x)^2}{\sigma_1^2} + \dots + \frac{(y_m - x)^2}{\sigma_m^2} \right)$$

$$\hat{x}_{\text{MLE}} = \hat{x}_{\text{LS}} = \underset{x}{\operatorname{argmin}} \mathcal{L}_{\text{LS}}(x) = \underset{x}{\operatorname{argmax}} \mathcal{S}_{\text{MLE}}(x)$$

Good, but sensitive for outliers

5 Kalman

5.1 Linear dynamical system



$$\begin{array}{lll} \text{Motion model:} & \mathbf{x}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} & + \mathbf{G}_{k-1}\mathbf{u}_{k-1} + \mathbf{w}_{k-1} \\ \text{Measurement model:} & \mathbf{y}_k = \mathbf{H}_k\mathbf{x}_k & + \mathbf{v}_k \end{array}$$

$$\begin{array}{ll} \text{Process or Motion Noise} & \mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_k) \\ \text{Measurement Noise} & \mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k) \end{array}$$

$\check{\mathbf{x}}_k$ Prediction (given motion model) at time k
 $\hat{\mathbf{x}}_k$ Corrected prediction (given measurement) at time k

$$\text{Prediction} \quad \begin{cases} \check{\mathbf{x}}_k = \mathbf{F}_{k-1}\mathbf{x}_{k-1} + \mathbf{G}_{k-1}\mathbf{u}_{k-1} \\ \check{\mathbf{P}}_k = \mathbf{F}_{k-1}\hat{\mathbf{P}}_{k-1}\mathbf{F}_{k-1}^T + \mathbf{Q}_{k-1} \end{cases}$$

$$\begin{array}{l} \text{Measurement \& Correction} \quad \begin{cases} \mathbf{K}_k = \check{\mathbf{P}}_k \mathbf{H}_k^T (\mathbf{H}_k \check{\mathbf{P}}_k \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \\ \hat{\mathbf{x}}_k = \check{\mathbf{x}}_k + \mathbf{K}_k (\mathbf{y}_k - \mathbf{H}_k \check{\mathbf{x}}_k) \\ \hat{\mathbf{P}}_k = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \check{\mathbf{P}}_k \end{cases} \end{array}$$