A Parallel Ensemble Kalman Filter Implementation Based On Modified Cholesky Decomposition

Elias D. Nino-Ruiz and Adrian Sandu
Computational Science Laboratory,
Virginia Polytechnic Institute and State University,
Blacksburg, VA 24060, USA
enino@vt.edu, sandu@cs.vt.edu

November 16, 2015
Table of contents

Problem Formulation

Local Ensemble Transform Kalman Filter

EnKF Based on Modified Cholesky Decomposition
   Estimating $\mathbf{B}^{-1}$
   EnKF implementations based on MC
   Domain decomposition

Experimental Settings

Conclusions
Problem Formulation I

We want to estimate $\mathbf{x}^* \in \mathbb{R}^{n \times 1}$,

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

model dimension: $n$.

Based on (assuming Gaussian errors):

A prior estimate (best estimate prior measurements):

$$
\mathbf{x}^b = \mathbf{x}^* + \xi, \text{ with } \xi \sim \mathcal{N} (\mathbf{0}_n, \mathbf{B})
$$

where $\mathbf{B} \in \mathbb{R}^{n \times n}$ is unknown.

A noisy observation:

$$
\mathbf{y} = \mathcal{H} (\mathbf{x}^*) + \mathbf{e}, \text{ with } \mathbf{e} \sim \mathcal{N} (\mathbf{0}_m, \mathbf{R}),
$$

number of observed model components: $m$. $\mathbf{R}^{m \times m}$ is the data error covariance matrix. $\mathcal{H} : \mathbb{R}^{n \times 1} \rightarrow \mathbb{R}^{m \times 1}$. With $m \ll n$ or $m < n$. 
Bayesian approximation (posterior state):

\[ x^a = x^b + B \cdot H^T \cdot \left[ R + H \cdot B \cdot H^T \right]^{-1} \cdot d \in \mathbb{R}^{n \times 1} \]

where \( d = y - \mathcal{H}(x^b) \in \mathbb{R}^{m \times 1} \) and \( \mathcal{H}' \approx H \in \mathbb{R}^{m \times N} \).

How can we estimate \( B \)?

Background error statistics of any model state \( x \in \mathbb{R}^{n \times 1} \):

\[ x \sim \mathcal{N}(x^b, B) \]
Empirical moments of an ensemble:

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

\[ \mathbf{x}^b \approx \bar{x}^b = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}^{b[i]} \in \mathbb{R}^{n \times 1}, \quad \mathbf{B} \approx \mathbf{P}^b = \mathbf{S} \cdot \mathbf{S}^T \in \mathbb{R}^{n \times n} \]

where \( \mathbf{S} = \frac{1}{\sqrt{N-1}} \cdot \left[ \mathbf{X}^b - \bar{x}^b \otimes \mathbf{1}_N^T \right] \in \mathbb{R}^{n \times N}. \)
Problem Formulation IV

- The posterior (analysis) ensemble:

\[
X^a = X^b + P^b \cdot H^T \cdot \left[ R + H \cdot P^b \cdot H^T \right]^{-1} \cdot D \in \mathbb{R}^{n \times N},
\]

the \( i \)-th column of \( D \in \mathbb{R}^{m \times N} \) reads:

\[
d[i] = y + \epsilon[i] - H \cdot x^b[i] \in \mathbb{R}^{m \times 1}, \text{ for } 1 \leq i \leq N,
\]

with (stochastic version of the filter):

\[
\epsilon[i] \sim \mathcal{N}(0_m, R).
\]
Unfortunately, the number of samples $N$ is much lower than the model dimension $n \gg N$.

- $P^b$ is low-rank. (Spurious correlations)
- $X^a$ is computed in the ensemble space (few degrees of freedom)

Model dimensions are in the order of billions, while ensemble sizes in the order of hundreds.

Model propagations are computationally expensive.

Computational effort of the analysis is high.

We do need HPC not only to speedup computations but to have enough memory to represent ensemble members and to perform linear algebra computations.
Local Ensemble Transform Kalman Filter (LETKF) I

- One of the best parallel ensemble based implementations.

- Analysis equations:
  - Perturbations: \( \mathbf{U} = \mathbf{X}^b - \bar{x}^b \otimes \mathbf{1}_N^T \in \mathbb{R}^{n \times N} \).
  - Optimality in ensemble space: \( \mathbf{Q} = \mathbf{H} \cdot \mathbf{U} \in \mathbb{R}^{m \times N} \),
    \[
    \tilde{\mathbf{P}}^a = \left[ (N - 1) \cdot \mathbf{I}_{N \times N} + \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot \mathbf{Q} \right]^{-1} \in \mathbb{R}^{N \times N}
    \]
    \[
    \mathbf{w}^a = \tilde{\mathbf{P}}^a \cdot \mathbf{Q}^T \cdot \mathbf{R}^{-1} \cdot [\mathbf{y} - \mathbf{H} \cdot \bar{x}^b]
    \]
    \[
    \mathbf{W} = \mathbf{w}^a \otimes \mathbf{1}_N^T + \mathbf{W}^a \in \mathbb{R}^{N \times N}, \quad \mathbf{W}^a = \left[ (N - 1) \cdot \tilde{\mathbf{P}}^a \right]^{1/2} \in \mathbb{R}^{N \times N}
    \]

- Analysis ensemble:
  \[
  \mathbf{X}^a = \bar{x}^b \otimes \mathbf{1}_N^T + \mathbf{U} \cdot \mathbf{W} \in \mathbb{R}^{n \times N}.
  \]

- Domain localization [OHS+04, Kep00]:

Virginia Tech
Compute the Future
Local Ensemble Transform Kalman Filter [8/29]
Local Ensemble Transform Kalman Filter (LETKF) II

(d) $r = 1$

(e) $r = 3$
When $m_i$ and $m_j$ are conditionally independent, $C^{-1}_{m_i,m_j} = 0$.

We want to estimate $B^{-1}$:

- Recall $U = X^b - \bar{x}^b \otimes 1_N^T \in \mathbb{R}^{n \times N}$. Thus, $u[i] \sim \mathcal{N}(0_n, B)$, for $1 \leq i \leq N$.
- Let $x[i] \in \mathbb{R}^{N \times 1}$ the vector holding the $i$-th row across all columns of $U$, for $1 \leq i \leq n$.
- Then, the approximation of $B^{-1}$ arises from:

$$x[i] = \sum_{j=1}^{i-1} x[j] \cdot \beta_{i,j} + \xi[i] \in \mathbb{R}^{N \times 1}$$
Estimating $\mathbf{B}^{-1}$ II

- By the modified Cholesky (MC) decomposition for inverse covariance matrix estimation:

\[
\mathbf{B}^{-1} \approx \hat{\mathbf{B}}^{-1} = \mathbf{T}^T \cdot \mathbf{D}^{-1} \cdot \mathbf{T} \in \mathbb{R}^{n \times n}
\]
\[
\mathbf{B} \approx \hat{\mathbf{B}} = \mathbf{T}^{-1} \cdot \mathbf{D} \cdot \mathbf{T}^{-1}^T \in \mathbb{R}^{n \times n}
\]

where $\mathbf{T} \in \mathbb{R}^{n \times n}$ is an unitary lower triangular matrix with \(\{\mathbf{T}\}_{i,j} = -\beta_{i,j}\) and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with $\{\mathbf{D}\}_{i,i} = \text{var}(\xi_i)$, for $1 \leq j < i \leq n$.

- $\hat{\mathbf{B}}^{-1}$ can be sparse, $\hat{\mathbf{B}}$ is not necessarily sparse. Structure of $\hat{\mathbf{B}}^{-1}$ depends on $\mathbf{T}$. 

EnKF Based on Modified Cholesky Decomposition
Choosing the predecessors

(f) Row-major

<table>
<thead>
<tr>
<th>1</th>
<th>5</th>
<th>9</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>6</td>
<td>10</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
<td>16</td>
</tr>
</tbody>
</table>

(g) Column-major

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

(h) $r = 1$

(i) Predecess.
EnKF Based on Modified Cholesky Decomposition

estimating $B^{-1}$ [13/29]

EnKF formulations based on modified Cholesky decomposition for inverse background error estimation

▶ Primal:

\[ X^a = X^b + \left( \hat{B}^{-1} + H^T \cdot R^{-1} \cdot H \right)^{-1} \cdot H^T \cdot R^{-1} \cdot \left[ Y_s - H \cdot X^b \right] \]

▶ Dual:

\[ X^a = X^b + X \cdot V^T \cdot \left[ R + V \cdot V^T \right]^{-1} \cdot \left[ Y_s - H \cdot H \cdot X^b \right] \]

where \( T \cdot X = D^{1/2} \in \mathbb{R}^{n \times n} \) and \( V = H \cdot X \in \mathbb{R}^{m \times n} \).

▶ Efficient implementations [NRSA14, NRS15].
Domain Decomposition
Boundary Information
SPEEDY is a simplified GCM developed at ICTP by Franco Molteni and Fred Kucharski.
Nicknamed SPEEDY, for "Simplified Parameterizations, primitive-Equation DYnamics"
It is a hydrostatic, s-coordinate, spectral-transform model in the vorticity-divergence form, with semi-implicit treatment of gravity waves.
- 8 layers, $u$, $v$, $T$ and $sh$.
- T-63 resolution (96 x 192)
BlueRidge is a 408-node Cray CS-300 cluster.
Each node is outfitted with two octa-core Intel Sandy Bridge CPUs and 64 GB of memory.
Total of 6,528 cores and 27.3 TB of memory systemwide.
Eighteen nodes have 128 GB of memory.
In addition, 130 nodes are outfitted with two Intel MIC (Xeon Phi) coprocessors.
Experimental settings

- Number of ensemble members 96.
- 3 radius of influence are considered: 3, 4, 5.
- Model is propagated for 2 days and then observations are assimilated.
- Number of processors: 6 computing nodes (96 processors) up to 128 computing nodes (2048 processors)
- Fortran 90 and 77, MPI, LAPACK and BLAS.
- 3 different observational networks.
Observational networks

Figure: Sparse observational networks. Observed components in black. $p$ denotes percentage of observed model components.
RMSE for some configurations.

(a) $r = 3$, $p \sim 12\%$, $u$

(b) $r = 4$, $p \sim 12\%$, $v$

(c) $r = 5$, $p \sim 6\%$, $u$
Initial snapshots for $r = 5$ and $p \sim 4\%$ for $\nu$

(a) Reference  
(b) Background  
(c) EnKF-MC  
(d) LETKF
Initial snapshots for $r = 5$ and $p \sim 4\%$ for $u$

(a) Reference

(b) Background

(c) EnKF-MC

(d) LETKF
RMSE for different variables and # of computing nodes.

(a) $r = 5$, $p \sim 4\%$, $u$

(b) $r = 5$, $p \sim 4\%$, $v$

(c) $r = 5$, $p \sim 4\%$, $sh$
Elapsed time.

![Graph showing elapsed time for EnkF-MC and LETKF (No Cov. Estimation) over varying computing nodes. The x-axis represents computing nodes (x 16 processors) ranging from 0 to 150, and the y-axis represents time (s) ranging from 300 to 0. The graph includes two lines: one for EnkF-MC and another for LETKF (No Cov. Estimation).]
Conclusions

- The proposed implementations outperform the LETKF under the RMSE metric.
- Parallel resources and domain decompositions can be exploited in order to speed up the assimilation process.
- Localization is implicit. Domain decomposition is used just for computational reasons.
- The computational effort of the proposed method makes it attractive for the use under realistic scenarios.
Thank You.

*(John 3:16)* For God so loved the world that he gave his one and only Son, that whoever believes in him shall not perish but have eternal life.


\[
\left[ \hat{B}^{-1} + H^T \cdot R^{-1} \cdot H \right]^{-1} = \left[ X^T \cdot X + H^T \cdot R^{-1} H \right]^{-1} \\
= \left\{ X^T \cdot \left[ I_{n \times n} + Q \cdot Q^T \right] \cdot X \right\}^{-1} \\
= X^{-1} \cdot \left[ I_{n \times n} + Q \cdot Q^T \right]^{-1} \cdot X^{-T}
\]

where \( X^T \cdot Q = H^T \cdot R^{-1/2} \).