Chapter 6

STATE ESTIMATION

In practice, it is unrealistic to assume that all the disturbances and states can be measured. In general, one must *estimate* the states from the measured input / output sequences. This is called *state estimation*.

Let us assume the standard state-space system description we developed in the previous chapter:

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) + \varepsilon_1(k) \\ y(k) &= Cx(k) + \varepsilon_2(k) \end{aligned} \tag{6.1}$$

$\varepsilon_1(k)$ and $\varepsilon_2(k)$ are mutually independent white noise sequences of covariances R_1 and R_2 respectively. The problem of state estimation is to estimate $x(k+i),\ i \geq 0$, given $\{y(j), u(j), j \leq k\}$ (i.e., inputs and outputs up to the kth sample time). Estimating $x(k+i)$ for $i > 0$ is called *prediction*, while that for $i = 0$ is called *filtering*. Some applications require $x(k+i), i < 0$ to be estimated and this is referred to as *smoothing*.

There are many state estimation techniques, ranging from a simple open-loop observer to more sophisticated optimal observers like the Kalman filter. Since state estimation is an integral part of a model predictive controller, we examine some popular techniques in this chapter. These techniques are also useful for parameter estimation problems, such as those arise in system identification discussed in the next chapter.

An extremely important, but often overlooked point is the importance of correct disturbance modelling. Simply adding white noises into the state and output equations, as often done by those who misunderstand the role of white noise in a standard system description, can result in extreme bias. In general, to obtain satisfactory results, disturbances (or their effects) must be modelled as appropriate stationary / nonstationary stochastic processes and the system equations must be augmented with their describing stochastic equations before a state estimation technique is applied.

6.1 LINEAR OBSERVER STRUCTURE

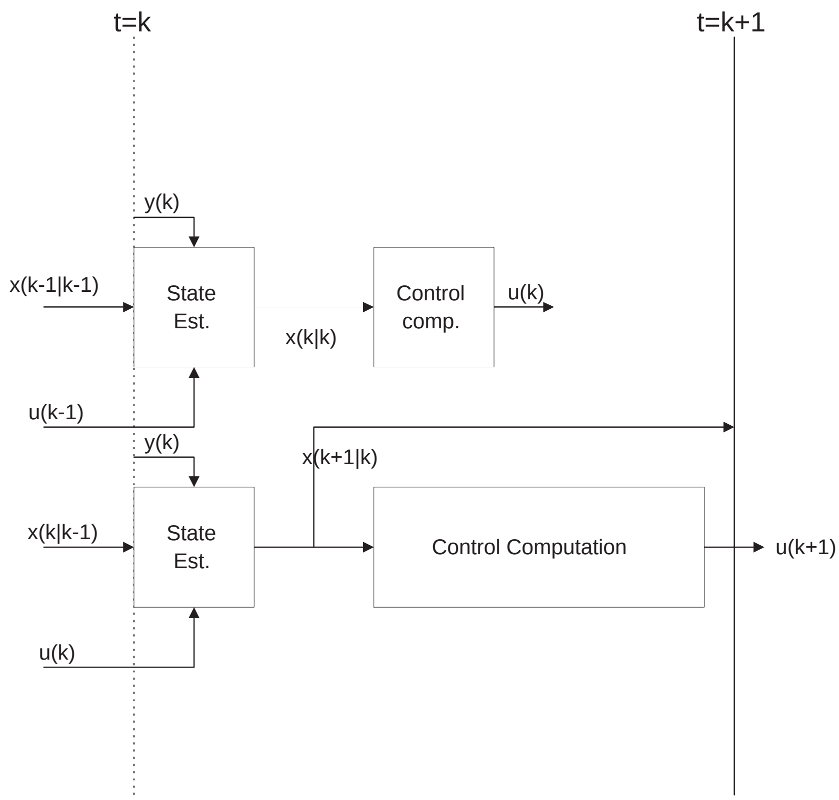
A linear observer for system (6.1) takes the form of

$$\begin{aligned} \hat{x}(k|k-1) &= A\hat{x}(k-1|k-1) + Bu(k-1) \\ \hat{x}(k|k) &= \hat{x}(k|k-1) + K(y(k) - \hat{x}(k|k-1)) \end{aligned} \tag{6.2}$$

In the above, $x(i|j)$ represents an estimate of $x(i)$ constructed using measurements up to time j. The above equations can be used to construct the filtered estimate $\hat{x}(k|k)$ recursively.

Comments:

- In some applications, one may need to compute the one-step-ahead prediction $\hat{x}(k+1|k)$ rather than the filtered estimate. For instance, in a control application, the control computation may require one sample period to complete and in this case, one may want to compute $\hat{x}(k+1|k)$ at time k in order to begin the computation for the control input $u(k+1)$.



Notice that (6.2) can be rewritten as a one-step-ahead predictor simply by switching the order of the two equations:

$$\begin{aligned} \hat{x}(k|k) &= \hat{x}(k|k-1) + K(y(k) - \hat{x}(k|k-1)) \\ \hat{x}(k+1|k) &= A\hat{x}(k|k) + Bu(k) \end{aligned} \tag{6.3}$$

- The free parameter in the above is K, which is called the *observer gain matrix*. What remains to be discussed is how to choose K. In general, it should be chosen so that the estimation error ($x_e(k) \triangleq x(k) - \hat{x}(k|k)$ or $\hat{x}_e(k+1) \triangleq x(k+1) - \hat{x}(k+1|k)$) is minimized in some sense.

- Equations for error dynamics can be easily derived. For instance, the equations for the filter estimation error is

$$x_e(k) = (A - KCA)x_e(k-1) + (I - KC)\varepsilon_1(k-1) + K\varepsilon_2(k) \tag{6.4}$$

The above can be derived straightforwardly by replacing $y(k)$ in the observer equation (6.2) with $CAx(k-1) + CBu(k-1) + C\varepsilon_1(k-1)$. The equation for prediction error $\hat{x}_e(k)$ can be derived similarly as

$$\hat{x}_e(k+1) = (A - AKC)\hat{x}_e(k) + \varepsilon_1(k-1) + AK\varepsilon_2(k) \tag{6.5}$$

- In some cases, it is advantageous to allow K to vary with time. This results in a time varying observer.

6.2 POLE PLACEMENT

From (6.4), it is clear that the eigenvalues of the transition matrix $A - KCA$ determine how the estimation error propagates. For instance, one must take care that all the eigenvalues lie strictly inside the unit circle in order to ensure stable error dynamics (i.e., asymptotically vanishing initialization error, finite error variance, etc.). The eigenvalues of $A - KCA$ are called observer poles and determining K on the basis of prespecified observer pole location is called *pole placement*. For instance, if (C, A) is an observable pair, the observer poles can be placed in an arbitrary manner through K.

One can also work with the one-step-ahead prediction error equation (6.5). In this case one can let $AK = \hat{K}$ and determine $\hat{K}$ so that the eigenvalues of $A - \hat{K}C$ are placed at desired locations. Again, with an observer system, the eigenvalues can be placed at arbitrary locations.

Pole placement is most coveniently carried out by first putting the system in an observer canonical form through an appropriate coordinate transformation (given by the observability matrix). For instance, consider the following observer canonical form for a single-input, single-output system:

$$\begin{aligned} x(k+1) &= \begin{bmatrix} -a_1 & 1 & 0 & \cdots & 0 \\ -a_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -a_{n-1} & 0 & 0 & \cdots & 1 \\ -a_n & 0 & 0 & \cdots & 0 \end{bmatrix} x(k) + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{n-1} \\ b_n \end{bmatrix} u(k) \\ y(k) &= \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \end{bmatrix} x(k) \end{aligned} \tag{6.6}$$

Then, assuming $K = [k_1 \; k_2 \; \cdots \; k_{n-1} \; k_n]^T$, we have

$$A - KC = \begin{bmatrix} -(a_1 + k_1) & 1 & 0 & \cdots & 0 \\ -(a_2 + k_2) & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ -(a_{n-1} + k_{n-1}) & 0 & 0 & \cdots & 1 \\ -(a_n + k_n) & 0 & 0 & \cdots & 0 \end{bmatrix} \tag{6.7}$$

The characteristic polynomial for the above matrix is

$$z^n + (a_1 + k_1) z^{n-1} + \cdots + (a_{n-1} + k_{n-1}) z + (a_n + k_n) = 0 \tag{6.8}$$

Hence, $k_1, \cdots, k_n$ can be easily determined to place the roots at desired locations.

6.3 KALMAN FILTER

An observer gain can also be determined from a stochastic optimal estimation viewpoint. For example, the observer gain for the linear observer structure can be chosen to minimize the variance of the estimation error. The resulting estimator is the celebrated *Kalman filter*, which has by far been the most popular state estimation technique. When the additional assumption is made that the disturbances are Gaussian, the Kalman filter is indeed the optimal estimator (not just the optimal linear estimator).

6.3.1 KALMAN FILTER AS THE OPTIMAL LINEAR OBSERVER

Note that the the linear observer (6.3) can be written in the following one-step-ahead predictor form:

$$\hat{x}(k+1|k) = A\hat{x}(k|k-1) + Bu(k) + \underbrace{AK(k)}_{\hat{K}(k)}\{y(k) - C\hat{x}(k|k-1)\} \qquad (6.9)$$

In the above, we allowed the observer gain to vary with time for generality.

Recall that the error dynamics for $\hat{x}_e(k) = x(k) - \hat{x}(k|k-1)$ are given by

$$\hat{x}_e(k+1) = (A - \hat{K}(k)C)\hat{x}_e(k) + \varepsilon_1(k) + \hat{K}(k)\varepsilon_2(k) \qquad (6.10)$$

Let

$$\begin{aligned} P(k) &= \text{Cov}\{\hat{x}_e(k)\} & (6.11)\\ &= E\left\{(\hat{x}_e(k) - E\{\hat{x}_e(k)\})\,(\hat{x}_e(k) - E\{\hat{x}_e(k)\})^T\right\} & (6.12)\end{aligned}$$

Assuming that the initial guess is chosen so that $E\{\hat{x}_e(0)\} = 0$, $E\{\hat{x}_e(k)\} = 0$ for all $k \geq 0$ and

$$\begin{aligned} P(k+1) &= \left\{\hat{x}_e(k+1)\hat{x}_e^T(k+1)\right\} & (6.13)\\ &= (A - \hat{K}(k)C)P(k)(A - \hat{K}(k)C)^T + R_1 - \hat{K}(k)R_2\hat{K}^T(k) \end{aligned}$$

In the above, we used the fact that $\hat{x}_e(k), \varepsilon_1(k)$ and $\varepsilon_2(k)$ in (6.10) are mutually independent.

Now let us choose $K(k)$ such that $\alpha^T P(k+1)\alpha$ is minimized for an arbitrary choice of α. Since α is an arbitrary vector, this choice of $K(k)$ minimizes the expected value of any norm of $\hat{x}_e$ (including the 2-norm which represents the error variance). Now, it is straightforward algebra to show that

$$\begin{aligned} \alpha^T P(k+1)\alpha &= \alpha^T \Big[AP(k)A^T + R_1 - \hat{K}(k)CP(k)A^T \\ &\quad -AP(k)C^T\hat{K}^T(k) + \hat{K}(k)(R_2 + CP(k)C^T)\hat{K}^T(k)\Big]\alpha \end{aligned} \tag{6.14}$$

Completing the square on the terms involving $\hat{K}(k)$, we obtain

$$\begin{aligned} \alpha^T P(k+1)\alpha &= \alpha^T \Big\{\Big[\hat{K}(k) - AP(k)C^T(R_2 + CP(k)C^T)^{-1}\Big]\left[R_2 + CP(k)C^T\right] \\ &\quad \times \Big[\hat{K}(k) - AP(k)C^T(R_2 + CP(k)C^T)^{-1}\Big]^T\Big\}\alpha \\ &\quad +\alpha^T\left[AP(k)A^T + R_1 - AP(k)C^T(R_2 + CP(k)C^T)^{-1}CP(k)A^T\right]\alpha \end{aligned} \tag{6.15}$$

Hence, $\hat{K}(k)$ minimizing the above is

$$\hat{K}(k) = AP(k)C^T(R_2 + CP(k)C^T)^{-1} \tag{6.16}$$

and

$$P(k+1) = AP(k)A^T + R_1 - AP(k)C^T\left(R_2 + CP(k)C^T\right)^{-1}CP(k)A^T \tag{6.17}$$

Given $x(1|0)$ and $P(1)$, the above equations can be used along with (6.9) to recursively compute $\hat{x}(k+1|k)$. They are referred to as the *time-varying* Kalman filter equations.

Note:

- For detectable systems, it can be shown that $P(k)$ converges to a constant matrix $\bar{P}$ as $K \to \infty$. Hence, for linear time-invariant systems, it is customary to implement an observer with a constant gain matrix derived from $\bar{P}$ according to (6.16). This is referred to as the *steady-state* Kalman filter.

- Also, recall the relationship between the one-step-ahead predictor gain $\hat{K}(k)$ and the filter gain $K(k)$ ($\hat{K}(k) = AK(k)$). Hence,

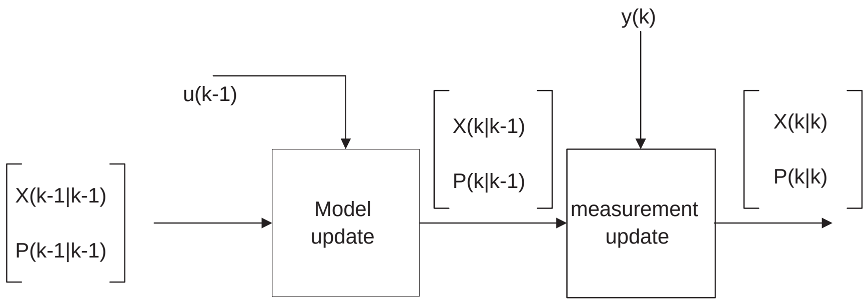
$$K(k) = P(k)C^T(R_2 + CP(k)C^T)^{-1} \tag{6.18}$$

This gain can be used to implement a filter of the form (6.2) that recursively computes $\hat{x}(k|k)$ rather than $\hat{x}(k+1|k)$.

6.3.2 KALMAN FILTER AS THE OPTIMAL ESTIMATOR FOR GAUSSIAN SYSTEMS

In the previous section, we assumed a linear observer structure and posed the problem as a parametric optimization where the expected value of the estimation error variance is minimized with respect to the observer gain. In fact, the Kalman filter can be derived from an entirely probabilistic argument, i.e., by deriving a Bayesian estimator that recursively computes the conditional density of $x(k)$.

Assume that $\varepsilon_1(k)$ and $\varepsilon_2(k)$ are Gaussian noise sequences. Then, assuming $x(0)$ is also a Gaussian variable, $x(k)$ and $y(k)$ are jointly-Gaussian sequences. Now we can simply formulate the state estimation problem as computing the conditional expectation $E\{x(k) \mid Y(k)\}$ where $Y(k) = [y^T(0),\ y^T(1),\ \cdots, y^T(k)]^T$. Let us denote $E\{x(i) \mid Y(j)\}$ as $x(i|j)$. We divide the estimation into the following two steps.



- **Model Update:** *Compute* $E\{x(k)|Y(k-1)\}$ *given* $E\{x(k-1)|Y(k-1)\}$, $P(k-1|k-1)$, *and* $u(k-1)$.
  Since $x(k) = Ax(k-1) + Bu(k-1) + \varepsilon_1(k-1)$ and $\varepsilon_1(k-1)$ is a zero-mean variable independent of $y(0), \cdots, y(k-1)$,

$$\begin{aligned} \hat{x}(k|k-1) &= E\{Ax(k-1) + Bu(k-1) + e(k-1) \mid Y(k-1)\} \\ &= AE\{x(k-1) \mid Y(k-1)\} + Bu(k-1) \qquad (6.19) \end{aligned}$$

Hence, we obtain

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1) \qquad (6.20)$$

In addition, note that

$$x(k) - \hat{x}(k|k-1) = A\left(x(k) - \hat{x}(k-1|k-1)\right) + \varepsilon_1(k-1) \qquad (6.21)$$

Therefore,

$$\begin{aligned} P(k|k-1) &= E\left\{\left(x(k) - \hat{x}(k|k-1)\right)\left(x(k) - \hat{x}(k|k-1)\right)^T\right\} \qquad (6.22) \\ &= AP(k-1|k-1)A^T + R_1 \qquad (6.23) \end{aligned}$$

Since the conditional density for $x(k)$ given $Y(k-1)$ is Gaussian, it is completely specified by $\hat{x}(k|k-1)$ and $P(k|k-1)$.

- **Measurement Update:** *Compute* $E\{x(k)|Y(k)\}$ *given* $E\{x(k)|Y(k-1)\}$ $P(k|k-1)$ *and* $y(k)$.

The conditional density $\mathcal{P}\{x(k) \mid Y(k)\}$ is equivalent to the conditional density $\mathcal{P}\{x(k) \mid y(k)\}$ with the prior density of $x(k)$ given instead by $\mathcal{P}\{x(k) \mid Y(k-1)\}$. Note that $\mathcal{P}\{x(k) \mid Y(k-1)\}$ is a Gaussian density of mean $\hat{x}(k|k-1)$ and covariance $P(k|k-1)$. In other words, we view $x(k)$ as a Gaussian variable of mean $\hat{x}(k|k-1)$ and covariance $P(k|k-1)$.

In addition, $y(k) = Cx(k) + \varepsilon_2(k)$ and therefore is also Gaussian.

$$E\{y(k)\} = CE\{x(k)\} + E\{\varepsilon(k)\} = C\hat{x}(k|k-1)$$

$$E\left\{(y(k) - E\{y(k)\})\,(y(k) - E\{y(k)\})^T\right\} = CP(k|k-1)C^T + R_2$$

In fact, $x(k)$ and $y(k)$ are jointly Gaussian with the following covariance:

$$\begin{aligned} & E\left\{\begin{bmatrix} x(k) - \hat{x}(k|k-1) \\ y(k) - y(k|k-1) \end{bmatrix}\begin{bmatrix} x(k) - \hat{x}(k|k-1) \\ y(k) - y(k|k-1) \end{bmatrix}^T\right\} \\ = & \begin{bmatrix} P(k|k-1) & P(k|k-1)C^T \\ CP(k|k-1) & CP(k|k-1)C^T + R_2 \end{bmatrix} \end{aligned} \tag{6.24}$$

Recall the earlier results for jointly Gaussian variables:

$$E\{x|y\} = E\{x\} + R_{xy}R_y^{-1}(y - E\{y\}) \tag{6.25}$$

$$\mathrm{Cov}\{x|y\} = R_x - R_{xy}R_y^{-1}R_{yx} \tag{6.26}$$

Applying the above to $x(k)$ and $y(k)$,

$$\begin{aligned} \hat{x}(k|k) &= E\{x(k)|y(k)\} \\ &= \hat{x}(k|k-1) \\ &\quad + P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}(y(k) - C\hat{x}(k|k-1)) \\ P(k|k) &= \mathrm{Cov}\{x(k)|y(k)\} \\ &= P(k|k-1) - P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}CP(k|k-1) \end{aligned} \tag{6.27}$$

In short, for Gaussian systems, we can compute the conditional mean and covariance of $x(k)$ recursively using

$$\hat{x}(k|k-1) = A\hat{x}(k-1|k-1) + Bu(k-1)$$

$$\hat{x}(k|k) = \hat{x}(k|k-1) + \underbrace{P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}}_{K(k)}(y(k) - C\hat{x}(k|k-1))$$

and

$$P(k|k-1) = AP(k-1|k-1)A^T + R_1$$

$$P(k|k) = P(k|k-1) - P(k|k-1)C^T\left(CP(k|k-1)C^T + R_2\right)^{-1}CP(k|k-1)$$

Note that this above has a linear observer structure with the observer gain given by the Kalman filter equations derived earlier ($P(k|k-1)$ in the above is $P(k)$ in Eq. (6.16)–(6.17)).