Chapter 11

Tutorial: The Kalman Filter

Tony Lacey.

11.1 Introduction

The Kalman filter [1] has long been regarded as the optimal solution to many tracking and data prediction tasks, [2]. Its use in the analysis of visual motion has been documented frequently. The standard Kalman filter derivation is given here as a tutorial exercise in the practical use of some of the statistical techniques outlied in previous sections. The filter is constructed as a mean squared error minimiser, but an alternative derivation of the filter is also provided showing how the filter relates to maximum likelihood statistics. Documenting this derivation furnishes the reader with further insight into the statistical constructs within the filter.

The purpose of filtering is to extract the required information from a signal, ignoring everything else. How well a filter performs this task can be measured using a cost or loss function. Indeed we may define the goal of the filter to be the minimisation of this loss function.

11.2 Mean squared error

Many signals can be described in the following way;

$$y_k = a_k x_k + n_k \tag{11.1}$$

where; y_k is the time dependent observed signal, a_k is a gain term, x_k is the information bearing signal and n_k is the additive noise.

The overall objective is to estimate x_k. The difference between the estimate of $\hat{x}_k$ and x_k itself is termed the error;

$$f(e_k) = f(x_k - \hat{x}_k) \tag{11.2}$$

The particular shape of $f(e_k)$ is dependent upon the application, however it is clear that the function should be both positive and increase monotonically [3]. An error function which exhibits these characteristics is the squared error function;

$$f(e_k) = (x_k - \hat{x}_k)^2 \tag{11.3}$$

Since it is necessary to consider the ability of the filter to predict many data over a period of time a more meaningful metric is the expected value of the error function;

$$lossfunction = E\left(f\left(e_k\right)\right) \tag{11.4}$$

This results in the *mean squared error* (MSE) function;

$$\epsilon\left(t\right) = E\left(e_k^2\right) \tag{11.5}$$

11.3 Maximum likelihood

The above derivation of mean squared error, although intuitive is somewhat heuristic. A more rigorous derivation can be developed using *maximum likelihood statistics*. This is achieved by redefining the goal of the filter to finding the $\hat{x}$ which maximises the probability or likelihood of y. That is;

$$max\left[P(y|\hat{x})\right] \tag{11.6}$$

Assuming that the additive random noise is Gaussian distributed with a standard deviation of σ_k gives;

$$P(y_k|\hat{x}_k) = K_k exp - \left(\frac{(y_k - a_k\hat{x}_k)^2}{2\sigma_k^2}\right) \tag{11.7}$$

where K_k is a normalisation constant. The maximum likelihood function of this is;

$$P(y|\hat{x}) = \prod_k K_k exp - \left(\frac{(y_k - a_k\hat{x}_k)^2}{2\sigma_k^2}\right) \tag{11.8}$$

Which leads to;

$$logP(y|\hat{x}) = -\frac{1}{2}\sum_k\left(\frac{(y_k - a_k\hat{x}_k)^2}{\sigma_k^2}\right) + constant \tag{11.9}$$

The driving function of equation 11.9 is the MSE, which may be maximised by the variation of $\hat{x}_k$. Therefore the mean squared error function is applicable when the expected variation of y_k is best modelled as a Gaussian distribution. In such a case the MSE serves to provide the value of $\hat{x}_k$ which maximises the likelihood of the signal y_k.

In the following derivation the *optimal* filter is defined as being that filter, from the set of all possible filters which minimises the mean squared error.

11.4 Kalman Filter Derivation

Before going on to discuss the Kalman filter the work of Norbert Wiener [4], should first be acknowledged . Wiener described an optimal *finite impulse response* (FIR) filter in the mean squared error sense. His solution will not be discussed here even though it has much in common with the Kalman filter. Suffice to say that his solution uses both the auto correlation and the cross correlation of the received signal with the original data, in order to derive an impulse response for the filter.

Kalman also presented a prescription of the optimal MSE filter. However Kalman's prescription has some advantages over Weiner's; it sidesteps the need to determine the impulse response of the filter, something which is poorly suited to numerical computation. Kalman described his filter using state space techniques, which unlike Wiener's perscription, enables the filter to be used as either a smoother, a filter or a predictor. The latter of these three, the ability of the Kalman filter to be used to predict data has proven to be a very useful function. It has lead to the Kalman filter being applied to a wide range of tracking and navigation problems. Defining the filter in terms of state space methods also simplifies the implementation of the filter in the discrete domain, another reason for its widespread appeal.

11.5 State space derivation

Assume that we want to know the value of a variable within a process of the form;

$$x_{k+1} = \Phi x_k + w_k \tag{11.10}$$

where; x_k is the state vector of the process at time k, (nx1); Φ is the state transition matrix of the process from the state at k to the state at $k+1$, and is assumed stationary over time, (nxm); w_k is the associated white noise process with known covariance, (nx1).

Observations on this variable can be modelled in the form;

$$z_k = Hx_k + v_k \tag{11.11}$$

where; z_k is the actual measurement of x at time k, (mx1); H is the noiseless connection between the state vector and the measurement vector, and is assumed stationary over time (mxn); v_k is the associated measurement error. This is again assumed to be a white noise process with known covariance and has zero cross-correlation with the process noise, (mx1).

As was shown in section **??** for the minimisation of the MSE to yield the optimal filter it must be possible to correctly model the system errors using Gaussian distributions. The covariances of the two noise models are assumed stationary over time and are given by;

$$Q = E\left[w_k w_k^T\right] \tag{11.12}$$

$$R = E\left[v_k v_k^T\right] \tag{11.13}$$

The mean squared error is given by 11.5. This is equivalent to;

$$E\left[e_k e_k^T\right] = P_k \tag{11.14}$$

where; P_k is the error covariance matrix at time k, (nxn).

Equation 11.14 may be expanded to give;

$$P_k = E\left[e_k e_k^T\right] = E\left[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^T\right] \tag{11.15}$$

Assuming the prior estimate of $\hat{x}_k$ is called $\hat{x}'_k$, and was gained by knowledge of the system. It posible to write an update equation for the new estimate, combing the old estimate with measurement data thus;

$$\hat{x}_k = \hat{x}'_k + K_k(z_k - H\hat{x}'_k) \tag{11.16}$$

where; K_k is the Kalman gain, which will be derived shortly. The term $z_k - H\hat{x}'_k$ in eqn. 11.16 is known as the *innovation* or *measurement residual*;

$$i_k = z_k - H\hat{x}_k \tag{11.17}$$

Substitution of 11.11 into 11.16 gives;

$$\hat{x}_k = \hat{x}'_k + K_k (H x_k + v_k - H \hat{x}'_k) \tag{11.18}$$

Substituting 11.18 into 11.15 gives;

$$\begin{aligned} P_k = E\ [&[(I - K_k H)(x_k - \hat{x}'_k) - K_k v_k] \\ &[(I - K_k H)(x_k - \hat{x}'_k) - K_k v_k]^T\Big] \end{aligned} \tag{11.19}$$

At this point it is noted that $x_k - \hat{x}'_k$ is the error of the prior estimate. It is clear that this is uncorrelated with the measurement noise and therefore the expectation may be re-written as;

$$\begin{aligned} P_k &= (I - K_k H) E\left[(x_k - \hat{x}'_k)(x_k - \hat{x}'_k)^T\right](I - K_k H) \\ &+ K_k E\left[v_k v_k^T\right] K_k^T \end{aligned} \tag{11.20}$$

Substituting equations 11.13 and 11.15 into 11.19 gives;

$$P_k = (I - K_k H) P'_k (I - K_k H)^T + K_k R K_k^T \tag{11.21}$$

where P'_k is the prior estimate of P_k.

Equation 11.21 is the error covariance update equation. The diagonal of the covariance matrix contains the mean squared errors as shown;

$$P_{kk} = \begin{bmatrix} E\left[e_{k-1} e_{k-1}^T\right] & E\left[e_k e_{k-1}^T\right] & E\left[e_{k+1} e_{k-1}^T\right] \\ E\left[e_{k-1} e_k^T\right] & E\left[e_k e_k^T\right] & E\left[e_{k+1} e_k^T\right] \\ E\left[e_{k-1} e_{k+1}^T\right] & E\left[e_k e_{k+1}^T\right] & E\left[e_{k+1} e_{k+1}^T\right] \end{bmatrix} \tag{11.22}$$

The sum of the diagonal elements of a matrix is the *trace* of a matrix. In the case of the error covariance matrix the trace is the sum of the mean squared errors. Therefore the mean squared error may be minimised by minimising the trace of P_k which in turn will minimise the trace of P_{kk}.

The trace of P_k is first differentiated with respect to K_k and the result set to zero in order to find the conditions of this minimum.

Expansion of 11.21 gives;

$$P_k = P'_k - K_k H P'_k - P'_k H^T K_k^T + K_k \left(H P'_k H^T + R\right) K_k^T \tag{11.23}$$

Note that the trace of a matrix is equal to the trace of its transpose, therefore it may written as;

$$T\left[P_k\right] = T\left[P'_k\right] - 2T\left[K_k H P'_k\right] + T\left[K_k \left(H P'_k H^T + R\right) K_k^T\right] \tag{11.24}$$

where; $T\left[P_k\right]$ is the trace of the matrix P_k.

Differentiating with respect to K_k gives;

$$\frac{dT\left[P_k\right]}{dK_k} = -2(H P'_k)^T + 2K_k \left(H P'_k H^T + R\right) \tag{11.25}$$

Setting to zero and re-arranging gives;

$$\left(HP'_k\right)^T \;=\; K_k\left(HP'_kH^T \;+\; R\right) \tag{11.26}$$

Now solving for K_k gives;

$$K_k \;=\; P'_kH^T\left(HP'_kH^T \;+\; R\right)^{-1} \tag{11.27}$$

Equation 11.27 is the Kalman gain equation. The innovation, i_k defined in eqn. 11.17 has an associated measurement prediction covariance. This is defined as;

$$S_k \;=\; HP'_kH^T \;+\; R \tag{11.28}$$

Finally, substitution of equation 11.27 into 11.23 gives;

$$\begin{aligned} P_k &= P'_k \;-\; P'_kH^T\left(HP'_kH^T \;+\; R\right)^{-1}HP'_k \\ &= P'_k \;-\; K_kHP'_k \\ &= \left(I \;-\; K_kH\right)P'_k \end{aligned} \tag{11.29}$$

Equation 11.29 is the update equation for the error covariance matrix with optimal gain. The three equations 11.16, 11.27, and 11.29 develop an estimate of the variable x_k. State projection is achieved using;

$$\hat{x}'_{k+1} \;=\; \Phi\hat{x}_k \tag{11.30}$$

To complete the recursion it is necessary to find an equation which projects the error covariance matrix into the next time interval, $k+1$. This is achieved by first forming an expressions for the prior error;

$$\begin{aligned} e'_{k+1} &= x_{k+1} \;-\; \hat{x}'_{k+1} \\ &= \left(\Phi x_k \;+\; w_k\right) \;-\; \Phi\hat{x}_k \\ &= \Phi e_k \;+\; w_k \end{aligned} \tag{11.31}$$

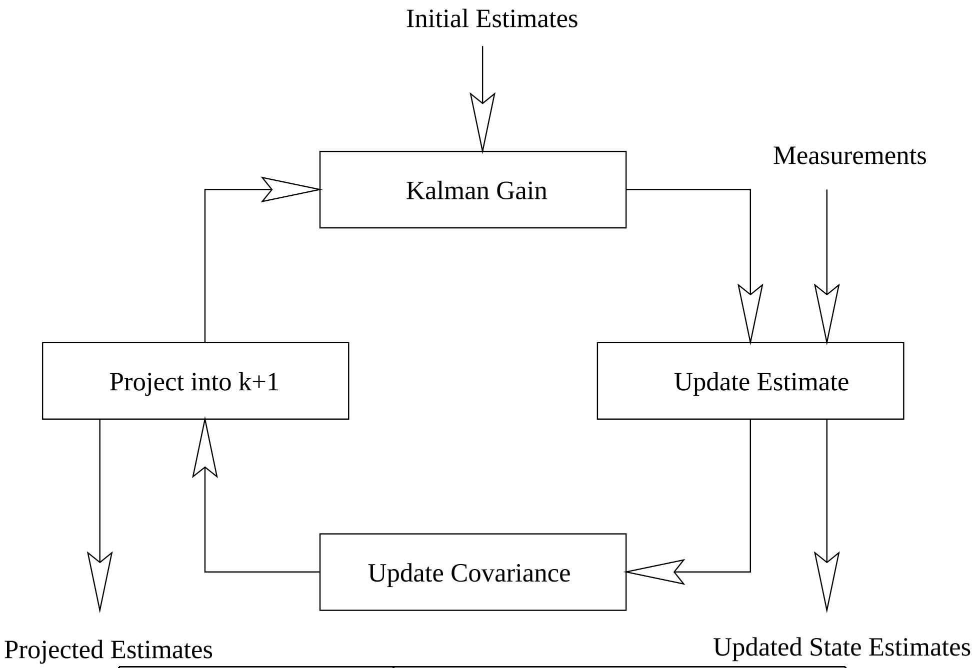
Extending equation 11.15 to time $k+1$;

$$P'_{k+1} \;=\; E\left[e'_{k+1}e^{T\prime}_{k+1}\right] \;=\; E\left[\left(\Phi e_k \;+\; w_k\right)\left(\Phi e_k \;+\; w_k\right)^T\right] \tag{11.32}$$

Note that e_k and w_k have zero cross-correlation because the noise w_k actually accumulates between k and $k+1$ whereas the error e_k is the error up until time k. Therefore;

$$\begin{aligned} P'_{k+1} &= E\left[e'_{k+1}e^{T\prime}_{k+1}\right] \\ &= E\left[\Phi e_k\left(\Phi e_k\right)^T\right] \;+\; E\left[w_kw_k^T\right] \\ &= \Phi P_k\Phi^T \;+\; Q \end{aligned} \tag{11.33}$$

This completes the recursive filter. The algorithmic loop is summarised in the diagram of figure 11.5.



Description	Equation
Kalman Gain	$K_k = P_k' H^T \left(H P_k' H^T + R \right)^{-1}$
Update Estimate	$\hat{x}_k = \hat{x}_k' + K_k \left(z_k - H \hat{x}_k' \right)$
Update Covariance	$P_k = \left(I - K_k H \right) P_k'$
Project into $k+1$	$\hat{x}_{k+1}' = \Phi \hat{x}_k$ $P_{k+1} = \Phi P_k \Phi^T + Q$

Figure 11.1: Kalman Filter Recursive Algorithm

11.6 The Kalman filter as a chi-square merit function

The objective of the Kalman filter is to minimise the mean squared error between the actual and estimated data. Thus it provides the best estimate of the data in the mean squared error sense. This being the case it should be possible to show that the Kalman filter has much in common with the *chi-square*. The chi-square merit function is a maximum likelihood function, and was derived earlier, equation 11.9. It is typically used as a criteria to fit a set of model parameters to a model a process known as *least squares* fitting. The Kalman filter is commonly known as a *recursive least squares* (RLS) fitter. Drawing similarities to the chi-square merit function will give a different perspective on what the Kalman filter is doing.

The chi-square merit function is;

$$\chi^2 = \sum_{i=1}^{k} \left[\frac{z_i - h(a_i, x)}{\sigma_i} \right]^2 \tag{11.34}$$

where; z_i is the measured value; h_i is the data model with parameters x, assumed linear in a; σ_i is the variance associated with the measured value.

The optimal set of parameters can then be defined as that which minimises the above function. Expanding out the variance gives;

$$\chi^2 = \sum_{i=1}^{k} \frac{1}{\sigma_i \sigma_i} \left[z_i - h\left(a_i, x\right) \right]^2 \tag{11.35}$$

Representing the chi-square in vector form and using notation from the earlier Kalman derivation;

$$\chi_k^2 = \left[z_k - h\left(a, x_k\right) \right] R^{-1} \left[z_k - h\left(a, x_k\right) \right]^T \tag{11.36}$$

where; R^{-1} is the matrix of inverse squared variances, i.e. $1/\sigma_i \sigma_i$.

The above merit function is the merit function associated with the latest, kth, measurement and provides a measure of how accurately the model predicted this measurement. Given that the inverse model covariance matrix is known up to time k, the merit function up to time k may be re-written as;

$$\chi_{k-1}^2 = \left(x_{k-1} - \hat{x}_{k-1}\right) P_{k-1}^{\prime -1} \left(x_{k-1} - \hat{x}_{k-1}\right)^T \tag{11.37}$$

To combine the new data with the previous, fitting the model parameters so as to minimise the overall chi-square function, the merit function becomes the summation of the two;

$$\chi^2 = \left(x_{k-1} - \hat{x}_{k-1}\right) P_{k-1}^{\prime -1} \left(x_{k-1} - \hat{x}_{k-1}\right)^T + \left[z_k - h\left(a, x_k\right) \right] R^{-1} \left[z_k - h\left(a, x_k\right) \right]^T \tag{11.38}$$

Where the first derivative of this is given by;

$$\frac{d\chi^2}{dx} = 2P_{k-1}^{\prime -1} \left(x_{k-1} - \hat{x}_{k-1}\right) - 2\nabla_x h\left(a, x_k\right)^T R^{-1} \left[z_k - h\left(a, x_k\right) \right] \tag{11.39}$$

The model function $h(a, x_k)$ with parameters fitted from information to date, may be considered as;

$$h(a, x_k) = h\left(a, \left(\hat{x}_k + \Delta x_k\right)\right) \tag{11.40}$$

where $\Delta x_k = x_k - \hat{x}_k$. The Taylor series expansion of the model function to first order is;

$$h\left(\hat{x}_k + \Delta x\right) = h\left(\hat{x}_k\right) + \Delta x \nabla_x h\left(\hat{x}_k\right) \tag{11.41}$$

Substituting this result into the derivative equation 11.39 gives;

$$\begin{aligned} \frac{d\chi^2}{dx} &= 2P_k^{\prime -1} \left(x_k - \hat{x}_k\right) \\ &- 2\nabla_x h\left(a, \hat{x_k}\right)^T R^{-1} \left[z_k - h\left(a, \hat{x_k}\right) - \left(x_k - \hat{x}_k\right) \nabla_x h\left(a, \hat{x_k}_k\right) \right] \end{aligned} \tag{11.42}$$

It is assumed that the estimated model parameters are a close approximation to the actual model parameters. Therefore it may be assumed that the derivatives of the actual model and the estimated model are the same. Further, for a system which is linear in a the model derivative is constant and may be written as;

$$\nabla_x h\left(a, x_k\right) = \nabla_x h\left(a, \hat{x}_k\right) = H \tag{11.43}$$

Substituting this into equation 11.39 gives;

$$\frac{d\chi^2}{dx} = 2P_k^{\prime -1} \Delta x_k + 2H^T R^{-1} H \Delta x_k - 2H^T R^{-1} \left[z_k - h\left(a, \hat{x}_k\right) \right] \tag{11.44}$$

Re-arranging gives;

$$\frac{d\chi^2}{dx} = 2\left[P_k^{\prime -1} + H^T R^{-1} H\right]\Delta x_k - 2H^T R^{-1}\left[z_k - h\left(a, \hat{x}_k\right)\right] \tag{11.45}$$

For a minimum the derivative is zero, rearrange in terms of Δx_k gives;

$$\Delta x_k = \left[P_k^{\prime -1} + H^T R^{-1} H\right]^{-1} H^T R^{-1}\left[z_k - h\left(a, \hat{x}_k\right)\right] \tag{11.46}$$

$$x = \hat{x}_k + \left[P_k^{\prime -1} + H^T R^{-1} H\right]^{-1} H^T R^{-1}\left[z_k - h\left(a, \hat{x}\right)\right] \tag{11.47}$$

Comparison of equation 11.47 to 11.16 allows the gain, K_k to be identified as;

$$K_k = \left[P_k^{\prime -1} + H^T R^{-1} H\right]^{-1} H^T R^{-1} \tag{11.48}$$

Giving a parameter update equation of the form;

$$x_k = \hat{x_k} + K_k\left[z_k - h\left(a, \hat{x}_k\right)\right] \tag{11.49}$$

Equation 11.49 is identical to 11.16 and describes the improvement of the parameter estimate using the error between measured and model projected values.

11.7 Model covariance update

The model parameter covariance has been considered in its inverted form where it is known as the *information matrix*[1]. It is possible to formulate an alternative update equation for the covariance matrix using standard error propogation;

$$P_k^{-1} = P_k^{\prime -1} + H R^{-1} H^T \tag{11.50}$$

It is possible to show that the covariance updates of equation 11.50 and equation 11.29 are equivalent. This may be achieved using the identity $P_k \times P_k^{-1} = I$. The original, eqn 11.29 and alternative, eqn 11.50 forms of the covariance update equations are;

$P_k = \left(I - K_k H\right) P_k^{\prime}$ and $P_k^{-1} = P_k^{\prime -1} + H R^{-1} H^T$

Therefore;

$$\left(I - K_k H\right) P_k^{\prime} \times P_k^{\prime -1} + H R^{-1} H^T = I \tag{11.51}$$

Substituting for K_k gives;

$$\begin{aligned}
&\left[P_k^{\prime} - P_k^{\prime} H^T \left(H P_k^{\prime} H^T + R\right)^{-1} H P_k^{\prime}\right]\left[P_k^{\prime -1} + H^T R^{-1} H\right] \\
= \;& I - P_k^{\prime} H^T \left[\left(H P_k^{\prime} H^T + R\right)^{-1}\right. \\
- \;& \left. R^{-1} + \left(H P_k^{\prime} H^T + R\right)^{-1} H P_k^{\prime} H^T R^{-1}\right] H \\
= \;& I - P_k^{\prime} H^T \left[\left(H P_k^{\prime} H^T + R\right)^{-1}\left(I + H P_k^{\prime} H^T R^{-1}\right) - R^{-1}\right] H \\
= \;& I - P_k^{\prime} H^T \left[R^{-1} - R^{-1}\right] H \\
= \;& I
\end{aligned} \tag{11.52}$$

[1]when the Kalman filter is built around the information matrix it is known as the *information filter*