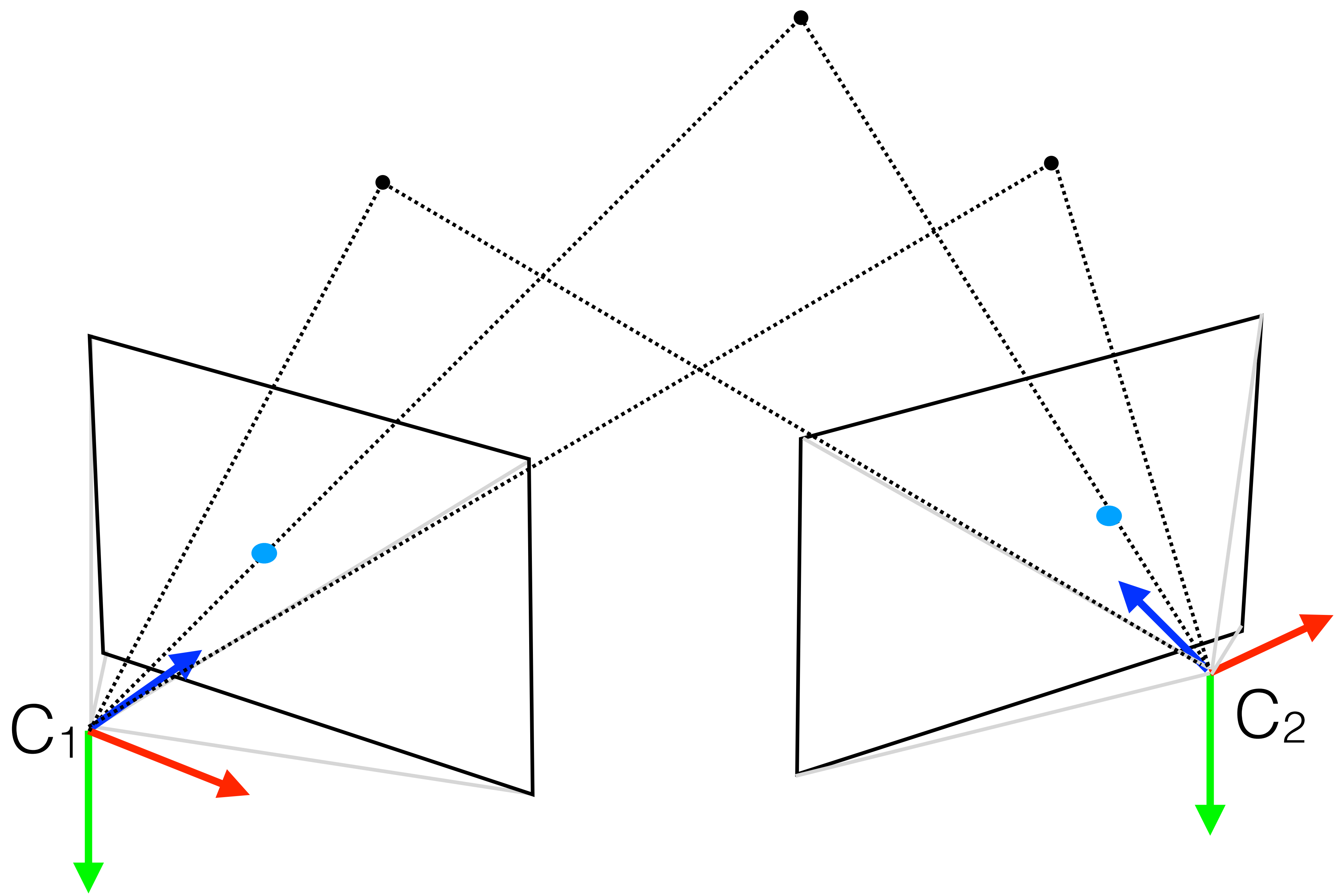
16.485: VNAV - Visual Navigation for Autonomous Vehicles

Luca Carlone

Lecture 16: From Optimization To Estimation Theory and Back

Recap: 2-view Geometry from 3D-3D Correspondences



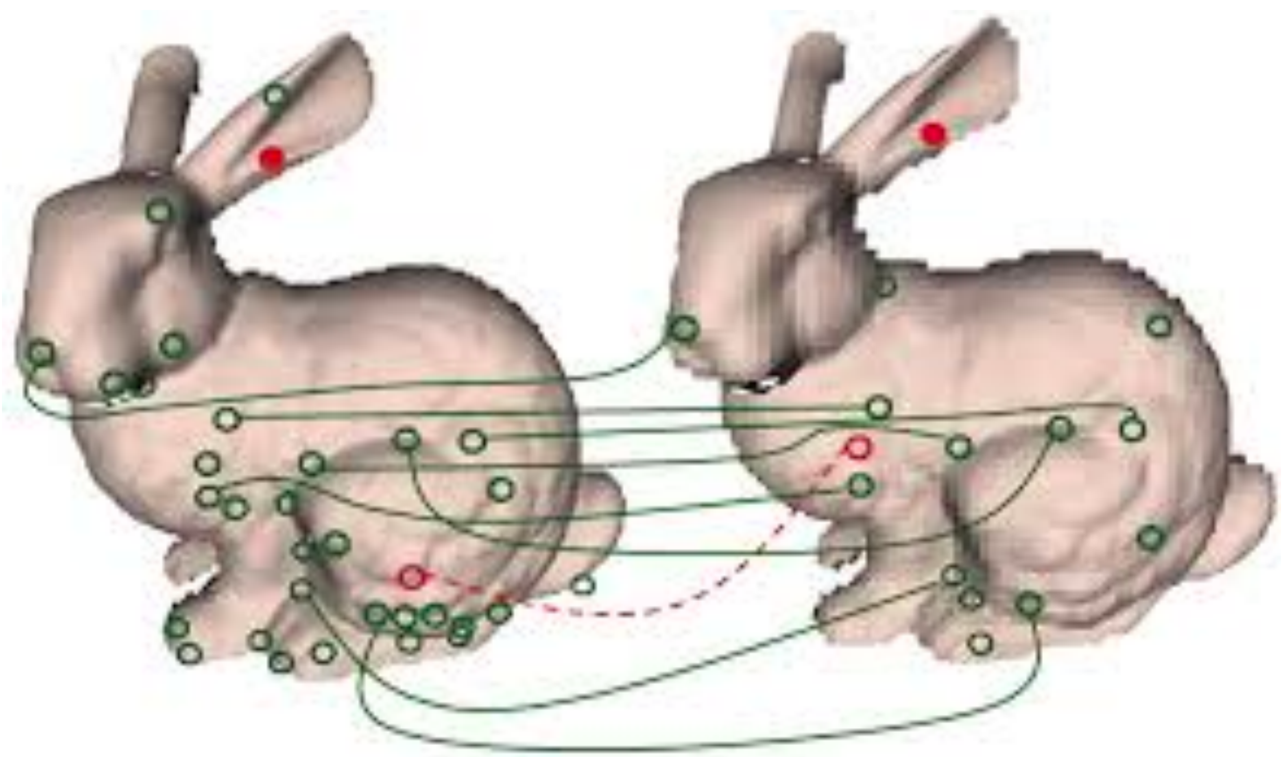
How to estimate the relative pose between the cameras from 3D-3D correspondences $(\boldsymbol{p}_{1,k}, \boldsymbol{p}_{2,k})$ **with** $k = 1, \ldots, N$ **?**

Few More Comments:

3 points are sufficient to compute the relative pose from 3D-3D correspondences

We can use the solver seen today as a 3-point minimal solver within a **RANSAC** method

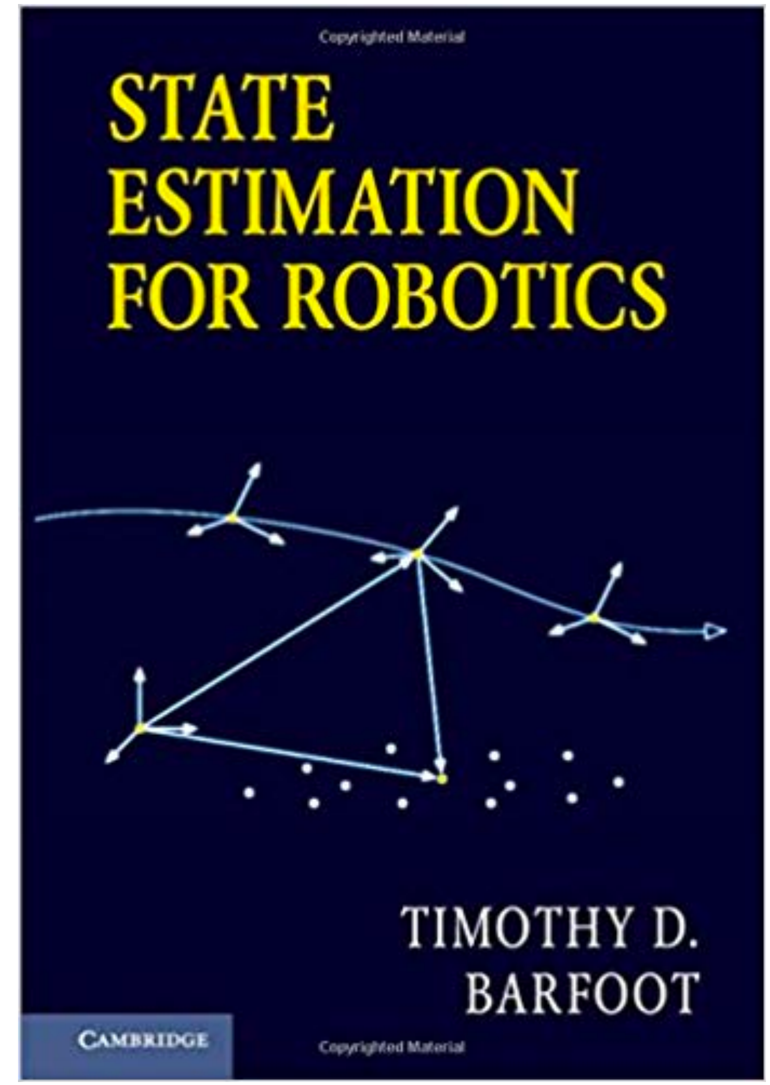
Also useful for 3D objects localization:



Other names: vector registration, point cloud alignment, ..

Today

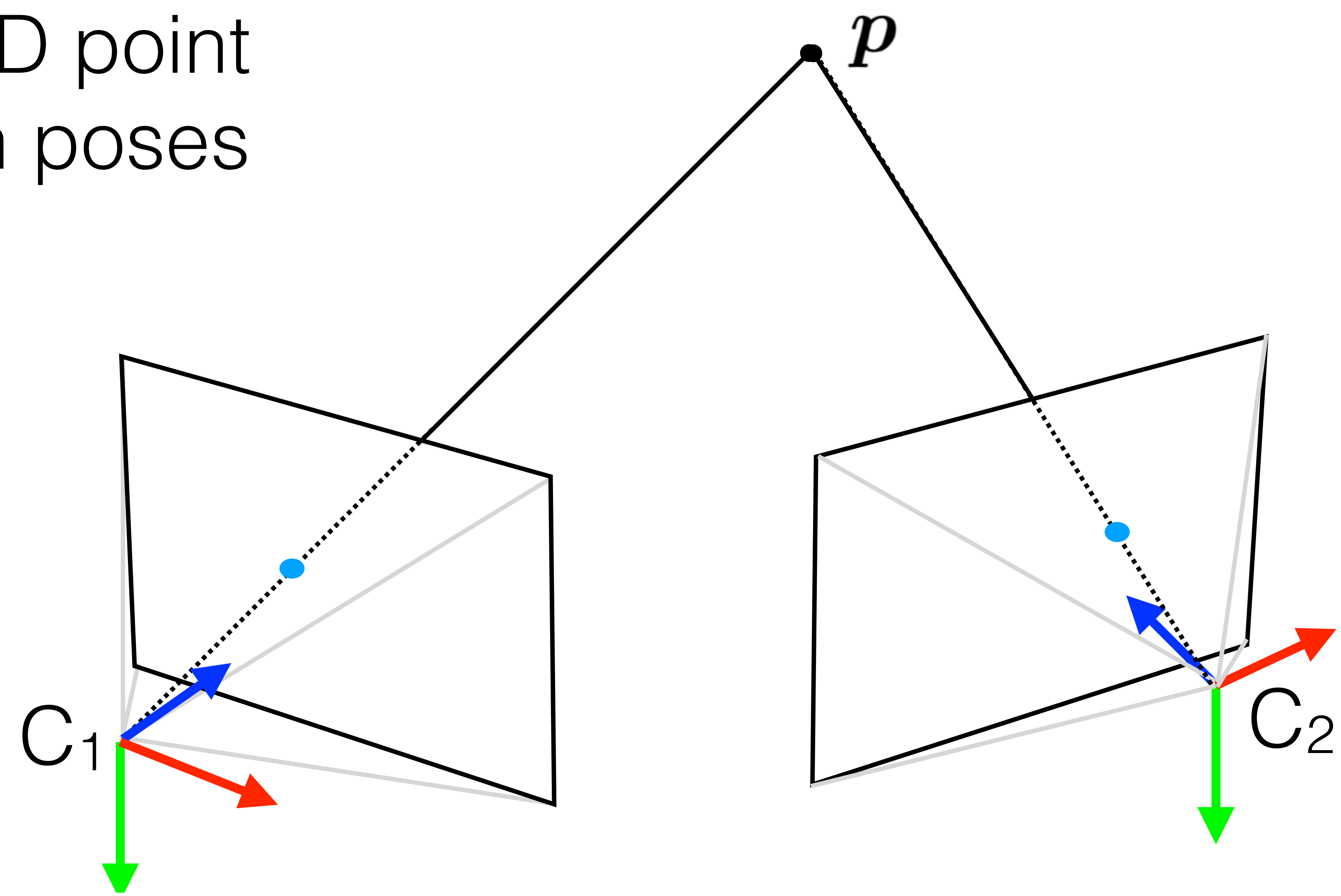
- Optimization examples
- Estimation Basics



Part I: Estimation Machinery
(more than what we need)

Example **1a**: Triangulation (Structure Reconstruction)

Compute 3D point from known poses

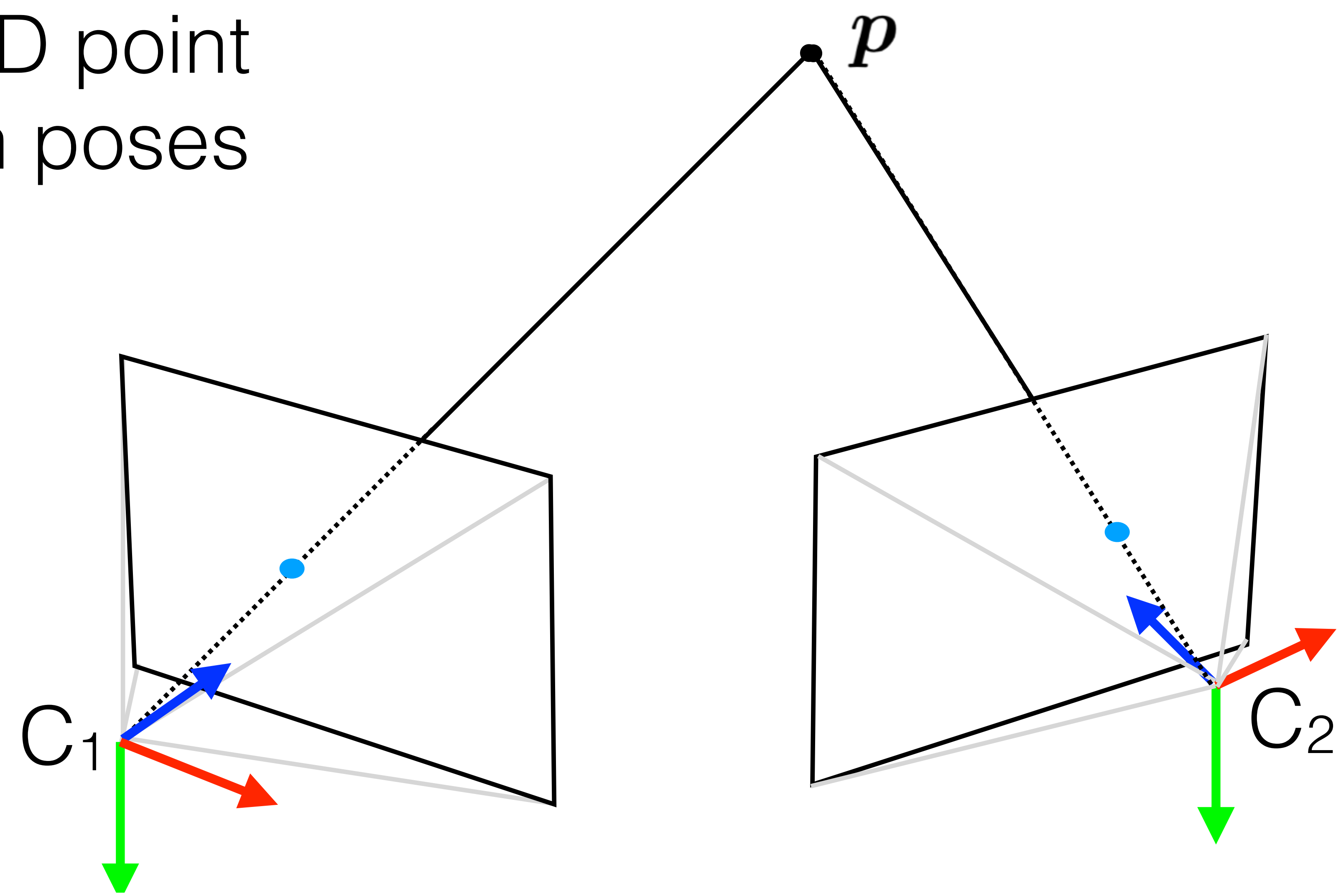


$$\lambda_1 \tilde{\boldsymbol{x}}_1 = \underbrace{\boldsymbol{K}_1[\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}_1}\ \boldsymbol{t}_{\mathrm{w}}^{\mathrm{c}_1}]}_{\boldsymbol{\Pi}_1}\tilde{\boldsymbol{p}}^{\mathrm{w}} \qquad \lambda_2 \tilde{\boldsymbol{x}}_2 = \underbrace{\boldsymbol{K}_2[\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}_2}\ \boldsymbol{t}_{\mathrm{w}}^{\mathrm{c}_2}]}_{\boldsymbol{\Pi}_2}\tilde{\boldsymbol{p}}^{\mathrm{w}}$$

Linear triangulation: $\min_{\|\tilde{\boldsymbol{p}}^{\mathrm{w}}\|=1} \|\boldsymbol{A}\tilde{\boldsymbol{p}}^{\mathrm{w}}\|^2$

Example **1b**: Triangulation (Structure Reconstruction)

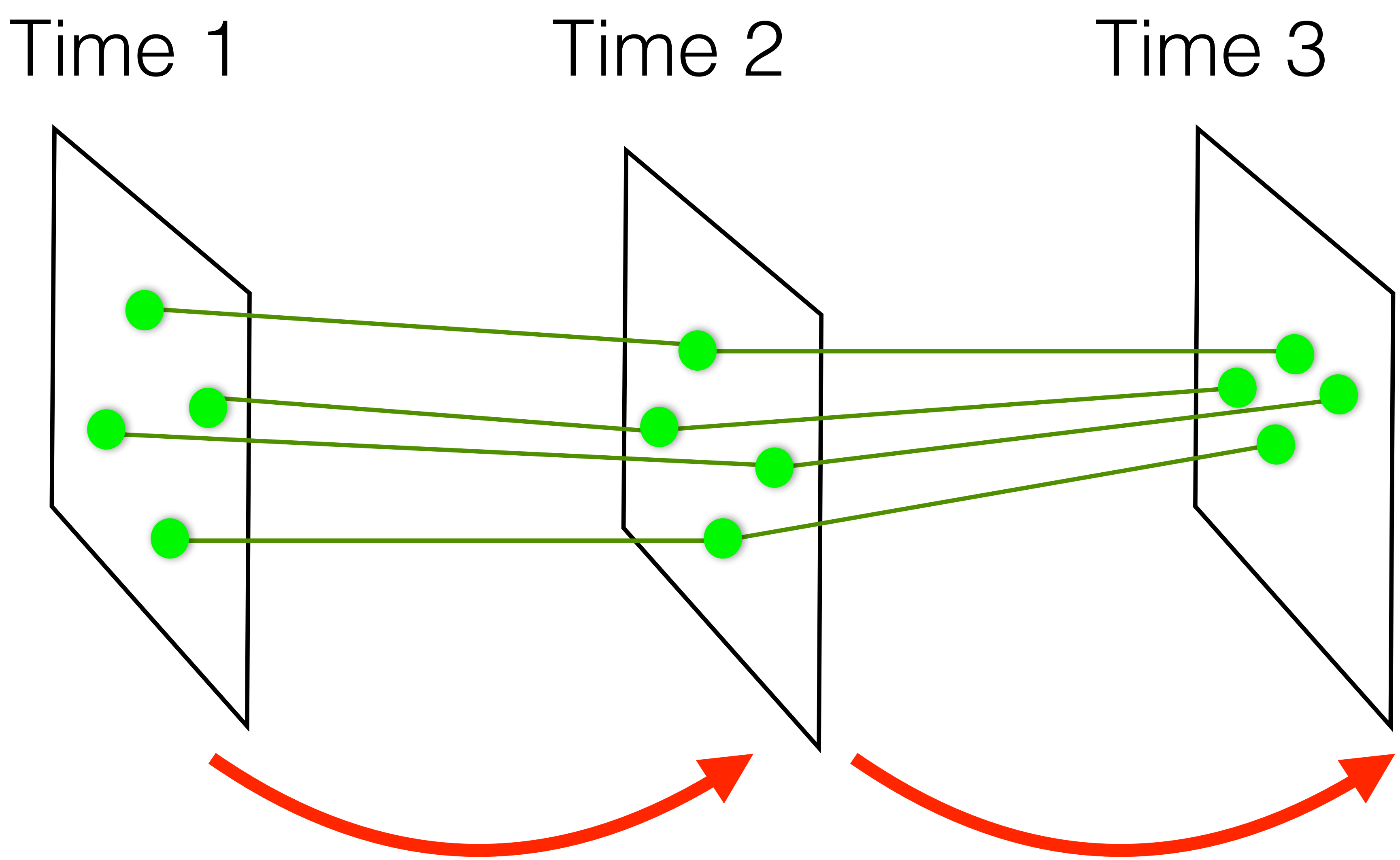
Compute 3D point from known poses



$$\lambda_1 \tilde{\boldsymbol{x}}_1 = \underbrace{\boldsymbol{K}_1 [\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}_1} \ \boldsymbol{t}_{\mathrm{w}}^{\mathrm{c}_1}]}_{\boldsymbol{\Pi}_1} \tilde{\boldsymbol{p}}^{\mathrm{w}} \qquad \lambda_2 \tilde{\boldsymbol{x}}_2 = \underbrace{\boldsymbol{K}_2 [\boldsymbol{R}_{\mathrm{w}}^{\mathrm{c}_2} \ \boldsymbol{t}_{\mathrm{w}}^{\mathrm{c}_2}]}_{\boldsymbol{\Pi}_2} \tilde{\boldsymbol{p}}^{\mathrm{w}}$$

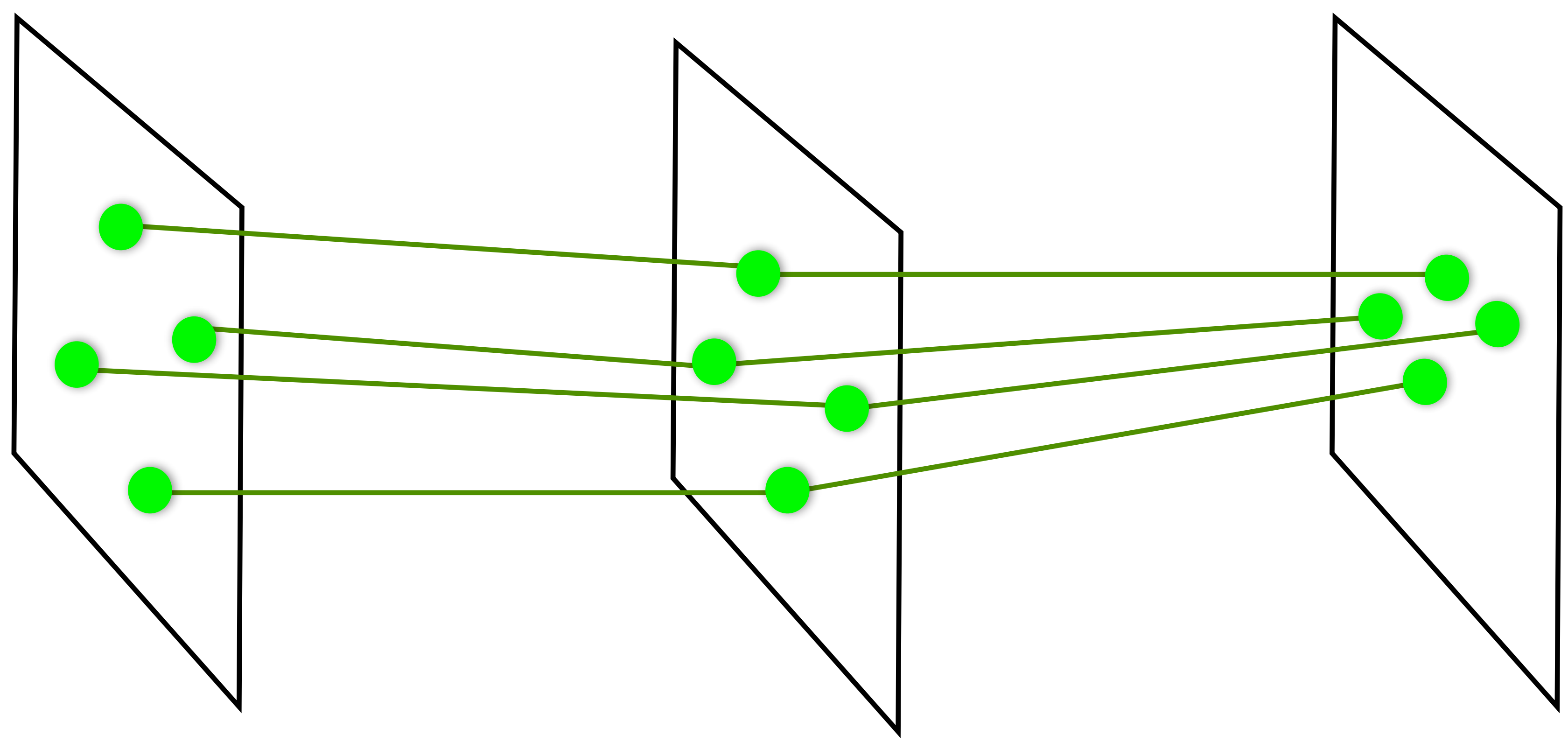
$$\min_{\boldsymbol{p}^{\mathrm{w}}} \|\boldsymbol{x}_1 - \pi(\boldsymbol{R}_{\mathrm{c}_1}^{\mathrm{w}}, \boldsymbol{t}_{\mathrm{c}_1}^{\mathrm{w}}, \boldsymbol{p}^{\mathrm{w}})\|^2 + \|\boldsymbol{x}_2 - \pi(\boldsymbol{R}_{\mathrm{c}_2}^{\mathrm{w}}, \boldsymbol{t}_{\mathrm{c}_2}^{\mathrm{w}}, \boldsymbol{p}^{\mathrm{w}})\|^2$$

Example **2a**: Motion Estimation



$$\boldsymbol{E}_{12} = \underset{\boldsymbol{E}_{12} \in \mathcal{S}_E}{\arg\min} \sum_{k=1}^{N} |\tilde{\boldsymbol{y}}_{k,2}^{\mathsf{T}} \boldsymbol{E}_{12} \tilde{\boldsymbol{y}}_{k,1}|^2 \qquad \boldsymbol{E}_{23} = \underset{\boldsymbol{E}_{23} \in \mathcal{S}_E}{\arg\min} \sum_{k=1}^{N} |\tilde{\boldsymbol{y}}_{k,3}^{\mathsf{T}} \boldsymbol{E}_{23} \tilde{\boldsymbol{y}}_{k,2}|^2$$

Example **2b**: Motion Estimation

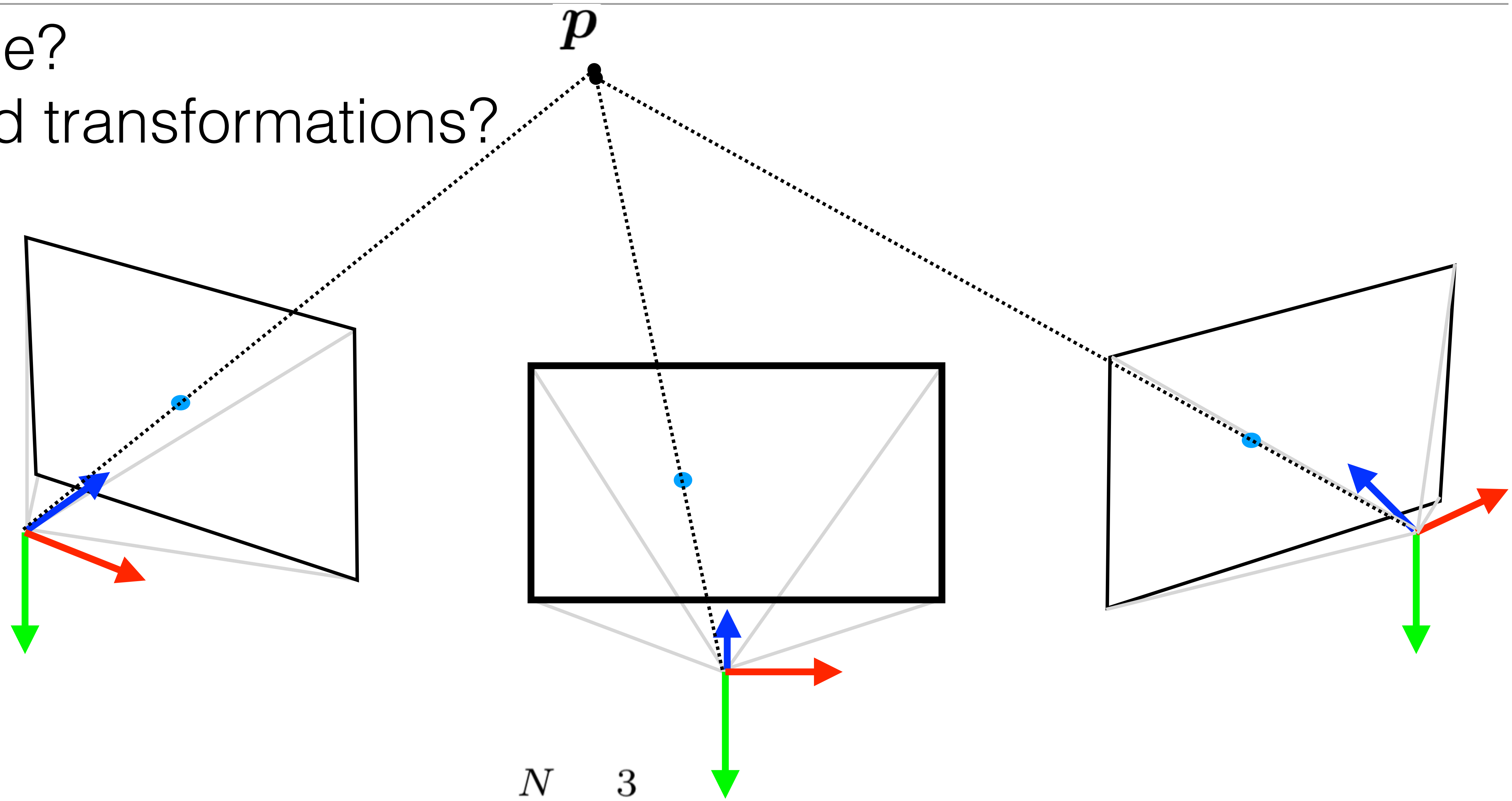


$$\min_{\substack{(\boldsymbol{R}_{\mathrm{c}_i}^{\mathrm{w}}, \boldsymbol{t}_{\mathrm{c}_i}^{\mathrm{w}}), i=1,2,3 \\ \boldsymbol{p}_k^{\mathrm{w}}, k=1,\ldots,N}} \sum_{k=1}^{N} \sum_{i=1}^{3} \| \boldsymbol{x}_{k,i} - \pi(\boldsymbol{R}_{\mathrm{c}_i}^{\mathrm{w}}, \boldsymbol{t}_{\mathrm{c}_i}^{\mathrm{w}}, \boldsymbol{p}_k^{\mathrm{w}}) \|^2$$

Generalizes to K cameras: **Bundle adjustment**

Example **2b**: Motion **and Structure** Estimation

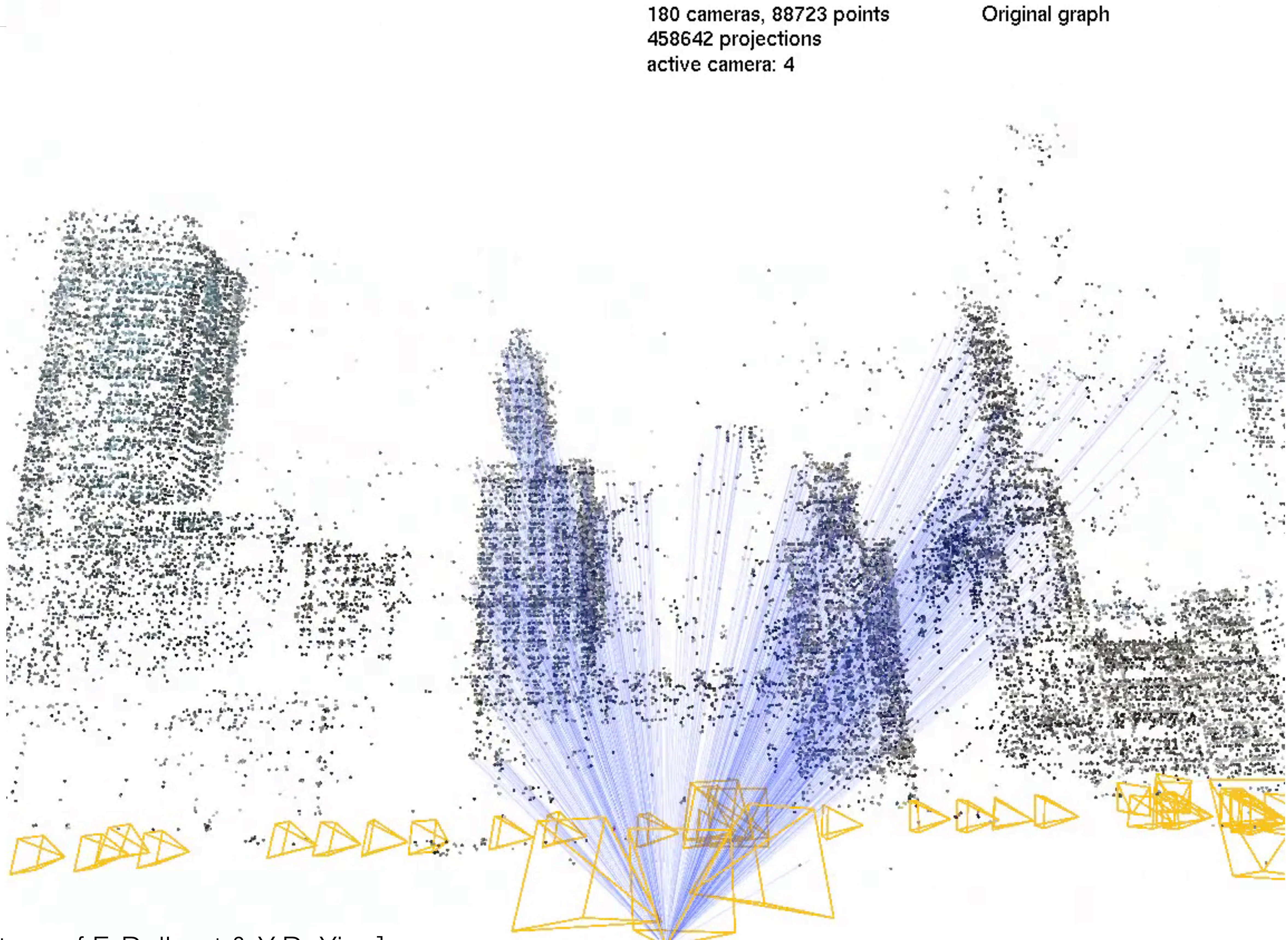
Scale?
Rigid transformations?



$$\min_{\substack{(\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}_i}, \boldsymbol{t}^{\mathrm{w}}_{\mathrm{c}_i}), i=1,2,3 \\ \boldsymbol{p}^{\mathrm{w}}_k, k=1,\ldots,N}} \sum_{k=1}^{N} \sum_{i=1}^{3} \| \boldsymbol{x}_{k,i} - \pi(\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}_i}, \boldsymbol{t}^{\mathrm{w}}_{\mathrm{c}_i}, \boldsymbol{p}^{\mathrm{w}}_k) \|^2$$

Generalizes to K cameras: **Bundle adjustment**

Structure from Motion



[courtesy of F. Dellaert & Y-D. Yian]

Estimation Theory

Concerned with the estimation of unknown variables given (noisy) measurements and prior information

Estimator: a function of the measurements that approximates the unknown variables

Measurements that depend on some unknown variable **x**:

$$z_1, \ldots, z_N$$

Estimator for **x**:

$$x^\star = \mathcal{F}(z_1, \ldots, z_N)$$

$$x^\star \approx x$$

Maximum Likelihood Estimation (MLE)

Assume we are given N measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ (e.g., pixel measurements) that are function of a variable we want to estimate $\boldsymbol{x}$ (e.g., camera poses, points). Assume that we are also given the conditional distributions:

$$\mathbb{P}\left(\boldsymbol{z}_j | \boldsymbol{x}\right)$$

Than the *maximum likelihood* estimator (MLE) is defined as:

$$\boldsymbol{x}_{\text{MLE}} = \arg\max_{\boldsymbol{x}} \underline{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}$$

Measurement likelihood

where $\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)$ is also called the *likelihood* of the measurements given $\boldsymbol{x}$. Equivalently:

$$\boldsymbol{x}_{\text{MLE}} = \arg\min_{\boldsymbol{x}} \underline{-\log \mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}$$

Negative log-likelihood

Maximum Likelihood Estimation (MLE)

Assume we are given N measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ (e.g., pixel measurements) that are function of a variable we want to estimate $\boldsymbol{x}$ (e.g., camera poses, points). Assume that we are also given the conditional distributions:

$$\mathbb{P}\left(\boldsymbol{z}_j | \boldsymbol{x}\right)$$

Than the *maximum likelihood* estimator (MLE) is defined as:

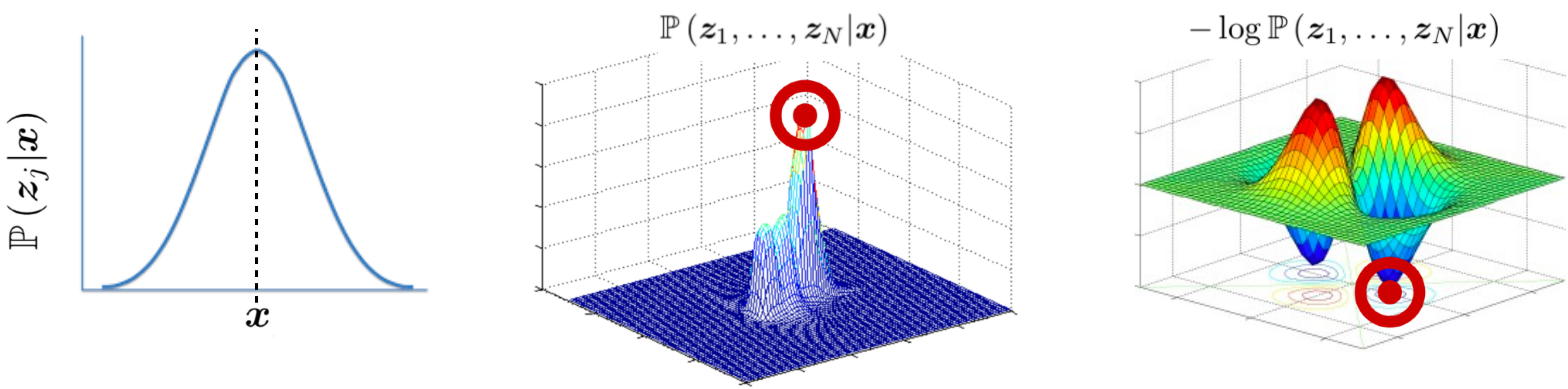
$$\boldsymbol{x}_{\text{MLE}} = \arg\max_{\boldsymbol{x}} \underline{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}$$

Measurement likelihood

where $\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)$ is also called the *likelihood* of the measurements given $\boldsymbol{x}$. Equivalently:

$$\boldsymbol{x}_{\text{MLE}} = \arg\min_{\boldsymbol{x}} \underline{-\log \mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}$$

Negative log-likelihood



Maximum a Posteriori Estimation (MAP)

Assume we are given N measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ (e.g., pixel measurements) that are function of a variable we want to estimate $\boldsymbol{x}$ (e.g., camera poses, points). *Maximum a Posteriori Estimation* (MAP) is a generalization of MLE. Then the MAP estimator is:

$$\underbrace{\arg\max_{\boldsymbol{x}} \mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}_{\text{Measurement likelihood}} \underbrace{\mathbb{P}\left(\boldsymbol{x}\right)}_{\text{Priors}}$$

Measurement likelihood

Priors

Maximum a Posteriori Estimation (MAP)

Assume we are given N measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ (e.g., pixel measurements) that are function of a variable we want to estimate $\boldsymbol{x}$ (e.g., camera poses, points). *Maximum a Posteriori Estimation* (MAP) is a generalization of MLE. Then the MAP estimator is:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \mathbb{P}\left(\boldsymbol{x} | \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right)$$

Using Bayes rule:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \mathbb{P}\left(\boldsymbol{x} | \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right) =$$

$$\arg\max_{\boldsymbol{x}} \frac{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right) \mathbb{P}\left(\boldsymbol{x}\right)}{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right)} =$$

$$\arg\max_{\boldsymbol{x}} \underbrace{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}_{\text{Measurement likelihood}} \underbrace{\mathbb{P}\left(\boldsymbol{x}\right)}_{\text{Priors}}$$

Measurement likelihood

Priors

Maximum a Posteriori Estimation (MAP)

Assume we are given N measurements $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N$ (e.g., pixel measurements) that are function of a variable we want to estimate $\boldsymbol{x}$ (e.g., camera poses, points). *Maximum a Posteriori Estimation* (MAP) is a generalization of MLE. Then the MAP estimator is:

$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \mathbb{P}\left(\boldsymbol{x} | \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right)$$

Using Bayes rule:

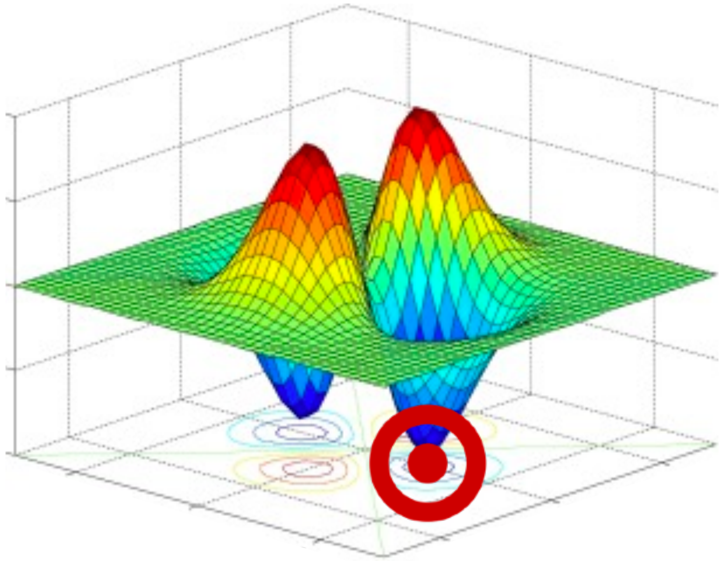
$$\boldsymbol{x}_{\text{MAP}} = \arg\max_{\boldsymbol{x}} \mathbb{P}\left(\boldsymbol{x} | \boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right) = \arg\max_{\boldsymbol{x}} \frac{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right) \mathbb{P}\left(\boldsymbol{x}\right)}{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\right)} = \arg\max_{\boldsymbol{x}} \underbrace{\mathbb{P}\left(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N | \boldsymbol{x}\right)}_{\text{Measurement likelihood}} \underbrace{\mathbb{P}\left(\boldsymbol{x}\right)}_{\text{Priors}}$$

Measurement likelihood

Priors

Assuming independence between measurements:

$$\boldsymbol{x}_{\text{MAP}} = \arg\min_{\boldsymbol{x}} - \sum_{j=1}^{N} \log \mathbb{P}\left(\boldsymbol{z}_j | \boldsymbol{x}\right) - \log \mathbb{P}\left(\boldsymbol{x}\right)$$



Optimization

Linear triangulation:

$$\min_{\|\tilde{\boldsymbol{p}}^{\mathrm{w}}\|=1} \|\boldsymbol{A}\tilde{\boldsymbol{p}}^{\mathrm{w}}\|^2$$

Nonlinear triangulation:

$$\min_{\boldsymbol{p}^{\mathrm{w}}} \|\boldsymbol{x}_1 - \pi(\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}_1}, \boldsymbol{t}^{\mathrm{w}}_{\mathrm{c}_1}, \boldsymbol{p}^{\mathrm{w}})\|^2 +$$

$$+ \|\boldsymbol{x}_2 - \pi(\boldsymbol{R}^{\mathrm{w}}_{\mathrm{c}_2}, \boldsymbol{t}^{\mathrm{w}}_{\mathrm{c}_2}, \boldsymbol{p}^{\mathrm{w}})\|^2$$