Module 1 | Lesson 2

Recursive Least Squares

Batch Least Squares

In our previous formulation, we assumed we had all of our measurements available when we computed our estimate:

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

'Batch Solution' $\hat{x}_{\text{WLS}} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}$

Recursive Estimation

- What happens if we have a *stream* of data? Do we need to re-solve for our solution every time? Can we do something smarter?

$$\hat{x}_1 = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_1$$

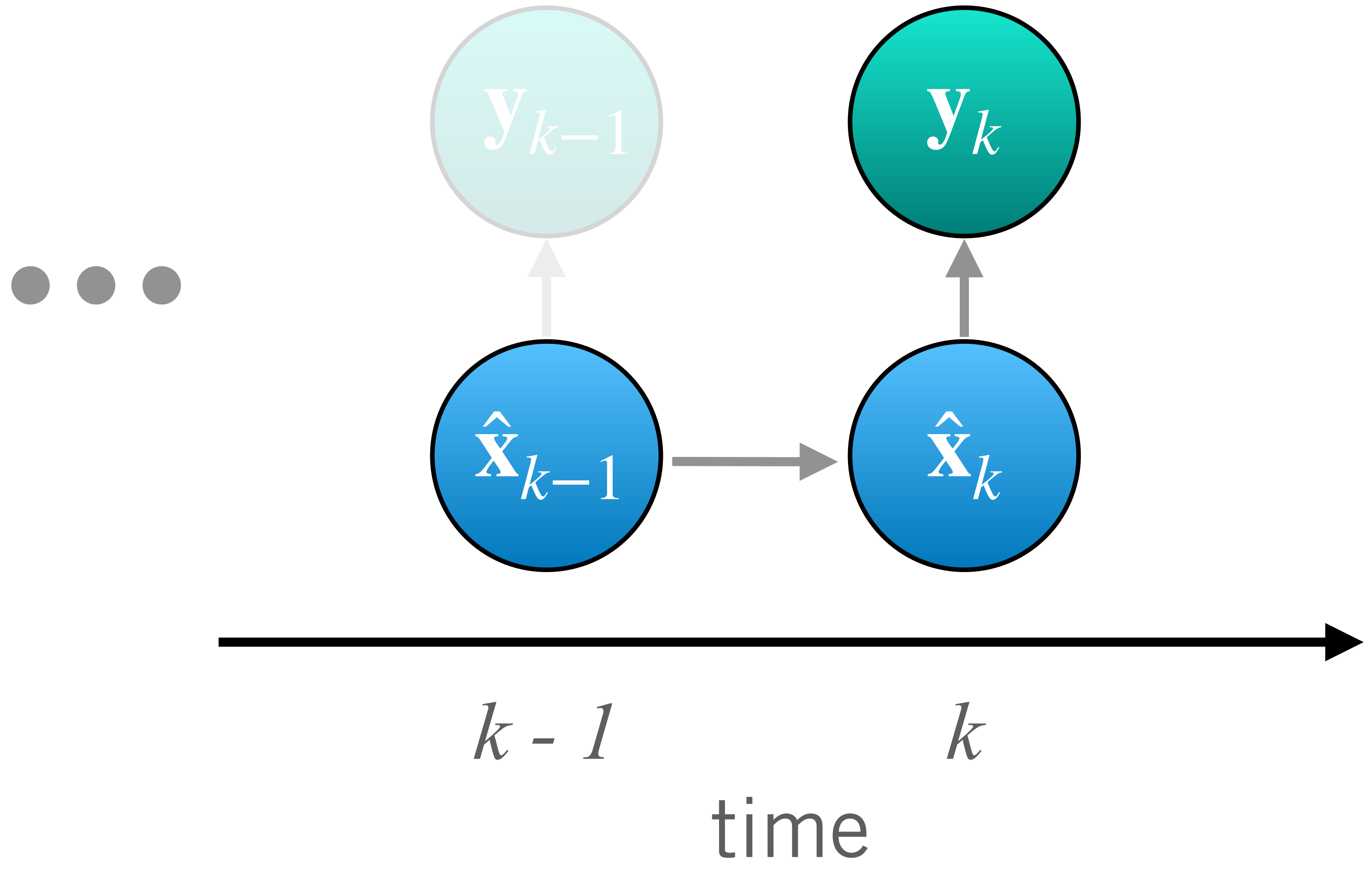
$$\hat{x}_2 = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\mathbf{H}^T\mathbf{R}^{-1}\mathbf{y}_{1:2}$$

$$\vdots$$

	Resistance (Ohms)	
Time	Multimeter A	Multimeter B
t = 1 sec	1068	
t = 2 sec	988	
t = 3 sec		1002
t = 4 sec		996

Linear Recursive Estimator

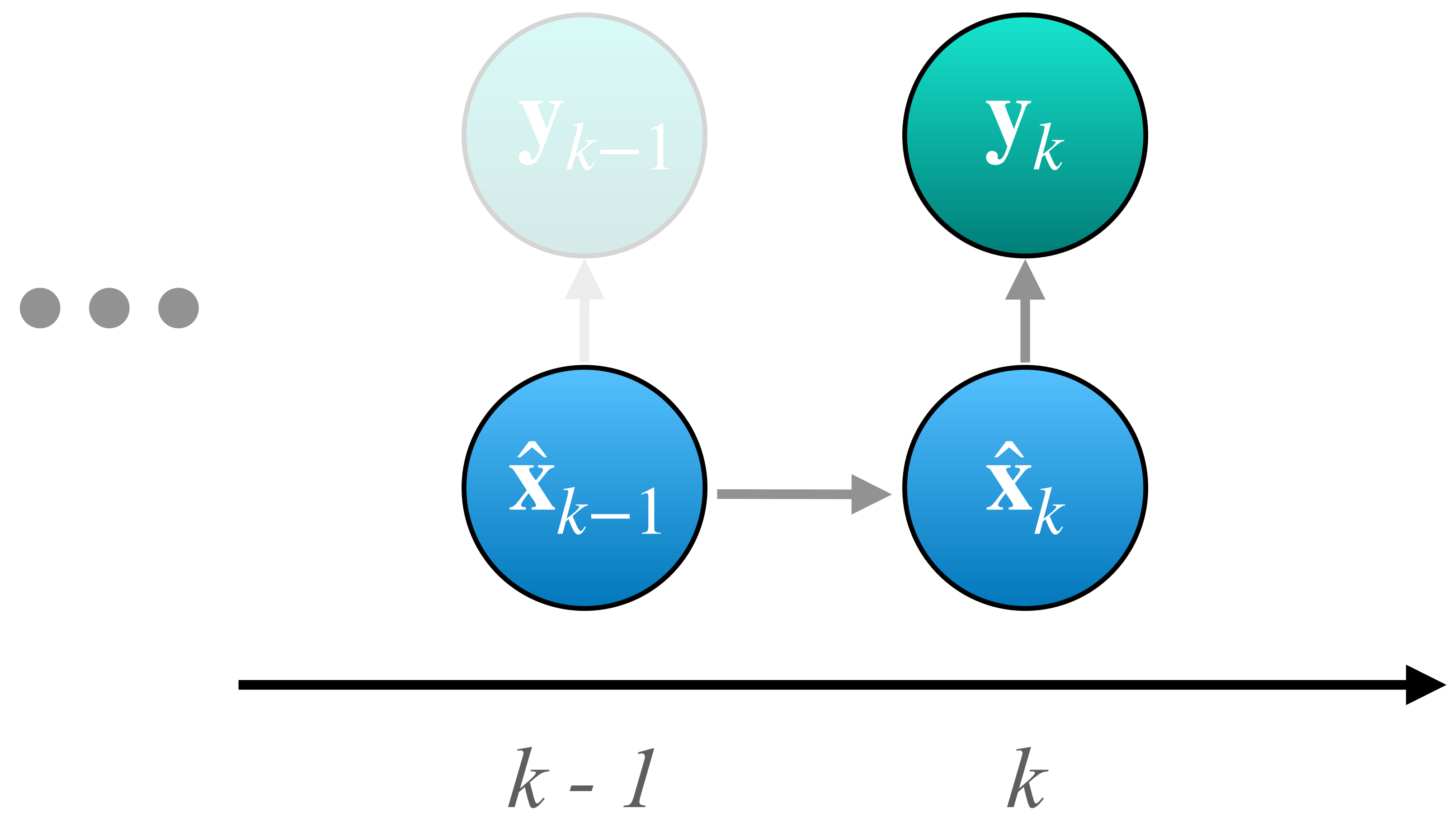
- We can use a *linear recursive estimator*
- Suppose we have an optimal estimate, $\hat{\mathbf{x}}_{k-1}$, of our unknown parameters at time interval $k-1$
- Then we obtain a new measurement at time k : $\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{v}_k$



Goal: compute $\hat{\mathbf{x}}_k$ as a function of $\mathbf{y}_k$ and $\hat{\mathbf{x}}_{k-1}$!

Given

Linear Recursive Estimator



- We can use a *linear recursive update:*

Best estimate

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k \left(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k-1}\right)$$

innovation

measurement

- We update our new state as a linear combination of the previous best guess and the current measurement *residual (or error)*, weighted by a gain matrix $\mathbf{K}_k$

Recursive Least Squares

- But what is the gain matrix $\mathbf{K}_k$?
- We can compute it by minimizing a similar least squares criterion, but this time we'll use a probabilistic formulation.
- We wish to minimize the **expected value of the sum of squared errors** of our current estimate at time step k:

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

- If we have n unknown parameters at time step k, we generalize this to

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

Estimator *covariance*

Recursive Least Squares

- Using our linear recursive formulation, we can express covariance as a function of $\mathbf{K}_k$

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

- We can show (through matrix calculus) that this is minimized when

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

- With this expression, we can also simplify our expression for $\mathbf{P}_k$:

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

Our covariance 'shrinks' with each measurement

Recursive Least Squares | Algorithm

1. Initialize the estimator

$$\hat{\mathbf{x}}_0 = \mathbb{E}[\mathbf{x}]$$

$$\mathbf{P}_0 = \mathbb{E}[(\mathbf{x} - \hat{\mathbf{x}}_0)(\mathbf{x} - \hat{\mathbf{x}}_0)^T]$$

2. Set up the measurement model, defining the Jacobian and the measurement covariance matrix:

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

3. Update the estimate of $\hat{\mathbf{x}}_k$ and the covariance $\mathbf{P}_k$ using:

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

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k-1} + \mathbf{K}_k\left(\mathbf{y}_k - \mathbf{H}_k\hat{\mathbf{x}}_{k-1}\right)$$

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

Important! Our parameter covariance ‘shrinks’ with each measurement

Summary | Recursive Least Squares

- RLS produces a ‘running estimate’ of parameter(s) for *a stream of measurements*

- RLS is a linear recursive estimator that minimizes the (co)variance of the parameter(s) at the current time

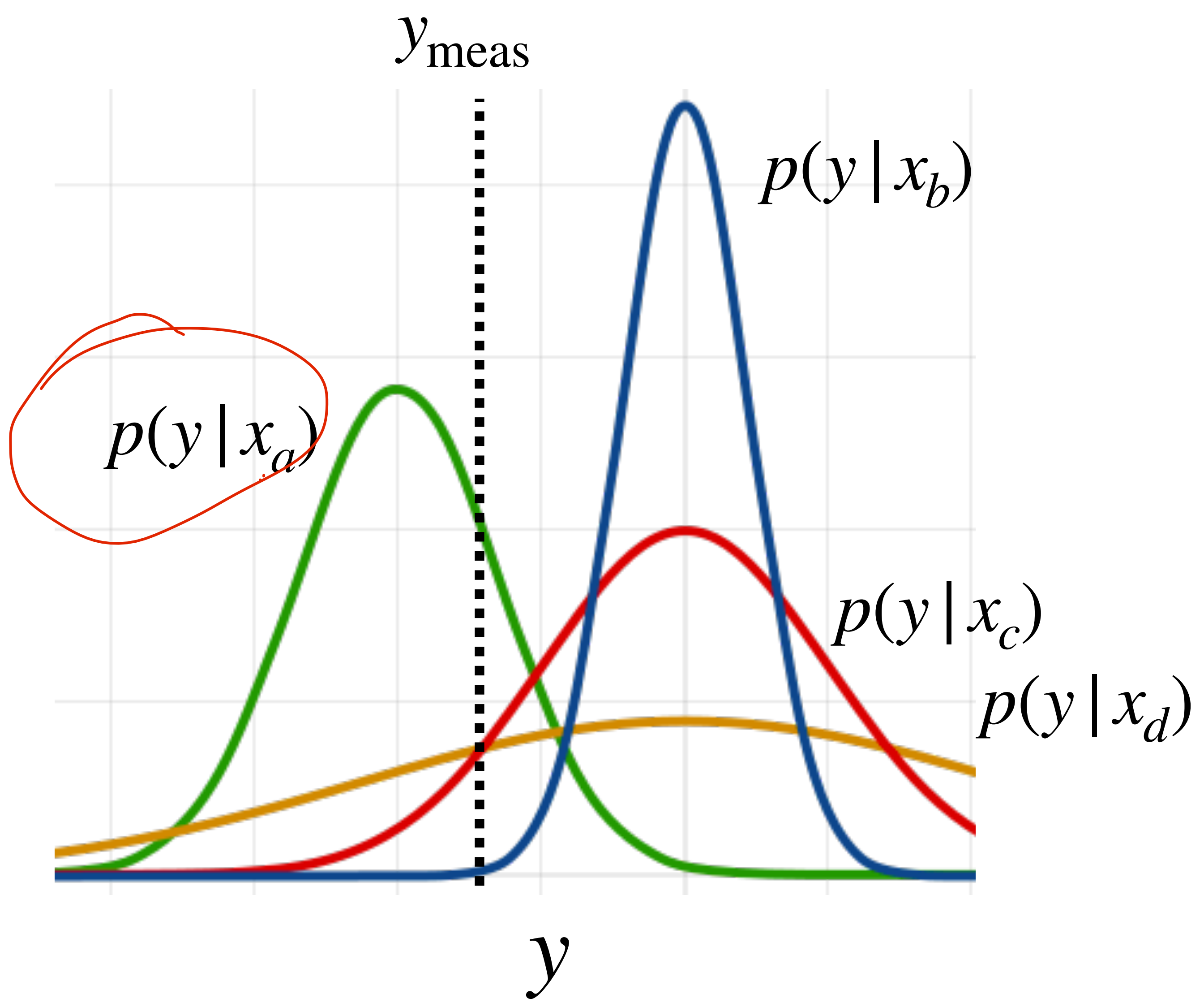
Module 1 | Lesson 3

Least Squares and Maximum Likelihood

The Method of Maximum Likelihood

- We can ask which x makes our measurement *most likely*. Or, in other words, which x maximizes the conditional probability of y:

$$\hat{x} = \text{argmax}_{\,x}\; p(y|x)$$



Which x is the most likely given the measurement?

Measurement Model

- Recall our simple measurement model:

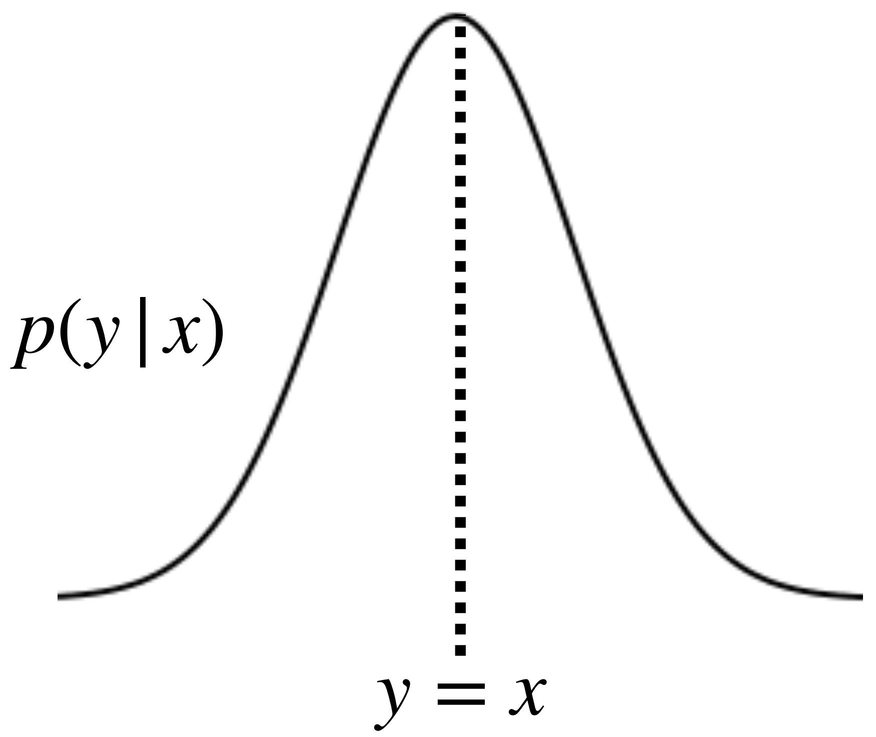
$$y = x + v$$

- We can convert this into a conditional probability on our measurement, by assuming some probability density for v. For example, if

$$v \sim \mathcal{N}(0,\ \sigma^2)$$

- Then:

$$p(y\,|\,x) = \mathcal{N}(x, \sigma^2)$$



Least Squares and Maximum Likelihood

- Probability density function of a Gaussian is:

$$\mathcal{N}(z;\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(z-\mu)^2}{2\sigma^2}}$$

- Our conditional measurement likelihood is

$$\begin{aligned} p(y \mid x) &= \mathcal{N}(y;x,\sigma^2) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(y-x)^2}{2\sigma^2}} \end{aligned}$$

- If we have multiple independent measurements, then:

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

Least Squares and Maximum Likelihood

- The maximal likelihood estimate (MLE) is given by

$$\hat{x}_{\mathrm{MLE}} = \operatorname{argmax}_x \, p(\mathbf{y} \,|\, x)$$

- Instead of trying to optimize the likelihood directly, we can take its logarithm:

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

The logarithm is monotonically increasing!

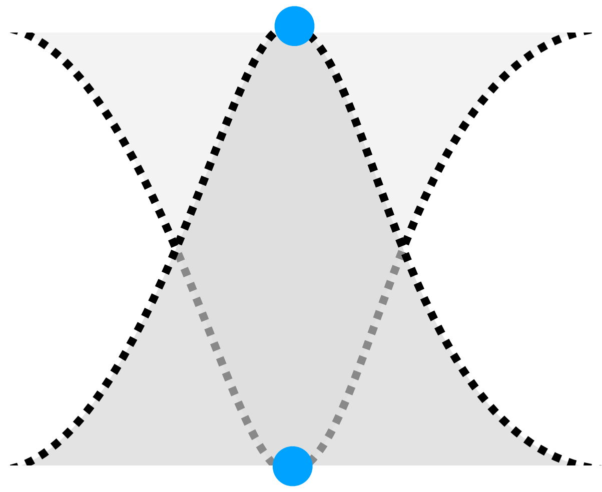
- Resulting in:

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

Least Squares and Maximum Likelihood

- Since

$$\operatorname{argmax}_z f(z) = \operatorname{argmin}_z \left(-f(z)\right)$$



- The maximal likelihood problem can therefore be written as

$$\begin{aligned}\hat{x}_{\mathrm{MLE}} &= \operatorname{argmin}_x -\left(\log p(\mathbf{y} \mid x)\right) \\ &= \operatorname{argmin}_x \frac{1}{2\sigma^2}\left((y_1 - x)^2 + \ldots + (y_m - x)^2\right)\end{aligned}$$

Least Squares and Maximum Likelihood

- So:

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

- Finally, if we assume each measurement has a different variance, we can derive

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

In both cases,

$$\hat{x}_{\text{MLE}} = \hat{x}_{\text{LS}} = \operatorname{argmin}_x \mathscr{L}_{\text{LS}}(x) = \operatorname{argmin}_x \mathscr{L}_{\text{MLE}}(x)$$

The Central Limit Theorem

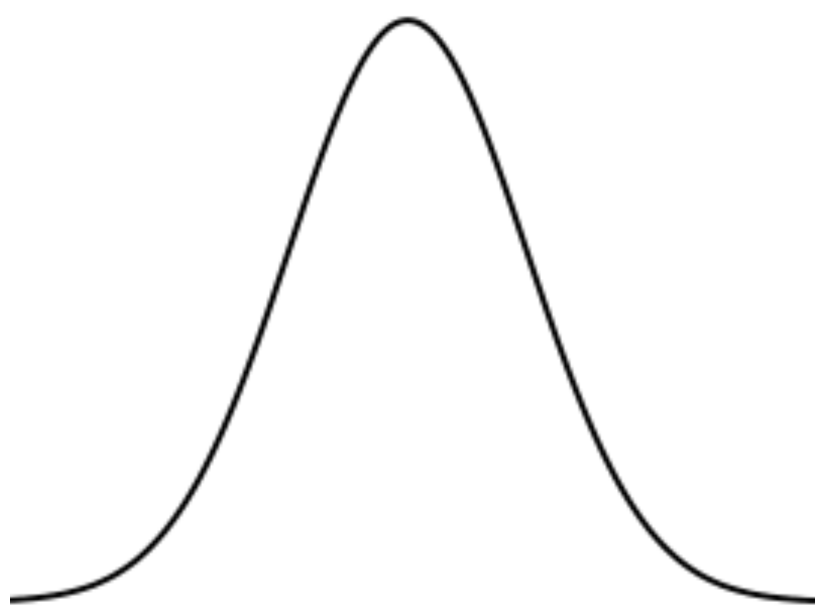
- In realistic systems like self driving cars, there are many sources of ‘noise’

> *Central Limit Theorem: When independent random variables are added, their normalized sum tends towards a normal distribution.*

- Why use the method of least squares?
  1. Central Limit Theorem: sum of different errors will tend be ‘Gaussian’-ish
  2. Least squares is equivalent to maximum likelihood under Gaussian noise

Least Squares | Some Caveats

- 'Poor' measurements (e.g. outliers) have a significant effect on the method of least squares
- It's important to check that the measurements roughly follow a Gaussian distribution



Under the Gaussian PDF, samples 'far away' from the mean are 'very improbable'

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

$\hat{x} = 1013.5$

#	Resistance (Ohms)
1	1068
2	988
3	1002
4	996
5 (outlier)	1430

$\hat{x} = 1096.8$

Summary | Least Squares and Maximum Likelihood

- LS and WLS produce the same estimates as maximum likelihood assuming Gaussian noise
- Central Limit Theorem states that complex errors will tend towards a Gaussian distribution.
- Least squares estimates are significantly affected by *outliers*