Chapter 4

PREDICTION AND MINIMUM VARIANCE CONTROL FOR ARMAX SYSTEMS

In this chapter, we discuss a class of estimation and control problems which can be solved using somewhat special tools. The estimation problem is that of optimal prediction of a scalar-valued process described by an ARMAX system. The optimal control problem is that of minimizing the output variance of the ARMAX system. Although the problem can be re-formulated into state space form (see Exercise 4.1), polynomial methods are simpler and easier to apply.

4.1 Prediction Theory for ARMAX Systems

Let z^{-1} denote the backward shift operator, i.e. $z^{-1}x_k = x_{k-1}$. Define the polynomials

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \\ B(z^{-1}) &= b_0 + b_1 z^{-1} + \cdots + b_m z^{-m} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \cdots + c_n z^{-n} \end{aligned}$$

and consider ARMAX systems of the form (y_k, u_k, w_k all scalar-valued)

$$A(z^{-1})y_k = z^{-d}B(z^{-1})u_k + C(z^{-1})w_k \tag{4.1}$$

where $d > 0$ is the delay in the system, $b_0 \neq 0$, w_k is a zero mean independent identically distributed white noise process with variance σ^2, and y and u form a jointly stationary process. We also assume that $C(z^{-1})$ is stable, i.e., that the poles of $C(z^{-1})$ all lie strictly inside the unit circle, and that at time k, $\{y_s, u_s, -\infty < s \leq k\}$ are known. The process $\{u_s, -\infty < s \leq k-1\}$ is assumed to be independent of $\{w_s, s \geq k\}$. The problem is to optimally predict the value of y_{k+d} in the sense of minimum mean square error based on the knowledge of $\{y_s, u_s, -\infty < s \leq k\}$.

By the division algorithm (i.e. use long division), we can always write

$$C(z^{-1}) = A(z^{-1})F(z^{-1}) + z^{-d}G(z^{-1}) \tag{4.2}$$

where $F(z^{-1}) = f_0 + f_1 z^{-1} + \cdots + f_{d-1}z^{-(d-1)}$ is a polynomial of degree $(d-1)$ and $G(z^{-1}) = g_0 + g_1 z^{-1} + \cdots + g_{n-1}z^{-(n-1)}$ is a polynomial of degree $(n-1)$. By applying $F(z^{-1})$ to both sides of (4.1), we get

$$F(z^{-1})A(z^{-1})y_k = z^{-d}B(z^{-1})F(z^{-1})u_k + F(z^{-1})C(z^{-1})w_k$$

so that

$$C(z^{-1})[y_k - F(z^{-1})w_k] = z^{-d}B(z^{-1})F(z^{-1})u_k + z^{-d}G(z^{-1})y_k$$

Hence

$$y_{k+d} - F(z^{-1})w_{k+d} = \frac{1}{C(z^{-1})}\{G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k\} \tag{4.3}$$

Since $C(z^{-1})$ is asymptotically stable, the right hand side of (4.3) is of the form $\sum_{i=0}^{\infty} l_i y_{k-i} + \sum_{i=0}^{\infty} m_i u_{k-i}$. If y_{k+d} is any estimate based on $\{y_s, u_s, -\infty < s \le k\}$, we get

$$\begin{aligned} E(y_{k+d} - \hat{y}_{k+d})^2 &= E\left\{\frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k] - \hat{y}_{k+d} + F(z^{-1})w_{k+d}\right\} \\ &= E\left\{\frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k] - \hat{y}_{k+d}\right\}^2 \\ &\quad + E[F(z^{-1})w_{k+d}]^2 \end{aligned} \tag{4.4}$$

using the fact that w_k is zero mean white noise and that $\{y_s, u_s, -\infty < s \le k\}$ is independent of $\{w_s, k+1 \le s \le k+d\}$. It is clear from (4.4) that the optimal estimate in the minimum mean square error sense is given by

$$\hat{y}_{k+d|k} = \frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k] \tag{4.5}$$

Alternatively, we can write (4.5) as

$$C(z^{-1})\hat{y}_{k+d|k} = G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k \tag{4.6}$$

The optimal prediction error is given by

$$E(y_{k+d} - \hat{y}_{k+d|k})^2 = E[F(z^{-1})w_{k+d}]^2 = \sum_{j=0}^{d-1} f_j^2\sigma^2 \tag{4.7}$$

The ARMAX prediction formula (4.6) is very useful in many applications.

If the stationarity assumption is relaxed, then the predictor given by (4.6) can be shown to be the predictor which minimizes the average prediction error criterion $\lim_{N\to\infty}\frac{1}{N}\sum_{k=1}^{N} E(y_{k+d} - \hat{y}_{k+d|k})^2$ (See Ex. 4).

4.2 Minimum Variance Control of ARMAX Systems

Consider again the ARMAX system (4.1). We now would like to determine the control input u_k as a function of past information for the purposes of optimizing a performance criterion. Admissible control laws are taken to be those of the form

$$u_k = f(k, y^k, u^{k-1}) \text{ where } y^k = \{y_k, y_{k-1}, \cdots\} \text{ etc.}$$

The performance criterion is to minimize Ey_{k+d}^2.

Again, we make use of (4.3). From (4.3), we have

$$y_{k+d} = \frac{1}{Cz^{-1}}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k] + F(z^{-1})w_{k+d} \tag{4.8}$$

From this or from the system equation (4.1), we see that u_k affects y_{k+d} but not previous outputs. We therefore try to choose u_k to minimize Ey^2_{k+d}. But

$$\begin{aligned} Ey^2_{k+d} &= E\left\{\frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k] + F(z^{-1})w_{k+d}\right\}^2 \\ &= E\left\{\frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k]\right\}^2 + E[F(z^{-1})w_{k+d}]^2 \end{aligned} \tag{4.9}$$

due to the independence of the sequence w_k. Hence (4.9) implies that the optimal control is given by

$$B(z^{-1})F(z^{-1})u_k = -G(z^{-1})y_k \tag{4.10}$$

with the optimal

$$Ey^2_{k+d} = \sum_{j=0}^{d-1} f_j^2\sigma^2$$

The optimal control law has the following interpretation. From the previous section, we know that the optimal prediction $\hat{y}_{k+d|k}$ is

$$\hat{y}_{k+d|k} = \frac{1}{C(z^{-1})}[G(z^{-1})y_k + B(z^{-1})F(z^{-1})y_k].$$

The minimum variance control law is thus obtained by setting the optimal predicted output equal to the desired output ($y_{k=d} = 0$).

If processes are not stationary, it can be shown, similar to the optimal prediction problem, that the control law (4.10) minimizes the average cost $\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E(y^2_{k+d})$ (See Ex. 4).

As an example, consider the difference equation

$$\begin{aligned} y_k + a_1y_{k-1} + \cdots + a_ny_{k-n} &= b_0u_{k-1} + \cdots + b_mu_{k-m-1} \\ &\quad +w_k + c_1w_{k-1} + \cdots + c_nw_{k-n} \\ &\qquad\qquad\qquad\qquad b_0 \neq 0 \end{aligned}$$

Since $d = 1$, it is readily seen that $F(z^{-1}) = 1$

$$\begin{aligned} G(z^{-1}) &= z[C(z^{-1} - A(z^{-1})] \\ &= (c_1 - a_1) + (c_2 - a_2)z^{-1} + \cdots + (c_n - a_n)z^{-(n-1)} \end{aligned}$$

(4.10) now gives

$$\begin{aligned} b_0u_k + b_1u_{k-1} + \cdots + b_mu_{k-m} &= (a_1 - c_1)y_k + (a_2 - c_2)y_k + \cdots \\ &\quad +(a_n - c_n)y_{k-n+1} \end{aligned}$$

The optimal control law is thus given by

$$u_k = -\frac{1}{b_0}[(c_1 - a_1)y_k + \cdots + (c_n - a_n)y_{k-n+1} + b_1u_{k-1} + \cdots + b_mu_{k-m}] \tag{4.11}$$

4.3 Limitations of Minimum Variance Control

While the minimum variance control law (4.10) is optimal, it can only be implemented when the polynomial $B(z^{-1})$ is stable. To see this, note that the closed loop system satisfy the following equation

$$\begin{bmatrix} A(z^{-1}) & -z^{-d}B(z^{-1}) \\ G(z^{-1}) & B(z^{-1})F(z^{-1}) \end{bmatrix} \begin{bmatrix} y_k \\ u_k \end{bmatrix} = \begin{bmatrix} C(z^{-1}) \\ 0 \end{bmatrix} w_k \tag{4.12}$$

The determinant of the matrix on the left hand side of (4.12) is given by $ABF + z^{-d}GB = CB$. Since $C(z^{-1})$ is assumed to be stable, the minimum variance control law gives rise to a stable closed loop system if and only if $B(z^{-1})$ is stable. Since the roots of $B(z^{-1})$ correspond to the zeros of the open loop transfer function from u to y, the requirement that $B(z^{-1})$ be stable is referred to as the minimum phase condition. Solving the closed loop system equation (4.12) explicitly yields

$$y_k = F(z^{-1})w_k \tag{4.13}$$

$$u_k = -\frac{G(z^{-1})}{B(z^{-1})}w_k \tag{4.14}$$

We see that if $B(z^{-1})$ is not stable, y_k will be bounded and have minimum variance, but u_k will generally be unstable and grow without bound.

4.4 Computing the G and F Polynomials

The determination of the G and F polynomials for the d-step ahead prediction formula through (4.2) is somewhat tedious and not readily done on a computer. There is a simple linear algebraic formula which determines the coefficients of the G and F polynomials. To simplify the notation, assume that the C polynomial is normalized to have leading coefficient $c_0 = 1$. This can be done without loss of generality. With this normalization, the constant term f_0 in the F polynomial is also 1. Since the G and F polynomials depend on the value of d, we write them as

$$G_d(z^{-1}) = g_{d,1} + g_{d,2}z^{-1} + \cdots + g_{d,n}z^{-(n-1)} = \sum_{j=1}^{n} g_{d,j}z^{-(j-1)}$$

$$F_d(z^{-1}) = 1 + f_{d,1}z^{-1} + f_{d,2}z^{-2} + \cdots + f_{d,d-1}z^{-(d-1)} = \sum_{j=0}^{d-1} f_{d,j}z^{-j}$$

Note that the degree of F_d increases with d, but the degree of G_d remains the same. Note also that G_d and F_d only depend on $A(z^{-1})$, $C(z^{-1})$, and d, but not on $B(z^{-1})$.

The observable representation of (4.1), for $B(z^{-1}) = 0$, is given by

$$\begin{aligned} x_{k+1} &= Ax_k + Kw_k \\ y_k &= Cx_k + w_k \end{aligned}$$

where

$$A = \begin{bmatrix} 0 & 0 & 0 & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & \ddots & 0 & \cdots & -a_{n-2} \\ \vdots & \cdots & \ddots & \cdots & \vdots \\ 0 & \cdots & \cdots & 1 & -a_1 \end{bmatrix} \tag{4.15}$$

$$K = \begin{bmatrix} c_n - a_n \\ c_{n-1} - a_{n-1} \\ \vdots \\ \vdots \\ c_1 - a_1 \end{bmatrix} \tag{4.16}$$

$$C = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$$

The following formulas determine the coefficients of the G_d and F_d polynomials. Let $p(k)$ be the solution of the equation

$$p(k+1) = Ap(k), \quad k \geq 1 \tag{4.17}$$

$$p(1) = K \tag{4.18}$$

Then

$$p(d) = A^{d-1}K = \begin{bmatrix} g_{d,n} \\ g_{d,n-1} \\ \vdots \\ g_{d,1} \end{bmatrix} \tag{4.19}$$

$$f_{d,j} = Cp(j), \quad 1 \leq j \leq d-1 \tag{4.20}$$

As an illustration, let

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + a_2 z^{-2} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + c_2 z^{-2} \end{aligned}$$

Let $d = 2$ so that we want to do 2-step ahead prediction. Then

$$f_{2,1} = Cp(1) = CK = c_1 - a_1$$

$$\begin{bmatrix} g_{2,2} \\ g_{2,1} \end{bmatrix} = p(2) = AK = \begin{bmatrix} -a_2(c_1 - a_1) \\ c_2 - a_2 - a_1(c_1 - a_1) \end{bmatrix}$$

This gives

$$F_d(z^{-1}) = 1 + (c_1 - a_1)z^{-1}$$

$$G_d(z^{-1}) = c_2 - a_2 - a_1(c_1 - a_1) - a_2(c_1 - a_1)z^{-1}$$

One can easily verify that the same polynomials are obtained using the division algorithm.

4.5 Exercises

1. Consider the scalar stationary ARMA process y_k described by

$$y_k + a_1 y_{k-1} + \ldots + a_n y_{k-n} = w_k + c_1 w_{k-1} + \ldots + c_n w_{k-n}$$

where w_k is a zero mean i.i.d. sequence with variance σ^2. We know that y_k can be described by the following state space model

$$x_{k+1} = \begin{bmatrix} 0 & & & 0 & -a_n \\ 1 & & & & -a_{n-1} \\ 0 & & & & \\ \vdots & & & 0 & \vdots \\ 0 & \cdots & 0 & 1 & -a_1 \end{bmatrix} x_k + \begin{bmatrix} c_n - a_n \\ \vdots \\ c_1 - a_1 \end{bmatrix} w_k$$

$$y_k = [0 \cdots 0 \quad 1] x_k + w_k$$

   (a) Verify that the system is always detectable.

   (b) What is the desired stabilizability condition to guarantee the uniqueness of the positive semidefinite solution to the ARE for this problem?

   (c) Assume, henceforth, that the stabilizibility condition you found in (b) holds. Determine the unique $P \geq 0$ solving the ARE.

   (d) Determine the steady state one-step ahead Kalman filter. Denote the one-step ahead predictor by $\hat{x}_{k+1/k}$. How is it related to $\hat{y}_{k+1/k}$?

   (e) Solve the one-step ahead prediction problem for y_k, i.e. determine $\hat{y}_{k+1/k}$, directly using ARMA prediction formula. Verify that the $\hat{y}_{k+1/k}$ determined in (d) and (e) are identical.

2. It is sometimes convenient to use a different state space representation than the observable representation for ARMA processes. Here we describe another state space representation. Consider the ARMA process

$$A(z^{-1}) y_k = C(z^{-1}) w_k \qquad (ex4.1)$$

where

$$\begin{aligned} A(z^{-1}) &= 1 + a_1 z^{-1} + \ldots + a_n z^{-n} \\ C(z^{-1}) &= 1 + c_1 z^{-1} + \ldots + c_n z^{-n} \end{aligned}$$

$C(z^{-1})$ is assumed to be a stable polynomial. Let

$$A = \begin{bmatrix} -a_1 & 1 & & 0 \\ \vdots & & \ddots & \\ -a_n & & & 1 \\ 0 & & & 0 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 \\ c_1 \\ \vdots \\ c_n \end{bmatrix}$$

$$C = [1 \quad 0 \ldots 0]$$

Note that A is $(n+1) \times (n+1)$.

(a) Show that the output y_k of the state space system

$$\begin{aligned} x_{k+1} &= Ax_k + Gw_{k+1} \\ y_k &= Cx_k \end{aligned} \qquad (ex4.2)$$

is the same as that of the ARMA representation (ex4.1).
(Hint: Find $C(zI - A)^{-1} = [\xi_1(z) \dots \xi_n(z)]$ by solving

$$[\xi_1(z) \dots \xi_n(z)](zI - A) = C \;)$$

(b) Write down the steady state Kalman filter equations for (ex4.2). Find a solution to the algebraic Riccati equation by inspection.

(c) Can you show that it is the unique solution?
(Hint: You need to transform the equation so that it is not a singular Riccati equation. To that end, set $\tilde{P} = P - GG^T$, derive an equation for $\tilde{P}$, and show that $\tilde{P} = 0$ is the unique solution.)

3. Consider the scalar stationary ARMAX process y_k described by

$$y_k + a_1 y_{k-1} + \dots + a_n y_{k-n} = u_{k-1} + w_k + c_1 w_{k-1} + \dots + c_n w_{k-n}$$

where w_k is a zero mean i.i.d. sequence with variance σ^2, and $C(z^{-1})$ is assumed to be a stable polynomial. Determine the minimum variance control law in this special case and the behaviour of the closed-loop system.

4. If the stationarity assumption made in this chapter is relaxed, we can, with a more complicated argument, show that the predictor given by (4.6) minimizes an average mean square error criterion. This problem guides you through the argument.

(a) From (4.3), we see that any $\hat{y}_{k+d|k}$ gives rise to the equation

$$C(z^{-1})[y_{k+d} - \hat{y}_{k+d|k} - F(z^{-1})w_{k+d}] = G(z^{-1})y_k + B(z^{-1})F(z^{-1})u_k - C(z^{-1})\hat{y}_{k+d|k}$$

Show that this results in

$$y_{k+d} - \hat{y}_{k+d|k} - F(z^{-1})w_{k+d} = \gamma_k + \eta_k \qquad (ex4.3)$$

where γ_k depends on initial conditions and converges to 0 geometrically fast, while η_k depends on $y_s,\ u_s,\ 0 \le s \le k$.

(b) Use $(ex4.3)$ to show that for any $\hat{y}_{k+d|k}$

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E(y_{k+d} - \hat{y}_{k+d|k})^2 \ge \sigma^2 \sum_{j=0}^{d-1} f_j^2$$

(c) Finally, show that the predictor given by (4.6) satisfies

$$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E(y_{k+d} - \hat{y}_{k+d|k})^2 = \sigma^2 \sum_{j=0}^{d-1} f_j^2$$

This proves that this predictor minimizes the average mean square error criterion
$\lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} E(y_{k+d} - \hat{y}_{k+d|k})^2$.

(d) Imitate the above argument to show that if stationarity is not assumed, the control law (4.10) minimizes the average variance criterion $\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E(y_{k+d}^2)$.

5. The results presented in this chapter consider the minimum variance control problem of minimizing $E(y_{k+d}^2)$. The results can be generalized to minimizing the mean square tracking error with respect to a known deterministic reference trajectory. Let y_{k+d}^* be the known deterministic reference trajectory. For the standard ARMAX model described in this chapter, determine the control law which minimizes $E(y_{k+d} - y_{k+d}^*)^2$.