Lecture 14 Kalman filtering

Tiejun Li

1 Problem setup

To introduce the Kalman filtering, let us consider an interesting problem which is as follows. Suppose we have a state equation

\[ v_n = v_{n-1} + \epsilon_n, \]
\[ x_n = x_{n-1} + v_{n-1} + \frac{1}{2} \epsilon_n. \]

and an observation equation

\[ y_n = x_n + e_n, \]

where \( \epsilon_n \sim N(0, \sigma^2_a I) \), \( e_n \sim N(0, \sigma^2_b I) \). Here \( v \) is the velocity, \( x \) is the real position of the object, and \( y \) is the observed position. The task here is to predict the mean value of \( X_n \) in the future.

There are three typical model problems are of fundamental interest. The first is the continuous-time dynamical model

\[ dX(t) = F(t) \cdot X(t) + B(t) \cdot u(t) + G(t) \cdot dW(t) \]

from which sampled-data measurements are available at times \( t_1, \ldots \), as

\[ z(t_i) = H(t_i)X_{t_i} + v(t_i). \]

The second model of interest is a discret-time dynamics model

\[ X(t_{i+1}) = \Phi(t_{i+1}, t_i)X(t_i) + B_d(t_i)u(t_i) + W_d(t_i) \]

where

\[ B_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, \tau)B(\tau)d\tau, \quad W_d(t_i) = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, \tau)G(\tau)dW(\tau). \]

The third model is a continuous-time dynamics model with continuous measurements as

\[ z(t) = H(t)x(t) + v(t). \]

We assume \( v(t) \) are Gaussian random variables with mean 0 and covariance kernel

\[ E_v^T(t)v(t') = R(t)\delta(t - t'). \]
2 General framework

The above dynamical and observation process can be sketched with the above graphical model, which can also be called a hidden Markov model (HMM). Because of the Markov assumption, the true state is conditionally independent of all earlier states given the immediately previous state.

\[ p(x_k | x_1, \ldots, x_{k-1}) = p(x_k | x_{k-1}) \]

Similarly the measurement at the k-th timestep is dependent only upon the current state and is conditionally independent of all other states given the current state.

\[ p(z_k | x_1, \ldots, x_k) = p(z_k | x_k) \]

Using these assumptions the probability distribution over all states of the hidden Markov model can be written simply as:

\[ p(x_1, \ldots, x_k, z_1, \ldots, z_k) = p(x_1) \left( \prod_{i=1}^{k-1} p(z_i | x_i) p(x_{i+1} | x_i) \right) p(z_k | x_k) \]

However, when we apply the Kalman filter (will be introduced later) to estimate the state \( X \), the probability distribution of interest is that associated with the current states conditioned on the measurements up to the current timestep. (This is achieved by marginalizing out the previous states and dividing by the probability of the measurement set.)

This leads to the predict and update steps of the Kalman filter written probabilistically. The probability distribution associated with the predicted state is the product of the probability distribution associated with the transition from the (k-1)th timestep to the kth and the probability distribution associated with the previous state, with the true state at \((k-1)\) integrated out.

\[ p(x_k | Z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | Z_{1:k-1}) \, dx_{k-1} \]

The measurement set up to time t is

\[ Z_{1:t} = \{ z_1, \ldots, z_t \} \]
The probability distribution of the update is proportional to the product of the measurement likelihood and the predicted state.

\[ p(x_k|Z_{1:k}) = \frac{p(z_k|x_k)p(x_k|Z_{1:k-1})}{p(z_k|Z_{1:k-1})} \]

The denominator

\[ p(z_k|Z_{1:k-1}) = \int p(z_k|x_k)p(x_k|Z_{1:k-1})dx_k \]

is a normalization term.

The remaining probability density functions are

\[ p(x_k|x_{k-1}) = N(\Phi(t_k, t_{k-1})x_{k-1}, G_kG_k^T), \]
\[ p(z_k|x_k) = N(H_kx_k, R_k) \]
\[ p(x_{k-1}|Z_{1:k-1}) = N(\hat{x}_{k-1}, P_{k-1}) \quad \text{(will be proved later)}, \]

Note that the PDF at the previous timestep is inductively assumed to be the estimated state and covariance.

This is justified because, as an optimal estimator, the Kalman filter makes best use of the measurements, therefore the PDF for \( X_k \) given the measurements \( Y_{1:k} \) is the Kalman filter estimate.

3 Optimal prediction and conditional expectation

Lemma 1 (Conditional expectation and optimal prediction). Let \( X \) and \( Y \) are two random variables, then the mean squared error (MSE)

\[ \hat{X}(y) = \arg \min E((X - g(y))^2|Y = y) \]

is minimized at \( \hat{X}(y) = E(X|Y = y) \). Furthermore, the least squares estimate of a random variable \( X \) given another random variable \( Y \) is the conditional expectation \( E(X|Y) \).

Proof. We only need to determine \( \min E(X - c)^2 \) where \( c \) is a constant. We have

\[ E(X - c)^2 = c^2 - 2cEX + EX^2. \]

This minimization is achieved when \( c = EX \). For the other statements, the proof is similar.

4 Estimation with static linear Gaussian system models

Let \( X \) and \( Y \) be jointly Gaussian vectors mapping \( \Omega \) into \( \mathbb{R}^n \) and \( \mathbb{R}^m \), respectively. So the joint pdf can be written as

\[ f_{X,Y}(x, y) = (2\pi)^{-\frac{n+m}{2}} \left| P_{xx} P_{xy} P_{yx} P_{yy} \right|^{-\frac{1}{2}} \exp \left(-\frac{1}{2} \begin{pmatrix} x - m_x \\ \frac{y - m_y} \end{pmatrix}^T \begin{pmatrix} P_{xx} & P_{xy} \\ P_{yx} & P_{yy} \end{pmatrix}^{-1} \begin{pmatrix} x - m_x \\ y - m_y \end{pmatrix} \right). \]

Here we assume the matrix \( P \) is positive definite.
From Bayesian rule and the algebraic manipulations we obtain

\[
f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)} = (2\pi)^{-\frac{d}{2}}|P_{xy}|^{-\frac{1}{2}} \exp \left( -\frac{1}{2}(x-m_{x|y})^TP_{x|y}^{-1}(x-m_{x|y}) \right),
\]
where

\[
m_{x|y} = m_x + P_{xy}P_{yy}^{-1}(y-m_y)
\]

\[
P_{x|y} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}
\]

Thus \(X|Y\) is also Gaussian.

The conditional mean and covariance are

\[
E_x(X|Y = y) = m_{x|y} = m_x + P_{xy}P_{yy}^{-1}(y-m_y)
\]
\[
E_x((X - m_{x|y})(X - m_{x|y})^T|Y = y) = P_{x|y} = P_{xx} - P_{xy}P_{yy}^{-1}P_{yx}.
\]

**Example 1** (Estimation of a 1D static object). Suppose \(x\) is the 1D position, \(z\) is the location measured by means of the model as

\[z(t_1) = x + v_1, \quad z(t_2) = x + v_2.\]

We assume that we had no a priori information about \(z\). \(v_1\) and \(v_2\) could be modeled as zero-mean Gaussian random variables with variances \(\sigma^2_1\) and \(\sigma^2_2\), respectively.

One means of solving for the optimal estimate of position is based on the Bayesian viewpoint. From the first observation we have

\[
f_{X|Z(t_1)}(x|z_1) = (2\pi)^{-\frac{d}{2}} \exp \left( -\frac{(x-z_1)^2}{2\sigma^2_1} \right)
\]

Since the second observation is independent of the first one, then we have

\[
f_{X|Z(t_1),Z(t_2)}(x|z_1, z_2) \propto (2\pi)^{-\frac{d}{2}} \exp \left( -\frac{(x-z_1)^2}{2\sigma^2_1} \right) \cdot (2\pi)^{-\frac{d}{2}} \exp \left( -\frac{(x-z_2)^2}{2\sigma^2_2} \right)
\]
\[
\propto \exp \left( -\frac{1}{2} \left( \frac{1}{\sigma^2_1} + \frac{1}{\sigma^2_2} \right) x^2 - 2 \left( \frac{z_1}{\sigma^2_1} + \frac{z_2}{\sigma^2_2} \right) x + \left( \frac{z_1^2}{\sigma^2_1} + \frac{z_2^2}{\sigma^2_2} \right) \right).
\]

This means the conditional mean of \(X\) after these two observations is

\[
\hat{X} = \frac{z_1}{\frac{1}{\sigma^2_1} + \frac{1}{\sigma^2_2}} = z_1 \frac{\sigma^2_2}{\sigma^2_1 + \sigma^2_2} + z_2 \frac{\sigma^2_1}{\sigma^2_1 + \sigma^2_2} = z_1 + \sigma^2_1(\sigma^2_2 + \sigma^2_1)^{-1}(z_2 - z_1)
\]

and the conditional variances

\[
P = \frac{1}{\frac{1}{\sigma^2_1} + \frac{1}{\sigma^2_2}} = \frac{\sigma^2_1\sigma^2_2}{\sigma^2_1 + \sigma^2_2} = \sigma^2_1 - \sigma^2_2(\sigma^2_2 + \sigma^2_1)^{-1}\sigma^2_1.
\]

**Example 2** (Estimation of a static object: general case). The problem may be stated as follows:

- The variables to be estimated is the \(n\)-dimensional vector \(x\). The true values of it will remain constant, but we do not know the exact values.

- There will be \(m\) measurements available, and these will be made the components of an \(m\)-dimensional vector \(z\).
• \( z = Hx + v \), where \( H \) is a \( m \times n \) matrix, and \( v \) is assumed as a Gaussian random vector with mean \( 0 \) and covariance \( R \).

• Our priori knowledge about \( X \) is a Gaussian random vector with mean \( \hat{x}^- \) and covariance \( P^- \).

The problem is to optimally estimate \( X \) based on the observation \( z \).

To solve this problem, we need to describe the a posteriori distribution \( f_{X|Z}(x|z) \). We first consider the joint distribution of \( x \) and \( v \). It is evident that

\[
 f_{X,V}(x,v) = f_X(x) f_V(v)
\]

by the independence. Define

\[
 U = \begin{pmatrix} X \\ V \end{pmatrix}
\]

Then the mean and covariance of \( U \) are

\[
 m_U = \begin{pmatrix} \hat{x}^- \\ 0 \end{pmatrix}, \quad P_{UU} = \begin{pmatrix} P^- & 0 \\ 0 & R \end{pmatrix}.
\]

For the linear combination of \( U \) we define

\[
 W = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} X \\ V \end{pmatrix} = \begin{pmatrix} X \\ HX + V \end{pmatrix} = \begin{pmatrix} X \\ Z \end{pmatrix},
\]

which is Gaussian and it has the mean and covariance

\[
 m_W = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} \hat{x}^- \\ 0 \end{pmatrix} = \begin{pmatrix} \hat{x}^- \\ H\hat{x}^- \end{pmatrix},
\]

\[
 P_{WW} = \begin{pmatrix} I & 0 \\ H & I \end{pmatrix} \begin{pmatrix} P^- & 0 \\ 0 & R \end{pmatrix} \begin{pmatrix} I & H^T \\ 0 & I \end{pmatrix} = \begin{pmatrix} P^- & P^-H^T \\ HP^- & HP^-H^T + R \end{pmatrix}.
\]

From the obtained result in this section, we have the conditional mean and covariance

\[
 \hat{x}^+ = \mathbb{E}(X|Z = z) = \hat{x}^- + (P^-H^T)(HP^-H^T + R)^{-1}(z - H\hat{x}^-)
\]

\[
 P^+ = \mathbb{E}((X - \hat{x}^+)(X - \hat{x}^+)^T|Z = z) = P^- + (P^-H^T)(HP^-H^T + R)^{-1}(HP^-).
\]

If we define the matrix (which will be called the “Kalman gain matrix” later)

\[
 K = (P^-H^T)(HP^-H^T + R)^{-1}
\]

then

\[
 \hat{x}^+ = \hat{x}^- + K(z - H\hat{x}^-), \quad P^+ = P^- - K(HP^-).
\]

5
5 Kalman filtering

Now we consider the dynamical optimal estimation. Suppose we have a linear stochastic differential equations:

\[ dX(t) = F(t) \cdot X(t) + G(t) \cdot dW(t) \]

where \( W \) is the \( s \)-dimensional canonical Brownian motion, \( F(t) \) and \( G \) are \( n \)-by-\( n \) and \( n \)-by-\( s \) matrices, respectively. To give an “explicit” solution of the above system, we define the biparameter-semigroup \( \Phi(t, \tau) \) which satisfies

\[ \frac{d\Phi(t, \tau)}{dt} = F(t)\Phi(t, \tau) \]

and \( \Phi(t_0, t_0) = I \). Then

\[ X(t) = \Phi(t, t_0)X(t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)dW(\tau). \]

It is straightforward to check that

\[ \mathbb{E}X(t) = \Phi(t, t_0)\mathbb{E}X(t_0) \]

and

\[ \mathbb{E}X(t)X^T(t) = \Phi(t, t_0)\mathbb{E}X^T(t_0)X(t_0)\Phi^T(t, t_0) + \int_{t_0}^{t} \Phi(t, \tau)G(\tau)G^T(\tau)\Phi^T(t, \tau)d\tau. \]

5.1 Discrete-time optimal estimator

We will consider the discrete model in this subsection. We are going to consider two measurement times, \( t_{i-1} \) and \( t_i \), and will propagate optimal estimates from the point just after the measurement at time \( t_{i-1} \) has been incorporated into the estimate, to the point just after the measurement at time \( t_i \) is incorporated. This is named as propagating from time \( t_{i-1}^{+} \) to time \( t_i^{+} \).

Before the estimates, we define the composite vector which comprises the entire measurement history to the current time, and denote it as \( Z(t_i) \), where

\[ Z(t_i) = (z(t_1), \ldots, z(t_i)). \]

This is a vector with growing dimension. Now suppose we are at time \( t_{i-1} \) and have just taken and processed the measurement \( z(t_{i-1}) = z_{i-1} \). From a Bayesian point of view, we are really interested in the probability density of \( x(t_i) \) conditioned on the entire measurement history to that time,

\[ f_{x(t_i)|Z(t_i)}(x_{i-1}|Z_{i-1}), \]

and how this density can be propagated forward through the next measurement time to generate \( f_{x(t_i)|Z(t_i)}(x_i|Z_i) \).

To start the derivation, we will assume that \( f_{x(t_i)|Z(t_i)}(x_{i-1}|Z_{i-1}) \) is a Gaussian conditional density:

\[ f_{x(t_i)|Z(t_i)}(x_{i-1}|Z_{i-1}) = (2\pi)^{-\frac{s}{2}} |P(t_{i-1}^{+})|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x_{i-1} - \hat{x}(t_{i-1}^{+}))^T P^{-1}(t_{i-1}^{+})(x_{i-1} - \hat{x}(t_{i-1}^{+})) \right] \]
where we define $\hat{x}(t_i^-)$ and $P(t_i^-)$ as the conditional mean and conditional covariance, respectively:

$$\hat{x}(t_i^-) = E(X(t_i^-)|Z(t_i^-) = Z_{t_i^-})$$

$$P(t_i^-) = E((X(t_i^-) - \hat{x}(t_i^-))(X(t_i^-) - \hat{x}(t_i^-))^T|Z(t_i^-) = Z_{t_i^-})$$

Now we want to propagate the conditional density and associated estimate from $t_i^-$ just after incorporating the measurement $Z(t_i^-) = z_{t_i^-}$, to time $t_i^+$ just after incorporating $z_i$. Let us decompose this into two steps:

- **Step 1:** a time propagation from $t_i^-$ to $t_i^+$, a time $t_i$ just before the measurement $z_i$ is incorporated;
- **Step 2:** a measurement update from $t_i^+$ to $t_i^+$.

To make the derivation simpler, we will at first neglect the deterministic mean equations of the algorithm: these known inputs have no effect on the spread of density functions, only on their location.

**Step 1:** **First consider the time propagation from $t_i^-$ to $t_i^+$.** From Bayesian point of view, we want to establish the conditional density $f_{x(t_i^-)|z(t_i^-)}(x_i|Z_{t_i^-})$ of the state at time $t_i$, conditioned on the measurement history up through the previous sample time $t_{i-1}$. We will prove that this density is Gaussian.

For the discrete-time model, we have

$$X(t_i) = \Phi(t_i, t_{i-1})X(t_{i-1}) + W_d(t_{i-1})$$

To obtain the density $f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-})$, we observe that $X(t_i)$ is a linear combination of $X(t_{i-1})$ and $W_d(t_{i-1})$, and we have

$$f_{x(t_i)|w(t_i), z(t_i^-)}(x_i, w_i^-|Z_{t_i^-}) = \frac{f_{x(t_i), w_d(t_{i-1}), z(t_i^-)}(x_i, w_i^-|Z_{t_i^-})}{f_Z(t_{i-1})(Z_{t_i^-})} = \frac{f_{x(t_i), z(t_i^-)}(x_i, Z_{t_i^-})}{f_Z(t_{i-1})(Z_{t_i^-})} \cdot f_{W_d(t_{i-1})}(w_i^-) = f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-}) f_{W_d(t_{i-1})}(w_i^-)$$

The density $f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-})$ has been assumed to be Gaussian, and $f_{W_d(t_{i-1})}(w_i^-)$ is Gaussian according to the stochastic integral. So we obtain $f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-})$ is Gaussian.

To specify the density completely, its mean and covariance should be computed. We have

$$E(X(t_i)|Z(t_i^-) = Z_{t_i^-}) = E(\Phi(t_i, t_{i-1})X(t_{i-1}) + W_d(t_{i-1})|Z(t_i^-) = Z_{t_i^-}) = \Phi(t_i, t_{i-1})E(X(t_{i-1})|Z(t_i^-) = Z_{t_i^-})$$

Nowe we define $\hat{x}(t_i^-)$ and $P(t_i^-)$ as the conditional mean and covariance of $x(t_i)$ before the measurement $z(t_i) = z_i$ is taken and processed, $\hat{x}(t_i^+)$ and $P(t_i^+)$ as the conditional mean and covariance of $x(t_i)$ after the measurement $z(t_i) = z_i$ is taken and processed. We have

$$\hat{x}(t_i^-) = E(X(t_i)|Z(t_i^-) = Z_{t_i^-}) = \Phi(t_i, t_{i-1})\hat{x}(t_{i-1}^-)$$

$$P(t_i^-) = \frac{f_{x(t_i), z(t_i^-)}(x_i, Z_{t_i^-})}{f_Z(t_{i-1})(Z_{t_i^-})} \cdot f_{W_d(t_{i-1})}(w_i^-) = f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-}) f_{W_d(t_{i-1})}(w_i^-)$$

$$E(X(t_i)|Z(t_i^-) = Z_{t_i^-}) = \Phi(t_i, t_{i-1})E(X(t_{i-1})|Z(t_i^-) = Z_{t_i^-})$$

$$P(t_i^-) = \frac{f_{x(t_i), z(t_i^-)}(x_i, Z_{t_i^-})}{f_Z(t_{i-1})(Z_{t_i^-})} \cdot f_{W_d(t_{i-1})}(w_i^-) = f_{x(t_i)|z(t_i^-)}(x_i|Z_{t_i^-}) f_{W_d(t_{i-1})}(w_i^-)$$

7
and

\[ P(t_i^-) = \mathbb{E} \left( (X(t_i) - \hat{x}(t_i^-))(X(t_i) - \hat{x}(t_i^-))^T \big| Z(t_i) = Z_{t_i^-} \right) = \Phi(t_i, t_{i-1}) P(t_{i-1}^-) \Phi^T(t_i, t_{i-1}) + \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau) G(\tau) G^T(\tau) \Phi^T(t_i, \tau) d\tau \]

Summarizing the result obtained above, we have the density function

\[ f_{x(t_i) \mid Z(t_{i-1})}(x_i \mid Z_{t_i^-}) = (2\pi)^{-\frac{d}{2}} |P(t_i^-)|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x_i - \hat{x}(t_i^-))^T P^{-1}(t_i^-) (x_i - \hat{x}(t_i^-)) \right] . \]

**Step 2:** Now we want to consider the incorporating the measurement that becomes available at time \( t_i \) so as to generate the density \( f_{x(t_i) \mid Z(t_i) \mid Z_{t_i^-}}(x_i \mid Z_i) \).

From Bayesian rule we have

\[ f_{x(t_i) \mid Z(t_i) \mid Z_{t_i^-}}(x_i \mid Z_i) = \frac{f_{x(t_i) \mid Z(t_i) \mid Z_{t_i^-}}(x_i \mid Z_t)}{f_{Z(t_i) \mid Z_{t_i^-}}(Z_t)} \]

\[ = \frac{f_{x(t_i) \mid Z(t_i), Z(t_{i-1})}(x_i \mid Z_t)}{f_{Z(t_i) \mid Z_{t_i^-}}(Z_t)} \]

\[ = \frac{f_{x(t_i) \mid Z(t_i), Z(t_{i-1})} f_{Z(t_i)}(Z(t_i) \mid Z_{t_i^-})}{f_{Z(t_i)}(Z(t_i) \mid Z_{t_i^-}) f_{Z(t_i)}(Z(t_i) \mid Z_{t_i^-})} \]

\[ = \frac{f_{x(t_i) \mid Z(t_i), Z(t_{i-1})} f_{Z(t_i)}(Z(t_i) \mid Z_{t_i^-})}{f_{Z(t_i)}(Z(t_i) \mid Z_{t_i^-})} \]

The density \( f_{x(t_i) \mid Z(t_{i-1})} \) has been known from the first step. Now let us consider \( f_{x(t_i) \mid Z(t_i), Z(t_{i-1})} \).

From the observation equation

\[ z(t_i) = H(t_i) x_i + v(t_i) \]

we know \( f_{z(t_i) \mid x(t_i), Z(t_{i-1})} \) is still a Gaussian. We can obtain its mean and covariance as follows.

\[ \mathbb{E}(z(t_i) \mid x_i, Z(t_{i-1}) = Z_{t_i^-}) = \mathbb{E}(H(t_i) x_i + v(t_i) \mid x_i) = x_i, Z(t_{i-1}) = Z_{t_i^-}) = H(t_i) x_i, \]

\[ \mathbb{E}((z(t_i) - H(t_i) x_i) (z(t_i) - H(t_i) x_i)^T \mid x_i) = R(t_i). \]

Thus we have

\[ f_{z(t_i) \mid x(t_i), Z(t_{i-1})}(z_i \mid x_i, Z_{t_i^-}) = (2\pi)^{-\frac{d}{2}} |R(t_i)|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (z_i - H(t_i) x_i)^T R^{-1}(t_i) (z_i - H(t_i) x_i) \right] . \]

Now let us consider \( f_{z(t_i) \mid Z(t_{i-1})} \). To show it is Gaussian we first note that \( z(t_i) \) is a linear combination of \( x(t_i) \) and \( v(t_i) \). If we can show that, conditioned on \( Z(t_{i-1}) \), \( x(t_i) \) and \( v(t_i) \) are jointly Gaussian, then we achieve the goal. We have

\[ f_{x(t_i), v(t_i) \mid Z(t_{i-1})}(x_i, v_i \mid Z_{t_i^-}) = f_{v(t_i) \mid x(t_i), Z(t_{i-1})}(v_i \mid x_i, Z_{t_i^-}) f_{x(t_i) \mid Z(t_{i-1})}(x_i \mid Z_{t_i^-}) \]

\[ = f_{v(t_i)} f_{x(t_i) \mid Z(t_{i-1})}(x_i \mid Z_{t_i^-}) \]
Since both densities on the right hand side of the equation above are Gaussian, \( f_{x(t_i), u(t_i)}(z(t_{i-1})) \) is Gaussian, too. Furthermore, we can obtain the mean and covariance as

\[
E(z(t_i) | Z(t_{i-1}) = Z_{i-1}) = H(t_i) \hat{x}(t_i^-)
\]

and

\[
E \left[ (z(t_i) - H(t_i) \hat{x}(t_i^-))^T (z(t_i) - H(t_i) \hat{x}(t_i^-)) \right] | Z(t_{i-1}) = Z_{i-1} = H(t_i) P(t_i^-) H^T(t_i) + R(t_i).
\]

So we can write

\[
f_{x(t_i) | Z(t_{i-1})}(x_i | Z_i) = (2\pi)^{-\frac{d}{2}} |P(t_i^-)|^\frac{1}{2} \exp \left\{ -\frac{1}{2} (x_i - H(t_i) \hat{x}(t_i^-))^T (H(t_i) P(t_i^-) H^T(t_i) + R(t_i))^{-1} (x_i - H(t_i) \hat{x}(t_i^-)) \right\} \cdot
\]

Combing all of the results above, we have

\[
f_{x(t_i) | Z(t_{i-1})}(x_i | Z_i) = \frac{|H(t_i) P(t_i^-) H^T(t_i) + R(t_i)|^\frac{1}{2}}{(2\pi)^{\frac{d}{2}} |P(t_i^-)|^\frac{1}{2}} \exp \left\{ -\frac{1}{2} K \right\}
\]

where

\[
K = (x_i - \hat{x}(t_i^-))^T P^{-1}(t_i^-) (x_i - \hat{x}(t_i^-)) + (z_i - H(t_i) x_i)^T R^{-1}(t_i)(z_i - H(t_i) x_i)
\]

\[
- (z_i - H(t_i) \hat{x}(t_i^-))^T (H(t_i) P(t_i^-) H^T(t_i) + R(t_i))^{-1} (z_i - H(t_i) \hat{x}(t_i^-))
\]

It is not immediately evident that the above density is in fact of Gaussian form. But we will show this in what follows. For convenience, we will omit the time notation. Expand \( K \) we obtain

\[
K = x_i^T [(P^{-1})^{-1} + H^T R^{-1} H] x_i - 2 x_i^T [((P^{-1})^{-1} \hat{x}^- + H^T R^{-1} z_i]
\]

\[
+ z_i^T [R^{-1} - (H P H^T + R)^{-1}] z_i + 2 \hat{x}^- T H^T [H P H^T + R]^{-1} z_i
\]

\[
+ \hat{x}^- T [(P^{-1})^{-1} - H^T (H P H^T + R)^{-1} H] \hat{x}^-
\]

**Lemma 2.** Suppose the matrices \( P \) and \( R \) are both positive definite, then

\[
(P^{-1} + H^T R^{-1} H)^{-1} = P - P H^T (H P H^T + R)^{-1} H P.
\]

And it is straightforward to check the following two identities based on the above result:

\[
(P^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} = P H^T (H P H^T + R)^{-1},
\]

\[
H (P^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} = R - R (H P H^T + R)^{-1} R.
\]

From the above equations, we have

\[
(P^{-1})^{-1}(P^{-1} + H^T R^{-1} H)^{-1} (P^{-1})^{-1} = (P^{-1})^{-1} - H^T (H P H^T + R)^{-1} H,
\]

\[
(P^{-1})^{-1}((P^{-1})^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} = H^T (H P H^T + R)^{-1},
\]
\[ R^{-1} H ((P^{-1} + H^T R^{-1} H)^{-1} H^T R^{-1} = R^{-1} - (HPH^T + R)^{-1}. \]

Substituting the results into \( K \), we obtain
\[
K = x_i^T [(P^{-1} + H^T R^{-1} H) x_i - 2x_i^T [(P^{-1})^{-1} \hat{x}^- + H^T R z_i]
+ [z_i^T R^{-1} H + x_i^{-T} (P^{-1})] [(P^{-1})^{-1} + H^T R^{-1} H]^{-1} [H^T R^{-1} z_i + (P^{-1})^{-1} \hat{x}^-].
\]

To simplify the algebra, we define the \( n \)-vector \( a \) and \( n \)-by-\( n \) matrix \( A \) as
\[
a = (P^{-1})^{-1} \hat{x}^- + H^T R^{-1} z_i, \quad A = (P^{-1})^{-1} + H^T R^{-1} H.
\]

With this notation we have
\[
K = x_i^T A x_i + 2x_i^T a + a^T A^{-1} a = (x_i - A^{-1} a)^T A (x_i - A^{-1} a).
\]

This in fact tells us that the conditional density \( f_{x(t_i)|Z(t_i)}(x_i|Z_i) \) is indeed Gaussian with mean \( A^{-1} a \) and covariance \( A^{-1} \).

Now take advantage of the previous notation we obtain
\[
\hat{x}(t_i^+) = \mathbb{E}(X(t_i)|Z(t_i)) = Z_i = A^{-1} a
\]
and
\[
P(t_i^+) = \mathbb{E}((X(t_i) - \hat{x}(t_i^+))(X(t_i) - \hat{x}(t_i^+))^T|Z(t_i) = Z_i) = A^{-1}
\]
(1)

So we get the conditional density
\[
f_{x(t_i)|Z(t_i)}(x_i|Z_i) = (2\pi)^{-\frac{n}{2}} |P(t_i^+)|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x_i - \hat{x}(t_i^+))^T P(t_i^+)^{-1} (x_i - \hat{x}(t_i^+)) \right].
\]

Substituting the results in Lemma 1 into the conditional mean and covariance we obtain
\[
\hat{x}(t_i^+) = [(P^{-1})^{-1} + H^T R^{-1} H]^{-1} [(P^{-1})^{-1} \hat{x}^- + H^T R^{-1} z_i]
= [P^- - P^- H^T (HPH^T + R)^{-1} H P^-] [(P^-)^{-1} \hat{x}^- + H^T R^{-1} z_i]
= [P^- - P^- H^T (HPH^T + R)^{-1} H P^-] (P^-)^{-1} \hat{x}^- + [P^- H^T (HPH^T + R)^{-1}] z_i
\]

Thus we have
\[
\hat{x}(t_i^+) = \hat{x}(t_i^-) + P(t_i^-) H^T (t_i) (H(t_i) P(t_i^-) H^T(t_i) + R(t_i))^{-1} (z_i - H(t_i) \hat{x}(t_i^-)),
\]
\[
P(t_i^+) = P(t_i^-) - P(t_i^-) H^T (t_i) (H(t_i) P(t_i^-) H^T (t_i) + R(t_i))^{-1} H(t_i) P(t_i^-).
\]

To complete the derivation, let us add the effects of deterministic control inputs. As derived previously, the only change in the estimator is that the state estimate time propagation relations
\[
\hat{x}(t_i^-) = \Phi(t_i, t_{i-1}) \hat{x}(t_{i-1}^-) + \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau) B(\tau) u(\tau) d\tau,
\]
Two sufficient conditions would be if $P$ reason, consider the initial time

Continuous-time filter: Kalman-Bucy filter

In the continuous form we have the problem set up

$$dX(t) = F(t)X(t)dt + B(t)u(t)dt + G(t)dW(t)$$
\[ z_c(t) = H(t)x(t) + v_c(t) \]

where \( v(t) \) is a zero-mean Gaussian white noise with
\[
\mathbb{E}(v_c(t)v_T(t')) = R_c(t)\delta(t - t').
\]

To derive the desired result, we will consider a discrete-time measurement process and examine the result of letting the time stepsize decrease to zero. Thus the measurements are described by
\[ z(t_i) = H(t_i)x(t_i) + v(t_i) \]

where \( v \) is a zero mean Gaussian sequence with
\[
\mathbb{E}(v(t_i)v^T(t_j)) = R_c(t_i)/\Delta t_i, \quad \mathbb{E}(v(t_i)v^T(t_j)) = 0, \ i \neq j.
\]

Without real loss of generality, let \( \Delta t_1 = \Delta t \).

From the former results
\[
\hat{x}(t^-) = \Phi(t_i, t_{i-1})\hat{x}(t_{i-1}^+) + \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau)B(\tau)u(\tau)d\tau,
\]
\[
P(t^-) = \Phi(t_i, t_{i-1})P(t_{i-1}^+)\Phi^T(t_i, t_{i-1}) + \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau)G(\tau)G^T(\tau)\Phi^T(t_i, \tau)d\tau
\]
and
\[
K(t_i) = P(t_i^-)H^T(t_i)(H(t_i)P(t_i^-)H^T(t_i) + R(t_i))^{-1}
\]
\[
= P(t_i^-)H^T(t_i)(H(t_i)P(t_i^-)H^T(t_i)\Delta t + R_c(t_i))^{-1} \Delta t.
\]

So
\[
\hat{x}(t^+) = \hat{x}(t^-) + K(t_i) [z_i - H(t_i)\hat{x}(t^-)]
\]
\[
= \Phi(t_i, t_{i-1})\hat{x}(t_{i-1}^+) + \int_{t_{i-1}}^{t_i} \Phi(t_i, \tau)B(\tau)u(\tau)d\tau + K(t_i) [z_i - H(t_i)\hat{x}(t^-)]
\]

As \( \Delta t \) goes to zero, we obtain the continuous equation
\[
\frac{d\hat{x}(t)}{dt} = F(t)\hat{x}(t) + B(t)u(t) + P(t)H^T(t)R_c^{-1}(t)(z(t) - H(t)\hat{x}(t))
\]
where in the limit, \( \hat{x}(t_{i-1}^+) \to \hat{x}(t^-) \to \hat{x}(t^+) = \hat{x}(t) \), \( P(t_{i-1}^+) \to P(t_i^-) \to P(t_i^+) = P(t) \). The equation for \( P(t) \) reads
\[
\frac{dP(t)}{dt} = F(t)P(t) + P(t)F^T(t) + G(t)G^T(t) - P(t)H^T(t)R_c^{-1}(t)H(t)P(t)
\]
Similar as the discrete case we define
\[
K(t) = P(t)H^T(t)R_c^{-1}(t)
\]
then the final optimal prediction equation is
\[
\frac{d\hat{x}(t)}{dt} = F(t)\hat{x}(t) + B(t)u(t) + K(t)(z(t) - H(t)\hat{x}(t))
\]
\[
\frac{dP(t)}{dt} = F(t)P(t) + P(t)F^T(t) + G(t)G^T(t) - K(t)R_c(t)K^T(t)
\]
7 Examples

Example 3 (Noisy observation of a constant process). Consider the simplest case:

\[ dX(t) = 0, \quad \text{i.e. } X(t) = X(0), \quad E\{X(0)\} = 0, E\{X^2(0)\} = a^2 \]

\[ dZ(t) = X(t)dt + mdV(t); \quad Z(0) = 0. \]

In this model we have \( R_c(t) = m^2, F(t) = 0, B(t) = 0, G(t) = 0, H(t) = 1, z_c(t) = \dot{Z}(t). \) So we have \( K(t) = P(t)/m^2, \) and solving

\[ \frac{dP(t)}{dt} = -P^2(t)/m^2, \quad P(0) = a^2 \]

gives

\[ P(t) = \frac{a^2m^2}{m^2 + a^2t}, \quad t \geq 0. \]

For \( \dot{x}(t) \) we have

\[ \frac{d\dot{x}(t)}{dt} = K(t)(\dot{Z}(t) - H(t)x(t)) \]
\[ = -\frac{a^2}{m^2 + a^2t} \dot{x}(t) + \frac{a^2}{m^2 + a^2t} \dot{Z}(t). \]

This gives

\[ \dot{x}(t) = \frac{a^2}{m^2 + a^2t} Z(t). \]

Example 4 (Estimation of a parameter). Consider the simplest case:

\[ d\theta = 0, \quad \text{assuming } E\{\theta(0)\} = \hat{\theta}_0, \quad \text{Var}\{\theta^2(0)\} = P_0 \]

\[ dZ(t) = \theta M(t)dt + N(t)dV(t); \quad Z(0) = 0. \]

In this model we have \( R_c(t) = N^2(t), F(t) = 0, B(t) = 0, G(t) = 0, H(t) = M(t), z_c(t) = \dot{Z}(t). \) So we have \( K(t) = P(t)M(t)/N^2(t), \) and solving

\[ \frac{dP(t)}{dt} = -\left(\frac{P(t)M(t)}{N(t)}\right)^2, \quad P(0) = P_0 \]

gives gives

\[ P(t) = \left(P_0^{-1} + \int_0^t M^2(s)N^{-2}(s)ds\right)^{-1}, \quad t \geq 0. \]

For \( \dot{\theta}(t) \) we have

\[ \frac{d\dot{\theta}(t)}{dt} = \frac{M(t)P(t)}{N^2(t)}(\dot{Z}(t) - M(t)\dot{\theta}(t)) \]

Solving this equation we get

\[ \dot{\theta}(t) = \frac{\hat{\theta}_0P_0^{-1} + \int_0^t M(s)N^{-2}(s)dsZ(s)}{P_0^{-1} + \int_0^t M^2(s)N^{-2}(s)ds}. \]

References
