<u>Derivation of the Kalman Recursive Filter Equations</u>

(Gelb, A. (ed.), **Applied Optimal Estimation,** MIT Press, Cambridge, MA, 1974)

A recursive estimator is one in which there is no need to store all past measurements for the purpose of computing current estimates. The Kalman filter is a recursive Bayesian estimation technique used to track the states of dynamic systems being observed with noisy sensors.

Let us define an *estimate*, $\hat{x}$, to be the computed value of a state, x, based on a set of measurements, z. From a *maximum likelihood* standpoint, we take $\hat{x}$ to be the quantity which maximizes the probability of the measurement z that actually occurred. That is,

$$\hat{x} = \arg\max_x \left[p(x|z) \right],$$

where

$$p(x|z) = \frac{p(z|x)\ p(x)}{p(z)},$$

$p(x)$ is the *a priori* probability density function of x, $p(z)$ is probability density of the measurements, and $p(z|x)$ is the data model.

From a *minimum variance* point of view, an estimate is one whose error variance is less than or equal to that of any other unbiased estimate. To find a generalized minimum variance Bayes' estimate is to minimize the cost function

$$J = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\hat{x} - x)^T S (\hat{x} - x) p(x|z) dx_1 dx_2 \dots dx_n,$$

where S is an arbitrary, positive semidefinite matrix. Simply set $\partial J / \partial \hat{x} = 0$ to find, independent of S, that the Bayes estimate of x given measurement z is

$$\hat{x} = E[x|z].$$

where $E[]$ is the expectation operation.

In keeping with the Bayesian approach we assume that we have an initial statistical model, the prior model, $p(x)$, for the state. We will assume that this is a Gaussian with mean $\hat{x}_0$ and covariance P_0. That is,

$$p(x) = \frac{1}{(2\pi)^{N/2} |P_0|^{1/2}} \exp\left[-\frac{1}{2} (x - \hat{x}_0)^T P_0^{-1} (x - \hat{x}_0) \right]$$

N is the dimension of the state vector.

We also need to specify the measurement model, which, in Bayesian terms, is captured by the conditional probability $p(z|x)$. To determine this, we assume that the measurements are a linear function of the state with added noise. That is,

$$z = Hx + \nu,$$

where H is the observation matrix and ν is assumed to be a Gaussian random variable with zero mean and covariance R. Thus we have that,

$$p(z|x) = \frac{1}{(2\pi)^{N/2}|R|^{1/2}} \exp\left[-\frac{1}{2}(z - Hx)^T R^{-1}(z - Hx)\right]$$

The minimum variance Bayes' estimate of x given the measurement z_1 is the value of x which maximizes $p(x|z)$. In the case of Gaussian models this can be shown to be the same as the mean of $p(x|z)$.

The *a posteriori* distribution $p(x|z)$ can be thought of as an updated statistical model for the state x. If $p(x), p(z), p(z|x)$ are all Gaussian, then $p(x|z)$ will be Gaussian as well. The mean of this Gaussian will be our updated state estimate, and its covariance will be our updated state covariance. We can obtain these updated values by looking at $p(z|x)p(x)$ ($p(z)$ will not affect the mean or the covariance, so we will leave it out of the calculation that follows).

$$p(z|x)p(x) = \frac{1}{(2\pi)^{N}|R|^{1/2}|P_0|^{1/2}} \exp\left[-\frac{1}{2}\left((z - Hx)^T R^{-1}(z - Hx) + (x - \hat{x}_0)^T P_0^{-1}(x - \hat{x}_0)\right)\right]$$

Since we know that $p(z|x)p(x)/p(z)$ must be Gaussian we know that $p(z|x)p(x)$ must have the form:

$$p(z|x)p(x) = K \exp\left[-\frac{1}{2}(x - \hat{x}_1)^T P_1^{-1}(x - \hat{x}_1)\right]$$

where $\hat{x}_1$ and P_1 are the updated state estimate and covariance estimate respectively and K is some constant factor.

Expanding the quadratic terms in these two formulae yields:

$$\begin{aligned} & x^T P_0^{-1} x - \hat{x}_0^T P_0^{-1} x - x^T P_0^{-1} \hat{x}_0 + \hat{x}_0^T P_0^{-1} \hat{x}_0 + x^T H^T R^{-1} H x - x^T H^T R^{-1} z - z R^{-1} H x + z^T R^{-1} z \\ &= x^T (P_0^{-1} + H^T R^{-1} H) x - x^T (P_0^{-1} \hat{x}_0 + H^T R^{-1} z) - (\hat{x}_0^T P_0^{-1} + z R^{-1} H) x + (\hat{x}_0^T P_0^{-1} \hat{x}_0 + z^T R^{-1} z) \end{aligned}$$

which is equal to:

$$x^T P_1^{-1} x - x^T P_1^{-1} \hat{x}_1 - \hat{x}_1^T P_1^{-1} x + \hat{x}_1^T P_1^{-1} \hat{x}_1$$

Equating terms we see that

$$P_1 = [P_0^{-1} + H_1^T R^{-1} H_1]^{-1}.$$

and

$$\hat{x}_1 = [P_0^{-1} + H_1^T R_1^{-1} H_1]^{-1} (H_1^T R_1^{-1} z_1 + P_0^{-1} \hat{x}_0),$$

The estimate $(\hat{x}_1, P_1)$ can be then used with another measurement z_2 to get a further estimate $(\hat{x}_2, P_2)$, and so on, therefore forming a *recursive filter*: Given $\hat{x}_k$, the estimate of the system state at time t_k, we seek an update estimate, $\hat{x}_{k+1}$, based on the measurement, z_k, which has the form

$$\hat{x}_k = K_k z_k + K_k' \hat{x}_{k-1},$$

where K_k and K_k' are time-varying weight matrices. We can see that

$$K_k = [P_{k-1}^{-1} + H_k^T R_k^{-1} H_k]^{-1} H_k^T R_k^{-1}$$

and

$$K'_k = [P_{k-1}^{-1} + H_k^T R_k^{-1} H_k]^{-1} P_{k-1}^{-1}$$

We can rewrite these equations in another form as follows. Recalling the matrix inversion identity $(AB)^{-1} = B^{-1}A^{-1}$, we can write: $H^T R^{-1} = (RH^{-T})^{-1}$ (where H^{-T} is a shorthand notation for the transpose of the inverse of H). This means that K_k can be rewritten: (we drop the subscripts for the time being, to improve clarity)

$$K = (P^{-1} + H^T R^{-1} H)^{-1} (RH^{-T})^{-1} = (RH^{-T}(P^{-1} + H^T R^{-1} H))^{-1}$$

which is equal to

$$(RH^{-T}P^{-1} + H)^{-1} = ((R + HPH^T)H^{-T}P^{-1})^{-1}$$

and which is, in turn, equal to

$$(H^{-T}P^{-1})^{-1}(R + HPH^T)^{-1} = PH^T(R + HPH^T)^{-1}$$

Therefore

$$K_k = P_{k-1} H_k^T [H_k P_{k-1} H_k^T + R_k]^{-1}$$

We can see that K and K' are related by:

$$K'_k = K_k (R_k H_k^{-T} P_{k-1})$$

We saw earlier that

$$K_k = (R_k H_k^{-T} P_{k-1}^{-1} + H_k)^{-1}$$

Thus

$$K_k (R_k H_k^{-T} P_{k-1}^{-1} + H_k) = I$$

or

$$K_k (R_k H_k^{-T} P_{k-1}) = (I - K_k H_k)$$

Thus we see that

$$K'_k = (I - K_k H_k)$$

Using these new forms for K_k and K'_k the estimate update equation can be rewritten as

$$\hat{x}_k = \hat{x}_{k-1} + K_k [z_k - H_k \hat{x}_{k-1}],$$

with error covariance propagation

$$P_k = [I - K_k H_k] P_{k-1}.$$

The matrix K_k is called the *Kalman gain matrix.*

In the case of nonlinear measurements of the form

$$z = h(x, u, t) + \nu,$$

where u is a control vector, one can linearize h around the current estimate, $\hat{x}$, as follows:

$$h_k(x_k) = h_k(\hat{x}_k) + H_k(\hat{x}_k)(x_k - \hat{x}_k) + \cdots,$$

where

$$H_k(\hat{x}_k) = \left.\frac{\partial h_k(x)}{\partial x}\right|_{x=\hat{x}_k}.$$

Substituting for the H_k's in the equations above we get an *extended Kalman filter.*

To improve things further one can repeatedly calculate $\hat{x}_k$ and K_k, each time linearizing about the most recent estimate. Let $\hat{x}_{k,i}$ denote the i^{th} estimate of $\hat{x}_k$, $i = 1, 2, \ldots, i_{\max}$, with $\hat{x}_{k,0} = \hat{x}_{k-1}$. Then we have an *iterated extended Kalman filter* which takes on the form

$$\begin{aligned} \hat{x}_{k,i+1} &= \hat{x}_{k-1} + K_{k,i}[z_k - h_k(\hat{x}_{k,i}) - H_k(\hat{x}_{k,i})(\hat{x}_{k-1} - \hat{x}_{k,i})] \\ K_{k,i} &= P_k H_k^T(\hat{x}_{k,i})[H_k(\hat{x}_{k,i}) P_k H_k^T(\hat{x}_{k,i}) + R_k]^{-1}, \end{aligned}$$

for $i = 1, 2, \ldots, i_{\max}$. The state estimate error covariance propagation is

$$P_{k+1} = [I - K_{k,i_{\max}} H_k(\hat{x}_{k,i_{\max}})] P_k.$$