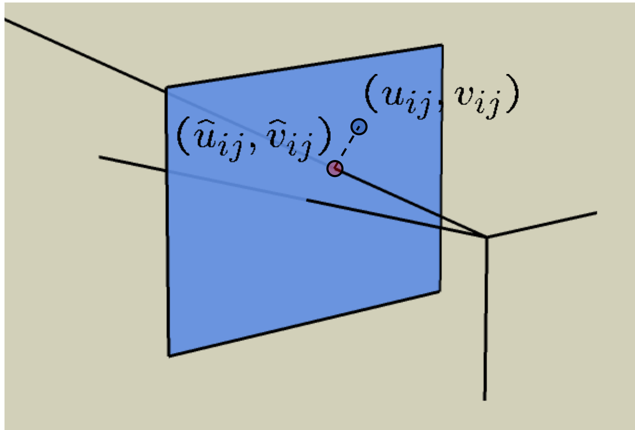
Numerical computation II

- Reprojection error
- Bundle adjustment
- Family of Newton's methods
- Statistical background
- Maximum likelihood estimation

Reprojection error

- Reprojection error = Distance between the observation and its estimation (reproduction) measured on the image

$$E(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} (u_{ij} - \hat{u}_{ij})^2 + (v_{ij} - \hat{v}_{ij})^2$$



$$\mathbf{x} \leftarrow \{\mathtt{P}_1, \ldots, \mathtt{P}_\mathtt{m}, \mathbf{X}_1, \ldots, \mathbf{X}_\mathtt{n}\}$$

- We search for x that minimizes E(x); Called bundle adjustment
- Minimization is performed by Newton’s method etc.
  - Good initial values necessary

Bundle adjustment

- Minimizing the sum of reprojection errors with all the unknown parameters, i.e., camera parameters and point coordinates

Geometric model

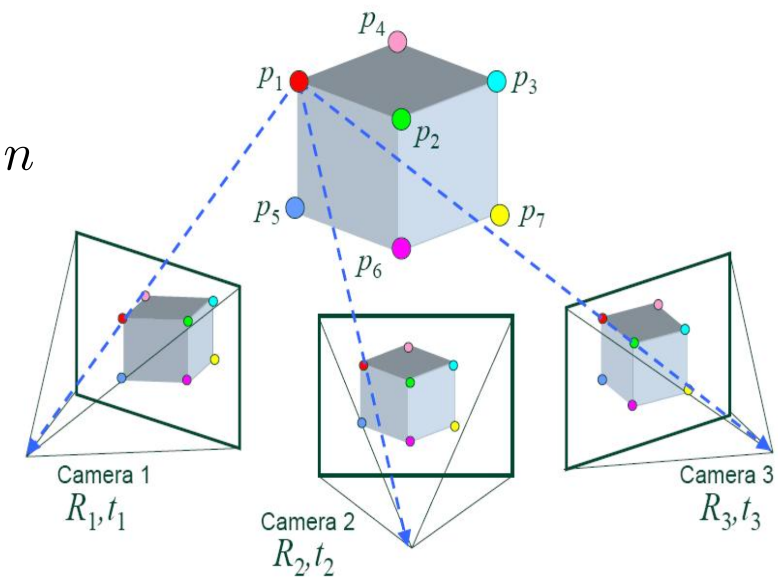
$$\mathbf{x}_j^{(i)} \propto \mathtt{P}_i \mathbf{X}_j$$

Input

$$\mathbf{x}_j^{(i)} = \begin{bmatrix} x_j^{(i)} & y_j^{(i)} & 1 \end{bmatrix}^\top$$

$$i = 1, \ldots, m, \quad j = 1, \ldots, n$$

$$2mn$$



Output

$$\mathbf{X}_j = \begin{bmatrix} X_j & Y_j & Z_j & 1 \end{bmatrix}^\top$$

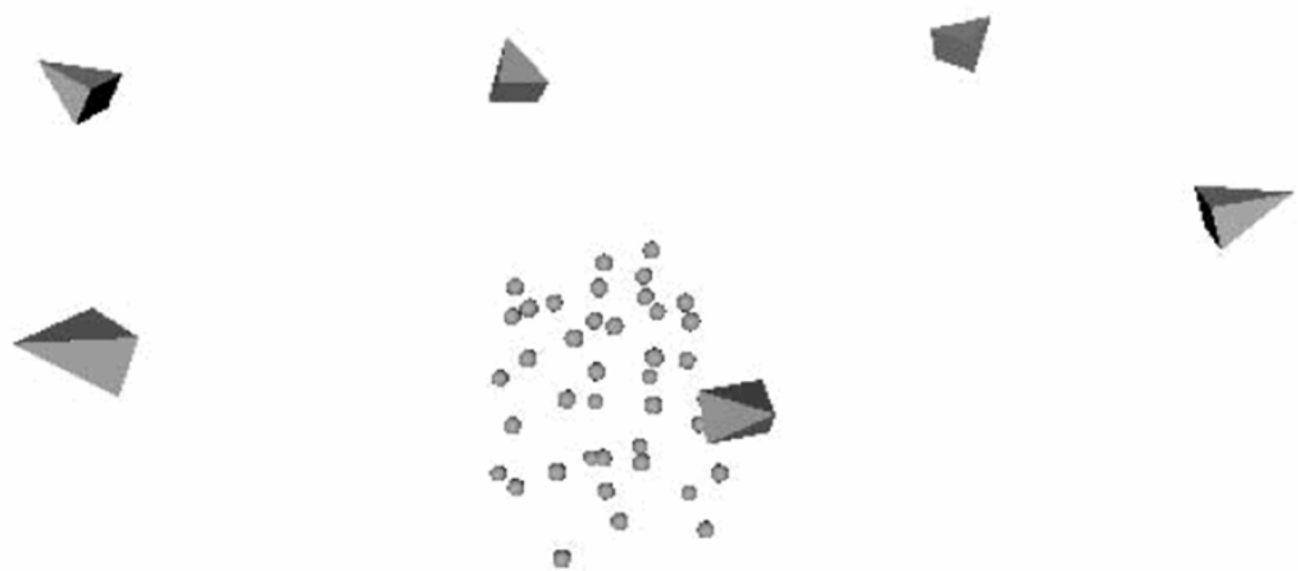
$$3n$$

$$\mathtt{P}_i = \mathtt{K}_i \begin{bmatrix} \mathtt{R}_i & \mathbf{t}_i \end{bmatrix}$$

$$11m$$

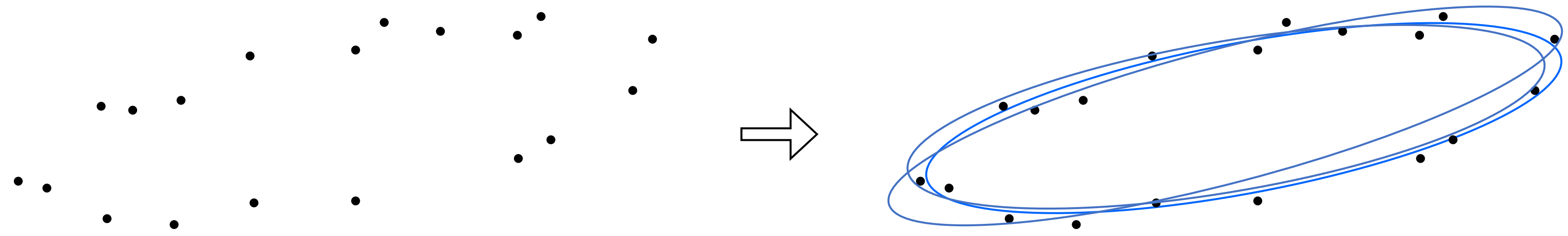
Example: Calibration of multiple surveillance cameras

- Observation: point correspondences among multi-view images
- Parameters to estimate: Poses and focal lengths of the cameras plus the 3D coordinates of scene points



$$E(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} (u_{ij} - \hat{u}_{ij})^2 + (v_{ij} - \hat{v}_{ij})^2$$

Example: Fitting an ellipse to a set of points



- Observation: points
- To estimate: the shape of ellipse and the true positions of the points

$$E(\mathbf{x}) = \sum_{i=1}^{m} \sum_{j=1}^{n} (u_{ij} - \hat{u}_{ij})^2 + (v_{ij} - \hat{v}_{ij})^2$$

Minimization of $E(\mathbf{x})$

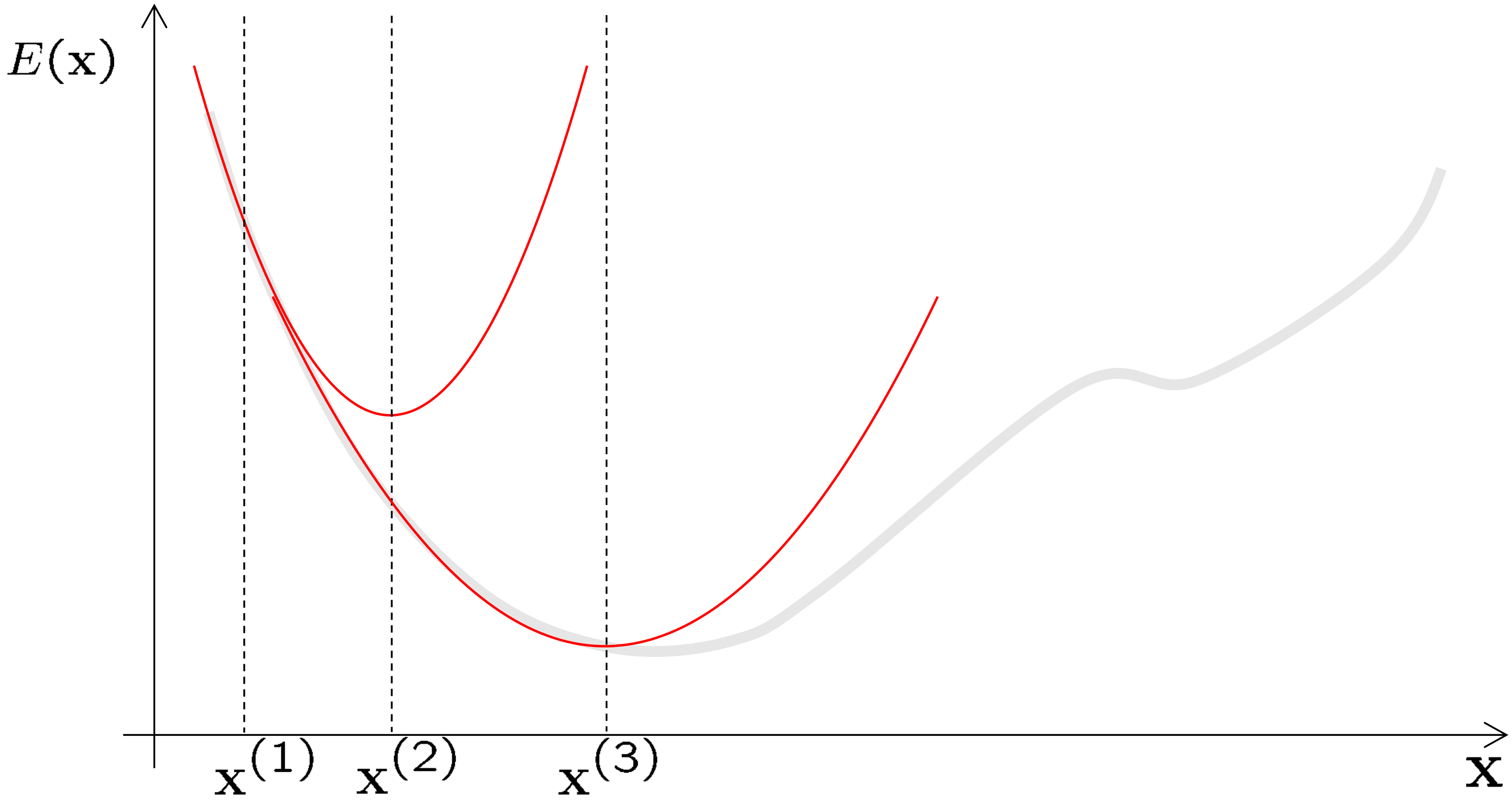
- Any numerical method for minimization can generally be used
  - Family of Newton's method
  - Other approaches
- Reprojection error has the form of a sum of squares; nonlinear least squares methods are a natural choice

$$E(\mathbf{x}) = \frac{1}{2}\|\mathbf{e}(\mathbf{x})\|^2 = \frac{1}{2}\sum_k e_k^2$$

- Standard methods:
  - Gauss-Newton method
  - Levenberg-Marquardt method

Using Newton’s method for minimization

- We approximate the local shape of E(x) by a quadratic function; then we find the minimum of the quadratic function
- Starting an initial value, we repeat this procedure until convergence



Using Newton's method for minimization

- We approximate the local shape of E(x) by a quadratic function; then we find the minimum of the quadratic function
- Starting an initial value, we repeat this procedure until convergence

$$E(\mathbf{x}+\delta\mathbf{x}) \approx E(\mathbf{x}) + \mathbf{g}^\top \delta\mathbf{x} + \frac{1}{2}\delta\mathbf{x}^\top \mathtt{H}\delta\mathbf{x}$$

Gradient: $\mathbf{g} = \left.\frac{dE}{d\mathbf{x}}\right|_{\mathbf{x}}$ Hessian: $\mathtt{H} = \left.\frac{d^2E}{d\mathbf{x}^2}\right|_{\mathbf{x}}$

Compute the gradient and Hessian

Update solution $\mathbf{x} + \delta\mathbf{x} \rightarrow \mathbf{x}$

$$\delta\mathbf{x} = -\mathtt{H}^{-1}\mathbf{g}$$

Gauss-Newton method

$$E(\mathbf{x}) = \frac{1}{2}\|\mathbf{e}(\mathbf{x})\|^2$$

- Approximates the Hessian utilizing E being a sum of squares

$$\mathtt{H} = \left.\frac{d^2 E}{d\mathbf{x}^2}\right|_{\mathbf{x}} = \left(\frac{d\mathbf{e}}{d\mathbf{x}}\right)^\top \left(\frac{d\mathbf{e}}{d\mathbf{x}}\right) + \left(\frac{d}{d\mathbf{x}}\left(\frac{d\mathbf{e}}{d\mathbf{x}}\right)^\top\right)\mathbf{e} \approx \left(\frac{d\mathbf{e}}{d\mathbf{x}}\right)^\top \left(\frac{d\mathbf{e}}{d\mathbf{x}}\right)$$

$$\left(\mathbf{g} = \frac{dE}{d\mathbf{x}} = \left(\frac{d\mathbf{e}}{d\mathbf{x}}\right)^\top \mathbf{e}\right)$$

- Or equivalently

$$\mathtt{H} \approx \mathtt{A} \equiv \mathtt{J}^\top \mathtt{J} \quad \text{where} \quad \mathtt{J} = \left.\frac{\mathrm{d}\mathbf{e}}{\mathrm{d}\mathbf{x}}\right|_{\mathbf{x}}$$

Parameter update: $\mathtt{A}\delta\mathbf{x} = -\mathbf{g}$

Levenberg-Marquardt method

- Update the parameter with

[Levenberg1944, Marquardt1963]

$$(\mathtt{A} + \lambda \mathtt{I})\delta\mathbf{x} = -\mathbf{g}$$

- λ is called the damping factor
  - $\lambda = 0$ → coincides with the Gauss-Newton update
  - $\lambda = \infty$ → coincides with the steepest descent method

- λ is adaptively chosen depending on if E(x) decreases
  - Increase λ when $E(\mathbf{x} + \delta\mathbf{x}) > E(\mathbf{x})$
  - Decrease λ when $E(\mathbf{x} + \delta\mathbf{x}) < E(\mathbf{x})$

Updating parameters

- Parameter update → Solve a linear equation $\mathtt{A}\delta\mathbf{x} = \mathbf{a}$
- Matrix inverse should not be used; inefficient, inaccurate

$$\mathtt{A}\delta\mathbf{x} = \mathbf{a} \quad \Rightarrow \quad \delta\mathbf{x} = \mathtt{A}^{-1}\mathbf{a}$$

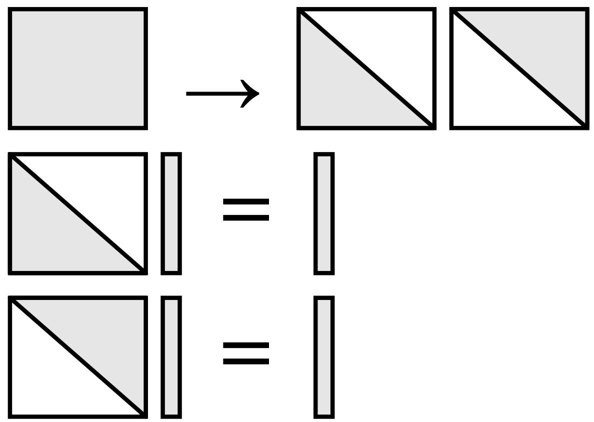
(NG!)

- A standard approach = Cholesky decomposition + forward/backward substitution

1) $\mathtt{A} = \mathtt{L}\mathtt{L}^\top \quad \left(\mathtt{L}(\mathtt{L}^\top\delta\mathbf{x}) = \mathbf{a}\right)$

2) $\mathtt{L}\mathbf{b} = \mathbf{a}$

3) $\mathtt{L}^\top\delta\mathbf{x} = \mathbf{b}$



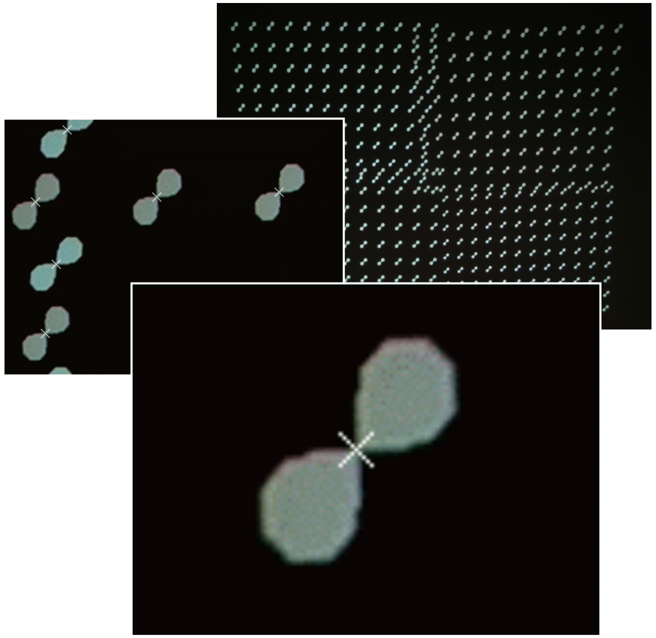
- Remark: Preconditioned Conjugate Gradient (PCG) method is preferred when A is very large

Statistical explanation of bundle adjustment

- Observation inevitably has errors

$$\boxed{\begin{matrix} u_i \\ v_i \end{matrix}} = \boxed{\begin{matrix} \bar{u}_i \\ \bar{v}_i \end{matrix}} + \boxed{\begin{matrix} \varepsilon_i \\ \varepsilon'_i \end{matrix}}$$

observation true val. error

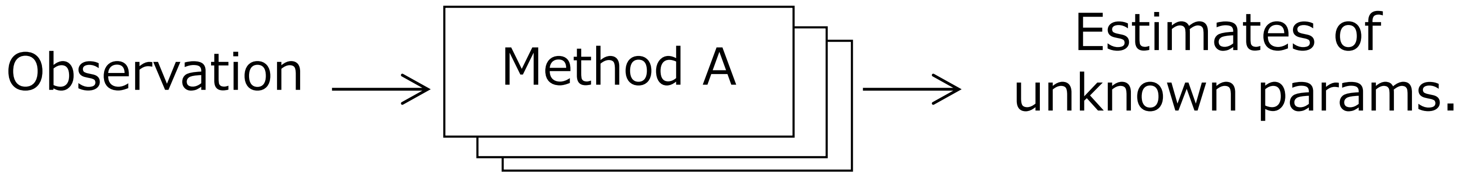


- Statistical model of the errors
  - They are random variables following a Gaussian distribution

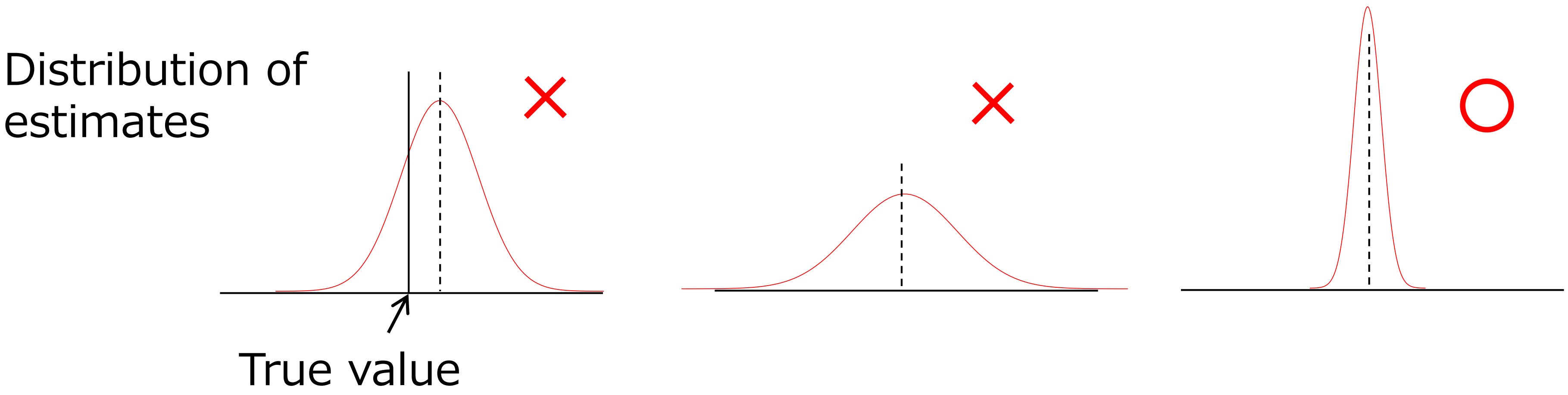
$$\varepsilon_i, \varepsilon'_i \sim N(0, \sigma^2)$$

What is good estimation?

- There are an infinite number of estimating methods



- A shared view of “What’s good estimate?”
  - Mean of the estimates coincides with the true value
  - Variance of the estimates is as small as possible
    - There indeed exists a theoretical lower bound for the variance, named the Cramer-Rao lower bound



Maximum likelihood estimation

- Maximum likelihood = Selects the parameter value that maximizes the likelihood as an estimate:

$$\hat{\mathbf{x}} = \operatorname*{argmax}_{\mathbf{x}} L(\mathbf{x}; \mathbf{Z})$$

- The likelihood is defined as

$$L(\underbrace{\mathbf{x}}_{\text{parameter}}; \underbrace{\mathbf{Z}}_{\text{all observations}}) \equiv p(\mathbf{Z}; \mathbf{x})$$

- The estimate is called the maximum likelihood estimate
- It can be shown that it attains the Cramer-Rao bound asymptotically as the number of observations goes to infinity

Bundle adjustment as maximum likelihood

Model of an observation:

$$\mathbf{z}_i = \bar{\mathbf{z}}_i + \varepsilon_i \quad \text{where the error} \quad \varepsilon_i \sim N(0, \sigma^2 \mathtt{I})$$

Then, the PDF of an observation is given by

$$p(\mathbf{z}_i) = C \exp\left\{ -\frac{(\mathbf{z}_i - \bar{\mathbf{z}}_i)^2}{2\sigma^2} \right\} \qquad C = \frac{1}{(2\pi)^{d/2}\sigma^d}$$

The PDF of all observations $\mathbf{z}_1, \ldots, \mathbf{z}_N$

$$p(\mathbf{z}_1, \ldots, \mathbf{z}_N) = \prod_{i=1}^{N} C \exp\left\{ -\frac{(\mathbf{z}_i - \bar{\mathbf{z}}_i)^2}{2\sigma^2} \right\}$$

Maximize the likelihood

$$L(\bar{\mathbf{z}}_1, \ldots, \bar{\mathbf{z}}_N; \mathbf{z}_1, \ldots, \mathbf{z}_N)$$

is equivalent to minimizing negative log-likelihood: $\sum_{i=1}^{N} (\mathbf{z}_i - \bar{\mathbf{z}}_i)^2$