6.4 Kalman Filter Equations

6.4.1 Recap: Auxiliary variables

Recall the definition of the auxiliary random variables $x_p(k)$ and $x_m(k)$:

Init: $x_m(0) := x(0)$

$$\left.\begin{array}{ll} \textbf{S1:} & x_p(k) := A(k\text{–}1)x_m(k\text{–}1) + B(k\text{–}1)u(k\text{–}1) + v(k\text{–}1) \\ \textbf{S2:} & z_m(k) := H(k)x_p(k) + w(k) \\ & x_m(k) \text{ defined via its PDF} \\ & \quad f_{x_m(k)}(\xi) := f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) \quad \forall \xi \end{array}\right\} \quad k = 1, 2, \ldots$$

We proved last lecture that $x_p(k)$ is identical to the random variable $x(k)$ conditioned on $z(1{:}k\text{–}1)$, and $x_m(k)$ is identical to the random variable $x(k)$ conditioned on $z(1{:}k)$. That is, for all ξ and $k = 1, 2, \ldots,$

$$\begin{aligned} f_{x_p(k)}(\xi) &= f_{x(k)|z(1:k-1)}(\xi|\mathbf{z}(1{:}k\text{–}1)) \\ f_{x_m(k)}(\xi) &= f_{x(k)|z(1:k)}(\xi|\mathbf{z}(1{:}k))\,. \end{aligned}$$

Hence, $x_p(k)$ and $x_m(k)$ capture the prior update and the measurement update, respectively, of the Bayesian state estimator. We can work with these random variables in the following derivation of the Kalman filter.

First, we show that $x_p(k)$ and $x_m(k)$ are GRVs. As a consequence, their PDFs are fully characterized by mean and variance, which we then compute as a second step. The resulting recursive equations for mean and variance are the Kalman filter equations.

We use the following notation to denote mean and variance of the auxiliary variables:

$$\begin{aligned} \hat{x}_p(k) &= \mathrm{E}[x_p(k)], & P_p(k) &= \mathrm{Var}[x_p(k)], \\ \hat{x}_m(k) &= \mathrm{E}[x_m(k)], & P_m(k) &= \mathrm{Var}[x_m(k)]. \end{aligned}$$

6.4.2 Auxiliary variables are GRVs

Claim: $x_p(k)$ and $x_m(k)$ are GRVs.

Proof:
We prove by induction. The claim is true for $x_m(0)$ by definition. We assume $x_m(k\text{–}1)$ is a GRV and show that this implies that $x_p(k)$ and $x_m(k)$ are GRVs.

S1: By definition, $x_p(k)$ is an affine combination of the random variables $x_m(k\text{–}1)$ and $v(k\text{–}1)$. The process noise $v(k\text{–}1)$ is a GRV, and $x_m(k\text{–}1)$ is a GRV by induction assumption. The random variables $x_m(k\text{–}1)$ and $v(k\text{–}1)$ are independent since $x_m(k\text{–}1)$ depends on $x(0)$, $v(0:k\text{–}2)$, and $w(1:k\text{–}1)$, which are independent of $v(k\text{–}1)$. Hence, $x_m(k\text{–}1)$ and $v(k\text{–}1)$ are jointly Gaussian. Using Property 2 of last lecture (Sec. 6.2.1), it follows that $x_p(k)$ is a GRV.

S2: To simplify the notation, we drop the time index k (all random variables in question are indexed by k). We then have, by Bayes' rule,

$$f_{x_m}(\xi) := f_{x_p|z_m}(\xi|\mathbf{z}) = \frac{f_{z_m|x_p}(\mathbf{z}|\xi)\, f_{x_p}(\xi)}{f_{z_m}(\mathbf{z})}.$$

Since x_p is a GRV (proved in **S1** above), we have

$$f_{x_p}(\xi) \propto \exp\left(-\frac{1}{2}(\xi - \hat{x}_p)^T\, P_p^{-1}\,(\xi - \hat{x}_p)\right)$$

Date compiled: April 16, 2013

and, by definition of z_m,

$$f_{z_m|x_p}(\mathbf{z}|\xi) \propto \exp\left(-\frac{1}{2}(\mathbf{z}-H\xi)^T R^{-1}(\mathbf{z}-H\xi)\right).$$

The PDF $f_{z_m}(\mathbf{z})$ does not depend on ξ and is just a constant, therefore:

$$f_{x_m}(\xi) \propto \exp\left(-\frac{1}{2}\left((\xi-\hat{x}_p)^T P_p^{-1}(\xi-\hat{x}_p) + (\mathbf{z}-H\xi)^T R^{-1}(\mathbf{z}-H\xi)\right)\right). \tag{6.1}$$

Notice that the argument of the exponential is a quadratic in ξ. Therefore (we don't care about constants), there exist μ and Σ such that

$$f_{x_m}(\xi) \propto \exp\left(-\frac{1}{2}(\xi-\mu)^T \Sigma^{-1}(\xi-\mu)\right). \tag{6.2}$$

That is, $x_m(k)$ is a GRV, which completes the induction.

6.4.3 Mean and variance of auxiliary variables

We compute mean and variance of $x_p(k)$ and $x_m(k)$.

S1: By direct calculation:

$$\begin{aligned}
\hat{x}_p(k) &= \mathrm{E}[x_p(k)] = A(k\text{-}1)\mathrm{E}[x_m(k\text{-}1)] + B(k\text{-}1)u(k\text{-}1) + \mathrm{E}[v(k\text{-}1)] \\
&= A(k\text{-}1)\hat{x}_m(k\text{-}1) + B(k\text{-}1)u(k\text{-}1)
\end{aligned}$$

$$\begin{aligned}
P_p(k) &= \mathrm{Var}[x_p(k)] = \mathrm{E}[(x_p(k)-\hat{x}_p(k))(x_p(k)-\hat{x}_p(k))^T] \\
&= A(k\text{-}1)\,\mathrm{E}[(x_m(k\text{-}1)-\hat{x}_m(k\text{-}1))(x_m(k\text{-}1)-\hat{x}_m(k\text{-}1))^T]\,A^T(k\text{-}1) + \mathrm{E}[v(k\text{-}1)v^T(k\text{-}1)] \\
&= A(k\text{-}1)P_m(k\text{-}1)A^T(k\text{-}1) + Q(k\text{-}1).
\end{aligned}$$

Note that in the second to last equation, the cross-terms (for example, $\mathrm{E}[(x_m(k\text{-}1)-\hat{x}_m(k\text{-}1))v^T(k\text{-}1)]$) vanish because $x_m(k\text{-}1)-\hat{x}_m(k\text{-}1)$ and $v(k\text{-}1)$ are independent and zero-mean.

S2: We have $\hat{x}_m(k) = \mu$ and $P_m(k) = \Sigma$, with μ and Σ as in (6.2). We can compute μ and Σ by comparing the quadratic and linear terms of (6.1) and (6.2).

Quadratic term:

$$\begin{aligned}
&\xi^T(P_p^{-1} + H^T R^{-1} H)\xi = \xi^T \Sigma^{-1}\xi \quad \forall \xi \\
&\Leftrightarrow \quad \Sigma^{-1} = \left(P_p^{-1} + H^T R^{-1} H\right) \quad \text{(for symmetric } A,\ \xi^T A \xi = 0\ \forall\xi \text{ if and only if } A = 0)
\end{aligned}$$

Linear term:

$$\begin{aligned}
&-2\xi^T(P_p^{-1}\hat{x}_p + H^T R^{-1}\mathbf{z}) = -2\xi^T\Sigma^{-1}\mu \quad \forall\xi \\
&\Leftrightarrow \mu = \Sigma\left(P_p^{-1}\hat{x}_p + H^T R^{-1}\mathbf{z}\right) \\
&\qquad = \Sigma\left(P_p^{-1}\hat{x}_p + H^T R^{-1}(\mathbf{z} - H\hat{x}_p + H\hat{x}_p)\right) \\
&\qquad = \Sigma\left(\Sigma^{-1}\hat{x}_p + H^T R^{-1}(\mathbf{z} - H\hat{x}_p)\right) \\
&\qquad = \hat{x}_p + \Sigma H^T R^{-1}(\mathbf{z} - H\hat{x}_p).
\end{aligned}$$

6.4.4 Summary (Kalman Filter equations)

The Kalman filter is given by the recursive update equations for mean and variance above:

Initialization: $\hat{x}_m(0) = x_0$, $P_m(0) = P_0$.

Step 1 (S1): Prior update/Prediction step

$$
\begin{aligned}
\hat{x}_p(k) &= A(k\text{–}1)\hat{x}_m(k\text{–}1) + B(k\text{–}1)u(k\text{–}1) \\
P_p(k) &= A(k\text{–}1)P_m(k\text{–}1)A^T(k\text{–}1) + Q(k\text{–}1)
\end{aligned}
$$

Step 2 (S2): A posteriori update/Measurement update step

Results from above, re-introducing time index k and shorthand notation (in particular, $z(k)$ for $\mathbf{z}(k)$):

$$
\begin{aligned}
P_m(k) &= \big(P_p^{-1}(k) + H^T(k)R^{-1}(k)H(k)\big)^{-1} \\
\hat{x}_m(k) &= \hat{x}_p(k) + P_m(k)H^T(k)R^{-1}(k)\left(z(k) - H(k)\hat{x}_p(k)\right)
\end{aligned}
$$

Since the state $x(k)$ conditioned on past measurements $\mathbf{z}(1{:}k)$ is a GRV, its PDF is fully characterized by its mean $\hat{x}_m(k)$ and its variance $P_m(k)$ (and, analogously, the state $x(k)$ conditioned on $\mathbf{z}(1{:}k\text{–}1)$). This means that the Kalman Filter is the analytic solution to the Bayesian state estimation problem for a linear system with Gaussian distributions.

6.4.5 Alternative Equations

An alternative form of the measurement update equations (can be derived using the matrix inversion lemma): **(PSET 4: P2)**

$$
\begin{aligned}
K(k) &= P_p(k)H^T(k)\big(H(k)P_p(k)H^T(k) + R(k)\big)^{-1} && (6.3) \\
\hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(z(k) - H(k)\hat{x}_p(k)\right) && (6.4) \\
P_m(k) &= \big(I - K(k)H(k)\big)P_p(k) \\
&= \big(I - K(k)H(k)\big)P_p(k)\big(I - K(k)H(k)\big)^T + K(k)R(k)K^T(k) && (6.5)
\end{aligned}
$$

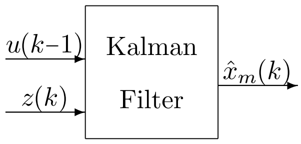
The matrix $K(k)$ is called the *Kalman Filter gain.*

6.5 Remarks

Implementation

If the problem data is known ahead of time (i.e. $A(k)$, $B(k)$, $H(k)$, $Q(k)$, and $R(k)$ for all k as well as P_0), all KF matrices (i.e. $P_p(k)$, $P_m(k)$, and $K(k)$) can be computed off-line. In particular, the KF gain does not depend on real-time measurement data!

$$
\begin{aligned}
\hat{x}_p(k) &= A(k\text{–}1)\hat{x}_m(k\text{–}1) + B(k\text{–}1)u(k\text{–}1) \\
\hat{x}_m(k) &= \hat{x}_p(k) + K(k)\left(z(k) - H(k)\hat{x}_p(k)\right) \\
\hat{x}_m(k) &= \big(I - K(k)H(k)\big)A(k\text{–}1)\hat{x}_m(k\text{–}1) \\
&\quad + \big(I - K(k)H(k)\big)B(k\text{–}1)u(k\text{–}1) + K(k)z(k) \\
&= \hat{A}(k)\hat{x}_m(k\text{–}1) + \hat{B}(k)u(k\text{–}1) + K(k)z(k)
\end{aligned}
$$



Obviously, the filter is linear and time varying.

Estimation error

Estimation error: $e(k) := x(k) - \hat{x}_m(k)$.

$$
\begin{aligned}
e(k) &= x(k) - \big(I - K(k)H(k)\big)\big(A(k\text{–}1)\hat{x}_m(k\text{–}1) + B(k\text{–}1)u(k\text{–}1)\big) - K(k)H(k)x(k) - K(k)w(k) \\
&= \big(I - K(k)H(k)\big)\big(A(k\text{–}1)x(k\text{–}1) + B(k\text{–}1)u(k\text{–}1) + v(k\text{–}1)\big) \\
&\quad - \big(I - K(k)H(k)\big)\big(A(k\text{–}1)\hat{x}_m(k\text{–}1) + B(k\text{–}1)u(k\text{–}1)\big) - K(k)w(k) \\
&= \big(I - K(k)H(k)\big)A(k\text{–}1)e(k\text{–}1) + \big(I - K(k)H(k)\big)v(k\text{–}1) - K(k)w(k)
\end{aligned}
$$

- $e(k)$ is a GRV.
- Error mean: $\mathrm{E}[e(k)] = (I - K(k)H(k))A(k\text{–}1)\mathrm{E}[e(k\text{–}1)]$. It follows that the filter is unbiased, i.e. $\mathrm{E}[e(k)] = 0$ (assuming initialization with $\hat{x}_m(0) = x_0$).
- Error variance: $\mathrm{Var}[e(k)] = \mathrm{E}[(x(k) - \hat{x}_m(k))(x(k) - \hat{x}_m(k))^T] = P_m(k)$.
- Note that stability of $(I - K(k)H(k))A(k\text{–}1)$ is important; for example, for the mean not to diverge if we do not initialize properly ($\mathrm{E}[e(0)] \neq 0$). Asymptotic properties of the KF are discussed in more detail in the next lecture.

Another interpretation

Note that the equations (6.3)–(6.5) are the recursive least squares (RLS) equations! This is a second interpretation of the KF, which also applies to non-Gaussian random variables: among the class of *linear*, *unbiased* estimators (specifically, those that have the structure $\hat{x}_m(k) = \hat{x}_p(k) + K(k)(z(k) - H(k)\hat{x}_p(k))$, compare RLS lecture #5), the KF is the one that *minimizes the mean squared error*.

- For a linear system and Gaussian noise, the KF is the best you can do (it keeps track of the full conditional PDFs). The linear estimator structure is optimal.
- Otherwise, one must be careful. The KF is no longer optimal (nonlinear may do better), but often reasonable (it is the best *linear* estimator in the MMSE sense).

Other KF formulations

There are other KF formulations that are equivalent to the equations presented herein, but may be advantageous for certain applications. One of them is the *information filter*.

Instead of working with the state variance, the information filter propagates the information matrix, which is defined as the inverse of the variance:

$$I_p(k) := P_p^{-1}(k), \quad I_m(k) := P_m^{-1}(k)$$

$$\text{Measurement update:} \quad I_m(k) = I_p(k) + H^T(k)R^{-1}(k)H(k).$$

While the variance captures the uncertainty in the state estimates, its inverse is a measure of the information that one has. Measurements increase the information. The information filter becomes computationally attractive when we have a large number of sensors, relative to the number of states.

Remark: Positive definiteness of variance matrices

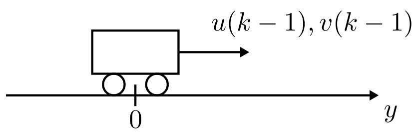
In the derivation of the KF, we assumed that $x(0)$, $v(k)$, and $w(k)$ are GRVs and therefore have positive definite variance matrices (i.e. $P_0 > 0$, $Q(k) > 0$, $R(k) > 0$). Note that the KF equations also make sense when some of the variance matrices are positive semidefinite (i.e. ≥ 0), as long as the involved matrix inversions are well defined (in the derivation of RLS, for example, we only require $H(k)P_p(k)H^T(k) + R(k) > 0$ in order to be invertible).

Examples:

- $P_0 = 0$ means that the initial state is perfectly known.
- $Q \geq 0$ with some zero eigenvalue means that there is no process noise in some direction of the state space.

6.6 Example

Consider an object moving along a line, with its position at time k denoted by $y(k)$ (in m).



- State: position $y(k)$ and velocity $\dot{y}(k)$.
- We can command the change of velocity with some uncertainty: $\dot{y}(k) = \dot{y}(k{-}1) + u(k{-}1) + v_{\dot{y}}(k{-}1)$ with $v_{\dot{y}}(k{-}1) \sim \mathcal{N}(0, 0.01)$.
- A position measurement, corrupted by noise, is available: $z(k) = y(k) + w(k)$ with $w(k) \sim \mathcal{N}(0, 0.5)$.
- Initial distribution: $x(0) \sim \mathcal{N}(0, I)$.

System equations

Assuming the object's velocity is constant for one time step T, we have $y(k) = y(k{-}1) + T\dot{y}(k{-}1)$. Let $x(k) = (y(k), \dot{y}(k))$. Combining all equations:

$$x(k) = \underbrace{\begin{bmatrix} 1 & T \\ 0 & 1 \end{bmatrix}}_{A} x(k{-}1) + \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{B} u(k{-}1) + v(k{-}1), \qquad v(k{-}1) \sim \mathcal{N}(0, Q), \quad Q = \begin{bmatrix} 0 & 0 \\ 0 & 0.01 \end{bmatrix},$$

$$z(k) = \underbrace{\begin{bmatrix} 1 & 0 \end{bmatrix}}_{H} x(k) + w(k), \qquad w(k) \sim \mathcal{N}(0, R), \quad R = 0.5.$$

A, B, H, Q, R, and initial mean and variance is all we need to implement the KF.

Simulations

(Matlab files to simulate this example are provided on the class website.)

- Simulate true state $x(k)$ and estimate $\hat{x}_m(k)$
  - KF provides estimates for position *and* velocity.
  - $P_m(k)$ captures uncertainty in estimates. For example, plot of mean plus/minus one standard deviation, $\hat{x}^i_m(k) \pm \sqrt{P^{ii}_m(k)}$: For any k, the probability of $x^i(k)$ being in the interval $[\hat{x}^i_m(k) - \sqrt{P^{ii}_m(k)}\ ,\ \hat{x}^i_m(k) + \sqrt{P^{ii}_m(k)}]$ is 68.2%.
  - $P_m(k)$ converges to a constant matrix as $k \to \infty$. (To be discussed in the next lecture.)
- Run simulation multiple times and plot distribution of final error $e(k_{\text{final}})$: a GRV with zero-mean and variance $P_m(k_{\text{final}})$, as expected.