Chapter 2
Discrete-Time System Models for Control

2.1 Deterministic Environment

2.1.1 Input-Output Difference Operator Models

We will consider single-input single-output time invariant systems described by input-output discrete-time models of the form:

$$y(t) = -\sum_{i=1}^{n_A} a_i y(t-i) + \sum_{i=1}^{n_B} b_i u(t-d-i) \tag{2.1}$$

where t denotes the normalized sampling time (i.e., $t = \frac{t}{T_S}$, $T_S =$ sampling period), $u(t)$ is the input, $y(t)$ is the output, d is the integer number of sampling periods contained in the time delay of the systems, a_i and b_i are the parameters (coefficients) of the models. As such the output of the system at instant t is a weighted average of the past output over an horizon of n_A samples plus a weighted average of past inputs over an horizon of n_B samples (delayed by d samples). This input-output model (2.1) can be more conveniently represented using a coding in terms of forward or backward shift operators defined as:

$$q \triangleq \text{forward shift operator } (qy(t) = y(t+1)) \tag{2.2}$$

$$q^{-1} \triangleq \text{backward shift operator } (q^{-1}y(t) = y(t-1)) \tag{2.3}$$

Using the notation:

$$1 + \sum_{i=1}^{n_A} a_i q^{-i} = A(q^{-1}) = 1 + q^{-1}A^*(q^{-1}) \tag{2.4}$$

where:

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_A} q^{-n_A} \tag{2.5}$$

$$A^*(q^{-1}) = a_1 + a_2 q^{-1} + \cdots + a_{n_A} q^{-n_A+1} \tag{2.6}$$

I.D. Landau et al., *Adaptive Control*, Communications and Control Engineering,
DOI 10.1007/978-0-85729-664-1_2,

and:

$$\sum_{i=1}^{n_B} b_i q^{-i} = B(q^{-1}) = q^{-1} B^*(q^{-1}) \tag{2.7}$$

where:

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + \cdots + b_{n_B} q^{-n_B} \tag{2.8}$$

$$B^*(q^{-1}) = b_1 + b_2 q^{-1} + \cdots + b_{n_B} q^{-n_B+1} \tag{2.9}$$

Equation (2.1) can be rewritten as:

$$A(q^{-1})y(t) = q^{-d} B(q^{-1})u(t) = q^{-d-1} B^*(q^{-1})u(t) \tag{2.10}$$

or forward in time:

$$A(q^{-1})y(t+d) = B(q^{-1})u(t) \tag{2.11}$$

as well as:[1]

$$y(t+1) = -A^* y(t) + q^{-d} B^* u(t) = -A^* y(t) + B^* u(t-d) \tag{2.12}$$

Observe that (2.12) can also be expressed as (the *regressor form*):

$$y(t+1) = \theta^T \varphi(t) \tag{2.13}$$

where θ defines the vector of parameters

$$\theta^T = [a_1, \ldots, a_{n_A}, b_1, \ldots, b_{n_B}] \tag{2.14}$$

and $\varphi(t)$ defines the vector of measurements (or the regressor)

$$\varphi^T(t) = [-y(t), \ldots, -y(t-n_A+1), u(t-d), \ldots, u(t-d-n_B+1)] \tag{2.15}$$

The form of (2.13) will be used in order to estimate the parameters of a system model from input-output data. Consider (2.10). Passing the quantities in the left and in the right through a filter $\frac{1}{A(q^{-1})}$ one gets:

$$y(t) = G(q^{-1})u(t) \tag{2.16}$$

where:

$$G(q^{-1}) = \frac{q^{-d} B(q^{-1})}{A(q^{-1})} \tag{2.17}$$

is termed the *transfer operator*.

Computing the z-transform of (2.1), one gets the pulse transfer function characterizing the input-output model of (2.1):[2]

$$G(z^{-1}) = \frac{z^{-d} B(z^{-1})}{A(z^{-1})} \tag{2.18}$$

[1] In many cases, the argument q^{-1} will be dropped out, to simplify the notation.

[2] A number of authors prefer to use the notation $G(z)$ for this quantity, instead of $G(z^{-1})$, in order to be coherent with the definition of the z-transform.

Observe that the transfer function of the input-output model of (2.1) can be formally obtained from the *transfer operator* by replacing the time operator q by the complex variable z. However, one should be careful since the domain of these variables is different. Nevertheless in the linear case with constant parameters one can use either one and their appropriate signification will result from the context.

Note also that the transfer operator $G(q^{-1})$ can be defined even if the parameters of the model (2.1) are time varying, while the concept of pulse transfer function does simply not exist in this case.

While in most of the developments throughout the book we will not need to associate a state-space form to the input-output model of (2.1), this indeed clarifies a number of properties of these models, in particular the definition of the "order" of the system.

Theorem 2.1 *The order r of the system model* (2.1), *is the dimension of the minimal state space representation associated to the input-output model* (2.1) *and in the case of irreducible transfer function it is equal to*:

$$r = \max(n_A, n_B + d) \tag{2.19}$$

which corresponds also to the number of the poles of the irreducible transfer function of the system.

The order of the system is immediately obtained by expressing the transfer operator (2.17) or the transfer function (2.18) in the forward operator q and respectively the complex variable z. The passage from $G(z^{-1})$ to $G(z)$ is obtained multiplying by z^r:

$$G(z) = \frac{\bar{B}(z)}{\bar{A}(z)} = \frac{z^{r-d}B(z^{-1})}{z^r A(z^{-1})} \tag{2.20}$$

Example

$$G(z^{-1}) = \frac{z^{-3}(b_1 z^{-1} + b_2 z^{-2})}{1 + a_1 z^{-1}} \quad \Longrightarrow \quad r = \max(1, 5) = 5$$

$$G(z) = \frac{b_1 z + b_2}{z^5 + a_1 z^4}$$

To see that r effectively corresponds to the dimension of the minimal state space representation let us consider the following observable canonical form:

$$x(t+1) = A_0 x(t) + B_0 u(t) \tag{2.21}$$

$$y(t) = C_0 x(t) \tag{2.22}$$

where $x(t)$ is the state vector and the matrices (or vectors) A_0, B_0, C_0 are given by:

(a) $n_B + d > n_A$

$$A_0 = \left[\begin{array}{ccccccc} -a_1 & 1 & & & & & \\ \vdots & & \ddots & & & & 0 \\ \vdots & & & \ddots & & & \\ -a_{n_A} & \cdots & \cdots & \cdots & 1 & \cdots & \cdots \\ 0 & & & & \ddots & & \\ \vdots & & 0 & & & \ddots & \\ 0 & & & & & & 1 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} n_A \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n_B+d-n_A \end{array}$$

$$B_0 = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ b_1 \\ \vdots \\ \vdots \\ b_{n_B} \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} d \\ \left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\} n_B \end{array}$$

$$C_0 = [1, 0, \ldots, 0]$$

(b) $n_A \geq n_B + d$

$$A_0 = \left[\begin{array}{cccccc} -a_1 & 1 & \cdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & & & \vdots \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & 0 \\ -a_{n_A} & & & & 0 & 1 \end{array}\right], \qquad B_0 = \left[\begin{array}{c} 0 \\ \vdots \\ 0 \\ b_1 \\ \vdots \\ b_{n_B} \\ 0 \\ \vdots \\ 0 \end{array}\right] \begin{array}{l} \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} d \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n_B \\ \left.\vphantom{\begin{array}{c}1\\1\\1\end{array}}\right\} n_A - (n_B + d) \end{array}$$

$$C_0 = [1, 0, \ldots, 0]$$

The input-output transfer function is given by:

$$G(z) = C_0(zI - A_0)^{-1} B_0 = \frac{\bar{B}(z)}{\bar{A}(z)} = \frac{z^{r-d} B(z^{-1})}{z^r A(z^{-1})} \tag{2.23}$$

Remarks

- If $n_B + d > n_A$, the system will have $n_B + d - n_A$ poles at the origin ($z = 0$).
- One has assumed that A_0, B_0, C_0 is a minimal state space realization of the system model (2.1), i.e., that the eventual common factors of $A(z^{-1})$ and $B(z^{-1})$ have been canceled.

In general we will assume that the model of (2.1) and the corresponding transfer function (2.18) is irreducible. However, situations may occur where this is indeed not the case (an estimated model may feature an almost pole zeros cancellation). The properties of the system model in such cases are summarized below:

- The presence of pole zeros cancellations correspond to the existence of unobservable or uncontrollable modes.
- If the common poles and zeros are stable (they are inside the unit circle) the system is termed *stabilizable*, i.e., there is a feedback law stabilizing the system.
- If the common poles and zeros are unstable, the system is *not stabilizable* (a feedback law stabilizing the system does not exist).

The co-primeness of $A(z^{-1})$ and $B(z^{-1})$ is an important property of the model. A characterization of the co-primeness of $A(z^{-1})$ and $B(z^{-1})$ without searching the roots of $A(z^{-1})$ and $B(z^{-1})$ is given by the Sylvester Theorem.

Theorem 2.2 (Wolowich 1974; Kailath 1980; Goodwin and Sin 1984) *The polynomials $A(q^{-1})$, $q^{-d}B(q^{-1})$ are relatively prime if and only if their eliminant matrix M (known also as the Sylvester Matrix) is nonsingular, where M is a square matrix $r \times r$ with $r = \max(n_A, n_B + d)$ given by*:

$$\left.\underbrace{\begin{array}{cc} \overbrace{\hphantom{xxxxxxxxxxxx}}^{n_B+d} & \overbrace{\hphantom{xxxxxxxxxxxx}}^{n_A} \\ \multicolumn{2}{c}{\begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ a_1 & 1 & & \vdots & b'_1 & & & \\ a_2 & & & 0 & b'_2 & & & b'_1 \\ & & & 1 & \vdots & & & b'_2 \\ & & & a_1 & \vdots & & & \vdots \\ a_{n_A} & & & a_2 & b'_{n_B} & & & \vdots \\ 0 & & & \vdots & 0 & \dots & \dots & \ddots \\ 0 & \dots & 0 & a_{n_A} & 0 & 0 & 0 & b'_{n_{B'}} \end{bmatrix}} \end{array}}_{n_A+n_B+d}\right\} n_A + n_B + d \tag{2.24}$$

where: $b'_i = 0$ *for* $i = 0, 1, \dots, d$; $b'_i = b_{i-d}$ *for* $i \geq d + 1$.

Remarks

- The nonsingularity of the matrix M implies the controllability and the observability of the associated state space representation.
- The condition number of M allows to evaluate the ill conditioning of the matrix M, i.e., the closeness of some poles and zeros.
- The matrix M is also used for solving the diophantine equation (Bezout identity):

$$A(z^{-1})S(z^{-1}) + z^{-d}B(z^{-1})R(z^{-1}) = P(z^{-1})$$

for S and R, given P (see Sect. 7.3 Pole Placement).

2.1.2 Predictor Form (Prediction for Deterministic SISO Models)

The model of (2.1) or of (2.12) is a one step ahead predictive form. A problem of interest is to predict from the model (2.1) future values of the output beyond $(t+1)$ based on the information available up to the instant t, assuming in the deterministic case that disturbances will not affect the system in the future. A typical example is the control of a system with delay d for which we would like either to predict its output $d+1$ step ahead based on the measurements and the controls applied up to and including the instant t or to compute $u(t)$ in order to reach a certain value of the output at $t+d+1$.

Therefore the objective is in general to compute (for $1 \leq j \leq d+1$):

$$\hat{y}(t+j/t) = f[y(t), y(t-1), \ldots, u(t), u(t-1), \ldots] \tag{2.25}$$

or a filtered predicted value:

$$P(q^{-1})\hat{y}(t+j/t) = f_P[y(t), y(t-1), \ldots, u(t), u(t-1), \ldots] \tag{2.26}$$

where:

$$\begin{aligned} &P(q^{-1}) = 1 + p_1 q^{-1} + \cdots + p_{n_p} q^{-n_p} = 1 + q^{-1} P^*(q^{-1}) \\ &n_p \leq n_A + j - 1 \end{aligned} \tag{2.27}$$

Note that in deterministic case, due to the hypothesis of the absence of disturbances in the future, the future values of the output can be exactly evaluated.

To simplify the notation, $\hat{y}(t+j/t)$ will be replaced by $\hat{y}(t+j)$. We will start with the case $j=1$, in order to emphasize some of the properties of the predictor. For the case $j=1$, taking account of (2.12) and (2.27), one has:

$$P(q^{-1})y(t+1) = (P^* - A^*)y(t) + q^{-d}B^*u(t) \tag{2.28}$$

and, therefore, this suggests to consider as predictor the form:

$$P(q^{-1})\hat{y}(t+1) = (P^* - A^*)y(t) + q^{-d}B^*u(t) \tag{2.29}$$

The prediction error:

$$\varepsilon(t+1) = y(t+1) - \hat{y}(t+1) \tag{2.30}$$

will be governed by:

$$P(q^{-1})\varepsilon(t+1) = 0 \tag{2.31}$$

which is obtained by subtracting (2.29) from (2.28). Observe also that the prediction equation has a "feedback" form since it can be rewritten as:

$$\hat{y}(t+1) = (P^* - A^*)[y(t) - \hat{y}(t)] - A^*\hat{y}(t) + q^{-d}Bu(t) \tag{2.32}$$

(the prediction error drives the prediction). Note also that for $P = A$, (2.32) leads to an "open-loop predictor" or "output error" predictor governed by:

$$\hat{y}(t+1) = -A^*(q^{-1})\hat{y}(t) + q^{-d}B^*(q^{-1})u(t) \tag{2.33}$$

i.e., the predicted output will depend only on the input and previous predictions.

A state space observer can be immediately associated with the predictor of (2.29). Associating to (2.1), the observable canonical state space representation (to simplify the presentation, it is assumed that $d = 0$ and $n_A = n_B = r$):

$$x(t+1) = \begin{bmatrix} -a_1 & \vdots & & \\ \vdots & \vdots & I_{r-1} & \\ \vdots & \ddots & \dots & \dots \\ -a_r & & & \end{bmatrix} x(t) + \begin{bmatrix} b_1 \\ \vdots \\ \vdots \\ b_r \end{bmatrix} u(t)$$

$$y(t) = [1, 0, \dots, 0]x(t) \tag{2.34}$$

the observer will have the form:

$$\hat{x}(t+1) = \begin{bmatrix} -a_1 & \vdots & & \\ \vdots & \vdots & I_{r-1} & \\ \vdots & \ddots & \dots & \dots \\ -a_r & 0 & & 0 \end{bmatrix} \hat{x}(t) + \begin{bmatrix} b_1 \\ \vdots \\ \vdots \\ b_r \end{bmatrix} u(t) + \begin{bmatrix} p_1 - a_1 \\ \vdots \\ \vdots \\ p_n - a_n \end{bmatrix} [y(t) - \hat{y}(t)]$$

$$\hat{y}(t) = [1, \dots, 0]\hat{x}(t) \tag{2.35}$$

and the state error $\tilde{x}(t) = x(t) - \hat{x}(t)$ is governed by:

$$\tilde{x}(t+1) = \begin{bmatrix} -p_1 & \vdots & & \\ \vdots & \vdots & I_{r-1} & \\ \vdots & \ddots & \dots & \dots \\ -p_r & 0 & 0 & \end{bmatrix} \tilde{x}(t) \tag{2.36}$$

(i.e., the polynomial $P(q^{-1})$ defines the dynamics of the observer).

For the case $1 \le j \le d + 1$, one has the following result:

Theorem 2.3 *The filtered predicted value $P(q^{-1})\hat{y}(t+j)$ can be expressed as*:

$$P(q^{-1})\hat{y}(t+j) = F_j(q^{-1})y(t) + E_j(q^{-1})B^*(q^{-1})u(t+j-d-1) \tag{2.37}$$

with:

$$\deg P(q^{-1}) \le n_A + j - 1 \tag{2.38}$$

where $E(q^{-1})$ and $F(q^{-1})$ are solutions of the polynomial equation.

$$AE_j(q^{-1}) + q^{-j}F_j(q^{-1}) = P(q^{-1}) \tag{2.39}$$

with:

$$\begin{aligned} E_j(q^{-1}) &= 1 + e_1 q^{-1} + \dots + e_{j-1}q^{-j+1} \\ F_j(q^{-1}) &= f_0^j + f_1^j q^{-1} + \dots + f_{n_F}^j q^{-n_f}; \quad n_F \le \max(n_A - 1, n_p - j) \end{aligned} \tag{2.40}$$

and the prediction error is governed by:

$$P(q^{-1})[y(t+j) - \hat{y}(t+j)] = 0 \tag{2.41}$$

Proof Using (2.39) one can write:

$$P(q^{-1})y(t+j) = A(q^{-1})E_j(q^{-1})y(t+j) + F_j(q^{-1})y(t) \tag{2.42}$$

But taking into account (2.10), one finally gets:

$$P(q^{-1})y(t+j) = F_j(q^{-1})y(t) + E_j(q^{-1})B^*(q^{-1})u(t+j-d-1) \tag{2.43}$$

from which (2.37) is obtained. Subtracting (2.37) from (2.43), the prediction error equation (2.41) is obtained. □

Since one takes advantage of (2.10), it is clear that if $j > d+1$ (long range prediction), the predicted values $\hat{y}(t+j)$ will depend also on future input values (beyond t) and a "scenario" for these future values is necessary for computing the prediction (see Sect. 7.7).

The orders of polynomials $E_j(q^{-1})$ and $F_j(q^{-1})$ assure the unicity of the solutions of (2.39).

Equation (2.39) can be expressed in matrix form:

$$Mx^T = p \tag{2.44}$$

where:

$$\begin{aligned} x^T &= [1, e_1, \ldots, e_{j-1}, f_o^j, \ldots, f_{n_F}^j] \\ p &= [1, p_1, \ldots, p_{n_A}, p_{n_A+1}, \ldots, p_{n_A+j-1}] \end{aligned} \tag{2.45}$$

and M is a lower-triangular matrix of dimension $(n_A + j) \times (n_A + j)$.

$$M = \left.\left[\begin{array}{ccccccccc}
1 & 0 & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
a_1 & 1 & 0 & & & & & & \vdots \\
a_2 & a_1 & 1 & & & & & & \vdots \\
\vdots & & & 0 & & & & & \vdots \\
a_{j-1} & a_{j-2} & & 1 & 0 & & & & \vdots \\
a_j & a_{j-1} & & a_1 & 1 & 0 & & & \vdots \\
a_{j+1} & a_j & & \vdots & a_1 & 1 & 0 & & \vdots \\
\vdots & \vdots & & \vdots & \vdots & 0 & 1 & & \vdots \\
a_{n_A} & \vdots & & \vdots & \vdots & \vdots & 0 & & \vdots \\
0 & \vdots & & \vdots & \vdots & \vdots & \vdots & 0 & \vdots \\
\vdots & \vdots & \cdots & \cdots & \vdots & \vdots & \vdots & 1 & 0 \\
0 & \cdots & \cdots & 0 & a_{n_A} & 0 & \vdots & 0 & 1 \\
\multicolumn{5}{c}{\underbrace{\hphantom{a_{j-1}\quad a_{j-2}\quad 0\quad 1\quad a_{n_A}}}_{j}} & \multicolumn{4}{c}{\underbrace{\hphantom{0\quad 0\quad 0\quad 0}}_{n_A}}
\end{array}\right]\right\} n_A + j \tag{2.46}$$

Because of the structure of the matrix M (lower-triangular), there is always an inverse and one has:

$$x^T = M^{-1} p \tag{2.47}$$

Remark The j step ahead predictor can be obtained by successive use of the one step ahead predictor (see Problem 2.1). One replaces: $y(t+j-1)$ by $\hat{y}(t+j-1) = f[y(t+j-2)\ \ldots]$, then $y(t+j-2)$ by $\hat{y}(t+j-2) = f[y(t+j-3)\ \ldots]$ and so on.

Regressor Form

Taking into account the form of (2.37), the j step ahead predictor can be expressed also in a regressor form as:

$$P(q^{-1})\hat{y}(t+j) = \theta^T \phi(t) \tag{2.48}$$

where:

$$\theta^T = [f_0^j, \ldots, f_{n_F}^j, g_0, \ldots, g_{n_B+j-2}] \tag{2.49}$$

$$\phi^T(t) = [y(t), \ldots, y(t-n_F), u(t+j-d-1), \ldots, u(t-n_B-d+1)] \tag{2.50}$$

$$G_j(q^{-1}) = E_j B^* = g_0 + g_1 q^{-1} + \cdots + g_{j+n_B-2} q^{-j-n_B+2} \tag{2.51}$$

2.2 Stochastic Environment

2.2.1 *Input-Output Models*

In many practical situations, the deterministic input-output model given in (2.1) cannot take into account the presence of stochastic disturbances. A model is therefore needed which accommodates the presence of such disturbances.

An immediate extension of the deterministic input-output model is:

$$y(t+1) = -A^*(q^{-1})y(t) + q^{-d}B^*(q^{-1})u(t) + v(t+1) \tag{2.52}$$

where $v(t)$ is a stochastic process which describes the effect upon the output of the various stochastic disturbances. However, we need to further characterize this disturbance in order to predict the behavior of the system, and to control it.

Stationary stochastic disturbances having a rational spectrum can be modeled as the output of a dynamic system driven by a Gaussian white noise sequence (Factorization Theorem—Åström 1970; Faurre et al. 1979; Åström et al. 1984).

The Gaussian discrete-time white noise is a sequence of independent, equally distributed (Gaussian) random variables of zero mean value and variance σ^2. This sequence will be denoted $\{e(t)\}$ and characterized by $(0, \sigma)$, where the first number indicates the mean value and second number indicates the standard deviation.

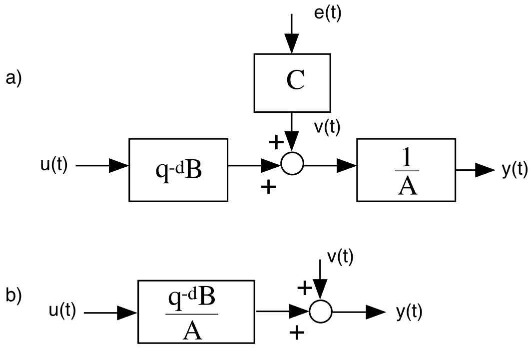

Fig. 2.1 Configuration for additive stochastic disturbances, (**a**) equation error model structure, (**b**) output error model structure

A large class of stochastic processes of interest for applications will be described by the output of a poles and zeros system driven by white noise called the Auto Regressive Moving Average (ARMA) process.

$$v(t) = \frac{C(q^{-1})}{D(q^{-1})} e(t) \tag{2.53}$$

where:

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_{n_C} q^{-n_C} \tag{2.54}$$

$$D(q^{-1}) = 1 + d_1 q^{-1} + \cdots + d_{n_D} q^{-n_D} \tag{2.55}$$

$D(z^{-1})$ has all its roots inside the unit circle and it will be assumed that $C(z^{-1})$ also has all its roots inside the unit circle.[3]

For the case $C(q^{-1}) = 1$, one has an autoregressive (AR) model and for the case $D(q^{-1}) = 1$ one has a moving average model (MA). However, $v(t)$ may act in different part of the system. Equation (2.52) corresponds to the block diagram shown in Fig. 2.1a (known also as the *equation error model*). If in addition $v(t)$ is modeled by (2.53) with $D(q^{-1}) = 1$, one obtains the Autoregressive Moving Average with Exogenous Input (ARMAX) model (Fig. 2.1a):

$$y(t+1) = -A^*(q^{-1})y(t) + q^{-d}B^*(q^{-1})u(t) + C(q^{-1})e(t+1) \tag{2.56}$$

or:

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})} u(t) + \frac{C(q^{-1})}{A(q^{-1})} e(t) \tag{2.57}$$

In general: $n_c \leq n_A$.

Equation (2.57) gives an equivalent form for (2.56) which corresponds to the disturbance added to the output, but filtered by $1/A(q^{-1})$. However, the disturbance may be represented as directly added to the output as shown in Fig. 2.1b.

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})} u(t) + v(t) \tag{2.58}$$

[3]Factorization theorem for rational spectrum allows for zeros of $C(z^{-1})$ on the unit circle (Åström et al. 1984).

This configuration is known as the *output error model* which can further be written as:

$$y(t+1) = -A^*(q^{-1})y(t) + q^{-d}B^*(q^{-1})u(t) + A(q^{-1})v(t) \tag{2.59}$$

A model of the form (2.53) can be associated to $v(t)$ in (2.58) (this is the Box-Jenkins model). However, in many cases, one simply makes the hypothesis that $\{u(t)\}$ and $\{v(t)\}$ are independent without considering a special model for $v(t)$ (except that it is assumed to be of finite power).

All these models allow a representation of the form:

$$y(t+1) = \theta^T \varphi(t) + w(t+1) \tag{2.60}$$

where $w(t+1)$ is different depending on the context, and θ and φ are given by (2.14) and (2.15), respectively.

2.2.2 Predictors for ARMAX Input-Output Models

In the stochastic context, the prediction takes its full significance since the future values of the output will be affected by the disturbances for $j \geq 1$. However, the information upon the disturbance model will be taken into account for constructing a predictor. We will consider the ARMAX model:

$$y(t+1) = -A^*(q^{-1})y(t) + q^{-d}B^*(q^{-1})u(t) + C(q^{-1})e(t+1) \tag{2.61}$$

The objective is to find a linear predictor as a function of the available data up to and including time t:

$$\hat{y}(t+j/t) = \hat{y}(t+j) = f[y(t), y(t-1), \ldots, u(t), u(t-1), \ldots] \tag{2.62}$$

such that:

$$\mathbf{E}\{[y(t+j) - \hat{y}(t+j)]^2\} = \min \tag{2.63}$$

One has the following results.

Theorem 2.4 (Åström 1970) *For the system of* (2.61) *provided that* $C(q^{-1})$ *is asymptotically stable and* $e(t)$ *is a discrete-time white noise*, *the optimal* j *step ahead predictor minimizing* (2.63) *is given by*:

$$\hat{y}(t+j) = \frac{F_j(q^{-1})}{C(q^{-1})}y(t) + \frac{E_j(q^{-1})B^*(q^{-1})}{C(q^{-1})}u(t+j-d-1) \tag{2.64}$$

where $F_j(q^{-1})$ *and* $E_j(q^{-1})$ *are solutions of the polynomial equation*:

$$C(q^{-1}) = A(q^{-1})E_j(q^{-1}) + q^{-j}F_j(q^{-1}) \tag{2.65}$$

where:

$$E_j(q^{-1}) = 1 + e_1 q^{-1} + \cdots + e_{j-1} q^{-j+1} \tag{2.66}$$

$$F_j(q^{-1}) = f_0^j + f_1^j q^{-1} + \cdots + f_{n_F} q^{-n_F}$$

$$n_F = n_A - 1;\ n_C \le n_A \tag{2.67}$$

and the optimal prediction error is a moving average given by:

$$\varepsilon(t+j)|_{opt} = y(t+j) - \hat{y}(t+j)\ |_{opt} = E_j e(t+j) \tag{2.68}$$

Note that for $j = 1$, *the optimal prediction error is a white noise.*

Proof From (2.65) one has:

$$C(q^{-1})y(t+j) = E_j A y(t+j) + F_j y(t) \tag{2.69}$$

and using (2.61) one has:

$$Ay(t+j) = B^* u(t+j-d-1) + C(q^{-1})e(t+j) \tag{2.70}$$

Introducing (2.70) in (2.69) one gets:

$$C(q^{-1})y(t+j) = F_j y(t) + E_j B^* u(t+j-d-1) + E_j C e(t+j) \tag{2.71}$$

and respectively:

$$y(t+j) = \frac{F_j}{C} y(t) + \frac{E_j B^*}{C} u(t+j-d-1) + E_j e(t+j) \tag{2.72}$$

The expression of $y(t+j)$ given by (2.72) has what is called an "innovation" form, i.e.:

$$\begin{aligned} y(t+j) &= f[y(t), y(t-1), \ldots, u(t+j-d-1), \ldots] \\ &\quad + g[e(t+1), e(t+2), \ldots, e(t+j)] \end{aligned} \tag{2.73}$$

where the first term depends on the data available up to and including time t, and the second term is the "true" stochastic term describing the future behavior which by no means can be predicted at time t. Introducing (2.72) in the criterion (2.63) leads to:

$$\begin{aligned} &\mathbf{E}\{[y(t+j) - \hat{y}(t+j)]^2\} \\ &\quad = \mathbf{E}\left\{\left[\frac{F_j}{C} y(t) + \frac{E_j B^*}{C} u(t+j-d-1) - \hat{y}(t+j)\right]^2\right\} \\ &\qquad + 2\mathbf{E}\left\{[E_j e(t+j)]\left[\frac{F_j}{c} y(t) + \frac{E_j B^*}{c} u(t+j-d-1)\right]\right\} \\ &\qquad + \mathbf{E}\{[E_j e(t+j)]^2\} \end{aligned} \tag{2.74}$$

The second term of the right hand side is null because $E_j e(t+j)$ contains $e(t+1)$, $\ldots, e(t+j)$, which are all independent of $y(t), y(t-1), \ldots, u(t+j-d-1), \ldots$. The third term does not depend on the choice of $\hat{y}(t+j)$ and therefore the minimum is obtained by choosing $\hat{y}(t+j)$ given in (2.64) such that the first term will become null.

Subtracting $\hat{y}(t+j)$ given by (2.64) from (2.72) gives the expression (2.68) of the prediction error. □

Rapprochement with the Deterministic Case

Comparing Theorem 2.4 with Theorem 2.3, one can see immediately that the predictors for deterministic case and the stochastic case are the same if $P(q^{-1}) = C(q^{-1})$.

In other words, a deterministic predictor is an optimal predictor in the stochastic environment for an ARMAX type model with $C(q^{-1}) = P(q^{-1})$. Conversely, the polynomial $C(q^{-1})$ will define the decaying dynamics of the prediction error in the deterministic case (and of the initial conditions in the stochastic case).

Rapprochement with the Kalman Predictor

Consider the ARMAX model of (2.61) with $j = 1, d = 0, n_B < n_A$, $n_C = n_A$ to simplify the discussion. This model admits a state space "innovation" representation in observable canonical form given by:

$$x(t+1) = \begin{bmatrix} -a_1 & \vdots & & \\ \vdots & \vdots & I_{n_A-1} & \\ \vdots & \vdots & \cdots & \cdots \\ -a_{n_A} & 0 & \cdots & 0 \end{bmatrix} x(t) + \begin{bmatrix} b_1 \\ \vdots \\ b_{n_B} \\ \vdots \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} c_1 - a_1 \\ \vdots \\ \vdots \\ c_{n_C} - a_{n_A} \end{bmatrix} e(t) \tag{2.75}$$

$$y(t) = [1, 0, \ldots, 0]x(t) + e(t) \tag{2.76}$$

The optimal predictor of (2.64) for $d = 0$, $j = 1$ can be rewritten as ($E_1 = 1$, $F_1 = C^* - A^*$):

$$C\hat{y}(t+1) = (C^* - A^*)y(t) + B^*u(t) \pm A^*\hat{y}(t) \tag{2.77}$$

yielding:

$$\hat{y}(t+1) = -A^*\hat{y} + B^*u(t) + (C^* - A^*)[y(t) - \hat{y}(t)] \tag{2.78}$$

i.e., the predictor is driven by the prediction error which is an innovation process.

The associated state space representation of the predictor takes the form:

$$\hat{x}(t+1) = \begin{bmatrix} -a_1 & \vdots & & \\ \vdots & \vdots & I_{n_A-1} & \\ \vdots & \ddots & \cdots & \cdots \\ a_{n_A} & 0 & \cdots & 0 \end{bmatrix} \hat{x}(t) + \begin{bmatrix} b_1 \\ \vdots \\ b_{n_B} \\ \vdots \\ 0 \end{bmatrix} u(t) + \begin{bmatrix} c_1 - a_1 \\ \vdots \\ \vdots \\ c_{n_A} - a_{n_A} \end{bmatrix} [y(t) - \hat{y}(t)] \tag{2.79}$$

$$\hat{y}(t) = [1, 0, \ldots, 0]\hat{x}(t) \tag{2.80}$$

which is nothing else than the steady state Kalman predictor for the system (2.75)–(2.76). The steady state Kalman predictor can be directly obtained by inspection, since the model (2.75)–(2.76) is in *innovation* form. The objective is to obtain asymptotically an output prediction error which is an innovation sequence, i.e.

$$\lim_{t\to\infty}[y(t)-\hat{y}(t)]=e(t) \tag{2.81}$$

Subtracting (2.79) from (2.75), one obtains for the dynamics of the state estimation error $\tilde{x}(t)=x(t)-\hat{x}(t)$:

$$\tilde{x}(t+1)=\begin{bmatrix} -c_1 & \vdots & & \\ \vdots & \vdots & I_{n_C-1} & \\ \vdots & \ddots & \dots & \dots \\ -c_{n_C} & 0 & \dots & 0 \end{bmatrix}\tilde{x}(t) \tag{2.82}$$

Therefore the poles associated with the state estimation error are defined by $C(z^{-1})$ which is assumed to be asymptotically stable.

Taking this into account, it results that:

$$\lim_{t\to\infty}[y(t)-\hat{y}(t)]=[1,0,\dots,0]\lim_{t\to\infty}\tilde{x}(t)+e(t)=e(t) \tag{2.83}$$

Regressor Form

Taking into account (2.64), the j step ahead predictor can be expressed as:

$$C\hat{y}(t+j)=F_j y(t)+G_j u(t+j-d-1) \tag{2.84}$$

where:

$$G_j=E_jB^*=g_0+g_1q^{-1}+\dots+g_{j+n_B-2}q^{-j-n_B+2} \tag{2.85}$$

from which one obtains:

$$C\hat{y}(t+j)=\theta^T\phi(t) \tag{2.86}$$

where:

$$\theta^T=[f_0^j,\dots,f_{n_F}^j,g_0,\dots,g_{j+n_B-2}] \tag{2.87}$$

$$\phi^T(t)=[y(t),\dots,y(t-n_F),u(t+j-d-1),\dots,u(t-n_B-d+1)] \tag{2.88}$$

An alternative regressor form is:

$$\hat{y}(t+j)=\theta_e^T\phi_e(t) \tag{2.89}$$

where:

$$\theta_e^T=[\theta^T,c_1,\dots,c_{n_C}] \tag{2.90}$$

$$\phi_e(t)=[\phi^T(t),-\hat{y}(t+j-1),\dots,-\hat{y}(t+j-n_C)] \tag{2.91}$$

For the case $j = 1$, the predictor equation takes the form:

$$\hat{y}(t+1) = (C^* - A^*)y(t) + q^{-d}B^*u(t) - C^*\hat{y}(t) \tag{2.92}$$

allowing the representation shown above. However, this predictor can also be rewritten as:

$$\hat{y}(t+1) = -A^*y(t) + q^{-d}B^*u(t) + C^*[y(t) - \hat{y}(t)] \tag{2.93}$$

leading to the regressor form representation:

$$\hat{y}(t+1) = \theta_e^T \phi_e(t) \tag{2.94}$$

where:

$$\theta_e^T = [a_1, \ldots, a_{n_A}, b_1, \ldots, b_{n_B}, c_1, \ldots, c_{n_C}] \tag{2.95}$$

$$\begin{aligned}\phi_e^T(t) = [&-y(t), \ldots, -y(t-n_A+1), u(t-d), \ldots, u(t-d-n_B+1),\\ &\varepsilon(t), \ldots, \varepsilon(t-n_C+1)]\end{aligned} \tag{2.96}$$

$$\varepsilon(t) = y(t) - \hat{y}(t) \tag{2.97}$$

Equation (2.93) can also be equivalently rewritten as (by adding $\pm A^*\hat{y}(t)$):

$$\hat{y}(t+1) = -A^*\hat{y}(t) + q^{-d}B^*u(t) + [C^* - A^*][y(t) - \hat{y}(t)] \tag{2.98}$$

leading to the regression form representation (2.94) where now:

$$\theta_e^T = [a_1, \ldots, a_{n_A}, b_1, \ldots, b_{n_B}, h_1, \ldots, h_{n_C}] \tag{2.99}$$

$$\begin{aligned}\phi_e^T(t) = [&-\hat{y}(t), \ldots, -\hat{y}(t-n_A+1), u(t-d), \ldots, u(t-d-n_B+1),\\ &\varepsilon(t), \ldots, \varepsilon(t-n_C+1)]\end{aligned} \tag{2.100}$$

$$h_i = c_i - a_i; \quad i = 1, \ldots, n_C \tag{2.101}$$

These different forms of one step ahead predictor will be used for the estimation of the parameters of a plant model in a stochastic environment.

An important characteristic of the predictors for ARMAX models is that the predicted values depend upon (i) current and previous values of the input and output and (ii) current and previous values of the prediction.

2.2.3 Predictors for Output Error Model Structure

Let us consider now the case where the stochastic disturbance $v(t)$ affects directly the output:

$$y(t+1) = \frac{q^{-d}B^*(q^{-1})}{A(q^{-1})}u(t) + v(t+1) \tag{2.102}$$

under the hypothesis that $\{v(t)\}$ and $\{u(t)\}$ are independent and that $\{v(t)\}$ has finite variance.

The objective is to find a linear predictor depending on the information available up to and including t which minimizes the criterion:

$$\mathbf{E}\{[y(t+1) - \hat{y}(t+1)]^2\} \tag{2.103}$$

Introducing (2.102) in (2.103), one gets:

$$\begin{aligned}
&\mathbf{E}\{[y(t+1) - \hat{y}(t+1)]^2\} \\
&\quad = \mathbf{E}\left\{\left[\frac{q^{-d}B^*}{A^*}u(t) - \hat{y}(t+1)\right]^2\right\} + \mathbf{E}\{v(t)^2\} \\
&\qquad + 2\mathbf{E}\left\{v(t+1)\left[\frac{q^{-d}B^*}{A^*}u(t) - \hat{y}(t+1)\right]\right\}
\end{aligned} \tag{2.104}$$

The third term of this expression is null, since $v(t)$ is independent with respect to $u(t), u(t-1), \ldots$ and linear combinations of these variables which will serve to generate $\hat{y}(t+1)$. The second term does not depend upon the choice of $\hat{y}(t+1)$ and the criterion will be minimized if the first term becomes null. This leads to:

$$\hat{y}(t+1) = \frac{q^{-d}B^*}{A}u(t) \tag{2.105}$$

or:

$$\hat{y}(t+1) = -A^*\hat{y}(t) + q^{-d}B^*u(t) \tag{2.106}$$

known as the *output error predictor*.

Its main characteristic is that the predicted output will depend only upon (i) the current and previous inputs and (ii) the current and previous predicted outputs (in the ARMAX case the predicted output depends also upon current and past measurements).

The output error predictor can be expressed also in the regressor form:

$$\hat{y}(t+1) = \theta^T\phi(t) \tag{2.107}$$

where:

$$\theta^T = [a_1, \ldots, a_{n_A}, b_1, \ldots, b_{n_B}] \tag{2.108}$$

$$\phi^T(t) = [-\hat{y}(t), \ldots, -\hat{y}(t-n_A+1), u(t-d), \ldots, u(t-d-n_B+1)] \tag{2.109}$$

2.3 Concluding Remarks

1. In a deterministic environment the discrete-time single input, single output models describing the systems to be controlled are of the form:

$$y(t) = -\sum_{i=1}^{n_A} a_i y(t-i) + \sum_{i=1}^{n_B} b_i u(t-d-i) \tag{$*$}$$

where $y(t)$ is the output and $u(t)$ is the input. Using the delay operator $q^{-1}(y(t-1) = q^{-1}y(t))$, the model $(*)$ takes the form:

$$y(t+1) = -A^*(q^{-1})y(t) + q^{-d}B^*(q^{-1})u(t)$$

where:

$$A(q^{-1}) = 1 + \sum_{i=1}^{n_A} a_i q^{-i} = 1 + q^{-1}A^*(q^{-1})$$

$$B(q^{-1}) = \sum_{i=1}^{n_B} b_i q^{-i} = q^{-1}B^*(q^{-1})$$

2. The filtered predicted values of $y(t)$, j steps ahead (for j up to $d+1$) can be expressed as a function of current and previous input-output measurements:

$$P(q^{-1})\hat{y}(t+j) = F_j(q^{-1})y(t+j) + E_j(q^{-1})B^*(q^{-1})u(t+j-d-1)$$

with:

$$\deg P(q^{-1}) \le n_A + j - 1$$

where $E_j(q^{-1})$ and $F_j(q^{-1})$ are solutions of the polynomial equation:

$$P(q^{-1}) = A(q^{-1})E_j(q^{-1}) + q^{-j}F_j(q^{-1})$$

3. In a stochastic environment, the input-output model takes the form:

$$y(t+1) = -A^*y(t) + q^{-d}B^*u(t) + v(t+1)$$

or:

$$y(t+1) = -A^*y(t) + q^{-d}B^*u(t) + Av(t+1)$$

where $v(t+1)$ represents the effect of a stochastic disturbance. The first model corresponds to the *equation error model* and the second model corresponds to the *output error model*.

4. Typical forms for the stochastic disturbances are:
   (i) $v(t) = \frac{C(^{-1})}{D(q^{-1})}e(t)$ where $e(t)$ is a zero mean white noise
   (ii) $\{v(t)\}$ is a zero mean stochastic process of finite power and independent of $u(t)$. For the case $v(t) = C(q^{-1})e(t)$, the *equation error model* is an ARMAX model:

$$y(t+1) = -A^*y(t) + q^{-d}B^*u(t) + Ce(t+1)$$

5. The optimal j step ahead predictor for ARMAX models has the form:

$$\hat{y}(t+j) = \frac{F_j}{C}y(t) + \frac{E_jB^*}{C}u(t+j-d-1)$$

where $E_j(q^{-1})$ and $F_j(q^{-1})$ are solutions of the polynomial equation:

$$C(q^{-1}) = A(q^{-1})E_j(q^{-1}) + q^{-j}E_j(q^{-1})$$

6. Predictors for the deterministic models and stochastic ARMA models are the same if $P(q^{-1}) = C(q^{-1})$.
7. The one step ahead optimal predictor for the *output error model*, when $v(t)$ is independent with respect to the input, is:

$$\hat{y}(t+1) = \frac{q^{-d}B^*}{A}u(t)$$

(it depends only upon the input).

2.4 Problems

2.1 Show that the j-step ahead predictor (2.37) for $y(t)$ given by (2.1) can be obtained by successive use of one step ahead predictors.

2.2 From (2.64) it results that the j-step ahead predictor of $y(t+1)$ given by (2.60) can be expressed as:

$$\hat{y}(t+j) = f[\hat{y}(t+j-1), \hat{y}(t+j-2), \ldots, y(t), y(t-1), \ldots, \\ u(t+j-d-1), u(t+j-d-2), \ldots]$$

Using a second polynomial division show that one can express $\hat{y}(t+j)$ as:

$$\hat{y}(t+j) = f[y(t), y(t-1), \ldots, u(t+j-d-1), u(t+j-d-2), \ldots]$$

2.3 Give a recursive formula for the computation of the polynomials $E_{j+1}(q^{-1})$ and $F_{j+1}(q^{-1})$ in (2.39), knowing the solutions $E_j(q^{-1})$ and $F_j(q^{-1})$.

2.4 Construct the one step ahead optimal predictor for:

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})}u(t) + \frac{1}{C(q^{-1})A(q^{-1})}e(t)$$

where $e(t)$ is a white noise.

2.5 Construct the one step ahead optimal predictor for:

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{D(q^{-1})}e(t)$$

where $e(t)$ is a white noise.

2.6 Consider the ARMAX model:

$$y(t) = \frac{q^{-d}B(q^{-1})}{A(q^{-1})}u(t) + \frac{C(q^{-1})}{A(q^{-1})}e(t)$$

where:

$$u(t) = -\frac{R(q^{-1})}{S(q^{-1})}y(t) + r(t)$$

Construct the one step ahead optimal predictor for $y(t)$.

2.7 Consider the continuous-time transfer function:

$$G(s) = \frac{Ge^{-s\tau}}{1 + sT}$$

Compute the zero-hold, discrete-time equivalent model. The relation between the continuous-time delay and the discrete-time delay is:

$$\tau = dT_S + L; \quad 0 < L < T_S$$

where T_S is the sampling period and L is the fractional delay. Examine the position of the discrete-time zeros for various values of L. Does the system has always stable discrete-time zeros?

2.8 Give examples of discrete-time models featuring:

(a) unstable discrete-time zeros but "minimum phase" behavior
(b) unstable discrete-time zeros and "non-minimum phase" behavior
(c) stable discrete-time zeros but "non-minimum phase" behavior

(non-minimum phase behavior = the time-response to a positive step starts with negative values).

2.9 Consider the discrete-time model:

$$\frac{b_1 q^{-1} + b_2 q^{-2}}{(1 + f_1 q^{-1})(1 + f_2 q^{-2})}$$

Verify that for $f_2 = -\frac{b_2}{b_1}$, the determinant of the Sylvester matrix (2.24) is null.

http://www.springer.com/978-0-85729-663-4

Adaptive Control
Algorithms, Analysis and Applications
Landau, I.D.; Lozano, R.; M'Saad, M.; Karimi, A.
2011, XXII, 587 p. 176 illus., 13 in color. With online files/update., Hardcover
ISBN: 978-0-85729-663-4