Brief Description of Kalman Filter

Phanindra Jampana

October 10, 2014

0.1 Kalman Filter

For vector valued random variables, the following definitions are used:

- $E(x) = (E(x_1)E(x_2)\cdots E(x_n))'$ (Just component wise expectation)
- $\mathrm{Cov}(x) = E((x - Ex)(x - Ex)^T)$
- $E(x|y) = (E(x_1|y)E(x_2|y)\cdots E(x_n|y))'$
- $\mathcal{L}^2 = \{x|Ex^2 < \infty\}$
- x has a multivariate Gaussian distribution if all linear combinations, $\sum_{k=1}^{n} a_k x_k$ are univariate Gaussian
- Also, if $x, y \in \mathcal{L}^2$ then $xy \in \mathcal{L}^1$

$$\begin{aligned}
x[n+1] &= A(\theta)x[n] + B(\theta)u[n] + w[n] \\
y[n] &= C(\theta)x[n] + D(\theta)u[n] + v[n] \\
Ew[n]w[n]^T &= R_1(\theta) \\
Ev[n]v[n]^T &= R_2(\theta) \\
Ew[n]v[n]^T &= R_{12}(\theta)
\end{aligned}$$

where $w[n], v[n]$ are multivariate Gaussian and $A(\theta) \in \mathbb{R}^{n\times n}, B(\theta) \in \mathbb{R}^{n\times r}, C(\theta) \in \mathbb{R}^{m\times n}, D(\theta) \in \mathbb{R}^{m\times r}$. We will denote by $x_i[n]$ the ith component of the vector $x[n]$.

We would like to derive the predictor $x[n|n-1] := E(x[n]|y[k], k \le n-1)$. The main idea used in the derivation of this predictor is that if X, Y are jointly Gaussian then $E(X|Y)$ can be treated as an orthogonal projection in $\mathcal{L}^2$. We will show this now.

The following theorem can be found in [1]:

Theorem 0.1.1. *If $\mathcal{K}$ is a complete vector subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, then given $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ there exists $Y \in \mathcal{K}$ such that*

- $\|X - Y\| := \inf\{\|X - W\| : W \in \mathcal{K}\}$
- $E((X - Y)Z) = 0 \ \forall Z \in \mathcal{K}$
- *Both the above are equivalent and if $\tilde{Y}$ is another r.v. which satisfies either the above, then $\|Y - \tilde{Y}\| = 0$ a.s.*

Let $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P), \mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$ and let $\mathcal{K} = \mathcal{L}^2(\Omega, \mathcal{G}, P)$. Then taking $Z = I_G, G \in \mathcal{G}$, we get that $Y = E(X|\mathcal{G})$. Hence, conditional expectation is equal to orthogonal projection for $\mathcal{L}^2$ random variables.

> In rough terms, the conditional expectation of an $\mathcal{L}^2$ random variable X given Y is the same as the orthogonal projection of X on the space of all random variables of the form $f(Y)$ where f is any (measurable)-function.

Now, let's consider the space $\mathcal{L}^2$ and define the inner product, $< X, Y >= E(XY)$. (This is not a true inner product as it does not satisfy the positive definiteness condition. However, if we identity equivalence classes with the a.s. relation, then it is a correct inner product).

So, essentially, we think of random variables as *vectors*. The difference from $\mathbb{R}^n$ is that no finite number of vectors span the entire space as $\mathcal{L}^2$ is infinite dimensional.

With the inner product, we can define the orthogonal projection of X onto the span$(Y_1, Y_2, \cdots, Y_n)$ denoted by $\hat{E}(X|Y_1, Y_2, \cdots, Y_n)$. From standard ideas in orthogonal projection we have

$$E((X - \hat{E}(X|Y_1, Y_2, \cdots, Y_n))Y_k) = 0 \ \forall 1 \leq k \leq n$$

If we now assume that $X, Y_1, Y_2, \cdots, Y_n$ are multivariate Gaussian, then by definition $X - \hat{E}(X|Y_1, Y_2, \cdots, Y_n)$ is also Gaussian (as $\hat{E}(X|Y_1, Y_2, \cdots, Y_n) = \sum_{k=1}^{n} a_k Y_k$ for some a_k's).

However, if X, Y are Gaussian and $E(XY) = 0$ then X, Y are independent. Hence, $X - \hat{E}(X|Y_1, Y_2, \cdots, Y_n)$ and Y_k are independent. This in-turn gives that $X - \hat{E}(X|Y_1, Y_2, \cdots, Y_n)$, $f(Y_1, Y_2, \cdots, Y_k)$ are independent for any measurable f. From the above theorem, it follows that

$$\hat{E}(X|Y_1, Y_2, \cdots, Y_n) = E(X_1|Y_1, Y_2, \cdots, Y_n) \tag{1}$$

Without multivariate Gaussianity, conditional expectation is the projection on to a very complicated space $f(Y_1, Y_2, \cdots, Y_n)$, where f is any measurable function such that $f(Y_1, Y_2, \cdots, Y_n)$ is square integrable. However, with the assumption of Gaussianity, the projection is onto a much simpler space, span$(Y_1, Y_2, \cdots, Y_n)$. Now, standard projection ideas can be easily applied. Therefore, this idea can be considered to central to the development of the theory.

First, it is to be noted that all random variables that we get during the computations are multivariate Gaussian. For example, $(x[n], y[1], y[2], \cdots, y[n])$ is multivariate Gaussian as $x[n]$ and each of the $y[k]$ are linear combinations of $w[n]$ and $v[n]$. Therefore, we can freely use (1).

Denoting span$(y[k], 1 \leq k \leq n) =: \mathcal{Y}[n]$ and following Kalman, we start with

$$\begin{aligned} E(x[n+1]|y[k], 1 \leq k \leq n) &= \hat{E}(x[n+1]|\mathcal{Y}[n]) \\ &= \hat{E}(x[n+1]|\mathcal{Y}[n-1]) + \hat{E}(x[n+1]|\mathcal{Z}[n]) \\ &= A(\theta)E(x[n]|y[k], 1 \leq k \leq n-1) + B(\theta)u[n] + \hat{E}(x[n+1]|\mathcal{Z}[n]) \end{aligned}$$

Note here that $\mathcal{Y}[n] = \mathcal{Y}[n-1] \oplus \mathcal{Z}[n]$, i.e. direct sum of two vector subspaces. $\mathcal{Z}[n]$ is spanned by $\tilde{y}[n|n-1]$, where $y[n] = \bar{y}[n|n-1] + \tilde{y}[n|n-1]$ and $\bar{y}[n|n-1] \in \mathcal{Y}[n-1]$. Hence, $\hat{E}(x[n+1]|\mathcal{Z}[n]) = \Delta^*[n]\tilde{y}[n|n-1]$ and

$$\begin{aligned} \tilde{y}[n|n-1] &= y[n] - \bar{y}[n|n-1] \\ &= y[n] - C(\theta)E(x[n]|y[k], 1 \leq k \leq n-1) - D(\theta)u[n] \end{aligned}$$

Therefore,

$$\begin{aligned} x[n+1|n] &= A(\theta)x[n|n-1] + B(\theta)u[n] + \Delta^*[n](y[n] - C(\theta)x[n|n-1] - D(\theta)u[n]) \\ &= (A(\theta) - \Delta^*[n]C(\theta))x[n|n-1] + B(\theta)u[n] + \Delta^*[n](y[n] - D(\theta)u[n]) \end{aligned}$$

where we used the notation, $x[n+1|n] = E(x[n+1]|y[k], 1 \leq k \leq n)$. In the following, we use $\tilde{x}[n+1|n] = x[n+1] - x[n+1|n]$. From the above

$$\begin{aligned} \tilde{x}[n+1|n] &= x[n+1] - (A(\theta) - \Delta^*[n]C(\theta))x[n|n-1] - B(\theta)u[n] - \Delta^*[n](y[n] - D(\theta)u[n]) \\ &= A(\theta)x[n] - (A(\theta) - \Delta^*[n]C(\theta))x[n|n-1] - \Delta^*[n](C(\theta)x[n] + v[n]) + w[n] \\ &= (A(\theta) - \Delta^*[n]C(\theta))\tilde{x}[n|n-1] - \Delta^*[n]v[n] + w[n] \end{aligned}$$

We will denote $(A(\theta) - \Delta^*[n]C(\theta)) =: A^*(\theta, n), E(\tilde{x}[n+1|n]\tilde{x}[n+1|n]^T) = P[n+1|n]$. We can now calculate this covariance matrix of the error in estimation, $\tilde{x}[n+1|n]$ assuming that $E(x[0]) = 0$.

$$P[n+1|n] = A^*(\theta, n)P[n|n-1]A^*(\theta, n) + R_1(\theta) + \Delta^*[n]R_2(\theta)\Delta^*[n] - \Delta^*[n](R_{12}(\theta))^T - R_{12}(\theta)(\Delta^*[n])^T$$

Now, we need to find $\Delta^*[n]$. Now, $x[n+1] - \hat{E}(x[n+1]|\mathcal{Z}[n])$ is orthogonal to $\tilde{y}[n|n-1]$. Therefore,

$$\begin{aligned}
0 &= E(x[n+1] - \hat{E}(x[n+1]|\mathcal{Z}[n]))\tilde{y}[n|n-1]^T \\
&= E(x[n+1] - \Delta^*[n]\tilde{y}[n|n-1])\tilde{y}[n|n-1]^T \\
&= E(x[n+1]\tilde{y}[n|n-1]^T) - E(\Delta^*[n]\tilde{y}[n|n-1]\tilde{y}[n|n-1]^T) \\
&= E(\tilde{x}[n+1|n-1]\tilde{y}[n|n-1]^T) - E(\Delta^*[n]\tilde{y}[n|n-1]\tilde{y}[n|n-1]^T)
\end{aligned}$$

Now,

$$\begin{aligned}
E(\tilde{x}[n+1|n-1]\tilde{y}[n|n-1]^T) &= x[n+1] - A(\theta)x[n|n-1] - B(\theta)u[n] \\
&= A(\theta)\tilde{x}[n|n-1] + w[n] \\
\tilde{y}[n|n-1] &= y[n] - C(\theta)x[n|n-1] - D(\theta)u[n] \\
&= C(\theta)\tilde{x}[n|n-1] + v[n]
\end{aligned}$$

Therefore,

$$\begin{aligned}
0 &= E(\tilde{x}[n+1|n-1]\tilde{y}[n|n-1]^T) - E(\Delta^*[n]\tilde{y}[n|n-1]\tilde{y}[n|n-1]^T) \\
&= E((A(\theta)\tilde{x}[n|n-1] + w[n])(\tilde{y}[n|n-1]^T) - E(\Delta^*[n]\tilde{y}[n|n-1]\tilde{y}[n|n-1]^T) \\
&= E(A(\theta)\tilde{x}[n|n-1] + w[n])(\tilde{x}[n|n-1]^T C(\theta)^T + v[n]^T) - E(\Delta^*[n](C(\theta)\tilde{x}[n|n-1] + v[n])(\tilde{x}[n|n-1]^T C(\theta)^T + v[n]^T) \\
&= A(\theta)P[n|n-1]C(\theta)^T + R_{12}(\theta) - \Delta^*[n]C(\theta)P[n|n-1]C(\theta)^T - \Delta^*[n]R_2(\theta)
\end{aligned}$$

which results in the Kalman gain,

$$\Delta^*[n] = (A(\theta)P[n|n-1]C(\theta)^T + R_{12}(\theta)) * (C(\theta)P[n|n-1]C(\theta)^T + R_2(\theta))^{-1}$$

Note that the prediction can be taken one step further by estimating $x[n|n]$ but this will not be done here as it is easy to derive once the concepts are understood.

Bibliography

[1] David Williams. *Probability with Martingales*. Cambridge University Press, 1991.