The Ensemble Kalman Filter and Friends

Dr Sarah Dance

September 2014

University of Reading

Contents

1 **Overview** 1

2 **Introduction** 1

3 **KF derivation** 3

4 **KF properties** 5

5 **EnKF** 6

6 **Perturbed obs EnKF** 9

7 **EnSRF** 11

8 **Statistical properties** 13

9 **References** 13

10 **App: Nonlinear obs** 14

1 Overview

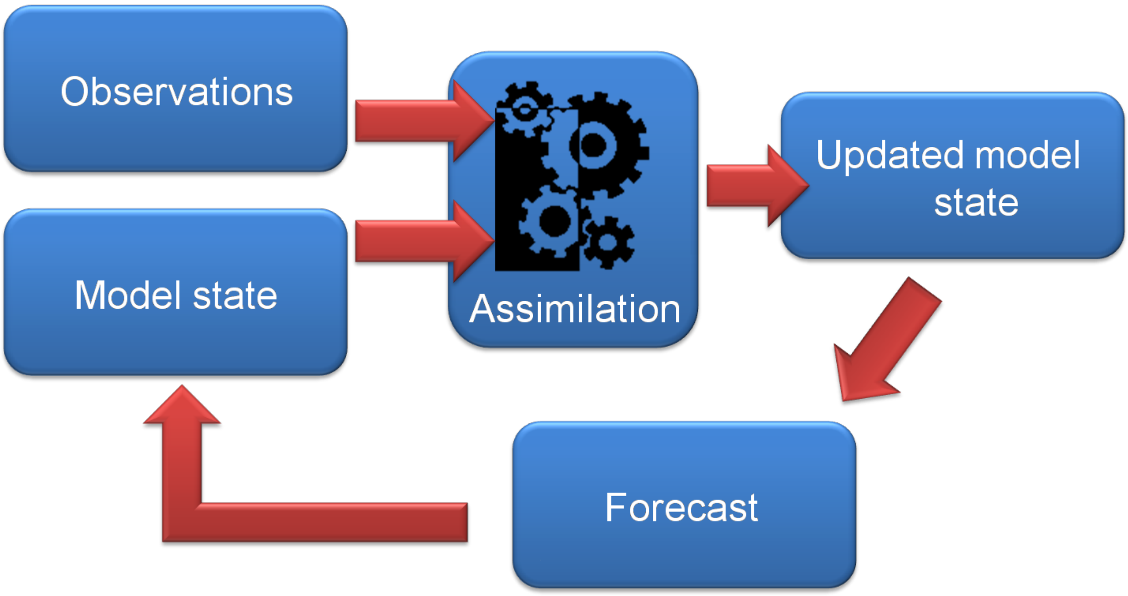
Overview

- Statement of problem and notation
- The Kalman filter
- The Ensemble Kalman filter
- Perturbed observation filters and square root filters

2 Introduction

In this section we will describe the filtering problem and establish the basic notation to be used throughout the lecture.

State estimation feedback loop



The Model

Consider the discrete, linear system,

$$\mathbf{x}_{k+1} = \mathbf{M}_k \mathbf{x}_k + \mathbf{w}_k, \quad k = 0, 1, 2, \ldots, \tag{1}$$

where

- $\mathbf{x}_k \in \mathbb{R}^n$ is the *state vector* at time t_k
- $\mathbf{M}_k \in \mathbb{R}^{n \times n}$ is the *state transition matrix* (mapping from time t_k to t_{k+1}) or *model*
- $\{\mathbf{w}_k \in \mathbb{R}^n; k = 0, 1, 2, \ldots\}$ is a white, Gaussian sequence, with $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, often referred to as *model error*
- $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite covariance matrix (known as the *model error covariance matrix*).

The Observations

We also have discrete, linear observations that satisfy

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad k = 1, 2, 3, \ldots, \tag{2}$$

where

- $\mathbf{y}_k \in \mathbb{R}^p$ is the vector of actual measurements or *observations* at time t_k
- $\mathbf{H}_k \in \mathbb{R}^{p \times n}$ is the *observation operator*. Note that this is not in general a square matrix.
- $\{\mathbf{v}_k \in \mathbb{R}^p; k = 1, 2, \ldots\}$ is a white, Gaussian sequence, with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, often referred to as *observation error*.
- $\mathbf{R}_k \in \mathbb{R}^{p \times p}$ is a symmetric positive definite covariance matrix (known as the *observation error covariance matrix*).

We assume that the initial state, $\mathbf{x}_0$ and the noise vectors at each step, $\{\mathbf{w}_k\}$, $\{\mathbf{v}_k\}$, are assumed mutually independent.

The Prediction and Filtering Problems

We suppose that there is some uncertainty in the initial state, i.e.,

$$\mathbf{x}_0 \sim N(0, \mathbf{P}_0) \tag{3}$$

with $\mathbf{P}_0 \in \mathbb{R}^{n\times n}$ a symmetric positive definite covariance matrix. The problem is now to compute an improved estimate of the stochastic variable $\mathbf{x}_k$, provided $\mathbf{y}_1, \ldots \mathbf{y}_j$ have been measured:

$$\widehat{\mathbf{x}}_{k|j} = \widehat{\mathbf{x}}_{k|y_1,\ldots,y_j}. \tag{4}$$

- When $j = k$ this is called the *filtered estimate*.
- When $j = k - 1$ this is the one-step predicted, or (here) the *predicted estimate*.

3 Derivation of the Kalman Filter

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem. (The Kalman-Bucy filter (Kalman and Bucy, 1961) provides a continuous time analogue).
- We will take a *maximum a posteriori (MAP)* approach to deriving the filter.
- We assume that all the relevant probability densities are Gaussian so that we can simply consider the mean and covariance.
- Rigorous justifcation and other approaches to deriving the filter are discussed by Jazwinski (1970), Chapter 7.

Prediction step

We first derive the equation for one-step prediction of the mean using the state propagation model (1).

$$\begin{aligned}\widehat{\mathbf{x}}_{k+1|k} &= \mathbb{E}\left[\mathbf{x}_{k+1}|\mathbf{y}_1,\ldots\mathbf{y}_k\right],\\ &= \mathbb{E}\left[\mathbf{M}_k\mathbf{x}_k + \mathbf{w}_k\right],\\ &= \mathbf{M}_k\widehat{\mathbf{x}}_{k|k}\end{aligned} \tag{5}$$

The one step prediction of the covariance is defined by,

$$\mathbf{P}_{k+1|k} = \mathbb{E}\left[(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})(\mathbf{x}_{k+1} - \widehat{\mathbf{x}}_{k+1|k})^T|\mathbf{y}_1,\ldots\mathbf{y}_k\right]. \tag{6}$$

Exercise: Using the state propagation model, (1), and one-step prediction of the mean, (5), show that

$$\mathbf{P}_{k+1|k} = \mathbf{M}_k\mathbf{P}_{k|k}\mathbf{M}_k^T + \mathbf{Q}_k. \tag{7}$$

Filtering Step

At the time of an observation we use Bayes theorem (see Ross' lecture) to write down the posterior density as

$$p(\mathbf{x}_{k|k}) = p(\mathbf{x}_k|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_k) = \frac{p(\mathbf{y}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_{k-1})}{p(\mathbf{y}_k|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_{k-1})} \tag{8}$$

Note that, by definition

$$\begin{aligned}p(\mathbf{y}_k|\mathbf{x}_k) &= N(\mathbf{H}_k\mathbf{x}_k, \mathbf{R})\\ &\propto \exp\left\{\frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)\right\}\end{aligned} \tag{9}$$

and

$$\begin{aligned}&p(\mathbf{x}_k|\mathbf{y}_1,\mathbf{y}_2,\ldots\mathbf{y}_{k-1}) = N(\mathbf{x}_{k|k-1}, \mathbf{P}_{k|k-1})\\ &\quad\propto \exp\left\{\frac{1}{2}(\mathbf{x}_k - \mathbf{x}_{k|k-1})^T\mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k - \mathbf{x}_{k|k-1})\right\}.\end{aligned} \tag{10}$$

The denominator $p(\mathbf{y}_k|\mathbf{y}_1, \mathbf{y}_2, \dots \mathbf{y}_{k-1})$ is independent of $\mathbf{x}_k$.

We now find the maximum a posteriori estimate by maximizing $p(\mathbf{x}_k|\mathbf{y}_1, \mathbf{y}_2, \dots \mathbf{y}_k)$ with respect to $\mathbf{x}_k$. This is equivalent to minimizing

$$\begin{aligned} J(\mathbf{x}_k) &= \frac{1}{2}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k)^T\mathbf{R}_k^{-1}(\mathbf{y}_k - \mathbf{H}_k\mathbf{x}_k) \\ &\quad + (\mathbf{x}_k - \mathbf{x}_{k|k-1})^T\mathbf{P}_{k|k-1}^{-1}(\mathbf{x}_k - \mathbf{x}_{k|k-1}) \end{aligned} \tag{11}$$

with respect to $\mathbf{x}_k$.

One can show that this is equal to

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \tag{12}$$

where $\mathbf{K}_k \in \mathbb{R}^{n\times p}$ is known as the *Kalman gain* and is equal to

$$\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1}. \tag{13}$$

- Note that since $\widehat{\mathbf{x}}_{k|k}$ has been derived as a MAP estimate it is equal to the *mode* of the distribution.
- Since the densities involved are Gaussian it is also equal to the *mean* of the distribution - a fact we will use in deriving the posterior covariance.

Posterior Covariance

At the time of an observation, we have seen that the update to the mean may be written as a linear combination of the observation and the previous estimate:

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \tag{14}$$

where $\mathbf{K}_k \in \mathbb{R}^{n\times p}$ is known as the *Kalman gain*. We now consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T|\mathbf{y}_1, \dots \mathbf{y}_k\right]. \tag{15}$$

Using the observation update for the mean (12) we have,

$$\begin{aligned} \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}) \\ &= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \\ &\quad \text{replacing the observations with their model equivalent,} \\ &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1}) - \mathbf{K}_k\mathbf{v}_k. \end{aligned} \tag{16}$$

Thus, since the error in the prior estimate, $\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1}$ is uncorrelated with the measurement noise we find

$$\begin{aligned} \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})^T\right](\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T \\ &\quad + \mathbf{K}_k\mathbb{E}\left[\mathbf{v}_k\mathbf{v}_k^T\right]\mathbf{K}_k^T. \end{aligned} \tag{17}$$

Remark

Using our established notation for the prior and observation error covariances. we obtain the formula

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T. \tag{18}$$

This is sometimes known as the *Joseph form* for the covariance update. It is valid for any value of $\mathbf{K}_k$. If we choose the optimal Kalman gain, it can be simplified further (see below).

Simplification of the a posteriori error covariance formula

Using the value of the Kalman gain we are in a position to simplify the Joseph form as

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{19}$$

Exercise: Show this.

Note that the covariance update equation is independent of the actual measurements: so $\mathbf{P}^{k|k}$ could be computed in advance.

Summary of the Kalman filter

Prediction step

Mean update: $\widehat{\mathbf{x}}_{k+1|k} = \mathbf{M}_k \widehat{\mathbf{x}}_{k|k}$

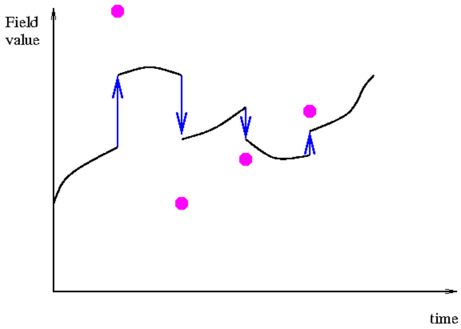
Covariance update: $\mathbf{P}_{k+1|k} = \mathbf{M}_k \mathbf{P}_{k|k} \mathbf{M}_k^T + \mathbf{Q}_k$.

Observation update step

Mean update: $\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \widehat{\mathbf{x}}_{k|k-1})$

Kalman gain: $\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^T(\mathbf{H}_k \mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1}$

Covariance update: $\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k)\mathbf{P}_{k|k-1}$.



Scalar Example

Exercise: Suppose we have a scalar, time-invariant perfect model system such that $\mathbf{M} = 1, \mathbf{w} = 0, \mathbf{Q} = 0, \mathbf{H} = 1, \mathbf{R} = r$. By combining the prediction and filtering steps, show that the following recurrence relations written in terms of prior quantities hold:

$$x_{k+1|k} = (1 - K_k)x_{k|k-1} + K_k y_k \quad (20)$$

$$K_k = \frac{p_{k|k-1}}{p_{k|k-1} + r} \quad (21)$$

$$p_{k+1|k} = \frac{p_{k|k-1} r}{p_{k|k-1} + r}. \quad (22)$$

If we divide the recurrence for $p_{k+1|k}$, (22), by r on each side, and write $\rho_k = p_{k+1|k}/r$ we have

$$\rho_k = \frac{\rho_{k-1}}{\rho_{k-1} + 1}. \quad (23)$$

Solving this nonlinear recurrence we find

$$\rho_k = \frac{\rho_0}{1 + k\rho_0}. \quad (24)$$

So $\rho_k \to 0$ as $k \to \infty$, i.e., over time the error in the state estimate becomes vanishingly small.

4 Kalman Filter properties

Maximum a posteriori and minimum variance

For our derivation we assumed

- Linear model (state propagation) and observation operator
- Gaussian statistics for the uncertainty in the initial state and observations

We found the Kalman filter as a *maximum a posteriori* estimate under these assumptions.

From the Joseph form of the posterior error covariance we can show that choosing the Kalman gain minimizes $\mathrm{trace}(\mathbf{P}_{k|k})$. Hence the Kalman filter is a *minimum variance estimate*.

Filter stability

Theorem 1 (see Jazwinski (1970)). *If the dynamical system,* (1), (2)*, is uniformly completely observable and uniformly completely controllable, then the Kalman filter is uniformly asymptotically stable.*

Observability and controllability

- Observability measures if there is enough observation information. It takes into account the propagation of information with the model.
- Controllability measures if it is possible to nudge the system to the correct solution by applying appropriate increments.
- Uniform asymptotic stability implies that regardless of the initial data $\mathbf{x}_0, \mathbf{P}_0$, with bounded observation errors, the errors in the output will remain bounded.
- Even with an unstable model $\mathbf{M}$, the Kalman filter will stabilize the system!

Filter stability

- We have seen that the Kalman filter is an optimal filter.
- However, optimality does not imply stability.
- The Kalman filter is a stable filter in exact arithmetic
- Stability in exact arithmetic does not imply numerical stability!

Filter divergence

- Despite the nice stability properties of the filter in exact arithmetic, in practice the Kalman filter does suffer from *filter divergence*.
- Filter divergence is often made manifest through overconfidence in the filter prediction ($\mathbf{P}_{k|k}$ too small), with subsequent observations having little effect on the estimate.
- Filter divergence can be causes by inaccurate descriptions of the model (and model error) dynamics, biased observations etc, as well as due to numerical roundoff errors.

5 The Ensemble Kalman filter

We now know something about the Kalman filter! This is only valid for

- Linear state evolution models
- Linear observation operators

For large systems, it is also rather computationally expensive to store and evolve the covariance $\mathbf{P}$ in time since this may contain $n(n+1)/2)$ independent elements.

The Extended Kalman filter

- The *Extended Kalman filter(EKF)* (see, e.g., Grewal and Andrews (2008)) was developed to get around the linearity problem.
- This required tangent linear and adjoint models for the state propagation and observations
- It did not get around the need to store and propagate $\mathbf{P}$
- There are numerous examples in the literature where the EKF fails

The Ensemble Kalman filter

- The *Ensemble Kalman filter (EnKF)* was developed by Evensen (1994)
- The idea is to use an ensemble (statistical sample) of states to represent the evolution of the filtered state pdf.
- Evensen (2009) is a textbook devoted to the Ensemble Kalman filter.
- Review articles on the EnKF include Ehrendorfer (2007); Evensen (2003); Lorenc (2003); Houtekamer and Mitchell (2005).
- Note that there are several different variants of the filter.
- We will start by looking at the perturbed observation filter.
- We may look at some square root forms of the filter.
- We will not be able to cover all variants in the time available.

Notation for the EnKF

- We now set up the notation for the dynamical system we are trying to estimate
- We generalize to a nonlinear state evolution model.
- For now, we stick with a linear observation operator - although we will discuss later how to generalize for nonlinear observations.

The Nonlinear Model
Consider the discrete, nonlinear system,

$$\mathbf{x}_{k+1} = \mathscr{M}(\mathbf{x}_k, k) + \mathbf{w}_k, \quad k = 0, 1, 2, \ldots, \tag{25}$$

where

- $\mathbf{x}_k \in \mathbb{R}^n$ is the *state vector* at time t_k
- $\mathscr{M} : \mathbb{R}^n \times 1 \to \mathbb{R}^n$ is the *nonlinear state transition model* (mapping from time t_k to t_{k+1}) or *model*
- $\{\mathbf{w}_k \in \mathbb{R}^n; k = 0, 1, 2, \ldots\}$ is a white, Gaussian sequence, with $\mathbf{w}_k \sim N(\mathbf{0}, \mathbf{Q}_k)$, often referred to as *model error*
- $\mathbf{Q}_k \in \mathbb{R}^{n \times n}$ is a symmetric positive definite covariance matrix (known as the *model error covariance matrix*).

Remarks
Note that:

- Another common notation in the ensemble Kalman filter literature is to use ψ for the state (e.g., Evensen, 2009).
- Additive model error is not the most general (or indeed realistic) scenario. However since model error is typically unknown it is sensible to treat it in a simple way

The Observations

We also have discrete, linear observations that satisfy

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k, \quad k = 1, 2, 3, \ldots, \tag{26}$$

where

- $\mathbf{y}_k \in \mathbb{R}^p$ is the vector of actual measurements or *observations* at time t_k
- $\mathbf{H}_k \in \mathbb{R}^{p \times n}$ is the *observation operator*. Note that this is not in general a square matrix.
- $\{\mathbf{v}_k \in \mathbb{R}^p; k = 1, 2, \ldots\}$ is a white, Gaussian sequence, with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$, often referred to as *observation error*.
- $\mathbf{R}_k \in \mathbb{R}^{p \times p}$ is a symmetric positive definite covariance matrix (known as the *observation error covariance matrix*).

We assume that the initial state, $\mathbf{x}_0$ and the noise vectors at each step, $\{\mathbf{w}_k\}$, $\{\mathbf{v}_k\}$, are assumed mutually independent.

The ensemble and ensemble mean

Let $\{\mathbf{x}^{(i)}\}$ $(i = 1, \ldots, m)$ be an m-*member ensemble* in an n-dimensional state space.

The *ensemble mean* is the vector defined by

$$\overline{\mathbf{x}} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}^{(i)}. \tag{27}$$

If the members of $\{\mathbf{x}^{(i)}\}$ are drawn independently from the same probability distribution, then $\overline{\mathbf{x}}$ is an unbiased estimator of the population mean (e.g., Barlow, 1989, Chapter 5), although in practice the estimates obtained thereby may be subject to large sampling errors. In fact the expected value of the RMSE for the ensemble mean is $O(m^{-1/2})$.

The ensemble covariance

The ensemble covariance is computed as

$$\mathbf{P} = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}^{(i)} - \overline{\mathbf{x}})(\mathbf{x}^{(i)} - \overline{\mathbf{x}})^T, \tag{28}$$

where choosing the factor $\frac{1}{m-1}$ (rather than $\frac{1}{m}$) ensures that this is an unbiased estimator of the population covariance matrix (e.g., Barlow, 1989, Chapter 5).

The ensemble perturbation matrix

It is convenient to introduce the *ensemble perturbation matrix* as the $n \times m$ matrix defined by

$$\mathbf{X} = \frac{1}{\sqrt{m-1}} \begin{pmatrix} \mathbf{x}^{(1)} - \overline{\mathbf{x}} & \mathbf{x}^{(2)} - \overline{\mathbf{x}} & \ldots & \mathbf{x}^{(m)} - \overline{\mathbf{x}} \end{pmatrix}. \tag{29}$$

Note that this definition incorporates the factor $1/\sqrt{m-1}$, as in e.g., Bishop et al. (2001); Lorenc (2003).

Then the ensemble covariance matrix is the $n \times n$ matrix

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}^{(i)} - \overline{\mathbf{x}})(\mathbf{x}^{(i)} - \overline{\mathbf{x}})^T. \tag{30}$$

- $\mathbf{X}$ is sometimes called a *matrix square root* of $\mathbf{P}$.

- This is inconsistent with the mathematical literature, where a square root of the matrix $\mathbf{P}$ is often defined to be a matrix $\mathbf{X}$ such that $\mathbf{P} = \mathbf{X}^2$ (see, for example,Golub and Van Loan (1996, section 4.2.10)).
- However, the usage is well-established in the engineering literature (as in Gelb (1974, section 8.4)) and is also common in geophysical applications (as in Tippett et al. (2003)).
- If $\mathbf{X}$ is symmetric then the definitions coincide.

Exercise
Consider the ensemble given by

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}.$$

Compute $\overline{\mathbf{x}}$ and $\mathbf{X}$. What rank is the matrix $\mathbf{X}$?

6 The perturbed observation ensemble Kalman filter

Prediction step
We denote forecast quantities with the superscript f and analysis quantities with the superscript a. One-step prediction of the ensemble is accomplished simply using the state propagation model (25).

$$\mathbf{x}_{k+1}^{(i),f} = \mathscr{M}(\mathbf{x}_k^{(i),a}, k) + \mathbf{w}_k, \quad i = 1, 2, \ldots, m; \quad k = 0, 1, 2, \ldots, \tag{31}$$

The one step prediction of the ensemble mean is simply accomplished by taking the mean of the forecast ensemble

$$\overline{\mathbf{x}^f} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}^{(i),f}. \tag{32}$$

Forecast covariance
The one step prediction of the covariance is simply accomplished by taking the covariance of the forecast ensemble,

$$\mathbf{P}^f = \mathbf{X}^f (\mathbf{X}^f)^T = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}^{(i),f} - \overline{\mathbf{x}^f})(\mathbf{x}^{(i),f} - \overline{\mathbf{x}^f})^T. \tag{33}$$

Filtering Step
We would now like to use an ensemble approximation to the Kalman filter to update the ensemble. Recall the Kalman gain

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}^T + \mathbf{R}_k)^{-1}. \tag{34}$$

We propose an ensemble approximation of this matrix by replacing $\mathbf{P}_{k|k-1}$ with the ensemble approximation $\mathbf{P}^f$, so that

$$\mathbf{K}_e = \mathbf{P}^f \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}^f \mathbf{H}^T + \mathbf{R}_k)^{-1}. \tag{35}$$

Then the ensemble update is given by

$$\mathbf{x}_k^{(i),a} = \mathbf{x}_k^{(i),f} + \mathbf{K}_e (\mathbf{y}_k - \mathbf{H} \mathbf{x}_k^{(i),f}). \tag{36}$$

The update for the mean is simply the mean of this analysis ensemble

$$\overline{\mathbf{x}^a} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}^{(i),a}. \tag{37}$$

Exercise: Calculate $\overline{\mathbf{x}^a}$ in terms of $\overline{\mathbf{x}^f}$.

Q: Does this result in the "correct" analysis error covariance matrix?

According to the linear Kalman filter equations we have the relationship

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{38}$$

Does this hold for the ensemble version of the filter?

Exercise

Show that the analysis ensemble covariance

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{K}_e\mathbf{H})\mathbf{P}^f(\mathbf{I} - \mathbf{K}_e\mathbf{H}).$$

Remark: If we compare with the Joseph form of the analysis error covariance (18) we see that this is missing a factor $\mathbf{K}_e\mathbf{R}\mathbf{K}_e$.

Perturbed observation algorithm

To correct the algorithm Burgers et al. (1998) showed that the observations should also be treated as random variables

$$\mathbf{d}_k^{(i)} = \mathbf{y}_k + \mathbf{v}_k^{(i)} \tag{39}$$

where

- $\mathbf{y}_k \in \mathbb{R}^p$ is the vector of actual measurements or *observations* at time t_k
- $\{\mathbf{v}_k^{(i)} \in \mathbb{R}^p; i = 1, \ldots, m\}$ is an ensemble of samples of observation error with $\mathbf{v}_k \sim N(\mathbf{0}, \mathbf{R}_k)$.

The sample covariance is given by

$$\mathbf{R}_e = \frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{d}^{(i),f} - \overline{\mathbf{d}^f})(\mathbf{d}^{(i),f} - \overline{\mathbf{d}^f})^T. \tag{40}$$

Note that due to sampling error it may be the case that $\overline{\mathbf{d}} \neq \mathbf{y}_k$ if $\overline{\mathbf{v}} \neq 0$. In practice, the requirement for zero mean measurement noise is imposed in the simulated sample.

Perturbed observation filter algorithm

Using the perturbed observation errors results in the following EnKF algorithm - here compared alongside the linear Kalman filter:

EnKF	KF
Prediction step	
$\mathbf{x}_{k+1}^{(i),f} = \mathscr{M}(\mathbf{x}_k^{(i),f}, k) + \mathbf{w}_k,$	$\widehat{\mathbf{x}}_{k+1\|k} = \mathbf{M}_k\widehat{\mathbf{x}}_{k\|k}$
$\mathbf{P}^f = \mathbf{X}^f(\mathbf{X}^f)^T$	$\mathbf{P}_{k+1\|k} = \mathbf{M}_k\mathbf{P}_{k\|k}\mathbf{M}_k^T + \mathbf{Q}_k.$
Observation update step	
$\mathbf{x}_k^{(i),a} = \mathbf{x}_k^{(i),f} + \mathbf{K}_e(\mathbf{d}_k^{(i)} - \mathbf{H}\mathbf{x}_k^{(i),f})$	$\widehat{\mathbf{x}}_{k\|k} = \widehat{\mathbf{x}}_{k\|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k\|k-1})$
$\mathbf{K}_e = \mathbf{P}^f\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}^f\mathbf{H}^T + \mathbf{R}_{e,k})^{-1}$	$\mathbf{K}_k = \mathbf{P}_{k\|k-1}\mathbf{H}_k^T(\mathbf{H}_k\mathbf{P}_{k\|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1}$
$\mathbf{P}^a = \mathbf{X}^a(\mathbf{X}^a)^T$	$\mathbf{P}_{k\|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k\|k-1}.$

Note that the matrix $(\mathbf{H}_k\mathbf{P}^f\mathbf{H}^T + \mathbf{R}_{e,k})$ may be singular and pseudo-inversion may be required.

Pseudo-inverse

We note that $\mathbf{D} = (\mathbf{H}_k\mathbf{P}^f\mathbf{H}^T + \mathbf{R}_{e,k})$ is symmetric positive semi-definite so $\mathbf{D} \in \mathbb{R}^{p\times p}$ has an eigenvalue decomposition of the form

$$\mathbf{D} = \mathbf{Z}\Lambda\mathbf{Z}^\mathbf{T},$$

where $\mathbf{Z} \in \mathbb{R}^{p\times p}$ is an orthogonal matrix and $\Lambda \in \mathbb{R}^{p\times p}$ is diagonal of rank r with diagonal elements $\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots, 0$. A common choice of pseudo-inverse would be

$$\mathbf{D}^+ = \mathbf{Z}\Lambda^+\mathbf{Z}^\mathbf{T}$$

where Λ^+ is diagonal of rank r with diagonal elements $\lambda_1^{-1}, \lambda_2^{-1}, \ldots, \lambda_r^{-1}, 0, \ldots, 0$.

Note, for practical implementations, using the *SVD* is usually preferable to the eigenvalue decomposition- for reasons of numerical accuracy.

7 Ensemble square root filtering framework

- Last chapter was all about the perturbed observation ensemble Kalman filter (EnKF).
- This required us to perturb the observations in order to get the "correct" analysis error covariance statistics.
- This has a few undesirable properties:
  - Introduction of additional noise by perturbing the observations
  - "Inversion" of the rank-deficient matrix $\mathbf{H}\mathbf{P}^f\mathbf{H}^T + \mathbf{R}_e$
- Square root forms of the ensemble filter are sometimes called "deterministic" since they avoid the need to stochastically perturb the observations.

The ensemble perturbation matrix

Recall that the *ensemble perturbation matrix* is the $n \times m$ matrix defined by

$$\mathbf{X} = \frac{1}{\sqrt{m-1}} \left(\begin{array}{cccc} \mathbf{x}^{(1)} - \overline{\mathbf{x}} & \mathbf{x}^{(2)} - \overline{\mathbf{x}} & \dots & \mathbf{x}^{(m)} - \overline{\mathbf{x}} \end{array} \right). \tag{41}$$

Note that $\mathbf{X}$ has rank at most $m-1$:

defining $\mathbf{1}_m = (1, 1, \dots, 1)^T \in \mathbb{R}^m$, we see that $\mathbf{X}\mathbf{1}_m = 0$ by construction (Livings et al., 2008; Wang et al., 2004).

Ensemble Covariance

Then the ensemble covariance matrix is the $n \times n$ matrix

$$\mathbf{P} = \mathbf{X}\mathbf{X}^T = \frac{1}{m-1} \sum_{i=1}^{m} (\mathbf{x}^{(i)} - \overline{\mathbf{x}})(\mathbf{x}^{(i)} - \overline{\mathbf{x}})^T. \tag{42}$$

Writing the formulation in terms of $\mathbf{X}$ is what gives the name *square root filter*.

This matrix inherits its rank properties from $\mathbf{X}_k$ and also has rank at most $m-1$.

Prediction step

We denote forecast quantities with the superscript f and analysis quantities with the superscript a. One-step prediction of the ensemble is accomplished simply using the state propagation model (25).

$$\mathbf{x}_{k+1}^{(i),f} = \mathscr{M}(\mathbf{x}_k^{(i),a}, k) + \mathbf{w}_k, \quad i = 1, 2, \dots, m; \quad k = 0, 1, 2, \dots, \tag{43}$$

The one step prediction of the ensemble mean is simply accomplished by taking the mean of the forecast ensemble

$$\overline{\mathbf{x}^f} = \frac{1}{m} \sum_{i=1}^{m} \mathbf{x}^{(i),f}. \tag{44}$$

Forecast-observation ensemble

The ensemble update may be written in terms of a forecast observation ensemble $\{\mathbf{y}_{i,k}^f\}$ defined by

$$\mathbf{y}_{i,k}^f = \mathscr{H}_k(\mathbf{x}_{i,k}^f). \tag{45}$$

Like any other ensemble, the forecast observation ensemble has an ensemble mean $\overline{\mathbf{y}_k^f}$ and an ensemble perturbation matrix $\mathbf{Y}_k^f$.

Ensemble square root relationships

The update for the ensemble mean and perturbation matrix should satisfy (Tippett et al., 2003)

$$\overline{\mathbf{x}_k^a} = \overline{\mathbf{x}_k^f} + \mathbf{K}_k(\mathbf{y}_k - \overline{\mathbf{y}_k^f}), \tag{46}$$
$$\mathbf{X}_k^a = \mathbf{X}_k^f \mathbf{T}_k,, \tag{47}$$

where,

$$\mathbf{K}_k = \mathbf{X}_k^f (\mathbf{Y}_k^f)^T \mathbf{D}_k^{-1}, \tag{48}$$
$$\mathbf{D}_k = \mathbf{Y}_k^f (\mathbf{Y}_k^f)^T + \mathbf{R}_k. \tag{49}$$
$$\tag{50}$$

Notes:

- The $n \times p_k$ matrix $\mathbf{K}_k$ is known as the gain matrix.
- The $p_k \times p_k$ matrix $\mathbf{D}_k$ is positive definite and invertible since $\mathbf{R}_k$ is a positive definite covariance matrix, and the product $\mathbf{Y}_k^f(\mathbf{Y}_k^f)^T$ is positive semi-definite.

The T-Matrix

The matrix $\mathbf{T}_k$ is an $m \times m$, such that

$$\mathbf{T}_k \mathbf{T}_k^T = \mathbf{I} - (\mathbf{Y}_k^f)^T \mathbf{D}_k^{-1} \mathbf{Y}_k^f. \tag{51}$$

This definition of $\mathbf{T}_k$ implies that

$$\mathbf{P}_k^a = (\mathbf{X}_k^f - \mathbf{K}_k \mathbf{Y}_k^f)(\mathbf{X}_k^f)^T. \tag{52}$$

- Equation (51) does not define $\mathbf{T}_k$ uniquely (Tippett et al., 2003), so there are various possible implementations of an SRF, some of which may result in a biased filter (Livings et al., 2008; SAKOV and OKE, 2008).
- However, Livings et al. (2008) showed that any $\mathbf{T}_k$ satisfying (51) is an invertible matrix.

Finally, given $\overline{\mathbf{x}_k^a}$ and $\mathbf{X}_k^a$, the analysis ensemble $\{\mathbf{x}_{i,k}\}$ is obtained as

$$\mathbf{x}_{i,k} = \overline{\mathbf{x}_k^a} + \mathbf{x}'_{i,k} \tag{53}$$

for $i = 1, 2, \ldots, m$, where the column n-vector $\mathbf{x}'_{i,k}$ is the $i-$th column of the $n \times m$ matrix $\mathbf{X}_k^a$.

The preceding discussion has given us a set of relationships that we must satisfy, but it has not given us an algorithm that we can implement.

We need a method to calculate $\mathbf{T}$!

Literature

- Tippett et al. (2003) review several square root filters, places them in a common framework, and compares their numerical efficiency.
- Nerger and Hiller (2013) discuss strategies for parallel implementation.
- The Ensemble Transform Kalman filter of Bishop et al. (2001) and Wang et al. (2004). This is one of the most popular implementations due to its numerical efficiency. Note that Bishop et al. (2001) predates Livings et al. (2008) discovery of the results needed to ensure an unbiased implementation - so don't use this version!
- The EAKF (ensemble adjustment Kalman filter) (Anderson, 2001) and Whitaker and Hamill (2002) are both written in a pre-multiplier form $\mathbf{X}^a = \mathbf{A}\mathbf{X}^f$.
- The DEnKF (deterministic ensemble Kalman filter) by SAKOV and OKE (2008) is used by the Met Office.

8 Statistical properties of the ensemble Kalman filter

Convergence to the Kalman filter for large ensemble sizes

Butala et al. (2009) and Mandel et al. (2011) showed that with linear forecast and observation models, in the limit of large ensemble size the ensemble Kalman filter converges in probability to the Kalman filter.

The proof is beyond the scope of this lecture.

Ensemble spread

We now restrict ourselves to the case of linear $\mathbf{H}$ again and write $\mathbf{P}^a_t$ for the true analysis error covariance.

Theorem 2 (Furrer and Bengtsson (2007), Corollary 2)**.** *Let* $\mathbf{H}^T\mathbf{H} = \mathbf{I}_{n\times n}$ *and* $\mathbf{R} = \sigma^2\mathbf{I}_{p\times p}$. *Then*

$$\mathrm{trace}[\mathbb{E}(\mathbf{P}^a)] < \mathrm{trace}[\mathbf{P}^a_t].$$

Remarks:

- It may seem that that the assumptions are restictive, but it is always possible to define a projected space dynamical system by letting $\widetilde{\mathbf{y}} = \mathbf{R}^{-1/2}\mathbf{y}$ and $\widetilde{\mathbf{x}} = \mathbf{R}^{-1/2}\mathbf{H}\mathbf{x}$ and for this system the assumptions of the corollary are satisfied.
- This result shows that the EnKF yields ensemble members with too little spread. Furrer and Bengtsson (2007) show that this lack of spread is due entirely to the inverse in the sample Kalman gain, and that this holds even when as $\mathbb{E}(\mathbf{P}^f) = \mathbf{P}^f_t$.

The proof of this theorem is beyond the scope of this lecture. Ross' lecture this afternoon will talk about some methods to deal with this problem!

9 References

References

Jeffrey L. Anderson. An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, 129:2884–2903, 2001.

R. J. Barlow. *Statistics: a guide to the use of statistical methods in the physical sciences.* John Wiley & Sons Ltd., Chichester, 1989.

Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar. Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Mon. Wea. Rev.*, 129:420–436, 2001.

Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble Kalman filter. *Mon. Wea. Rev.*, 126: 1719–1724, 1998.

Mark D. Butala, Richard A. Frazin, Yuguo Chen, and Farzad Kamalabadi. Tomographic imaging of dynamic objects with the ensemble kalman filter. *IEEE Trans. on Image Processing*, 2009. To appear.

Martin Ehrendorfer. A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift*, 16:795–818, 2007. doi: 10.1127/0941-2948/2007/0256.

Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99(C5):10143–10162, 1994.

Geir Evensen. The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367, 2003.

Geir Evensen. *Data Assimilation: The Ensemble Kalman Filter*. Springer, 2nd edition edition, 2009 2009.

R. Furrer and T. Bengtsson. Estimation of high-dimensional prior and posterior covariance matrices in Kalman filter variants. *J. Multivariate Analysis*, 98:227–255, 2007.

Arthur Gelb, editor. *Applied Optimal Estimation*. The M.I.T. Press, 1974. 374 pp.

Gene H. Golub and Charles F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996. 694 pp.

M.S. Grewal and A. P. Andrews. *Kalman Filtering:Theory and Practice using MATLAB*. Wiley, New Jersey, 2008.

P. L. Houtekamer and Herschel L. Mitchell. Ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 131 (613):3269–3289, 2005. ISSN 1477-870X. doi: 10.1256/qj.05.135. URL `http://dx.doi.org/10.1256/qj.05.135`.

Andrew H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, 1970. 376 pp.

R.E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):32–45, 1960.

R.E. Kalman and R.S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83:95–108, 1961.

David M Livings, Sarah L Dance, and Nancy K Nichols. Unbiased ensemble square root filters. *Physica D: Nonlinear Phenomena*, 237(8):1021–1028, 2008.

A. C. Lorenc. The potential of the ensemble Kalman filter for NWP, a comparison with 4D-Var. *Q. J. R. Meteorol. Soc.*, 129(595): 3183–3203, 2003.

Jan Mandel, Loren Cobb, and Jonathan Beezley. On the convergence of the ensemble kalman filter. *Applications of Mathematics*, 56: 533–541, 2011. ISSN 0862-7940. URL `http://dx.doi.org/10.1007/s10492-011-0031-2`. 10.1007/s10492-011-0031-2.

Lars Nerger and Wolfgang Hiller. Software for ensemble-based data assimilation systems: Implementation strategies and scalability. *Computers & Geosciences*, 55(0):110 – 118, 2013. ISSN 0098-3004. doi: http://dx.doi.org/10.1016/j.cageo.2012.03.026. URL `http://www.sciencedirect.com/science/article/pii/S0098300412001215`. Ensemble Kalman filter for data assimilation.

PAVEL SAKOV and PETER R. OKE. A deterministic formulation of the ensemble kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371, 2008. ISSN 1600-0870. doi: 10.1111/j.1600-0870.2007.00299.x. URL `http://dx.doi.org/10.1111/j.1600-0870.2007.00299.x`.

Michael K. Tippett, Jeffrey L. Anderson, Craig H. Bishop, Thomas M. Hamill, and Jeffrey S. Whitaker. Ensemble square root filters. *Mon. Wea. Rev.*, 131:1485–1490, 2003.

Xuguang Wang, Craig H. Bishop, and Simon J. Julier. Which is better, an ensemble of positive-negative pairs or a centered spherical simplex ensemble? *Mon. Wea. Rev.*, 132:1590–1605, 2004.

Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble data assimilation without perturbed observations. *Mon. Wea. Rev.*, 130: 1913–1924, 2002.

10 Appendix: Nonlinear observation operators

We largely dealt only with linear observation operators. It is straightforward to generalize to nonlinear observation operators, $\mathscr{H}$. There are two ways that this is typically done:

- State augmentation
- Creation of a "forecast-observation" ensemble. (This is not commonly used terminology but gives us a label to refer to the method!)

State augmentation, Evensen (2003), section 4.5

In state-augmentation we append any nonlinear observations to the state vector at observation time and define a linear observation operator on the augmented space:

$$\widetilde{\mathbf{x}} = \begin{pmatrix} \mathbf{x} \\ \mathscr{H}(\mathbf{x}) \end{pmatrix}; \ \mathbf{H}\widetilde{\mathbf{x}} = \begin{pmatrix} 0 & \mathbf{I} \end{pmatrix} \widetilde{\mathbf{x}} = \mathbf{y}. \tag{54}$$

The observation update equations are now applied to the augmented system.

Forecast-observation ensemble

The other method is to use a forecast-observation ensemble $\{\mathbf{y}_i^f\}$ defined by

$$\mathbf{y}^{(i),f} = H(\mathbf{x}^{(i),f}). \tag{55}$$

Like any other ensemble, the forecast observation ensemble has an ensemble mean

$$\overline{\mathbf{y}^f} = \overline{H(\mathbf{x}^f)}.$$

Note that this is *NOT* equal to $H(\overline{\mathbf{x}^f})$ in general.

The forecast observation ensemble perturbation matrix

$$\mathbf{Y}^f = \frac{1}{\sqrt{m-1}} \begin{pmatrix} \mathbf{y}^{(1),f} - \overline{\mathbf{y}^f} & \mathbf{y}^{(2),f} - \overline{\mathbf{y}^f} & \dots & \mathbf{y}^{(m),f} - \overline{\mathbf{x}^f} \end{pmatrix}.$$

is defined as any other ensemble perturbation matrix.

Exercise: Show that in the special case of a linear observation operator, $\mathbf{H}$, $\overline{\mathbf{y}^f} = \mathbf{H}\overline{\mathbf{x}^f}$ and $\mathbf{Y}^f = \mathbf{H}\mathbf{X}^f$.

We then have the observation update step as

$$\mathbf{x}^{(i),a} = \overline{\mathbf{x}^f} + \mathbf{K}_e(\mathbf{d}^{(i)} - \mathbf{y}^{(i),f}), \quad (56)$$

$$\mathbf{K}_e = \mathbf{X}^f(\mathbf{Y}^f)^T(\mathbf{Y}^f(\mathbf{Y}^f)^T + \mathbf{R}_e)^{-1}. \quad (57)$$