Kalman Filter

Matthew Cocci

November 5, 2013

1. Basic Idea and Terminology

Here's the basic procedure associated with the Kalman Filter:

1. Start with a prior for some variable of interest in the current period, $p(x)$.
2. Observe the current measurement y_t.
3. "Filter" out the noise and compute the filtering distribution: $p_t(x|y)$.
4. Compute the predictive distribution $p_{t+1}(x)$ from the filtering distribution and your model.
5. Increment t by one, and go back to step 1, taking the predictive distribution as your prior.

2. Normal Example

Suppose we want to measure some variable x. We will assume a *prior* that is multivariate normal such that

$$x \sim \mathrm{N}(\hat{x}, \Sigma)$$

Next, we "measure" x by matching it to an observable in a *measurement equation*:

$$y = Gx + v \qquad v \sim \mathrm{N}(0, R)$$

where R is positive definite, while G and R are both 2×2. This forms the *likelihood*.

We then "filter" out the noise, updating our view of x in light of the data in the filtering step using Bayes' Rule:

$$\begin{aligned} p(x \mid y) &= \frac{p(y \mid x) \cdot p(x)}{p(y)} \propto p(y \mid x) \cdot p(x) \qquad (1) \\ &\propto \exp\left\{-\frac{1}{2}\,(y - Gx)'\,R\,(y - Gx)\right\} \exp\left\{-\frac{1}{2}(x - \hat{x})'\Sigma(x - \hat{x})\right\} \end{aligned}$$

Now let's expand the term the lefthand exponential:

$$\begin{aligned} A = (y - Gx)' \, R \, (y - Gx) &= (y' - x'G') \, R \, (y - Gx) \\ &= (y'R - x'G'R) \, (y - Gx) \\ &= (y'Ry - y'RGx - x'G'Ry + x'G'RGx) \end{aligned}$$

And now the same for the righthand exponential:

$$\begin{aligned} B = (x - \hat{x})'\Sigma(x - \hat{x}) &= (x' - \hat{x}')\Sigma(x - \hat{x}) \\ &= (x'\Sigma - \hat{x}'\Sigma)(x - \hat{x}) \\ &= x'\Sigma x - x'\Sigma\hat{x} - \hat{x}'\Sigma x + \hat{x}'\Sigma\hat{x} \end{aligned} \tag{2}$$

Adding the two exponentials, we get:

$$\begin{aligned} C = A + B = x' \, (\Sigma + G'RG) \, x &- x'(\Sigma\hat{x} + G'Ry) - (\hat{x}'\Sigma + y'RG)x \\ &+ \hat{x}'\Sigma\hat{x} + y'Ry \end{aligned}$$

Now notice that Expression 1 is the probability distribution of x *conditional* on y and pretty much anything else that isn't x. And because of the wonderful properties of the exponential function and the black-hole nature of the proportionality constant, we'll be able to simplify things nicely (and we'll worry that the distribution $p(x \mid y)$ integrates to one later on).

Specifically, in the expression for C, the two terms in the second row *don't* depend upon x. Therefore, letting $C(x)$ be the portion of C that depends upon x, and letting $C(\neg x)$ bet the additive terms which don't depend upon x, we can simplify

$$\begin{aligned} p(x \mid y) \propto \exp\left\{-\frac{1}{2}C\right\} &= \exp\left\{-\frac{1}{2}\left[C(x) + C(\neg x)\right]\right\} \\ \propto \exp\left\{-\frac{1}{2}C(x)\right\} &+ \exp\left\{-\frac{1}{2}C(\neg x)\right\} \\ \propto \exp\left\{-\frac{1}{2}C(x)\right\} & \end{aligned}$$

We just absorb the portion not relevant to $p(x \mid y)$ into the proportionality constant. This means our the work we did above to get C simplifies our target expression to

$$p(x \mid y) \propto \exp\left\{-\frac{1}{2}\left[x' \, (\Sigma + G'RG) \, x - x'(\Sigma\hat{x} + G'Ry) - (\hat{x}'\Sigma + y'RG)x\right]\right\} \tag{3}$$

Now this doesn't look too helpful, but with a little bit of work, we can turn this into the probability distribution for a multivariate normal random variable. So let's do it.

If we take a second and compare that Expression 3 to Expression 2, it's becomes clear from inspection that we must have

$$(\Sigma\hat{x} + G'Ry) = (\Sigma + G'RG) \, Z \tag{4}$$

To see this, liken the lefthand side of Equation 4 to the result of the matrix multiplication $\Sigma\hat{x}$ in Equation 2. To get the righthand side, use the fact that we *know* the Equation 4 analogue to Equation 2's Σ: it's sandwiched between x and x'. So all that's left to do is solve for Z.

And so we solve Equation 2 by using the Woodbury matrix identity, stated in the appendix:

$$\begin{aligned}
(\Sigma\hat{x} + G'Ry) &= (\Sigma + G'RG)\,Z \\
\Rightarrow \quad Z &= (\Sigma + G'RG)^{-1}\,(\Sigma\hat{x} + G'Ry) \\
Z &= \left(\Sigma^{-1} - \Sigma^{-1}G'(R^{-1} + G\Sigma^{-1}G')^{-1}G\Sigma^{-1}\right)(\Sigma\hat{x} + G'Ry)
\end{aligned}$$

Now do some heavy simplifying, using the result $(AB)^{-1} = B^{-1}A^{-1}$ often and, typically, in reverse:

$$\begin{aligned}
Z &= \left(\Sigma^{-1} - \Sigma^{-1}G'(R^{-1} + G\Sigma^{-1}G')^{-1}G\Sigma^{-1}\right)(\Sigma\hat{x} + G'Ry) \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\Sigma^{-1}G'(R^{-1} + G\Sigma^{-1}G')^{-1}G\Sigma^{-1}\right]\Sigma\hat{x} \\
&\quad - \left[\Sigma^{-1}G'(R^{-1} + G\Sigma^{-1}G')^{-1}G\Sigma^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\Sigma^{-1}G'(R^{-1} + G\Sigma^{-1}G')^{-1}G\right]\hat{x} \\
&\quad - \left[\left(G'^{-1}\Sigma\right)^{-1}(R^{-1} + G\Sigma^{-1}G')^{-1}\left(\Sigma G^{-1}\right)^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[(G'^{-1}\Sigma)^{-1}(R^{-1} + G\Sigma^{-1}G')^{-1}G\right]\hat{x} \\
&\quad - \left[\left\{\left(\Sigma G^{-1}\right)(R^{-1} + G\Sigma^{-1}G')\left(G'^{-1}\Sigma\right)\right\}^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\left\{(R^{-1} + G\Sigma^{-1}G')(G'^{-1}\Sigma)\right\}^{-1}G\right]\hat{x} \\
&\quad - \left[\left\{\left(\Sigma G^{-1}\right)R^{-1} + \left(\Sigma G^{-1}\right)G\Sigma^{-1}G'\left(G'^{-1}\Sigma\right)\right\}^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\left\{R^{-1}G'^{-1}\Sigma + G\Sigma^{-1}G'G'^{-1}\Sigma\right\}^{-1}G\right]\hat{x} \\
&\quad - \left[\left\{\Sigma G^{-1}R^{-1} + \Sigma\right\}^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\left\{R^{-1}G'^{-1}\Sigma + G\right\}^{-1}\left(G^{-1}\right)^{-1}\right]\hat{x} \\
&\quad - \left[\left\{\Sigma G^{-1}R^{-1} + \Sigma\right\}^{-1}\right]G'Ry \\
&= \hat{x} + \Sigma^{-1}G'Ry - \left[\left\{G^{-1}\left(R^{-1}G'^{-1}\Sigma + G\right)\right\}^{-1}\right]\hat{x} \\
&\quad - \left[\left\{\Sigma G^{-1}R^{-1} + \Sigma\right\}^{-1}\right]G'Ry
\end{aligned}$$

A. Woodbury Matrix Identity

For matrices A, U, C, and V:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{5}$$