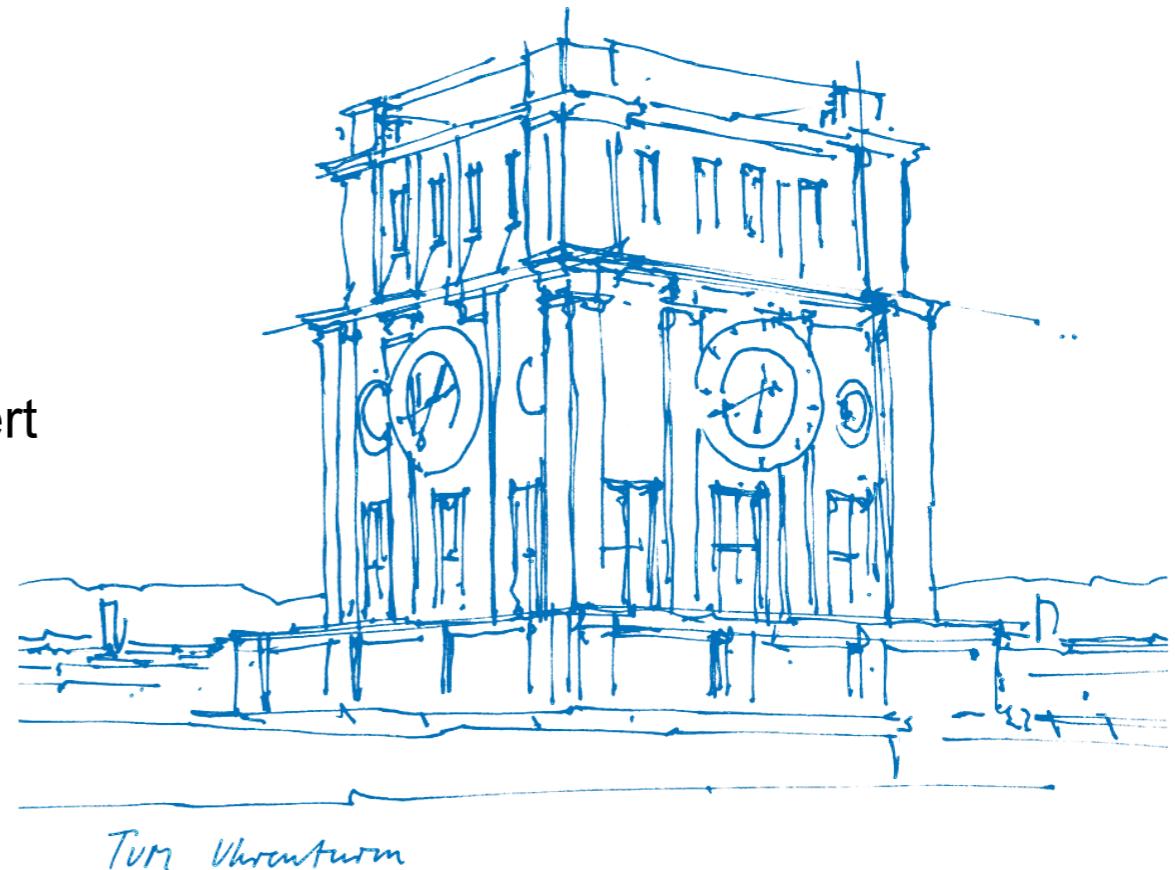


Practical Course: Vision Based Navigation

Lecture 4: Structure from Motion (SfM)

Dr. Vladyslav Usenko, Nikolaus Demmel, David Schubert
Prof. Dr. Daniel Cremers



Topics Covered

- Introduction
 - Structure from Motion (SfM)
 - Simultaneous Localization and Mapping (SLAM)
- Bundle Adjustment
 - Energy Function
 - Non-linear Least Squares
 - Exploiting the Sparse Structure
- Triangulation

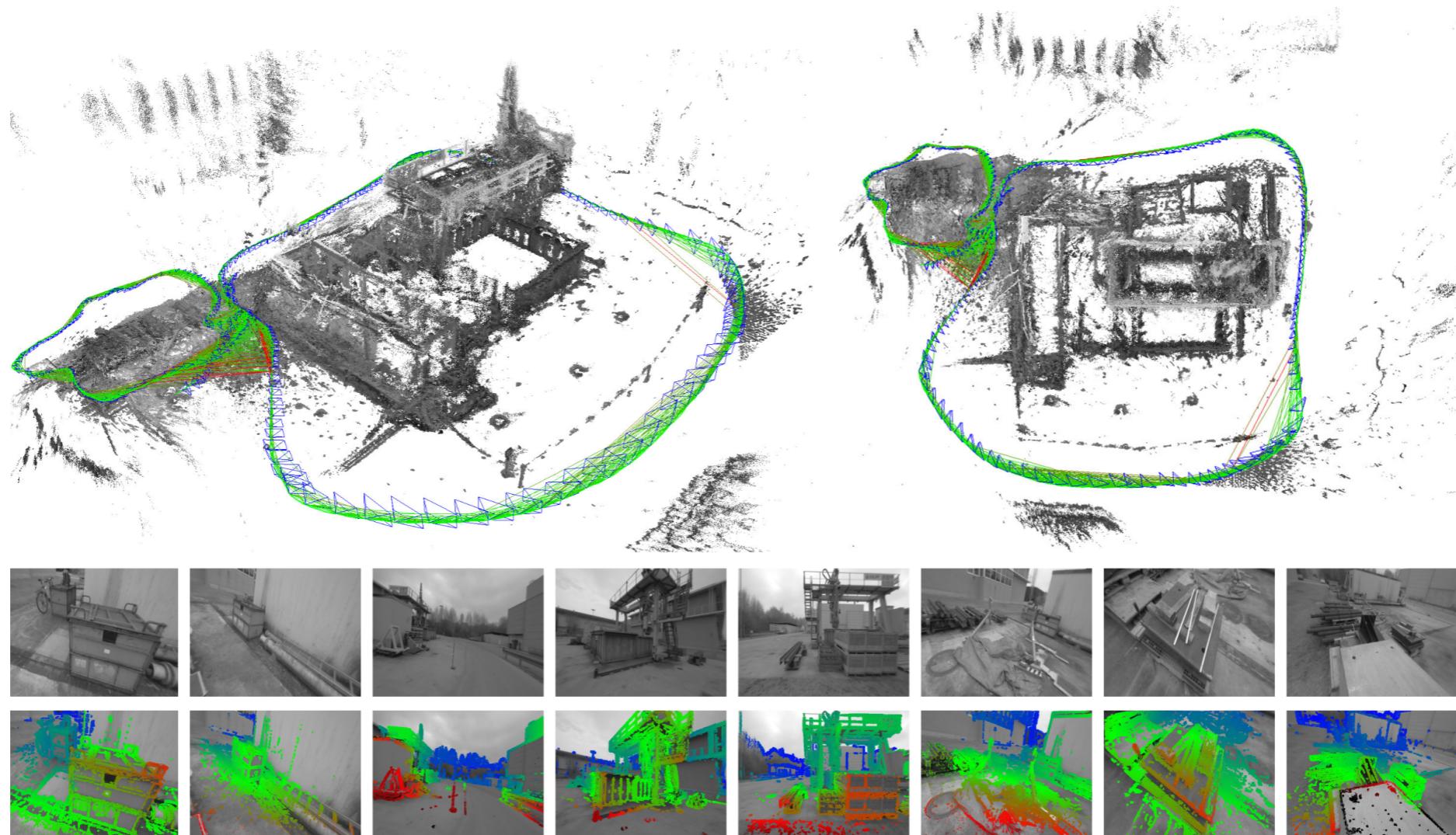
Structure from Motion



Agarwal et al., “Building Rome in a day”, ICCV 2009, “Dubrovnik” image set

- 3D reconstruction using a set of **unordered images**
- Requires estimation of 6DoF poses

Simultaneous Localization and Mapping (SLAM)



Engel et al., “LSD-SLAM: Large-Scale Direct Monocular SLAM”, ECCV 2014

- Estimate 6DoF poses and map from sequential image data
- Update once new frames arrive

Problem Definition SfM / Visual SLAM

Estimate camera poses and map from a set of images

- Input

Set of images $I_{0:t} = \{I_0, I_1, \dots, I_t\}$



Additional input possible

- Stereo
- Depth
- Inertial measurements
- Control input

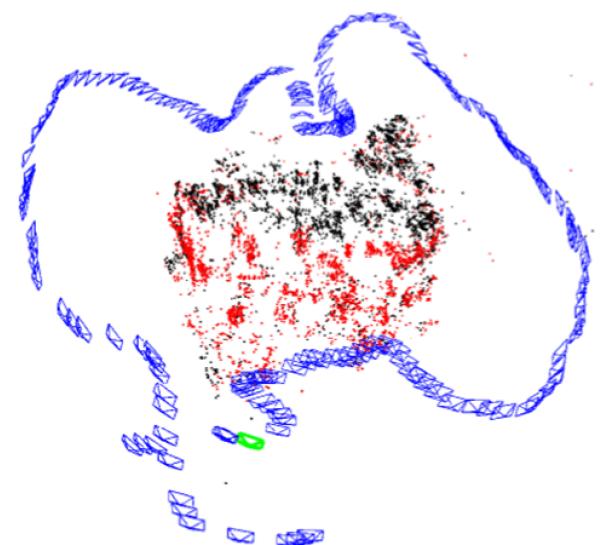


fr3/long_office_household sequence,
TUM RGB-D benchmark

- Output

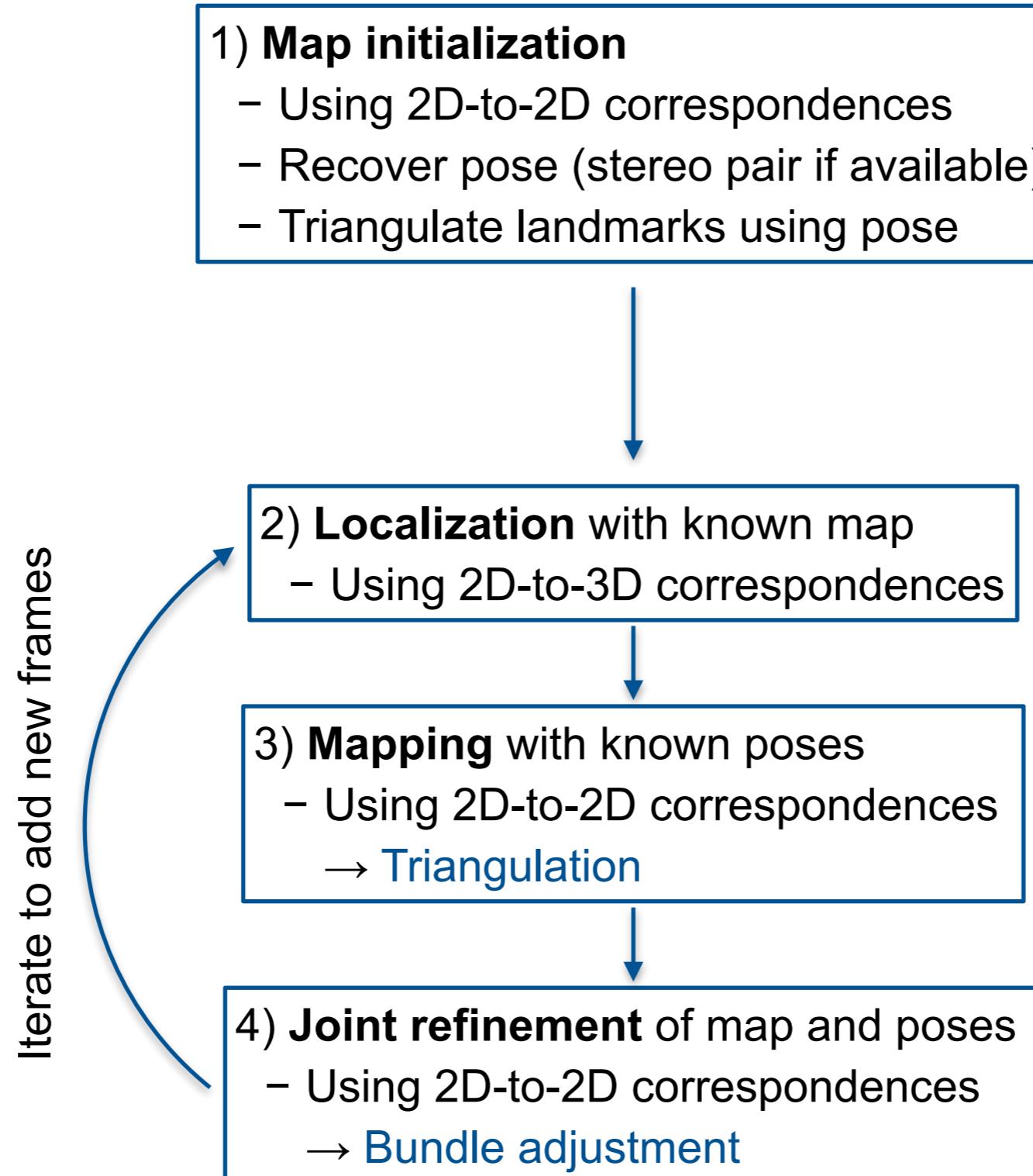
Camera pose estimates $\mathbf{T}_i \in \text{SE}(3)$,
also written as $\xi_i = (\log \mathbf{T}_i)^\vee$ $i \in \{0, 1, \dots, t\}$

Environment map M



Mur-Artal et al., 2015

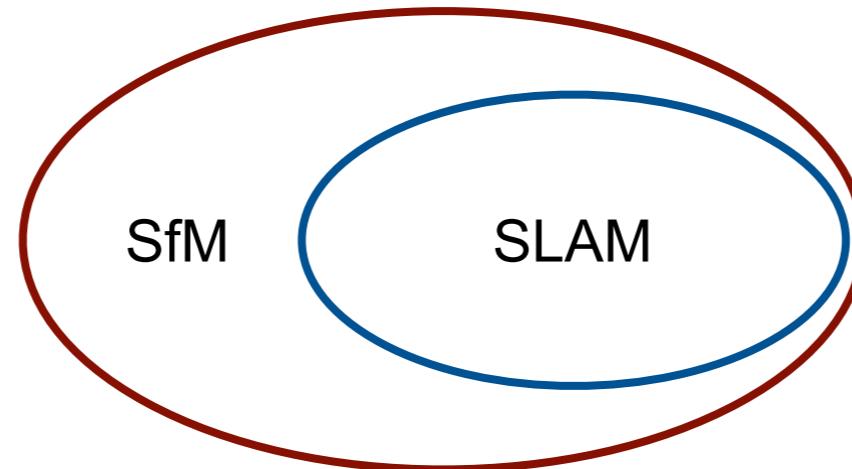
Typical SfM Pipeline



Visual SLAM

SLAM \subset SfM, with special focus:

- Sequential image data
- Data arrives sequentially
- Preferably realtime
- More focus on trajectory



Technical solutions:

- Windowed optimization
- Selection of keyframes
- Removal of keyframes (e.g. marginalization)

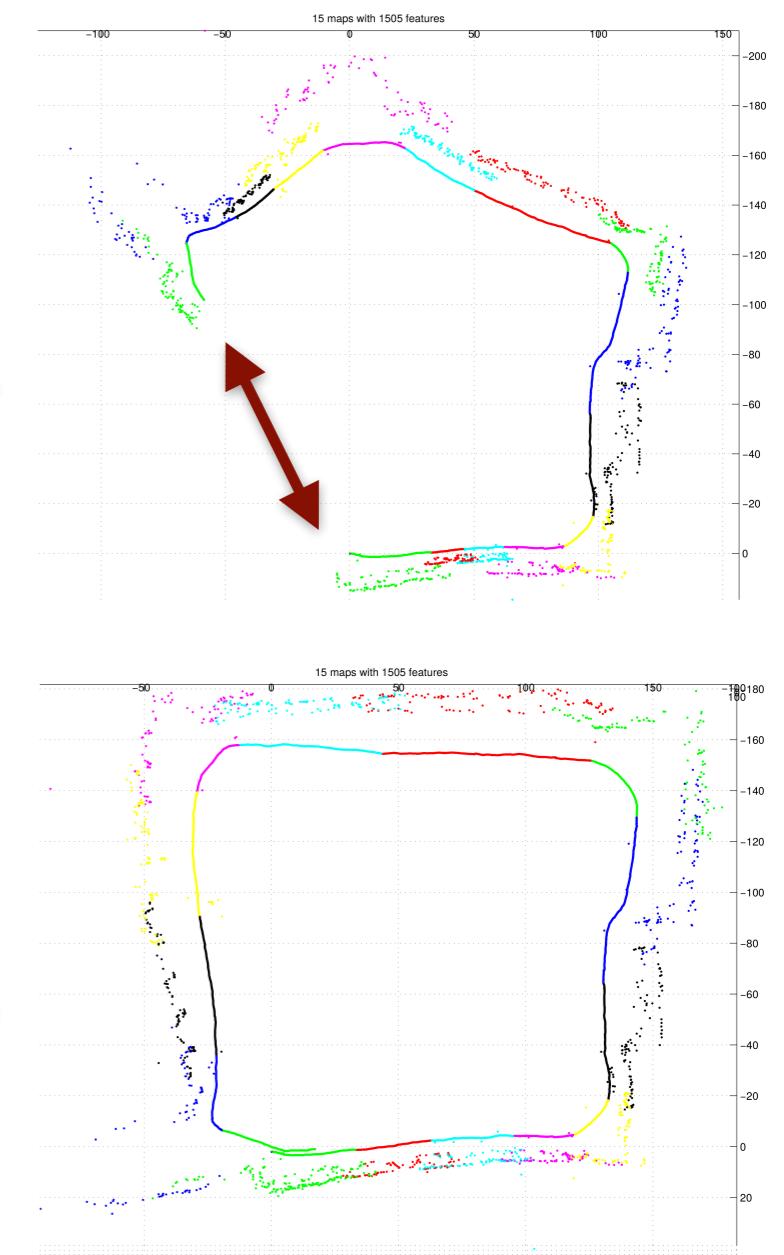
→ Accumulation of drift

- Detect loop closures
- Global mapping in separate thread
(e.g. pose graph optimization)

Odometry

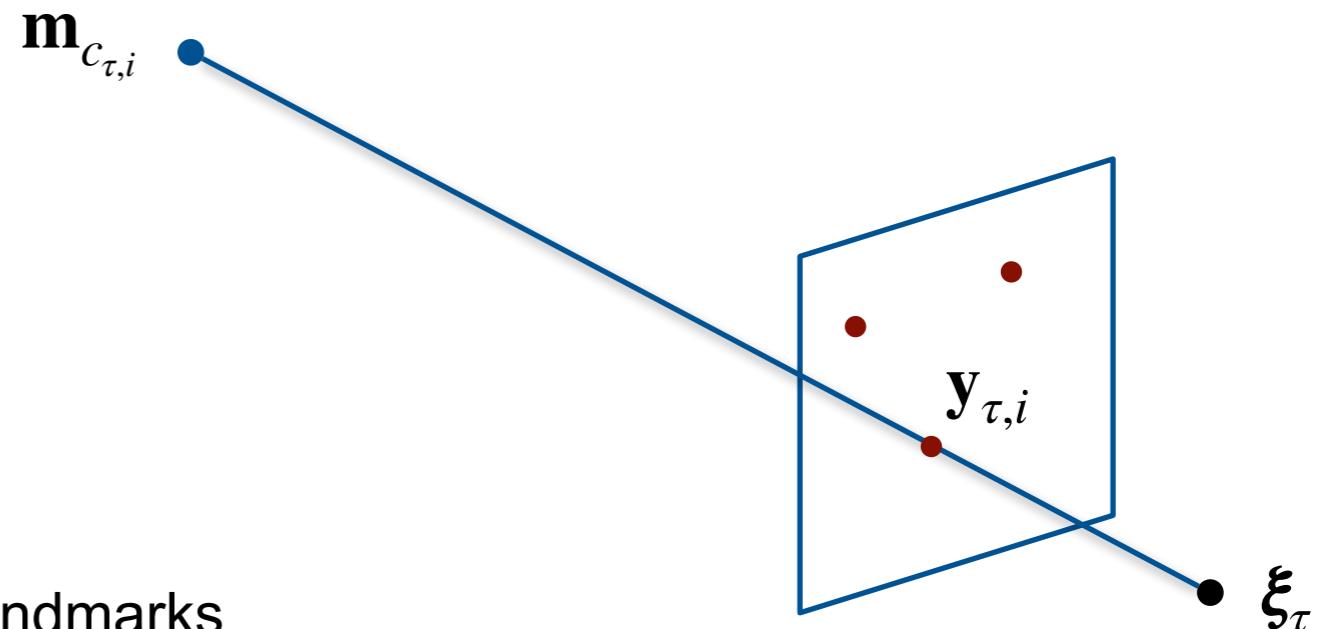
- No global mapping
- Incremental tracking only
- Local map possible

Loop closure



Clemente et al., RSS 2007

Landmarks and Features



- The map consists of 3D locations of landmarks

$$M = \{\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_S\}$$

- For image τ , the set of 2D image coordinates of detected features is denoted

$$Y_{\tau} = \{\mathbf{y}_{\tau,1}, \mathbf{y}_{\tau,2}, \dots, \mathbf{y}_{\tau,N}\}$$

- Known data association:

Feature i in image τ corresponds to landmark $j = c_{\tau,i}$ $(1 \leq i \leq N, 1 \leq j \leq S)$

Bundle Adjustment Energy

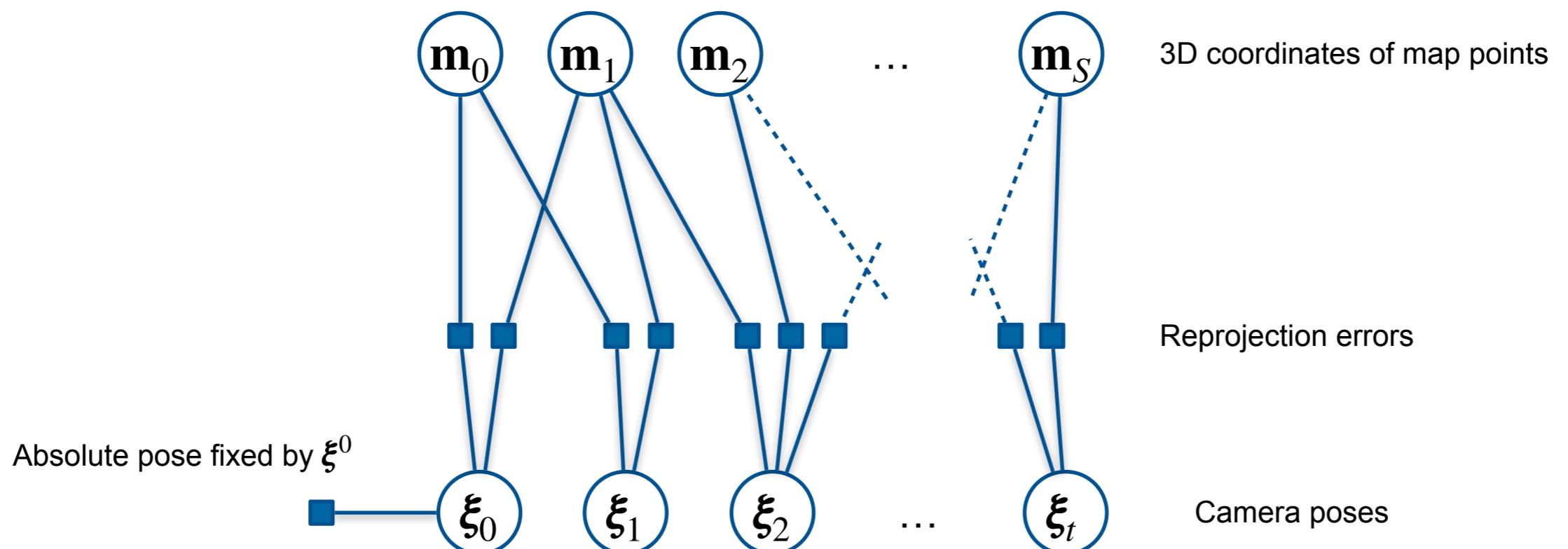
$$E(\xi_{0:t}, M) = \frac{1}{2} (\xi_0 \ominus \xi^0)^\top \Sigma_{0,\xi}^{-1} (\xi_0 \ominus \xi^0)$$

Absolute pose prior

$$+ \frac{1}{2} \sum_{\tau=0}^t \sum_{i=1}^{N_\tau} \left(\mathbf{y}_{\tau,i} - h(\xi_\tau, \mathbf{m}_{c_{\tau,i}}) \right)^\top \Sigma_{\mathbf{y}_{\tau,i}}^{-1} \left(\mathbf{y}_{\tau,i} - h(\xi_\tau, \mathbf{m}_{c_{\tau,i}}) \right)$$

Reprojection error

- Pose prior: Fix absolute pose ambiguity
 - In this case equivalent to keeping $\xi_0 = \xi^0$
 - Keep absolute pose information e.g. when first frame is marginalized
- Additional prior to fix scale ambiguity might be necessary



Energy Function as Non-linear Least Squares

- Residuals as function of state vector \mathbf{x}

$$\mathbf{r}^0(\mathbf{x}) := \xi_0 \ominus \xi^0$$

$$\mathbf{r}_{t,i}^y(\mathbf{x}) := \mathbf{y}_{t,i} - h(\xi_t, \mathbf{m}_{c_{t,i}})$$

$$\mathbf{x} := \begin{pmatrix} \xi_0 \\ \vdots \\ \xi_t \\ \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_S \end{pmatrix}$$

- Stack the residuals in a vector-valued function und collect the residual covariances on the diagonal blocks of a square matrix

$$\mathbf{r}(\mathbf{x}) := \begin{pmatrix} \mathbf{r}^0(\mathbf{x}) \\ \mathbf{r}_{0,1}^y(\mathbf{x}) \\ \vdots \\ \mathbf{r}_{t,N_t}^y(\mathbf{x}) \end{pmatrix}$$

$$\mathbf{W} := \begin{pmatrix} \Sigma_{0,\xi}^{-1} & 0 & \cdots & 0 \\ 0 & \Sigma_{\mathbf{y}_{0,1}}^{-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \Sigma_{\mathbf{y}_{t,N_t}}^{-1} \end{pmatrix}$$

- Rewrite energy function as $E(\mathbf{x}) = \frac{1}{2} \mathbf{r}(\mathbf{x})^\top \mathbf{W} \mathbf{r}(\mathbf{x})$

Recap: Gauss-Newton Method

- Idea: Approximate Newton's method to minimize $E(\mathbf{x})$
- Approximate $E(\mathbf{x})$ through linearization of residuals

$$\begin{aligned}
 \tilde{E}(\mathbf{x}) &= \frac{1}{2} \tilde{\mathbf{r}}(\mathbf{x})^\top \mathbf{W} \tilde{\mathbf{r}}(\mathbf{x}) && k \text{ iteration index} \\
 &= \frac{1}{2} \left(\mathbf{r}(\mathbf{x}_k) + \mathbf{J}_k (\mathbf{x} - \mathbf{x}_k) \right)^\top \mathbf{W} \left(\mathbf{r}(\mathbf{x}_k) + \mathbf{J}_k (\mathbf{x} - \mathbf{x}_k) \right) \\
 &= \frac{1}{2} \mathbf{r}(\mathbf{x}_k)^\top \mathbf{W} \mathbf{r}(\mathbf{x}_k) + \underbrace{\mathbf{r}(\mathbf{x}_k)^\top \mathbf{W} \mathbf{J}_k (\mathbf{x} - \mathbf{x}_k)}_{=: \mathbf{b}_k^\top} + \frac{1}{2} (\mathbf{x} - \mathbf{x}_k)^\top \underbrace{\mathbf{J}_k^\top \mathbf{W} \mathbf{J}_k}_{=: \mathbf{H}_k} (\mathbf{x} - \mathbf{x}_k)
 \end{aligned}$$

- Finding root of gradient as in Newton's method leads to update rule

$$\nabla_{\mathbf{x}} \tilde{E}(\mathbf{x}) = \mathbf{b}_k^\top + (\mathbf{x} - \mathbf{x}_k)^\top \mathbf{H}_k$$

$$\nabla_{\mathbf{x}} \tilde{E}(\mathbf{x}) = 0 \quad \text{iff} \quad \mathbf{x} = \mathbf{x}_k - \mathbf{H}_k^{-1} \mathbf{b}_k$$

$$\boxed{\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{H}_k^{-1} \mathbf{b}_k}$$

- Pros:
 - Faster convergence than gradient descent (approx. quadratic convergence rate)
- Cons:
 - Divergence if too far from local optimum (\mathbf{H} not positive definite)
 - Solution quality depends on initial guess

Structure of the Bundle Adjustment Problem

- \mathbf{b}_k and \mathbf{H}_k sum terms from individual residuals:

$$\mathbf{b}_k = \mathbf{b}_k^0 + \sum_{\tau=0}^t \sum_{i=1}^{N_\tau} \mathbf{b}_k^{\tau,i} = (\mathbf{J}_k^0)^\top \boldsymbol{\Sigma}_{0,\xi}^{-1} \mathbf{r}^0(\mathbf{x}_k) + \sum_{\tau=0}^t \sum_{i=1}^{N_\tau} (\mathbf{J}_k^{\tau,i})^\top \boldsymbol{\Sigma}_{\mathbf{y}_{\tau,i}}^{-1} \mathbf{r}_{\tau,i}^{\mathbf{y}}(\mathbf{x}_k)$$

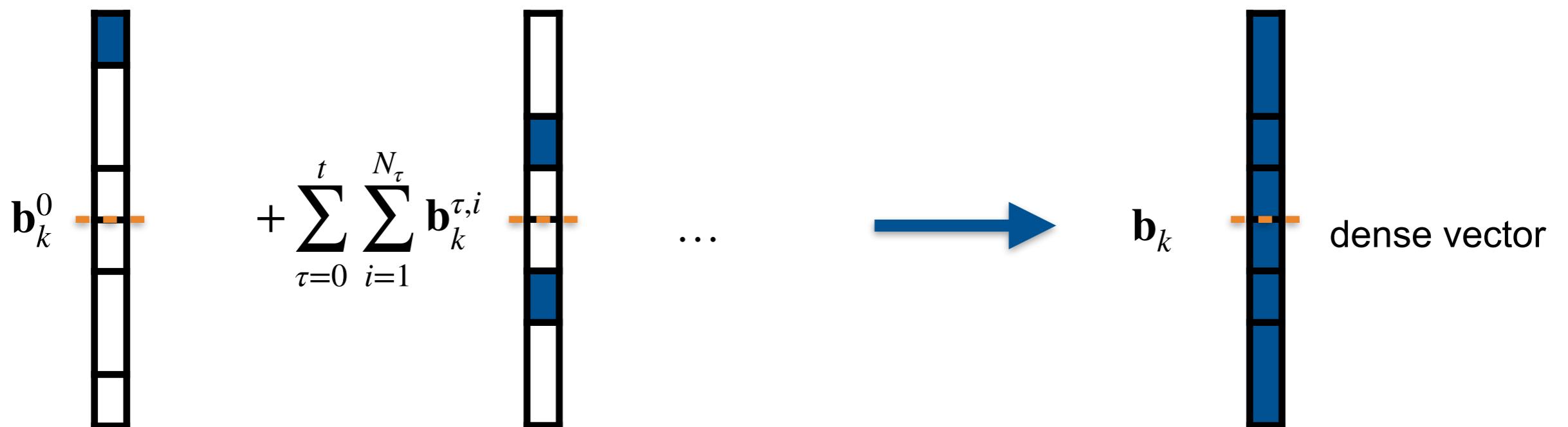
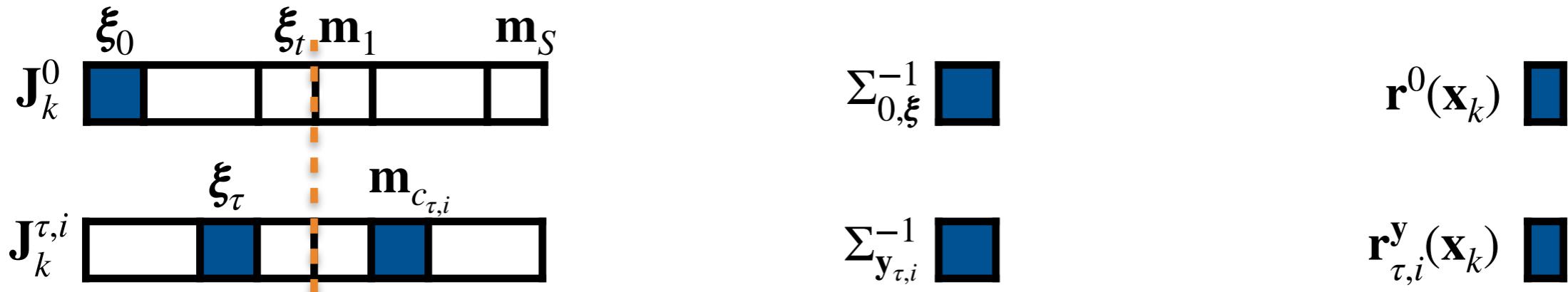
$$\mathbf{H}_k = \mathbf{H}_k^0 + \sum_{\tau=0}^t \sum_{i=1}^{N_\tau} \mathbf{H}_k^{\tau,i} = (\mathbf{J}_k^0)^\top \boldsymbol{\Sigma}_{0,\xi}^{-1} (\mathbf{J}_k^0) + \sum_{\tau=0}^t \sum_{i=1}^{N_\tau} (\mathbf{J}_k^{\tau,i})^\top \boldsymbol{\Sigma}_{\mathbf{y}_{\tau,i}}^{-1} (\mathbf{J}_k^{\tau,i})$$

\mathbf{J}_k^0 Jacobian of pose prior

$\mathbf{J}_k^{\tau,i}$ Jacobian of residuals for feature i in image τ

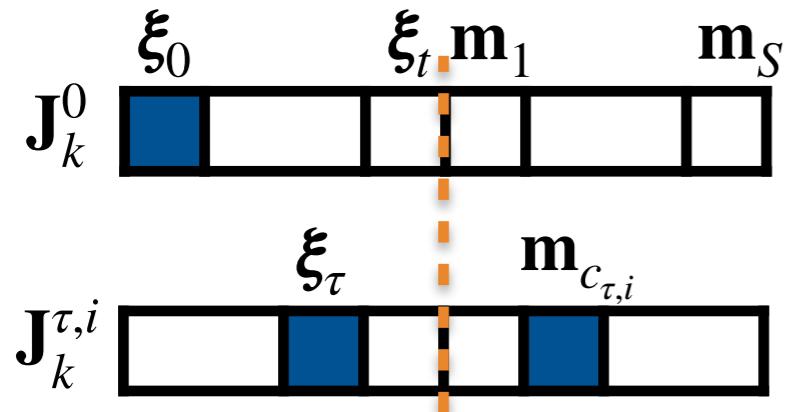
- What is the structure of these terms?

Structure of the Bundle Adjustment Problem



$$\mathbf{b}_k = \mathbf{b}_k^0 + \sum_{\tau=0}^t \sum_{i=1}^{N_{\tau}} \mathbf{b}_k^{\tau,i} = (\mathbf{J}_k^0)^{\top} \Sigma_{0,\xi}^{-1} \mathbf{r}^0(\mathbf{x}_k) + \sum_{\tau=0}^t \sum_{i=1}^{N_{\tau}} (\mathbf{J}_k^{\tau,i})^{\top} \Sigma_{\mathbf{y}_{\tau,i}}^{-1} \mathbf{r}_{\tau,i}^y(\mathbf{x}_k)$$

Structure of the Bundle Adjustment Problem

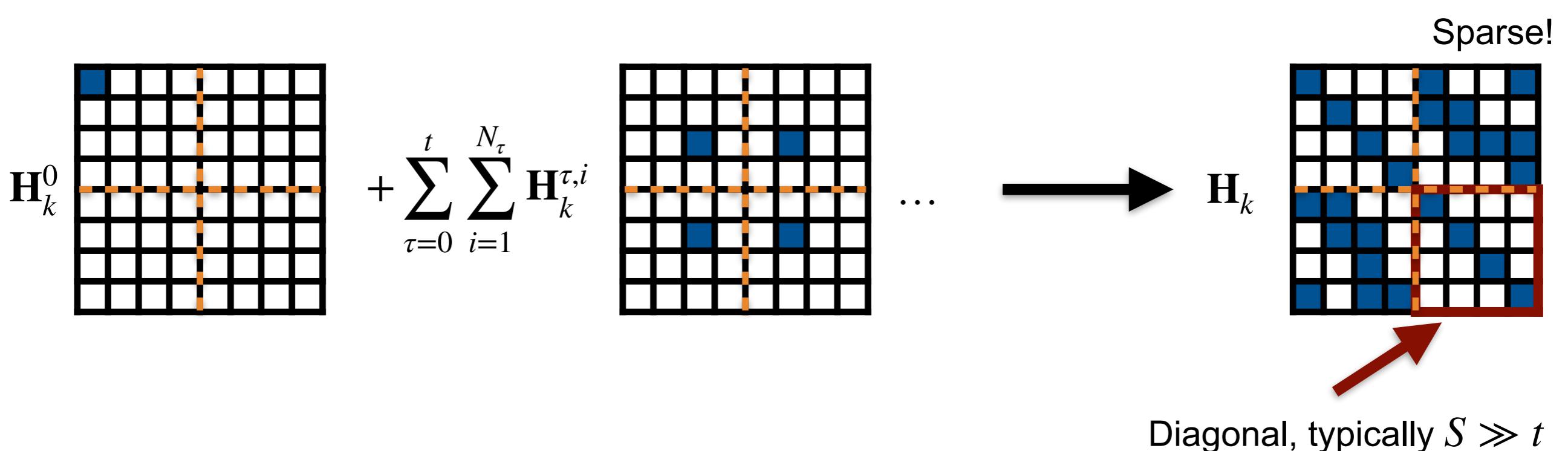


$$\Sigma_{0,\xi}^{-1}$$

$$\mathbf{r}^0(\mathbf{x}_k)$$

$$\Sigma_{\mathbf{y}_{\tau,i}}^{-1}$$

$$\mathbf{r}_{\tau,i}^{\mathbf{y}}(\mathbf{x}_k)$$

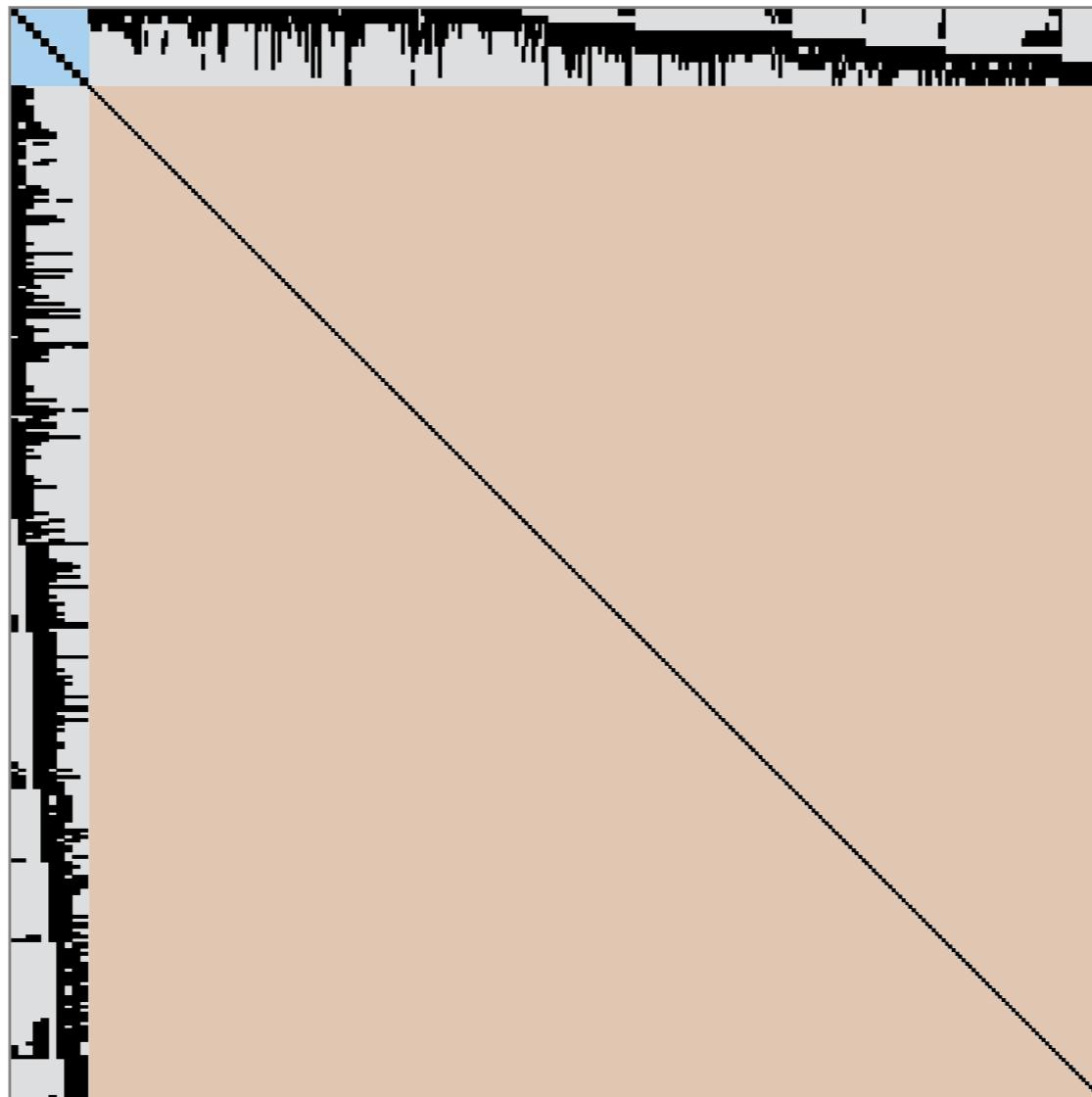


$$\mathbf{H}_k = \mathbf{H}_k^0 + \sum_{\tau=0}^t \sum_{i=1}^{N_{\tau}} \mathbf{H}_k^{\tau,i} = (\mathbf{J}_k^0)^T \Sigma_{0,\xi}^{-1} (\mathbf{J}_k^0) + \sum_{\tau=0}^t \sum_{i=1}^{N_{\tau}} (\mathbf{J}_k^{\tau,i})^T \Sigma_{\mathbf{y}_{\tau,i}}^{-1} (\mathbf{J}_k^{\tau,i})$$

Example Hessian of a BA Problem

Pose dimensions
(10 poses)

$$H_k =$$



Lourakis et al., 2009

Large, but sparse!

How to invert efficiently?

Exploiting the Sparse Structure

- Idea:

Apply the Schur complement to solve the system in a partitioned way

$$\mathbf{H}_k \Delta \mathbf{x} = -\mathbf{b}_k \quad \longrightarrow \quad \begin{pmatrix} \mathbf{H}_{\xi\xi} & \mathbf{H}_{\xi m} \\ \mathbf{H}_{m\xi} & \mathbf{H}_{mm} \end{pmatrix} \begin{pmatrix} \Delta \mathbf{x}_\xi \\ \Delta \mathbf{x}_m \end{pmatrix} = - \begin{pmatrix} \mathbf{b}_\xi \\ \mathbf{b}_m \end{pmatrix}$$

$$\longrightarrow \Delta \mathbf{x}_\xi = - \left(\mathbf{H}_{\xi\xi} - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{H}_{m\xi} \right)^{-1} \left(\mathbf{b}_\xi - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{b}_m \right)$$

$$\longrightarrow \Delta \mathbf{x}_m = - \mathbf{H}_{mm}^{-1} \left(\mathbf{b}_m + \mathbf{H}_{m\xi} \Delta \mathbf{x}_\xi \right)$$

- Is this any better?

Exploiting the Sparse Structure

- What is the structure of the two sub-problems?

- Poses:

$$\Delta \mathbf{x}_\xi = - \left(\mathbf{H}_{\xi\xi} - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{H}_{m\xi} \right)^{-1} \left(\mathbf{b}_\xi - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{b}_m \right)$$

$$\mathbf{H}_{\xi\xi} - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{H}_{m\xi} = \mathbf{H}_{\xi\xi} - \sum_{j=1}^S \mathbf{H}_{\xi m_j} \mathbf{H}_{m_j m_j}^{-1} \mathbf{H}_{m_j \xi}$$

Reduced pose Hessian

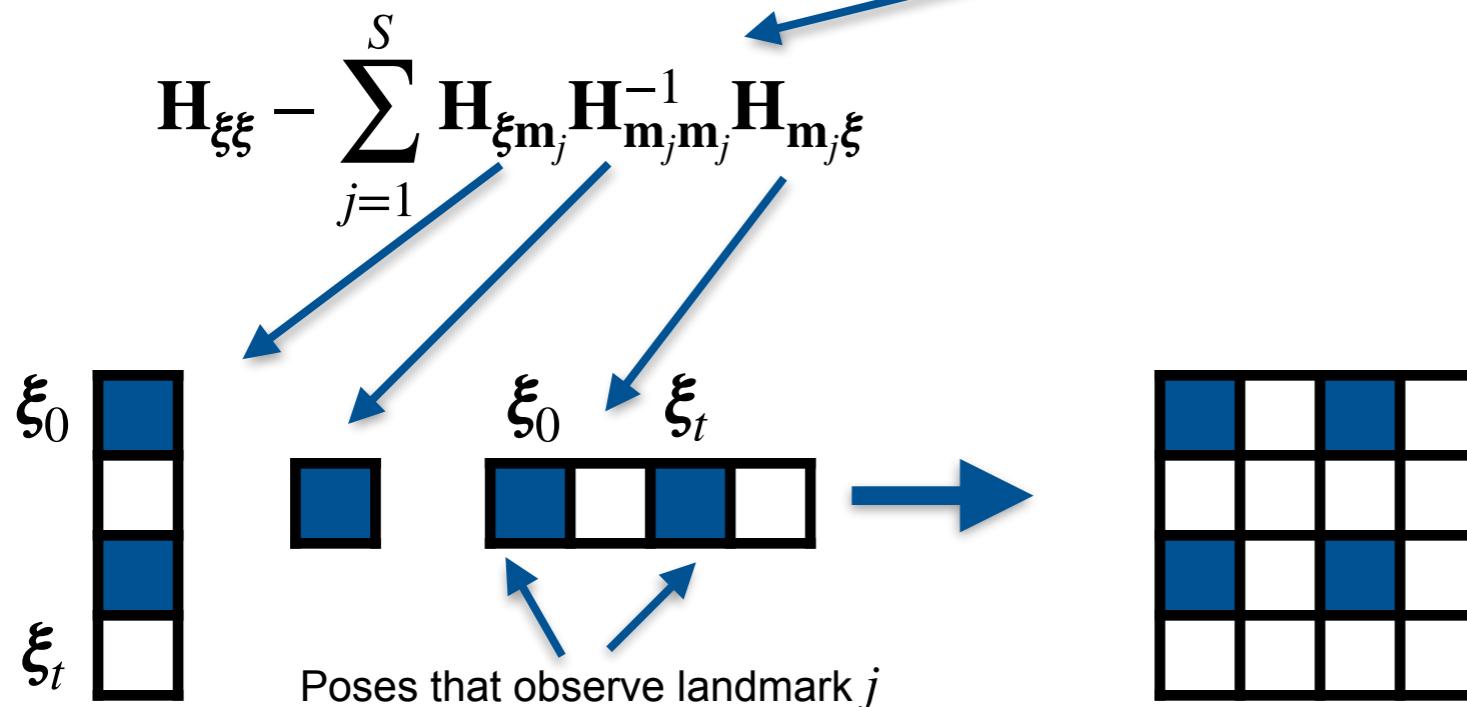
$$\mathbf{b}_\xi - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{b}_m = \mathbf{b}_\xi - \sum_{j=1}^S \mathbf{H}_{\xi m_j} \mathbf{H}_{m_j m_j}^{-1} \mathbf{b}_{m_j}$$

Exploiting the Sparse Structure

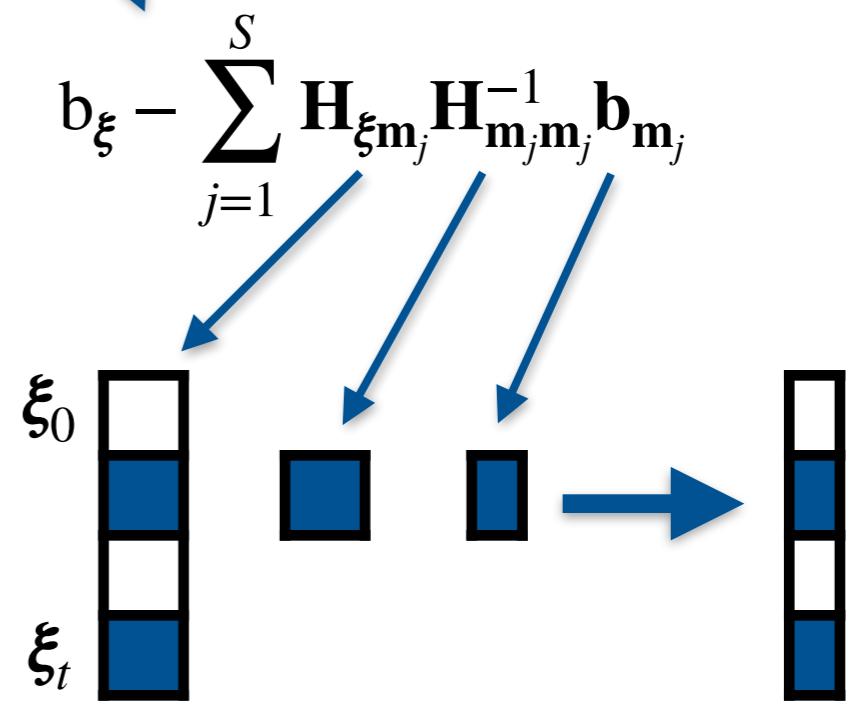
- What is the structure of the two sub-problems?

- Poses:

$$\Delta \mathbf{x}_\xi = - \left(\underline{\mathbf{H}_{\xi\xi} - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{H}_{m\xi}} \right)^{-1} \left(\underline{\mathbf{b}_\xi - \mathbf{H}_{\xi m} \mathbf{H}_{mm}^{-1} \mathbf{b}_m} \right)$$



$$\mathbf{H}_{\xi\xi} - \sum_{j=1}^S \mathbf{H}_{\xi m_j} \mathbf{H}_{m_j m_j}^{-1} \mathbf{H}_{m_j \xi} =$$



$$\mathbf{b}_\xi - \sum_{j=1}^S \mathbf{H}_{\xi m_j} \mathbf{H}_{m_j m_j}^{-1} \mathbf{b}_{m_j} =$$

Exploiting the Sparse Structure

- What is the structure of the two sub-problems?
- Landmarks:

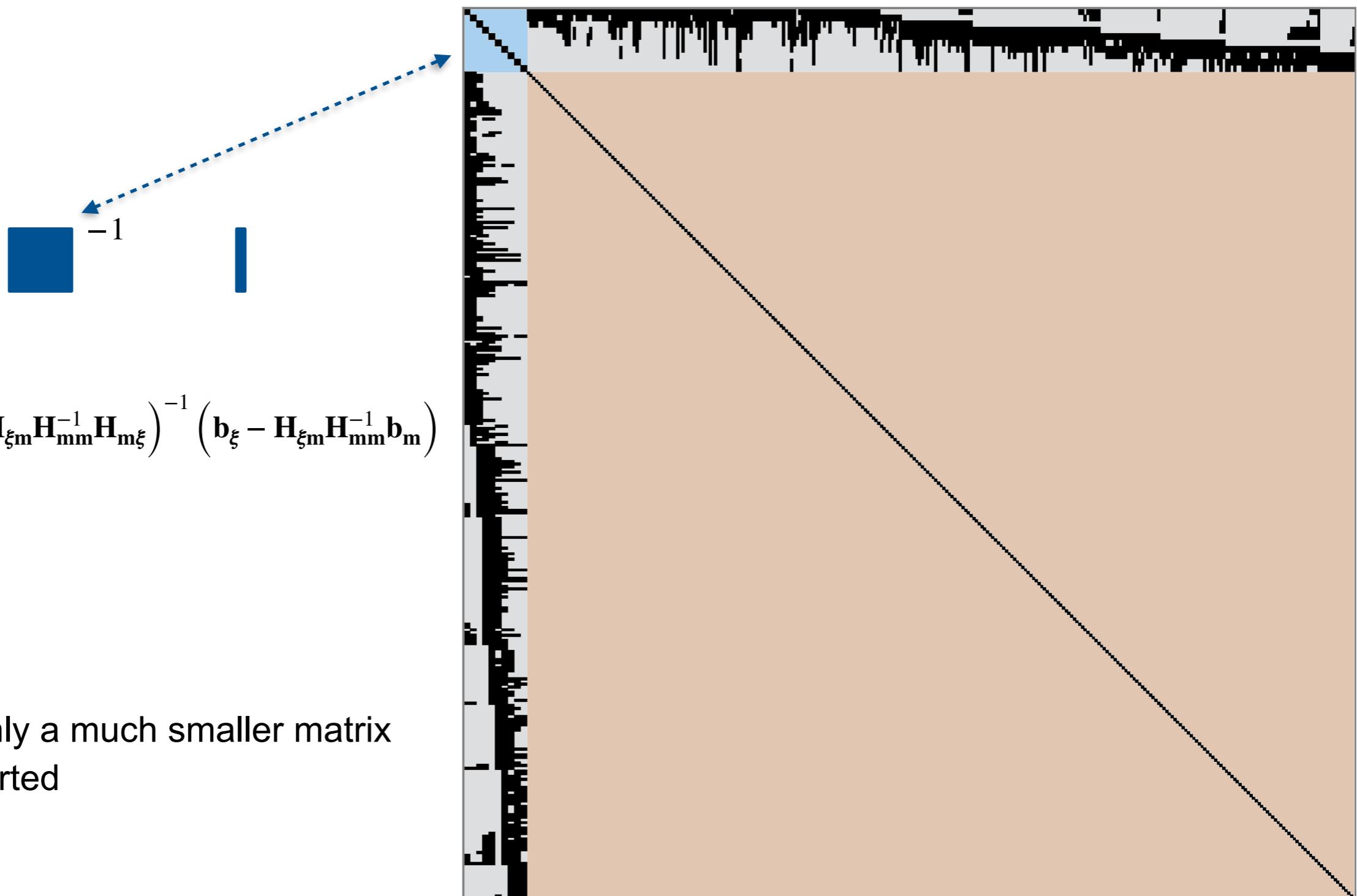
$$\Delta \mathbf{x}_m = -\mathbf{H}_{mm}^{-1} (\mathbf{b}_m + \mathbf{H}_{m\xi} \Delta \mathbf{x}_\xi)$$

→

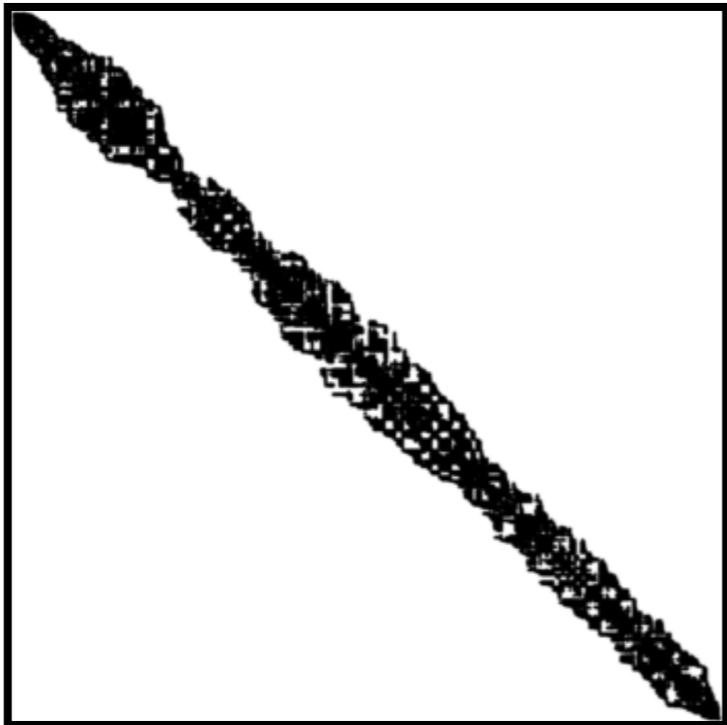
$$\Delta \mathbf{x}_{m_j} = -\mathbf{H}_{m_j m_j}^{-1} (\mathbf{b}_{m_j} + \mathbf{H}_{m_j \xi} \Delta \mathbf{x}_\xi)$$
$$= - \left[\begin{array}{c} \boxed{} \\ \vdots \\ \boxed{} \end{array} \right] \left(\left[\begin{array}{c} \boxed{} \\ \vdots \\ \boxed{} \end{array} \right] + \left[\begin{array}{cc|cc} \xi_0 & \xi_t & & \\ \hline & & \ddots & \\ & & & \ddots \end{array} \right] \right)$$

- Landmark-wise solution
- Comparably small matrix operations
- Only involves poses that observe the landmark

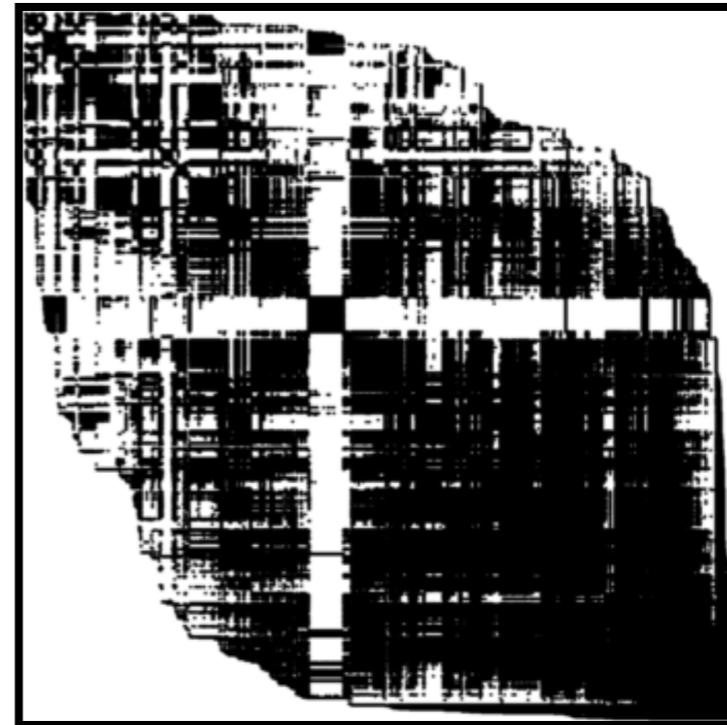
Exploiting the sparse structure



Exploiting the Sparse Structure



Camera on a moving vehicle
(6375 images)



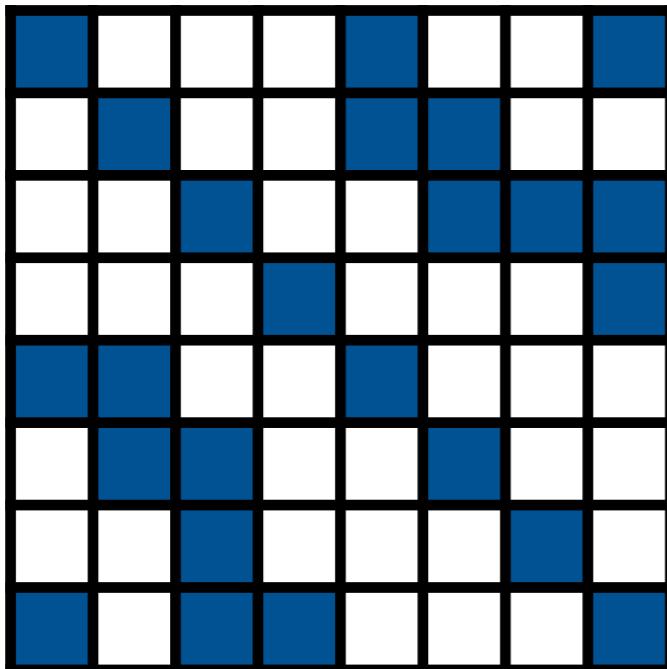
Flickr image search “Dubrovnik”
(4585 images)

Agarwal et al., ECCV 2010

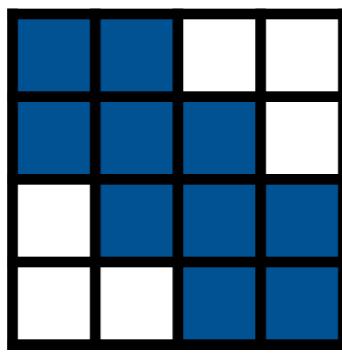
- Reduced pose Hessian can still have a sparse structure
- For many camera poses with many shared observations, the inversion of the reduced pose Hessian is still computationally expensive!
- Exploit further structure, e.g. using variable reordering or hierarchical decomposition

Effect of Loop Closures on the Hessian

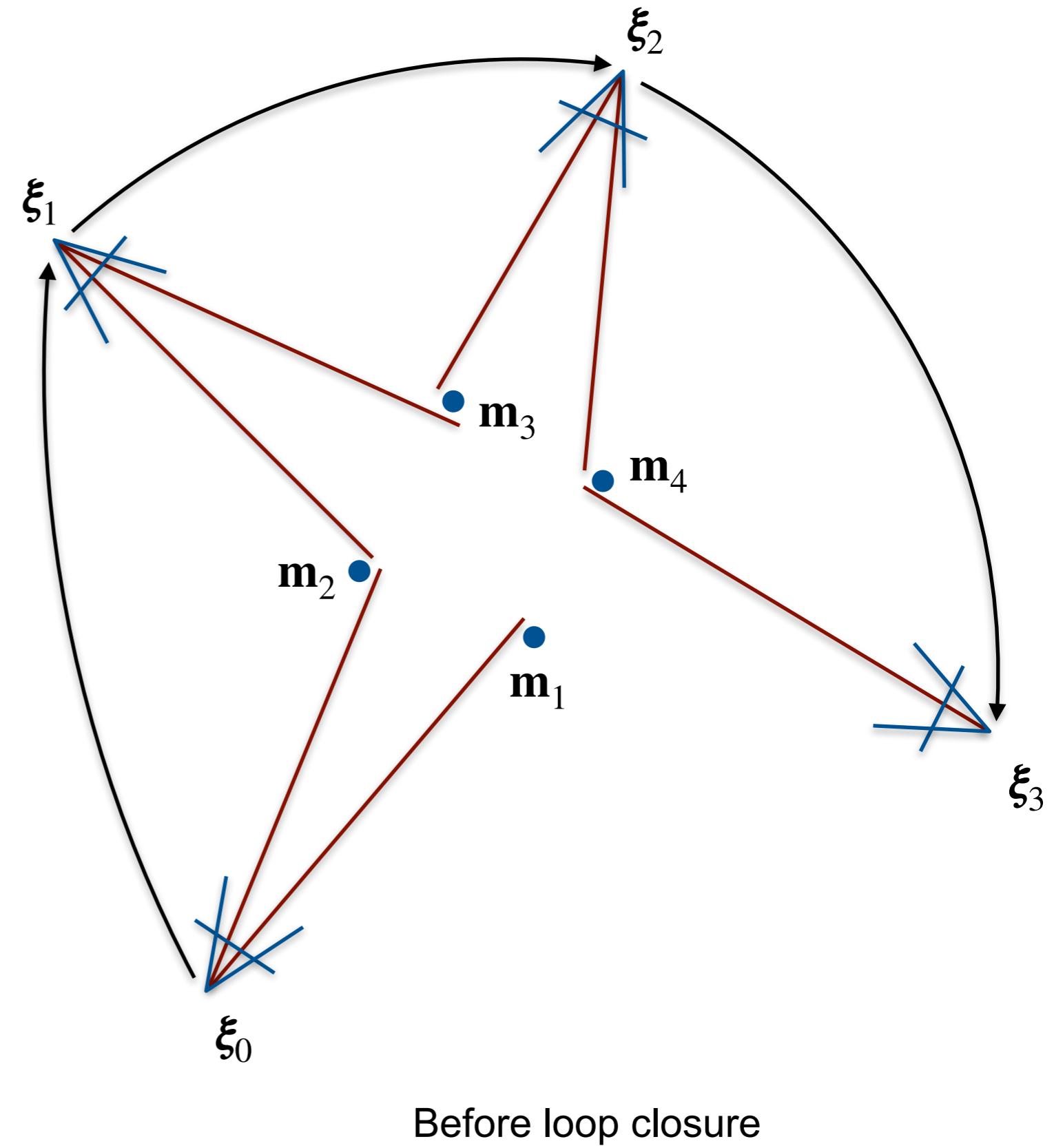
Full Hessian



Reduced pose Hessian

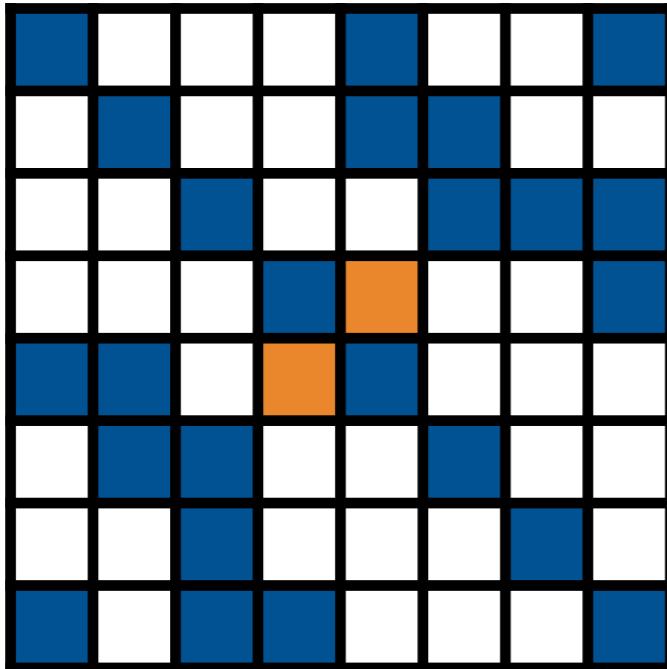


Band matrix

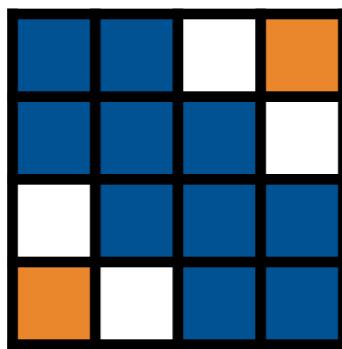


Effect of Loop Closures on the Hessian

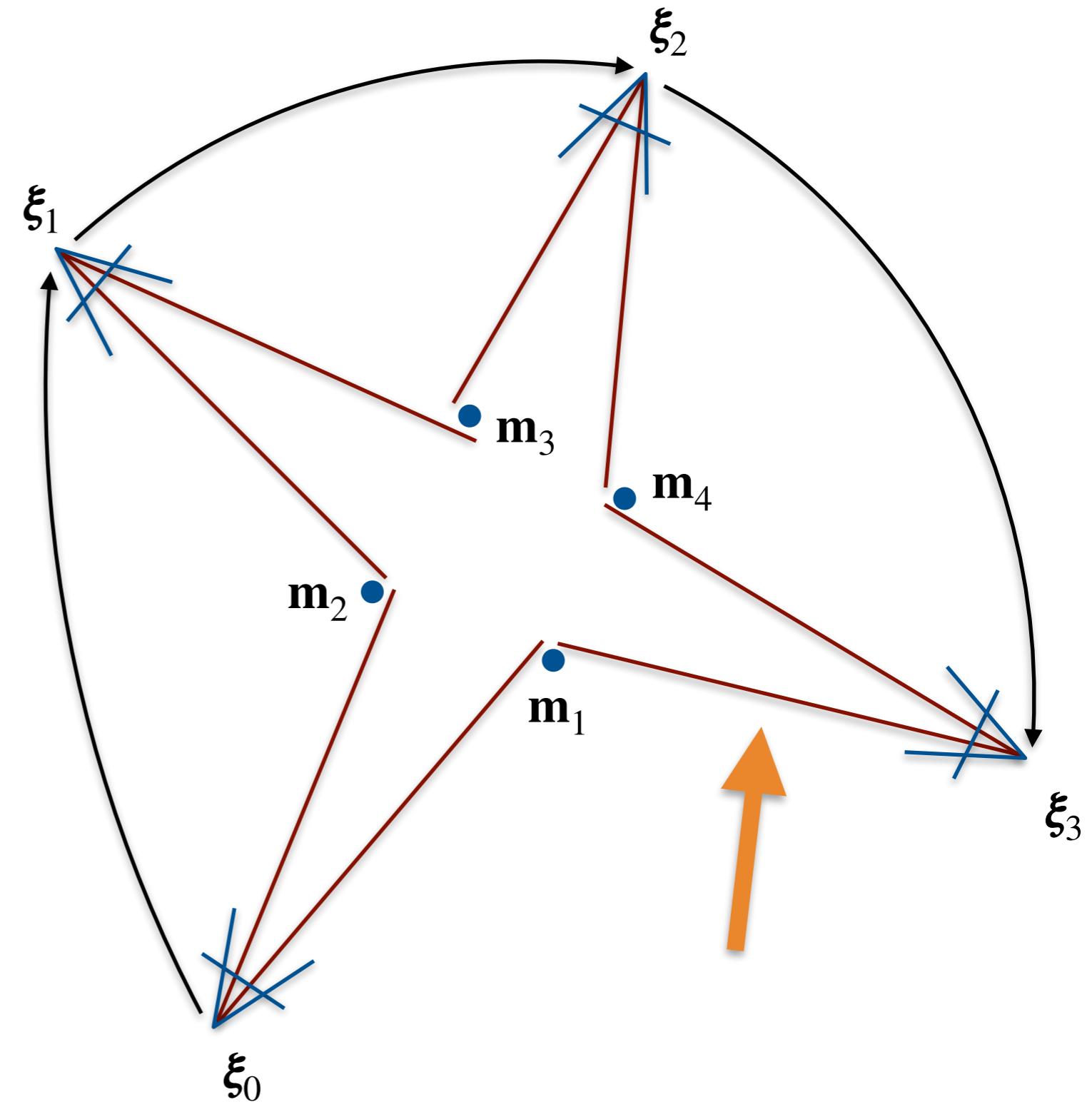
Full Hessian



Reduced pose Hessian



No band matrix: costlier to solve



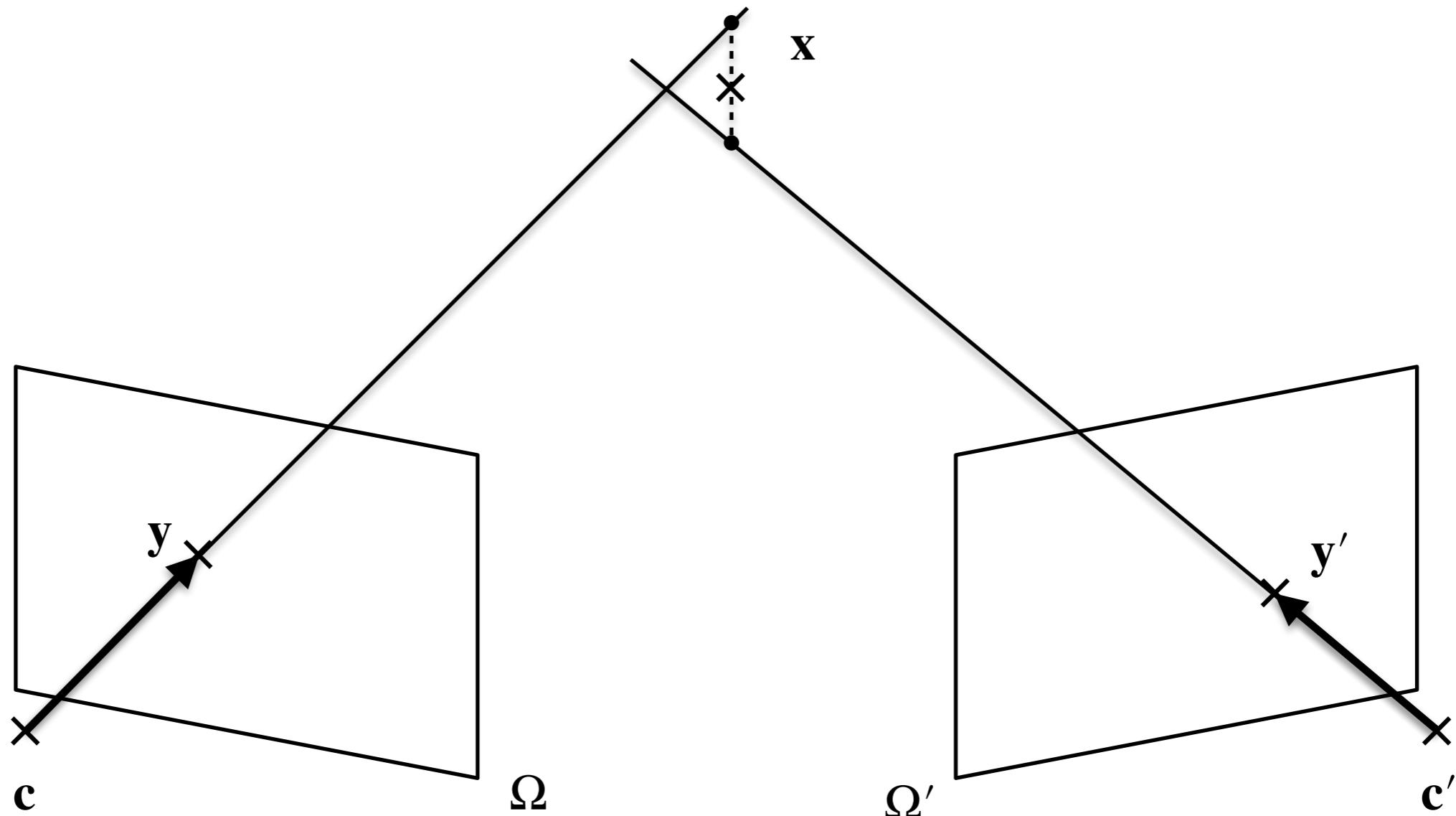
After loop closure

Further Considerations

Many methods to improve convergence / robustness / run-time efficiency, e.g.

- Use matrix decompositions (e.g. Cholesky) to perform inversions
- Levenberg-Marquardt optimization improves basin of convergence
- Heavier-tail distributions / robust norms on the residuals can be implemented using iteratively reweighted least squares
- Preconditioning
- Hierarchical optimization
- Variable reordering
- Delayed relinearization

Triangulation



- Find landmark position given the camera poses
- Ideally, the rays should intersect
- In practice, many sources of error: pose estimates, feature detections and camera model / intrinsic parameters

Triangulation

- Goal: Reconstruct 3D point $\tilde{\mathbf{x}} = (x, y, z, w)^\top \in \mathbb{P}^3$ from 2D image observations $\{\mathbf{y}_1, \dots, \mathbf{y}_N\}$ for known camera poses $\{\mathbf{T}_1, \dots, \mathbf{T}_N\}$
- Linear solution: Find 3D point such that reprojections equal its projection

– For each image i , let $\mathbf{T}_i = \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ and $\mathbf{y}_i = \begin{pmatrix} u \\ v \end{pmatrix}$

– Projecting $\tilde{\mathbf{x}}$ yields $\mathbf{y}'_i = \pi(\mathbf{T}_i \tilde{\mathbf{x}}) = \begin{pmatrix} \mathbf{p}_1 \tilde{\mathbf{x}} / \mathbf{p}_3 \tilde{\mathbf{x}} \\ \mathbf{p}_2 \tilde{\mathbf{x}} / \mathbf{p}_3 \tilde{\mathbf{x}} \end{pmatrix}$

– Requiring $\mathbf{y}'_i = \mathbf{y}_i$ gives two linear equations per image:

$$\mathbf{p}_1 \tilde{\mathbf{x}} = u \mathbf{p}_3 \tilde{\mathbf{x}}$$

$$\mathbf{p}_2 \tilde{\mathbf{x}} = v \mathbf{p}_3 \tilde{\mathbf{x}}$$

- Leads to system of linear equations $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{0}$, two approaches to solve:
- Set $w = 1$ and solve non-homogeneous least squares problem
 - Find nullspace of \mathbf{A} using SVD, then scale such that $w = 1$

- Non-linear least squares on reprojection errors (more accurate):

$$\min_{\mathbf{x}} \left\{ \sum_{i=1}^N \|\mathbf{y}_i - \mathbf{y}'_i\|_2^2 \right\}$$

- Different solutions for different methods in the presence of noise

Exercises

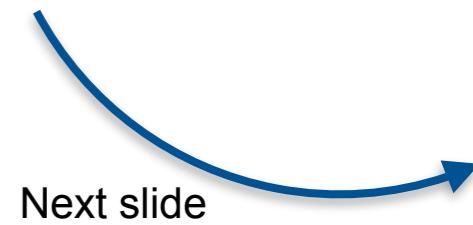
Exercise sheet 4

- Implement components of SfM pipeline
- BA: Ceres can do the Schur complement
- Triangulation: use OpenGV's triangulate function



```
ceres::Solver::Options ceres_options;
ceres_options.max_num_iterations = 20;
ceres_options.linear_solver_type =
ceres::SPARSE_SCHUR;
ceres_options.num_threads = 8;
ceres::Solver::Summary summary;
Solve(ceres_options, &problem,
&summary);
std::cout << summary.FullReport() <<
std::endl;
```

Next slide



Exercise sheet 5

- Implement components of odometry
- Similar to sheet 4, but:
 - More efficient 2D-3D matching using approximate pose of previous frame
 - New keyframe if number of matches too small
 - New landmarks by triangulation from stereo pair
 - Keep runtime bounded by removing old keyframes

	Original	Reduced
Parameter blocks	4896	4892
Parameters	15354	15324
Effective parameters	15190	15162
Residual blocks	24014	24014
Residuals	48028	48028
Minimizer	TRUST_REGION	
Sparse linear algebra library	SUITE_SPARSE	
Trust region strategy	LEVENBERG_MARQUARDT	
	Given	Used
Linear solver	SPARSE_SCHUR	SPARSE_SCHUR
Threads	8	8
Linear solver ordering	AUTOMATIC	4730,162
Schur structure	2,3,6	2,3,6
Cost:		
Initial	3.979886e+03	
Final	3.766801e+03	
Change	2.130843e+02	
Minimizer iterations	21	
Successful steps	21	
Unsuccessful steps	0	
Time (in seconds):		
Preprocessor	0.048047	
Residual only evaluation	0.069569 (20)	
Jacobian & residual evaluation	0.388923 (21)	
Linear solver	0.586967 (20)	
Minimizer	1.134797	
Postprocessor	0.001068	
Total	1.183913	
Termination:	NO_CONVERGENCE (Maximum number of iterations reached. Number of iterations: 20.)	

Summary

SfM

- Estimate map and camera poses from set of images
- SLAM: Sequential data, real-time
- Odometry: No global mapping

Bundle Adjustment

- Non-linear least squares problem
- Sparse structure of Hessian can be exploited for efficient inversion

Triangulation

- Linear and non-linear algorithms
- Differences in the presence of noise