NC STATE UNIVERSITY

Paper number: 1376

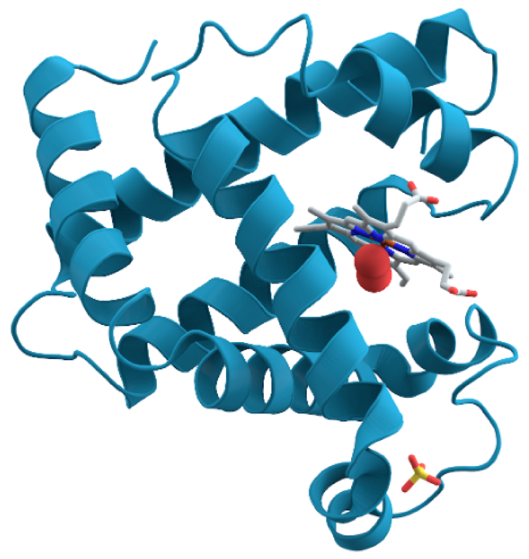
Mismatched Estimation in the Distance Geometry Problem

Mahmoud Abdelkhalek, Dror Baron, Chau-Wai Wong

Electrical and Computer Engineering, North Carolina State University

Overview

- The Distance Geometry Problem (DGP): determine the locations of points in Euclidean space given noisy pairwise distance measurements.
- Applications of the DGP include:



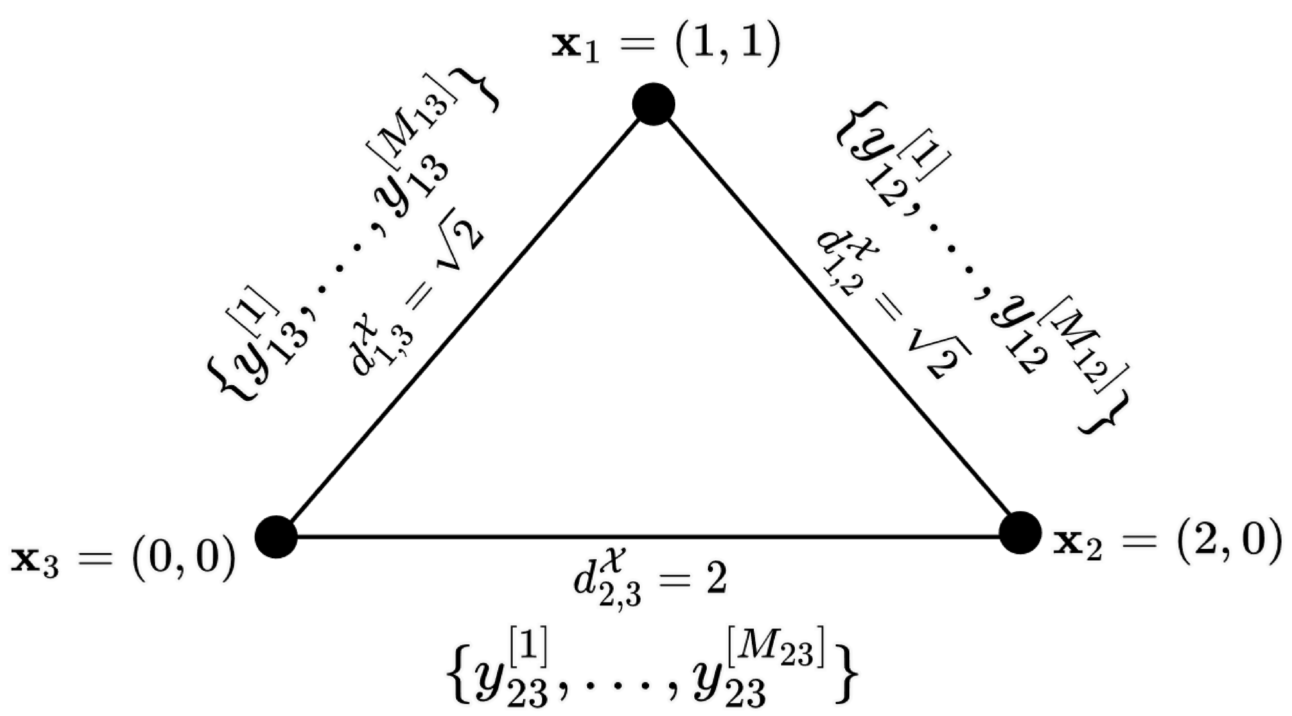
Computational Biology

Wireless Networks

Robotics

Problem Formulation

- The DGP consists of:
  - $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$
  - $\mathcal{E}(\mathcal{X}) = \{(i,j) \mid (i,j) \in \{1,\ldots,N\}^2, i<j\}$
  - $d_{ij}^{\mathcal{X}} = \|\mathbf{x}_i - \mathbf{x}_j\|_2$
  - $\left\{\left\{y_{ij}^{[1]},\ldots,y_{ij}^{[M_{ij}]}\right\} \;\middle|\; (i,j) \in \mathcal{E}(\mathcal{X})\right\}$
- For example:



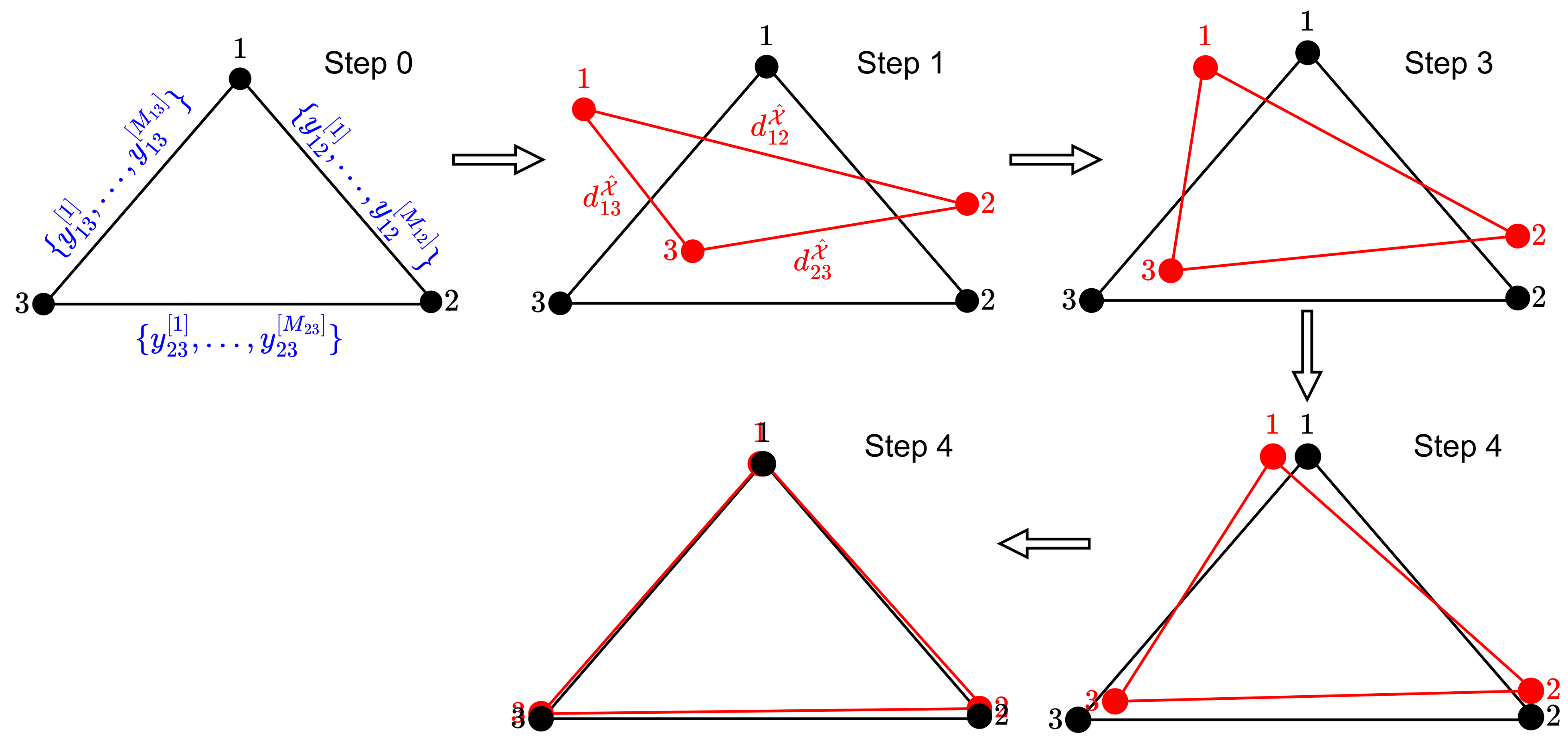
- The objective is to estimate $\mathcal{X}$ given noisy measurements of the lengths of edges in $\mathcal{E}(\mathcal{X})$.

Estimating the locations of the points

- Given noisy measurements associated with edges (shown in blue in step 0), common approach to estimating $\mathcal{X}$ follows these steps:
  1. Make an initial guess and compute its edge lengths.
  2. Compute the sum of squared errors (SSE) as

$$\sum_{(i,j)\in\mathcal{E}(\mathcal{X})} \sum_{m=1}^{M_{ij}} \left(y_{ij}^{[m]} - d_{ij}^{\hat{\mathcal{X}}}\right)^2$$

  3. Adjust the locations of the points to minimize this squared error.
  4. Repeat until convergence.



Maximum Likelihood Estimation

- Minimizing SSE cost function implies Gaussian noise assumption; what if measurement noise isn't Gaussian?
- Locations of points in $\mathcal{X}$ can be estimated by maximizing the likelihood function:

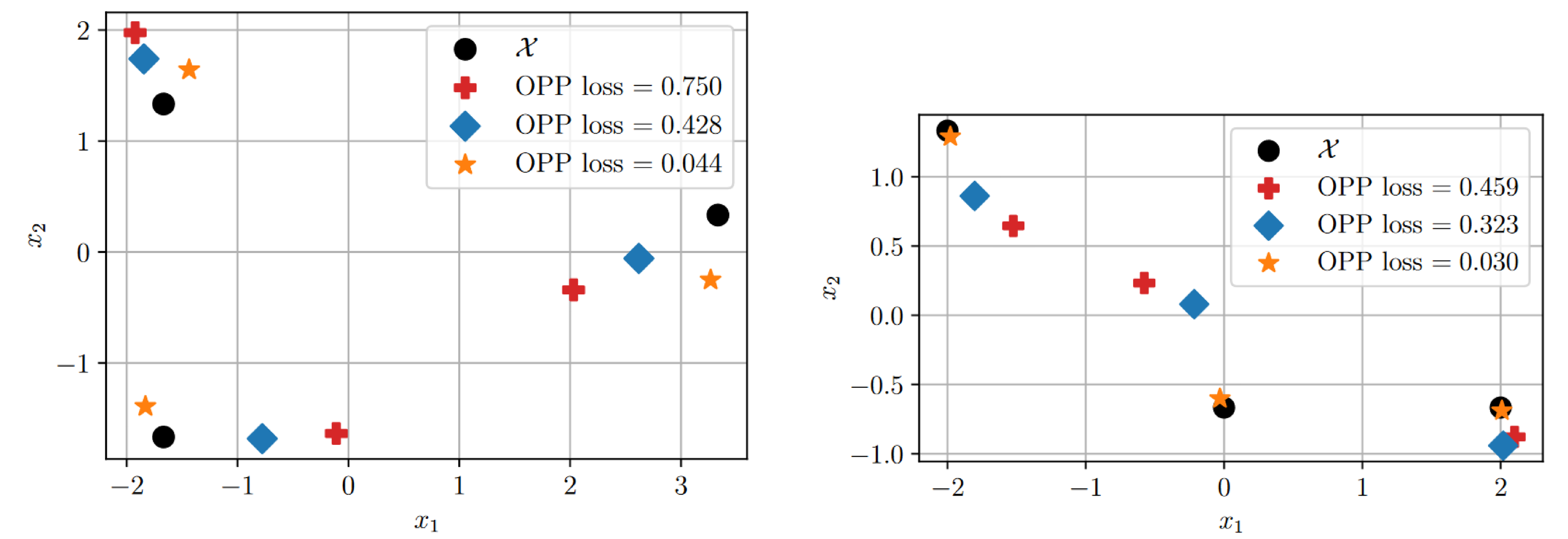
$$f_{\mathcal{Y}}\left(\mathcal{Y} \mid d_{\mathcal{E}(\mathcal{X})}\right) = \prod_{(i,j)\in\mathcal{E}(\mathcal{X})} \prod_{m=1}^{M} f_{Y_{ij}^{[m]}}\left(y_{ij}^{[m]} \;\middle|\; d_{ij}^{\mathcal{X}}\right)$$

- SSE cost function arises when $Y_{ij}^{[m]}$ follows a Gaussian distribution.
- $Y_{ij}^{[m]}$ does not always follow a Gaussian distribution, so minimization of SSE cost function may not lead to a maximum likelihood estimate.
- Idea: choose cost function to be minimized depending on distribution of $Y_{ij}^{[m]}$.

OPP Loss

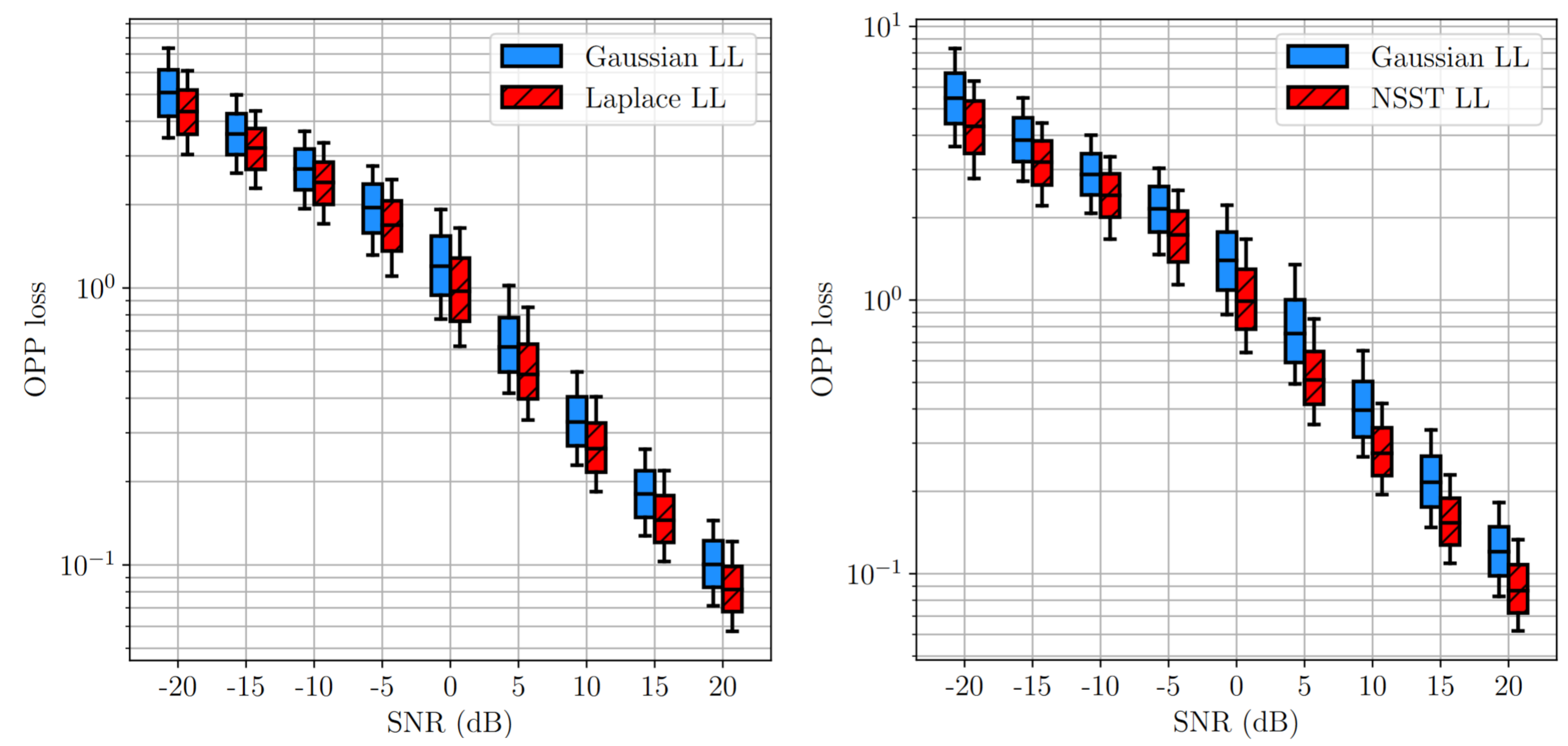
- Measure of how close an estimate $\hat{\mathcal{X}}$ is to $\mathcal{X}$.

$$\text{OPP loss} = \min_{\mathbf{R}} \left\|\mathbf{R}\hat{\mathbf{X}}_c - \mathbf{X}_c\right\|_F, \quad \text{s.t.} \quad \mathbf{R}^T\mathbf{R} = \mathbf{I}$$



Results

Results for 30 10-point structures:



Results for different number of measurements (M) per edge: