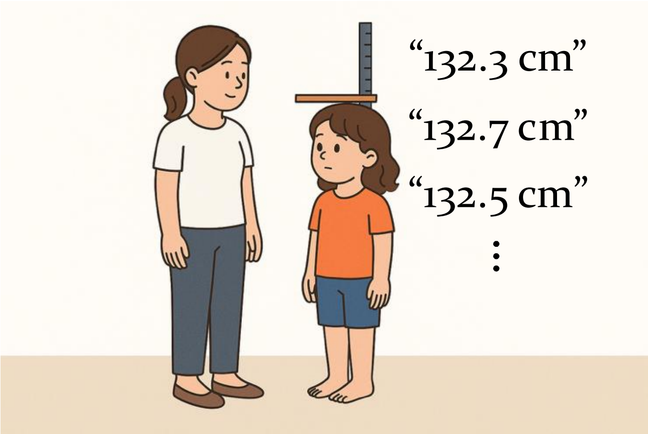
Basic SLAM Tutorials for Everyone
Lec. #4. Maximum Likelihood Estimation (MLE) and Maximum A Posteriori (MAP)

Hyungtae Lim, Ph.D.

<shapelim [at] mit [dot] edu> / <fudxo5143+slam [at] gmail [dot] com>

ORORA lab
Open Robotics and Resilient AI

Introduction: MLE? MAP?



Example A:
- Measurements: {132.3, 132.7, 132.5}

$\Longrightarrow$ *Maximum likelihood estimation (MLE)*

Example B:
- Measurements: {132.3, 132.7, 132.5}
- Prior: 130

$\Longrightarrow$ *Maximum A Posteriori (MAP)*

Probability vs. Likelihood

- Example: Coin flip experiment
- Observation (Fixed Data): 8 Heads and 2 Tails in 10 flips.
- θ: the probability of obtaining heads in a single trial
- Such results can occur even if the coin is fair (i.e., θ = 0.5).

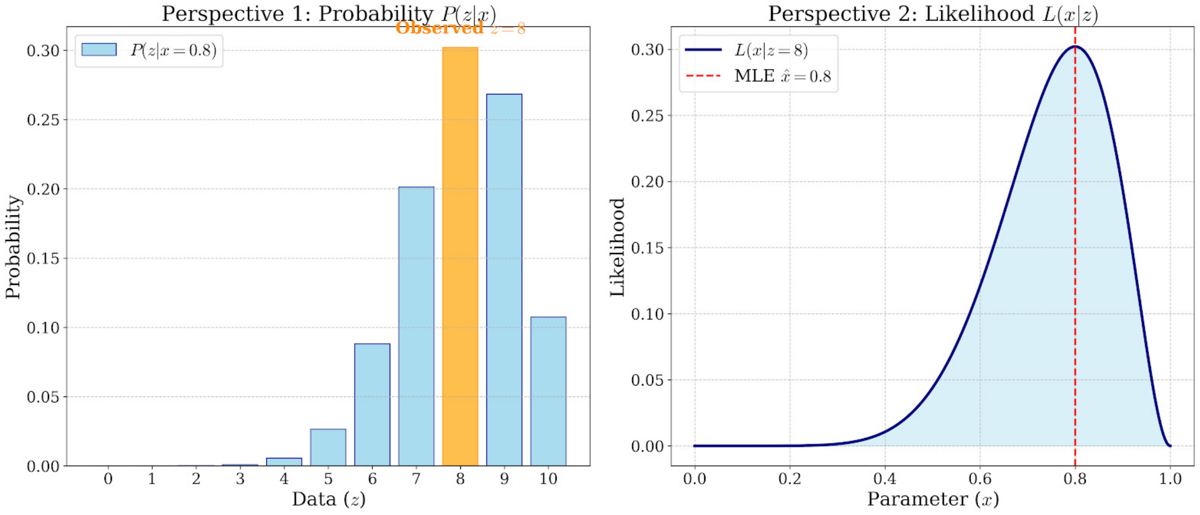
- Question: Which hypothesis is more "plausible"?
  - θ = 0.5 (Fair coin) vs. θ = 0.8 (Biased coin, prone to Heads)
- Introducing *Likelihood*
  - Definition: Varying the parameter θ for a fixed set of data to see "who explains it better."
    - How well does each parameter θ explain the observed data?

Probability vs. Likelihood (Cont'd)

- Observation (Fixed Data): 8 Heads and 2 Tails in 10 flips.
- We use the **same mathematical formula** as probability,
  - but likelihood: we vary parameter θ while keeping observation z:

- Let us set $x \leftrightarrow \theta$

$$\mathcal{L}(x|z) \propto p(z|x) = \binom{10}{8} x^8 (1-x)^2$$



ORORA lab
Open Robotics and Resilient AI

Probability vs. Likelihood (Cont'd)

Perspective	Given	Variable	Goal	Expression
Probability	Parameter θ	Observed data z	Predicting data	$p(z\|\theta)$
Likelihood	Observed data z	Parameter θ	Estimating parameters	$\mathcal{L}(\theta\|z)$

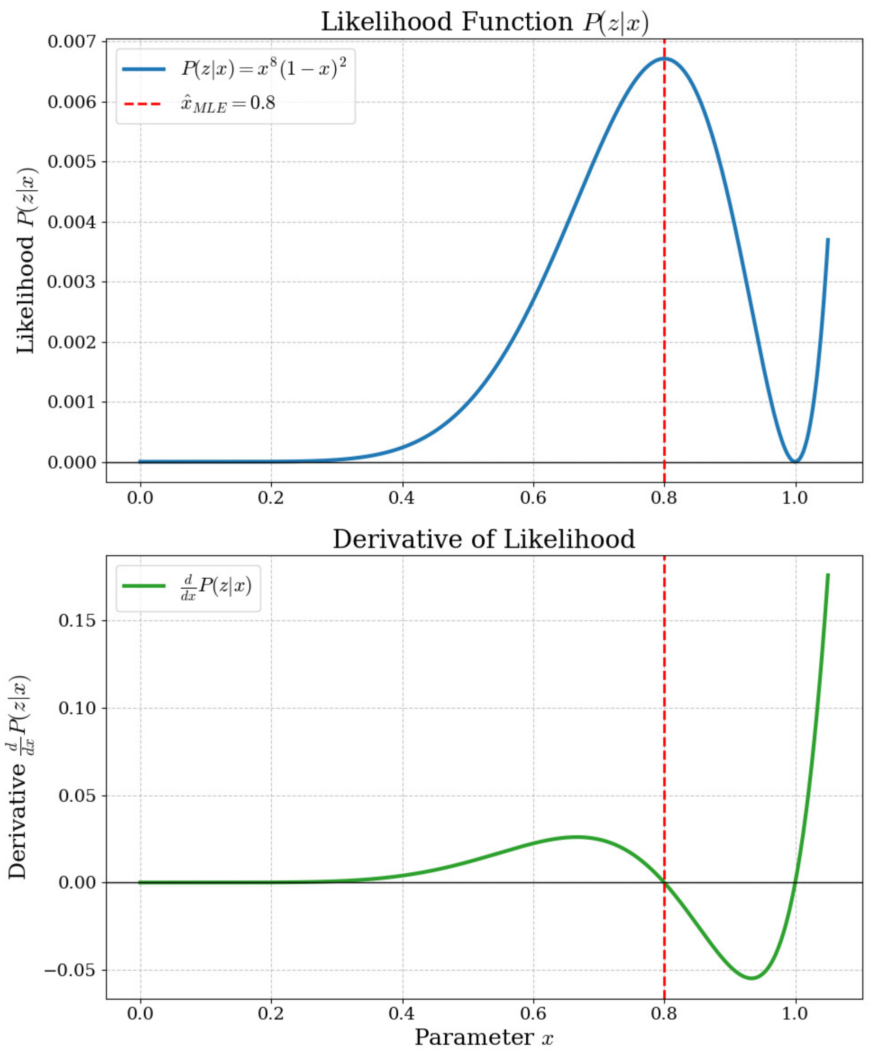
- **1. Probability: *Forward Looking (prediction)***
  - **Given**: A fixed model/parameter (θ).
  - **Question**: What is the chance of observing various data (z)?
  - **Constraint**: Sum of all possible outcomes must be 1 ($\Sigma p(z|\theta) = 1$).
  - **Purpose**: Prediction of *future events*.
- **2. Likelihood: *Backward Looking (inference)***
  - **Given**: Observed, fixed data (z).
  - **Question**: How well does each candidate parameter (θ) explain this specific data?
  - **Constraint**: Does not necessarily sum to 1.
  - **Purpose**: *Estimation* of the underlying cause (*state*).

ORORA lab
Open Robotics and Resilient AI

Maximum Likelihood Estimation (MLE)

- Now, let us set $x \leftrightarrow \theta$

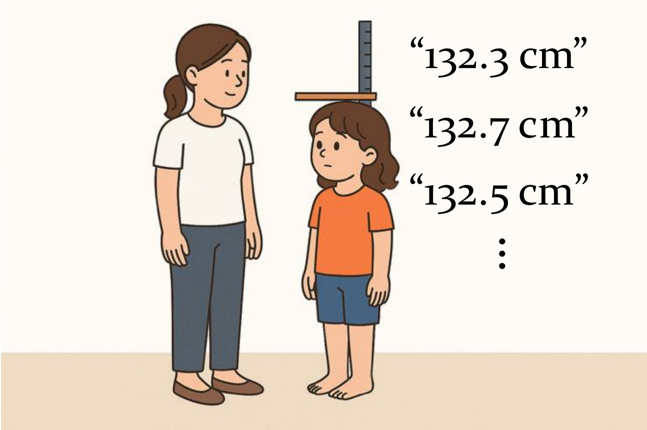
$$\hat{x}_{\mathrm{MLE}} = \arg\max_{x} \mathcal{L}(x|z)$$
$$\propto \arg\max_{x} p(z|x)$$
$$= 0.8$$



MLE in Height Estimation

$$\hat{x}_{\mathrm{MLE}} = \arg\max_x \mathcal{L}(x|z)$$
$$= \arg\max_x p(z|x)$$

Q. How to model $p(z|x)$?



Case A:

- Measurements: {132.3, 132.7, 132.5}

From Measurements to a Probabilistic Model

Noise model: A measurement = true state + noise:

$$z = x + \epsilon, \qquad \epsilon : \text{measurement noise}$$

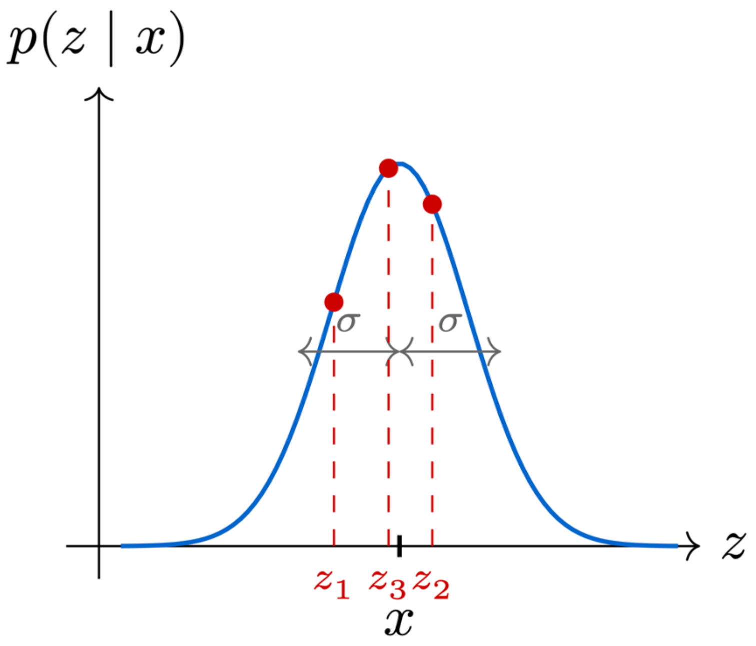
$\Rightarrow \; p(z \mid x) = p_\epsilon(z - x).$ So we need a model for ϵ.

Why Gaussian?

- Sensor noise $=$ sum of many small, independent error sources.
- By the **Central Limit Theorem (CLT)**, such sums $\rightarrow$ Gaussian.

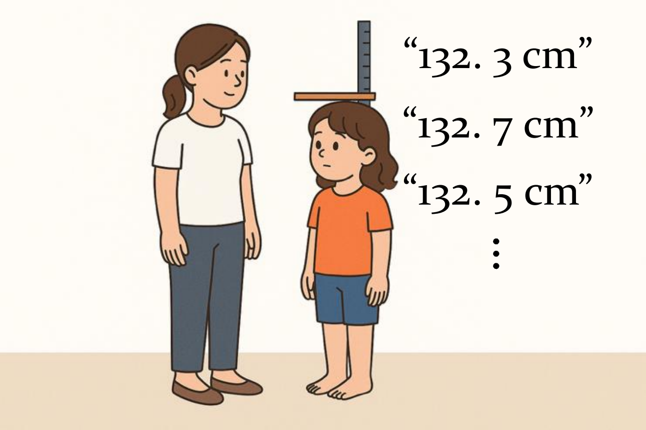
If $\epsilon \sim \mathcal{N}(0, \sigma^2)$, then:

$$p(z_i \mid x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(z_i - x)^2}{2\sigma^2}\right)$$



Each z_i is a sample from $\mathcal{N}(x, \sigma^2)$

MLE for Height Estimation (Example A)



Three independent height measurements with equal noise σ:
$z_1 = 132.3,\; z_2 = 132.7,\; z_3 = 132.5$ cm.
Joint likelihood (independence):

$$p(z_1, z_2, z_3 \mid x) = p(z_1 \mid x)p(z_2 \mid x)p(z_3 \mid x) = \prod_{i=1}^{3} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(z_i - x)^2}{2\sigma^2}\right)$$

Likelihood to Negative log-likelihood:

$$-\ln p(z_1, z_2, z_3 \mid x) = \frac{1}{2\sigma^2} \sum_{i=1}^{3} (z_i - x)^2 + \text{const.}$$

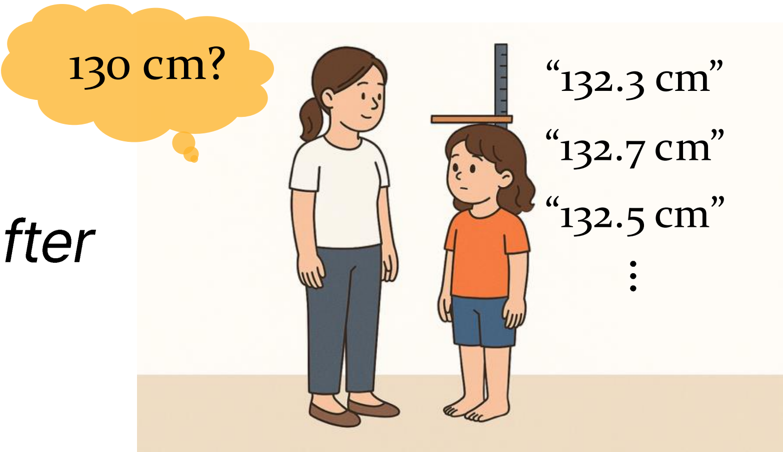
MLE $=$ minimize the sum of squared errors (σ cancels):

$$\hat{x}_{\text{MLE}} = \arg\min_x \sum_{i=1}^{3} (z_i - x)^2 \quad \Longrightarrow \quad \hat{x} = \tfrac{1}{3} \sum_{i=1}^{3} z_i = 132.5 \text{ cm}$$

Equal σ $\Rightarrow$ **equal weights** $\Rightarrow$ **simple average** (Example A).
Different σ_i $\Rightarrow$ weight $\omega_i = 1/\sigma_i^2$ $\Rightarrow$ **weighted average** (Example B).

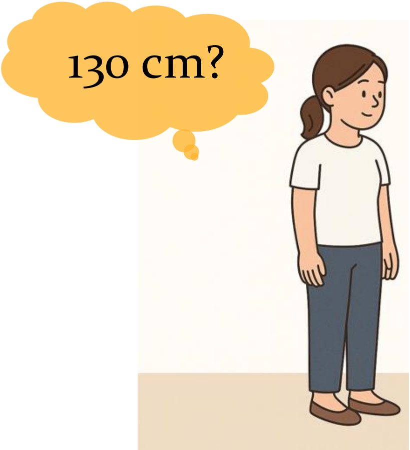
ORORA lab
Open Robotics and Resilient AI

Maximum a Posteriori (MAP)



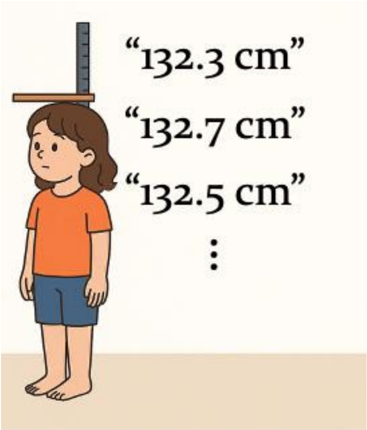
- A Posteriori: Derived from Latin, meaning *from what comes after*
  - Knowledge or reasoning derived from empirical evidence or observed data.

Prior
(previous information/belief)



X

Likelihood
(Observation)



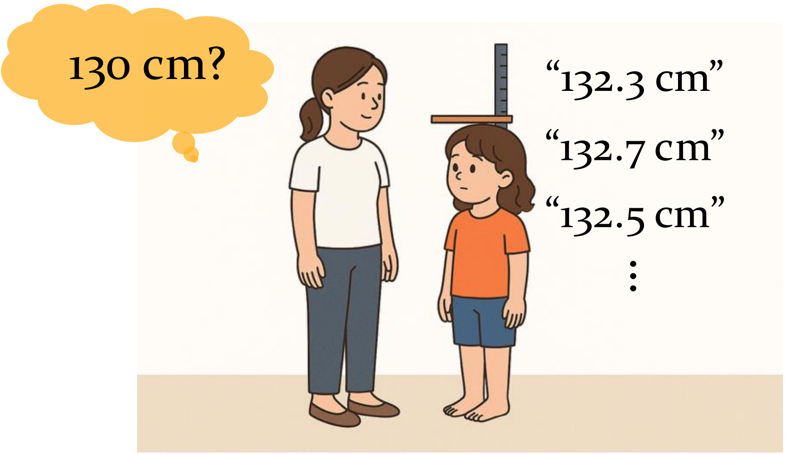
= *Posterior*

MAP for Height Estimation (Example B)

$$\hat{x}_{\text{MAP}} = \arg\max_x p(x|z)$$

$$= \arg\max_x \frac{p(z|x)p(x)}{p(z)}$$

$$\propto \arg\max_x p(z|x)p(x)$$



Case B:

	Prior: 130	Meas. avg.: 132.5
Case 1: Reliability ⇒ Estimate	$\omega_0 = 10$ $\frac{10\cdot130+3\cdot0.5\cdot132.5}{10+3\cdot0.5} = 130.33$ cm	$\omega_m = 0.5$ (each)
Case 2: Reliability ⇒ Estimate	$\omega_0 = 0.1$ $\frac{0.1\cdot130+3\cdot0.5\cdot132.5}{0.1+3\cdot0.5} = 132.34$ cm	$\omega_m = 0.5$ (each)

∴ Gaussian Noise Leads MLE (and MAP) to *Least Squares*

Assume independent measurements with Gaussian noise: $z_i = h(\boldsymbol{x}) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma_i^2)$.

Likelihood of a single measurement:

$$p(z_i \mid \boldsymbol{x}) = \frac{1}{\sqrt{2\pi}\,\sigma_i} \exp\left(-\frac{(z_i - h(\boldsymbol{x}))^2}{2\sigma_i^2}\right)$$

Joint likelihood (independence): $p(\boldsymbol{z}_{1:N} \mid \boldsymbol{x}) = \prod_{i=1}^{N} p(z_i \mid \boldsymbol{x})$.

Negative log-likelihood:

$$-\ln p(\boldsymbol{z}_{1:N} \mid \boldsymbol{x}) = \sum_{i=1}^{N} \frac{(z_i - h(\boldsymbol{x}))^2}{2\sigma_i^2} + \text{const.}$$

MLE under Gaussian noise $=$ Weighted Least Squares:

$$\hat{\boldsymbol{x}}_{\text{MLE}} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{N} \frac{1}{2\sigma_i^2} (z_i - h(\boldsymbol{x}))^2$$

The weight $\omega_i = 1/\sigma_i^2$ is the **information** (inverse variance) of measurement i.

Least Squares in Multi-Dimensional Space

MLE under Gaussian noise = Weighted Least Squares:

$$\hat{\boldsymbol{x}}_{\mathsf{MLE}} = \arg\min_{\boldsymbol{x}} \sum_{i=1}^{N} \frac{1}{2\sigma_i^2} \left(z_i - h(\boldsymbol{x})\right)^2$$

⬇

$$\hat{\mathbf{x}}_{\mathrm{MLE}} = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \frac{1}{2} \left(\mathbf{z}_i - \mathbf{h}_i(\mathbf{x})\right)^T \boldsymbol{\Sigma}_i^{-1} \left(\mathbf{z}_i - \mathbf{h}_i(\mathbf{x})\right)$$

or

$$\hat{\mathbf{x}}_{\mathrm{MLE}} = \arg\min_{\mathbf{x}} \sum_{i=1}^{N} \frac{1}{2} \left\|\mathbf{z}_i - \mathbf{h}_i(\mathbf{x})\right\|_{\boldsymbol{\Sigma}_i}^2$$

Least Squares in Multi-Dimensional Space (Cont'd)

Non-linear Least Squares

Introduction

Ceres can solve bounds constrained robustified non-linear least squares problems of the form

(1)

$$\min_{\mathbf{x}} \quad \frac{1}{2}\sum_i \rho_i\left(\left\|f_i\left(x_{i_1},\ldots,x_{i_k}\right)\right\|^2\right)$$
$$\text{s.t.} \quad l_j \le x_j \le u_j$$

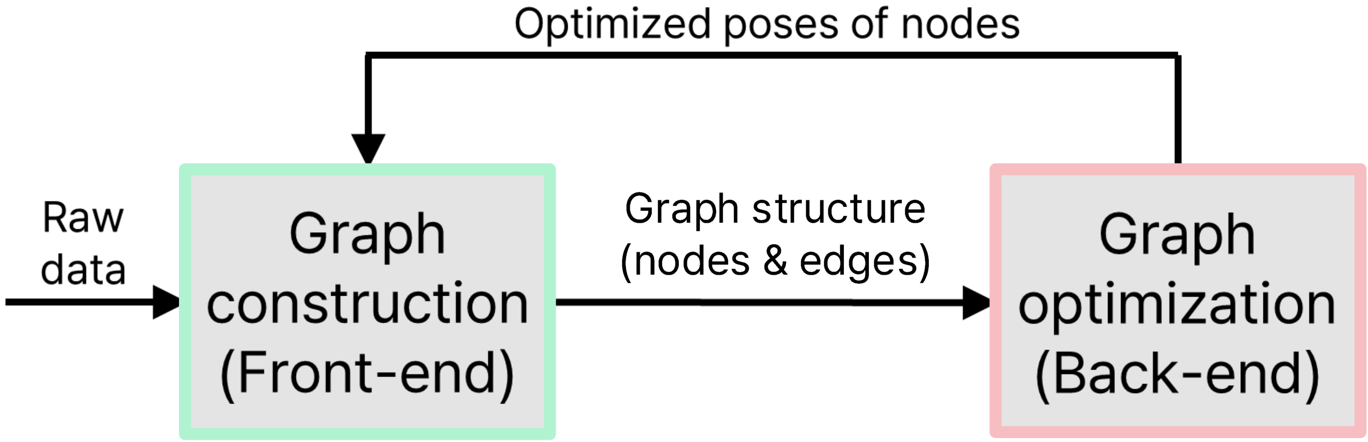
Problems of this form comes up in a broad range of areas across science and engineering - from fitting curves in statistics, to constructing 3D models from photographs in computer vision.

In this chapter we will learn how to solve (1) using Ceres Solver. Full working code for all the examples described in this chapter and more can be found in the examples directory.

http://ceres-solver.org/nnls_tutorial.html

A First Look at Least Squares in Graph SLAM

- Graph-based SLAM
  - = Front-End (Graph construction)
  - + Back-End (Optimization)



$$\mathcal{X}^{\text{MAP}} = \underset{\mathbf{X}\in\mathcal{X}}{\text{argmin}} \sum_i \left\| h\left(\mathbf{X}_i, \mathbf{X}_{i+1}\right) \boxminus \mathbf{z}_i^{\text{odom}} \right\|^2_{\Sigma_{\text{odom}_i}} + \sum_{(j,k)\in\mathcal{I}} \left\| s_{jk} \left(h\left(\mathbf{X}_j, \mathbf{X}_k\right) \boxminus \mathbf{z}_{jk}^{\text{loop}} \right) \right\|^2_{\Sigma_{\text{loop}_{jk}}}$$

where $j+1 \neq k$

$\mathbf{X}$: State $\in$ SE(n)

Σ: $n \times n$ covariance

$\mathcal{I}$: A set of index pairs from loop detection

$\mathbf{z}^{\text{odom}}$: Measurements from odometry

$\mathbf{z}^{\text{loop}}$: Measurements from loop closure

$s_j k$: Scaling factor

$\boxminus$: Minus operation in SE(n)

ORORA lab
Open Robotics and Resilient AI

Conclusion

- Probability: forward looking vs. likelihood: backward looking
- Likelihood: *how probable the observation is* given a state
- Posterior: *how probable a state is* given the observation

$$\underbrace{p(\boldsymbol{x} \mid \boldsymbol{z}_{1:N})}_{\text{Posterior}} = \frac{\overbrace{p(\boldsymbol{z}_{1:N} \mid \boldsymbol{x})}^{\text{Likelihood}} \cdot \overbrace{p(\boldsymbol{x})}^{\text{Prior}}}{\underbrace{p(\boldsymbol{z}_{1:N})}_{\text{Evidence (const. w.r.t. } \boldsymbol{x})}}$$

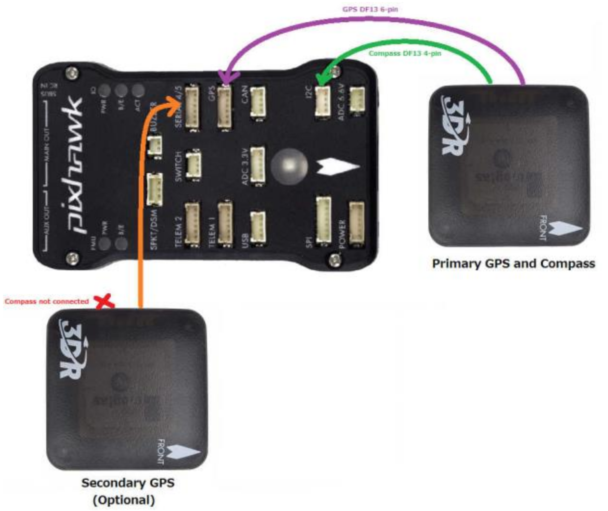
Likelihood $p(\boldsymbol{z}_{1:N} \mid \boldsymbol{x})$	Prior $p(\boldsymbol{x})$
How probable are the observed measurements, given a hypothesized state $\boldsymbol{x}$?	What do we believe about $\boldsymbol{x}$ before seeing data? E.g., last year's height + growth model.

- Estimation = weighted least squares
  - By leveraging Gaussian noise modeling

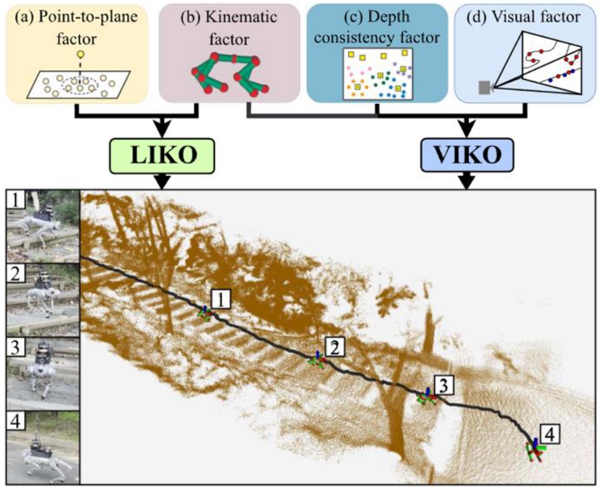
ORORA lab
Open Robotics and Resilient AI

Conclusion (Cont'd)

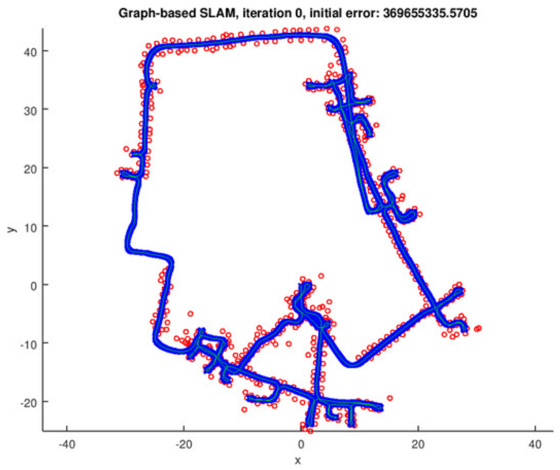
TL;DR: State estimation is fundamentally a game of *weighted averages*.



Dual GPS
https://ardupilot.org/copter/docs/common-gps-blending.html



Sensor fusion
Marsim et al., LVI-Q, *RA-L*, 2025



Graph optimization