Data Assimilation
Matrix-Free EnKF Implementations

Elias D. Nino-Ruiz, Ph.D.

Universidad del Norte

enino@uninorte.edu.co

August 26, 2020

The Ensemble Kalman Filter (EnKF)

- The Ensemble Kalman Filter (EnKF) is a sequential Monte Carlo method for parameter and state estimation in highly non-linear models.
- Suppose we want to estimate the state of a system $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$ which approximately evolves according to some imperfect numerical operator,

$$\mathbf{x}_k^* = \mathcal{M}_{t_{k-1} \rightarrow t_k}\left(\mathbf{x}_{k-1}^*\right) \,,$$

- As we discussed previosuly, n is the number of model components (i.e., the number of grid points), we seek to approximate the actual values onto our grid even though they can belong to a continuous field.

The Ensemble Kalman Filter (EnKF)

- An ensemble of model realizations,

$$\mathbf{X}^b = \left[\mathbf{x}^{b[1]},\, \mathbf{x}^{b[2]},\, \ldots,\, \mathbf{x}^{b[N]}\right] \in \mathbb{R}^{n \times N}\,,$$

where $\mathbf{x}^{b[i]} \in \mathbb{R}^{n\times 1}$ is the i-th ensemble member, for $1 \le i \le N$, is used in order to estimate the moments of the background error distribution,

$$\mathbf{x} \sim \mathcal{N}\left(\mathbf{x}^b,\, \mathbf{B}\right)\,,$$

where

$$\mathbf{x}^b = \mathbf{x}^* + \boldsymbol{\theta}^b \in \mathbb{R}^{n\times 1},$$

with

$$\boldsymbol{\theta}^b \sim \mathcal{N}\left(\mathbf{0}_n,\, \mathbf{B}\right)\,.$$

The Ensemble Kalman Filter (EnKF)

- The approximation is performed making use of the empirical moments of the ensemble,

$$\begin{aligned} \mathbf{x}^b &\approx \overline{\mathbf{x}}^b = \frac{1}{N} \cdot \sum_{i=1}^{N} \mathbf{x}^{b[i]} \in \mathbb{R}^{n \times 1}, \\ \mathbf{B} &\approx \mathbf{P}^b = \frac{1}{N-1} \cdot \boldsymbol{\Delta}\mathbf{X} \cdot \boldsymbol{\Delta}\mathbf{X}^T \in \mathbb{R}^{n \times n}, \end{aligned}$$

with the matrix of member deviations being,

$$\boldsymbol{\Delta}\mathbf{X} = \mathbf{X}^b - \overline{\mathbf{x}}^b \cdot \mathbf{1}_N^T \in \mathbb{R}^{n \times N},$$

- When an observation

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x}^* + \boldsymbol{\theta}^o \in \mathbb{R}^{m \times 1},$$

the analysis ensemble can be computed as follows,

$$\mathbf{X}^a = \left[\mathbf{P}^{b^{-1}} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \left[\mathbf{P}^{b^{-1}} \cdot \mathbf{X}^b + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Y}^s\right]$$

where $\mathbf{R} \in \mathbb{R}^{m \times m}$ is the estimated data error covariance matrix, $\mathbf{H} \in \mathbb{R}^{m \times n}$ is a linear observation operator,

$$\boldsymbol{\theta}^o \sim \mathcal{N}\left(\mathbf{0}_m,\, \mathbf{R}\right)\,,$$

and the i-th column of $\mathbf{Y}^s \in \mathbb{R}^{m \times N}$ reads,

$$\mathbf{y}^{s[i]} \sim \mathcal{N}\left(\mathbf{y},\, \mathbf{R}\right)\,.$$

The Ensemble Kalman Filter (EnKF)

- Here we have several opportunities and limitations:
  - The data error covariance matrix is usually diagonal, block diagonal or it is easy to decompose.
  - The dimension of the state vector n is several times greater than the ensemble size N. In operational data assimilation, $\mathbf{P}^b$ is usually low-rank.
  - The low-rank deficiency of $\mathbf{P}^b$ can be exploited in order to develop efficient matrix-free EnKF implementations.
- It can be easily show that, the analysis members are (approximately) minimizers of the 3D-Var cost function,

$$\mathcal{J}\left(\mathbf{x};\,\mathbf{x}^b,\mathbf{y},\mathbf{B},\mathbf{R}\right) = \frac{1}{2}\cdot\left\|\mathbf{x}-\mathbf{x}^b\right\|^2_{\mathbf{B}^{-1}} + \frac{1}{2}\cdot\left\|\mathbf{y}-\mathbf{H}\cdot\mathbf{x}\right\|^2_{\mathbf{R}^{-1}}\,.$$

for instance,

$$\mathbf{x}^{a[i]} = \arg\min_{\mathbf{x}}\mathcal{J}\left(\mathbf{x};\mathbf{x}^{b[i]},\mathbf{y}^{s[i]},\mathbf{P}^b,\mathbf{R}\right)\,.$$

you can also, note that,

$$\mathcal{P}_i^a(\mathbf{x}) \propto \exp\left(-\mathcal{J}\left(\mathbf{x}; \mathbf{x}^{b[i]}, \mathbf{y}^{s[i]}, \mathbf{P}^b, \mathbf{R}\right)\right) \,.$$

- Now, recall your first exam...

$$\mathcal{P}^a\left(\boldsymbol{\beta}|\mathbf{X},\, \mathbf{y}\right) \propto \mathcal{P}^b(\boldsymbol{\beta}) \cdot \mathcal{L}\left(\boldsymbol{\beta}|\mathbf{X},\, \mathbf{y}\right)$$

where

$$\mathcal{P}^b(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2} \cdot \left\|\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\right\|_{\mathbf{Q}^{-1}}^2\right)$$

and

$$\mathcal{L}\left(\boldsymbol{\beta}|\mathbf{X},\, \mathbf{y}\right) \propto \exp\left(-\frac{1}{2} \cdot \left\|\mathbf{y} - \mathbf{X} \cdot \boldsymbol{\beta}\right\|_{\mathbf{Z}^{-1}}^2\right) \,,$$

and therefore,

$$\mathcal{P}\left(\boldsymbol{\beta}|\mathbf{X},\, \mathbf{y}\right) \propto \exp\left(-\frac{1}{2}\cdot\left\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}\right\|_{\mathbf{A}^{-1}}^{2}\right)$$

where

$$\begin{aligned}\mathbf{A} &= \left[\mathbf{Q}^{-1}+\mathbf{X}^{T}\cdot\mathbf{Z}^{-1}\cdot\mathbf{X}\right]^{-1}\in\mathbb{R}^{p\times p}\,,\\ \widehat{\boldsymbol{\beta}} &= \mathbf{A}\cdot\left[\mathbf{Q}^{-1}\cdot\tilde{\boldsymbol{\beta}}+\mathbf{X}^{T}\cdot\mathbf{Z}^{-1}\cdot\mathbf{y}\right]\in\mathbb{R}^{p\times 1}\,.\end{aligned}$$

- Does this look familiar to you?
- So, at the end of the day, a Data Assimilation problem is nothing but...

The Sherman Morrison Formula

- The Woodbury Matrix Identity reads,

$$
\begin{aligned}
[\mathbf{A}+\mathbf{U}\cdot\mathbf{C}\cdot\mathbf{V}]^{-1} &= \mathbf{A}^{-1}-\mathbf{A}^{-1}\cdot\mathbf{U} \\
&\cdot \left[\mathbf{C}^{-1}+\mathbf{V}\cdot\mathbf{A}^{-1}\cdot\mathbf{U}\right]^{-1}\cdot\mathbf{V}\cdot\mathbf{A}^{-1}
\end{aligned}
$$

- The Sherman Morrison formula reads,

$$
\left[\mathbf{A}+\mathbf{u}\cdot\mathbf{v}^{T}\right]^{-1} = \mathbf{A}^{-1}-\frac{\mathbf{A}^{-1}\cdot\mathbf{u}\cdot\mathbf{v}^{T}\cdot\mathbf{A}^{-1}}{1+\mathbf{v}^{T}\cdot\mathbf{A}^{-1}\cdot\mathbf{u}}\,.
$$

Equivalent EnKF Formulations

- Making use of the SMF and some matrix properties, you can show that, the analysis step of the EnKF can be written in the following ways,

$$
\begin{aligned}
\mathbf{X}^a &= \left[{\mathbf{P}^b}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \left[{\mathbf{P}^b}^{-1} \cdot \mathbf{X}^b + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Y}^s\right] \\
\mathbf{X}^a &= \mathbf{X}^b + \left[{\mathbf{P}^b}^{-1} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b\right] \\
\mathbf{X}^a &= \mathbf{X}^b + \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T + \mathbf{R}\right]^{-1} \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b\right]
\end{aligned}
$$

- In practice, the most common formulation reads,

$$
\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T + \mathbf{R}\right]^{-1} \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b\right] .
$$

Efficient Implementations of the EnKF

- Note that,

$$\begin{aligned}\mathbf{X}^a - \mathbf{X}^b &= \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T + \mathbf{R}\right]^{-1} \left[\mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b\right] \\ &= \widehat{\boldsymbol{\Delta}\mathbf{X}} \cdot \mathbf{V}^T \cdot \left[\mathbf{R} + \mathbf{V} \cdot \mathbf{V}^T\right]^{-1} \cdot \mathbf{D}\end{aligned}$$

where $\widehat{\boldsymbol{\Delta}\mathbf{X}} = \frac{1}{\sqrt{N-1}} \cdot \boldsymbol{\Delta}\mathbf{X} \in \mathbb{R}^{n \times N}$, $\mathbf{V} = \mathbf{H} \cdot \widehat{\boldsymbol{\Delta}\mathbf{X}}$, and $\mathbf{D} = \mathbf{Y}^s - \mathbf{H} \cdot \mathbf{X}^b \in \mathbb{R}^{m \times N}$, and therefore,

$$\mathbf{x}^{a[i]} - \mathbf{x}^{b[i]} \in \mathrm{range}\left\{\widehat{\boldsymbol{\Delta}\mathbf{X}}\right\}.$$

which yields to,

$$\mathbf{X}^a - \mathbf{X}^b = \widehat{\boldsymbol{\Delta}\mathbf{X}} \cdot \mathbf{W}$$

with

$$\mathbf{W} = \mathbf{V}^T \cdot \left[\mathbf{R} + \mathbf{V} \cdot \mathbf{V}^T\right]^{-1} \cdot \mathbf{D}.$$

Efficient Implementations of the EnKF

- What are your conclusions so far?

- Lets project everything onto the ensemble space...
- Recall

$$\mathbf{X}^a = \mathbf{X}^b + \left[\mathbf{P}^{b^{-1}} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}\right]^{-1} \cdot \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{D}$$

- The inverse of the (estimated) background error covariance matrix $\mathbf{P}^{b^{-1}} \in \mathbb{R}^{n \times n}$ onto such space reads,

$$(N-1) \cdot \mathbf{I}$$

- The data error component $\mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H} \in \mathbb{R}^{n\times n}$ in the ensemble space reads,

$$[\mathbf{H} \cdot \mathbf{\Delta X}]^T \, \mathbf{R}^{-1} \cdot [\mathbf{H} \cdot \mathbf{\Delta X}]$$

  and therefore $\mathbf{P}^{b^{-1}} + \mathbf{H}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{H}$, in the ensemble space reads,

$$\mathbf{W} = (N-1) \cdot \mathbf{I} + \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Q} \in \mathbb{R}^{N\times N} \, .$$

  where $\mathbf{Q} = \mathbf{H} \cdot \mathbf{\Delta X} \in \mathbb{R}^{m\times N}$.

- Lets perform a Cholesky decomposition on $\mathbf{W}$

$$\mathbf{L} \cdot \mathbf{L}^T = \mathrm{Cholesky}\left(\mathbf{W}\right) \, .$$

- Hence,

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{\Delta X} \cdot \mathbf{Z} \,,$$

with

$$\mathbf{L} \cdot \mathbf{L}^T \cdot \mathbf{Z} = \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{D} \,.$$

- The computational effort of this method reads,

$$\mathcal{O}\left(m \cdot N^2 + N^3 + n \cdot N^2\right) \,.$$

- Consider the formulation,

$$\mathbf{X}^a = \mathbf{X}^b + \mathbf{P}^b \cdot \mathbf{H}^T \cdot \left[\mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T + \mathbf{R} \right]^{-1} \cdot \mathbf{D}$$

- Since background and data error are uncorrelated, the next factorization is possible,

$$\left[\mathbf{R} + \mathbf{H} \cdot \mathbf{P}^b \cdot \mathbf{H}^T \right] = \left[\mathbf{H} \cdot \widehat{\boldsymbol{\Delta}\mathbf{X}} + \mathbf{R}^{1/2} \right] \cdot \left[\mathbf{H} \cdot \widehat{\boldsymbol{\Delta}\mathbf{X}} + \mathbf{R}^{1/2} \right]^T \in \mathbb{R}^{m \times m} \,.$$

- Consider the Singular Value Decomposition,

$$\mathbf{H} \cdot \widehat{\mathbf{\Delta X}} + \mathbf{R}^{1/2} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{E}^T \,.$$

Hence,

$$\left[\mathbf{H} \cdot \widehat{\mathbf{\Delta X}} + \mathbf{R}^{1/2}\right] \cdot \left[\mathbf{H} \cdot \widehat{\mathbf{\Delta X}} + \mathbf{R}^{1/2}\right]^T = \mathbf{U} \cdot \mathbf{\Sigma}^2 \cdot \mathbf{U}^T$$

and consequently,

$$\mathbf{X}^a = \mathbf{X}^b + \widehat{\mathbf{\Delta X}} \cdot \mathbf{V}^T \cdot \mathbf{Z}$$

where here,

$$\mathbf{U} \cdot \mathbf{\Sigma}^2 \cdot \mathbf{U}^T \cdot \mathbf{Z} = \mathbf{D} \,.$$

- The computational effort of this method reads,

$$\mathcal{O}\left(N^2 \cdot n + m \cdot N^2 + m \cdot N \cdot n + m\right)$$

- Yet, another efficient implementation of the EnKF can be obtained making use of the Sherman Morrison formula.
- Consider the covariance matrix,

$$\mathbf{W} = \mathbf{R} + \mathbf{V} \cdot \mathbf{V}^T \in \mathbb{R}^{m \times m}$$

notice, this can be written as follows,

$$\mathbf{W} = \mathbf{R} + \sum_{i=1}^{N} \mathbf{v}_i \cdot \mathbf{v}_i^T \,,$$

where $\mathbf{v}_i \in \mathbb{R}^{m \times 1}$ is the i-th column of matrix $\mathbf{V} = \mathbf{H} \cdot \widehat{\mathbf{\Delta X}} \in \mathbb{R}^{m \times N}$. from here, the analysis step of the EnKF can be written as follows,

$$\mathbf{X}^a = \mathbf{X}^b + \widehat{\mathbf{\Delta X}} \cdot \mathbf{V}^T \cdot \mathbf{Z} \,,$$

where $\mathbf{Z} \in \mathbb{R}^{m \times N}$ is the solution of the linear system,

$$\mathbf{W} \cdot \mathbf{Z} = \mathbf{D} \Leftrightarrow \left[\mathbf{R} + \sum_{i=1}^{N} \mathbf{v}_i \cdot \mathbf{v}_i^T \right] \cdot \mathbf{Z} = \mathbf{D} \,. \tag{1}$$

- For the solution of linear system (1), consider the sequence of matrices,

$$\begin{aligned} \mathbf{W}^{(0)} &= \mathbf{R} \\ \mathbf{W}^{(1)} &= \mathbf{W}^{(0)} + \mathbf{v}_1 \cdot \mathbf{v}_1^T \\ \mathbf{W}^{(2)} &= \mathbf{W}^{(1)} + \mathbf{v}_2 \cdot \mathbf{v}_2^T \\ \vdots &= \vdots \\ \mathbf{W}^{(N)} &= \mathbf{W}^{(N-1)} + \mathbf{v}_N \cdot \mathbf{v}_N^T \,. \end{aligned}$$

Hence, the linear system (1) can be written as follows,

$$\left[\mathbf{W}^{(N-1)} + \mathbf{v}_N \cdot \mathbf{v}_N^T\right] \cdot \mathbf{z}_i = \mathbf{d}_i \,.$$

where we have considered only one column of $\mathbf{D}$.

- Making use of the Sherman Morrison formula, we have,

$$\begin{aligned} \mathbf{z}_i &= \left[\mathbf{W}^{(N-1)} + \mathbf{v}_N \cdot \mathbf{v}_N^T\right]^{-1} \cdot \mathbf{d}_i \\ &= \left[\mathbf{W}^{(N-1)}\right]^{-1} \cdot \mathbf{d}_i - \frac{1}{\gamma} \cdot \left\{\left[\mathbf{W}^{(N-1)}\right]^{-1} \cdot \mathbf{v}_N\right\} \cdot \mathbf{v}_N^T \\ &\cdot \left[\mathbf{W}^{(N-1)}\right]^{-1} \cdot \mathbf{d}_i \end{aligned}$$

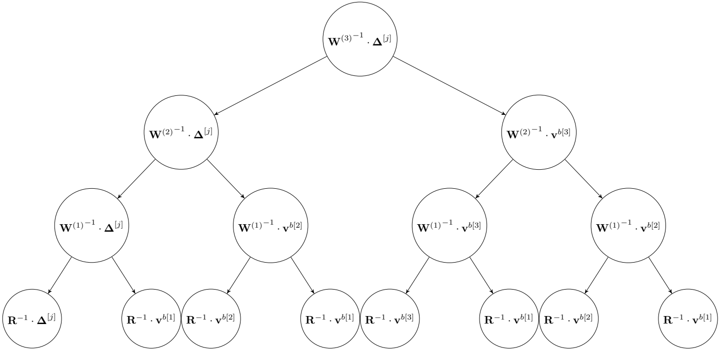
where

$$\gamma = 1 + \mathbf{v}_N^T \cdot \left[\mathbf{W}^{(N-1)}\right]^{-1} \cdot \mathbf{v}_N \,.$$

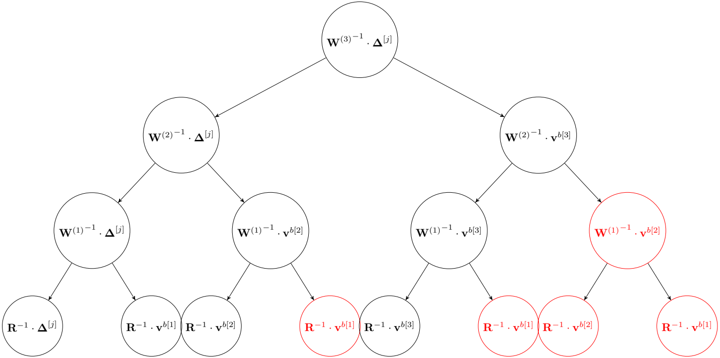
- The new linear systems to solve can be computed making use of the same idea, from the "recursive Sherman Morrison EnKF" is derived,

$$\mathrm{Analysis\_SMF}(\mathbf{x}, \mathrm{i}) = \begin{cases} {\color{red}\mathbf{z}} = \mathbf{R}^{-1} \cdot \mathbf{x} & \text{, for } i = 0 \\ \mathbf{f} = \mathrm{Analysis\_SMF}(\mathbf{x}, \mathrm{i}-1) & \\ \mathbf{g} = \mathrm{Analysis\_SMF}(\mathbf{v}_\mathrm{i}, \mathrm{i}-1) & \text{, otherwise} \\ \gamma = 1 + \mathbf{v}_i^T \cdot \mathbf{g} & \\ {\color{red}\mathbf{z}} = \mathbf{f} - \frac{1}{\gamma} \cdot \mathbf{g} \cdot \left[\mathbf{v}_i^T \cdot \mathbf{f}\right] \; . & \end{cases}$$

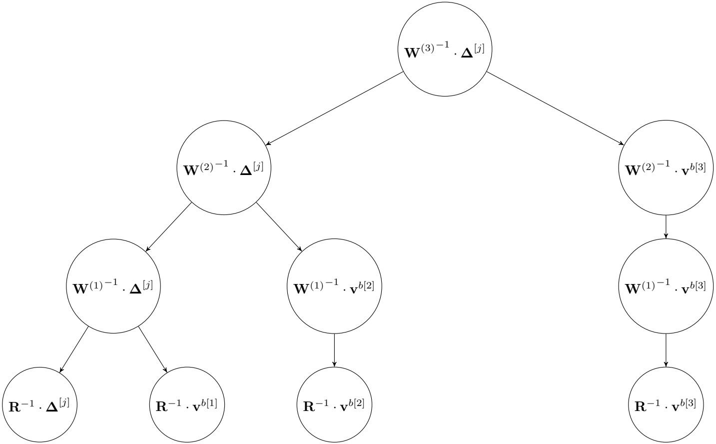
- The computations derived by this subroutine are shown next,



(a) All computations derived from the Sherman Morrison EnKF



(a) Common computations derived from the Sherman Morrison EnKF



(a) Optimal computations derived from the SM-EnKF.

- We can elaborate an iterative version which mimics the behaviour of the recursive one and even more, it is able to avoid common computations,

1. Level 0

$$
\begin{aligned}
\mathbf{Z}^{(0)} &= \mathbf{R}^{-1} \cdot \mathbf{D} \\
\mathbf{U}^{(0)} &= \mathbf{R}^{-1} \cdot \mathbf{V}
\end{aligned}
$$

2. Level $1 \leq i \leq N$

$$
\begin{aligned}
\mathbf{h}^{(k)} &= \left[1 + \mathbf{v}_i^T \cdot \mathbf{u}_i^{(i-1)}\right]^{-1} \cdot \mathbf{u}_i^{(i-1)} \in \mathbb{R}^{m \times 1} \\
\mathbf{Z}^{(i)} &= \mathbf{Z}^{(i-1)} - \mathbf{h}^{(i)} \cdot \left[\mathbf{v}_i^T \cdot \mathbf{Z}^{(i-1)}\right] \in \mathbb{R}^{m \times N} \\
\mathbf{u}_j^{(i)} &= \mathbf{u}_j^{(i)} - \mathbf{h}^{(i)} \cdot \left[\mathbf{v}_i^T \cdot \mathbf{u}_j^{(i)}\right]
\end{aligned}
$$

where $i + 1 \leq j \leq N$.

Avoiding Filter Divergence

Avoiding Filter Divergence

Localization