Kalman Filter

Matthew Cocci

November 6, 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:

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

Now let's expand the term the lefthand exponential:

$$\begin{aligned} A = (y - Gx)' R^{-1} (y - Gx) &= (y' - x'G') R^{-1} (y - Gx) \\ &= \left(y'R^{-1} - x'G'R^{-1}\right)(y - Gx) \\ &= \left(y'R^{-1}y - y'R^{-1}Gx - x'G'R^{-1}y + x'G'R^{-1}Gx\right) \end{aligned}$$

And now the same for the righthand exponential:

$$\begin{aligned} B = (x - \hat{x})'\Sigma^{-1}(x - \hat{x}) &= (x' - \hat{x}')\Sigma^{-1}(x - \hat{x}) \\ &= (x'\Sigma^{-1} - \hat{x}'\Sigma^{-1})(x - \hat{x}) \\ &= x'\Sigma^{-1}x - x'\Sigma^{-1}\hat{x} - \hat{x}'\Sigma^{-1}x + \hat{x}'\Sigma^{-1}\hat{x} \end{aligned} \tag{2}$$

Adding the two exponentials, we get:

$$\begin{aligned} C = A + B = x'\left(\Sigma^{-1} + G'R^{-1}G\right)x - x'(\Sigma^{-1}\hat{x} + G'R^{-1}y) - (\hat{x}'\Sigma^{-1} + y'R^{-1}G)x \\ + \hat{x}'\Sigma^{-1}\hat{x} + y'R^{-1}y \end{aligned}$$

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

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

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.

First, the variance of the normal distribution corresponding to $p(x \mid y)$ can be derived by examining Equation 3 and likening it to Equation 2 (which has the contents of the exponential in the prior distribution of x). Namely, the inverse of the new variance, which we'll denote as Σ_F will be sandwiched in between x' and x in Equation 3, just as it was sandwiched between x' and x in Equation 2. We use this fact, along with the the Woodbury matrix identity, stated in the appendix, to derive:

$$\begin{aligned} \Sigma_F &= \left(\Sigma^{-1} + G'R^{-1}G\right)^{-1} \\ \text{Woodbury Identity} \Rightarrow \qquad &= \Sigma - \Sigma G'(R + G\Sigma G')^{-1}G\Sigma \end{aligned}$$

Next, we want to get the mean of the distribution of $p(x|y)$. Again, once we take a second and compare Expression 3 to Expression 2, it's becomes clear from inspection that we must have

$$(\Sigma^{-1}\hat{x} + G'R^{-1}y) = \left(\Sigma^{-1} + G'R^{-1}G\right) Z \tag{4}$$

To see this, liken the lefthand side of Equation 4 to the result of the matrix multiplication $\Sigma^{-1}\hat{x}$ in Equation 2. To get the righthand side, use the fact that we *know* the Equation 4 analogue to Equation 2's Σ^{-1}, which we just derived.

So all that's left to do is solve for Z in Equation 4. And so we solve Equation 2 by using the Woodbury matrix identity representation from above:

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

Now let's simplify this representation a bit, particularly using the trick $(AB)^{-1} =$

$B^{-1}A^{-1}$ and especially in reverse:

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

A. Woodbury Matrix Identity

For matrices A, U, C, and V:

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

Now consider the special case we have above with the Kalman filter:

$$(A + V'CV)^{-1} = A^{-1} - A^{-1}V'(C^{-1} + VA^{-1}V')^{-1}VA^{-1} \tag{6}$$

We'll use the trick $(AB)^{-1} = B^{-1}A^{-1}$ often, and typically in reverse, going from the right side of the equality to the left to simplify Equation 6.

$$\begin{aligned}
(A + V'CV)^{-1} &= A^{-1} - A^{-1}V'(C^{-1} + VA^{-1}V')^{-1}VA^{-1} \\
&= A^{-1} - \left(V'^{-1}A\right)^{-1}(C^{-1} + VA^{-1}V')^{-1}VA^{-1} \\
&= A^{-1} - \left\{(C^{-1} + VA^{-1}V')\left(V'^{-1}A\right)\right\}^{-1}VA^{-1} \\
&= A^{-1} - \left\{C^{-1}\left(V'^{-1}A\right) + VA^{-1}V'\left(V'^{-1}A\right)\right\}^{-1}VA^{-1} \\
&= A^{-1} - \left\{C^{-1}\left(V'^{-1}A\right) + V\right\}^{-1}\left(AV^{-1}\right)^{-1} \\
&= A^{-1} - \left\{\left(AV^{-1}\right)\left[C^{-1}\left(V'^{-1}A\right) + V\right]\right\}^{-1} \\
&= A^{-1} - \left\{AV^{-1}C^{-1}\left(V'^{-1}A\right) + AV^{-1}V\right\}^{-1} \\
&= A^{-1} - \left\{AV^{-1}C^{-1}V'^{-1}A + A\right\}^{-1} \\
&= A^{-1} - \left\{AV^{-1}C^{-1}V'^{-1}A + A\right\}^{-1}
\end{aligned}$$