Kalman Filter

Matthew Cocci

November 11, 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 predictive distribution $p_{t+1}(x)$ from filtering distribution and your model.
5. Increment t and return to step 1, taking the predictive distribution as your prior.

2. Normal Example, Filtering Step

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. Note, this is called "filtering" because we don't use the prior and likelihood forecast into the future. We combine the prior with the likelihood to filter out noise and get closer to the true value of x based on the data, summarzed in the posterior, or the *filtering distribution*:

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

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:

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

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 $p(x|y)$ 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

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

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. In fact, the rest of the section may look complicated, but keep in mind the big picture: the likelihood and prior were both multivariate normal, so the posterior $p(x|y)$ is going to be normal. We just want to identify the mean vector and variance-covariance matrix; then we're home.

Variance So first, the variance of the normal distribution corresponding to $p(x|y)$ 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:

$$\begin{aligned} \Sigma_F &= \left(\Sigma^{-1} + G'R^{-1}G\right)^{-1} \\ \text{Woodbury Identity} \Rightarrow \qquad &= \Sigma - \Sigma G'(R + G\Sigma G')^{-1}G\Sigma \end{aligned} \tag{4}$$

Mean Next, we want to get the mean of the distribution of $p(x|y)$, which we'll denote by $\hat{x}_F$. Again, once we take a second and compare Expression 3 to Expression 2, it's becomes clear from inspection that we must have

$$(\Sigma^{-1}\hat{x} + G'R^{-1}y) = \left(\Sigma^{-1} + G'R^{-1}G\right) Z \tag{5}$$

To see this, liken the lefthand side of Equation 5 (which itself comes from Expression 3) 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 5 analogue to Equation 2's Σ^{-1}, which we just derived and called Σ_F.

So all that's left to do is solve for Z in Equation 5. The result will turn out to be our mean vector for the posterior, $\hat{x}_F$. 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 $\hat{x}_F = Z$ a bit, expanding out the multiplication:

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

Okay, now let's take a breather. We'll make this a bit easier on ourselves, and just consider simplifying the guy in the brackets, $\{\}$:

$$
\begin{aligned}
\left\{R^{-1} - (R + G\Sigma G')^{-1} G\Sigma G' R^{-1}\right\} &= (R + G\Sigma G')^{-1}(R + G\Sigma G')R^{-1} \\
&\quad - (R + G\Sigma G')^{-1} G\Sigma G' R^{-1} \\
&= (R + G\Sigma G')^{-1}\left[(R + G\Sigma G')R^{-1} - G\Sigma G' R^{-1}\right] \\
&= (R + G\Sigma G')^{-1}\left[RR^{-1} + G\Sigma G' R^{-1} - G\Sigma G' R^{-1}\right] \\
&= (R + G\Sigma G')^{-1}\left[I + 0\right] \\
&= (R + G\Sigma G')^{-1}
\end{aligned}
$$

Substituting back in above for the term in braces, $\{\}$, we get the following expression for $Z = \hat{x}_F$:

$$
\begin{aligned}
\hat{x}_F = Z &= \hat{x} - \left[\Sigma G'(R + G\Sigma G')^{-1}\right](G\hat{x}) + \Sigma G'(R + G\Sigma G')^{-1} y \\
&= \hat{x} + \left[\Sigma G'(R + G\Sigma G')^{-1}\right](y - G\hat{x})
\end{aligned} \tag{6}
$$

Putting together the expressions for the mean and variance (see Equations 6 and 4, respectively) of the posterior estimate for $x|y$, we get that

$$
x|y \sim \mathrm{N}\left(\hat{x}_F,\ \Sigma_F\right) \tag{7}
$$

$$
\begin{aligned}
\text{where} \quad \hat{x}_F &= \hat{x} + \left[\Sigma G'(R + G\Sigma G')^{-1}\right](y - G\hat{x}) \\
\Sigma_F &= \Sigma - \Sigma G'(R + G\Sigma G')^{-1} G\Sigma
\end{aligned}
$$

Conclusion Okay, so what did we just do?

1. We took a Multivariate Normal (MVN) prior to summarize our beliefs about x.

2. Knowing that we'll observe some data, y, that provides a "noisy" measure of x, we postulated a likelihood $p(y|x)$ that is also MVN.

3. Then, using Bayes' Rule, we combine the information contained in our prior $p(x)$ and the data (via the likelihhod $p(y|x)$) to get a "filtered" distribution of x, $p(x|y)$, given the data and our prior.

Why might this long, tortuous, painful process help us in economics? Well, imagine that in our model, there's some state for the level of output, denoted by x. Now we'll have economic statistics, like measurements of GDP itself along with other informative statistics such as unemployment and investment, which might provide information about output. But of course, those statistics are imperfect and noisy. The Kalman Filter gives us a way to combine those noise estimates in with our beliefs in a principled, sensible manner.

3. Normal Expample, Forecasting Step

Now let's make our model a little more dynamic and include forecasting in our method. To do so, we specify a model of how the state, x evolves. To make it easy on ourselves, let's assume everything's Gaussian (woohoo! that's easy):

$$x_{t+1} = Ax_t + w_{t+1} \qquad w_t \sim \mathrm{N}(0, Q) \tag{8}$$

A. Woodbury Matrix Identity

For matrices A, U, C, and V:

$$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1} \tag{9}$$

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

$$(A + V'CV)^{-1} = A^{-1} - A^{-1}V'(C^{-1} + VA^{-1}V')^{-1}VA^{-1} \tag{10}$$