The Kalman Filter

Data Assimilation & Inverse Problems from Weather Forecasting to Neuroscience

Sarah Dance

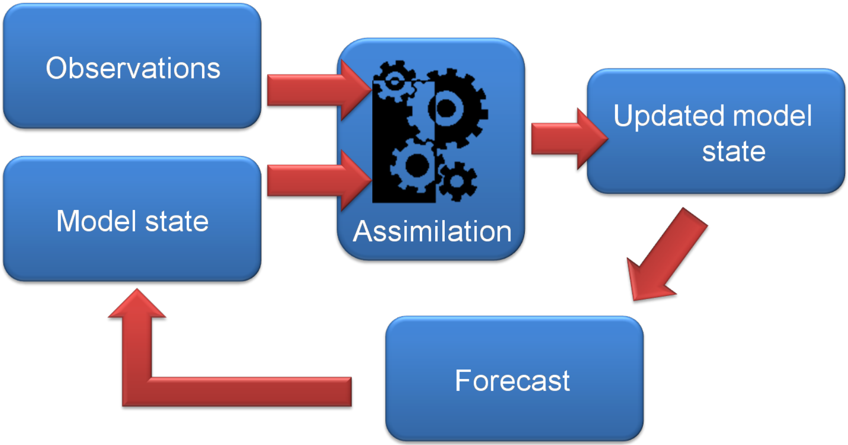
School of Mathematical and Physical Sciences, University of Reading
s.l.dance@reading.ac.uk

July 2013

University of Reading

Outline

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

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

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

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

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 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

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

where

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

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
- $\mathbf{H}_k \in \mathbb{R}^{n \times p}$ is the observation operator. Note that this is not in general a square 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
- $\mathbf{H}_k \in \mathbb{R}^{n \times p}$ 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.

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
- $\mathbf{H}_k \in \mathbb{R}^{n \times p}$ 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).

The Observations

We also have discrete, linear observations that satisfy

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

where

- $\mathbf{y}_k \in \mathbb{R}^p$ is the vector of actual measurements or observations at time t_k
- $\mathbf{H}_k \in \mathbb{R}^{n \times p}$ is the observation operator. Note that this is not in general a square matrix.
- $\{\mathbf{v}_k \in \mathbb{R}^p; k = 1, 2, \dots\}$ 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 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:

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

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.

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.

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem.

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem.
- We will take a minimum variance approach to deriving the filter.

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem.
- We will take a minimum variance 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.

- The Kalman filter (Kalman, 1960) provides estimates for the linear discrete prediction and filtering problem.
- We will take a minimum variance 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).

$$\widehat{\mathbf{x}}_{k+1|k} = \mathbb{E}\left[\mathbf{x}_{k+1}|\mathbf{y}_1, \ldots \mathbf{y}_k\right],$$

Prediction step

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

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

Prediction step

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

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

The one step prediction of the covariance is defined by,

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

The one step prediction of the covariance is defined by,

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

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

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

Filtering Step

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

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \tag{8}$$

where $\mathbf{K}_k \in \mathbb{R}^{n \times p}$ is known as the Kalman gain and will be derived shortly.

Filtering Step

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

$$\widehat{\mathbf{x}}_{k|k} = \widehat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \tag{8}$$

where $\mathbf{K}_k \in \mathbb{R}^{n\times p}$ is known as the Kalman gain and will be derived shortly.

But first we consider the covariance associated with this estimate:

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

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T | \mathbf{y}_1, \dots \mathbf{y}_k\right]. \tag{9}$$

Using the observation update for the mean (8) we have,

$$\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k} \quad =$$

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T | \mathbf{y}_1, \ldots \mathbf{y}_k\right]. \tag{9}$$

Using the observation update for the mean (8) we have,

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

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T | \mathbf{y}_1, \ldots \mathbf{y}_k\right]. \quad (9)$$

Using the observation update for the mean (8) we have,

$$\begin{aligned}
\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}) \\
&= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}),
\end{aligned}$$

replacing the observations with their model equivalent,

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T|\mathbf{y}_1, \ldots \mathbf{y}_k\right]. \tag{9}$$

Using the observation update for the mean (8) we have,

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

But first we consider the covariance associated with this estimate:

$$\mathbf{P}_{k|k} = \mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k})^T | \mathbf{y}_1, \dots \mathbf{y}_k\right]. \qquad (9)$$

Using the observation update for the mean (8) we have,

$$\begin{aligned}
\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k} &= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}) \\
&= \mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1} - \mathbf{K}_k(\mathbf{H}_k\mathbf{x}_k + \mathbf{v}_k - \mathbf{H}_k\widehat{\mathbf{x}}_{k|k-1}), \\
&\quad \text{replacing the observations with their model equivalent,} \\
&= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1}) - \mathbf{K}_k\mathbf{v}_k. \qquad (10)
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbb{E}\left[(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k-1})^T\right](\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T \\
&\quad + \mathbf{K}_k\mathbb{E}\left[\mathbf{v}_k\mathbf{v}_k^T\right]\mathbf{K}_k^T. \qquad (11)
\end{aligned}$$

Remark

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

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T. \qquad (12)$$

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).

To provide a minimum variance estimate we must minimize $\mathrm{trace}(\mathbf{P}_{k|k})$ over all possible values of the gain matrix $\mathbf{K}_k$.

Exercise: Show that

$$\frac{\partial \mathrm{trace}(\mathbf{P}_{k|k})}{\partial \mathbf{K}_k} = -2(\mathbf{H}_k\mathbf{P}_{k|k-1})^T + 2\mathbf{K}_k(\mathbf{H}_k\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k).$$

To provide a minimum variance estimate we must minimize $\mathrm{trace}(\mathbf{P}_{k|k})$ over all possible values of the gain matrix $\mathbf{K}_k$.

Exercise: Show that

$$\frac{\partial \mathrm{trace}(\mathbf{P}_{k|k})}{\partial \mathbf{K}_k} = -2(\mathbf{H}_k \mathbf{P}_{k|k-1})^T + 2\mathbf{K}_k(\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}^T + \mathbf{R}_k). \quad (13)$$

Setting this expression to zero and solving for $\mathbf{K}_k$ yields the Kalman gain

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}^T + \mathbf{R}_k)^{-1}. \quad (14)$$

Simplification of the a posteriori error covariance formula

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

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)^T + \mathbf{K}_k\mathbf{R}_k\mathbf{K}_k^T = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_{k|k-1}. \tag{15}$$

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

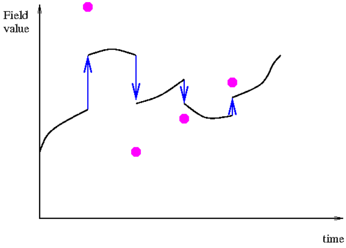
Covariance update: $\mathbf{P}_{k+1|k} = \mathbf{M}_k \mathbf{P}_{k|k} \mathbf{M}_k^T + \mathbf{Q}_k.$

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:

$$x_{k+1|k} = (1 - K_k) x_{k|k-1} + K_k y_k \tag{16}$$

$$K_k = \frac{p_{k|k-1}}{p_{k|k-1} + r} \tag{17}$$

$$p_{k+1|k} = \frac{p_{k|k-1} r}{p_{k|k-1} + r}. \tag{18}$$

If we divide the recurrence for $p_{k+1|k}$, (18), by r on each side, and write $\rho_k = p_{k+1|k}/r$ we have

$$\rho_k = \frac{\rho_{k-1}}{\rho_{k-1} + 1}.$$

If we divide the recurrence for $p_{k+1|k}$, (18), by r on each side, and write $\rho_k = p_{k+1|k}/r$ we have

$$\rho_k = \frac{\rho_{k-1}}{\rho_{k-1} + 1}. \tag{19}$$

Solving this nonlinear recurrence we find

$$\rho_k = \frac{\rho_0}{1 + k\rho_0}.$$

If we divide the recurrence for $p_{k+1|k}$, (18), by r on each side, and write $\rho_k = p_{k+1|k}/r$ we have

$$\rho_k = \frac{\rho_{k-1}}{\rho_{k-1} + 1}. \tag{19}$$

Solving this nonlinear recurrence we find

$$\rho_k = \frac{\rho_0}{1 + k\rho_0}. \tag{20}$$

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

Minimum variance and MMSE

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 minimum variance estimate under these assumptions.

Since the posterior estimation error is $\mathbf{x}_k - \widehat{\mathbf{x}}_{k|k}$, $\mathrm{trace}(\mathbf{P}_{k|k})$ also represents the expected mean square error. Hence our derivation shows the Kalman filter is a minimum mean square error (MMSE) estimate.

BLUE

If we stick with a linear model, but allow for non-Gaussian statistics we find that the Kalman filter provides the Best Linear Unbiased Estimate or linear minimum mean square error (LMMSE) estimate.

Least squares - MAP estimate

Consider the least squares problem:

$$
\begin{aligned}
J_k(\mathbf{x}_0) \quad = \quad & \frac{1}{2}\left(\mathbf{x}_0 - \widehat{\mathbf{x}}_0\right)^T \mathbf{P}_0^{-1} \left(\mathbf{x}_0 - \widehat{\mathbf{x}}_0\right) \\
& + \frac{1}{2}\sum_{l=1}^{k} \left(\mathbf{y}_l - \mathbf{H}_l \mathbf{x}_l\right)^T \mathbf{R}_l^{-1} \left(\mathbf{y}_l - \mathbf{H}_l \mathbf{x}_l\right) \\
& + \frac{1}{2}\sum_{l=0}^{k-1} \mathbf{w}_l^T \mathbf{Q}_l^{-1} \mathbf{w}_l, \qquad (21)
\end{aligned}
$$

subject to the constraint

$$
\mathbf{x}_{l+1} = \mathbf{M}_l \mathbf{x}_l + \mathbf{w}_l, \; l = 0, 1, \ldots, k-1. \qquad (22)
$$

Remarks

- This is the same as the weak constraint 4D-Var cost function

Remarks

- This is the same as the weak constraint 4D-Var cost function
- The minimizing solution to the least squares problem $\mathbf{x}_0^a$ is the maximizer of the conditional probability

$$p(\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_k | \mathbf{y}_1, \ldots \mathbf{y}_k)$$

This is known as the MAP (maximum a posteriori) estimate.

Remarks

- This is the same as the weak constraint 4D-Var cost function
- The minimizing solution to the least squares problem $\mathbf{x}_0^a$ is the maximizer of the conditional probability

$$p(\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_k | \mathbf{y}_1, \ldots \mathbf{y}_k)$$

  This is known as the MAP (maximum a posteriori) estimate.
- The solution of the least squares problem, evolved to the end of the time window, i.e., $\mathbf{x}_k^a$ can be computed exactly as $\widehat{\mathbf{x}}_{k|k}$ in the corresponding Kalman filter problem.

Remarks

- This is the same as the weak constraint 4D-Var cost function
- The minimizing solution to the least squares problem $\mathbf{x}_0^a$ is the maximizer of the conditional probability

$$p(\mathbf{x}_0, \mathbf{x}_1, \ldots \mathbf{x}_k | \mathbf{y}_1, \ldots \mathbf{y}_k)$$

  This is known as the MAP (maximum a posteriori) estimate.
- The solution of the least squares problem, evolved to the end of the time window, i.e., $\mathbf{x}_k^a$ can be computed exactly as $\widehat{\mathbf{x}}_{k|k}$ in the corresponding Kalman filter problem.
- In view of the constraints, we could alternatively minimize the cost function with respect to $\mathbf{x}_k$. This leads to other forms of Kalman filter known as the recursive least squares (RLS) and information filters. This is covered by Simon (2006).

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!

Theorem (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.*

Remarks

- 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 $\mathbf{x}_0, \mathbf{P}_0$, the errors in the Kalman filter solution will decay exponentially fast with time.
- Even with an unstable model $\mathbf{M}$, the Kalman filter will stabilize the system!

- 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.

Compensating for numerical and other errors

- Naive implementations of the Kalman filter can blow-up (see for example Verhaegen and van Dooren (1986) for error analysis).

Compensating for numerical and other errors

- Naive implementations of the Kalman filter can blow-up (see for example Verhaegen and van Dooren (1986) for error analysis).
- This can be improved by implementing a square root or *UDU* form of filter. (This is discussed by Bierman (1977) - more recent books provide short summaries e.g. Simon (2006), chapter 6; Grewal and Andrews (2008), chapter6).

Compensating for numerical and other errors

- Naive implementations of the Kalman filter can blow-up (see for example Verhaegen and van Dooren (1986) for error analysis).
- This can be improved by implementing a square root or *UDU* form of filter. (This is discussed by Bierman (1977) - more recent books provide short summaries e.g. Simon (2006), chapter 6; Grewal and Andrews (2008), chapter6).
- Other errors may be compensated for by increasing the model error - or equivalently using a fading memory version of the filter. This has the effect of reducing the filter's confidence in the model and hence recent observations are given more weight.

Conclusions

In this lecture we have

- Derived the Kalman filter equations
- Seen the optimality of the filter as a minimum variance estimate
- Briefly discussed the stability of the filter in exact arithmetic
- Pointed out the dangers of naive implementation of the Kalman filter

Some important aspects that we didn't cover

- Kalman smoothing
- Nonlinear extensions of the Kalman filter such as the Extended Kalman filter (but see Janjic lecture!)

G.J. Bierman. *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York, 1977.

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

A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

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

D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley, New Jersey, 2006.

M. Verhaegen and P. van Dooren. Numerical aspects of different kalman filter implementations. *IEEE Trans. Auto. Control*, AC-31 (10), 1986.