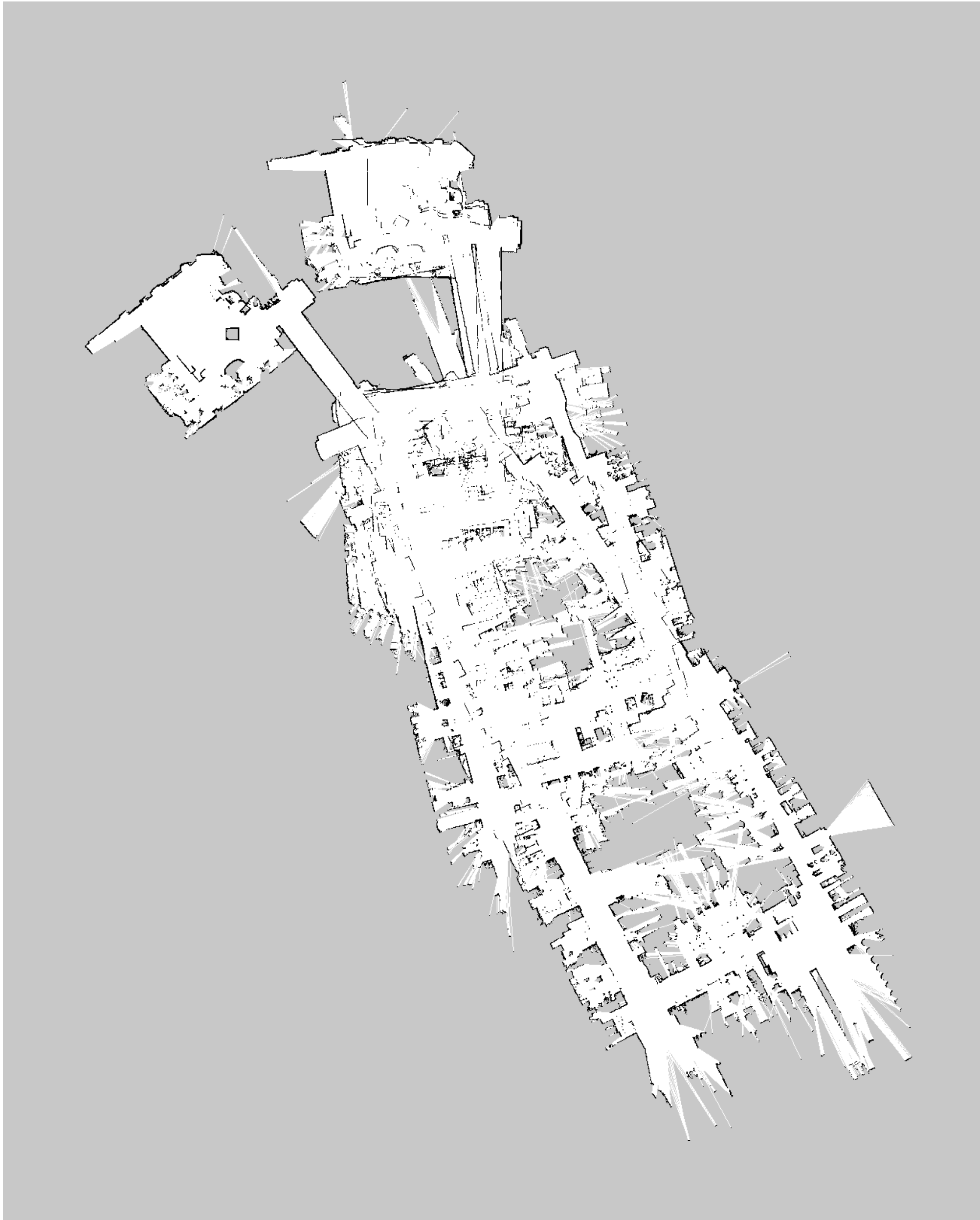
Least Squares and SLAM
Pose-SLAM

Giorgio Grisetti

Part of the material of this course is taken from the Robotics 2 lectures given by G.Grisetti, W.Burgard, C.Stachniss, K.Arras, D. Tipaldi and M.Bennewitz

Graph-Based SLAM in a Nutshell

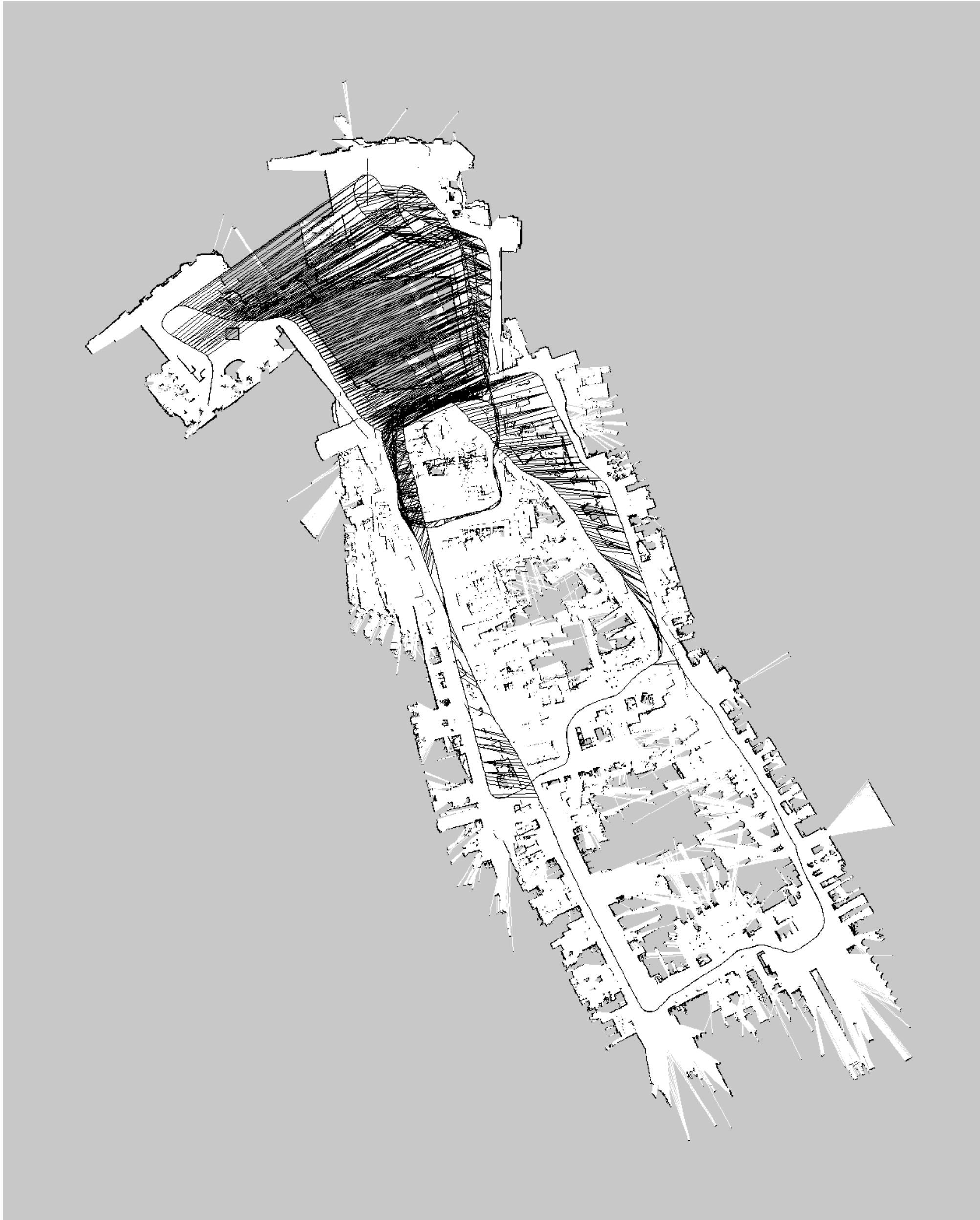
- Problem described as a graph
  - Every node corresponds to a robot position and to a laser measurement
  - An edge between two nodes represents a data-dependent spatial constraint between the nodes



KUKA Halle 22, courtesy of the P. Pfaff

Graph-Based SLAM in a Nutshell

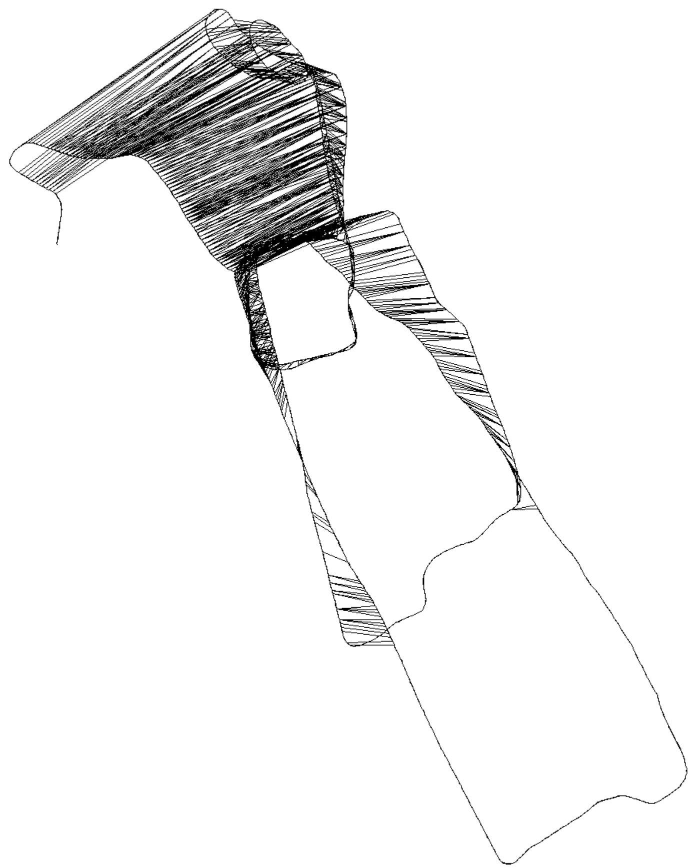
- Problem described as a graph
  - Every node corresponds to a robot position and to a laser measurement
  - An edge between two nodes represents a data-dependent spatial constraint between the nodes



KUKA Halle 22, courtesy of the P. Pfaff

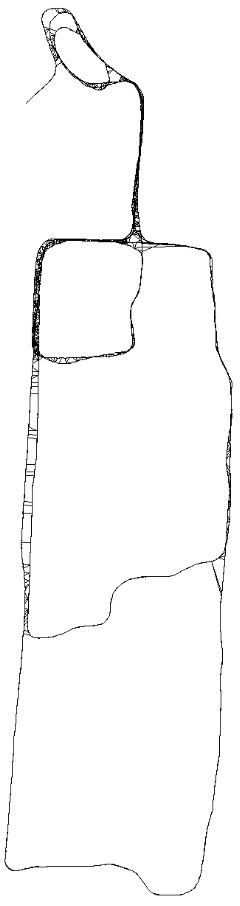
Graph-Based SLAM in a Nutshell

- Once we have the graph we determine the most likely map by "moving" the nodes



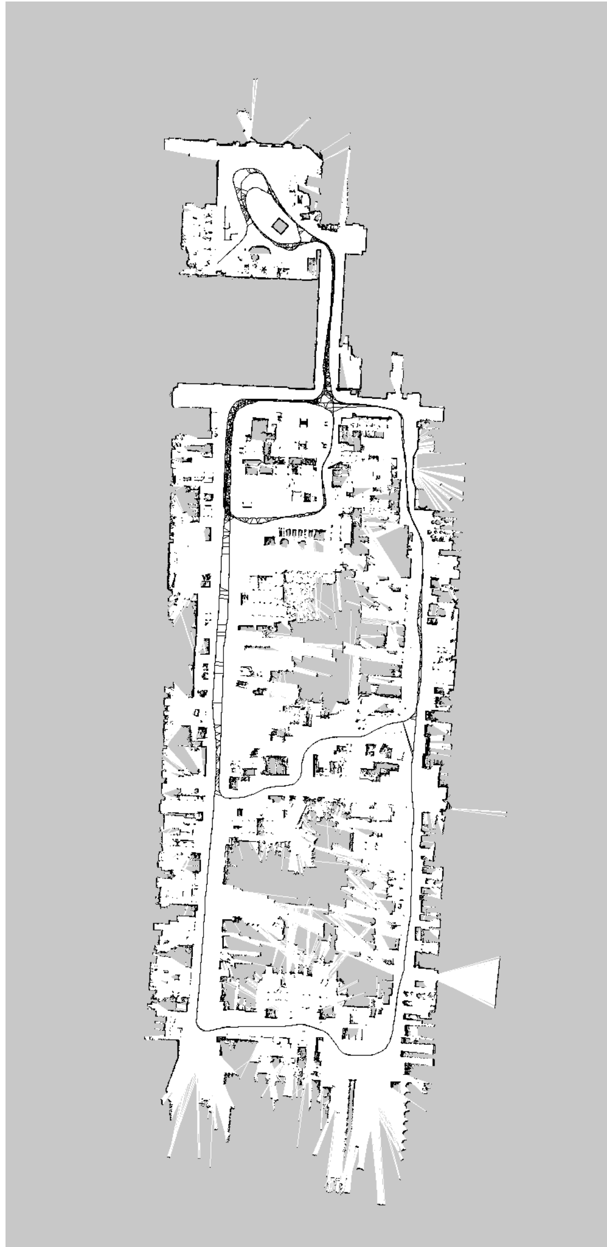
Graph-Based SLAM in a Nutshell

- Once we have the graph we determine the most likely map by "moving" the nodes
- … like this



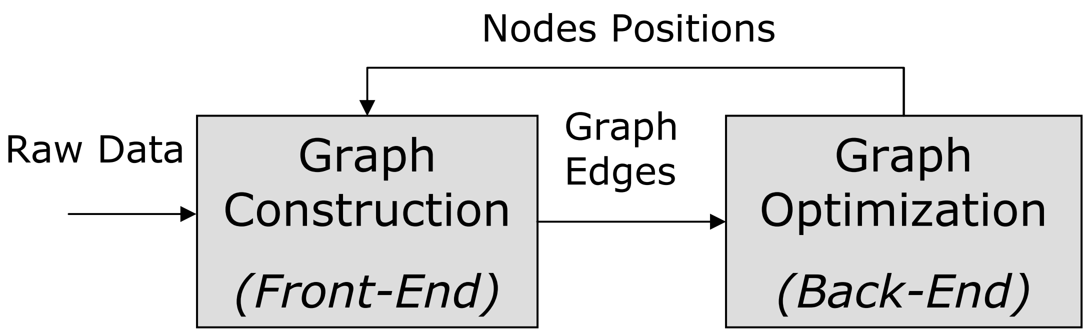
Graph-Based SLAM in a Nutshell

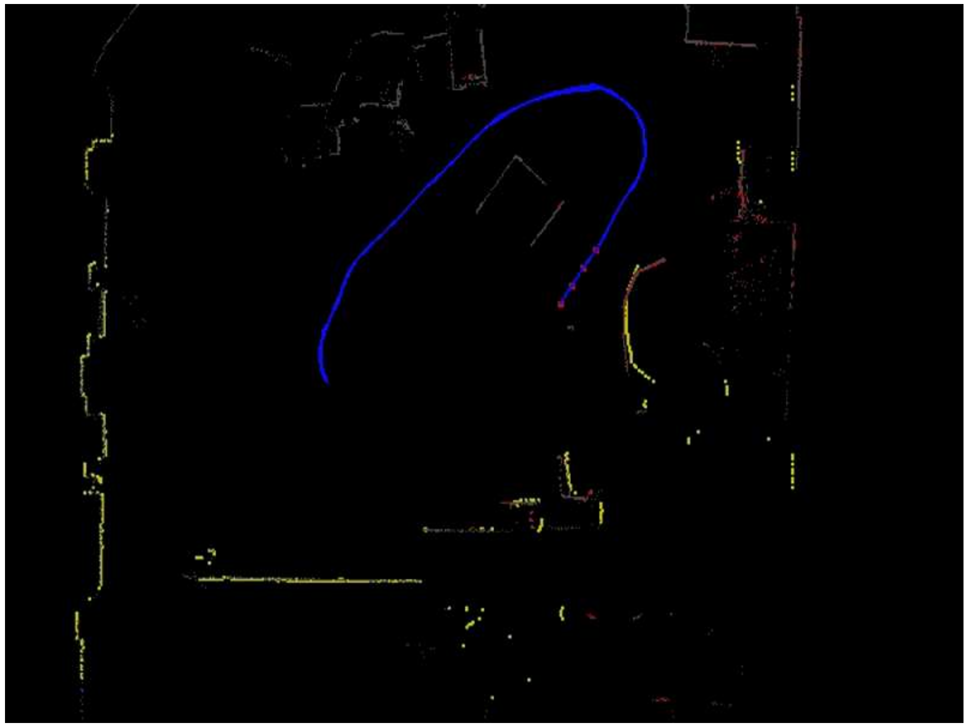
- Once we have the graph we determine the most likely map by “moving” the nodes
- … like this
- Then, we can render a map based on the known poses



Graph Optimization

- In this lecture, we will ***not*** address the how to construct the graph but how to retrieve the position of its nodes which is maximally consistent the observations in the edges.
- A general Graph-based SLAM algorithm interleaves the two steps
  - Graph construction
  - Graph optimization
- A consistent map helps in determining the new constraints by reducing the search space.



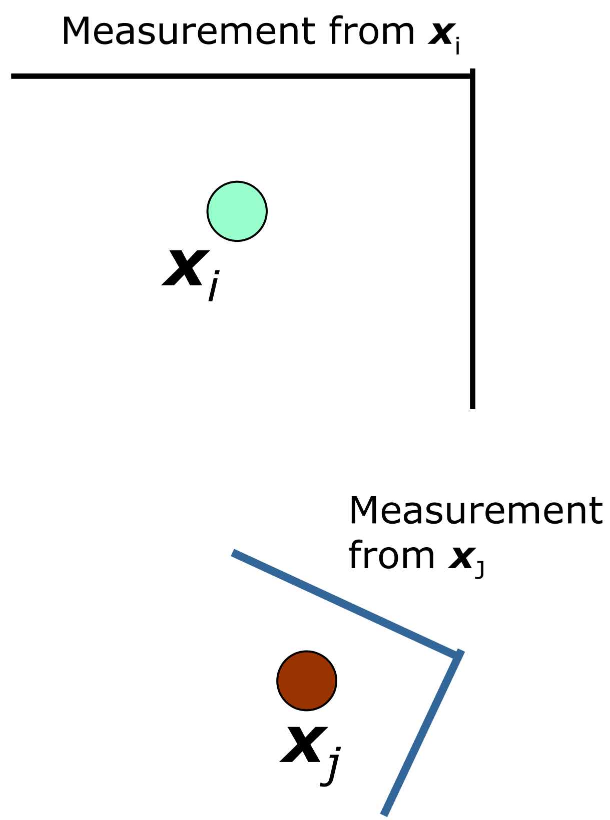


What Does the Graph Look Like?

- It has $\boldsymbol{n}$ nodes $\boldsymbol{x}=\boldsymbol{x}_{1:n}$
  - Each node $\boldsymbol{x}_i$ is a 2D or 3D transformation representing the pose of the robot at time $\boldsymbol{t}_i$.
- There is a constraint e_{ij} between the node $\boldsymbol{x}_i$ and the node $\boldsymbol{x}_j$ if
  - either
    - the robot observed the same part of the environment from both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and,
    - via this common observation it constructs a "virtual measurement" about the position of $\boldsymbol{x}_j$ seen from.
  - Or
    - the positions are subsequent in time and there is an odometry measurement between the two.

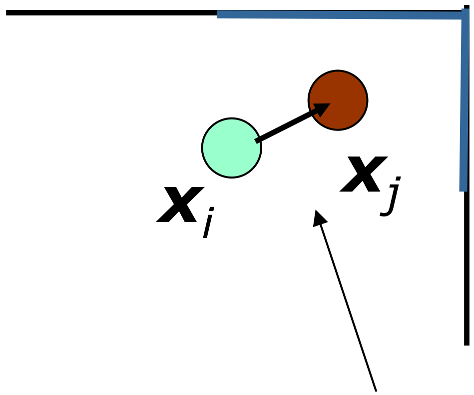
What Does the Graph Look Like?

- It has n nodes $x = x_{1:n}$
  - Each node x_i is a 2D or 3D transformation representing the pose of the robot at time t_i.
- There is a constraint e_{ij} between the node x_i and the node x_j if
  - either
    - the robot observed the same part of the environment from both x_i and x_j and,
    - via this common observation it constructs a "virtual measurement" about the position of x_j seen from.
  - Or
    - the positions are subsequent in time and there is an odometry measurement between the two.



What Does the Graph Look Like?

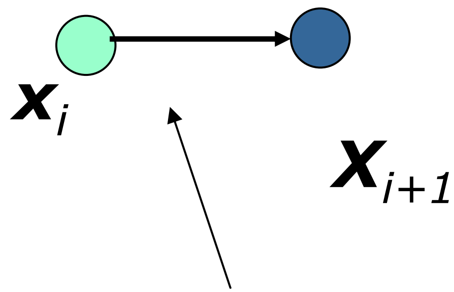
- It has ***n*** nodes $\boldsymbol{x}=\boldsymbol{x}_{1:n}$
  - Each node $\boldsymbol{x}_i$ is a 2D or 3D transformation representing the pose of the robot at time $\boldsymbol{t}_i$.
- There is a constraint e_{ij} between the node $\boldsymbol{x}_i$ and the node $\boldsymbol{x}_j$ if
  - either
    - the robot observed the same part of the environment from both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and,
    - via this common observation it constructs a "virtual measurement" about the position of $\boldsymbol{x}_j$ seen from.
  - Or
    - the positions are subsequent in time and there is an odometry measurement between the two.



The edge represents the position of $\boldsymbol{x}_j$ seen from $\boldsymbol{x}_i$, based on the ***observations***

What Does the Graph Look Like?

- It has $\boldsymbol{n}$ nodes $\boldsymbol{x}=\boldsymbol{x}_{1:n}$
  - Each node $\boldsymbol{x}_i$ is a 2D or 3D transformation representing the pose of the robot at time $\boldsymbol{t}_i$.
- There is a constraint e_{ij} between the node $\boldsymbol{x}_i$ and the node $\boldsymbol{x}_j$ if
  - either
    - the robot observed the same part of the environment from both $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ and,
    - via this common observation it constructs a "virtual measurement" about the position of $\boldsymbol{x}_j$ seen from.
  - Or
    - the positions are subsequent in time and there is an odometry measurement between the two.



The edge represents the **odometry** measurement

The Edge Information Matrices

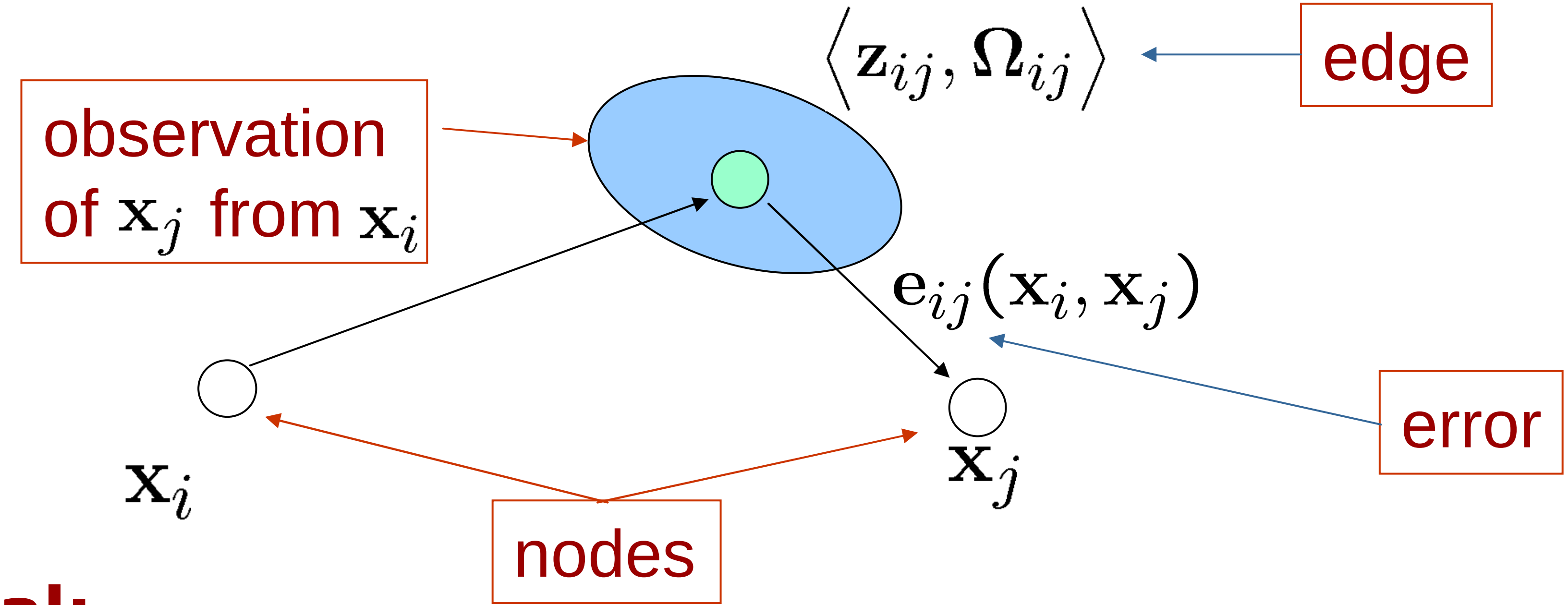
- To account for the different nature of the observations we add to the edge an information matrix $\boldsymbol{\Omega}_{ij}$ to encode the uncertainty of the edge.
- The "bigger" (in matrix sense) $\boldsymbol{\Omega}_{ij}$ is, the more the edge "matters" in the optimization procedure.

Questions:

- Any idea about the information matrices of the system in case we use scan-matching and odometry?
- What should these matrices look like in an endless corridor in both cases?

Pose Graph

- The input for the optimization procedure is a graph annotated as follows:



- **Goal:**
  - Find the assignment of poses to the nodes of the graph which minimizes the negative log likelihood of the observations:

$$\hat{\mathbf{x}} = \text{argmin} \sum_{ij} \mathbf{e}_{ij}^T \mathbf{\Omega}_{ij} \mathbf{e}_{ij}$$

SLAM as a Least Square Problem

- The error function looks suitable for least squares

$$\begin{aligned}\hat{\mathbf{x}} &= \operatorname{argmin} \sum_{ij} \mathbf{e}_{ij}^T(\mathbf{x}_i, \mathbf{x}_j)\boldsymbol{\Omega}_{ij}\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j) \\ &= \operatorname{argmin} \sum_{k} \mathbf{e}_{k}^T(\mathbf{x})\boldsymbol{\Omega}_{k}\mathbf{e}_{k}(\mathbf{x})\end{aligned}$$

- We can regard each edge as a measurement, and use what we already now
- Questions:
  - What is the state vector?
  - What is the error function?

SLAM as a Least Square Problem

- The error function looks suitable for least squares

$$\hat{\mathbf{x}} = \operatorname{argmin} \sum_{ij} \mathbf{e}_{ij}^T(\mathbf{x}_i, \mathbf{x}_j)\boldsymbol{\Omega}_{ij}\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

$$= \operatorname{argmin} \sum_{k} \mathbf{e}_{k}^T(\mathbf{x})\boldsymbol{\Omega}_{k}\mathbf{e}_{k}(\mathbf{x})$$

- We can regard each edge as a measurement, and use what we already now
- Questions:
  - What is the state vector?

$$\mathbf{x}^T = \begin{pmatrix} \mathbf{x}_1^T & \mathbf{x}_2^T & \cdots & \mathbf{x}_n^T \end{pmatrix}$$

One block for each node of the graph

  - What is the error function?

The Error Function

- The generic error function of a constraint characterized by a mean $\boldsymbol{z}_{ij}$ and an information matrix $\boldsymbol{\Omega}_{ij}$ is a vector of the same size as $\boldsymbol{x}_i$

$$\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \text{t2v}(\underbrace{\mathbf{Z}_{ij}^{-1}}_{\text{measurement}}(\underbrace{\mathbf{X}_i^{-1} \cdot \mathbf{X}_j}_{\mathbf{x}_j \text{ in the reference of } \mathbf{x}_i}))$$

- We can write the error as a function of all the state $\boldsymbol{x}$.

$$\mathbf{e}_{ij}(\mathbf{x}) = \text{t2v}(\mathbf{Z}_{ij}^{-1}(\mathbf{X}_i^{-1} \cdot \mathbf{X}_j))$$

- Note that the error function is 0 when

$$\mathbf{Z}_{ij} = (\mathbf{X}_i^{-1} \cdot \mathbf{X}_j)$$

The Overall Procedure (Recap)

To find a minimum

- Define the error function
- Linearize the error function (Taylor)
- Compute its derivative
- Set it to zero
- Solve the linear system
- Iterate this procedure until convergence

Linearizing the Error Function

- We can approximate the error functions around an initial guess ***x*** via Taylor expansion

$$\mathbf{J}_{ij} = \frac{\partial \mathbf{e}_{ij}(\mathbf{x})}{\partial \mathbf{x}}$$

$$\mathbf{e}_{ij}(\mathbf{x} + \boldsymbol{\Delta}\mathbf{x}) = \mathbf{e}_{ij}(\mathbf{x}) + \mathbf{J}_{ij}\boldsymbol{\Delta}\mathbf{x}$$

The Derivative of the Error Function

- Does one error function $\boldsymbol{e_{ij}(x)}$ depend on all state variables?

The Derivative of the Error Function

- Does one error function $\boldsymbol{e_{ij}(x)}$ depend on all state variables?
  - No, only on $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$
- Is there any consequence on the *structure* of the Jacobian?

The Derivative of the Error Function

- Does one error function $\mathbf{e}_{ij}(\mathbf{x})$ depend on all state variables?
  - No, only on $\mathbf{x}_i$ and $\mathbf{x}_j$
- Is there any consequence on the *structure* of the Jacobian?
  - Yes, it will be non-zero only in the rows corresponding to $\mathbf{x}_i$ and $\mathbf{x}_j$!

$$
\begin{aligned}
\frac{\partial \mathbf{e}_{ij}(\mathbf{x})}{\partial \mathbf{x}} &= \left(\mathbf{0} \cdots \frac{\partial \mathbf{e}_{ij}(\mathbf{x}_i)}{\partial \mathbf{x}_i} \cdots \frac{\partial \mathbf{e}_{ij}(\mathbf{x}_j)}{\partial \mathbf{x}_j} \cdots \mathbf{0} \right) \\
\mathbf{J}_{ij} &= \left(\mathbf{0} \cdots \mathbf{A}_{ij} \cdots \mathbf{B}_{ij} \cdots \mathbf{0} \right)
\end{aligned}
$$

Consequences of the Sparsity

- To apply least squares we need to compute the coefficient vectors and the coefficient matrices:

$$
\begin{aligned}
\mathbf{b}^T &= \sum_{ij} \mathbf{b}_{ij}^T = \sum_{ij} \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \\
\mathbf{H} &= \sum_{ij} \mathbf{H}_{ij} = \sum_{ij} \mathbf{J}_{ij}^T \boldsymbol{\Omega} \mathbf{J}_{ij}
\end{aligned}
$$

- The sparse structure of $\boldsymbol{J}_{ij}$, will result in a sparse structure of the linear system
- This structure will reflect the topology of the graph

Jacobians and Sparsity

- In error function $\mathbf{e}_{ij}$ of a constraint depends only on the two parameter blocks $\mathbf{x}_i$ and $\mathbf{x}_j$

$$\mathbf{e}_{ij}(\mathbf{x}) = \mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j)$$

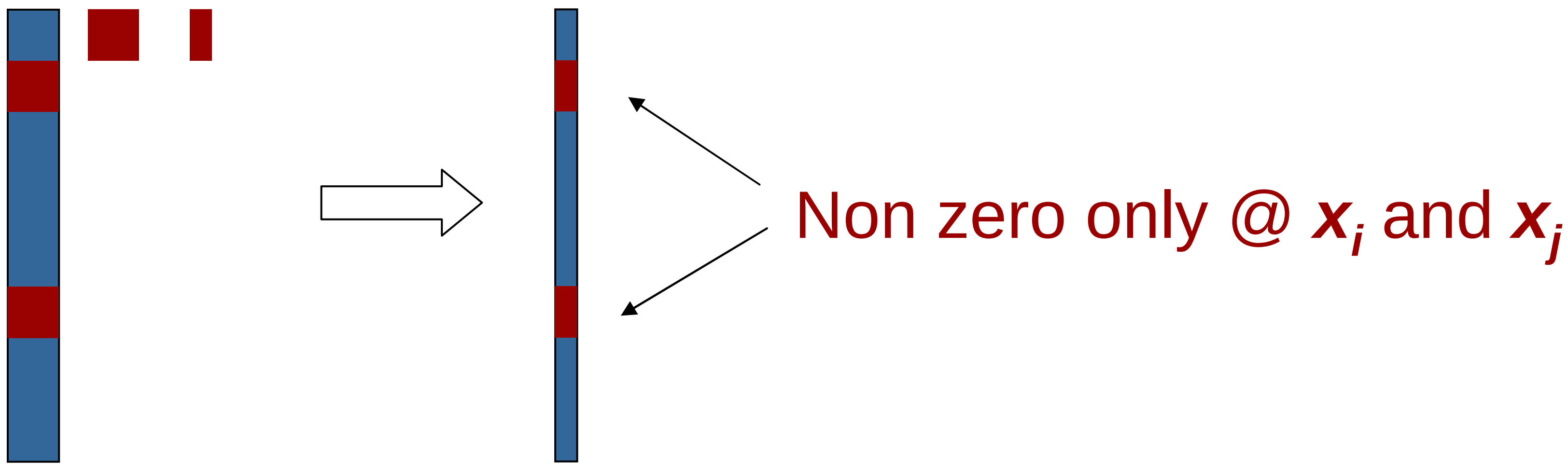
- Thus the Jacobian will be 0 everywhere but in the columns of $\mathbf{x}_i$ and $\mathbf{x}_j$.

$$\mathbf{J}_{ij} = \left(\mathbf{0} \cdots \mathbf{0} \quad \underbrace{\frac{\partial \mathbf{e}(\mathbf{x}_i)}{\partial \mathbf{x}_i}}_{\mathbf{A}_{ij}} \quad \mathbf{0} \cdots \mathbf{0} \quad \underbrace{\frac{\partial \mathbf{e}(\mathbf{x}_j)}{\partial \mathbf{x}_j}}_{\mathbf{B}_{ij}} \quad \mathbf{0} \cdots \mathbf{0} \right)$$

- This leads to a sparse pattern in the matrix $\boldsymbol{H}$, that reflects the adjacency matrix of the graph.

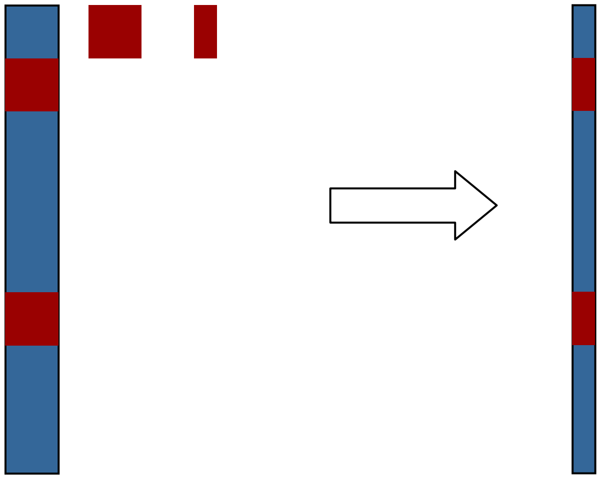
Consequences on the Structure

$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}$$



Consequences on the Structure

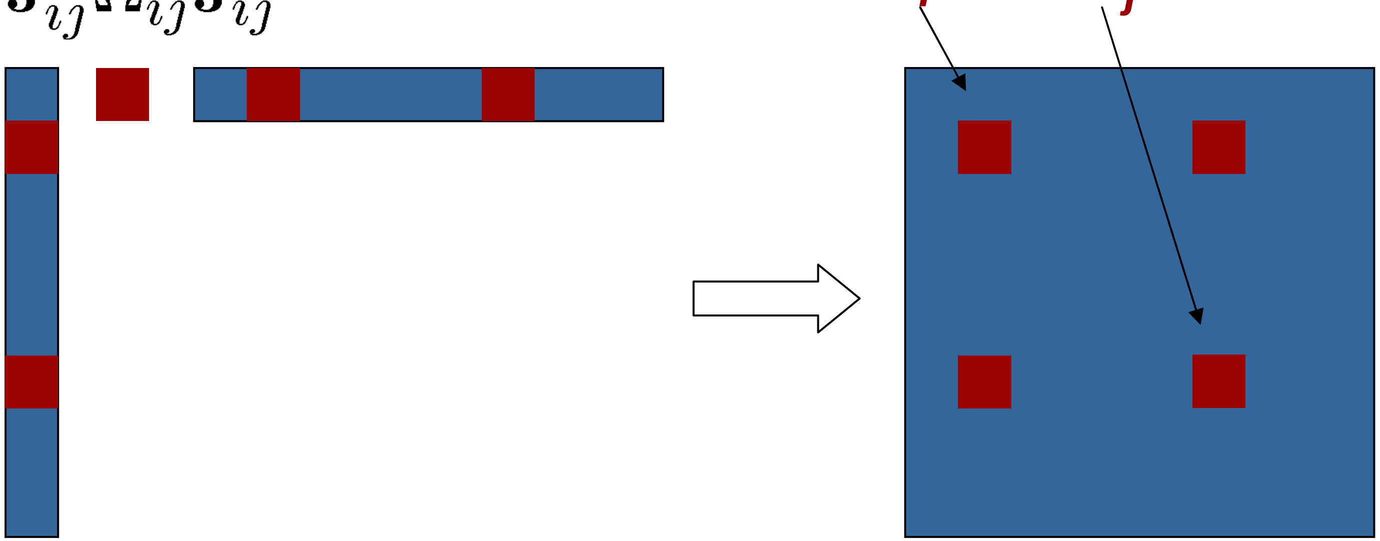
$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}$$



Non zero only @ $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

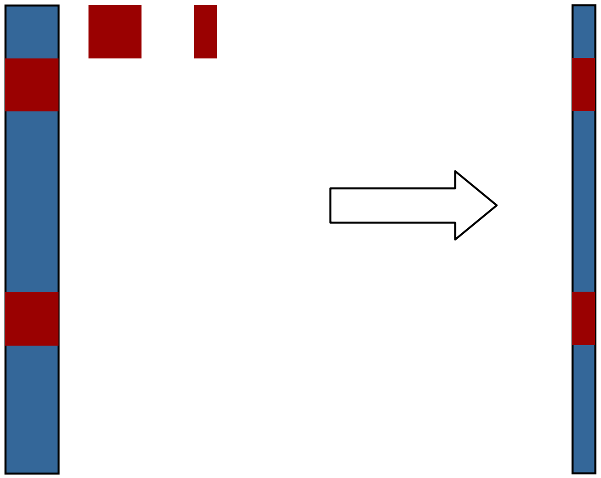
$$\mathbf{H}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij}$$

Non zero on the main diagonal @ $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$



Consequences on the Structure

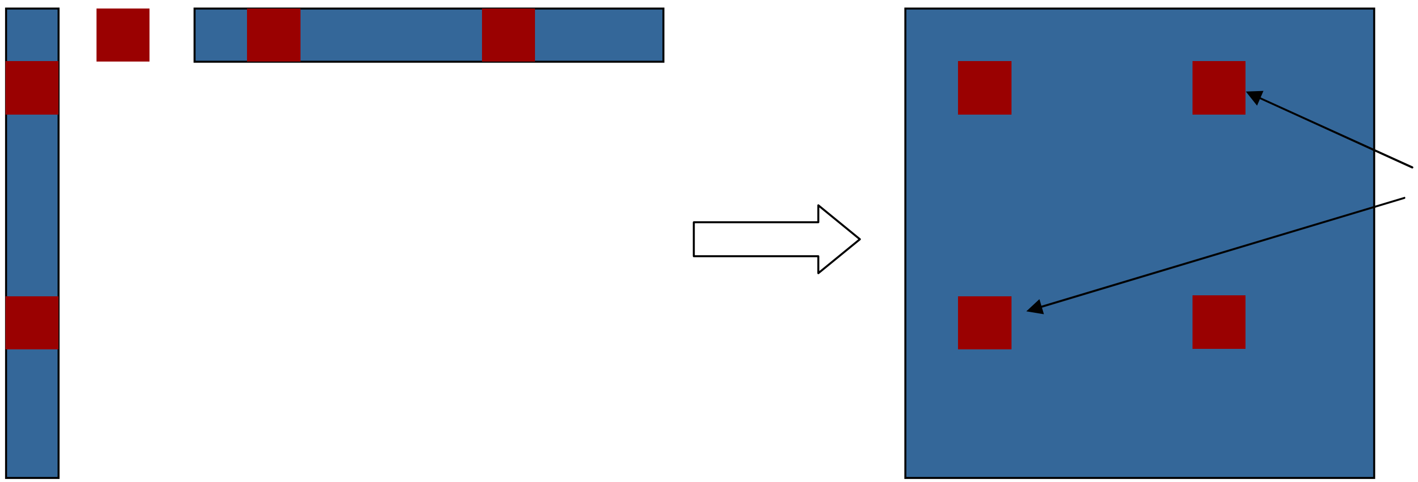
$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}$$



Non zero only @ $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

Non zero on the main diagonal @ $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$

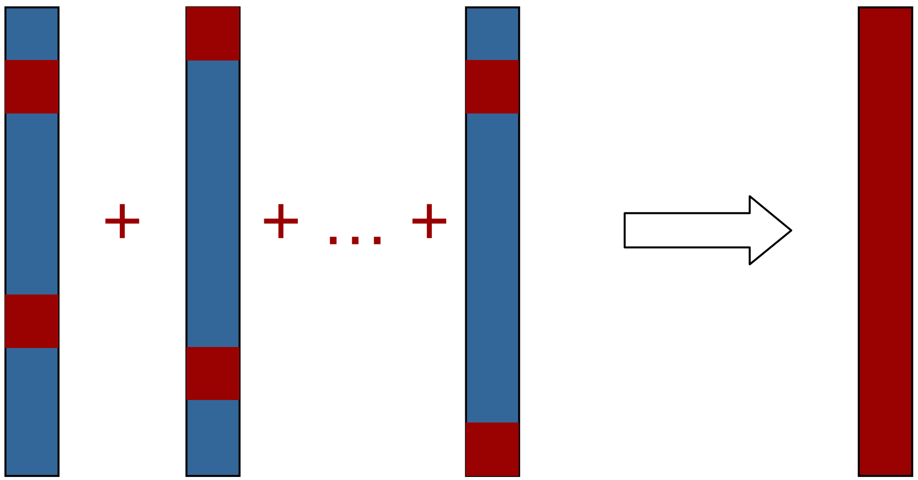
$$\mathbf{H}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij}$$



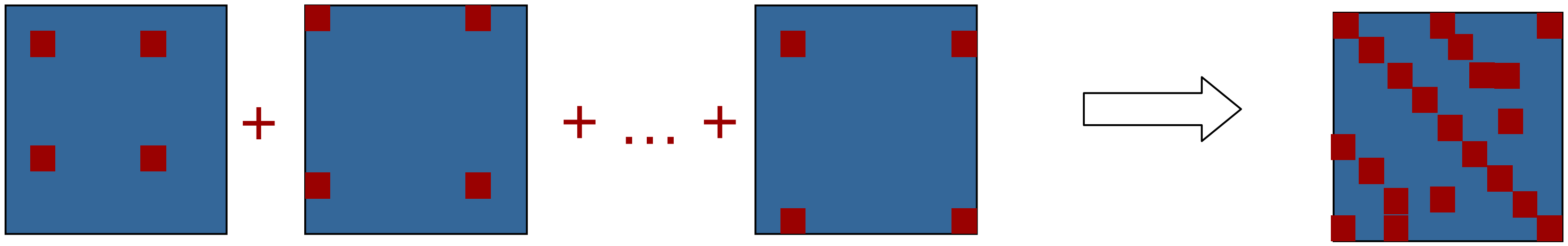
.. and at the blocks *ij*,*ji*

Consequences on the Structure

$$\mathrm{b} = \sum_{ij} \mathrm{b}_{ij}$$



$$\mathbf{H} = \sum_{ij} \mathbf{H}_{ij}$$



Consequences of the Sparsity

- An edge of the graph contribute s to the linear system via its coefficient vector $\boldsymbol{b_{ij}}$ and its coefficient matrix $\boldsymbol{H_{ij}}$.
  - The coefficient vector is:

$$\begin{aligned}
\mathbf{b}_{ij}^T &= \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \\
&= \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \left(\, 0 \cdots \mathbf{A}_{ij} \cdots \mathbf{B}_{ij} \cdots 0 \, \right) \\
&= \left(\, 0 \cdots \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} \cdots \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} \cdots 0 \, \right)
\end{aligned}$$

  - It is non-zero only in correspondence of $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$

Consequences of the Sparsity (cont.)

- The coefficient matrix of an edge is:

$$
\begin{aligned}
\mathbf{H}_{ij} &= \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \\
&= \begin{pmatrix} \vdots \\ \mathbf{A}_{ij}^T \\ \vdots \\ \mathbf{B}_{ij}^T \\ \vdots \end{pmatrix} \boldsymbol{\Omega}_{ij} \begin{pmatrix} \cdots \mathbf{A}_{ij} \cdots \mathbf{B}_{ij} \cdots \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} & \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} \\ \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} & \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij} \end{pmatrix}
\end{aligned}
$$

- Is non zero only in the blocks ***i,j.***

Consequences of the Sparsity (cont.)

- An edge between $\mathbf{x}_i$ and $\mathbf{x}_j$ in the graph contributes only
  - to the i^{th} and the j^{th} blocks of the coefficient vector,
  - to the blocks ***ii, jj, ij*** and ***ji*** of the coefficient matrix.
- The resulting system is sparse, and can be computed by iteratively "accumulating" the contribution of each edge
- Efficient solvers can be used
  - Sparse Cholesky decomposition with COLAMD
  - Conjugate Gradients
  - … many others

The Linear System

- Vector of the states increments:

$$\Delta\mathbf{x}^T = \begin{pmatrix} \Delta\mathbf{x}_1^T & \Delta\mathbf{x}_2^T & \cdots & \Delta\mathbf{x}_n^T \end{pmatrix}$$

- Coefficient vector:

$$\mathbf{b}^T = \begin{pmatrix} \bar{\mathbf{b}}_1^T & \bar{\mathbf{b}}_2^T & \cdots & \bar{\mathbf{b}}_n^T \end{pmatrix}$$

- System Matrix:

$$\mathbf{H} = \begin{pmatrix} \bar{\mathbf{H}}^{11} & \bar{\mathbf{H}}^{12} & \dots & \bar{\mathbf{H}}^{1n} \\ \bar{\mathbf{H}}^{21} & \bar{\mathbf{H}}^{22} & \dots & \bar{\mathbf{H}}^{2n} \\ \vdots & \ddots & & \vdots \\ \bar{\mathbf{H}}^{n1} & \bar{\mathbf{H}}^{n2} & \dots & \bar{\mathbf{H}}^{nn} \end{pmatrix}$$

- The linear system is a block system with ***n*** blocks, one for each node of the graph.

Building the Linear System

- $\mathbf{x}$ is the current linearization point
- Initialization

$$\mathbf{b} = 0 \qquad \mathbf{H} = 0$$

- For each constraint
  - Compute the error

$$\mathbf{e}_{ij} = \mathrm{t2v}(\mathbf{Z}_{ij}^{-1}(\mathbf{X}_i^{-1} \cdot \mathbf{X}_j))$$

  - Compute the blocks of the Jacobian:

$$\mathbf{A}_{ij} = \frac{\partial \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i} \qquad \mathbf{B}_{ij} = \frac{\partial \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_j}$$

  - Update the coefficient vector:

$$\bar{\mathbf{b}}_i^T + = \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} \qquad \bar{\mathbf{b}}_j^T + = \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij}$$

  - Update the system matrix:

$$\bar{\mathbf{H}}^{ii} + = \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} \qquad \bar{\mathbf{H}}^{ij} + = \mathbf{A}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij}$$

$$\bar{\mathbf{H}}^{ji} + = \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{A}_{ij} \qquad \bar{\mathbf{H}}^{jj} + = \mathbf{B}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{B}_{ij}$$

Algorithm

- ***x***: the initial guess
- While (! converged)
  - ***<H,b>*** = buildLinearSystem(***x***);
  - ***Δx*** = solveSparse(***H Δx = -b***);
  - x += ***Δx;***

Exercise(s)

- Consider a 2D graph, where each pose $\boldsymbol{x_i}$ is parameterized as

$$\mathbf{x}_i^T = (x_i \; y_i \; \theta_i)$$

- Consider the error function

$$\mathbf{e}_{ij} = \text{t2v}(\mathbf{Z}_{ij}^{-1}(\mathbf{X}_i^{-1} \cdot \mathbf{X}_j))$$

- Compute the blocks of the Jacobian $\boldsymbol{J}$

$$\mathbf{A}_{ij} = \frac{\partial \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_i} \qquad \mathbf{B}_{ij} = \frac{\partial \mathbf{e}(\mathbf{x}_i, \mathbf{x}_j)}{\partial \mathbf{x}_j}$$

- Hint: write the error function by using rotation matrices and translation vectors

$$\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j) \;=\; \mathbf{Z}_{ij}^{-1} \begin{pmatrix} \mathbf{R}_i^T(\mathbf{t}_j - \mathbf{t}_i) \\ \theta_j - \theta_i \end{pmatrix}$$

Conclusions

- A part of the SLAM problem can be effectively solved with least square optimization.
- The algorithm described in this lecture has been entirely implemented in octave. Get the package from the web-page of the course.
- Play with the example, and figure out the relation between
  - the connectivity of the graph and
  - the structure of the matrix $\boldsymbol{H}$.