Kalman Filtering: Part II
Ensemble-based Kalman Filters

Instructor: Istvan Szunyogh

Lecture: July 30 2007

Summary of Part I: The cost function

The most likely trajectory of the system is the one that minimizes the cost function

$$J^o(\{\mathbf{x}(t)\}) = \sum_{j=1}^{n} [\mathbf{y}_j^o - H_j(\mathbf{x}(t_j))]^T \mathbf{R}_j^{-1} [\mathbf{y}_j^o - H_j(\mathbf{x}(t_j))]. \tag{1}$$

Thus, the "most likely" trajectory is also the one that best fits the observations in a least square sense. Three assumptions were made to obtain this cost function:

- The observation errors is Gaussian
- The the errors of the observations taken at different times t_j $(j = 1, \ldots, n)$ are uncorrelated
- The observed quantities depend on the system state in a known way $H_j(\mathbf{x}(t_j)$

Summary of Part I: Extended Kalman Filter

$$\bar{\mathbf{x}}_n^b = M_{t_{n-1},t_n}(\bar{\mathbf{x}}_{n-1}^a), \tag{2}$$

$$\mathbf{P}_n^b = \mathbf{M}_{t_{n-1},t_n}\mathbf{P}_{n-1}^a\mathbf{M}_{t_{n-1},t_n}^T. \tag{3}$$

$$\begin{aligned} J_{t_n}^o(\mathbf{x}) \;=\; & \\ & [\mathbf{x}-\bar{\mathbf{x}}_n^b]^T(\mathbf{P}_n^b)^{-1}[\mathbf{x}-\bar{\mathbf{x}}_n^b] \\ + \;& [\mathbf{y}_n^o - H_n(\mathbf{x})]^T\mathbf{R}_n^{-1}[\mathbf{y}_n^o - H_n(\mathbf{x})] + c. \end{aligned} \tag{4}$$

From (4) we determine the state estimate $\bar{\mathbf{x}}_n^a$ and its covariance $\mathbf{P}_n^a$.

- $\mathbf{M}_{t_{n-1},t_n}$ is the linearization of M_{t_{n-1},t_n} around $\bar{\mathbf{x}}_{n-1}^a$. (To obtain $\bar{\mathbf{x}}_n^a$ and its covariance $\mathbf{P}_n^a$ we also need the linearization $\mathbf{H}_n$ of H_n around $\bar{\mathbf{x}}_n^b$.)

- The main obstacle to a practical implementation is the computational cost of (3).

Summary of Part I: Strategies for Approximate Solutions

- Direct Minimization of (4): Variational data assimilation schemes (3DVar and 4DVar)

- Solving the equations that provide the minimum of (4) for the linear case (OI and PSAS): $\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{P}_n^a \mathbf{H}_n^T \mathbf{R}_n^{-1} (\mathbf{y}_n^o - \mathbf{H}_n \bar{\mathbf{x}}_n^b)$, where $\mathbf{P}_n^a = (\mathbf{I} + \mathbf{P}_n^b \mathbf{H}_n^T \mathbf{R}_n^{-1} \mathbf{H}_n)^{-1} \mathbf{P}_n^b$.

For both types of schemes, in the operational practice

- The state estimate is evolved from one analysis cycle to the next using (2) [$\bar{\mathbf{x}}_n^b = M_{t_{n-1},t_n}(\bar{\mathbf{x}}_{n-1}^a)$]

- $\mathbf{P}_j^b$ is kept constant and an equivalent of (3) is not needed

An Alternative Strategy: Ensemble-Based Kalman Filters

- Assume that we have an ensemble $\{\mathbf{x}_{n-1}^{a(i)} : i = 1, 2, \ldots, k\}$ of model state vectors at time t_{n-1}

- The background ensemble $\{\mathbf{x}_n^{b(i)} : i = 1, 2, \ldots, k\}$ at time t_n is obtain by:

$$\mathbf{x}_n^{b(i)} = M_{t_{n-1},t_n}(\mathbf{x}_{n-1}^{a(i)}).$$

- For the background state estimate and its covariance, we use the sample mean and covariance of the background ensemble:

$$\bar{\mathbf{x}}^b = k^{-1} \sum_{i=1}^{k} \mathbf{x}^{b(i)},$$

$$\mathbf{P}^b = (k-1)^{-1} \sum_{i=1}^{k} (\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b)(\mathbf{x}^{b(i)} - \bar{\mathbf{x}}^b)^T$$

An Alternative Strategy: Ensemble-Based Kalman Filters (continued)

Based on the above information, an ensemble-based Kalman scheme returns an ensemble $\{\mathbf{x}^{a(i)} : i = 1, 2, \ldots, k\}$ of analyses with the appropriate sample mean and covariance:

$$\bar{\mathbf{x}}^a = k^{-1} \sum_{i=1}^{k} \mathbf{x}^{a(i)},$$

$$\mathbf{P}^a = (k-1)^{-1} \sum_{i=1}^{k} (\mathbf{x}^{a(i)} - \bar{\mathbf{x}}^a)(\mathbf{x}^{a(i)} - \bar{\mathbf{x}}^a)^T$$

The challenge: to find an algorithm that provides an accurate $\bar{\mathbf{x}}^a$ estimate of the state and a good representation of $\mathbf{P}^a$ using a relatively small ensemble (e.g. $k \leq 100$)

Potential advantage: a spatio-temporally varying representation of $\mathbf{P}^b$ and $\mathbf{P}^a$ that depends both on the flow and on the observing network

A Remark on the Rank of $\mathbf{P}^b$

The rank of $\mathbf{P}^b$ is equal to the rank of $\mathbf{X}^b$, which is at most $k-1$ because the sum of its columns is $\mathbf{0}$. Thus, the ensemble size limits the rank of the background covariance matrix. This can be both a serious limitation and an advantage:

- **Potential Limitation:** For a k-member ensemble the ensemble-based estimate of $\mathbf{P}^b$ can capture uncertainties in $k-1$ state space direction. The ensemble based estimate can be potentially rank-defficient.

- **Potential Advantage:** Doing the matrix calculation in a $k-1$-dimensional space is computationally extremely cheap compared to doing the calculation in a $10^6 - 10^8$ dimensional space.

Covariance localization: An Efficient Approach to Address Rank Defficiency

The covariance is considered only between state variables at nearby locations. (Formally, most off-diagonal entries of $\mathbf{P}^b$ are replaced with zeros, which restores the high rank of $\mathbf{P}^b$.)

Different functions have been used for localization, e.g., step functions that drop from from 1 to 0 at a given distance, functions that decrease gradually from 1 to 0 with a Gaussian shape.

An additional advantage of the localization is that it decouples the analysis process at the locations that are further apart from each other than the localization radiance. This allows for an efficient implementation on a parallel computer.

Classification of the Schemes based on the approach they obtain the analysis ensemble

- Perturbed-observation schemes
  - k sets of observations are obtained by perturbing the observations by k sets of random observational noise generated according to $\mathbf{R}$
  - The gain matrix $\mathbf{K}$ is applied to each background ensemble member and a set of observations to obtain the associated analysis ensemble member
  - For an infinitely large ensemble, the ensemble of analyses generated such a way would accurately represent $\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})$

- This is the original Evensen (1993) formulation, except that he did not perturbed the observations, which led to the incorrect $\mathbf{P}_n^a = (\mathbf{I}-\mathbf{KH})\mathbf{P}_n^b(\mathbf{I}-\mathbf{KH})^T$

- Square -Root Filters

  - $\mathbf{P}^a$ is calculated from the standard equation $\mathbf{P}_n^a = (\mathbf{I} + \mathbf{P}_n^b\mathbf{H}_n^T\mathbf{R}_n^{-1}\mathbf{H}_n)^{-1}\mathbf{P}_n^b$.

  - An ensemble is generated such that it exactly represent $\mathbf{P}_n^a$. (This involves the calculation of a matrix square-root).

  - Because $\mathbf{P}_n^a$ is exactly represented independently of k, this approach provides much more accurate analyses for small ensemble sizes (for an illustration of this effect see Whitaker and Hamill 2002).

Examples for Perturbed-Observation Schemes and Square-Root Filters

- Perturbed-Observation Schemes (also often called Ensemble Kalman Filter, EnKF): Burgers et al. 1998; Houtekamer and Mitchell 1998 and later papers: papers by Snyder, Zhang, Hakim and different coauthors.

- Square-Root Filters: Ensemble Adjustment Kalman Filter (EnAKF, Anderson 2001); Ensemble Transform Kalman Filter (ETKF, Bishop et al. 2001); Ensemble Square-Root Filter (EnSQR, Whitaker and Hamill 2002); Local Ensemble Kalman Filter (LEKF, Ott et al. 2002 and 2004) and Local Ensemble Transform Kalman Filter (LETKF, Hunt et al. 2007); Maximum Lekelihood Ensemble Filter (MLEF, Zupanski 2005)

Classification of the Schemes based on the processing of the observational information

- Serial Schemes: Assimilate the observations one by one or by batches of correlated observations; all state vector component is updated that may be affected by the given observation (examples are all schemes with the exceptions below)

- Local Schemes: The state is updated independently for each grid point (or for each state vector component) assimilating all observations simultaneously, which may affect the state estimate at the given grid point (LEKF and LETKF)

- Global Schemes: Assimilate all observations without covariance localization (ETKF and MLEF)

Classification of the Schemes based on the analysis update method

All schemes obtain the analysis by an equivalent of $\bar{\mathbf{x}}_n^a = \bar{\mathbf{x}}_n^b + \mathbf{P}_n^a \mathbf{H}_n^T \mathbf{R}_n^{-1} (\mathbf{y}_n^o - \mathbf{H}_n \bar{\mathbf{x}}_n^b)$, except for the MLEF scheme that is based on the variational approach

Which scheme would be the best for my application?

- When carefully implemented the different formulations should be equally accurate and the advantages and disadvantages are primarily associated with the computational efficiency. In particular,
  - Global schemes are not competitive for the affordable ensemble sizes ($k \leq 100$)
  - Square-Root Filters work well for small ensembles ($20 \leq k \leq 60$), which makes them more efficient than the perturbed-observation schemes

- For a large number of observations (e.g., for more than 10^5 observations) on a parallel computer architecture the LETKF is

by far the fastest, though serial schemes allow for an efficient data thinning procedure, which may eliminate most of the advantage of the LETKF when there is a lot of redundancy between the observations

- On a single processor machine serial schemes should be the fastest, since in that case the LETKF cannot take advantage of the fact that it can process the different grid points independently