7. The Kalman Filter: Algorithm

7.1 The Kalman Filter equations

Recall the auxiliary random variable description of the system:

$$x_m(0) := x(0), \quad x(0) \sim \mathcal{N}(x_0, P_0)$$

S1: $x_p(k) := A(k)x_m(k\text{-}1) + B(k)u(k) + v(k), \quad v(k) \sim \mathcal{N}(0, Q(k)) \qquad k = 1, 2, \ldots$

S2: $z_m(k) := H(k)x_p(k) + w(k), \quad w(k) \sim \mathcal{N}(0, R(k))$
$x_m(k)$ defined via its PDF $\qquad k = 1, 2, \ldots$

$$f_{x_m(k)}(\xi) := f_{x_p(k)|z_m(k)}(\xi|\mathbf{z}(k)) \quad \forall \xi$$

The objective is to calculate the PDFs of $x_p(k)$ and $x_m(k)$. Define

$$\hat{x}_p(k) = \mathrm{E}[x_p(k)], \qquad P_p(k) = \mathrm{Var}[x_p(k)]$$
$$\hat{x}_m(k) = \mathrm{E}[x_m(k)], \qquad P_m(k) = \mathrm{Var}[x_m(k)]$$

We claim that $x_p(k)$ and $x_m(k)$ are Gaussian random variables, which we will prove by induction. The importance of this property is that the random variables are then fully characterized by their mean and variance.

Proof:
The claim is true for $x_m(0)$, by definition. Assume $x_m(k\text{-}1)$ is GRV, show $x_p(k)$ and $x_m(k)$ are GRV.

S1: Linear combination of GRVs is a GRV, i.e. $x_p(k)$ is a GRV, by induction assumption.

S2: A bit more complicated. To simplify the notation, we use short-hand notation for PDFs again and will drop the time index k, since all random variables in question are indexed by k. We then have:

$$f(x_m) := f(x_p|z) = \frac{f(z|x_p)\, f(x_p)}{f(z)}$$

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

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

$$f(z|x_p) \propto \exp\left(-\frac{1}{2}\left((z - Hx_p)^T R^{-1} (z - Hx_p)\right)\right)$$

PDF $f(z)$ does not depend on x_p and is just a constant, therefore:

$$\begin{aligned} f(x_m) &\propto \exp\left(-\frac{1}{2}\left((x_p - \hat{x}_p)^T P_p^{-1} (x_p - \hat{x}_p) + (z - Hx_p)^T R^{-1} (z - Hx_p)\right)\right) \\ &\propto \exp\left(-\frac{1}{2}\left((x_p - \mu)^T \Sigma^{-1} (x_p - \mu)\right)\right) \quad \text{for some } \mu \text{ and } \Sigma \end{aligned}$$

Therefore $x_m(k)$ is a GRV! *This proves our induction.*

Date compiled: April 19, 2012

What are μ and Σ? Comparing terms, we get:

Quadratic term:

$$x_p^T \left(P_p^{-1} + H^T R^{-1} H\right) x_p = x_p^T \Sigma^{-1} x_p$$

$$\therefore \Sigma^{-1} = \left(P_p^{-1} + H^T R^{-1} H\right)$$

Linear Term:

$$-2x_p^T \left(P_p^{-1}\hat{x}_p + H^T R^{-1} z\right) = -2x_p^T \Sigma^{-1} \mu$$

$$\begin{aligned}\therefore \mu &= \Sigma \left(P_p^{-1}\hat{x}_p + H^T R^{-1} z\right) \\ &= \Sigma \left(P_p^{-1}\hat{x}_p + H^T R^{-1} \left(z - H\hat{x}_p + H\hat{x}_p\right)\right) \\ &= \Sigma \left(\Sigma^{-1}\hat{x}_p + H^T R^{-1} \left(z - H\hat{x}_p\right)\right) \\ &= \hat{x}_p + \Sigma H^T R^{-1} \left(z - H\hat{x}_p\right)\end{aligned}$$

Summary

Step 1: A priori state estimate/Prediction step.

The mean and variance of $x_p(k)$ can be readily calculated (**PSET 4: P4**).

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

Step 2: A posteriori state estimate/Measurement update step.

Results from above.

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

Done! Kalman Filter equations - simply bookkeeping of GRVs.

7.2 Alternative equations

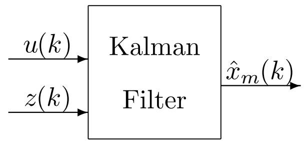
Alternative form of measurement update equations (**PSET 4: P5**):

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

$K(k)$: Kalman filter gain

- Note that these are the recursive least squares equations! This is a second interpretation of the Kalman Filter, which applies to non-Gaussian random variables: a *minimum mean squared error* (MMSE), optimal, *linear*, *unbiased* estimator.
- Gaussian random variables give us for free: linearity, and that the MMSE optimal value is just the mean, which by definition is unbiased.

- For Gaussian random variables, the Kalman Filter is the best you can do.
- For non-Gaussian random variables, one must be careful. The Kalman Filter is no longer optimal (nonlinear may do better), but reasonable.
- Information Filter: $I_p(k) := P_p^{-1}(k)$, $I_m(k) := P_m^{-1}(k)$.
  Example: $I_m(k) = I_p(k) + H^T(k)R^{-1}(k)H(k)$
  - The measurements give us information.
  - Becomes computationally attractive when we have a large number of sensors, relative to number of states.
- All matrices can be computed off-line!



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

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

7.3 The steady-state Kalman Filter

- All system and noise matrices are constant: $A(k) = A$, $Q(k) = Q$, etc.
- The Kalman Filter is still time varying:

$$\begin{aligned}P_p(k) &= AP_m(k\text{-}1)A^T + Q\\ P_m(k) &= \left(P_p^{-1}(k) + H^TR^{-1}H\right)^{-1}\end{aligned}$$

Assume $P_p(k)$ and $P_m(k)$ converge. Define $P_\infty := \lim_{k\to\infty} P_p(k)$. Then we have

$$P_\infty = A\left(P_\infty^{-1} + H^TR^{-1}H\right)^{-1}A^T + Q$$

We claim: $\left(P_\infty^{-1} + H^TR^{-1}H\right)^{-1} = P_\infty - P_\infty H^T\left(HP_\infty H^T + R\right)^{-1}HP_\infty$ (**PSET 4: P5**, matrix inversion lemma), from which follows (**PSET 4: P6**)

$$P_\infty = AP_\infty A^T - AP_\infty H^T\left(HP_\infty H^T + R\right)^{-1}HP_\infty A^T + Q,$$

a Discrete Algebraic Riccati Equation (DARE). We can then compute the steady-state gain:

$$K = P_\infty H^T\left(HP_\infty H^T + R\right)^{-1}.$$

This only makes sense if $P_\infty \geq 0$!

Theorem

The DARE has a unique solution $P_\infty \geq 0$ (all eigenvalues are non-negative) if and only if

1. (H, A) is detectable (can observe instability), and
2. (A, Q) is stabilizable (noise excites unstable parts)

Furthermore, the resulting $(I - KH)\,A$ is stable (all eigenvalues have magnitude less than unity).

The steady-state estimator

Let $\hat{x}(k) = \hat{x}_m(k)$ and $e(k) = x(k) - \hat{x}(k)$ (the estimation error).

$$\begin{aligned}\hat{x}(k) &= (I - KH)\, A\hat{x}(k\text{-}1) + (I - KH)\, Bu(k) + Kz(k) \\ &= (I - KH)\, \big(A\hat{x}(k\text{-}1) + Bu(k)\big) + KHx(k) + Kw(k)\end{aligned}$$

Therefore,

$$\begin{aligned}e(k) &= (I - KH)\, x(k) - (I - KH)\, \big(A\hat{x}(k\text{-}1) + Bu(k)\big) - Kw(k) \\ &= (I - KH)\, \big(Ax(k\text{-}1) + Bu(k) + v(k)\big) - (I - KH)\, \big(A\hat{x}(k\text{-}1) + Bu(k)\big) - Kw(k) \\ &= \underbrace{(I - KH)\, A}_{\text{stability important!}}\ e(k\text{-}1) + (I - KH)\, v(k) - Kw(k)\end{aligned}$$

Example

Let $A = 2$, $B = 0$, $H = 1$, $R = 1$, $Q = \varepsilon \geq 0$

$$\begin{aligned}x(k) &= 2x(k\text{-}1) + v(k) \\ z(k) &= x(k) + w(k)\end{aligned}$$

Case $\varepsilon = 0$: (A, Q) not stabilizable

$$P = 4P - \frac{4P^2}{P+1}, \quad P = 0 \text{ solution}$$

$$3 = \frac{4P}{P+1}, \quad 4P = 3P + 3, \quad P = 3 \text{ solution}$$

Two solutions! $P = 0$ seems to be better, since variance=0! $P = 0 \Rightarrow K = 0$, therefore:

$$\begin{aligned}\hat{x}(k) &= 2\hat{x}(k\text{-}1) \\ e(k) &= 2e(k\text{-}1) + v(k)\end{aligned}$$

But $v(k) = 0\ \forall\ k$ (0 variance). Therefore, if $e(0) = 0$, $e(k) = 0\ \forall\ k$. *But* if $e(0) \neq 0$ (or if a little bit of noise), we have a problem...

Case $\varepsilon > 0$: (A, Q) stabilizable

$$P = 4P - \frac{4P^2}{P+1} + \varepsilon, \quad \frac{4P^2}{P+1} = 3P + \varepsilon, \quad 4P^2 = 3P^2 + 3P + \varepsilon P + \varepsilon$$

$$P^2 - P(3 + \varepsilon) - \varepsilon = 0$$

$$\begin{aligned}P &= \frac{(3+\varepsilon) \pm \sqrt{(3+\varepsilon)^2 + 4\varepsilon}}{2} \\ &\approx \frac{(3+\varepsilon)}{2}\left(1 \pm \left(1 + \frac{1}{2}\frac{4\varepsilon}{(3+\varepsilon)^2}\right)\right) \\ &\approx 3 + \frac{4}{3}\varepsilon, \quad -\frac{\varepsilon}{3}\end{aligned}$$

Summary: The $P = 0$ solution is arbitrarily sensitive to noise.

$$P = 3: \quad (I - KH)\, A = \left(1 - \frac{3}{4}\right) 2 = \frac{1}{2}, \quad \text{a stable solution.}$$