# 3

# SIMULATION AND PREDICTION

The system descriptions given in Chapter 2 can be used for a variety of design problems related to the true system. In this chapter we shall discuss some such uses. The purpose of this is twofold. First, the idea of how to predict future output values will turn out to be most essential for the development of identification methods. The expressions provided in Section 3.2 will therefore be instrumental for the further discussion in this book. Second, by illustrating different uses of system descriptions, we will provide some insights into what is required for such descriptions to be adequate for their intended uses. A leading idea of our framework for identification will be that the effort spent in developing a model of a system must be related to the application it is going to be used for. Throughout the chapter we assume that the system description is given in the form (2.97):

$$y(t) = G(q)u(t) + H(q)e(t) \tag{3.1}$$

## 3.1 SIMULATION

The most basic use of a system description is to simulate the system's response to various input scenarios. This simply means that an input sequence $u^*(t)$, $t = 1, 2, \ldots, N$, chosen by the user is applied to (3.1) to compute the undisturbed output

$$y^*(t) = G(q)u^*(t), \qquad t = 1, 2, \ldots, N \tag{3.2}$$

This is the output that the system would produce had there been no disturbances, according to the description (3.1). To evaluate the disturbance influence, a random-number generator (in the computer) is used to produce a sequence of numbers $e^*(t)$, $t = 1, 2, \ldots, N$, that can be considered as a realization of a white-noise stochastic process with variance $\lambda$. Then the disturbance is calculated as

$$v^*(t) = H(q)e^*(t) \tag{3.3}$$

By suitably presenting $y^*(t)$ and $v^*(t)$ to the user, an idea of the system's response to $\{u^*(t)\}$ can be formed.

This way of experimenting on the model (3.1) rather than on the actual, physical process to evaluate its behavior under various conditions has become widely used in engineering practice of all fields and no doubt reflects the most common use of mathematical descriptions. To be true, models used in, say, flight simulators or nuclear power station training simulators are of course far more complex than (3.1), but they still follow the same general idea (see also Chapter 5).

## 3.2 PREDICTION

We shall start by discussing how future values of $v(t)$ can be predicted in case it is described by

$$v(t) = H(q)e(t) = \sum_{k=0}^{\infty} h(k)e(t-k) \tag{3.4}$$

For (3.4) to be meaningful, we assume that $H$ is stable; that is,

$$\sum_{k=0}^{\infty} |h(k)| < \infty \tag{3.5}$$

**Invertibility of the Noise Model**

A crucial property of (3.4), which we will impose, is that it should be *invertible*; that is, if $v(s)$, $s \le t$, are known, then we shall be able to compute $e(t)$ as

$$e(t) = \tilde{H}(q)v(t) = \sum_{k=0}^{\infty} \tilde{h}(k)v(t-k) \tag{3.6}$$

with

$$\sum_{k=0}^{\infty} \left|\tilde{h}(k)\right| < \infty$$

How can we determine the filter $\tilde{H}(q)$ from $H(q)$? The following lemma gives the answer.

**Lemma 3.1.** Consider $\{v(t)\}$ defined by (3.4) and assume that the filter $H$ is stable. Let

$$H(z) = \sum_{k=0}^{\infty} h(k)z^{-k} \tag{3.7}$$

and assume that the function $1/H(z)$ is analytic in $|z| \ge 1$:

$$\frac{1}{H(z)} = \sum_{k=0}^{\infty} \tilde{h}(k)z^{-k} \tag{3.8}$$

Define the filter $H^{-1}(q)$ by

$$H^{-1}(q) = \sum_{k=0}^{\infty} \tilde{h}(k) q^{-k} \tag{3.9}$$

Then $\tilde{H}(q) = H^{-1}(q)$ satisfies (3.6).

*Remark.* That (3.8) exists for $|z| \geq 1$ also means that the filter $H^{-1}(q)$ is stable. For convenience, we shall then say that $H(q)$ is an *inversely stable* filter.

***Proof.*** From (3.7) and (3.8) it follows that

$$1 = \sum_{k=0}^{\infty} \sum_{s=0}^{\infty} h(k)\tilde{h}(s) z^{-(k+s)} = [k + s = \ell] = \sum_{\ell=0}^{\infty} \sum_{k=0}^{\ell} h(k)\tilde{h}(\ell - k) z^{-\ell}$$

which implies that

$$\sum_{k=0}^{\ell} h(k)\tilde{h}(\ell - k) = \begin{cases} 1, & \text{if } \ell = 0 \\ 0, & \text{if } \ell \neq 0 \end{cases} \tag{3.10}$$

Now let $\{v(t)\}$ be defined by (3.4) and consider

$$\begin{aligned} \sum_{k=0}^{\infty} \tilde{h}(k) v(t - k) &= \sum_{k=0}^{\infty} \tilde{h}(k) \sum_{s=0}^{\infty} h(s) e(t - k - s) \\ &= \sum_{k=0}^{\infty} \sum_{s=0}^{\infty} \tilde{h}(k) h(s) e(t - k - s) = [k + s = \ell] \\ &= \sum_{\ell=0}^{\infty} \left[ \sum_{k=0}^{\ell} \tilde{h}(k) h(\ell - k) \right] e(t - \ell) = e(t) \end{aligned}$$

according to (3.10), which proves the lemma. $\square$

*Note:* The lemma shows that the properties of *the filter* $H(q)$ are quite analogous to those of *the function* $H(z)$. It is not a triviality that the inverse filter $H^{-1}(q)$ can be drived by inverting the function $H(z)$: hence the formulation of the result as a lemma. However, all similar relationships between $H(q)$ and $H(z)$ will also hold, and from a practical point of view it will be useful to switch freely between the filter and its $z$-transform. See also Problem 3D.1.

The lemma shows that the inverse filter (3.6) in a natural way relates to the original filter (3.4). In view of its definition, we shall also write

$$H^{-1}(q) = \frac{1}{H(q)} \tag{3.11}$$

for this filter. All that is needed is that the function $1/H(z)$ be analytic in $|z| \geq 1$; that is, it has no poles on or outside the unit circle. We could also phrase the condition as $H(z)$ must have no zeros on or outside the unit circle. This ties in very nicely with the spectral factorization result (see Section 2.3) according to which, for rational strictly positive spectra, we can always find a representation $H(q)$ with these properties.

**Example 3.1 A Moving Average Process**

Suppose that

$$v(t) = e(t) + ce(t-1) \tag{3.12}$$

That is,

$$H(q) = 1 + cq^{-1}$$

According to (2.87), this process is a *moving average of order 1*, MA(1). Then

$$H(z) = 1 + cz^{-1} = \frac{z+c}{z}$$

has a pole in $z = 0$ and a zero in $z = -c$, which is inside the unit circle if $|c| < 1$. If so, the inverse filter is determined as

$$H^{-1}(z) = \frac{1}{H(z)} = \frac{1}{1 + cz^{-1}} = \sum_{k=0}^{\infty} (-c)^k z^{-k}$$

and $e(t)$ is recovered from (3.12) as

$$e(t) = \sum_{k=0}^{\infty} (-c)^k v(t-k)$$

□

**One-step-ahead Prediction of $v$**

Suppose now that we have observed $v(s)$ for $s \leq t-1$ and that we want to predict the value of $v(t)$ based on these observations. We have, since $H$ is monic,

$$v(t) = \sum_{k=0}^{\infty} h(k)e(t-k) = e(t) + \sum_{k=1}^{\infty} h(k)e(t-k) \tag{3.13}$$

Now, the knowledge of $v(s)$, $s \leq t-1$ implies the knowledge of $e(s)$, $s \leq t-1$, in view of (3.6). The second term of (3.13) is therefore known at time $t-1$. Let us denote it, provisionally, by $m(t-1)$:

$$m(t-1) = \sum_{k=1}^{\infty} h(k)e(t-k)$$

Suppose that $\{e(t)\}$ are identically distributed, and let the probability distribution of $e(t)$ be denoted by $f_e(x)$:

$$P(x \leq e(t) \leq x + \Delta x) \approx f_e(x)\Delta x$$

This distribution is independent of the other values of $e(s)$, $s \neq t$, since $\{e(t)\}$ is a sequence of independent random variables. What we can say about $v(t)$ at time $t-1$ is consequently that the probability that $v(t)$ assumes a value between $m(t-1)+x$ and $m(t-1)+x+\Delta x$ is $f_e(x)\Delta x$. This could also be phrased as

> *the (posterior) probability density function of $v(t)$, given observations up to time $t-1$, is $f_v(x) = f_e(x - m(t-1))$.*

Formally, these calculations can be written as

$$\begin{aligned}
f_v(x)\Delta x &\approx P\left(x \le v(t) \le x + \Delta x | v_{-\infty}^{t-1}\right) \\
&= P\left(x \le m(t-1) + e(t) \le x + \Delta x\right) \\
&= P\left(x - m(t-1) \le e(t) \le x + \Delta x - m(t-1)\right) \\
&\approx f_e\left(x - m(t-1)\right)\Delta x
\end{aligned}$$

Here $P(A|v_{-\infty}^{t-1})$ means the conditional probability of the event $A$, given $v_{-\infty}^{t-1}$.

This is the most complete statement that can be made about $v(t)$ at time $t-1$. Often we just give one value that characterizes this probability distribution and hence serves as a *prediction* of $v(t)$. This could be chosen as the value for which the PDF $f_e(x - m(t-1))$ has its maximum, the most probable value of $v(t)$, which also is called the *maximum a posteriori (MAP)* prediction. We shall, however, mostly work with the mean value of the distribution in question, the *conditional expectation* of $v(t)$ denoted by $\hat{v}(t|t-1)$. Since the variable $e(t)$ has zero mean, we have

$$\hat{v}(t|t-1) = m(t-1) = \sum_{k=1}^{\infty} h(k)e(t-k) \tag{3.14}$$

It is easy to establish that the conditional expectation also minimizes the mean-square error of the prediction error:

$$\min_{\hat{v}(t)} E\left(v(t) - \hat{v}(t)\right)^2 \Rightarrow \hat{v}(t) = \hat{v}(t|t-1)$$

where the minimization is carried out over all functions $\hat{v}(t)$ of $v_{-\infty}^{t-1}$. See Problem 3D.3.

Let us find a more convenient expression for (3.14). We have, using (3.6) and (3.11),

$$\begin{aligned}
\hat{v}(t|t-1) &= \left[\sum_{k=1}^{\infty} h(k)q^{-k}\right] e(t) = [H(q) - 1]\, e(t) \\
&= \frac{H(q) - 1}{H(q)} v(t) = \left[1 - H^{-1}(q)\right] v(t) = \sum_{k=1}^{\infty} -\tilde{h}(k)v(t-k)
\end{aligned} \tag{3.15}$$

Applying $H(q)$ to both sides gives the alternative expression

$$H(q)\hat{v}(t|t-1) = [H(q) - 1]\,v(t) = \sum_{k=1}^{\infty} h(k)v(t-k) \tag{3.16}$$

**Example 3.2 A Moving Average Process**

Consider the process (3.12). Then (3.16) shows that the predictor is calculated as

$$\hat{v}(t|t-1) + c\hat{v}(t-1|t-2) = cv(t-1) \tag{3.17}$$

Alternatively we can determine $H^{-1}(q)$ from Example 3.1 and use (3.15):

$$\hat{v}(t|t-1) = -\sum_{k=1}^{\infty}(-c)^k v(t-k)$$

□

**Example 3.3 An Autoregressive Process**

Consider a process

$$v(t) = \sum_{k=0}^{\infty} a^k e(t-k), \qquad |a| < 1$$

Then

$$H(z) = \sum_{k=0}^{\infty} a^k z^{-k} = \frac{1}{1 - az^{-1}}$$

which gives

$$H^{-1}(z) = 1 - az^{-1}$$

and the predictor, according to (3.15),

$$\hat{v}(t|t-1) = av(t-1) \tag{3.18}$$

□

**One-step-ahead Prediction of $y$**

Consider the description (3.1), and assume that $y(s)$ and $u(s)$ are known for $s \le t-1$. Since

$$v(s) = y(s) - G(q)u(s) \tag{3.19}$$

this means that also $v(s)$ are known for $s \le t-1$. We would like to predict the value

$$y(t) = G(q)u(t) + v(t)$$

based on this information. Clearly, the conditional expectation of $y(t)$, given the information in question, is

$$\begin{aligned}\hat{y}(t|t-1) &= G(q)u(t) + \hat{v}(t|t-1)\\ &= G(q)u(t) + \left[1 - H^{-1}(q)\right]v(t)\\ &= G(q)u(t) + \left[1 - H^{-1}(q)\right]\left[y(t) - G(q)u(t)\right]\end{aligned}$$

using (3.15) and (3.19), respectively. Collecting the terms gives

$$\boxed{\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \left[1 - H^{-1}(q)\right]y(t) \qquad (3.20)}$$

or

$$\boxed{H(q)\hat{y}(t|t-1) = G(q)u(t) + [H(q) - 1]\,y(t) \qquad (3.21)}$$

Remember that these expressions are shorthand notation for expansions. For example, let $\{\ell(k)\}$ be defined by

$$\frac{G(z)}{H(z)} = \sum_{k=1}^{\infty} \ell(k)z^{-k} \tag{3.22}$$

[This expansion exists for $|z| \geq 1$ if $H(z)$ has no zeros and $G(z)$ no poles in $|z| \geq 1$.] Then (3.20) means that

$$\hat{y}(t|t-1) = \sum_{k=1}^{\infty} \ell(k)u(t-k) + \sum_{k=1}^{\infty} -\tilde{h}(k)y(t-k) \tag{3.23}$$

### Unknown Initial Conditions

In the reasoning so far we have made use of the assumption that the whole data record from time minus infinity to $t-1$ is available. Indeed, in the expression (3.20) as in (3.23) all these data appear explicitly. In practice, however, it is usually the case that only data over the interval $[0, t-1]$ are known. The simplest thing would then be to replace the unknown data by zero (say) in (3.23):

$$\hat{y}(t|t-1) \approx \sum_{k=1}^{t} \ell(k)u(t-k) + \sum_{k=1}^{t} -\tilde{h}(k)y(t-k) \tag{3.24}$$

One should realize that this is now only an approximation of the actual conditional expectation of $y(t)$, given data over $[0, t-1]$. The exact prediction involves time-varying filter coefficients and can be computed using the Kalman filter [see (4.94)]. For most practical purposes, (3.24) will, however, give a satisfactory solution. The reason is that the coefficients $\{\ell(k)\}$ and $\{\tilde{h}(k)\}$ typically decay exponentially with $k$ (see Problem 3G.1).

### The Prediction Error

From (3.20) and (3.1), we find that the prediction error $y(t) - \hat{y}(t|t-1)$ is given by

$$\boxed{y(t) - \hat{y}(t|t-1) = -H^{-1}(q)G(q)u(t) + H^{-1}(q)y(t) = e(t)} \tag{3.25}$$

The variable $e(t)$ thus represents that part of the output $y(t)$ that cannot be predicted from past data. For this reason it is also called the *innovation* at time $t$.

### $k$-step-ahead Prediction of $y$ (∗)

Having treated the problem of one-step-ahead prediction in some detail, it is easy to generalize to the following problem: Suppose that we have observed $v(s)$ for $s \leq t$ and that we want to predict the value $v(t+k)$. We have

$$\begin{aligned} v(t+k) &= \sum_{\ell=0}^{\infty} h(\ell)e(t+k-\ell) \\ &= \sum_{\ell=0}^{k-1} h(\ell)e(t+k-\ell) + \sum_{\ell=k}^{\infty} h(\ell)e(t+k-\ell) \end{aligned} \tag{3.26}$$

Let us define

$$\overline{H}_k(q) = \sum_{\ell=0}^{k-1} h(\ell)q^{-\ell}, \qquad \tilde{H}_k(q) = \sum_{\ell=k}^{\infty} h(\ell)q^{-\ell+k} \tag{3.27}$$

The second sum of (3.26) is known at time time $t$, while the first sum is independent of what has happened up to time $t$ and has zero mean. The conditional mean of $v(t+k)$, given $v_{-\infty}^{t}$ is thus given by

$$\hat{v}(t+k|t) = \sum_{\ell=k}^{\infty} h(\ell)e(t+k-\ell) = \tilde{H}_k(q)e(t) = \tilde{H}_k(q) \cdot H^{-1}(q)v(t)$$

This expression is the $k$-step-ahead predictor of $v$.

Now suppose that we have measured $y_{-\infty}^{t}$ and know $u_{-\infty}^{t+k-1}$ and would like to predict $y(t+k)$. We have, as before

$$y(t+k) = G(q)u(t+k) + v(t+k)$$

which gives

$$\begin{aligned}\hat{y}(t+k|y_{-\infty}^{t}, u_{-\infty}^{t+k-1}) \triangleq \hat{y}(t+k|t) &= G(q)u(t+k) + \hat{v}(t+k|t)\\ &= G(q)u(t+k) + \tilde{H}_k(q)H^{-1}(q)v(t) \qquad (3.28)\\ &= G(q)u(t+k) + \tilde{H}_k(q)H^{-1}(q)\,[y(t) - G(q)u(t)]\end{aligned}$$

Introduce

$$\begin{aligned}W_k(q) &\triangleq 1 - q^{-k}\tilde{H}_k(q)H^{-1}(q) = \left[H(q) - q^{-k}\tilde{H}_k(q)\right]H^{-1}(q)\\ &= \overline{H}_k(q)H^{-1}(q)\end{aligned} \qquad (3.29)$$

Then simple manipulation on (3.28) gives

$$\hat{y}(t+k|t) = W_k(q)G(q)u(t+k) + \tilde{H}_k(q)H^{-1}(q)y(t) \qquad (3.30)$$

or, using the first equality in (3.29),

$$\boxed{\hat{y}(t|t-k) = W_k(q)G(q)u(t) + [1 - W_k(q)]\,y(t) \qquad (3.31)}$$

This expression, together with (3.27) and (3.29), defines the $k$-step-ahead predictor for $y$. Notice that this predictor can also be viewed as a one-step-ahead predictor associated with the model

$$y(t) = G(q)u(t) + W_k^{-1}(q)e(t) \qquad (3.32)$$

The prediction error is obtained from (3.30) as

$$\begin{aligned}e_k(t+k) &\triangleq y(t+k) - \hat{y}(t+k|t) = -W_k(q)G(q)u(t+k)\\ &\quad + \left[q^k - \tilde{H}_k(q)H^{-1}(q)\right]y(t)\\ &= W_k(q)\,[y(t+k) - G(q)u(t+k)] = W_k(q)H(q)e(t+k)\\ &= \overline{H}_k(q)e(t+k)\end{aligned} \qquad (3.33)$$

Here we used (3.29) in the second and fourth equalities. According to (3.27), $\overline{H}_k(q)$ is a polynomial in $q^{-1}$ of order $k-1$. Hence the prediction error is a moving average of $e(t+k), \ldots, e(t+1)$.

### The Multivariable Case (*)

For a multivariable system description (3.1) (or 2.90), we define the $p \times p$ matrix filter $H^{-1}(q)$ as

$$H^{-1}(q) = \sum_{k=0}^{\infty} \tilde{h}(k) q^{-k}$$

Here $\tilde{h}(k)$ are the $p \times p$ matrices defined by the expansion of the matrix function

$$[H(z)]^{-1} = \sum_{k=0}^{\infty} \tilde{h}(k) z^{-k} \tag{3.34}$$

This expansion can be interpreted entrywise in the matrix $[H(z)]^{-1}$ (formed by standard manipulations for matrix inversion). It exists for $|z| \geq 1$ provided the function det $H(z)$ has no zeros in $|z| \geq 1$. With $H^{-1}(q)$ thus defined, all calculations and formulas given previously are valid also for the multivariable case.

## 3.3 OBSERVERS

In many cases in systems and control theory, one does not work with a full description of the properties of disturbances as in (3.1). Instead a noise-free or "*deterministic*" model is used:

$$y(t) = G(q)u(t) \tag{3.35}$$

In this case one probably keeps in the back of one's mind, though, that (3.35) is not really the full story about the input-output properties.

The description (3.35) can of course also be used for "computing," "guessing," or "predicting" future values of the output. The lack of noise model, however, leaves several possibilities for how this can best be done. The concept of *observers* is a key issue for these calculations. This concept is normally discussed in terms of state-space representations of (3.35) (see Section 4.3); see, for example, Luenberger (1971)or Åström and Wittenmark (1984), but it can equally well be introduced for the input-output form (3.35).

### An Example

Let

$$G(z) = b \sum_{k=1}^{\infty} (a)^{k-1} z^{-k} = \frac{bz^{-1}}{1 - az^{-1}} \tag{3.36}$$

This means that the input-output relationship can be represented either as

$$y(t) = b \sum_{k=1}^{\infty} (a)^{k-1} u(t-k) \tag{3.37}$$

that is

$$y(t) = \frac{bq^{-1}}{1 - aq^{-1}} u(t)$$

or as

$$(1 - aq^{-1})y(t) = bq^{-1}u(t)$$

i.e.

$$y(t) - ay(t-1) = bu(t-1) \tag{3.38}$$

Now, if we are given the description (3.35) and (3.36) together with data $y(s)$, $u(s)$, $s \le t-1$, and are asked to produce a "guess" or to "calculate" what $y(t)$ might be, we could use either

$$\hat{y}(t|t-1) = b\sum_{k=1}^{\infty}(a)^{k-1}u(t-k) \tag{3.39}$$

or

$$\hat{y}(t|t-1) = ay(t-1) + bu(t-1) \tag{3.40}$$

As long as the data and the system description are correct, there would also be no difference between (3.39) and (3.40); they are both "observers" (in our setting "predictors" would be a more appropriate term) for the system. The choice between them would be carried out by the designer in terms of how vulnerable they are to imperfections in data and descriptions. For example, if input-output data are lacking prior to time $s = 0$, then (3.39) suffers from an error that decays like $a^t$ (effect of wrong initial conditions), whereas (3.40) is still correct for $t \ge 1$. On the other hand, (3.39) is unaffected by measurement errors in the output, whereas such errors are directly transferred to the prediction in (3.40). From the discussion of Section 3.2, it should be clear that, if (3.35) is complemented with a noise model as in (3.1), then the choice of predictor becomes unique (cf. Problem 3E.3). This follows since the conditional mean of the output, computed according to the assumed noise model, is a uniquely defined quantity.

**A Family of Predictors for (3.35)**

The example (3.36) showed that the choice of predictor could be seen as a trade-off between sensitivity with respect to output measurement errors and rapidly decaying effects of erroneous initial conditions. To introduce design variables for this trade-off, choose a filter $W(q)$ such that

$$W(q) = 1 + \sum_{\ell=k}^{\infty} w_\ell q^{-\ell} \tag{3.41}$$

Apply it to both sides of (3.35):

$$W(q)y(t) = W(q)G(q)u(t)$$

which means that

$$y(t) = [1 - W(q)]\,y(t) + W(q)G(q)u(t)$$

In view of (3.41), the right side of this expression depends only on $y(s)$, $s \le t-k$, and $u(s)$, $s \le t-1$. Based on that information, we could thus produce a "guess" or prediction of $y(t)$ as

$$\hat{y}(t|t-k) = [1 - W(q)]\,y(t) + W(q)G(q)u(t) \tag{3.42}$$

The trade-off considerations for the choice of $W$ would then be:

$$\begin{array}{l} \textbf{1. } \text{Select } W(q) \text{ so that both } W \text{ and } WG \text{ have rapidly decaying filter coefficients in order to minimize the influence of erroneous initial conditions.} \\ \textbf{2. } \text{Select } W(q) \text{ so that measurement imperfections in } y(t) \text{ are maximally attenuated.} \end{array} \tag{3.43}$$

The later issue can be illuminated in the frequency domain: Suppose that $y(t) = y_M(t)+v(t)$, where $y_M(t) = G(q)u(t)$ is the useful signal and $v(t)$ is a measurement error. Then the prediction error according to (3.42) is

$$\varepsilon(t) = y(t) - \hat{y}(t|t-k) = W(q)v(t) \tag{3.44}$$

The spectrum of this error is, according to Theorem 2.2,

$$\Phi_\varepsilon(\omega) = \left|W(e^{i\omega})\right|^2 \Phi_v(\omega) \tag{3.45}$$

where $\Phi_v(\omega)$ is the spectrum of $v$. The problem is thus to select $W$, subject to (3.41), such that the error spectrum (3.45) has an acceptable size and suitable shape.

A comparison with the $k$-step prediction case of Section 3.2 shows that the expression (3.42) is identical to (3.31) with $W(q) = W_k(q)$. It is clear that the qualification of a complete noise model in (3.1) allows us to analytically compute the filter $W$ in accordance with aspect 2. This was indeed what we did in Section 3.2. However, aspect 1 was neglected there, since we assumed all past data to be available. Normally, as we pointed out, this aspect is also less important.

**Fundamental Role of the Predictor Filter**

It turns out that for most uses of system descriptions it is the predictor form (3.20), or as in (3.31) and (3.42), that is more important than the description (3.1) or (3.35) itself. We use (3.31) and (3.42) to predict, or "guess," future outputs; we use it for control design to regulate the predicted output, and so on. Now, (3.31) and (3.42) are just linear filters into which sequences $\{u(t)\}$ and $\{y(t)\}$ are fed, and that produce $\hat{y}(t|t-k)$ as output. The thoughts that the designer had when he or she selected this filter are immaterial once it is put to use: *The filter is the same whether $W = W_k$ was chosen as a trade-off (3.43) or computed from $H$ as in (3.27) and (3.29).* The noise model $H$ in (3.1) is from this point of view just an alibi for determining the predictor. This is the viewpoint we are going to adopt. The predictor filter is the fundamental system description (Figure 3.1). Our rationale for arriving at the filter is secondary. This also means that the difference between a "stochastic system" (3.1) and a "deterministic" one (3.35) is not fundamental. Nevertheless, we find it convenient to use the description (3.1) as the basic system description. It is in a one-to-one correspondence with the one-step-ahead predictor (3.20) (see Problem 3D.2) and relates more immediately to traditional system descriptions.

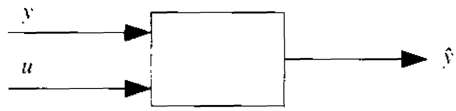


**Figure 3.1** The predictor filter.

## 3.4 SUMMARY

Starting from the representation

$$y(t) = G(q)u(t) + H(q)e(t)$$

we have derived an expression for the one-step-ahead prediction of $y(t)$ [i.e., the best "guess" of $y(t)$ given $u(s)$ and $y(s)$, $s \leq t - 1$]. This expression is given by

$$\hat{y}(t|t-1) = H^{-1}(q)G(q)u(t) + \left[1 - H^{-1}(q)\right]y(t) \tag{3.46}$$

We also derived a corresponding $k$-step-ahead predictor (3.31). We pointed out that one can arrive at such predictors also through deterministic observer considerations, not relying on a noise model $H$. We have stressed that the bottom line in most uses of a system description is how these predictions actually are computed; the underlying noise assumptions are merely vehicles for arriving at the predictors. The discussion of Chapters 2 and 3 can thus be viewed as a methodology for "guessing" future system outputs.

It should be noted that calculations such as (3.46) involved in determining the predictors and regulators are typically performed with greater computational efficiency once they are applied to transfer functions $G$ and $H$ with more specific structures. This will be illustrated in the next chapter.

## 3.5 BIBLIOGRAPHY

Prediction and control are standard textbook topics. Accounts of the $k$-step-ahead predictor and associated control problems can be found in Åström (1970)and Åström and Wittenmark (1984). Prediction is treated in detail in, for example, Anderson and Moore (1979)and Box and Jenkins (1970). An early account of this theory is Whittle (1963).

Prediction theory was developed by Kolmogorov (1941), Wiener (1949), Kalman (1960), and Kalman and Bucy (1961). The hard part in these problems is indeed to find a suitable representation of the disturbance. Once we arrive at (3.1) via spectral factorization, or at its time-varying counterpart via the Riccati equation [see (4.95) and Problem 4G.3], the calculation of a reasonable predictor is, as demonstrated here, easy. Note, however (as pointed out in Problem 2E.3), that for non-Gaussian processes normally only the second-order properties can be adequately described by (3.1), which consequently is too simple a representation to accommodate more complex noise structures. The calculations carried out in Section 3.2 are given in Åström (1970)for the case where $G$ and $H$ are rational with the same denominators. Rissanen and Barbosa (1969)have given expressions for the prediction in input-output models of this kind when the lack of knowledge of the infinite past is treated properly [i.e., when the ad hoc solution (3.24) is not accepted]. The result is, of course, a time-varying predictor.

## 3.6 PROBLEMS

**3G.1** Suppose that the transfer function $G(z)$ is rational and that its poles are all inside $|z| < \mu$, where $\mu < 1$. Show that

$$|g(k)| \leq c \cdot \mu^k$$

where $g(k)$ is defined as in (2.16).

**3G.2** Let $A(q)$ and $B(q)$ be two monic stable and inversely stable filters. Show that

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \left|A(e^{i\omega})\right|^2 \cdot \left|B(e^{i\omega})\right|^2 \, d\omega \geq 1$$

with equality only if $A(q) = 1/B(q)$.

**3E.1** Let

$$H(q) = 1 - 1.1q^{-1} + 0.3q^{-2}$$

Compute $H^{-1}(q)$ as an explicit infinite expansion.

**3E.2** Determine the 3-step-ahead predictors for

$$y(t) = \frac{1}{1 - aq^{-1}}e(t)$$

and

$$y(t) = (1 + cq^{-1})e(t)$$

respectively. What are the variances of the associated prediction errors?

**3E.3** Show that if (3.35) and (3.36) are complemented with the noise model $H(q) = 1$ then (3.39) is the natural predictor, whereas the noise model

$$H(q) = \sum_{k=0}^{\infty}(a^k)q^{-k}$$

leads to the predictor (3.40).

**3E.4** Let $e(t)$ have the distribution

$$e(t) = \begin{cases} 1, & \text{w.p. } 0.5 \\ -0.5, & \text{w.p. } 0.25 \\ -1.5, & \text{w.p. } 0.25 \end{cases}$$

Let

$$v(t) = H(q)e(t)$$

and let $\hat{v}(t|t-1)$ be defined as in the text. What is the most probable value (MAP) of $v(t)$ given the information $\hat{v}(t|t-1)$? What is the probability that $v(t)$ will assume a value between $\hat{v}(t|t-1) - \frac{1}{4}$ and $\hat{v}(t|t-1) + \frac{1}{4}$?

**3T.1** Suppose that $A(q)$ is inversely stable and monic. Show that $A^{-1}(q)$ is monic.

**3T.2** Suppose the measurement error spectrum of $v$ in (3.44) and (3.45) is given by

$$\Phi_v(\omega) = \lambda \left| R(e^{i\omega}) \right|^2$$

for some monic stable and inversely stable filter $R(q)$. Find the filter $W$, subject to (3.41) with $k = 1$, that minimizes

$$\overline{E}\varepsilon^2(t)$$

*Hint:* Use Problem 3G.2.

**3T.3** Consider the system description of Problem 2E.4:

$$x(t+1) = fx(t) + w(t)$$

$$y(t) = hx(t) + v(t)$$

($x$ scalar). Assume that $\{v(t)\}$ is white Gaussian noise with variance $R_2$ and that $\{w(t)\}$ is a sequence of independent variables with

$$w(t) = \begin{cases} 1, & \text{w.p. } 0.05 \\ -1, & \text{w.p. } 0.05 \\ 0, & \text{w.p. } 0.9 \end{cases}$$

Determine a monic filter $W(q)$ such that the predictor

$$\hat{y}(t) = (1 - W(q))\, y(t)$$

minimizes

$$\overline{E}\left(y(t) - \hat{y}(t)\right)^2$$

What can be said about

$$E\left(y(t) | y_{-\infty}^{t-1}\right)?$$

**3T.4** Consider the noise description

$$v(t) = e(t) + ce(t-1), \quad |c| > 1, \quad Ee^2(t) = \lambda \tag{3.47}$$

Show that $e(t)$ cannot be reconstructed from $v^t$ by a causal, stable filter. However, show that $e(t)$ can be computed from $v_{t+1}^{\infty}$ by an anticausal, stable filter. Thus construct a stable, anticausal predictor for $v(t)$ given $v(s)$, $s \geq t+1$.

Determine a noise $\overline{v}(t)$ with the same second-order properties as $v(t)$, such that

$$\overline{v}(t) = \overline{e}(t) + c^*\overline{e}(t-1), \quad |c^*| < 1, \quad E\overline{e}^2(t) = \lambda^* \tag{3.48}$$

Show that $\overline{v}(t)$ can be predicted from $\overline{v}^{t-1}$ by a stable, causal predictor. [Measuring just second-order properties of the noise, we cannot distinguish between (3.47) and (3.48). However, when $e(t)$ in (3.47) is a physically well defined quantity (although not measured by us), we may be interested in which one of (3.47) and (3.48) has generated the noise. See Benveniste, Goursat, and Ruget (1980).]

**3D.1** In the chapter we have freely multiplied, added, subtracted, and divided by transfer-function operators $G(q)$ and $H(q)$. Division was formalized and justified by Lemma 3.1 and (3.11). Justify similarly addition and multiplication.

**3D.2** Suppose a one-step-ahead predictor is given as

$$\hat{y}(t|t-1) = L_1(q)u(t-1) + L_2(q)y(t-1)$$

Calculate the system description (3.1) from which this predictor was derived.

**3D.3** Consider a stochastic process $\{v(t)\}$ and let

$$\hat{v}(t) = E\left(v(t)|v^{t-1}\right)$$

Define

$$e(t) = v(t) - \hat{v}(t)$$

Let $\bar{v}(t)$ be an arbitrary function of $v^{t-1}$. Show that

$$E\left(v(t) - \bar{v}(t)\right)^2 \geq Ee^2(t)$$

*Hint:* Use $Ex^2 = E_z E(x^2|z)$.