Sparse Extended Information Filters: Insights into Sparsification

August 3, 2005

Ryan Eustice*, Matthew Walter, & John Leonard

*Woods Hole Oceanographic Institution
Massachusetts Institute of Technology

Roadmap

I Introduction

a. Normal form vs. Information form in SLAM
b. Motivation for Sparse Extended Information Filters

II Sparsification in SEIFS

III Modified sparsification rule

IV Results

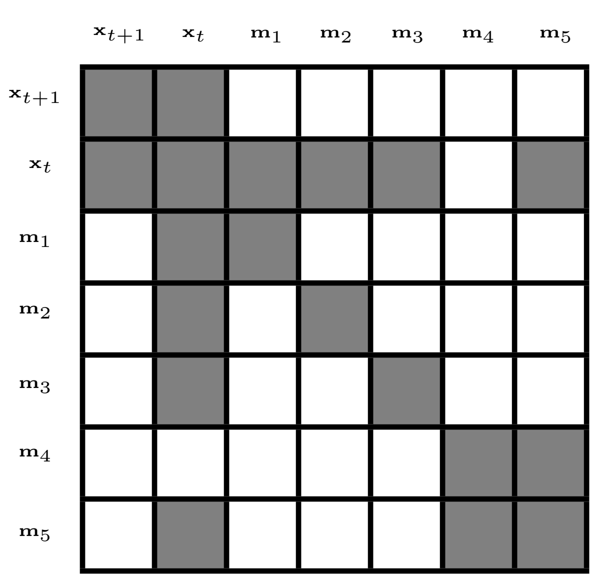
V Conclusions

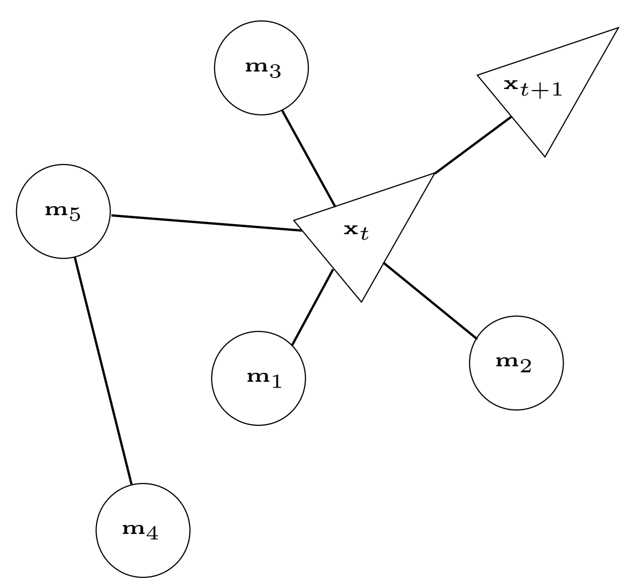
Canonical Gaussian Parameterization

$$\boldsymbol{\xi}_t \sim \mathcal{N}\left(\boldsymbol{\mu}_t, \Sigma_t\right)$$
$$\sim \mathcal{N}^{-1}\left(\boldsymbol{\eta}_t, \Lambda_t\right) \longleftarrow$$

$\Lambda_t = \Sigma_t^{-1}$	information matrix
$\boldsymbol{\eta}_t = \Lambda_t \boldsymbol{\mu}_t$	information vector

- Encodes Markov random field





Represents independence relationships

$$p\left(\mathbf{x}_{t+1} | \mathbf{x}_t, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4, \mathbf{m}_5\right) = p\left(\mathbf{x}_{t+1} | \mathbf{x}_t\right)$$

Duality of standard and information forms

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

	Covariance Form	Information Form
Marginalization $p(\boldsymbol{\alpha}) = \int p(\boldsymbol{\alpha},\boldsymbol{\beta})\, d\boldsymbol{\beta}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha$ $\Sigma' = \Sigma_{\alpha\alpha}$ easy	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda' = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$ hard
Conditioning $p(\boldsymbol{\alpha}\vert\boldsymbol{\beta}) = \frac{p(\boldsymbol{\alpha},\boldsymbol{\beta})}{p(\boldsymbol{\beta})}$	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

Consequences for EKF and EIF

	Covariance Form	Information Form
Marg.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha$ $\Sigma' = \Sigma_{\alpha\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda' = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$
Cond.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

I. Measurement update step: conditioning

$$\mathbf{z}_t = \mathbf{h}(\boldsymbol{\xi}_t) + \mathbf{v}_t$$
$$\approx \mathbf{h}(\boldsymbol{\mu}_t) + \mathrm{H}\left(\boldsymbol{\xi}_t - \boldsymbol{\mu}_t\right) + \mathbf{v}_t$$

$$\mathrm{H} = \begin{bmatrix} \mathbf{0} & \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_i} & \cdots & \mathbf{0} & \cdots & \frac{\partial \mathbf{h}}{\partial \mathbf{x}_j} & \cdots & \mathbf{0} \end{bmatrix}$$

$$\boldsymbol{\eta}_t' = \boldsymbol{\eta}_t + \mathrm{H}^\top \mathrm{R}^{-1}\left(\mathbf{z}_t - \mathbf{h}(\boldsymbol{\mu}_t) + \mathrm{H}\boldsymbol{\mu}\right)$$
$$\Lambda_t' = \Lambda_t + \mathrm{H}^\top \mathrm{R}^{-1}\mathrm{H}$$

↑ non-zero only for observed features

a. Computationally cheap*

b. Creates/strengthens links with observed features only

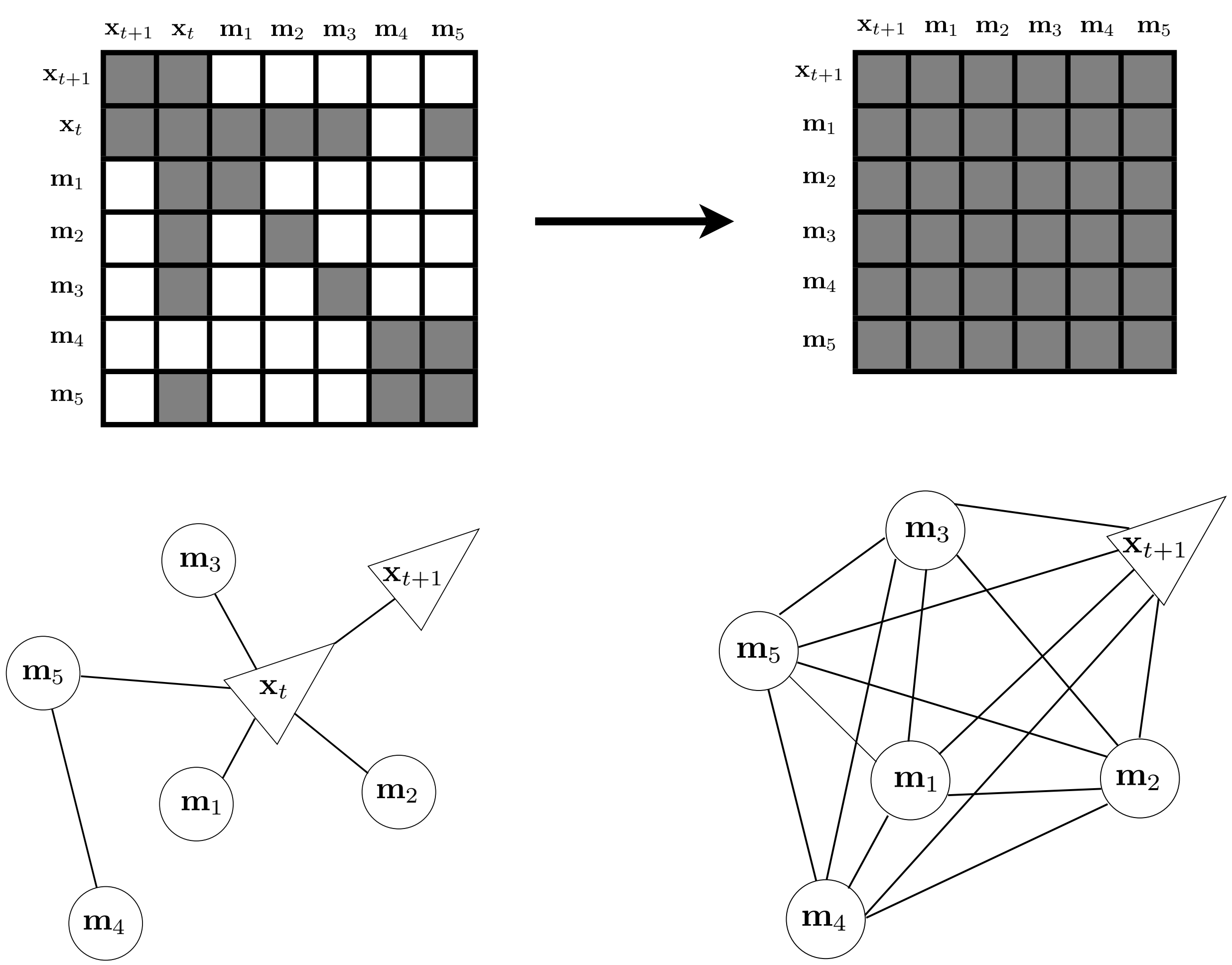
* Linearization requires knowledge of the mean
(Involves full matrix inversion: $\sim \mathcal{O}(n^3)$)

Consequences for EKF and EIF

	Covariance Form	Information Form
Marg.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha$ $\Sigma' = \Sigma_{\alpha\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda' = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$
Cond.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

II. Time projection step: state augmentation + marginalization

$$p\left(\mathbf{x}_{t+1}, \mathbf{M}\right) = \int p\left(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{M}\right) d\mathbf{x}_t$$



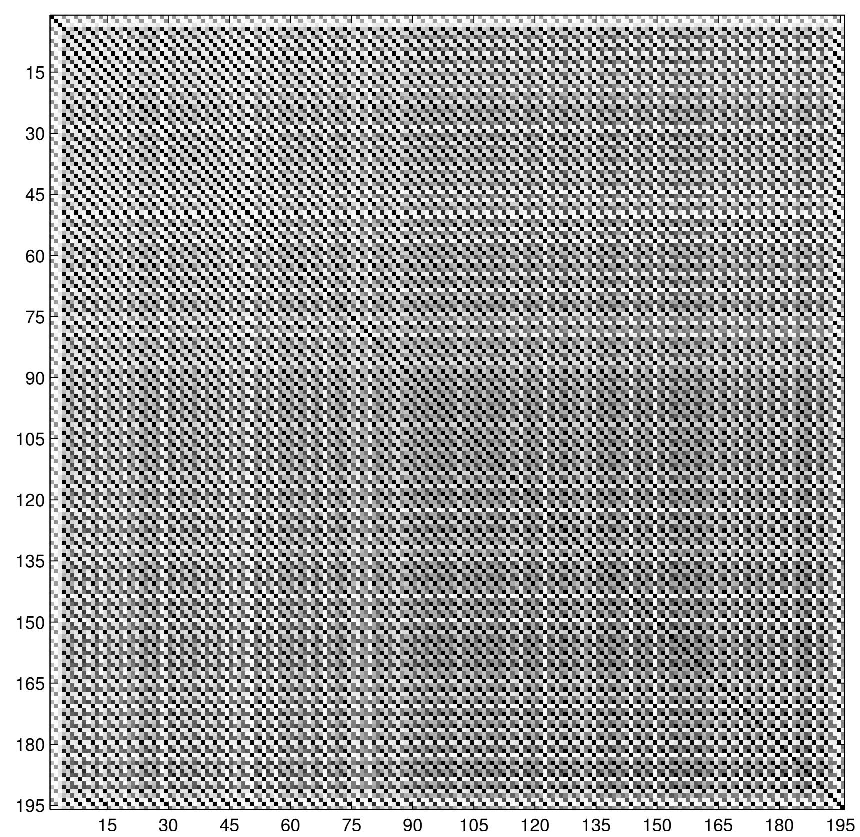
a. $\sim \mathcal{O}(n^3)$ cubic in the number of states

b. Populates the information matrix

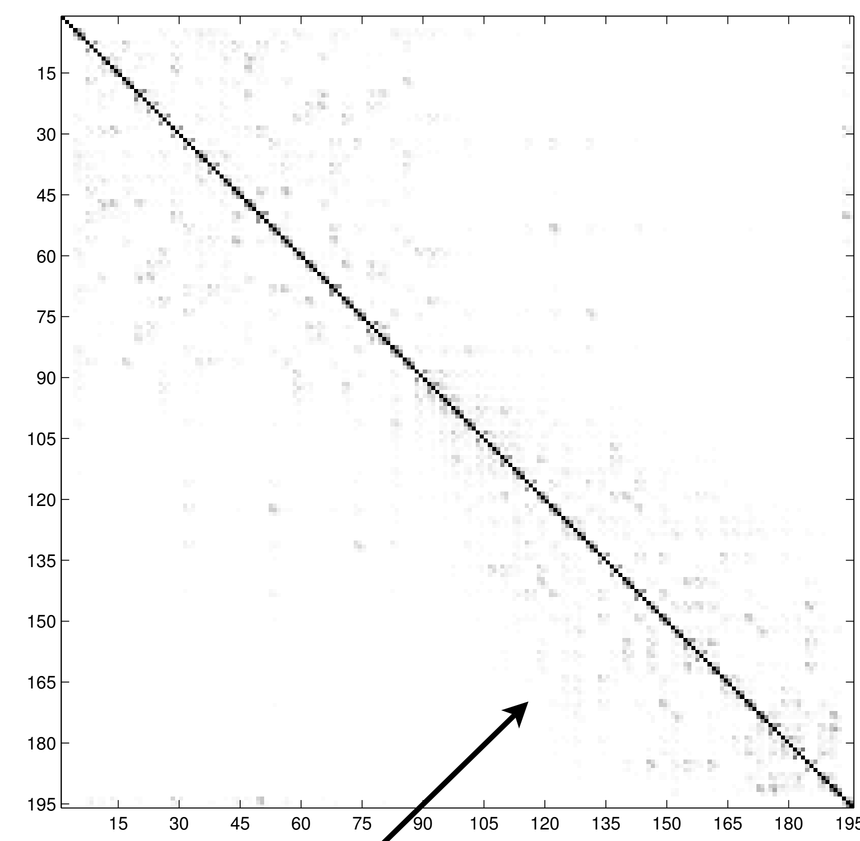
c. Weakens shared information.

SEIF's key insight:

<u>Information matrix is *relatively* sparse</u>



$$\Lambda_t = \Sigma_t^{-1}$$



small but <u>not zero</u>

Advantages when truly sparse

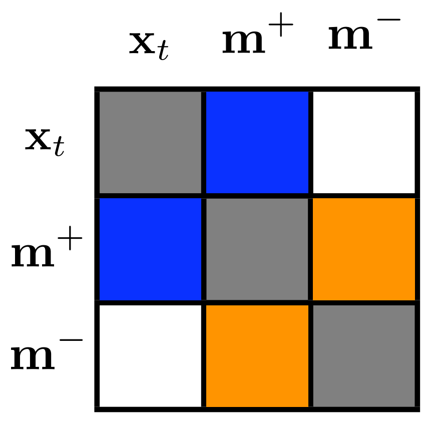
1. Linear time projection is constant-time.
2. Greatly reduced storage requirements.
3. Approximate mean calculation is efficient.
4. Given the mean, nonlinear time projection is constant-time.

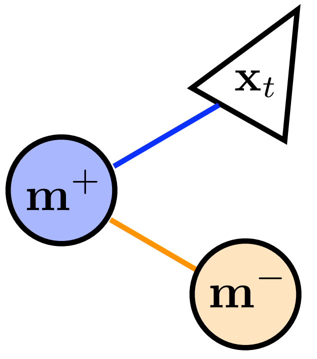
Advantages when truly sparse

1. Linear time projection is constant-time.
2. Greatly reduced storage requirements.
3. Approximate mean calculation is efficient.
4. Given the mean, nonlinear time projection is constant-time.

But the information matrix is fully populated!

Controlling density: limit number of active features

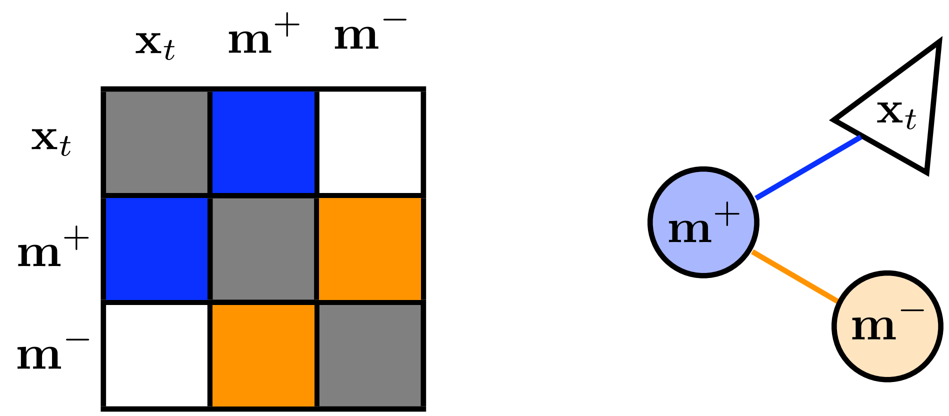




$\mathbf{m}^-$ active features

$\mathbf{m}^+$ passive features

Controlling density: limit number of active features

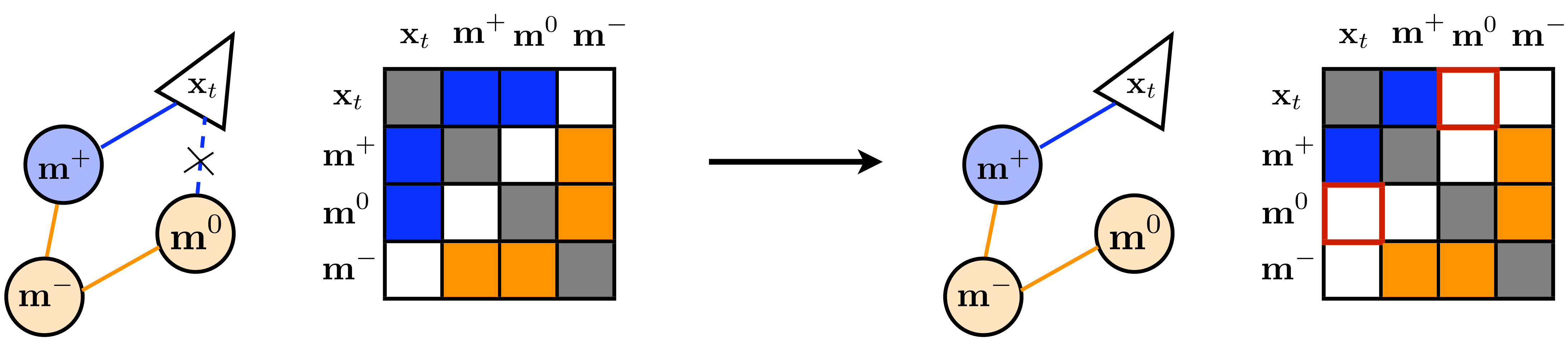


$\mathbf{m}^+$ active features

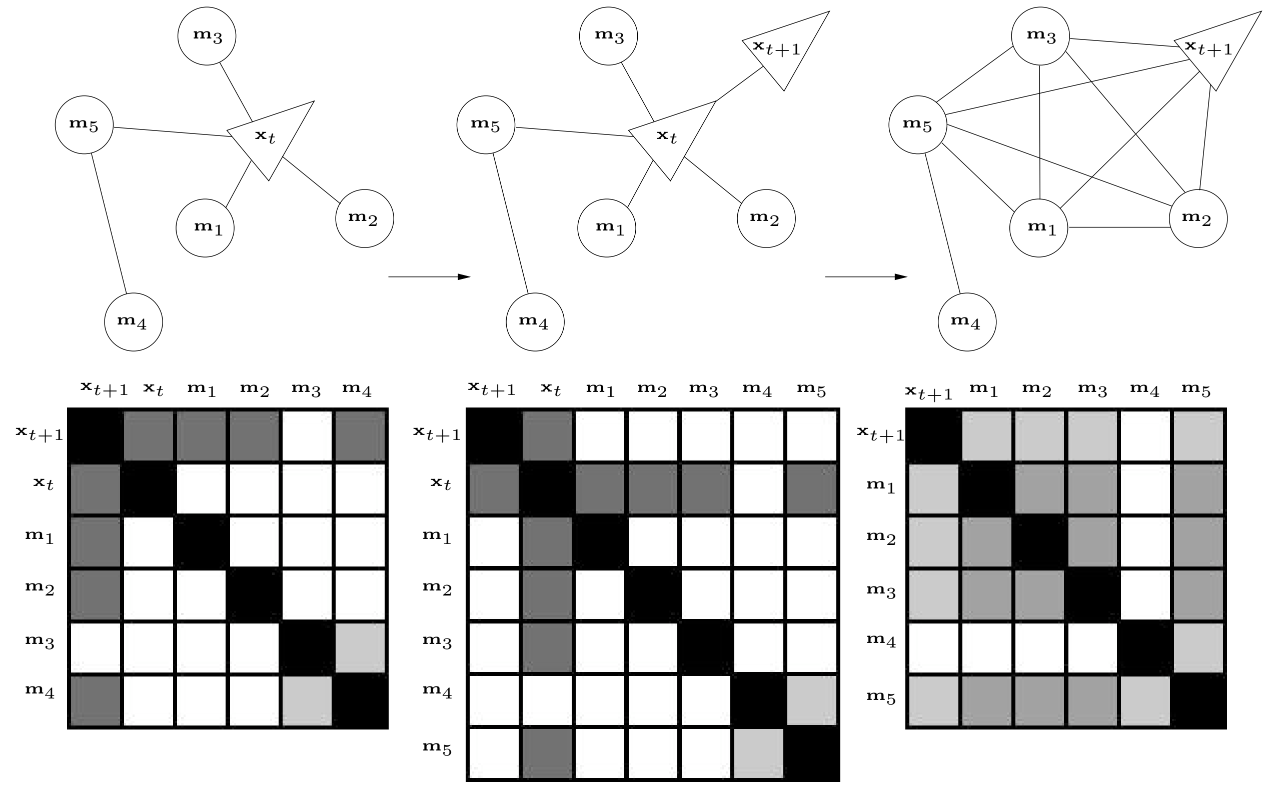
$\mathbf{m}^-$ passive features

Pacify active landmarks by breaking weak links

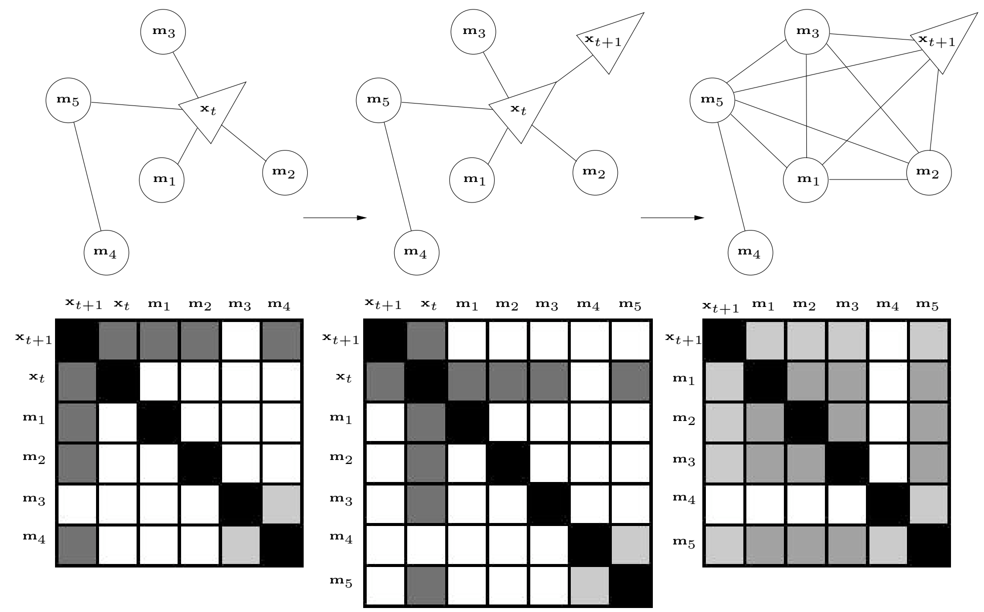
$\mathbf{m}^0$ active features to be made passive

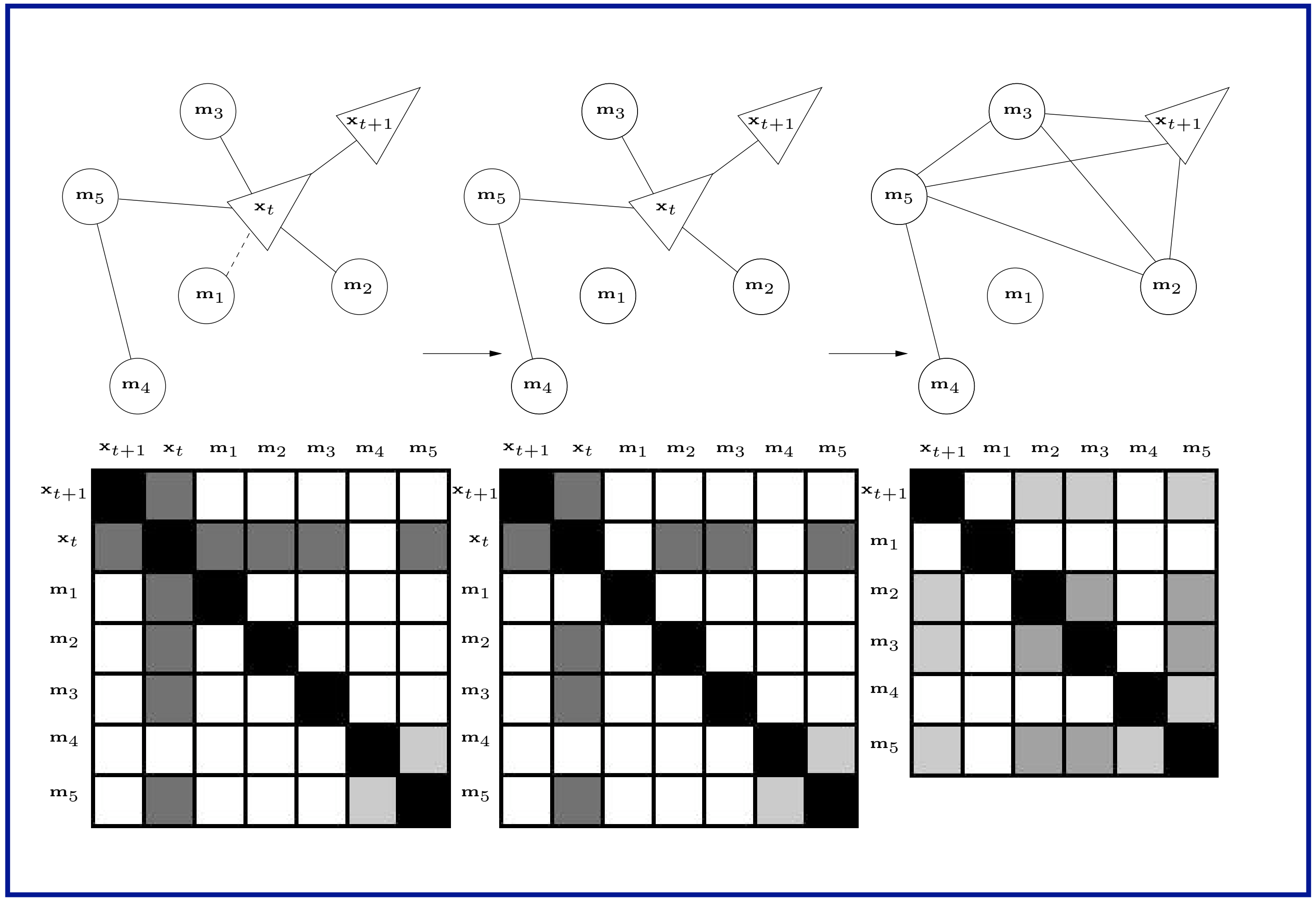


Controlling density: limit number of active features



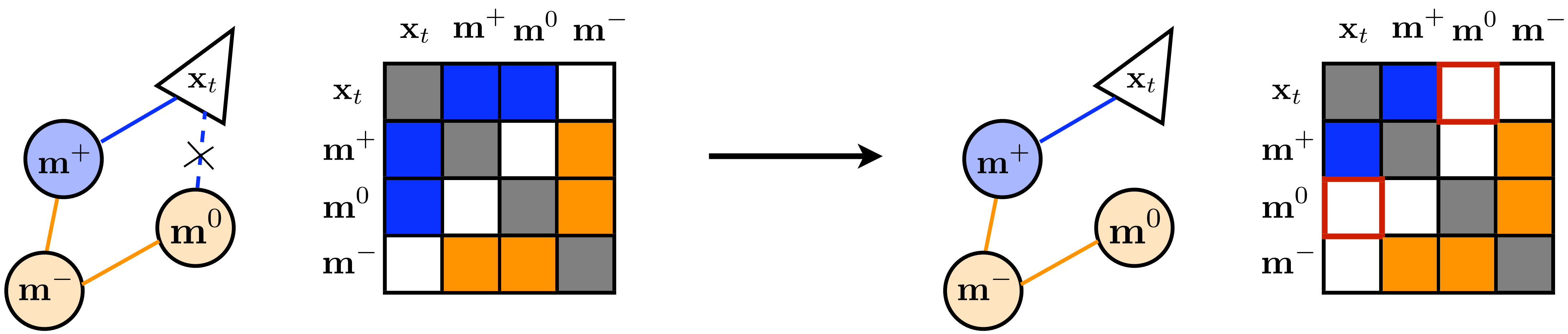
Controlling density: limit number of active features





Deactivation = Imposing conditional independence

$\mathbf{m}^+$	active features
$\mathbf{m}^0$	active features to be made passive
$\mathbf{m}^-$	passive features



$$p(\boldsymbol{\xi}_t) = p(\mathbf{x}_t|\mathbf{m}^+, \mathbf{m}^0)\, p(\mathbf{m}^+, \mathbf{m}^0, \mathbf{m}^-)$$

$$\tilde{p}(\boldsymbol{\xi}_t) = p(\mathbf{x}_t|\mathbf{m}^+)\, p(\mathbf{m}^+, \mathbf{m}^0, \mathbf{m}^-)$$

How do we force $p(\mathbf{x}_t|\mathbf{m}^+, \mathbf{m}^0) \longrightarrow p(\mathbf{x}_t|\mathbf{m}^+)$?

How we sparsify is nontrivial!

Roadmap

I Introduction

a. Normal form vs. Information form in SLAM

b. Motivation for Sparse Extended Information Filters

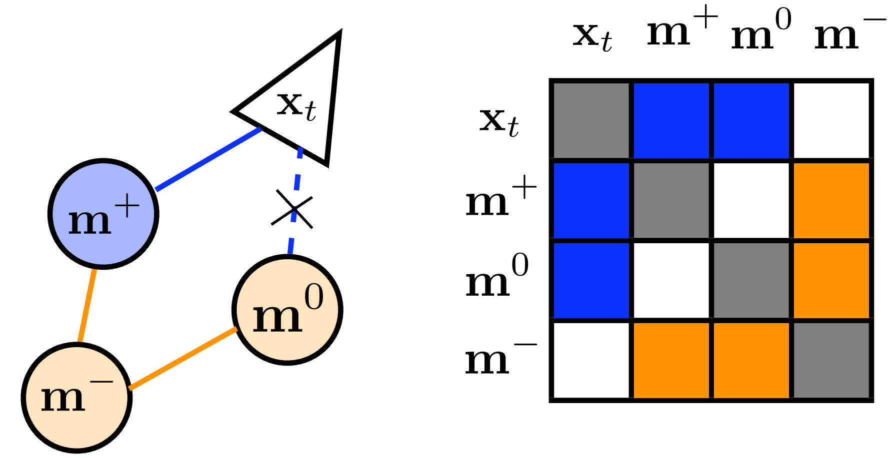
II Sparsification in SEIFS

III Modified sparsification rule

IV Results

V Conclusions

Sparsification in SEIFs

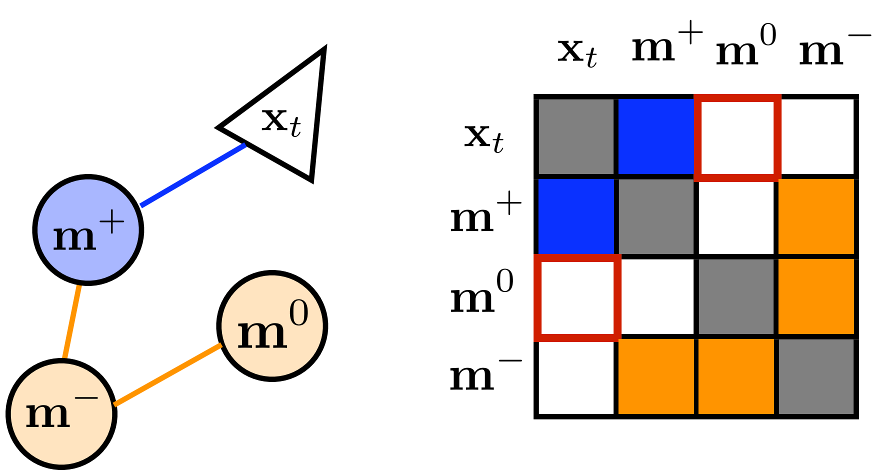


Bayes rule: $p(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) = p\left(\mathbf{x}_t \mid \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right) p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right)$

$$= p\left(\mathbf{x}_t \mid \mathbf{m}^0, \mathbf{m}^+, \underline{\mathbf{m}^- = \boldsymbol{\alpha}}\right) p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right)$$

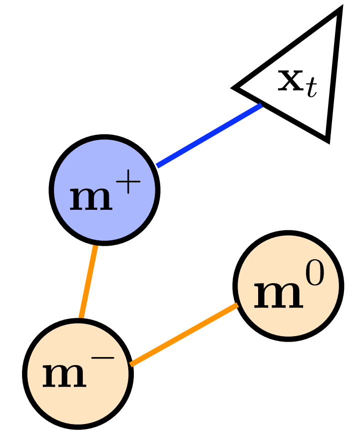
C.I.

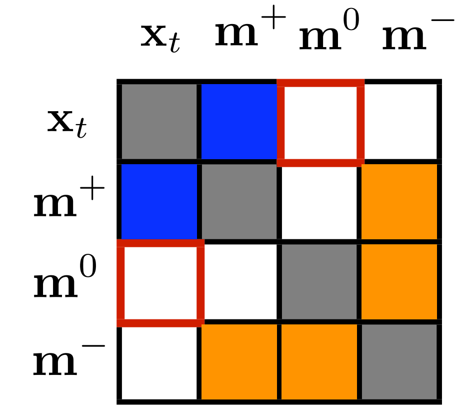
SEIF's rule deactivates link by forcing conditional independence to feature we want to deactivate



$$\tilde{p}_{\mathrm{SEIFs}}(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) = p\left(\mathbf{x}_t \mid \mathbf{m}^+, \mathbf{m}^- = \boldsymbol{\alpha}\right) p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right)$$

Sparsification in SEIFs





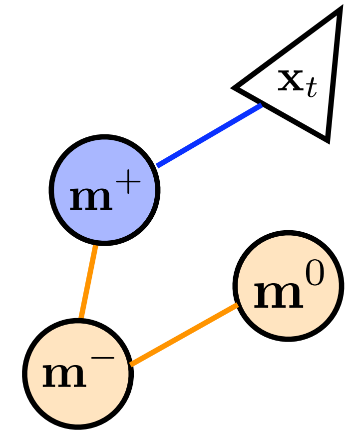
$$\tilde{p}_{\text{SEIFs}}(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) = \mathcal{N}^{-1}(\tilde{\boldsymbol{\eta}}_t, \tilde{\Lambda}_t)$$

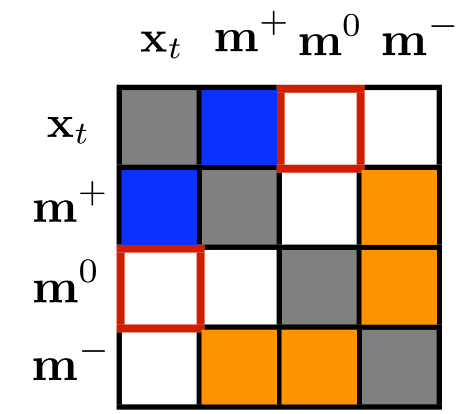
$$\begin{aligned}
\tilde{\Lambda}_t &= S_{x_t,m^+} \Lambda_B S_{x_t,m^+}^\top \\
&\qquad - S_{m^+} \Lambda_C S_{m^+}^\top + S_{m^0,m^+,m^-} \Lambda_D S_{m^0,m^+,m^-}^\top \\
\tilde{\boldsymbol{\eta}}_t &= S_{x_t,m^+} \boldsymbol{\eta}_B - S_{m^+} \boldsymbol{\eta}_C + S_{m^0,m^+,m^-} \boldsymbol{\eta}_D
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\eta}_\alpha &= \Sigma_t S_{m^-} \boldsymbol{\alpha} \\
\Lambda_B &= S_{x_t,m^+}^\top \left(\mathrm{I} - \Lambda_t S_{m^0} \underline{\left(S_{m^0}^\top \Lambda_t S_{m^0} \right)}^{-1} S_{m^0}^\top \right) \Lambda_t S_{x_t,m^+} \\
\boldsymbol{\eta}_B &= S_{x_t,m^+}^\top \left(\mathrm{I} - \Lambda_t S_{m^0} \left(S_{m^0}^\top \Lambda_t S_{m^0} \right)^{-1} S_{m^0}^\top \right) (\boldsymbol{\eta}_t - \boldsymbol{\eta}_\alpha) \\
\Lambda_C &= S_{m^+}^\top \left(\mathrm{I} - \Lambda_t S_{x_t,m^0} \underline{\left(S_{x_t,m^0}^\top \Lambda_t S_{x_t,m^0} \right)}^{-1} S_{x_t,m^0}^\top \right) \Lambda_t S_{m^+} \\
\boldsymbol{\eta}_C &= S_{m^+}^\top \left(\mathrm{I} - \Lambda_t S_{x_t,m^0} \left(S_{x_t,m^0}^\top \Lambda_t S_{x_t,m^0} \right)^{-1} S_{x_t,m^0}^\top \right) (\boldsymbol{\eta}_t - \boldsymbol{\eta}_\alpha) \\
\Lambda_D &= S_{m^0,m^+,m^-}^\top \left(\mathrm{I} - \Lambda_t S_{x_t} \left(S_{x_t}^\top \Lambda_t S_{x_t} \right)^{-1} S_{x_t}^\top \right) \Lambda_t S_{m^0,m^+,m^-} \\
\boldsymbol{\eta}_D &= S_{m^0,m^+,m^-}^\top \left(\mathrm{I} - \Lambda_t S_{x_t} \left(S_{x_t}^\top \Lambda_t S_{x_t} \right)^{-1} S_{x_t}^\top \right) \boldsymbol{\eta}_t
\end{aligned}$$

only requires matrix inversion on the order of the number of links we are breaking

Sparsification in SEIFs





$$\tilde{p}_{\text{SEIFs}}(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) = \mathcal{N}(\tilde{\boldsymbol{\mu}}_t, \tilde{\Sigma}_t)$$

$$\begin{aligned}
\tilde{\Sigma}_t &= \big(S_{x_t,m^+}\Sigma_B^{-1}S_{x_t,m^+}^\top - S_{m^+}\Sigma_C^{-1}S_{m^+}^\top \\
&\qquad + S_{m^0,m^+,m^-}\Sigma_D^{-1}S_{m^0,m^+,m^-}^\top\big)^{-1} \\
\tilde{\boldsymbol{\mu}}_t &= \boldsymbol{\mu}_t + \tilde{\Sigma}_t\big(S_{x_t,m^+}\Sigma_B^{-1}S_{x_t,m^+}^\top - S_{m^+}\Sigma_C^{-1}S_{m^+}^\top\big)\times \\
&\qquad \Sigma_t S_{m^-}\big(S_{m^-}^\top\Sigma_t S_{m^-}\big)^{-1}\big(\boldsymbol{\alpha} - S_{m^-}^\top\boldsymbol{\mu}_t\big)
\end{aligned}$$

Note: In general the mean will change!

$$\begin{aligned}
\Sigma_B &= S_{x,m^+}^\top\Big(\mathrm{I} - \Sigma_t S_{m^-}\big(S_{m^-}^\top\Sigma_t S_{m^-}\big)^{-1}S_{m^-}^\top\Big)\Sigma_t S_{x,m^+} \\
\Sigma_C &= S_{m^+}^\top\Big(\mathrm{I} - \Sigma_t S_{m^-}\big(S_{m^-}^\top\Sigma_t S_{m^-}\big)^{-1}S_{m^-}^\top\Big)\Sigma_t S_{m^+} \\
\Sigma_D &= S_{m^0,m^+,m^-}^\top\Sigma_t S_{m^0,m^+,m^-}
\end{aligned}$$

Roadmap

I Introduction

a. Normal form vs. Information form in SLAM
b. Motivation for Sparse Extended Information Filters

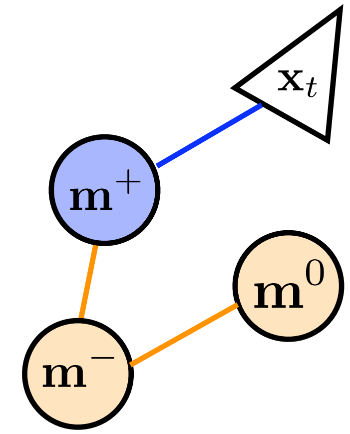
II Sparsification in SEIFS

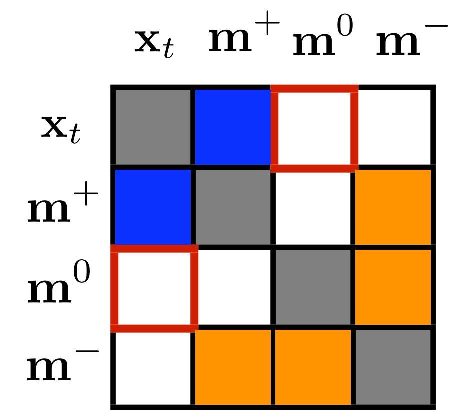
III Modified sparsification rule

IV Results

V Conclusions

Modified sparsification rule





$$\begin{aligned}\breve{p}(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) &= \frac{p\left(\mathbf{x}_t \mid \mathbf{m}^+\right) p\left(\mathbf{m}^0 \mid \mathbf{m}^+\right)}{p\left(\mathbf{m}^0 \mid \mathbf{m}^+\right)} p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right) \\ &= p\left(\mathbf{x}_t \mid \mathbf{m}^+\right) p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right) \\ &= \frac{p\left(\mathbf{x}_t, \mathbf{m}^+\right)}{p\left(\mathbf{m}^+\right)} p\left(\mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-\right)\end{aligned}$$

$$\breve{p}(\mathbf{x}_t, \mathbf{m}^0, \mathbf{m}^+, \mathbf{m}^-) = \mathcal{N}^{-1}\left(\breve{\boldsymbol{\eta}}_t, \breve{\Lambda}_t\right)$$

$$\begin{aligned}\breve{\Lambda}_t &= S_{x_t,m^+}\Lambda_U S_{x_t,m^+}^\top - S_{m^+}\Lambda_V S_{m^+}^\top \\ &\qquad + S_{m^0,m^+,m^-}\Lambda_D S_{m^0,m^+,m^-}^\top \\ \breve{\boldsymbol{\eta}}_t &= S_{x_t,m^+}\boldsymbol{\eta}_U - S_{m^+}\boldsymbol{\eta}_V + S_{m^0,m^+,m^-}\boldsymbol{\eta}_D\end{aligned}$$

$$\begin{aligned}\Lambda_U &= S_{x_t,m^+}^\top\Big(\mathrm{I} - \Lambda_t S_{m^0,m^-} \times \\ &\qquad \underline{\left(S_{m^0,m^-}^\top \Lambda_t S_{m^0,m^-}\right)^{-1}} S_{m^0,m^-}^\top\Big)\Lambda_t S_{x_t,m^+} \\ \boldsymbol{\eta}_U &= S_{x_t,m^+}^\top\Big(\mathrm{I} - \Lambda_t S_{m^0,m^-}\left(S_{m^0,m^-}^\top \Lambda_t S_{m^0,m^-}\right)^{-1} S_{m^0,m^-}^\top\Big)\boldsymbol{\eta}_t \\ \Lambda_V &= S_{m^+}^\top\Big(\mathrm{I} - \Lambda_t S_{x_t,m^0,m^-} \times \\ &\qquad \underline{\left(S_{x_t,m^0,m^-}^\top \Lambda_t S_{x_t,m^0,m^-}\right)^{-1}} S_{x_t,m^0,m^-}^\top\Big)\Lambda_t S_{m^+} \\ \boldsymbol{\eta}_V &= S_{m^+}^\top\Big(\mathrm{I} - \Lambda_t S_{x_t,m^0,m^-} \times \\ &\qquad \left(S_{x_t,m^0,m^-}^\top \Lambda_t S_{x_t,m^0,m^-}\right)^{-1} S_{x_t,m^0,m^-}^\top\Big)\boldsymbol{\eta}_t \\ \Lambda_D &= S_{m^0,m^+,m^-}^\top\Big(\mathrm{I} - \Lambda_t S_{x_t}\left(S_{x_t}^\top \Lambda_t S_{x_t}\right)^{-1} S_{x_t}^\top\Big)\Lambda_t S_{m^0,m^+,m^-} \\ \boldsymbol{\eta}_D &= S_{m^0,m^+,m^-}^\top\Big(\mathrm{I} - \Lambda_t S_{x_t}\left(S_{x_t}^\top \Lambda_t S_{x_t}\right)^{-1} S_{x_t}^\top\Big)\boldsymbol{\eta}_t\end{aligned}$$

requires matrix inversion on the order of the number of links being removed

No longer constant time

Roadmap

I Introduction

a. Normal form vs. Information form in SLAM
b. Motivation for Sparse Extended Information Filters

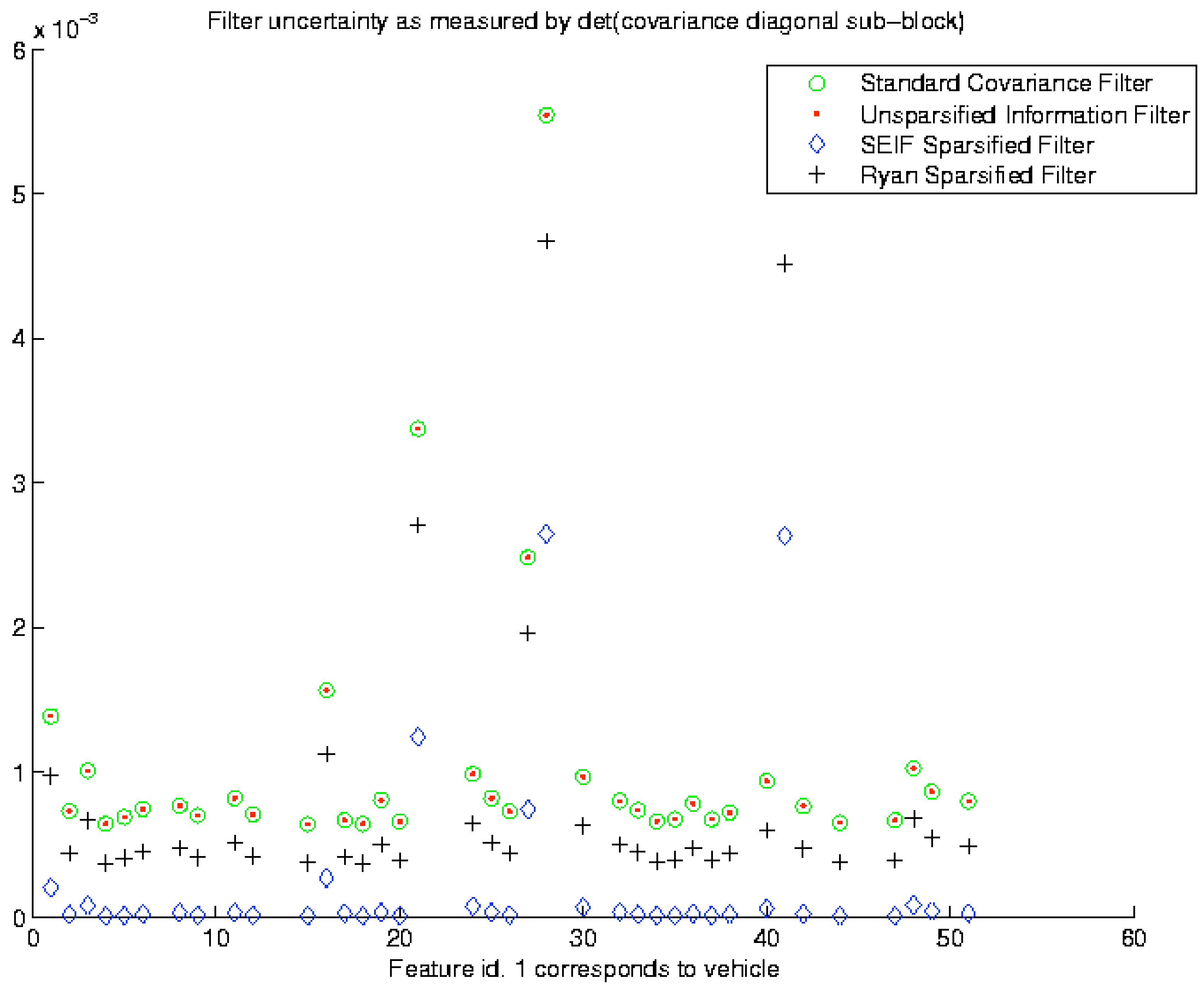
II Sparsification in SEIFS

III Modified sparsification rule

IV Results

V Conclusions

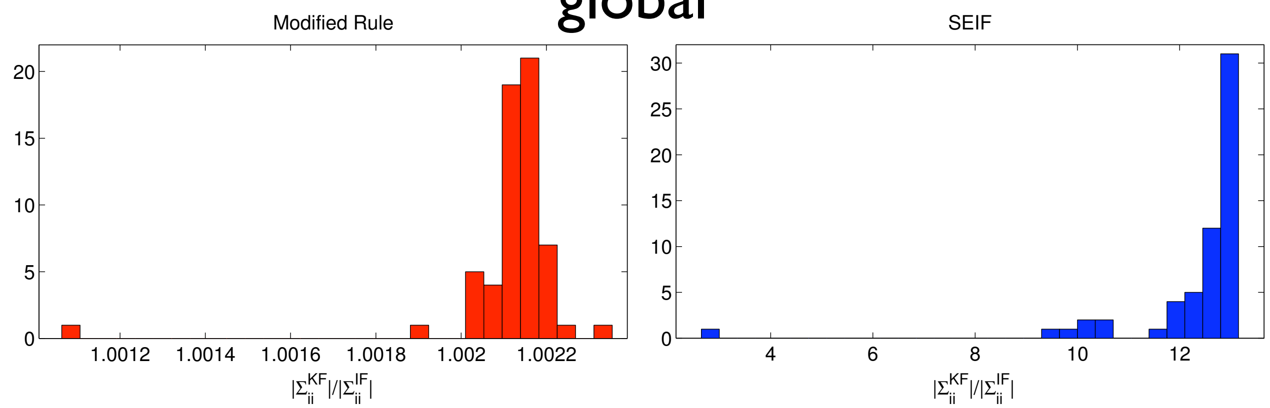
LG simulation results

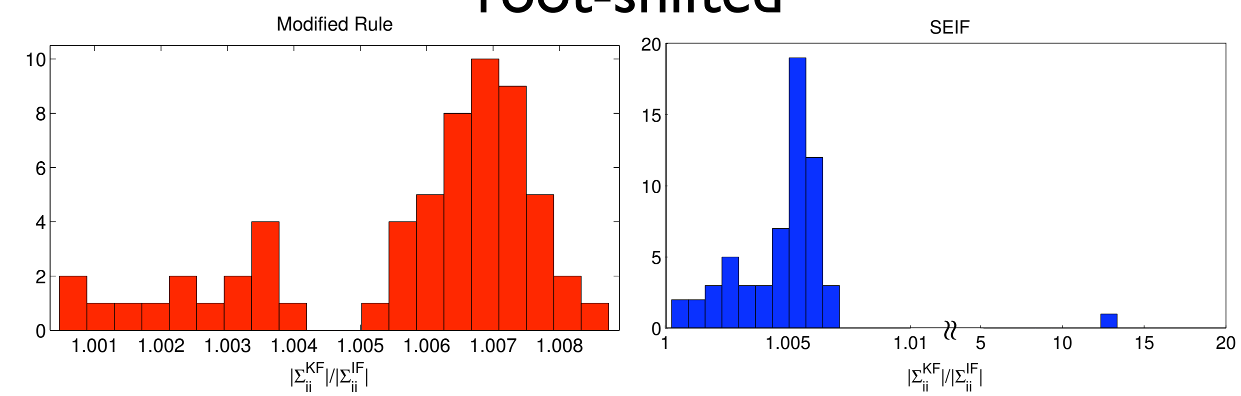


LG simulation results

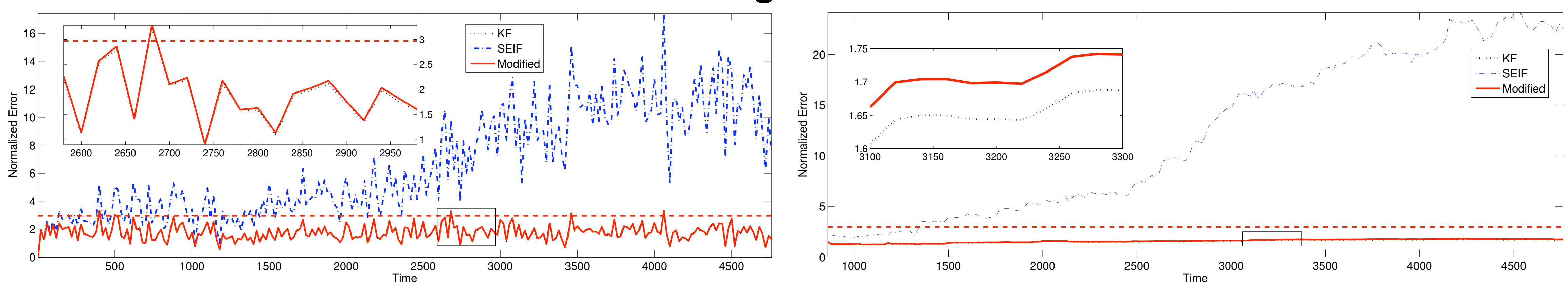
global

root-shifted

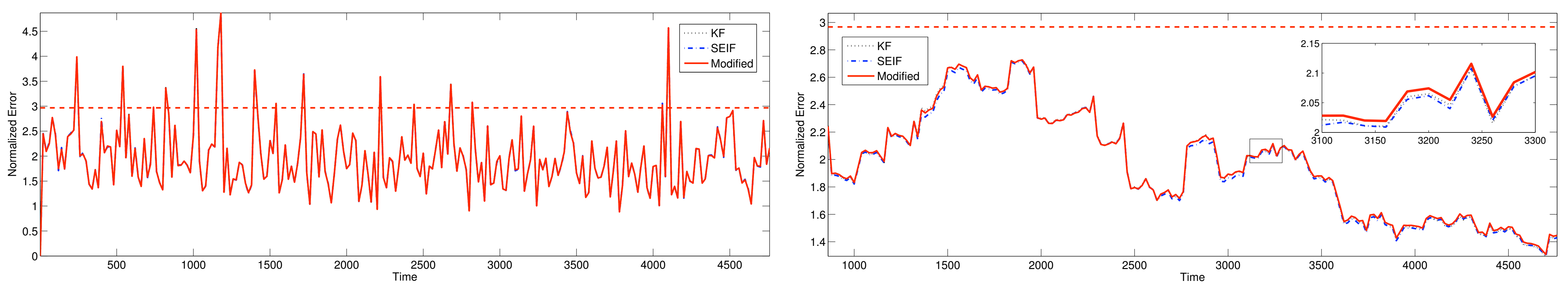




global

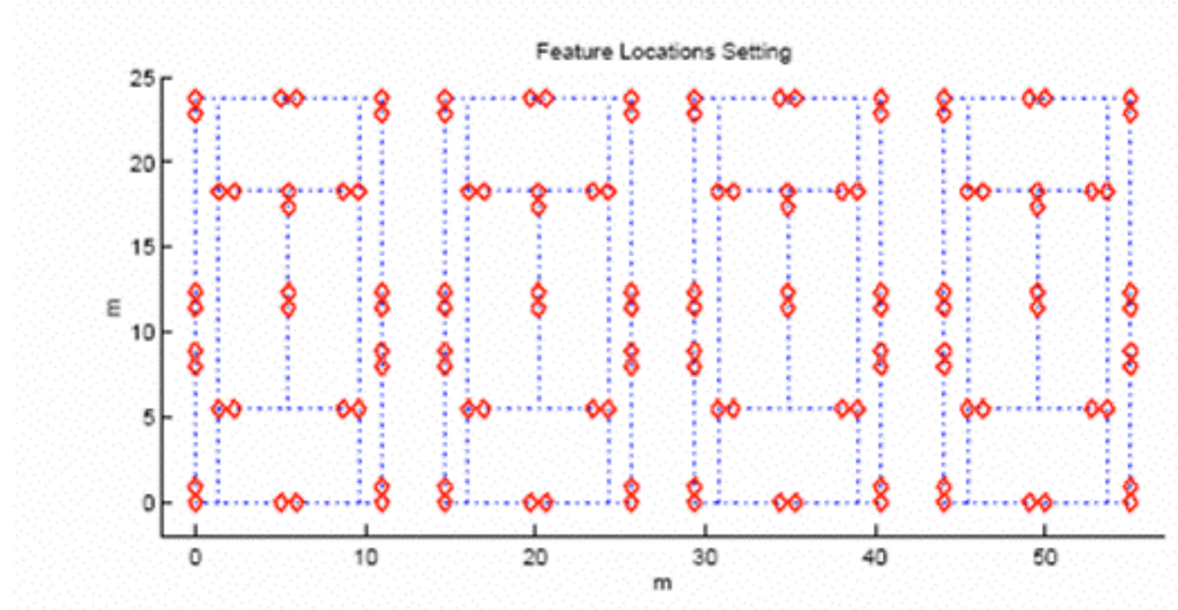


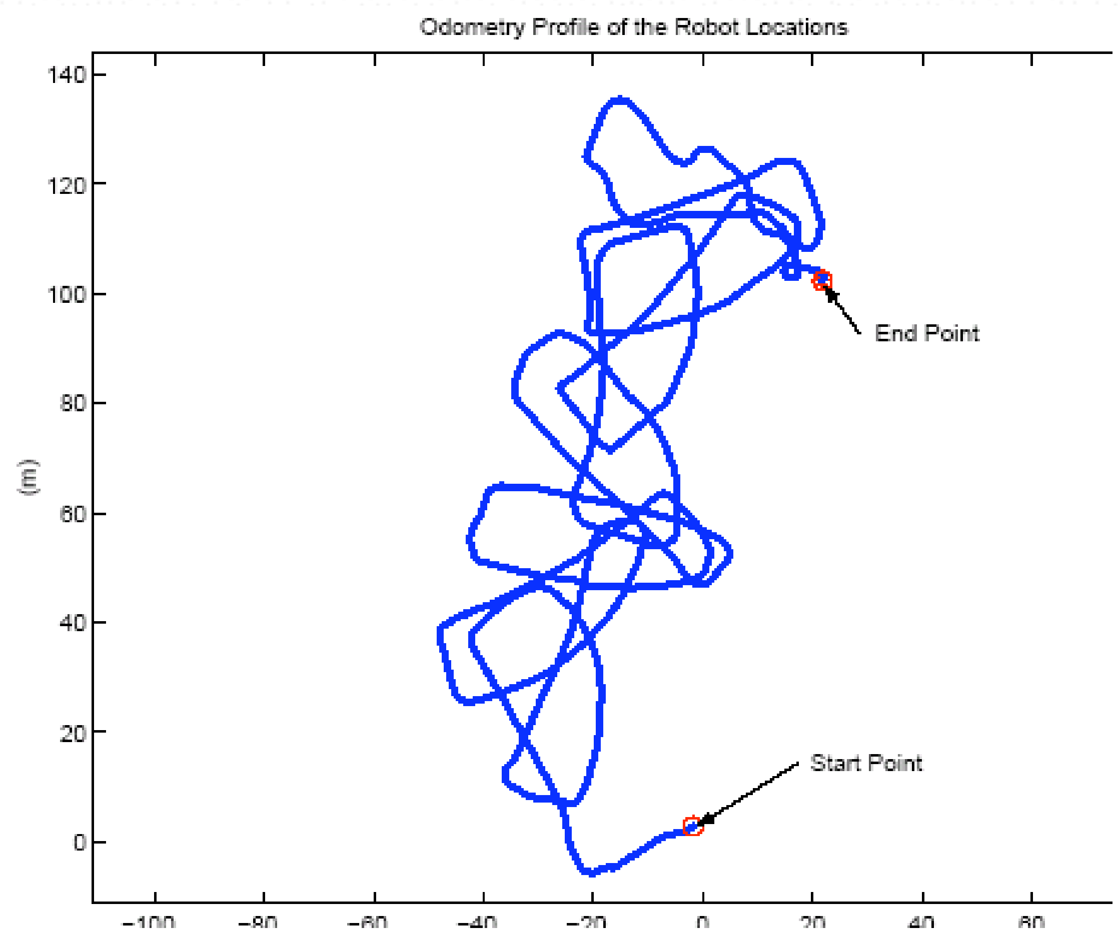
root-shifted



Nonlinear dataset

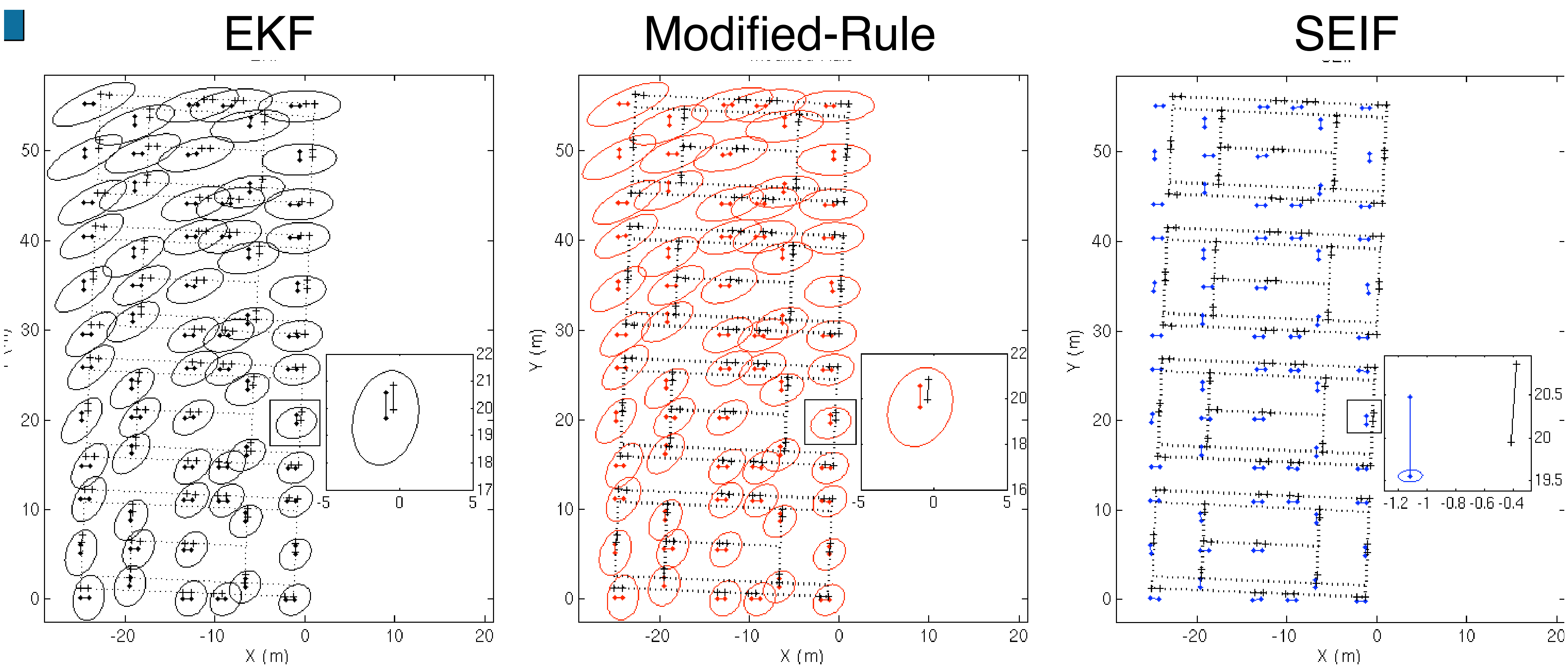
MIT Indoor Track





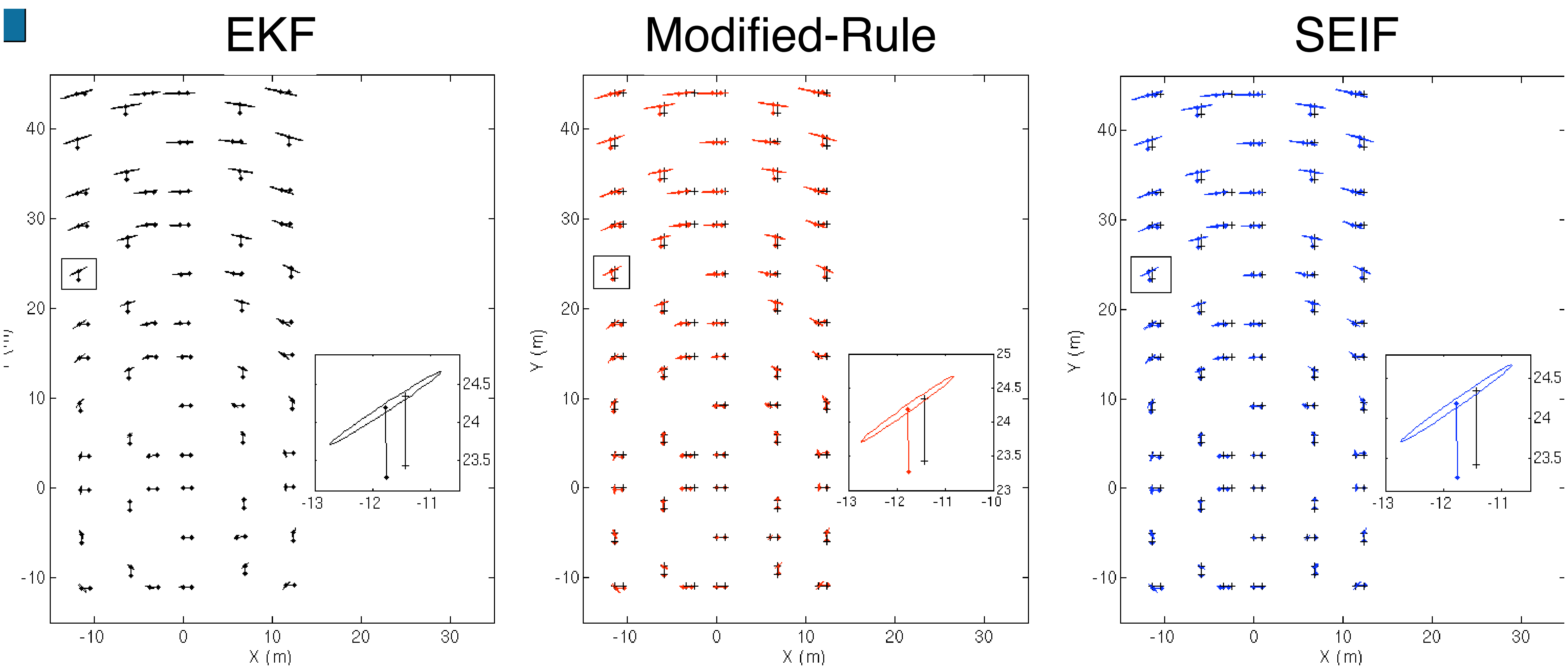
Nonlinear dataset

Global map



Nonlinear dataset

Root-shifted map



Roadmap

Conclusions

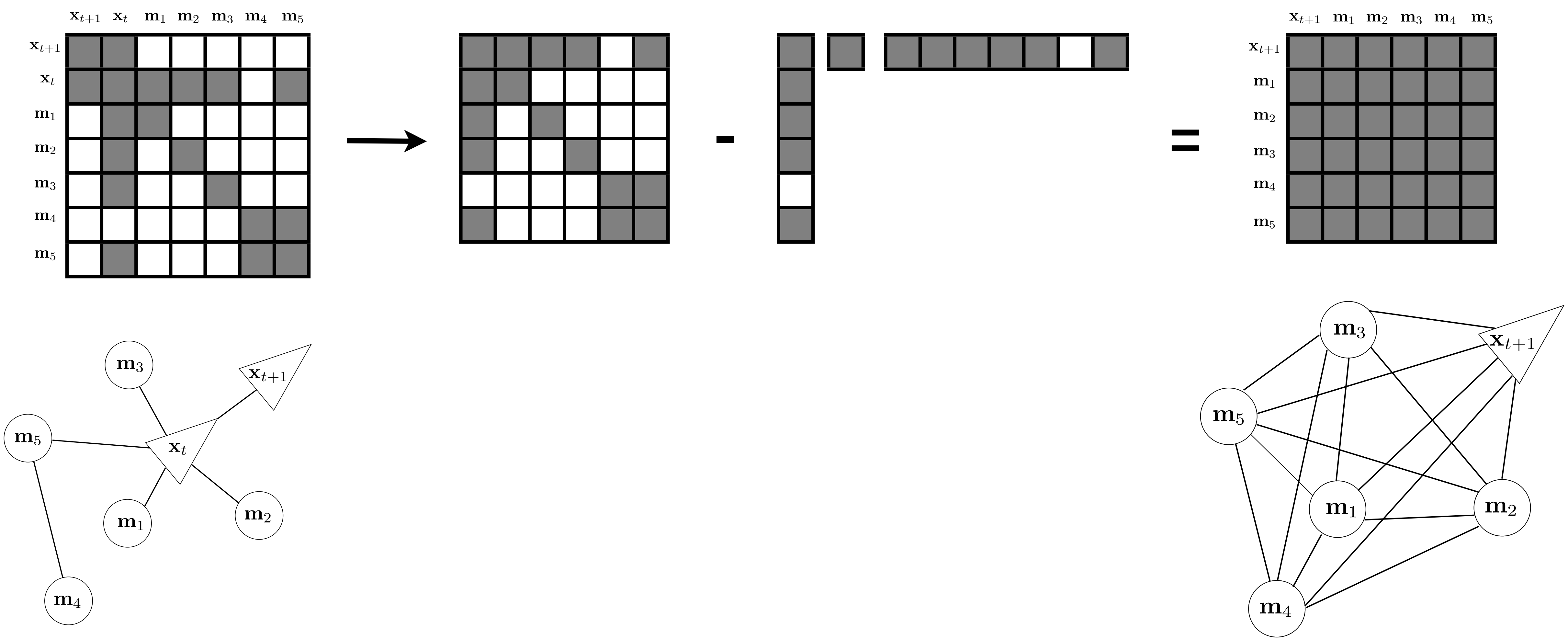
- Insights into Sparsification of Information Form SLAM algorithms
- Modified rule corrects problem in original SEIF derivation, but is no longer constant-time
- Experimental results show that SEIFs are globally inconsistent, however consistency is maintained when the map is expresed as a *relative* map in a local frame defined by one of the features

Consequences for EKF and EIF

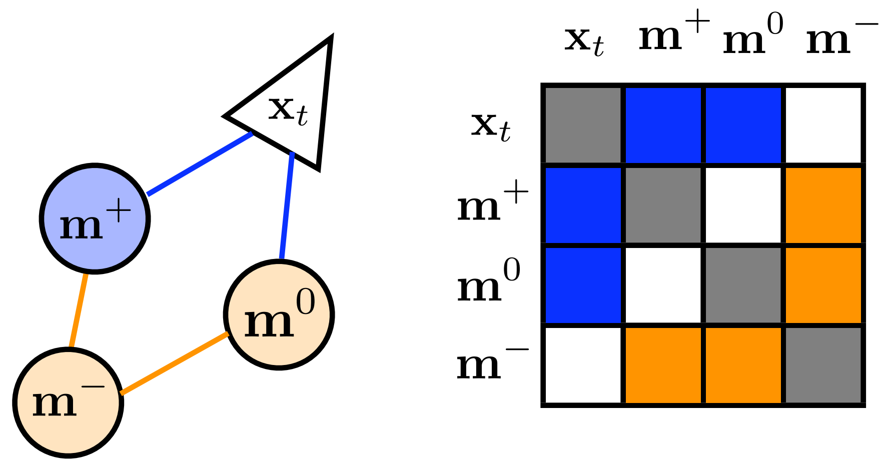
	Covariance Form	Information Form
Marg.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha$ $\Sigma' = \Sigma_{\alpha\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\boldsymbol{\eta}_\beta$ $\Lambda' = \Lambda_{\alpha\alpha} - \Lambda_{\alpha\beta}\Lambda_{\beta\beta}^{-1}\Lambda_{\beta\alpha}$
Cond.	$\boldsymbol{\mu}' = \boldsymbol{\mu}_\alpha + \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta)$ $\Sigma' = \Sigma_{\alpha\alpha} - \Sigma_{\alpha\beta}\Sigma_{\beta\beta}^{-1}\Sigma_{\beta\alpha}$	$\boldsymbol{\eta}' = \boldsymbol{\eta}_\alpha - \Lambda_{\alpha\beta}\boldsymbol{\beta}$ $\Lambda' = \Lambda_{\alpha\alpha}$

II. Time projection step: state augmentation + marginalization

$$p\left(\mathbf{x}_{t+1}, \mathbf{M}\right) = \int p\left(\mathbf{x}_t, \mathbf{x}_{t+1}, \mathbf{M}\right) d\mathbf{x}_t \qquad \boldsymbol{\alpha} = \{\mathbf{x}_{t+1}, \mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4, \mathbf{m}_5\} \qquad \boldsymbol{\beta} = \mathbf{x}_t$$



Robot's conditional independence from passive features

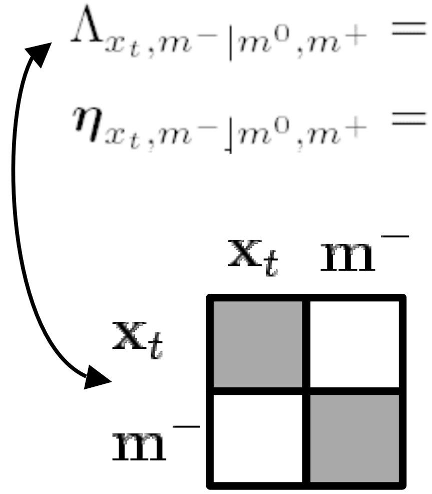


$$p\left(\mathbf{x}_t, \mathbf{m}^- \mid \mathbf{m}^0, \mathbf{m}^+\right) = \mathcal{N}^{-1}\left(S^\top_{x_t,m^-}\boldsymbol{\xi}_t; \boldsymbol{\eta}_{x_t,m^-|m^0,m^+}, \Lambda_{x_t,m^-|m^0,m^+}\right)$$

$$\Lambda_{x_t,m^-|m^0,m^+} = S^\top_{x_t,m^-}\Lambda_t S_{x_t,m^-}$$

extract $\mathbf{x}_t$ $\mathbf{m}^-$ sub-block

$$\boldsymbol{\eta}_{x_t,m^-|m^0,m^+} = S^\top_{x_t,m^-}\boldsymbol{\eta}_t - \left(S^\top_{x_t,m^-}\Lambda_t S_{m^0,m^+}\right)S^\top_{m^0,m^+}\boldsymbol{\xi}_t$$



The conditional information matrix is block-diagonal and the inverse of a block-diagonal matrix is also block-diagonal. This implies that the robot and passive features are uncorrelated which for a Gaussian RV implies independence. QED