Chapter 6

Kalman Filtering

The Kalman Filter is a discrete-time dynamical system of hidden random variables $\mathbf{x}$ and observation random variables $\mathbf{z}$ subject to Gaussian noise. The generative model for $\mathbf{x}$ and $\mathbf{z}$ is given by:

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{q}_t \tag{6.1}$$
$$\mathbf{z}_t = H\mathbf{x}_t + \mathbf{r}_t, \tag{6.2}$$

where A is a $k \times k$ state transition matrix and H is a $p \times k$ observation matrix. The variables $\mathbf{q} \sim N(0, Q)$ and $\mathbf{r} \sim N(0, R)$ are Gaussian white noise sources with mean zero and covariances Q and R, respectively. Typically, the goal of Kalman Filtering is to estimate hidden state sequences $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)$ from observation data $\mathbf{z} = (z_1, \ldots, z_T)$ and given model parameters A, H, Q, and R.

There exists an iterative scheme for the computation of the optimal hidden state sequence $\hat{\mathbf{x}}$ and its confidence. We express the optimal state $\hat{\mathbf{x}}_k$ and underlying covariance P_k given all measurements up to time k as a function of the optimal state $\hat{\mathbf{x}}_{k-1}$ and covariances P_{k-1} at the previous time step. The state probability conditional on the partial observation sequence $\mathbf{Z}_k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$ can be expressed by Bayes' Theorem as:

$$p(\mathbf{x}_k|Z_k) = \frac{p(z_k|\mathbf{x}_k, Z_{k-1})p(\mathbf{x}_k|Z_{k-1})}{p(z_k|Z_{k-1})} = \frac{p(z_k|\mathbf{x}_k, Z_{k-1})}{p(z_k|Z_{k-1})} \int d\mathbf{x}_{k-1} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|Z_{k-1}), \tag{6.3}$$

where

$$\begin{aligned} p(z_k|\mathbf{x}_k, Z_{k-1}) &= N(H\mathbf{x}_k, R) \\ p(\mathbf{x}_k|\mathbf{x}_{k-1}) &= N(A\mathbf{x}_{k-1}|Q) \\ p(\mathbf{x}_{k-1}|Z_{k-1}) &= N(\hat{\mathbf{x}}_{k-1}, P_{k-1}). \end{aligned}$$

The iterative scheme is obvious from Equation 6.3: the mean $\hat{\mathbf{x}}_k$ and the covariance P_k of $p(\mathbf{x}_k|Z_k)$ are expressed as a function of the mean $\hat{\mathbf{x}}_{k-1}$ and the covariance P_{k-1} of $p(\mathbf{x}_{k-1}|Z_{k-1})$. To compute the mean and covariance of $p(\mathbf{x}_k|Z_k)$ we can leave out the denominator, as it does neither depend on $\mathbf{x}_k$ nor on $\mathbf{x}_{k-1}$. The calculation leads to the following result:

$$\hat{\mathbf{x}}_k = A\hat{\mathbf{x}}_{k-1} + K(z_k - HA\hat{\mathbf{x}}_{k-1}) \tag{6.4}$$
$$P_k = (\mathrm{I} - KH)\overline{P}_k, \tag{6.5}$$

where

$$\begin{aligned} K &= \overline{P}_k H^T (R + H\overline{P}_k H^T)^{-1} \\ \overline{P}_k &= AP_{k-1}A^T + Q, \end{aligned}$$

and I is the identity matrix. The matrix $\overline{P}_k$ is known as the a priori covariance estimate of $\hat{\mathbf{x}}_k$ given $\hat{\mathbf{x}}_{k-1}$, but not $\mathbf{z}_k$; and the matrix K is known as the Kalman gain.

Note that the Kalman filtering problem 6.1 can be easily generalized to a case where A and H are time dependent (H_t, A_t) and where there is an additional (known) external input $\mathbf{b}_t$ ($\mathbf{x}_t = A_t\mathbf{x}_{t-1} + \mathbf{b}_t + \mathbf{q}_t$).

Proof:

To derive Equation 6.4 we can proceed in two ways. The first consists of brute force integration of Equation 6.3 and is a tricky exercise in matrix algebra, and the second consists of smart guessing and doing a simpler calculation the physicist's way. Let us first summarize the straight-froward calculation. We start with the integral:

$$\int d\mathbf{x}_{k-1} p(\mathbf{x}_k|\mathbf{x}_{k-1}) p(\mathbf{x}_{k-1}|Z_{k-1}) = \alpha \int d\mathbf{x}_{k-1} e^{\Phi/2}, \tag{6.6}$$

where α is a constant that guarantees proper normalization and Φ is

$$\Phi = -(\mathbf{x}_k - A\mathbf{x}_{k-1})^T Q^{-1}(\mathbf{x}_k - A\mathbf{x}_{k-1}) - (\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^T P_{k-1}^{-1}(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1}).$$

To calculate the integral, we complete the square of $\mathbf{x}_{k-1}$ and rewrite the exponent as:

$$\Phi = -(\mathbf{x}_{k-1} - M\mathbf{x}_k - N\hat{\mathbf{x}}_{k-1})^T U(\mathbf{x}_{k-1} - M\mathbf{x}_k - N\hat{\mathbf{x}}_{k-1}) \tag{6.7}$$

$$+\mathbf{x}_k^T M^T U M\mathbf{x}_k + \hat{\mathbf{x}}_{k-1}^T N^T U N\hat{\mathbf{x}}_{k-1} + \mathbf{x}_k^T M^T U N\hat{\mathbf{x}}_{k-1} + \hat{\mathbf{x}}_{k-1}^T N^T U M\mathbf{x}_k, \tag{6.8}$$

where

$$U = A^T Q^{-1} A, \quad UM = A^T Q^{-1}, \quad UN = P_{k-1}^{-1}. \tag{6.9}$$

Integration is now possible, the first term in Equation 6.7 integrates to a factor of one and we find

$$p(\mathbf{x}_k|Z_k) = \beta e^{\Psi/2}, \tag{6.10}$$

where

$$\begin{aligned} \Psi = &-(z_k - H\mathbf{x}_k)^T R^{-1}(z_k - H\mathbf{x}_k) - \mathbf{x}_k^T(Q^{-1} - M^T U M)\mathbf{x}_k \\ &-\hat{\mathbf{x}}_{k-1}^T(P_{k-1}^{-1} - N^T U N)\hat{\mathbf{x}}_{k-1} + \mathbf{x}_k^T M^T U N\hat{\mathbf{x}}_{k-1} + \hat{\mathbf{x}}_{k-1}^T N^T U M\mathbf{x}_k, \end{aligned}$$

and β is a suitable normalization constant. Again we complete the square, this time for $\mathbf{x}_k$ and find

$$\Psi = -(\mathbf{x}_k - Dz_k - E\hat{\mathbf{x}}_{k-1})^T P_K^{-1}(\mathbf{x}_k - Dz_k - E\hat{\mathbf{x}}_{k-1}) + \Omega, \tag{6.11}$$

where

$$P_k^{-1} = H^T R^{-1} H + Q^{-1} - M^T U M = H^T R^{-1} H + \overline{P}_k^{-1} \tag{6.12}$$

$$P_k^{-1} D = H^T R^{-1} \tag{6.13}$$

$$P_k^{-1} E = M^T U N = M^T U U^{-1} U N = \overline{P}_k^{-1} A, \tag{6.14}$$

$$\overline{P}_k = (AP_{k-1}A^T + Q)$$

and where Ω contains all terms not related to $\mathbf{x}_k$ and therefore do not contribute to either the the mean or the covariance of $p(\mathbf{x}_k|Z_k)$. Finally, after solving equations 6.13 and 6.14 we find

$$\begin{aligned} E &= (\overline{P}_k H^T R^{-1} H + \mathrm{I})^{-1} A \\ D &= \overline{P}_k H^T (R + H\overline{P}_k H^T)^{-1} = K, \end{aligned}$$

which, in combination with Equation 6.11 and after some algebra, leads to the desired result:

$$\begin{aligned} \hat{\mathbf{x}}_k &= (A - KHA)\hat{\mathbf{x}}_{k-1} + K z_k \\ P_k &= (\mathrm{I} - KH)\overline{P}_k. \end{aligned}$$

Note that there exists a simpler way of finding the same result, based on Equation 6.4 used as an Ansatz. This Ansatz is a smart choice because the dependences of $\hat{\mathbf{x}}_k$ on z_k and $\hat{\mathbf{x}}_{k-1}$ must be linear, and because the a priory estimate $A\hat{\mathbf{x}}_k$ of $\hat{\mathbf{x}}_{k+1}$ is subject to a correction proportional to $z_k - A\hat{\mathbf{x}}_k$ after knowledge of z_k (with unknown proportionality matrix K). Using this Ansatz we find that

$$P_k = \langle (\mathbf{x}_k - \hat{\mathbf{x}}_k)(\mathbf{x}_k - \hat{\mathbf{x}}_k)^T \rangle = \langle e_k e_k^T \rangle, \tag{6.15}$$

with

$$\begin{aligned} e_k &= A\mathbf{x}_{k-1} + q_k - \hat{\mathbf{x}}_k & (6.16)\\ &= A\mathbf{x}_{k-1} + q_k - A\hat{\mathbf{x}}_{k-1} - K(H\mathbf{x}_k + r_k - HA\hat{\mathbf{x}}_{k-1}) & (6.17)\\ &= A\mathbf{x}_{k-1} + q_k - A\hat{\mathbf{x}}_{k-1} - K[H(A\mathbf{x}_{k-1} + q_k) + r_k - HA\hat{\mathbf{x}}_{k-1}]. & (6.18) \end{aligned}$$

By evaluating the expression for P_k, using the definition $P_{k-1} = \langle (\mathbf{X}_{k-1} - \hat{\mathbf{x}}_{k-1})(\mathbf{x}_{k-1} - \hat{\mathbf{x}}_{k-1})^T \rangle$ and of $\overline{P}_k$ in Equation 6.5, we find:

$$P_k = (\mathrm{I} - KH)\overline{P}_k(\mathrm{I} - KH)^T + KRK^T \tag{6.19}$$

The Kalman gain K is the matrix that minimizes the square prediction error (the trace of the covariance matrix P_k):

$$K^* = \arg\min_K \langle e_k^T e_k \rangle. \tag{6.20}$$

Setting the derivative of the error to zero ($\nabla_K \langle e_k^T e_k \rangle = 0$) and using the useful trace identities $\nabla_X \mathrm{tr}(AXB) = A^T B^T$ and $\nabla_X \mathrm{tr}(XAX^T) = XA^T + XA$ leads to the desired result:

$$K = \overline{P}_k H^T (R + H\overline{P}_k H^T)^{-1}, \tag{6.21}$$

and, after replacing this identity into Equation 6.19 leads to Equation 6.5.

6.1 Simple examples

6.1.1 Link to Bayesian estimation

The initial conditions $\hat{\mathbf{x}}_0$ and P_0 can be thought of as prior information available to the filter estimation problem. For example, assume we have a one-dimensional and stationary estimation problem: $A_k = 1$, $q_t = 0$, $H_k = 1$. With these parameters, Kalman estimation simplifies to:

$$\begin{aligned} \overline{P}_1 &= P_0 = \sigma_0^2 & (6.22)\\ K &= \frac{\sigma_0^2}{\sigma^2 + \sigma_0^2}, & (6.23) \end{aligned}$$

and thus

$$\begin{aligned} \hat{x}_1 &= \frac{\sigma^2}{\sigma^2+\sigma_0^2}\hat{x}_0 + \frac{\sigma_0^2}{\sigma^2+\sigma_0^2} z_1 \\ P_k &= \sigma_1^2 = \frac{\sigma^2\sigma_0^2}{\sigma^2+\sigma_0^2}. \end{aligned} \tag{6.24}$$

It is interesting to note that Equation 6.24 exactly corresponds to maximum posterior estimation of the mean of a Gaussian random variable with Gaussian prior, Equation 5.46. If we have little prior information ($\sigma \ll \sigma_0$), then we trust the data and our updated estimate of the mean is $\hat{x}_1 \simeq z_1$. If on the other hand, our prior is good ($\sigma \gg \sigma_0$), then we trust it more than the data, $\hat{x}_1 \simeq \hat{x}_0$. No matter what we trust, the updated variance is always smaller than the smaller variance, $\sigma_1^2 < \min(\sigma_0^2, \sigma^2)$.