A Hybrid Covariance localization / local analysis algorithm for parallel Square-Root EnKF

Jeff Steward, Altug Aksoy, Ziad Haddad

EnKF Workshop
May 25, 2016

Kalman filter

- Assimilating satellite observations challenging
- One idea: use stats to extract maximally certain information (Fri. poster)
- More and more observations ($N_{\text{obs}} \to 10^6$)
- Truly a "big-data" problem
- Require new scalable algorithms to handle these sizes
- Not all implementations will scale well

Kalman filter

- The Square-Root Kalman Filter is given by

$$\mathbf{x}_m^{(a)} = \mathbf{x}_m^{(f)} + \mathbf{K}\left(\mathbf{y} - \mathbf{HX}\right) \tag{1}$$

$$\mathbf{X}_A^{(a)} = \mathbf{X}_A^{(f)} + \tilde{\mathbf{K}}\left(\mathbf{0} - \mathbf{HA}\right) \tag{2}$$

  where

  - $\mathbf{x}_m$ ensemble mean, $\mathbf{X}_A$ perturbations, (a) posterior, (f) prior
  - $\mathbf{K} = \mathbf{C}_{\mathbf{x},\mathbf{y}}\mathbf{D}^{-1}$, $\mathbf{C}_{\mathbf{x},\mathbf{y}} = \text{cov}\left(\mathbf{x}^{(f)}, \mathbf{H}\left(\mathbf{x}^{(f)}\right)\right)$, $\mathbf{D} = \mathbf{C}_{\mathbf{y},\mathbf{y}} + \mathbf{R}$
  - $(\mathbf{HX})_i = h_i\left(\mathbf{x}_m^{(f)}\right)$, possibly non-linear
  - $\mathbf{HA}_{i,j} = h_i\left(\mathbf{x}_m^{(f)} + \left(\mathbf{X}_A^{(f)}\right)_j\right) - (\mathbf{HX})_i$
  - $\tilde{\mathbf{K}} = \mathbf{C}_{\mathbf{x},\mathbf{y}}\,\mathbf{D}^{-1/2}\left(\sqrt{\mathbf{D}} + \sqrt{\mathbf{R}}\right)^{-1}$

- Involves the inverse and $\sqrt{}$ of $\mathbf{D}$, $(n_{obs} \times n_{obs})$

Serial implementation

- Idea: Assume obs uncorrelated, assimilate observations serially
- As implemented in HEDAS (Aksoy et al. 2012) for each obs i:

$$\mathbf{K}_i = \rho_i \circ \mathbf{P}_f^{(i-1)} \mathbf{H}_i^{\mathrm{T}} / (\mathbf{H}_i \mathbf{P}_f^{(i-1)} \mathbf{H}_i^{\mathrm{T}} + r_i) \tag{3}$$

$$\overline{\mathbf{x}_m}^{(i)} = \overline{\mathbf{x}_f}^{(i-1)} + \mathbf{K}_i (y_i - h_i(\overline{x})) \tag{4}$$

$$\mathbf{X}_A^{(i)} = \mathbf{X}_A^{(i-1)} - \frac{\mathbf{K}_i}{1 + r_i / (\mathbf{H}_i \mathbf{P}_f^{(i-1)} \mathbf{H}_i^{\mathrm{T}} + r_i)} \mathbf{H}_i \mathbf{A} \tag{5}$$

where

- $\mathbf{H}_i$ is the i^{th} is the possibly non-linear operator h_i
- $\mathbf{H}_{i,j}\mathbf{A}$ is $h_i(\mathbf{X}_j) - h_i(\mathbf{x}_m)$, $\mathbf{P}_f^{(i)} = \mathbf{X}_A^{(i)} \left(\mathbf{X}_A^{(i)}\right)^{\mathrm{T}} / (NE - 1)$
- r_i is the observation error variance for the i^{th} observation

- Assumes $\mathbf{HP}_f\mathbf{H}^{\mathrm{T}} + \mathbf{R}$ is diagonal, so inverse is scalar divide
- $\mathbf{R}$ is diagonal, but $\mathbf{HP}_f\mathbf{H}^{\mathrm{T}}$ is not $\Leftrightarrow (h_i(X), h_j(X))$ correlated
- Diagonally dominant, however

Simple example

- Test w/ simple EnKF, 4 points and 2 obs, identity obs op

```
K>> D   = H*Pf*H'  + R
D =
     0.0091    -0.0005
    -0.0005     0.0111
```

- Diagonal is 20 times larger than the off-diagonal terms
- Serial update actually recovers upper half of this matrix
- Misses lower half (update from obs not yet assimilated)
- Neglecting adds non-trivial error

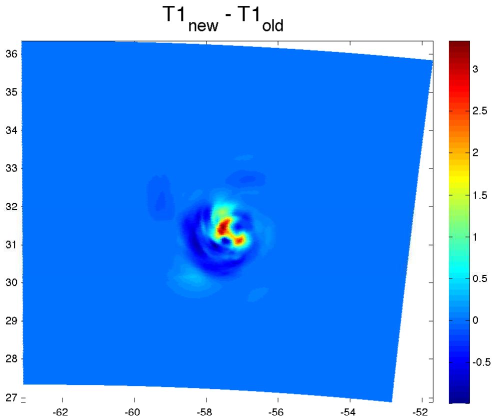


Figure: Difference in all 2×10^4 obs Edouardo case from two orderings, surface T

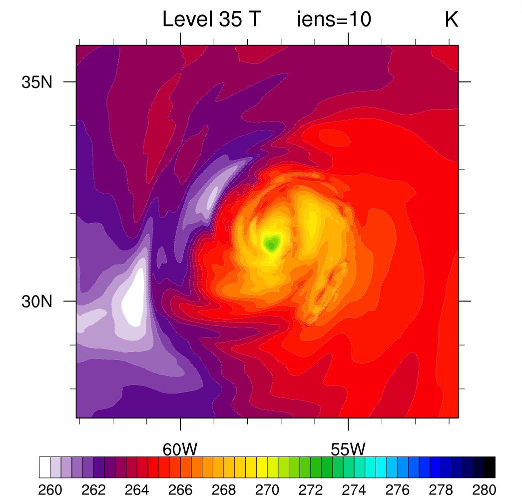


Figure: 10 Identity observations, ordering 1

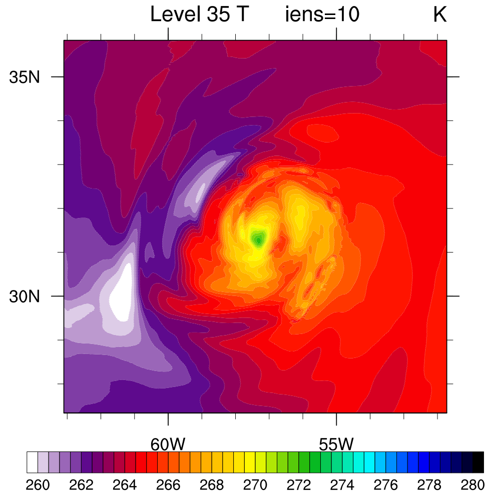


Figure: 10 Identity observations, ordering 2

Serial filter

- Ordering matters, can introduce non-trivial error
- Serial approach difficult to parallelize
- Requires “fixing up” unassimilated observations
- Does not solve Kalman equation (by neglecting off-diag)
- $\Rightarrow$ time to rethink serial approach to $\sqrt{\text{EnKF}}$?

$\sqrt{\text{EnKF}}$ Rethink

- Problem is $n_{obs} \times n_{obs}$ matrix invs in $\mathbf{K}$ and $\tilde{\mathbf{K}}$

$$\mathbf{K} = \mathbf{C_{x,y}} \mathbf{D}^{-1} \quad (6)$$

$$\tilde{\mathbf{K}} = \mathbf{C_{x,y}}\ \mathbf{D}^{-1/2} \left(\sqrt{\mathbf{D}} + \sqrt{\mathbf{R}} \right)^{-1} \quad (7)$$

  for $\mathbf{D} = \mathbf{C_{y,y}} + \mathbf{R}$
- Diagonal approximation accumulates errors. Solve directly?
- How? Square-root of **D** would be costly to evaluate, invert

$\sqrt{\text{EnKF}}$ Rethink

- Innovation in theory of matrix functions $f(\mathbf{A})$ by Hingham
- In our case, $\mathbf{M} = f(\mathbf{D}) = \mathbf{D} + \sqrt{\mathbf{D}}$, want $\mathbf{x}$ in $\mathbf{Mx} = \mathbf{y}$
- Eigenvectors of $f(\mathbf{D})$ are $\mathbf{v}_i$, same as $\mathbf{v}_i$ eigvecs of $\mathbf{D}$
- Eigenvalues of $f(\mathbf{D})$ are $f(\lambda_i)$ where λ_i are eigvals of $\mathbf{D}$
- Sufficient to find eigpairs of $\mathbf{D}$, $f(\lambda_i) = \lambda_i + \lambda_i^{1/2}$
  $\mathbf{Mv}_i = \mathbf{Dv}_i + \mathbf{D}^{1/2}\mathbf{v}_i = \left(\lambda_i + \lambda_i^{1/2}\right)\mathbf{v}_i$
- Inverse is $\mathbf{M}^{-1}\mathbf{v}_i = \lambda_i'\mathbf{v}_i$ for $\lambda_i' = 1/\left(\lambda_i + \lambda_i^{1/2}\right)$

$\sqrt{\text{EnKF}}$ Rethink

- To analyze the eigenpairs further, transform to a white **R**
- For $\mathbf{D} = \mathbf{C}_{\mathbf{y},\mathbf{y}} + \mathbf{R}$, transform $\mathbf{y}' = \mathbf{R}^{-1/2}\mathbf{y}$
- This gives $\mathbf{D}' = \mathbf{C}'_{\mathbf{y},\mathbf{y}} + \mathbf{R}' = \mathbf{C}'_{\mathbf{y},\mathbf{y}} + \mathbf{I}$
- Then $\lambda_{\mathbf{D}'_i} = \lambda_{\left(\mathbf{C}'_{y,y}\right)_i} + 1$ and eigvecs are same
- $\mathbf{D} + \mathbf{D}^{1/2} \to \mathbf{D} \to \mathbf{C}_{\mathbf{y},\mathbf{y}}$ share eigvecs, eigvals change
- In terms of $\mathbf{C}_{\mathbf{y},\mathbf{y}}$ (drop prime) eigenpairs,

$$\mathbf{M}^{-1}\mathbf{v}_i = \lambda'_i \mathbf{v}_i \tag{8}$$

for $\lambda'_i = 1/\left(\lambda_i + 1 + (\lambda_i + 1)^{1/2}\right)$

$\sqrt{\text{EnKF}}$ Rethink

- $\mathbf{M}^{-1}\mathbf{v}_i = \lambda_i'\mathbf{v}_i$ for $\lambda_i' = 1/\left(\lambda_i + 1 + (\lambda_i + 1)^{1/2}\right)$
- $\lambda_i \geq 0$ as $\mathbf{C}_{\mathbf{y},\mathbf{y}}$ is symmetric positive semi-definite
- As $\lambda_i \to \infty$, λ_i' goes to 0, while for $\lambda_i \to 0$, λ_i' goes to 1/2.
- Problem becomes: find eigenpairs of $\mathbf{C}_{\mathbf{y},\mathbf{y}}$
- $\Rightarrow$ can neglect large λ_i, small λ_i of $\mathbf{C}_{\mathbf{y},\mathbf{y}}$

$\sqrt{\text{EnKF}}$ Rethink

- $\mathbf{C}_{\mathbf{y},\mathbf{y}} = \rho_{\mathbf{y},\mathbf{y}} \circ \mathbf{Q}_{\mathbf{y},\mathbf{y}}$ for $\mathbf{Q}_{\mathbf{y},\mathbf{y}} = \frac{\mathbf{HA}(\mathbf{HA})^{\mathrm{T}}}{n_{\text{ens}}-1}$
- $(\rho_{\mathbf{y},\mathbf{y}})_{i,j} = \ell(d_{i,j}|L)$ for distance d and length scale L
- By Schur product rank inequality theorem, $\text{rank}\left(\mathbf{C}_{\mathbf{y},\mathbf{y}}\right) \leq \text{rank}\left(\rho_{\mathbf{y},\mathbf{y}}\right) \text{rank}\left(\mathbf{HA}\right)$
- As rank($\mathbf{Q}$) is N_{ens}, rank($\mathbf{C}_{\mathbf{y},\mathbf{y}}$) $\leq N_{\text{ens}}$ rank($\rho_{\mathbf{y},\mathbf{y}}$)
- Rank of $\rho_{\mathbf{y},\mathbf{y}}$ depends on L and obs distances.
- As $L \to 0$, $\ell_{i,j} = \delta_i^j$ (Kronecker δ), so $\mathbf{C}_{\mathbf{y},\mathbf{y}}$ is full rank.
- As $L \to \infty$, $\ell_{i,j} = 1$ (same as no loc), so rank($\mathbf{C}_{\mathbf{y},\mathbf{y}}$) $= n_{\text{ens}}$
- Therefore finding eigenpairs best when $L >> 0$

Direct $\sqrt{\text{EnKF}}$ Solution

- Now have an algorithm:
  - for obs with small L, use local analysis (e.g. LETKF, Hunt et al. 2007)
  - for obs with large L, find $\mathbf{K}$, $\tilde{\mathbf{K}}$ with eigpairs of $\mathbf{C}_{\mathbf{y},\mathbf{y}}$
  - Direct square-root removes small L obs from analysis
  - LETKF analysis “wins”
- Division between “large” and “small” depends on problem
- Requires scalable method for computing eigenpairs of sparse $n_{\text{obs}} \times n_{\text{obs}}$ matrix
- Library: SLEPc (Scalable Library for Eigen-Problem Computation – Hernandez, Roman, and Vidal 2005)
- Built upon PETSc (Portable Extensible Toolkit for Scientific Computing – Balay et al. 1997)

Direct $\sqrt{\text{EnKF}}$ Solution

- Use Multiple Shift-Invert Lanczos method of Aktulga et al. 2014
- First investigate eigval spectrum through inertia theorem
- To find number of eigvals of **A** between (α, β), first compute

$$\mathbf{A} - \alpha\mathbf{I} = \mathbf{L}_\alpha \mathbf{D}_\alpha \mathbf{L}_\alpha^{\mathrm{T}} \tag{9}$$

$$\mathbf{A} - \beta\mathbf{I} = \mathbf{L}_\beta \mathbf{D}_\beta \mathbf{L}_\beta^{\mathrm{T}} \tag{10}$$

  where for $\mathbf{L}_x \mathbf{D}_x \mathbf{L}_x^{\mathrm{T}}$, $\mathbf{L}_x$ is lower-triangular, $\mathbf{D}_x$ is diagonal
- The difference between negative entries of $\mathbf{D}_\alpha$, $\mathbf{D}_\beta$ is number of eigvals in (α, β)

Direct $\sqrt{\text{EnKF}}$ Solution

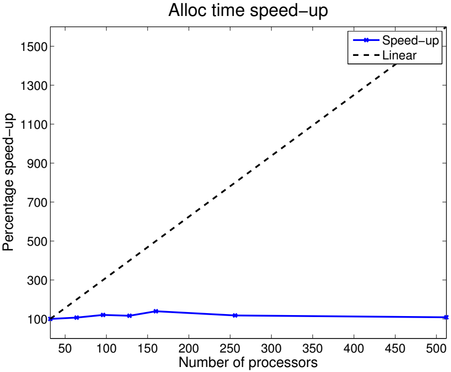
- Find N_{groups} (α, β) regions (monotonically increasing)
- Interpolate the spectrum for same number in each region
- Solve for all of the eigpairs in each of the N_{groups} in parallel

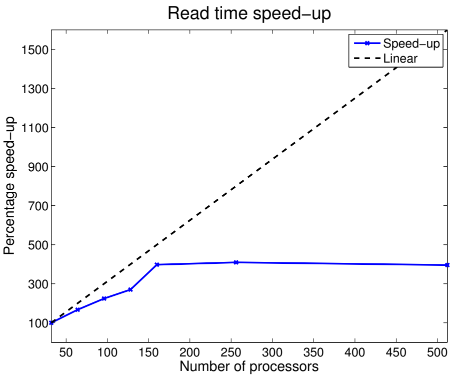
Hybrid direct $\sqrt{\text{EnKF}}$ Solution/LETKF

- Our hybrid parallel algorithm is therefore
  1. Divide computational domain (least-utilized round-robin)
  2. Divide small L, large L obs
  3. For large L obs:
     - 3.1 Slice the eigval spectrum by interpolating the inertia
     - 3.2 Compute in parallel all eigpairs within each region
     - 3.3 Compute $\mathbf{D}^{-1}(\mathbf{y} - \mathbf{HX})$ and $(\mathbf{D} + \mathbf{D}^{1/2})^{-1}(\mathbf{0} - \mathbf{HA})$ for $\mathbf{K}$ and $\tilde{\mathbf{K}}$
     - 3.4 Account for neglecting additional eigenpairs (MGS)
     - 3.5 Distribute and compute the analysis in an embarassingly parallel way
  4. For small L obs, perform LETKF

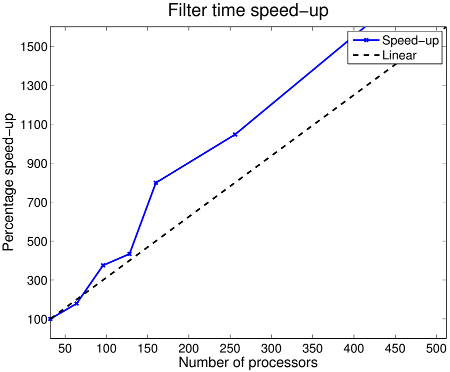
$\sqrt{\text{EnKF}}$ Rethink

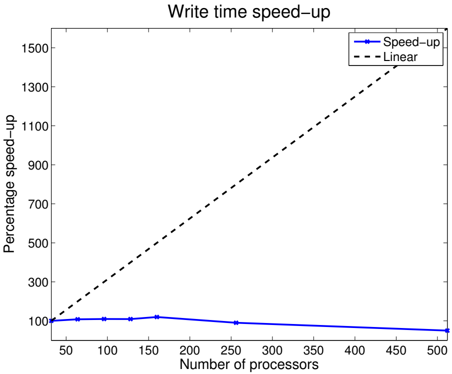
- Benefits of the new algorithm:
  - Scalable implementation
  - Ordering does not matter / no need to “patch up” observations
  - Fully consistent with EnKF update
  - Splits length scales for optimal efficiency
  - After eigenvalues computed, only store **M** and **Y** $(n_{obs} \times NE + 1)$
- Cons of the new algorithm:
  - More complex than serial
  - Requires integration, installation of PETSc, SLEPc
  - Setup and IO is fixed overhead, but the assimilation step is highly scalable


Alloc time speed−up
Speed−up
Linear
Percentage speed−up
1500
1300
1100
900
700
500
300
100
50
100
150
200
250
300
350
400
450
500
Number of processors



Read time speed−up
Speed−up
Linear
Percentage speed−up
1500
1300
1100
900
700
500
300
100
50
100
150
200
250
300
350
400
450
500
Number of processors


Filter time speed−up
Speed−up
Linear
Percentage speed−up
100
300
500
700
900
1100
1300
1500
50
100
150
200
250
300
350
400
450
500
Number of processors


Write time speed−up
Speed−up
Linear
Percentage speed−up
100
300
500
700
900
1100
1300
1500
50
100
150
200
250
300
350
400
450
500
Number of processors

References

Steward, J., A. Aksoy, and Z. Haddad. “Parallel Direct Solution of the Square-Root Ensemble Kalman Filter.” In preparation.

Aksoy, A, S. Lorsolo, T. Vukicevic, K. J. Sellwood, S. D. Aberson, and F.-Q. Zhang. “The HWRF Hurricane Ensemble Data Assimilation System (HEDAS) for High-Resolution Data: The Impact of Airborne Doppler Radar Observations in an OSSE.” Monthly Weather Review 2012 140:6, 1843-1862

Hingham, N. “Functions of Matrices: Theory and Computation.” Society for Industrial and Applied Mathematics, 978-0898716467, 2008.

Hernandez, V., J. Roman, and V. Vidal. “SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems.” ACM Transactions on Mathematical Software (TOMS) 31.3 (2005): 351-362.

Balay, S., et al. “Efficient management of parallelism in object-oriented numerical software libraries.” Modern software tools for scientific computing. Birkhäuser Boston, 1997. 163-202.

Hunt, B., E. Kostelich, and I. Szunyogh. “Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter.” Physica D: Nonlinear Phenomena 230.1 (2007): 112-126.

Aktulga, H., et al. “Parallel eigenvalue calculation based on multiple shift-invert Lanczos and contour integral based spectral projection method.” Parallel Computing 40.7 (2014): 195-212.