THE KALMAN FILTER[1]

The Kalman Filter[2] is a recursive program which estimates the "actual" value of a random variable (r.v.) vector **x** in a set of observations

$$\mathbf{z} = \mathbf{H}\mathbf{x} + \mathbf{v}$$

where **z** are the observations, **H** is the transformation matrix, and **v** is an error vector.

Background. The least squares estimate $\hat{\mathbf{x}}$ which minimizes a cost function

$$J = (\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})^{\mathbf{T}}\mathbf{R}^{-1}(\mathbf{z} - \mathbf{H}\hat{\mathbf{x}})$$

is

$$\hat{\mathbf{x}} = \left[\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1}\mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{z}$$

where **R** is a matrix of weights and the cost function J measures the weighted sum of squares of deviations. This is a purely deterministic solution, obtained by minimizing J by setting its gradient to zero and solving for $\hat{\mathbf{x}}$.

Alternatively, one can argue that the "best" estimate is the value $\hat{\mathbf{x}}$ which maximizes the probability of **z** having occurred, given known statistical properties of **v**. This is called the *maximum likelihood* estimate. This approach postulates any number of possible vectors **x** and *maximizes* the conditional probability $p(\mathbf{z}|\mathbf{x})$.

A third alternative is to use a Bayesian approach, in which the *a posteriori* probability of **x**, given observations **z**, is found from

$$p(\mathbf{x}|\mathbf{z}) = \frac{p(\mathbf{z}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{z})} \tag{1}$$

where $p(\mathbf{x})$ is an assumed *a priori* distribution on **x**. We can assume any *pdf* on **x** we choose; Bayes theorem will yield the appropriate conditional probability $p(\mathbf{x}|\mathbf{z})$, given appropriate distributions on **z**|**x** and **z**.

The distribution on **z**|**x** can be evaluated from the observations, as can that on **z**. The the issue is finding a method to maximize equation (1). By choosing a loss function in the form of a variance

---

[1] [CAL] C. A. La Varre, 29 Farewell Street, Newport, RI 02840-2563, January 1994

[2] [Gelb] Arthur Gelb, ed., *Applied Optimal Estimation*, The MIT Press, 1974, ISBN 0 262-700008-5, *pp*. 103*ff*.

$$L(\tilde{\mathbf{x}}) = [\hat{\mathbf{x}} - \mathbf{x}]^{\mathbf{T}}\mathbf{S}[\hat{\mathbf{x}} - \mathbf{x}]$$

we can then compute a *variance* cost function as the expected value of the variance over all possible values of **x**, given a conditional probability of **x** on **z**:

$$J = \iiint_{\mathbf{x}} [\hat{\mathbf{x}} - \mathbf{x}]^{\mathbf{T}}\mathbf{S}[\hat{\mathbf{x}} - \mathbf{x}]\, p(\mathbf{x}|\mathbf{z})\, dx_1 dx_2 \ldots dx_n \tag{2}$$

where **S** is some positive semidefinite matrix. Minimizing J with respect to **x** gives a *minimum variance* estimate $\hat{\mathbf{x}}$. The minimization is accomplished again by setting the gradient of J to zero and solving for $\hat{\mathbf{x}}$. It is given that the solution of this minimization is

$$\hat{\mathbf{x}} = \left[\mathbf{P}_0^{-1} \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{H}\right]^{-1} \mathbf{H}^{\mathbf{T}}\mathbf{R}^{-1}\mathbf{z} \tag{3}$$

where $\mathbf{P}_0$ is the *a priori* covariance matrix of **x** and Gaussian distributions are assumed for **x** and **v**. This assumption is ameliorated by the observation that

> Most often all we know about the characterization of a given random proces is its autocorrelation function. But there always exists a Gaussian random process possessing the same autocorrelation function; we therefore might as well assume that the given random process is itself Gaussian. [3]

The Kalman Filter. The Kalman filter uses these principles to calculate the optimal value of two matrices $\mathbf{K}_k$ and $\mathbf{K'}_k$ which is used in a recursion formula to calculate the next value of the estimate $\hat{\mathbf{x}}$ in the form

$$\hat{\mathbf{x}}_k^{+} = \mathbf{K'}_k \hat{\mathbf{x}}_k^{-} + \mathbf{K}_k \mathbf{z}_k \tag{4}$$

where $\hat{\mathbf{x}}_k^{+}$ is the *a posteriori* estimate and $\hat{\mathbf{x}}_k^{-}$ is the *a priori* estimate for the *k*-th estimate of **x**.

Each of these vary from the actual value **x** by some estimation error, respectively $\tilde{\mathbf{x}}_k^{+}$ and $\tilde{\mathbf{x}}_k^{-}$:

$$\begin{aligned} \hat{\mathbf{x}}_k^{+} &= \mathbf{x}_k + \tilde{\mathbf{x}}_k^{+} \\ \hat{\mathbf{x}}_k^{-} &= \mathbf{x}_k + \tilde{\mathbf{x}}_k^{-} \end{aligned} \tag{5}$$

The recursive nature of (4) indicates that successive observations $\mathbf{z}_k$ are obtained as

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{6}$$

---

[3] *Ibid., p.* 105.

An assumption of zero mean observation and estimation errors results in

$$\mathbf{K}'_k = \mathbf{I} - \mathbf{K}_k \mathbf{H}_k \tag{7}$$

Substituting (5), (6), and (7) in (4) results in an expression for the *a posteriori* error:

$$\tilde{\mathbf{x}}_k^{+} = \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right] \tilde{\mathbf{x}}_k^{-} + \mathbf{K}_k \mathbf{v}_k \tag{8}$$

and an expression for $\hat{\mathbf{x}}_k^{+}$:

$$\hat{\mathbf{x}}_k^{+} = \hat{\mathbf{x}}_k^{-} + \mathbf{K}_k \left[\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_k^{-}\right] \tag{9}$$

The task then is to find a value for $\mathbf{K}_k$. This is accomplished following equation (2) by minimizing $\mathbf{P}_k^{+}$, the expected variance of the *a posteriori* error $\tilde{\mathbf{x}}_k^{+}$. The error covariance matrix is

$$\mathbf{P}_k^{+} = E\left[\tilde{\mathbf{x}}_k^{+} \tilde{\mathbf{x}}_k^{+\mathbf{T}}\right] \tag{10}$$

Expanding (10) with (8) and substituting

$$E\left[\mathbf{v}_k \mathbf{v}_k^{\mathbf{T}}\right] = \mathbf{R}_k$$

gives an expression for $\mathbf{P}_k^{+}$ which can be minimized with respect to $\mathbf{K}_k$. resulting in

$$\mathbf{K}_{\mathbf{k}} = \mathbf{P}_{\mathbf{k}}^{-} \mathbf{H}_{\mathbf{k}}^{\mathbf{T}} \left[\mathbf{H}_{\mathbf{k}} \mathbf{P}_{\mathbf{k}}^{-} \mathbf{H}_{\mathbf{k}}^{\mathbf{T}} + \mathbf{R}_{\mathbf{k}}\right]^{-1} \tag{11}$$

where the generalized covariance matrix $\mathbf{R}_{\mathbf{k}}$ is estimated on each successive set of observations as[4] by

$$\mathbf{R}_k = \frac{1}{n-1} \mathbf{x}_k \left[\mathbf{I} - \frac{1}{n} \mathbf{1} \cdot \mathbf{1}^{\mathbf{T}}\right] \mathbf{x}^{\mathbf{T}} \tag{12}$$

and $\mathbf{1}$ is the nx1 vector of 1s.

A simpler expression for $\mathbf{P}_k^{+}$ is obtained from (11):

$$\mathbf{P}_k^{+} = \left[\mathbf{I} - \mathbf{K}_k \mathbf{H}_k\right] \mathbf{P}_k^{-} \tag{13}$$

---

[4] [Wichern] Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice Hall, 1988, ISBN 0-13-041146-9, *p.* 111

Strictly, we find the next *a priori* error covariance matrix, $\mathbf{P}_{\mathbf{k+1}}^{-}$ from the state transition matrix $\mathbf{F}_{\mathbf{k}}$, where $\mathbf{x}_{\mathbf{k+1}}^{-} = \mathbf{F}_{\mathbf{k}}\mathbf{x}_{\mathbf{k}}^{+}$ and

$$\mathbf{P}_{\mathbf{k+1}}^{-} = \mathbf{F}_{\mathbf{k}}\mathbf{P}_{\mathbf{k}}^{+}\mathbf{F}_{\mathbf{k}}^{\mathbf{T}} + \mathbf{Q}_{\mathbf{k}} \tag{14}$$

where $\mathbf{Q}_{\mathbf{k}}$ is the covariance of the system noise $\mathbf{w}_{\mathbf{k}}$. In practice we take $\mathbf{F}_{\mathbf{k}} = \mathbf{I}$ and $\mathbf{Q}_{\mathbf{k}} = 0$. Under this process, then

$$\mathbf{P}_{k+1}^{-} = \mathbf{P}_{k}^{+} \tag{15}$$

In summary, given initial conditions for the unknown variables $\mathbf{x}_0$, the error covariance matrix $\mathbf{P}_0$, and the observation error covariance matrix $\mathbf{R}_0$ with a given system model

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_{\boldsymbol{k}} \tag{6}$$

we calculate the *Kalman gain matrix* $\mathbf{K}_k$ with (11) and the updated variable covariance error matrix with(13), then calculate the next variable estimate from (9):

$$\hat{\mathbf{x}}_k^{+} = \hat{\mathbf{x}}_k^{-} + \mathbf{K}_k\left[\mathbf{z}_k - \mathbf{H}_k\hat{\mathbf{x}}_k^{-}\right] \tag{9}$$

We then set $\mathbf{P}_{k+1}^{-} = \mathbf{P}_{k}^{+}$ and continue iteration until (9) converges satisfactorily.