VAN course
Lesson 10

Dr. Refael Vivanti
vivanti@gmail.com

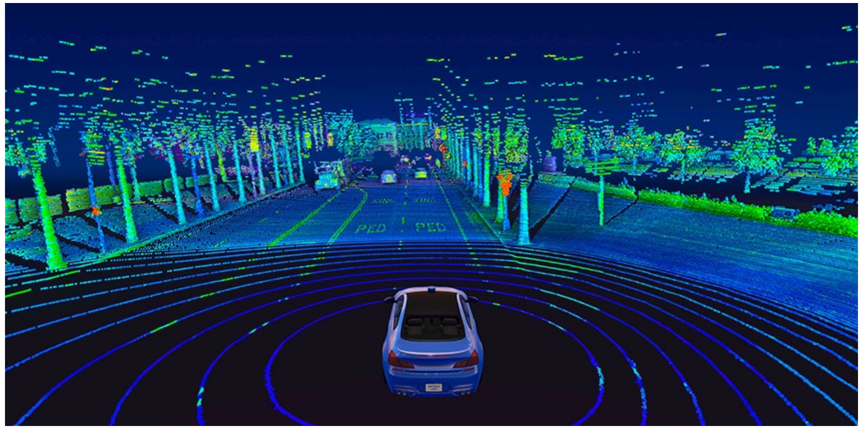
Today’s topics:

- Pose Graph
  - Depth cameras:
    - Lidar, Depth Cameras, TOF, RADAR, stereo
  - From point clouds to constraints
  - Our Pose Graph flavour
- How sparsity helps?
- Back to some statistics:
  - Information matrix and vector
  - Marginalization vs conditioning
- Compromises in our Pose Graph
- Our Pose Graph – how to

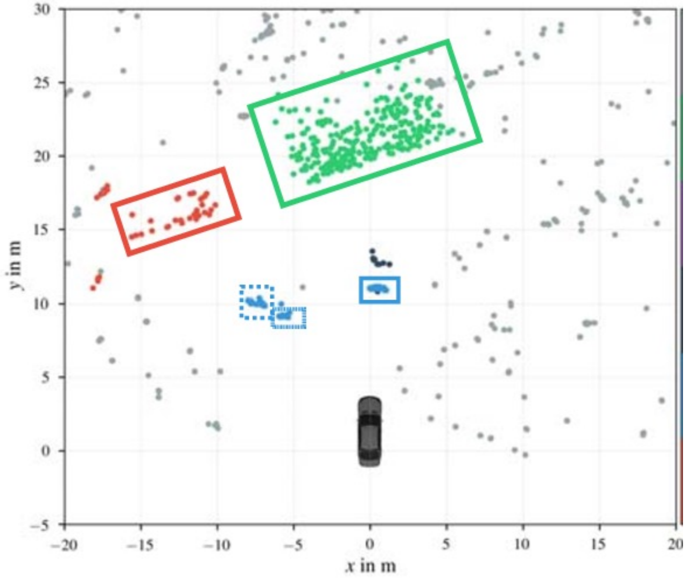
Pose Graph

Pose Graph

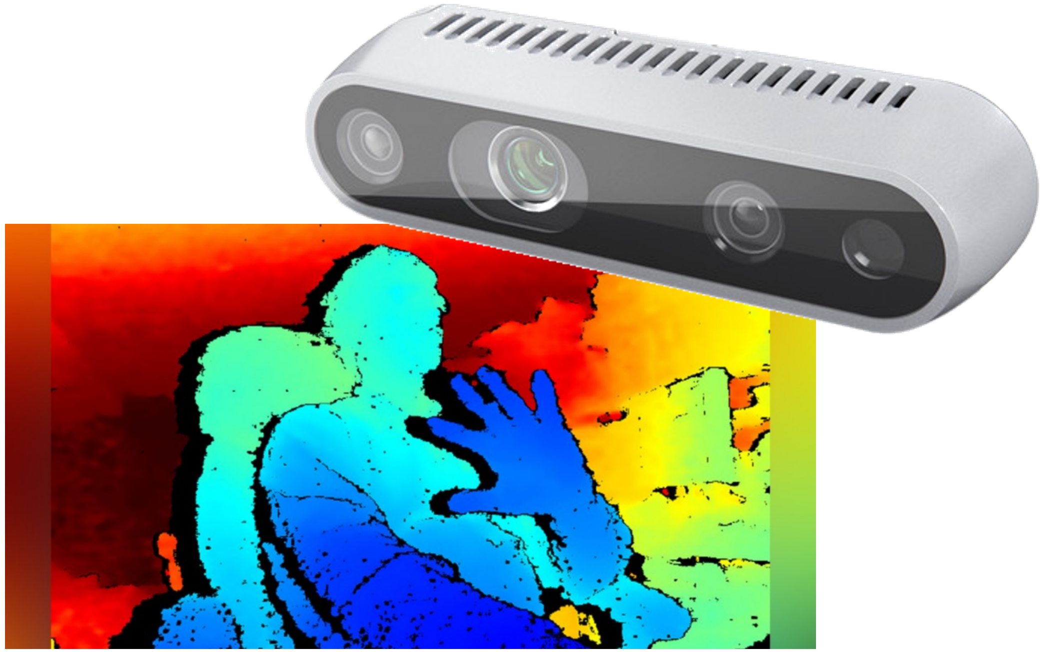
3D sensors:



LiDAR

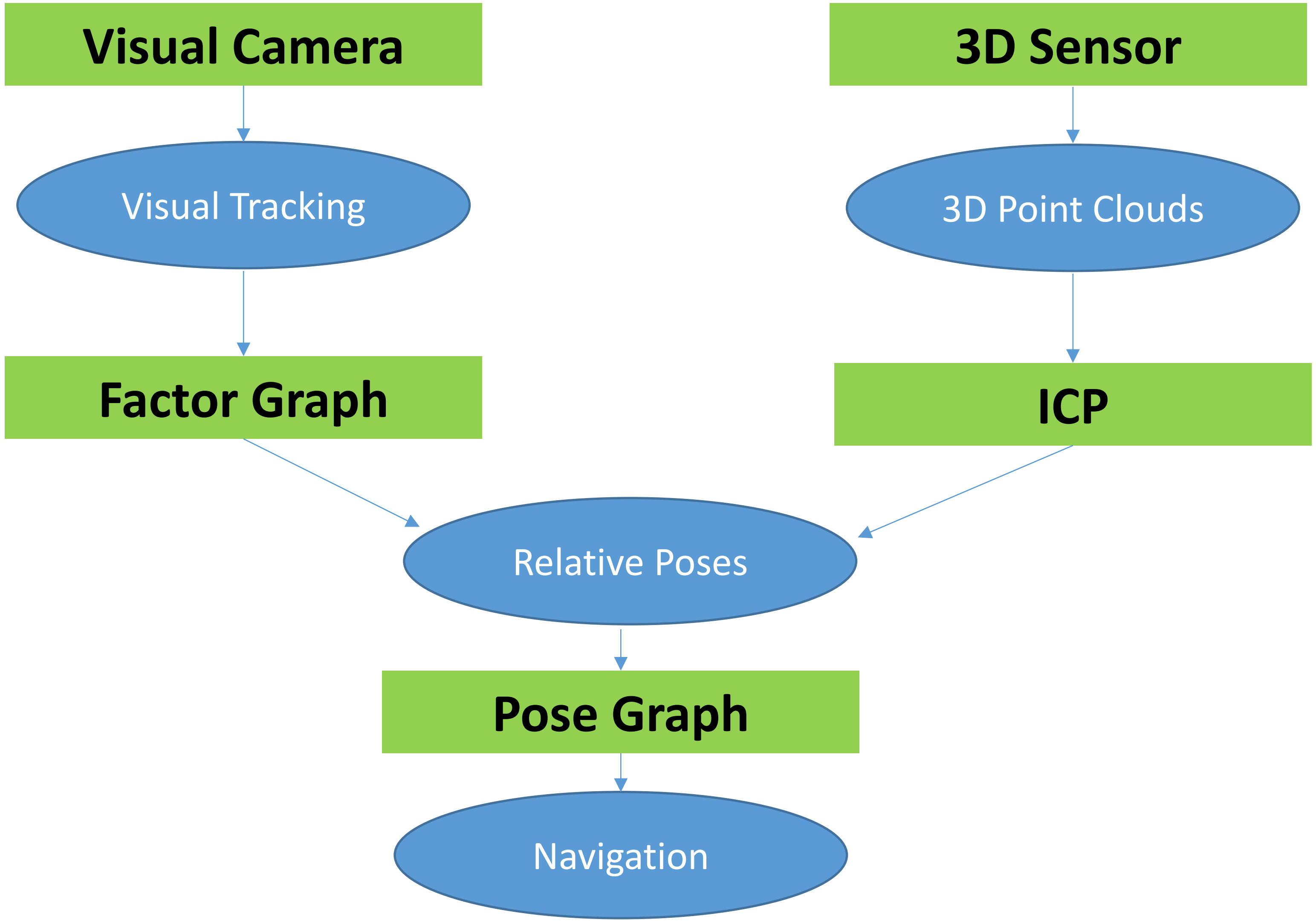


RADAR



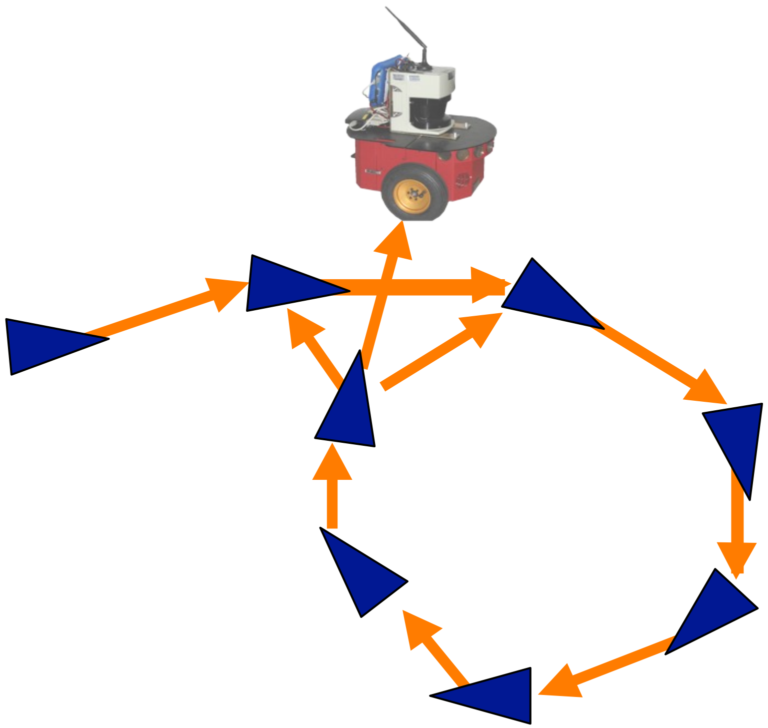
Depth Camera

Pose Graph


Visual Camera
Visual Tracking
Factor Graph
3D Sensor
3D Point Clouds
ICP
Relative Poses
Pose Graph
Navigation

Pose Graph

Graph-Based SLAM



Courtesy of Cyrill Stachniss

Robot pose

Constraint

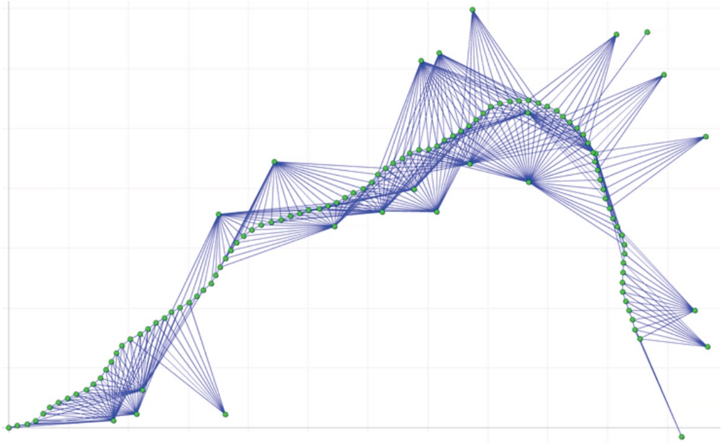
Pose Graph

Idea of Pose Graph SLAM

- Use a **graph** to represent the problem
- Every **node** in the graph corresponds to a pose of the robot during mapping
- Every **edge** between two nodes corresponds to a spatial constraint between them
- **Graph-Based SLAM:** Build the graph and find a node configuration that minimize the error introduced by the constraints

Pose Graph

How many computations did we save?

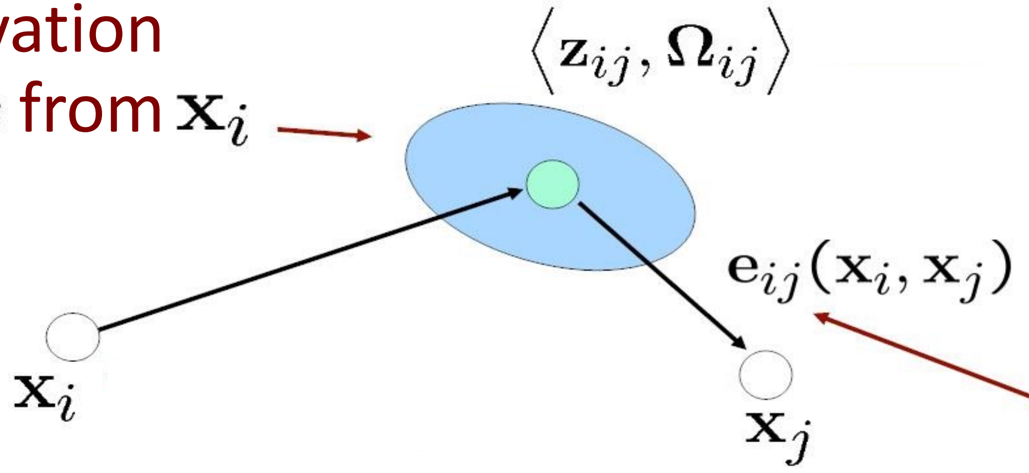


Courtesy of Dellaert06ijrr: "Square Root SAM"

- Case:
  - 1000 cameras, each sees 100 points.
- Full Factor graph:
  - Constraints: 10^5
  - Parameters: $6*10^3$ (cameras) + $3*10^5$ (3d points)
  - Jacobian: $\sim 10^{10}$, Information matrix: $\sim 10^{10}$
- Pose graph
  - Key Frame every 10 frames – 100 KFs
  - 100 constraints
    - Parameters: $6*10^2$ (cameras)
  - Jacobian: $\sim 10^4$, Information matrix: $\sim 10^5$, very sparse

Pose Graph

Edge: observation of $\mathbf{x}_j$ from $\mathbf{x}_i$



error

Goal: $\mathbf{x}^* = \underset{\mathbf{x}}{\operatorname{argmin}} \sum_{ij} \mathbf{e}_{ij}^T \mathbf{\Omega}_{ij} \mathbf{e}_{ij}$, $\Omega = \Sigma^{-1}$

The Error Function

Error function for a single constraint

$$\mathbf{e}_{ij}(\mathbf{x}_i, \mathbf{x}_j) = \text{t2v}(\underbrace{\mathbf{Z}_{ij}^{-1}}_{\text{measurement}}(\underbrace{\mathbf{X}_i^{-1}\mathbf{X}_j}_{x_j \text{ referenced w.r.t. } x_i}))$$

$$X = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$$

Error takes a value of zero if

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

t2v: X -> $(x, y, z, \alpha, \beta, \gamma)$

Gauss-Newton: The Overall Error Minimization Procedure

1. Define the error function
2. Linearize the error function
3. Compute its derivative
4. Set the derivative to zero
5. Solve the linear system
6. Iterate this procedure until convergence

Courtesy of Cyrill Stachniss

How Sparsity Helps?

How sparsity helps?

Jacobians and Sparsity

- Error $\mathbf{e}_{ij}(\mathbf{x})$ depends only on the two parameter blocks $\mathbf{x}_i$ and $\mathbf{x}_j$

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

- The Jacobian will be zero everywhere except in the columns of $\mathbf{x}_i$ and $\mathbf{x}_j$

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

Courtesy of Cyrill Stachniss

How sparsity helps?

Consequences of the Sparsity

- We need to compute the coefficient vector $\mathbf{b}$ and matrix $\mathbf{H}$:

$$\begin{aligned}\mathbf{b}^T &= \sum_{ij} \mathbf{b}_{ij}^T &= \sum_{ij} \mathbf{e}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij} \\ \mathbf{H} &= \sum_{ij} \mathbf{H}_{ij} &= \sum_{ij} \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{J}_{ij}\end{aligned} \qquad \Omega = \Sigma^{-1}$$

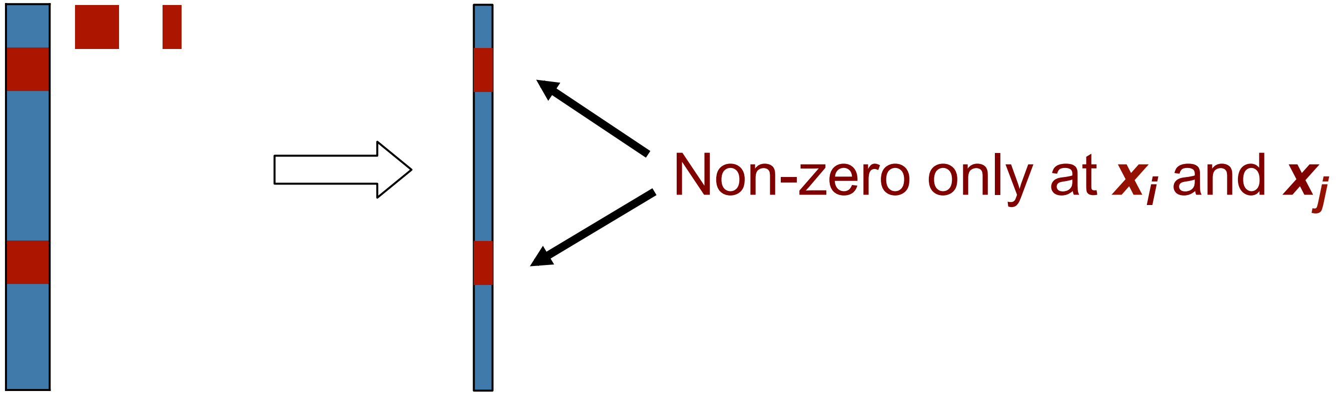
- The sparse structure of $\mathbf{J}_{ij}$ will result in a sparse structure of $\mathbf{H}$
- This structure reflects the adjacency matrix of the graph

Courtesy of Cyrill Stachniss

How sparsity helps?

Illustration of the Structure

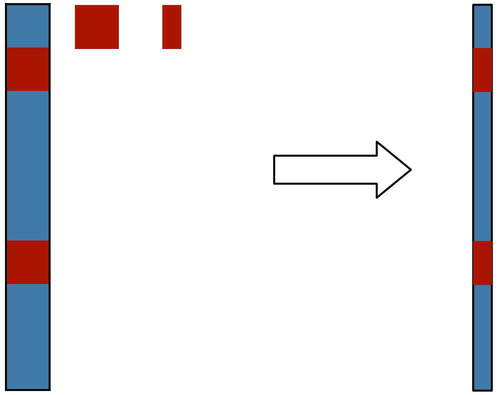
$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}$$



Courtesy of Cyrill Stachniss

How sparsity helps?
Illustration of the Structure

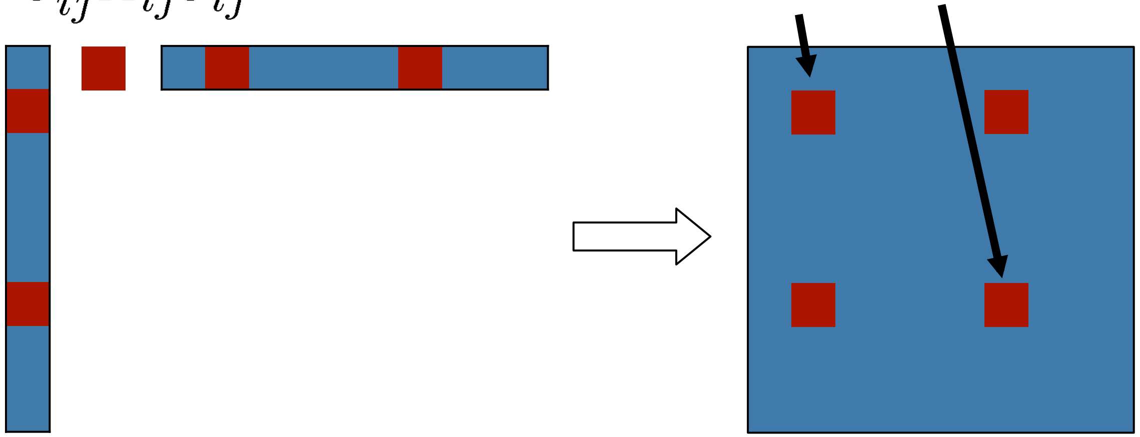
$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \boldsymbol{\Omega}_{ij} \mathbf{e}_{ij}$$



Non-zero only at $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$

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

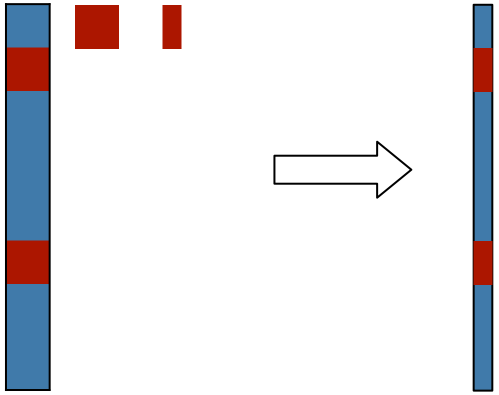
Non-zero on the main diagonal at $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$



Courtesy of Cyrill Stachniss

How sparsity helps?
Illustration of the Structure

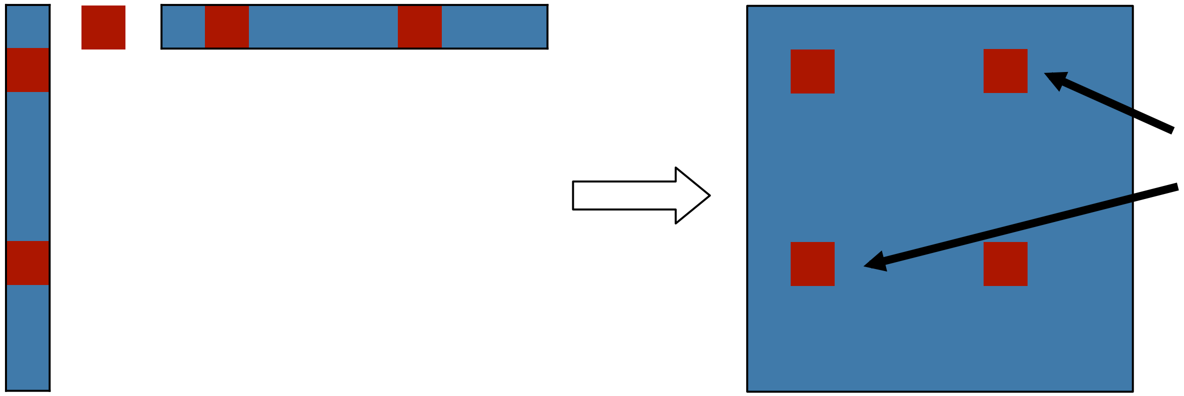
$$\mathbf{b}_{ij} = \mathbf{J}_{ij}^T \Omega_{ij} \mathbf{e}_{ij}$$



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

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

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

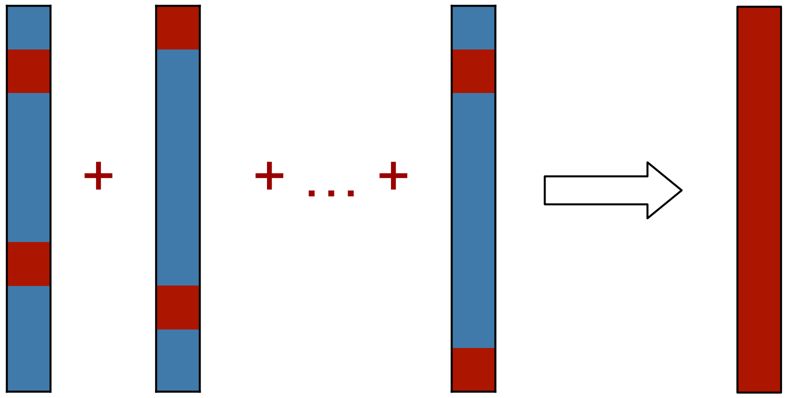


... and at the blocks *ij,ji*

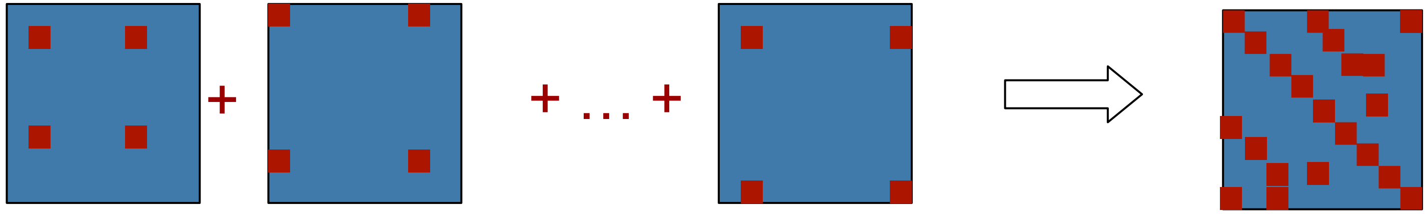
Courtesy of Cyrill Stachniss

How sparsity helps? Illustration of the Structure

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



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



Courtesy of Cyrill Stachniss

How sparsity helps?

Building the Linear System

For each constraint:

- Compute error $\mathbf{e}_{ij} = \text{t2v}(\mathbf{Z}_{ij}^{-1}(\mathbf{X}_i^{-1}\mathbf{X}_j))$
- Compute the building-blocks:

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

- 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:

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

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

Courtesy of Cyrill Stachniss

How sparsity helps?
Algorithm

optimize($\mathbf{x}$):

while (!*converged*)

$(\mathbf{H}, \mathbf{b}) = \text{buildLinearSystem}(\mathbf{x})$

$\boldsymbol{\Delta}\mathbf{x} = \text{solveSparse}(\mathbf{H}\boldsymbol{\Delta}\mathbf{x} = -\mathbf{b})$

$\mathbf{x} = \mathbf{x} + \boldsymbol{\Delta}\mathbf{x}$

less calculations

end

return $\mathbf{x}$

Courtesy of Cyrill Stachniss

Pose Graph

So we saved a lot of computation time:

- Dropping most of our information
  - Leaving only the Key frames and their relative poses
- Using the problem sparsity

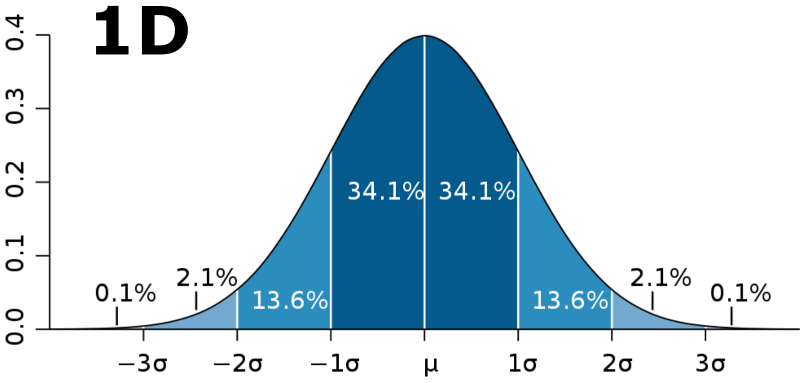
- But at what cost?

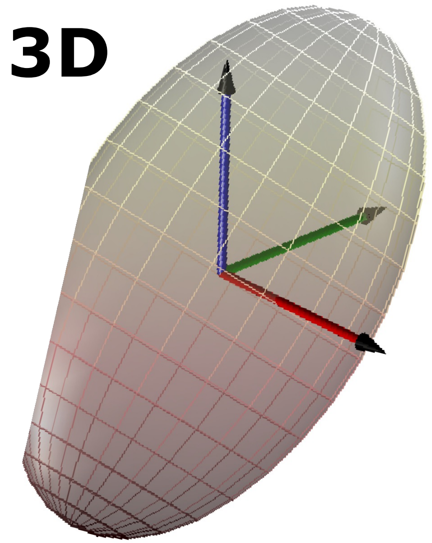
Back to
some statistics

Gaussians

- Gaussian described by **moments** μ, Σ

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$





Courtesy of Cyrill Stachniss

Canonical Parameterization

- Alternative representation for Gaussians
- Described by **information matrix** Ω and **information vector** ξ

Courtesy of Cyrill Stachniss

Canonical Parameterization

- Alternative representation for Gaussians
- Described by **information matrix** Ω

$$\Omega = \Sigma^{-1}$$

- and **information vector** ξ

$$\xi = \Sigma^{-1}\mu$$

Courtesy of Cyrill Stachniss

Complete Parameterizations

moments

$$\Sigma = \Omega^{-1}$$

$$\mu = \Omega^{-1}\xi$$

canonical

$$\Omega = \Sigma^{-1}$$

$$\xi = \Sigma^{-1}\mu$$

Courtesy of Cyrill Stachniss

Towards the Information Form

$$
\begin{aligned}
& p(x) \\
= \quad & \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)
\end{aligned}
$$

Towards the Information Form

$$
\begin{aligned}
p(x) & \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \Big) \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2} x^T \Sigma^{-1} x + x^T \Sigma^{-1} \mu - \frac{1}{2} \mu^T \Sigma^{-1} \mu \Big)
\end{aligned}
$$

Courtesy of Cyrill Stachniss

Towards the Information Form

$$
\begin{aligned}
&p(x) \\
&\quad = \quad \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big) \\
&\quad = \quad \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(\underline{-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu} - \underline{\frac{1}{2}\mu^T\Sigma^{-1}\mu}\Big) \\
&\quad = \quad \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}\mu^T\Sigma^{-1}\mu\Big) \\
&\qquad\qquad\qquad \exp\Big(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu\Big)
\end{aligned}
$$

Towards the Information Form

$$
\begin{aligned}
p(x) & \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu - \frac{1}{2}\mu^T\Sigma^{-1}\mu\right) \\
&= \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mu^T\Sigma^{-1}\mu\right) \\
&\qquad\qquad \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu\right) \\
&= \eta \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu\right)
\end{aligned}
$$

Courtesy of Cyrill Stachniss

Towards the Information Form

$$
\begin{aligned}
& p(x) \\
= \quad & \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right) \\
= \quad & \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu - \frac{1}{2}\mu^T\Sigma^{-1}\mu\right) \\
= \quad & \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\mu^T\Sigma^{-1}\mu\right) \\
& \qquad\qquad \exp\left(-\frac{1}{2}x^T\Sigma^{-1}x + x^T\Sigma^{-1}\mu\right) \\
= \quad & \eta \exp\left(-\frac{1}{2}x^T\underline{\Sigma^{-1}}x + x^T\underline{\Sigma^{-1}\mu}\right) \\
= \quad & \eta \exp\left(-\frac{1}{2}x^T\Omega x + x^T\xi\right)
\end{aligned}
$$

Dual Representation

$$p(x) = \frac{\exp(-\frac{1}{2}\mu^T\xi)}{\det(2\pi\Omega^{-1})^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}x^T\Omega x + x^T\xi\Big)$$

canonical parameterization

$$p(x) = \det(2\pi\Sigma)^{-\frac{1}{2}} \exp\Big(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\Big)$$

moments parameterization

Courtesy of Cyrill Stachniss

Marginalization vs. conditioning

- Both are dimension reduction: P(*x,y*) -> P(*x*)
  - **Marginalization** - summing over all y:

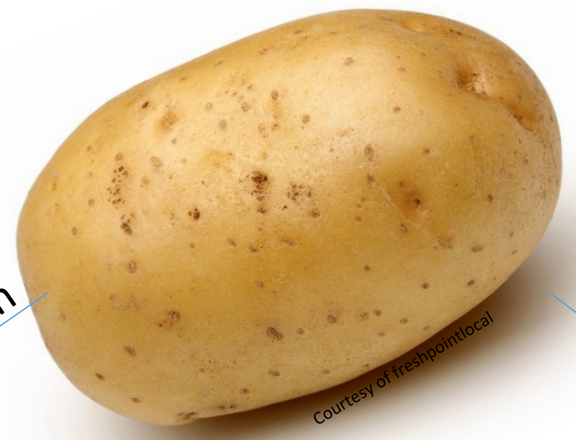
$$p(x) = \sum_y p(x, y)$$
$$= \sum_y p(x|y)\, p(y)$$

  - **Conditioning** – probability given a specific *y*:

$$p(x \mid y = y_0)$$

Courtesy of Cyrill Stachniss

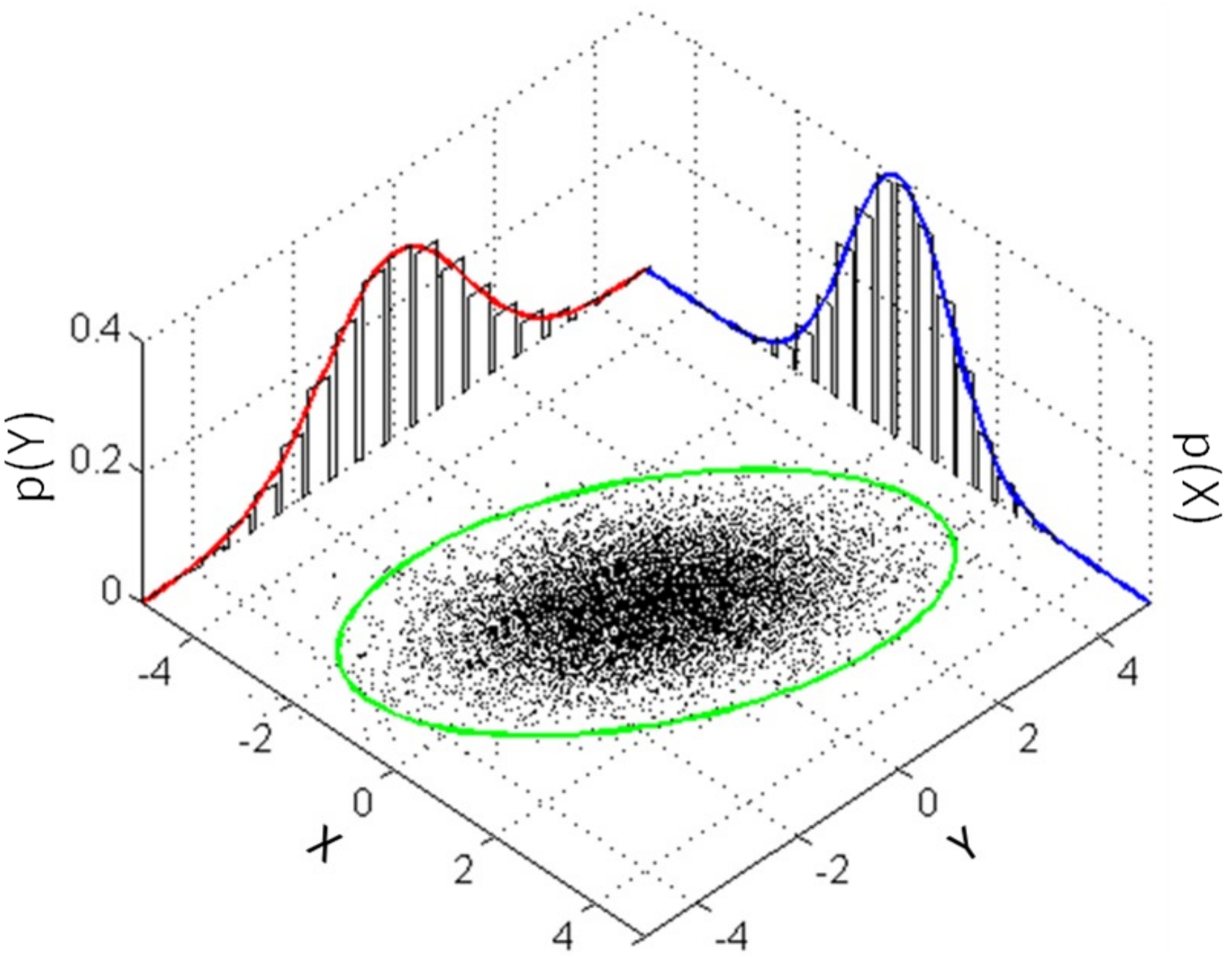
Marginalization vs. conditioning



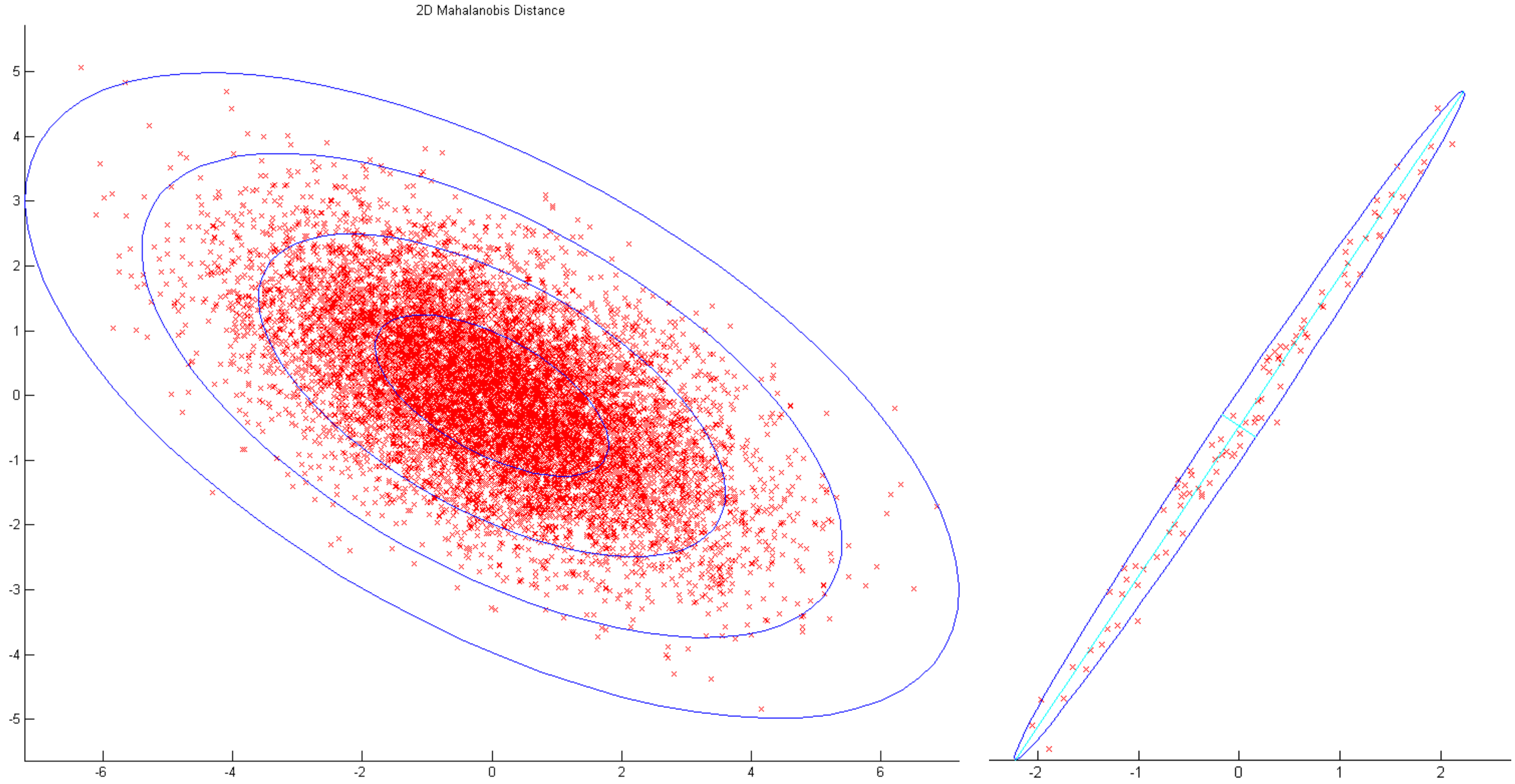
marginalization

conditioning

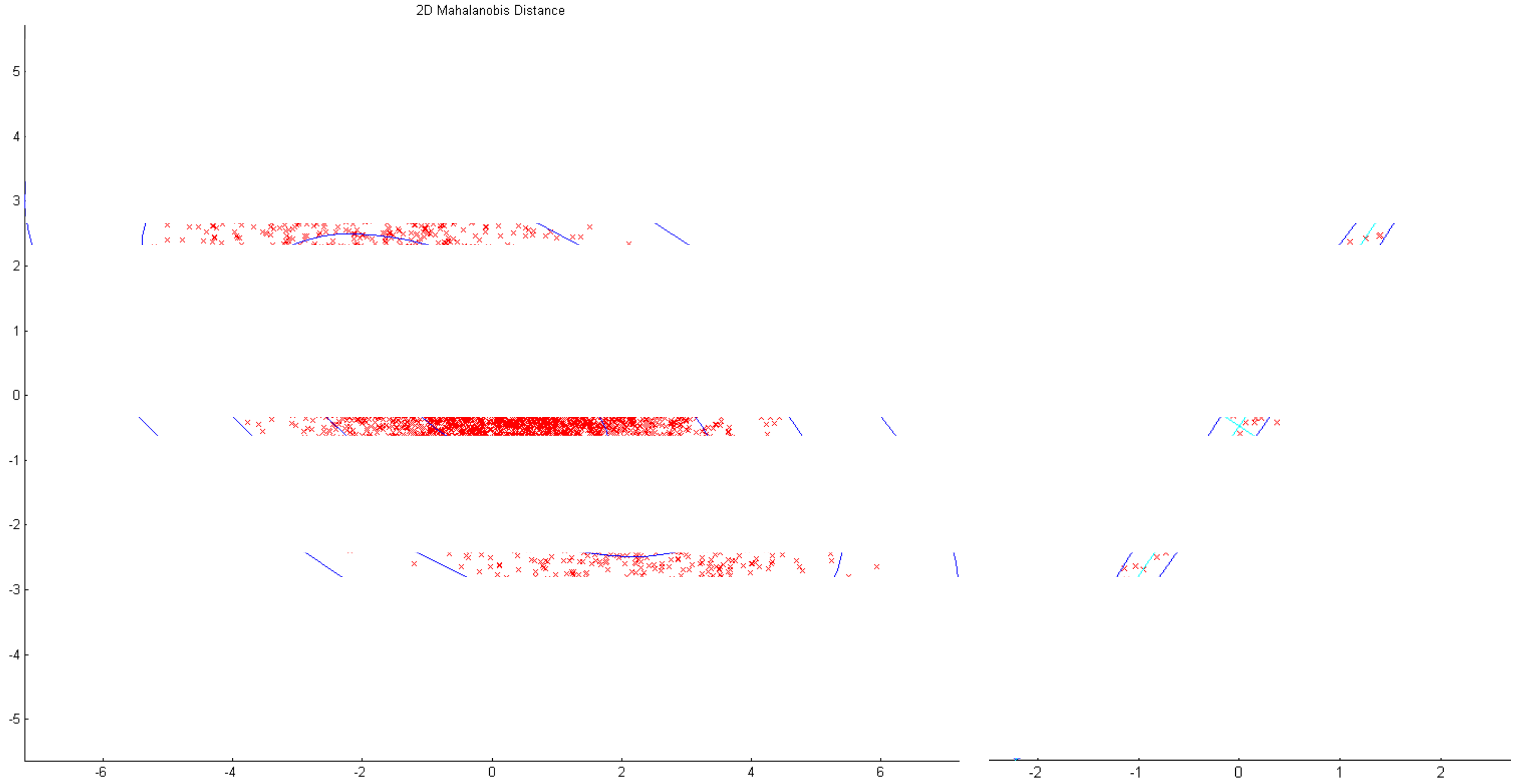
Marginalization – geometric intuition



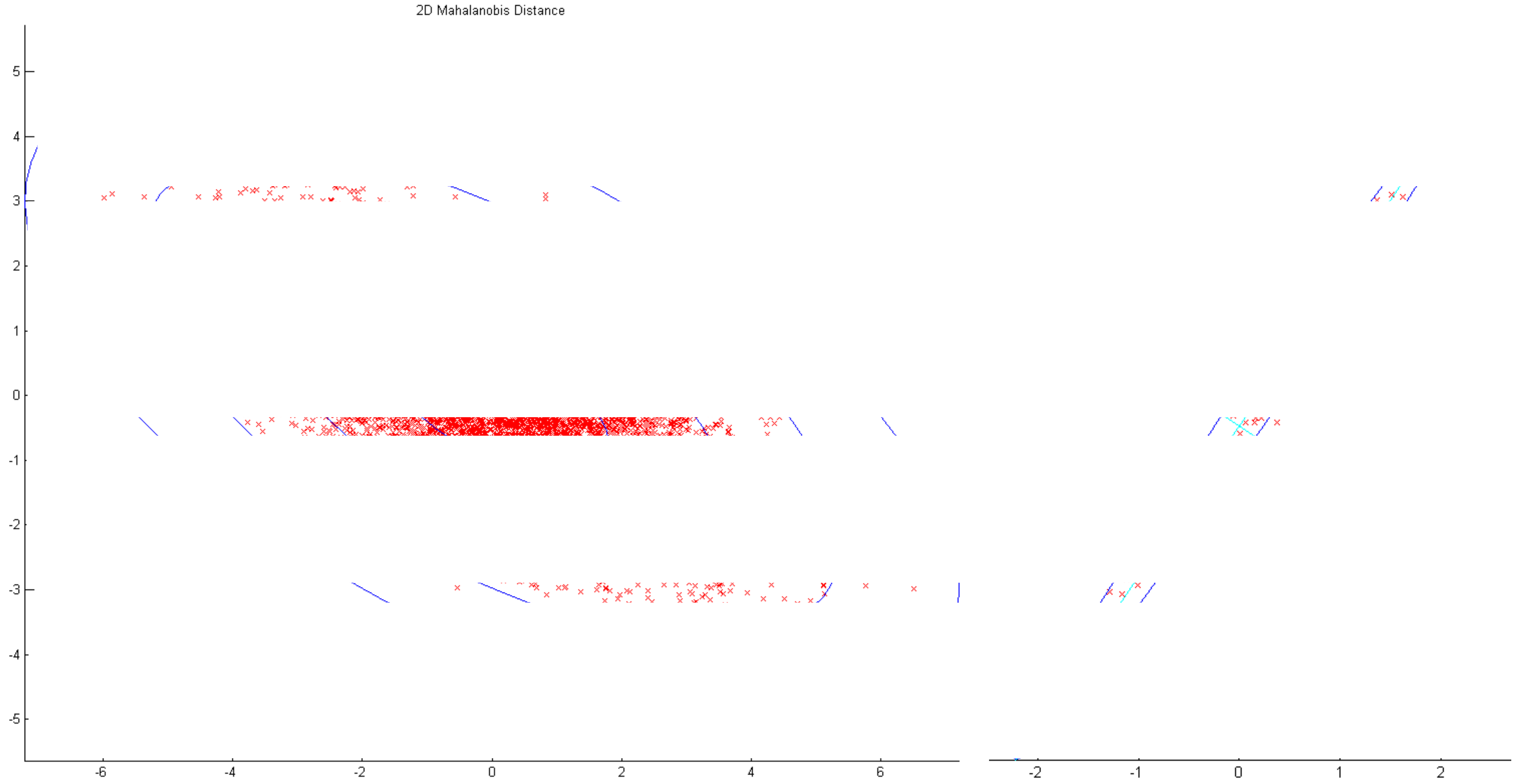
Conditioning – geometric intuition



Conditioning – geometric intuition

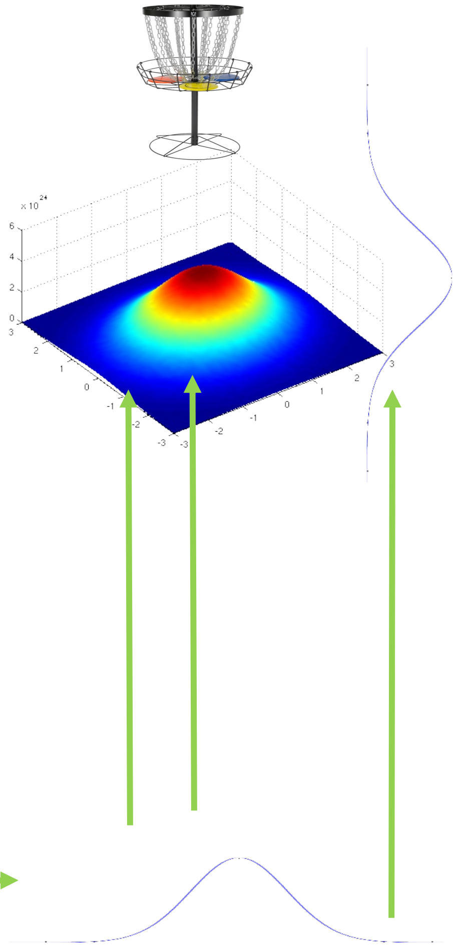


Conditioning – geometric intuition



Why covariance is constant?

Courtesy of hurriyetdailynews



Marginalization and Conditioning – how to

$\Lambda = \Omega$

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{\mu}_\alpha \\ \boldsymbol{\mu}_\beta \end{bmatrix}, \begin{bmatrix} \Sigma_{\alpha\alpha} & \Sigma_{\alpha\beta} \\ \Sigma_{\beta\alpha} & \Sigma_{\beta\beta} \end{bmatrix}\right) = \mathcal{N}^{-1}\left(\begin{bmatrix} \boldsymbol{\eta}_\alpha \\ \boldsymbol{\eta}_\beta \end{bmatrix}, \begin{bmatrix} \Lambda_{\alpha\alpha} & \Lambda_{\alpha\beta} \\ \Lambda_{\beta\alpha} & \Lambda_{\beta\beta} \end{bmatrix}\right)$$

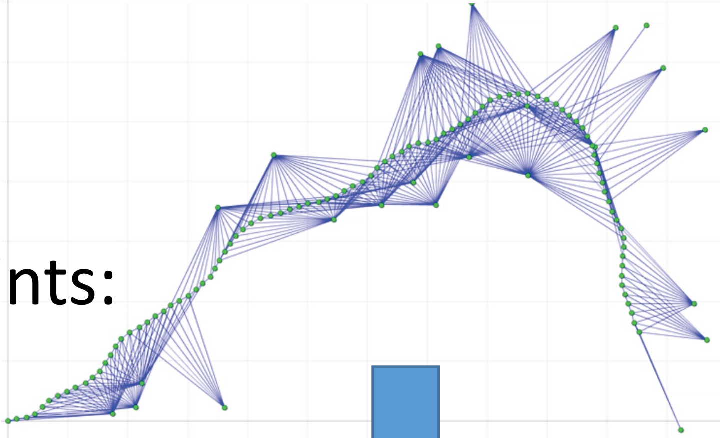
	MARGINALIZATION $p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha}, \boldsymbol{\beta})\, d\boldsymbol{\beta}$	CONDITIONING $p(\boldsymbol{\alpha} \mid \boldsymbol{\beta}) = p(\boldsymbol{\alpha}, \boldsymbol{\beta}) / p(\boldsymbol{\beta})$
COV. FORM	$\boldsymbol{\mu} = \boldsymbol{\mu}_\alpha$ $\Sigma = \Sigma_{\alpha\alpha}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma$ expensive $- \boldsymbol{\mu}_\beta)$ $\Sigma' =$ $_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$
INFO. FORM	$\boldsymbol{\eta} = \boldsymbol{\eta}_\alpha - \Lambda$ expensive $_\beta$ Λ $_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

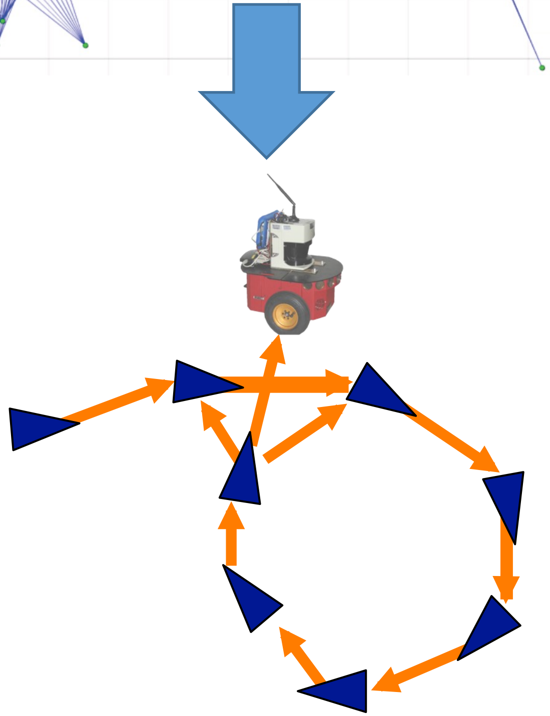
Courtesy: R. Eustice

Pose Graph - compromises

- We replaced our big factor graph with a pose graph
- How did we compromise?
- From each small factor graph of K cameras and P points:
  - We removed all points
  - We removed most cameras
  - $P(C_1,\dots,C_K,p_1,..p_P) \underbrace{\rightarrow}_{\text{marginalization}} P(C_1,C_K) \underbrace{\rightarrow}_{\text{conditioning}} P\,(C_K|C_1)$
- In the full factor graph, point p might have F cameras:
  - We used this track by parts, in each small factor graph:
  - $P(C_1,\dots,C_F,p_1) \rightarrow P(KF_2,p|KF_1) \cdot \cdot\, P(KF_N,\, p|KF_{N-1})$

Courtesy of Dellaert06ijrr: “Square Root SAM”





Courtesy of Cyrill Stachniss