Indirect Stochastic Adaptive Control Using Optimal Joint Parameter and State Estimation

by

Behnam Shahrrava

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Electrical Engineering

Waterloo, Ontario, Canada, 1998

0-612-32855-4

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

The SISO and MIMO adaptive problems are approached via the theory of stochastic optimal control. using results in the literature and previously-unexploited propertites of canonical singular-pencil dynamical models. For both the SISO and MIMO cases, it is shown how to choose the input to minimize the sample mean-square error between the filtered output and the filtered desired output.

In the SISO case, the proposed method. which is an indirect scheme. leads to control laws which not only use the estimates of the parameters and the states but also the variances of the estimation errors. In other words. the SISO cautious control problem is solved for the general delay-white noise case.

In the MIMO case. it is proved that the cautious control problem can be solved only if the interactor matrix of the system is known. However. if the interactor matrix is unknown, there will be no exact solution for the optimal problem. Hence. in this research the certainty equivalence principle is employed based on the result obtained by Das [17] and a priori knowledge of the degrees of the diagonal entries of the interactor matrix. It is shown that the conventional approaches discussed in the literature cannot be applied to a stochastic case. However, the indirect approach proposed in this research is capable of dealing with stochastic cases.

Simulation results show the performance of the proposed methods both in the SISO case and the MIMO case.

Acknowledgements

The author wishes to express his gratitude to those who have made this thesis possible.

In particular, I would like to thank my supervisor, Professor J. D. Aplevich, for his invaluable guidance and support throughout the duration of this thesis.

Thanks are also extended to Professors A. S. Morse, W. J. Wilson, A. J. Heunis, and G. R. Heppler for their useful comments and suggestions.

Finally, I wish to thank my wife, Farzaneh, and my daughter, Niki, for their encouragement and patience during the writing of this thesis.

To my family

Contents

List of Figures

Chapter 1

Introduction

This thesis is concerned with indirect stochastic adaptive prediction and control. Although there has been a considerable progress in stochastic adaptive prediction and control in the past twenty years, the challenge of extending existing results to general cases remains.

In this research, a number of open problems in single-input single-output (SISO) stochastic adaptive prediction and cautious control as well as in multi-input multi-output (MIMO) stochastic adaptive control have been tackled. The problems which have been solved are:

- Adaptive prediction that is optimal in a minimum variance sense at each step for SISO linear time-invariant discrete-time systems having general delay and white noise perturbation.
- Indirect SISO cautious control for the general delay-white noise case.
- MIMO stochastic adaptive control with unknown interactor matrix using an indirect approach.

- A new derivation of a previously-known important result on the optimality of the joint parameter and state estimation problem for a particular canonical form is given in Appendix A.

Stochastic adaptive control has become an important area of activity in control theory. This activity has developed especially during the past twenty years or so because many mathematical models of physical phenomena are stochastic systems which contain unknown parameters and for which control is required. With the maturity of this area of research there are a number of books and general surveys, for example, Wittenmark [53], Goodwin and Sin [28], Kumar [32], Caines [12], Åström and Wittenmark [10], Narendra, Ortega, and Dorato [41], and Pasik-Duncan [43]. The chief advantage of the stochastic formulation of adaptive problems is its ability to exploit the averaging properties of the disturbances [44]. This allows it to provide better performance in terms of disturbance rejection. The mathematical foundation of this field was laid by Goodwin *et al.* [27], and Solo [49], and the various ramifications were explored in Goodwin and Sin [28].

Adaptive prediction: In Chapter 2 a deficiency in SISO stochastic adaptive prediction related to transient performance is discussed. Using an indirect adaptive scheme, it is shown that in the general delay-white noise case how the optimal prediction can be obtained. First using state variables, the ARX model is transformed into a special canonical model proposed by Salut [45]. The identification is done by augmenting the state with the unknown parameters of the process. This usually leads to a nonlinear filtering problem. By choosing the model of special structure proposed by Salut [45] the problem can, however, be reduced to a linear problem, discussed in detail by Aplevich [3]. Taking advantage of the linearity of this canonical model, the optimal joint parameter and state estimation can be provided by a Kalman filter, [45, 46, 16, 4]. Although the use of the Kalman filter to identify system parameters is well known [42, 34], the joint estimation of the system parameters and states in the filter is not common in the literature.

For a unit delay system, the Kalman filter applied to the model proposed by Salut can be used as an adaptive optimal predictor, since the Kalman filter corresponds exactly to the optimal one-step-ahead predictor for the data observed.

For delays d greater than unity, the proposed approach leads to the necessity of using the theory of Gaussian random functions. The calculations outlined in the proposed approach are based on one of the most useful properties of Gaussian distribution which stems from the fact that the higher moments of the multidimensional Gaussian distribution depend only on the elements of the mean vector and the covariance matrix, [36]. Finally, it is described how the system parameter-state estimates and estimates of their uncertainty, provided by the Kalman filter, can be used to construct an optimal d-step-ahead predictor.

The conventional adaptive predictors discussed in Åström and Wittenmark [9], Wittenmark [52], Goodwin and Sin [28], and Ren and Kumar [44], approximately solve the optimal problem using the certainty equivalence principle. This approach makes sense when the parameter estimates are close to the actual plant parameters. However, in general there will be circumstances where the transient performance is unnecessarily poor. As a result, the conventional predictors are asymptotically optimal, proved by Sin, Goodwin, and Bitmead [47], and Ren and Kumar [44], whereas the proposed predictor is optimal in a minimum variance sense at each step resulting in good transient performance as well as asymptotic performance.

Indirect SISO cautious control: In a survey of stochastic adaptive control methods by Wittenmark in 1975, [53] the SISO cautious control problem has been discussed in Åström and Wittenmark [8], Wieslander and Wittenmark [51], and

Nahorski and Vidal [40]. The method proposed in this research and the method used by Åström and Wittenmark [8] are the same except that the proposed method can handle cases with uncertainty in initial conditions. Wieslander and Wittenmark [51] used a very complicated way to solve the adaptive problem for a unity delay system. Nahorski and Vidal [40] handle the case of general delay by using a special model.

In Chapter 3, an open problem in SISO adaptive control is solved. The adaptive problem is approached via the theory of stochastic optimal control following the philosophy of Åström [5, 6], and Wieslander and Wittenmark [51]. The system model and identification process are the same as employed in Chapter 2. The purpose of the control is to minimize the variance of the output around a desired value one step ahead. It is shown how the general one-step ahead optimization problem can be solved using stochastic control theory. Minimization over only one step leads to the one-step ahead or cautious controller. This controller takes the parameter uncertainties into account, in contrast to the certainty equivalence controller. Using an indirect scheme and the same approach as used in Chapter 2, it is proved that the optimal solution can be expressed analytically in a closed form. It is also shown if the conditional mean of the noise seen by the predictor obtained in Chapter 2 is zero, the optimal solution does not depend on the model used and the approach applied, direct or indirect. Finally, using the same property of Gaussian distribution employed in Chapter 2, the indirect method proposed in this research leads to control laws which not only use the estimates of the parameters and states but also the variances of the estimation errors. The simulation results show the performance of the proposed algorithm.

MIMO stochastic adaptive control: In Chapter 4 the MIMO stochastic adaptive control is discussed. There have been a number of important contributions to the multivariable adaptive control literature (e.g. Borrison. 1979 [11]; Koivo. 1980 [31]; Goodwin *et al.*, 1981 [27]; Morse. 1981 [39]; Elliot and Wolovich, 1982 [21]; Johansson, 1982 [30]; Dugard, Goodwin and deSouza. 1983 [19]; Goodwin and Dugard, 1983 [26]; Dugard, Goodwin and Xianya, 1984 [20]; Singh and Narendra. 1984 [48]; Elliot, Wolovich and Das, 1984 [50]; Goodwin and Sin, 1984 [28]; Das. 1986 [17]; Chen and Guo, 1987 [14]; Zhang and Lang, 1989 [55]; Guo and Chen, 1991 [29]; Mo and Bayoumi. 1993 [37]). Most of these have used direct strategies. In Chapter 4,the interest in direct strategies for MIMO systems is explained. It is also explained why an indirect scheme is the only way to solve the optimal problem considered.

The concept of an interactor matrix which is a generalized concept of a relative degree of a scalar system is reviewed. This matrix which describes the delay structure of MIMO system has an important role in multivariable control theory. [21. 28]. A procedure for constructing the interactor matrix was given by Wolovich and Falb (1976) [54]. Furuta and Kamiyama (1977) [25] gave a method for computing the Wolovich-Falb interactor from a state-space representation of the plant. Chang and Wang (1990) [13] have presented an algorithm for computing the interactor matrix form a right matrix fraction description (RMFD) of the plant.

As Dugard and Goodwin mentioned in one of their papers [20], if the interactor matrix is assumed known, then it is possible to develop a stochastic adaptive control algorithm which is globally convergent. However, the requirement of the knowledge of the interactor matrix imposes certain difficulties, in practice. In general, this matrix contains noninteger valued real variables, a knowledge of which will imply that the complete system transfer function is known. There is thus strong motivation to investigate ways in which the requirement of knowing the system interactor matrix might be removed. One idea, first proposed in Johansson (1982) [30] and further explored in Dugard, Goodwin, and deSouza (1983) [19], is to estimate the noninteger valued variables in the interactor along with the other system parameters. It was shown that one can still design a model-matching adaptive controller, if instead of knowing the complete interactor matrix, one presumes knowledge of the degrees of its diagonal entries as well as upper bounds on the degrees of its nondiagonal entries. Das (1986) [17] showed that the priori knowledge of the degrees of the diagonal entries of the interactor matrix alone is sufficient for proving global stability. Dion, Dugard, and Carrillo (1988) [18] verified the result obtained by Das [17] using a different approach. An alternative idea, due to Singh and Narendra (1984) [48], is to apply a suitably chosen precompensator to the system which transforms the interactor to a diagonal matrix having only integer valued parameters. This ingenious idea works for a large class of multivariable systems, though certain cases are excluded.

The new indirect approach proposed in this research is based on the result obtained by Das [17]. Here the proposed approach involves first estimating the parameters and states in a standard model such as the canonical ARX model or the canonical singular pencil (SP) model and then evaluating the control law by on-line calculations. These canonical models have been discussed in detail by Aplevich [4, 3]. Then it is shown how the entries of the interactor matrix are related to the entries of the canonical ARX model using two different methods. The first method uses the algorithm described in Furuta and Kamiyama (1977) [25] which involves nontrivial computations including tests for linear dependence. The second method is the method proposed in this research based on the result obtained by Das [17] and a priori knowledge of the degrees of the diagonal entries of the interactor matrix. The extension of the SISO cautious control to the MIMO case is examined. As for the SISO case, it is shown how the general optimal solution of the one-step-

step ahead criterion, defined for a MIMO system with unknown interactor matrix, can be obtained. Using this solution, it is shown that the MIMO cautious control can be designed only if the interactor matrix is known. However, if the interactor matrix is unknown, an exact solution for the MIMO one-step-ahead optimization problem cannot be obtained. Hence, the most popular approximation method, i.e. the certainty equivalence principle, is employed. Then it will be shown that the conventional approaches, [30, 19, 17], cannot be applied in the stochastic case. However, the proposed approach is capable of dealing with stochastic cases. At the end of Chapter 4, simulation results show the performance of the proposed algorithm.

A summary of the results obtained in this research and suggestions for further studies are given in Chapter 5.

Appendix A is concerned with estimating the state variables and parameters of discrete-time, linear stochastic SISO and MIMO time-varying systems. This topic has been discussed in detail by Salut [45]. Nevertheless, using a different method from that given by Salut, it is shown how the simultaneous estimation of state variables and system parameters with known noise statistics can be solved as an optimal linear filtering problem. In all the chapters to deal with cases with uncertainty in initial conditions, it is necessary that not only the parameters of the system but also the states of the system are estimated in an optimal sense. Therefore, the appendix has been added to this thesis.

Chapter 2

Adaptive optimal prediction

2.1 Introduction

This chapter is concerned with indirect stochastic adaptive prediction. The goal is to obtain an adaptive predictor that is optimal in a minimum variance sense at each step for SISO linear time-invariant discrete-time systems having general delay and white noise perturbation.

The conventional adaptive predictors discussed in Åström and Wittenmark [9], Wittenmark [52], Goodwin and Sin [28], and Ren and Kumar [44], approximately solve the adaptive optimal prediction problem using the certainty equivalence principle. This approach makes sense when the parameter estimates are close to the actual plant parameters. However, in general there will be circumstances where the transient performance is unnecessarily poor. As a result, the conventional predictors are asymptotically optimal, proved by Sin, Goodwin, and Bitmead [47], and Ren and Kumar [44], whereas the proposed predictor is optimal in a minimum variance sense at each step resulting in good transient performance as well as asymptotic performance.

This chapter is organized as follows. In Section 2.2 the system model is described. Using state variables, it is shown how an ARX model can be transformed into a special model proposed by Salut [45]. In Section 2.3 the identification is done by applying a Kalman filter to the model proposed by Salut to provide the optimal joint parameter and state estimation. Section 2.4 contains the estimation criterion of interest. In Section 2.5 it is shown how the system model can be transformed into various predictor forms.

Section 2.6 contains the main results. In Section 2.6.1 the conventional approach that is an approximation method based on the certainty equivalence principle is discussed. Section 2.6.2 contains the proposed approach based on one the most useful properties of Gaussian distribution. This property stems from the fact that the higher moments of the multi-dimensional Gaussian distribution depend only on the elements of the mean vector and the covariance matrix. Finally, it is described how the system parameter-state estimates and estimates of their uncertainty, provided by the Kalman filter, can be used to construct an optimal d-step-ahead predictor. A summary is given in Section 2.7.

2.2 System description

A time-invariant single-input, single-output system is assumed to be represented by the following ARX model:

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \tag{2.1}$$

where $\{y_k\}$, $\{u_k\}$, and $\{e_k\}$ denote the output, input and disturbance sequences respectively, and $A(q^{-1})$, and $B(q^{-1})$ are polynomial functions of q^{-1} which have the following forms:

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \tag{2.2a}$$

$$B(q^{-1}) = q^{-d}(b_0 + b_1 q^{-1} + \cdots + b_{n-d} q^{-(n-d)}) \; ; \; (n \geq d) \tag{2.2b}$$

where q^{-1} denotes the backward shift operator. The time delay is $d \geq 1$ and is chosen so that the leading coefficient of $B(q^{-1})$, b_0, in the model (2.1) is nonzero. The independent sequence $\{e_k\}$ is a gaussian zero mean white noise process with given covariance

$$E\{e_k e_l\} = R_k \delta_{k-l} \tag{2.3}$$

where δ_{k-l} is the Kronecker delta.

Using state variables, the model (2.1) can be written in the following canonical form:

$$\begin{bmatrix} x_{k+1}^{(n-1)} \\ x_{k+1}^{(n-2)} \\ \vdots \\ x_{k+1}^{(d-1)} \\ x_{k+1}^{(d-2)} \\ \vdots \\ x_{k+1}^{(0)} \end{bmatrix} = \begin{bmatrix} -a_n y_k + b_{n-d} u_k \\ x_k^{(n-1)} - a_{n-1} y_k + b_{n-d-1} u_k \\ \vdots \\ x_k^{(d)} - a_d y_k + b_0 u_k \\ x_k^{(d-1)} - a_{d-1} y_k \\ \vdots \\ x_k^{(1)} - a_1 y_k \end{bmatrix} \tag{2.4a}$$

$$y_k = x_k^{(0)} + e_k \tag{2.4b}$$

or putting

$$x_k = [\; x_k^{(n-1)} \; \ldots \; x_k^{(d-1)} \; x_k^{(d-2)} \; \ldots \; x_k^{(0)} \;]^T \tag{2.5}$$

Then the model (2.4) can be written in the following compact form:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k \tag{2.6a}$$

$$y_k = E_0 x_k + e_k \tag{2.6b}$$

where

$$E_* = \begin{bmatrix} 0 & . & . & . & . & 0 \\ 1 & & & & & \vdots \\ & . & & & & 0 \\ & & . & & & 0 \\ & & & . & & \vdots \\ & & & & 1 & 0 \end{bmatrix}, \; A_* = \begin{bmatrix} a_n \\ \vdots \\ a_d \\ a_{d-1} \\ \vdots \\ a_1 \end{bmatrix}, \; B_* = \begin{bmatrix} b_{n-d} \\ \vdots \\ b_0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

$$E_0 = [\, 0 \; . \; . \; . \; 0 \;\; 1 \,]. \tag{2.7}$$

Note that the model (2.6) is not a standard state-space model; however, a standard state-space model can easily be obtained by substituting (2.6b) into (2.6a) as follows:

$$x_{k+1} = \bar{A}x_k + \bar{B}u_k + w_k \tag{2.8a}$$

$$y_k = \bar{C}x_k + e_k \tag{2.8b}$$

where

$$\bar{A} = E_* - A_* E_0 \,, \quad \bar{B} = B_* \,, \quad \bar{C} = E_0 \,, \quad w_k = -A_* e_k \tag{2.8c}$$

and

$$E\left\{ \begin{bmatrix} w_k \\ e_k \end{bmatrix} \begin{bmatrix} w_l \\ e_l \end{bmatrix}^T \right\} = \begin{bmatrix} Q_k & S_k \\ S_k^T & R_k \end{bmatrix} \delta_{k-l} \tag{2.8d}$$

The model (2.8) represents the ARX model (2.1) in observer state-space form. In this model, the system matrix $\bar{A}$, Q_k, and S_k all depend on the system parameters in $A(q^{-1})$, and $\bar{B}$ depends on the parameters in $B(q^{-1})$.

In the next section, it will become clear that why model (2.6) is used to determine the optimal joint parameter and state of the system, rather than using the standard state-space model (2.8).

2.3 Parameter and state estimation

In this section, the approach proposed by Salut in 1976, [45], to determine the optimal joint parameter and state of the system is described. This section can also be considered as a preliminary section for the next chapter.

For parameter estimation and state estimation, the measurements $\{y_k\}$ and $\{u_k\}$ up to and including time k are assumed to be known. Furthermore, the matrices E_* and E_0 are known, such that the matrix $E = \begin{bmatrix} E_* \\ E_0 \end{bmatrix}$ is zero except for a positive unit element in each column. Thus in eqn. (2.6), the entries in the unknown state vector, x_k, and the unknown system parameters in A_* and B_* are multiplied by known values. This crucial fact allows us to solve the simultaneous state and parameter estimation problem as a linear filtering problem. On the other hand, using the standard state-space model (2.8) with unknown matrix $\bar{A} = E_* - A_* E_0$, the estimation problem leads to a nonlinear filtering problem due to the occurrence of products between parameters and states. Furthermore, the optimality of the Kalman filter depends on the a priori knowledge of the noise statistics. For the model (2.8), this knowledge is not available when the system parameters are unknown, because Q_k and S_k depend on the system parameters.

Now it is proceeded to describe the approach proposed by Salut. If θ consisting of the unknown parameters of the system (2.6) is defined as follows:

$$\theta = [a_1, \ldots, a_n, b_0, \ldots, b_{n-d}]^T \tag{2.9}$$

then the model (2.6) can be rearranged in the following form:

$$x_{k+1} = E_* x_k + G_*(y_k, u_k)\theta \tag{2.10a}$$

$$y_k = E_0 x_k + e_k \tag{2.10b}$$

where the matrix $G_*(y_k, u_k)$ is constructed to account for the order of the parameters in θ. Since the system is time invariant, $\theta_{k+1} = \theta_k = \theta$. By appending the parameter vectors θ_k to the state vector, the following model is obtained :

$$\begin{bmatrix} x_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} E_* & G_*(y_k, u_k) \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} \tag{2.11a}$$

$$y_k = \begin{bmatrix} E_0 & 0 \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + e_k \, . \tag{2.11b}$$

This model can be written in compact form as

$$s_{k+1} = F_k s_k \tag{2.12a}$$

$$y_k = H s_k + e_k \tag{2.12b}$$

where s_k is an $r = [n + (2n - d + 1)]$-dimensional state vector $[\, x_k^T \;\; \theta_k^T \,]^T$. Since the sequence $\{e_k\}$ is a gaussian zero mean white noise process with given covariance R_k, e_k has the following three properties which will be frequently used later:

Noise properties

$$E\{e_k | \mathcal{F}_{k-1}\} = 0 \qquad \text{a.s.} \tag{2.13a}$$

$$E\{e_k^2 | \mathcal{F}_{k-1}\} = R_k > 0 \qquad \text{a.s.} \tag{2.13b}$$

$$\lim_{N\to\infty} \sup \frac{1}{N} \sum_{k=1}^{N} e_k^2 < \infty \qquad \text{a.s.} \tag{2.13c}$$

where $\mathcal{F}_k$ denotes the sigma algebra generated by $y_0, \ldots, y_k$ and $u_0, \ldots, u_k$.

Now a Kalman filter for the augmented model (2.12) can be immediately obtained to give the optimal one-step-ahead estimate of the composite state vector. Let $\hat{s}_{k+1|k}$ denote the conditional mean of s_{k+1} given observations of $\{y_k\}$ and $\{u_k\}$ up to and including time k, that is,

$$\hat{s}_{k+1|k} = E\{s_{k+1} | \mathcal{F}_k\} \, . \tag{2.14}$$

In order to apply a Kalman filter to estimate the composite state s_k in (2.12), it is assumed that s_0 is a gaussian random variable of known mean $\bar{s}_0$ and known covariance Σ_0, i.e.,

$$\bar{s}_0 = E\{s_0 | \mathcal{F}_{-1}\} = \hat{s}_{0|-1} \tag{2.15}$$

$$\Sigma_0 = E\{(s_0 - \hat{s}_{0|-1})(s_0 - \hat{s}_{0|-1})^T | \mathcal{F}_{-1}\} = P_{0|-1} \geq 0 \,. \tag{2.16}$$

Further, it shall be assumed that s_0 is independent of e_k for any k.

The optimal predictor for the composite state has the following form, (see proof in Appendix A:

$$\hat{s}_{k+1|k} = F_k \hat{s}_{k|k-1} + K_k(y_k - H\hat{s}_{k|k-1})\,; \qquad \hat{s}_{0|-1} = \bar{s}_0 \tag{2.17a}$$

$$K_k = F_k P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1} \tag{2.17b}$$

$$P_{k+1|k} = F_k P_{k|k-1} F_k^T - K_k(H P_{k|k-1} H^T + R_k) K_k^T\,: \quad P_{0|-1} = \Sigma_0 \,. \tag{2.17c}$$

Because of the dependence of F_k on the actual measurements, u_k and y_k, the Kalman gain K_k, and the conditional error covariance $P_{k+1|k}$, are not precomputable off-line. The corresponding gain and state covariance matrix can be partitioned as follows:

$$K_k = \begin{bmatrix} K_{x_k} \\ K_{\theta_k} \end{bmatrix}, \qquad P_{k|k-1} = \begin{bmatrix} P_{xx_k} & P_{\theta x_k}^T \\ P_{\theta x_k} & P_{\theta\theta_k} \end{bmatrix} . \tag{2.18}$$

Using (2.18), (2.17b) can be rewritten as follows:

$$\begin{bmatrix} K_{x_k} \\ K_{\theta_k} \end{bmatrix} = \begin{bmatrix} E_* P_{xx_k} E_0^T + G_*(y_k, u_k) P_{\theta x_k} E_0^T \\ P_{\theta x_k} E_0^T \end{bmatrix} (E_0 P_{xx_k} E_0^T + R_k)^{-1} \,. \tag{2.19}$$

Due to the special structure of (2.12), it can be seen that only K_{x_k} depends on u_k and y_k, and the other gain, K_{θ_k}, is independent of u_k and y_k. To stress this, the notation $K_{x_k}(y_k, u_k)$ will be used instead of K_{x_k}. Thus K_{x_k} can be written as follows:

$$K_{x_k} = K_{x_k}(y_k, u_k) = E_* P_{xx_k} E_0^T (E_0 P_{xx_k} E_0^T + R_k)^{-1} + G_*(y_k, u_k) K_{\theta_k} \,. \tag{2.20}$$

Using this new notation and (2.19) the optimal predictor (2.17) can be written as follows:

$$\begin{bmatrix} \hat{x}_{k+1|k} \\ \hat{\theta}_{k+1|k} \end{bmatrix} = \begin{bmatrix} E_*\hat{x}_{k|k-1} + G_*(y_k, u_k)\hat{\theta}_{k|k-1} \\ \hat{\theta}_{k|k-1} \end{bmatrix} + \begin{bmatrix} K_{x_k}(y_k, u_k) \\ K_{\theta_k} \end{bmatrix} (y_k - E_0\hat{x}_{k|k-1}) . \tag{2.21}$$

Now let $\hat{s}_{k|k}$ be the estimate of s_k based on all measurements up to and including y_k. The optimal filter for this estimate has the following form:

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + P_{k|k-1}H^T(HP_{k|k-1}H^T + R_k)^{-1}(y_k - H\hat{s}_{k|k-1}) \tag{2.22a}$$

$$\hat{s}_{k+1|k} = F_k\hat{s}_{k|k} \tag{2.22b}$$

$$P_{k|k} = P_{k|k-1} - P_{k|k-1}H^T(HP_{k|k-1}H^T + R_k)^{-1}HP_{k|k-1} \tag{2.22c}$$

$$P_{k+1|k} = F_kP_{k|k}F_k^T \tag{2.22d}$$

The vector form of (2.22a) is

$$\begin{bmatrix} \hat{x}_{k|k} \\ \hat{\theta}_{k|k} \end{bmatrix} = \begin{bmatrix} \hat{x}_{k|k-1} \\ \hat{\theta}_{k|k-1} \end{bmatrix} + \begin{bmatrix} P_{xx_k}E_0^T \\ P_{\theta x_k}E_0^T \end{bmatrix} (E_0P_{xx_k}E_0^T + R_k)^{-1}(y_k - E_0\hat{x}_{k|k-1}) \tag{2.23}$$

The equation yielding $\hat{s}_{k+1|k}$ from $\hat{s}_{k|k}$, known as the time-update equation, is as follows:

$$\hat{s}_{k+1|k} = F_k\hat{s}_{k|k} , \tag{2.24}$$

or

$$\hat{x}_{k+1|k} = E_*\hat{x}_{k|k} + G_*(y_k, u_k)\hat{\theta}_{k|k} \tag{2.25a}$$

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k} . \tag{2.25b}$$

The equations (2.25b) are satisfied since the system is time-invariant. In order to simplify some of the equations which will be obtained in the next sections the following equation from (2.21) will frequently be used:

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k} = \hat{\theta}_{k|k-1} + K_{\theta_k}(y_k - E_0\hat{x}_{k|k-1}) . \tag{2.26}$$

2.4 Minimum mean square error prediction

Prediction is the forecasting side of information processing. The aim is to obtain at time k information about y_{k+d} for some $d \geq 1$, i.e., to obtain information about what y_k will be like subsequent to the time at which the measurement is available. In obtaining the information, measurements up to and including time k can be used. The criterion of interest is the minimum mean square error prediction based on the minimization of the cost function:

$$J_{k+d|k} = E\{(y_{k+d} - \hat{y}_{k+d|k})^2 | \mathcal{F}_k\} \tag{2.27}$$

where $\hat{y}_{k+d|k} = E\{y_{k+d}| \mathcal{F}_k\}$ is the optimal d-step-ahead prediction of y_k at time k in a minimum variance sense. Hence the objective is to produce $E\{y_{k+d}| \mathcal{F}_k\}$.

In the following sections it will be shown how this optimal problem can be solved without using any approximation methods.

2.5 Predictor forms

As seen in Section 2.2, the input-output behavior of a linear dynamical system can be described by an ARX model. The following lemma shows how an ARX model can be expressed in an alternative predictor form.

Lemma 2.1 *The output of the system (2.1), having an ARX model, at time $k+d$ can be expressed in the following predictor form:*

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{2.28}$$

where

$$x_{k+1} = E_*^{k+1} x_0 - \sum_{j=0}^{k} E_*^{k-j} A_* y_j + \sum_{j=0}^{k} E_*^{k-j} B_* u_j \tag{2.29}$$

$$E_0(E_* - A_*E_0)^{d-1} = [\underbrace{0,\ldots,0}_{n-d}, f_0, \ldots, f_{d-1}] \tag{2.30}$$

for $k \geq 0$ and $d \geq 1$.

Further, the coefficients $f_0, \ldots, f_{d-1}$ can be computed as follows:

$$f_i = \begin{cases} 1 & : \quad i = 0 \\ -\sum_{j=0}^{i-1} f_j a_{i-j} & : \quad i = 1, \ldots, d-1 \end{cases} \tag{2.31}$$

for $i = 0, \ldots, d-1$.

Proof: From (2.6) and using a sequential method, it will be shown how the predicted output at $k+d$, y_{k+d}, is obtained.

A sequence of inputs u_k, $u_0, u_1, \ldots, u_k$, is given, as well as a sequence of outputs y_k, $y_0, y_1, \ldots, y_k$, and the initial conditions x_0. Then

$$\begin{aligned} x_1 &= E_*x_0 - A_*y_0 + B_*u_0 \\ x_2 &= E_*x_1 - A_*y_1 + B_*u_1 = E_*^2x_0 - (E_*A_*y_0 + A_*y_1) + (E_*B_*u_0 + B_*u_1) \\ x_3 &= E_*x_2 - A_*y_2 + B_*u_2 \\ &= E_*^3x_0 - (E_*^2A_*y_0 + E_*A_*y_1 + A_*y_2) + (E_*^2B_*u_0 + E_*B_*u_1 + B_*u_2) \ . \end{aligned} \tag{2.32}$$

At step $k+1$, this leads to

$$x_{k+1} = E_*^{k+1}x_0 - \sum_{j=0}^{k} E_*^{k-j}A_*y_j + \sum_{j=0}^{k} E_*^{k-j}B_*u_j \tag{2.33}$$

for $k \geq 0$.

Following (2.8), d-step-ahead prediction may be developed as follows. First,

$$x_{k+2} = (E_* - A_*E_0)x_{k+1} + B_*u_{k+1} - A_*e_{k+1} \, .$$

Similarly,

$$\begin{aligned} x_{k+3} &= (E_* - A_*E_0)x_{k+2} + B_*u_{k+2} - A_*e_{k+2} \\ &= (E_* - A_*E_0)^2 x_{k+1} + (E_* - A_*E_0)B_*u_{k+1} + B_*u_{k+2} \\ &\quad - (E_* - A_*E_0)A_*e_{k+1} - A_*e_{k+2}\,. \end{aligned}$$

Hence

$$\begin{aligned} x_{k+d} &= (E_* - A_*E_0)^{d-1} x_{k+1} + \sum_{j=1}^{d-1}(E_* - A_*E_0)^{d-j-1}B_*u_{k+j} \\ &\quad - \sum_{j=1}^{d-1}(E_* - A_*E_0)^{d-j-1}A_*e_{k+j}\,. \end{aligned} \tag{2.34}$$

for $k \geq 0$ and $d \geq 1$, and also from (2.8b)

$$y_{k+d} = E_0 x_{k+d} + e_{k+d}\,. \tag{2.35}$$

Substituting (2.34) into (2.35) gives

$$\begin{aligned} y_{k+d} &= E_0(E_* - A_*E_0)^{d-1} x_{k+1} + \sum_{j=1}^{d-1} E_0(E_* - A_*E_0)^{d-j-1}B_*u_{k+j} \\ &\quad - \sum_{j=1}^{d-1} E_0(E_* - A_*E_0)^{d-j-1}A_*e_{k+j} + e_{k+d} \end{aligned} \tag{2.36}$$

for $k \geq 0$ and $d \geq 1$.

In order to simplify (2.36), the row vector $E_0(E_* - A_*E_0)^{d-j-1}$ is rewritten in the following form:

$$E_0(E_* - A_*E_0)^{d-j-1} = [\underbrace{0,\ldots,0}_{n-(d-j)},\underbrace{f_0,\ldots,f_{d-j-1}}_{d-j}] \tag{2.37}$$

where the sequence $\{f_0,\ldots,f_{d-1}\}$ is a scalar sequence and a function of the polynomial $A(q^{-1})$ appearing in (2.1) and can be computed from (2.31).

The following steps will show how (2.37) can be obtained.

Referring to (2.7), $E_* - A_* E_0$ is obtained as follows:

$$E_* - A_* E_0 = \begin{bmatrix} 0 & . & . & . & 0 \\ 1 & & & & . \\ & . & & & . \\ & & . & & . \\ & & & 1 & 0 \end{bmatrix} - \begin{bmatrix} a_n \\ . \\ . \\ . \\ a_1 \end{bmatrix} \begin{bmatrix} 0 & . & . & . & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & . & . & 0 & -a_n \\ 1 & & & & . \\ & . & & & . \\ & & . & & . \\ & & & 1 & -a_1 \end{bmatrix} . \tag{2.38}$$

Then

$$E_0(E_* - A_* E_0)^{d-j-1} = \begin{bmatrix} 0 & . & . & . & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & . & . & 0 & -a_n \\ 1 & & & & . \\ & . & & & . \\ & & . & & . \\ & & & 1 & -a_1 \end{bmatrix}^{d-j-1} . \tag{2.39}$$

Note that $E_0(E_* - A_* E_0)^{d-j-1}$ is the last row of $(E_* - A_* E_0)^{d-j-1}$ which establishes (2.37). Meanwhile, using (2.37) the following interesting result is obtained:

$$E_0(E_* - A_* E_0)^{d-j-1} B_* = 0 \; ; \quad j = 1, \dots, d-1 \tag{2.40}$$

because

$$E_0(E_* - A_* E_0)^{d-j-1} B_* = [\underbrace{0, \dots, 0}_{n-(d-j)}, \underbrace{f_0, \dots, f_{d-j-1}}_{d-j}] \begin{bmatrix} \begin{bmatrix} b_{n-d} \\ \vdots \\ b_0 \end{bmatrix}_{(n-d+1)\times 1} \\ \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}_{(d-1)\times 1} \end{bmatrix} = 0 \tag{2.41}$$

for $j = 1, \ldots, d-1$.

Now using (2.37) and (2.40), equation (2.36) is simplified as follows:

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} + \sum_{j=0}^{d-1} f_j e_{k+d-j} \ . \tag{2.42}$$

This gives (2.28).

Now it is shown how the proposed predictor form (2.42) can be related to the conventional predictor form discussed by Goodwin and Sin [28].

Substituting (2.33) into (2.42) gives

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1}(E_*^{k+1} x_0 - \sum_{j=0}^{k} E_*^{k-j} A_* y_j + \sum_{j=0}^{k} E_*^{k-j} B_* u_j) + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{2.43}$$

for $k \geq 0$.

Now let us introduce the following notations:

$$L_k = E_0(E_* - A_* E_0)^{d-1} E_*^{k+1} \tag{2.44a}$$

$$\alpha_{k-j} = -E_0(E_* - A_* E_0)^{d-1} E_*^{k-j} A_* \tag{2.44b}$$

$$\beta_{k-j} = E_0(E_* - A_* E_0)^{d-1} E_*^{k-j} B_* \tag{2.44c}$$

Then, (2.43) can be rewritten in terms of the new notations as follows:

$$y_{k+d} = L_k x_0 + \sum_{j=0}^{k} \alpha_{k-j} y_j + \sum_{j=0}^{k} \beta_{k-j} u_j + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{2.44d}$$

for $k \geq 0$.

Because of the special structure of matrix E_*

$$E_*^l = 0_{n \times n} \qquad \text{for } l \geq n \ . \tag{2.45}$$

Then for $k \geq n-1$ the effect of initial condition x_0 disappears and then (2.44d) can be simplified as follows:

$$y_{k+d} = \sum_{j=k-(n-1)}^{k} \alpha_{k-j} y_j + \sum_{j=k-(n-1)}^{k} \beta_{k-j} u_j + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{2.46}$$

Now the predictor (2.46) can be expressed in a conventional polynomial form, discussed by Goodwin and Sin [28], as follows:

$$y_{k+d} = \alpha(q^{-1}) y_k + \beta(q^{-1}) u_k + F(q) e_k \tag{2.47}$$

where

$$\alpha(q^{-1}) = \alpha_0 + \alpha_1 q^{-1} + \cdots + \alpha_{n-1} q^{-(n-1)} \tag{2.48}$$

$$\beta(q^{-1}) = \beta_0 + \beta_1 q^{-1} + \cdots + \beta_{n-1} q^{-(n-1)} \tag{2.49}$$

$$F(q) = f_0 q^d + f_1 q^{d-1} + \cdots + f_{d-1} q \tag{2.50}$$

where the coefficients in $\alpha(q^{-1})$ and $\beta(q^{-1})$, using (2.44b) and (2.44c), can be computed as follows:

$$\begin{aligned}\alpha_i &= -E_0(E_* - A_* E_0)^{d-1} E_*^i A_* \\ &= -[0, \ldots, 0, f_0, \ldots, f_{d-1}] E_*^i A_* = -a_{d+i} - \sum_{j=0}^{d-2} f_{j+1} a_{d+i-j-1} \end{aligned} \tag{2.51a}$$

$$\begin{aligned}\beta_i &= E_0(E_* - A_* E_0)^{d-1} E_*^i B_* \\ &= [0, \ldots, 0, f_0, \ldots, f_{d-1}] E_*^i B_* = b_i + \sum_{j=0}^{d-2} f_{j+1} b_{i-j-1} \end{aligned} \tag{2.51b}$$

where $i = 0, \ldots, n-1$ and the coefficients $f_0, \ldots, f_{d-1}$ are obtained from (2.31).

In the recursive equations above, (4.51a) and (4.51b) , whenever $a_{n+1}, a_{n+2}, \ldots$ and $b_{n-d+1}, b_{n-d+2}, \ldots$ appear in the formulas, they are set to zero.

As seen the predictor model above (2.47) is a special case of the predictor (2.44d) proposed in this research. In a zero initial condition case there is no difference between the two models. However, the proposed predictor is capable of dealing with a nonzero initial condition whereas the conventional predictor model lacks such a capability. In the following lemma, the technique used by Goodwin and Sin, [28], to express an ARX model in predictor form will be discussed.

Lemma 2.2 *For the system (2.1), the d-step-ahead predictor has the following form:*

$$y_{k+d} = \alpha(q^{-1})y_k + \beta(q^{-1})u_k + F(q)e_k \tag{2.52}$$

where

$$\alpha(q^{-1}) = \alpha_0 + \alpha_1 q^{-1} + \cdots + \alpha_{n-1} q^{-(n-1)} \tag{2.53a}$$

$$\beta(q^{-1}) = \beta_0 + \beta_1 q^{-1} + \cdots + \beta_{n-1} q^{-(n-1)}; \qquad \beta_0 \neq 0 \tag{2.53b}$$

$$F(q) = f_0 q^d + f_1 q^{d-1} + \cdots + f_{d-1} q; \qquad f_0 = 1 \tag{2.53c}$$

Proof: Multiplying (2.1) by $F(q)$ gives

$$F(q)A(q^{-1})y_k = F(q)B(q^{-1})u_k + F(q)e_k \tag{2.54}$$

Substituting the predicted output y_{k+d} into (2.54) gives

$$y_{k+d} - y_{k+d} + F(q)A(q^{-1})y_k = F(q)B(q^{-1})u_k + F(q)e_k \tag{2.55}$$

or

$$y_{k+d} = [q^d - F(q)A(q^{-1})]y_k + F(q)B(q^{-1})u_k + F(q)e_k \tag{2.56}$$

where

$$q^d - F(q)A(q^{-1}) = \alpha(q^{-1}) \tag{2.57a}$$

$$F(q)B(q^{-1}) = \beta(q^{-1}) . \tag{2.57b}$$

This establishes (2.52) as required. In the following it is shown how the coefficients in $F(q)$, $\alpha(q^{-1})$, and $\beta(q^{-1})$ can be computed uniquely.

Substituting (2.2a), (2.53a), and (2.53c) into (2.57a) gives

$$q^d-(f_0q^d+f_1q^{d-1}+\cdots+f_{d-1}q)(1+a_1q^{-1}+\cdots+a_nq^{-n}) = \alpha_0+\alpha_1q^{-1}+\cdots+\alpha_{n-1}q^{-(n-1)} \tag{2.58}$$

By equating the terms of equal degree, the following recursive equations are obtained for the coefficients of the polynomials $F(q)$ and $\alpha(q^{-1})$:

$$f_m = \begin{cases} 1 & : \; m=0 \\ -\sum_{j=0}^{m-1} f_j a_{m-j} & : \; m=1,\ldots,d-1 \end{cases} \tag{2.59}$$

and

$$\alpha_i = -a_{d+i} - \sum_{j=0}^{d-2} f_{j+1} a_{d+i-j-1} \, ; \qquad i=0,\ldots,n-1 \, . \tag{2.60}$$

Further, using (2.57b) the coefficients in $\beta(q^{-1})$ are determined as follows:

$$\beta_i = b_i + \sum_{j=0}^{d-2} f_{j+1} b_{i-j-1} \, ; \qquad i=0,\ldots,n-1 \, . \tag{2.61}$$

Note that, in the SISO case

$$\beta_0 = b_0 \, . \tag{2.62}$$

In the next section it will be shown how the parameter and state estimation technique of Section 2.3 can be used in the predictors above to design an adaptive predictor.

2.6 Adaptive prediction

There are two possible approaches to evaluate the conditional expected value of y_{k+d} given $\mathcal{F}_k$, $E\{y_{k+d}|\mathcal{F}_k\}$. These will be designated:

1. *Conventional approach* or *Approximation approach.*

2. *Proposed approach* or *Exact approach.*

The two approaches will be described and compared.

2.6.1 Conventional approach

It will be shown how the conditional expectation $E\{y_{k+d}|\mathcal{F}_k\}$ can be determined using an approximation method. Applying the following approximation method to the predictor model (2.52), it is possible to derive an optimal steady-state d-step-ahead predictor. The final result is the same as obtained by Åström [7], Wittenmark [52], and Goodwin and Sin [28].

Let us rewrite the conventional predictor (2.52) in the following form:

$$y_{k+d} = \alpha(q^{-1}, \theta)y_k + \beta(q^{-1}, \theta)u_k + F(q)e_k \tag{2.63}$$

where $\alpha(q^{-1}, \theta)$, $\beta(q^{-1}, \theta)$ denote polynomials having the same structure as $\alpha(q^{-1})$, $\beta(q^{-1})$ in (2.52) and having entries parametrized by the parameter vector θ, where θ contains the coefficients of $A(q^{-1})$ and $B(q^{-1})$ in (2.1).

The quantity $E\{y_{k+d}|\,\mathcal{F}_k\}$ is a nonlinear function of θ. The following notation is used to make this nonlinearity explicit:

$$E\{y_{k+d}|\,\mathcal{F}_k\} = f(\theta)\,, \tag{2.64}$$

where $f(\theta)$ is a nonlinear function of θ which is a random variable with the following conditional mean and covariance given $\mathcal{F}_k$:

$$E\{\theta|\,\mathcal{F}_k\} = E\{\theta_k|\,\mathcal{F}_k\} \;=\; \hat{\theta}_{k|k} \tag{2.65}$$

$$E\{(\theta - \hat{\theta}_{k|k})(\theta - \hat{\theta}_{k|k})^T|\,\mathcal{F}_k\} \;=\; \Sigma_{k|k} \tag{2.66}$$

where $\theta = \theta_k$ for a time-invariant system.

One of the approaches to determine the conditional mean of the random variable $f(\theta)$ is an approximation based on the truncation of the Taylor series expansion of $f(\theta)$ about the point $\hat{\theta}_{k|k} = E\{\theta|\mathcal{F}_k\}$ as follows:

The Taylor series expansion of $f(\theta)$ about $\hat{\theta}_{k|k} = E\{\theta_k|\,\mathcal{F}_k\}$ is

$$f(\theta) = f(\hat{\theta}_{k|k}) + \left.\frac{df}{d\theta^T}\right|_{\hat{\theta}_{k|k}} (\theta - \hat{\theta}_{k|k}) + \frac{1}{2!}(\theta - \hat{\theta}_{k|k})^T \left.\frac{d^2 f}{d\theta^2}\right|_{\hat{\theta}_{k|k}} (\theta - \hat{\theta}_{k|k}) + \cdots \,. \tag{2.67}$$

Taking conditional expectations given $\mathcal{F}_k$ yields

$$\begin{aligned} E\{f(\theta)|\,\mathcal{F}_k\} &= f(\hat{\theta}_{k|k}) + E\{\left.\frac{df}{d\theta^T}\right|_{\hat{\theta}_{k|k}} (\theta - \hat{\theta}_{k|k})|\,\mathcal{F}_k\} \\ &\quad + \frac{1}{2!}E\{(\theta - \hat{\theta}_{k|k})^T \left.\frac{d^2 f}{d\theta^2}\right|_{\hat{\theta}_{k|k}} (\theta - \hat{\theta}_{k|k})|\,\mathcal{F}_k\} + \cdots \,. \end{aligned} \tag{2.68}$$

Equation (2.68) can be simplified, using

$$E\{(\theta - \hat{\theta}_{k|k})|\,\mathcal{F}_k\} = 0 \tag{2.69}$$

$$E\{(\theta - \hat{\theta}_{k|k})(\theta - \hat{\theta}_{k|k})^T|\,\mathcal{F}_k\} = \Sigma_{k|k}\,, \tag{2.70}$$

as follows:

$$E\{f(\theta)|\,\mathcal{F}_k\} = f(\hat{\theta}_{k|k}) + \frac{1}{2!}tr(\left.\frac{d^2 f}{d\theta^2}\right|_{\hat{\theta}_{k|k}} \Sigma_{k|k}) + \cdots \,. \tag{2.71}$$

where $tr(.)$ is the trace operator.

If the conditional density function for θ given $\mathcal{F}_k$ is concentrated near $\hat{\theta}_{k|k}$, and $f(\theta)$ is smooth in the vicinity of this point, then, retaining only the first term of (2.71) gives

$$E\{f(\theta)|\,\mathcal{F}_k\} \approx f(\hat{\theta}_{k|k}) \tag{2.72}$$

or

$$E\{f(\theta)|\,\mathcal{F}_k\} \approx f(E\{\theta|\,\mathcal{F}_k\})\,. \tag{2.73}$$

Now taking conditional expectation on both sides of (2.63), gives

$$E\{y_{k+d}|\mathcal{F}_k\} = E\{\alpha(q^{-1},\theta)|\mathcal{F}_k\}y_k + E\{\beta(q^{-1},\theta)|\mathcal{F}_k\}u_k\,. \qquad (2.74)$$

since $E\{F(q)y_k|\mathcal{F}_k\} = 0$. Using the approximation above, (2.73), equation (2.74) can be simplified as follows:

$$E\{y_{k+d}|\mathcal{F}_k\} \approx \alpha(q^{-1},\hat{\theta}_{k|k})\,y_k + \beta(q^{-1},\hat{\theta}_{k|k})\,u_k \qquad (2.75)$$

and for simplicity if the right-hand side of the equation is called $\hat{y}_{k+d|k}$ we have

$$\hat{y}_{k+d|k} = \alpha(q^{-1},\hat{\theta}_{k|k})\,y_k + \beta(q^{-1},\hat{\theta}_{k|k})\,u_k \qquad (2.76)$$

where

$$\alpha(q^{-1},\hat{\theta}_{k|k}) = \hat{\alpha}_{0,k|k} + \hat{\alpha}_{1,k|k}\,q^{-1} + \cdots + \hat{\alpha}_{n-1,k|k}\,q^{-(n-1)} \qquad (2.77)$$

$$\beta(q^{-1},\hat{\theta}_{k|k}) = \hat{\beta}_{0,k|k} + \hat{\beta}_{1,k|k}\,q^{-1} + \cdots + \hat{\beta}_{n-1,k|k}\,q^{-(n-1)} \qquad (2.78)$$

and the coefficients in $\alpha(q^{-1},\hat{\theta}_{k|k})$ and $\beta(q^{-1},\hat{\theta}_{k|k})$ are computed as follows:

$$\hat{\alpha}_{i,k|k} = -\hat{a}_{d+i,k|k} - \sum_{j=0}^{d-2} \hat{f}_{j+1,k|k}\,\hat{a}_{d+i-j-1,k|k} \qquad (2.79a)$$

$$\hat{\beta}_{i,k|k} = \hat{b}_{i,k|k} + \sum_{j=0}^{d-2} \hat{f}_{j+1,k|k}\,\hat{b}_{i-j-1,k|k} \qquad (2.79b)$$

where $i = 0,\ldots,n-1$ and

$$\hat{f}_{m,k|k} = \begin{cases} 1 & ;\quad m = 0 \\ -\sum_{j=0}^{m-1} \hat{f}_{j,k|k}\,\hat{a}_{m-j,k|k} & ;\quad m = 1,\ldots,d-1\,. \end{cases} \qquad (2.80)$$

As before, in (2.79), whenever $\hat{a}_{n+1,k|k}, \hat{a}_{n+2,k|k}, \ldots$ and $\hat{b}_{n-d+1,k|k}, \hat{b}_{n-d+2,k|k}, \ldots$ appear in the formulas, they are set to zero.

Provided that the system input, output, and noise are absolutely bounded almost surely, the above indirect d-step-ahead predictor was proved by Sin, Goodwin, and Bitmead [47] to be globally convergent in the following sense:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} E\{(y_{k+d} - \hat{y}_{k+d|k})^2 | \mathcal{F}_k\} = \sum_{j=1}^{d-1} f_j^2 R_{k+d-j} \triangleq \gamma^2 \qquad \text{a.s.} \tag{2.81}$$

where γ^2 is the optimal d-step-ahead prediction error variance.

The adaptive predictor above uses the certainty equivalence principle. That is, it uses on-line parameter estimates in lieu of actual plant parameters. When the parameter estimates are close to the actual plant parameters (assuming the plant is in the model set), then this approach makes sense. However for adaptive schemes based on the certainty equivalence principle, in general there will be circumstances where the transient performance is unnecessarily poor, perhaps intolerably so.

2.6.2 Proposed approach

In this section it will be shown how a d-step-ahead predictor which is optimal at each step can be designed for the ARX model (2.1). As explained in the previous section the main objective is to evaluate $E\{y_{k+d}| \mathcal{F}_k\}$.

From Lemma 2.1, the ARX model (2.1) can be written in the following proposed predictor form, (2.42):

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} + \sum_{j=0}^{d-1} f_j e_{k+d-j} \qquad \text{for } k \geq 0 \,. \tag{2.82}$$

Now taking conditional expectation given $\mathcal{F}_k$ on both sides of (2.82), gives

$$E\{y_{k+d}| \mathcal{F}_k\} = E\{E_0(E_* - A_* E_0)^{d-1} x_{k+1} | \mathcal{F}_k\} + E\{\sum_{j=0}^{d-1} f_j e_{k+d-j} | \mathcal{F}_k\} \,. \tag{2.83}$$

Since $E\{\sum_{j=0}^{d-1} f_j e_{k+d-j} | \mathcal{F}_k\} = 0$, (2.83) can be simplified as follows:

$$E\{y_{k+d}| \mathcal{F}_k\} = E\{E_0(E_* - A_* E_0)^{d-1} x_{k+1} | \mathcal{F}_k\}. \tag{2.84}$$

Initially, assume that the system delay is unity [$d = 1$ in (2.1)]. Later it will be shown that this can be extended to the case $d \geq 1$.

Case $d = 1$: For $d = 1$, equation (2.84) finds the simplest possible form as follows:

$$E\{y_{k+1}| \mathcal{F}_k\} = E\{E_0 x_{k+1} | \mathcal{F}_k\} \tag{2.85}$$

or

$$\hat{y}_{k+1|k} = E_0 \hat{x}_{k+1|k} \tag{2.86}$$

Substituting (2.25a) into (2.86) gives

$$\hat{y}_{k+1|k} = E_0[E_* \hat{x}_{k|k} + G_*(y_k, u_k)\hat{\theta}_{k|k}] \quad \text{for} \quad k \geq 0 \tag{2.87}$$

where the estimates $\hat{x}_{k|k}$ and $\hat{\theta}_{k|k}$ satisfy the difference equations (2.22).

Using (2.46) and (4.51), for $k \geq n - 1$ the one-step-ahead predictor will have the following form:

$$y_{k+1} = - \sum_{j=k-(n-1)}^{k} a_{k-j} y_j + \sum_{j=k-(n-1)}^{k} b_{k-j} u_j + e_{k+1} \tag{2.88}$$

which can be written in a regression form

$$y_{k+1} = \phi_k^T \theta_k + e_{k+1} \tag{2.89}$$

where

$$\theta_k = [a_1, \ldots, a_{n-1}, b_0, \ldots, b_{n-d}]^T \tag{2.90}$$

$$\phi_k^T = [-y_k, \ldots, -y_{k-n+1}, u_k, \ldots, u_{k-n+1}]. \tag{2.91}$$

Taking conditional expectation yields

$$\hat{y}_{k+1|k} = \phi_k^T \hat{\theta}_{k|k} \quad \text{for} \quad k \geq n-1 \tag{2.92}$$

which coincides with the conventional one-step-ahead predictor discussed in the literature. The main difference between the proposed one-step-ahead predictor (2.87) and the conventional one (2.92) is that the proposed predictor is capable of dealing with cases with unknown initial conditions. However, the initial condition has to be known or zero to use the one-step-ahead predictor (2.92).

Case $d \geq 1$: As the number of time delays increases the adaptive optimal predictor becomes more complex. Substituting (2.30) into (2.84) gives

$$E\{y_{k+d}|\,\mathcal{F}_k\} = E\{[0,\ldots,0,f_0,\ldots,f_{d-1}]x_{k+1}|\,\mathcal{F}_k\} \tag{2.93}$$

or

$$\hat{y}_{k+d|k} = E\{f_0 x_{k+1}^{(d-1)} + f_1 x_{k+1}^{(d-2)} + \cdots + f_{d-1} x_{k+1}^{(0)}|\,\mathcal{F}_k\} \tag{2.94}$$

where $f_0, f_1, \ldots, f_{d-1}$ are iteratively calculated from (2.31), such that

$$\begin{aligned} f_0 &= 1 \\ f_1 &= -a_1 \\ f_2 &= -a_2 + a_1^2 \\ f_3 &= -a_3 + 2a_1a_2 - a_1^3 \\ &\vdots \\ f_{d-1} &= -\sum_{j=0}^{d-2} f_j a_{d-1-j}\,. \end{aligned} \tag{2.95}$$

Substituting (2.95) into (2.94) gives

$$\hat{y}_{k+d|k} = E\{x_{k+1}^{(d-1)} - a_1 x_{k+1}^{(d-2)} + \cdots - \sum_{j=0}^{d-2} f_j a_{d-1-j} x_{k+1}^{(0)}|\,\mathcal{F}_k\}\,. \tag{2.96}$$

Then according to the model (2.4) the equation above can be rewritten as follows:

$$\hat{y}_{k+d|k} = E\{(x_k^{(d)} - a_d y_k + b_0 u_k) - a_1(x_k^{(d-1)} - a_{d-1} y_k) + \cdots - \sum_{j=0}^{d-2} f_j a_{d-1-j}(x_k^{(1)} - a_1 y_k) | \mathcal{F}_k\} \tag{2.97}$$

provided that the system order satisfies $n > d$. If $n = d$, the first term in the equation above (2.97), $x_k^{(d)}$, is omitted.

Due to the linearity property of the expectation operator

$$\begin{aligned} \hat{y}_{k+d|k} &= E\{x_k^{(d)} | \mathcal{F}_k\} + E\{-a_d y_k | \mathcal{F}_k\} + E\{b_0 u_k | \mathcal{F}_k\} + E\{-a_1 x_k^{(d-1)} | \mathcal{F}_k\} \\ &+ E\{a_1 a_{d-1} y_k | \mathcal{F}_k\} + \cdots + E\{-\sum_{j=0}^{d-2} f_j a_{d-1-j} x_k^{(1)} | \mathcal{F}_k\} \\ &+ E\{\sum_{j=0}^{d-2} f_j a_{d-1-j} a_1 y_k | \mathcal{F}_k\} . \end{aligned} \tag{2.98}$$

Using the following lemma, it will be explained how the terms on the right-hand side of this equation can be evaluated from the elements of the mean vector $\hat{s}_{k|k}$ and the covariance matrix $P_{k|k}$.

Lemma 2.3 *If $Z = [z_1, z_2, \ldots, z_n]^T$ is a gaussian random vector with mean vector $M = [m_1, m_2, \ldots, m_n]^T$ and positive-definite covariance matrix Σ, then the higher moments depend only on the elements of the mean vector and the covariance matrix, [36], as follows:*

$$E\{z_i\} = m_i \tag{2.99a}$$

$$E\{z_i z_j\} = \Sigma_{ij} + m_i m_j \tag{2.99b}$$

$$E\{z_i z_j z_k\} = \Sigma_{ij} m_k + \Sigma_{ik} m_j + \Sigma_{jk} m_i + m_i m_j m_k \tag{2.99c}$$

$$\begin{aligned} E\{z_i z_j z_k z_l\} &= \Sigma_{ij}\Sigma_{kl} + \Sigma_{ik}\Sigma_{jl} + \Sigma_{il}\Sigma_{jk} + \Sigma_{ij} m_k m_l + \Sigma_{ik} m_j m_l + \Sigma_{il} m_j m_k \\ &+ \Sigma_{jk} m_i m_l + \Sigma_{jl} m_i m_k + \Sigma_{kl} m_i m_j + m_i m_j m_k m_l . \end{aligned} \tag{2.99d}$$

Proof: Using the moment generating function the lemma above is proved. If Z is normally distributed in n dimensions with mean vector M and positive definite covariance matrix Σ, the frequency function (probability density function) of Z is

$$f(Z) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(Z-M)^T\Sigma^{-1}(Z-M)} . \tag{2.100}$$

The moments of the distribution are defined by

$$E\{z_1^{r_1} z_2^{r_2} \cdots z_k^{r_k}\} = \int_{-\infty}^{\infty} z_1^{r_1} z_2^{r_2} \cdots z_k^{r_k} f(Z) dZ \tag{2.101}$$

where the r_j, $1 \leq j \leq k \leq n$, are nonnegative integers. Let $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_n]^T$ be a column vector. Then

$$\phi_Z(\lambda) = E\{e^{\lambda^T Z}\} \tag{2.102}$$

is called the moment generating function for Z since

$$\left. \frac{\partial^{r_1+r_2+\cdots+r_k}}{\partial\lambda_1^{r_1}\partial\lambda_2^{r_2}\cdots\partial\lambda_k^{r_k}} \phi_Z(\lambda) \right|_{\lambda=0} = E\{z_1^{r_1} z_2^{r_2} \cdots z_k^{r_k}\} . \tag{2.103}$$

An important property of moment generating functions is that the moment generating function determines the distribution. That is, there exists a one-to-one correspondence between the moment generating function and the distribution function of a random variable. The moment generating function for the case in which Z is normally distributed is calculated as follows. By definition of the moment generating function

$$\begin{aligned} \phi_Z(\lambda) &= E\{e^{\lambda^T Z}\} = \int_{-\infty}^{\infty} e^{\lambda^T Z} f(Z) dZ \\ &= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} \int_{-\infty}^{\infty} e^{\lambda^T Z - \frac{1}{2}(Z-M)^T\Sigma^{-1}(Z-M)} dZ . \end{aligned} \tag{2.104}$$

Hence

$$\phi_Z(\lambda) = e^{\lambda^T M + \frac{1}{2}\lambda^T\Sigma\lambda} . \tag{2.105}$$

For example

$$
\begin{aligned}
\frac{\partial}{\partial\lambda_i}\phi_Z(\lambda) &= \eta_i\phi_Z(\lambda) \\
\frac{\partial^2}{\partial\lambda_i\partial\lambda_j}\phi_Z(\lambda) &= (\Sigma_{ij}+\eta_i\eta_j)\phi_Z(\lambda) \\
\frac{\partial^3}{\partial\lambda_i\partial\lambda_j\partial\lambda_k}\phi_Z(\lambda) &= (\Sigma_{ij}\eta_k+\Sigma_{ik}\eta_j+\Sigma_{jk}\eta_i+\eta_i\eta_j\eta_k)\phi_Z(\lambda) \\
\frac{\partial^4}{\partial\lambda_i\partial\lambda_j\partial\lambda_k\partial\lambda_l}\phi_Z(\lambda) &= (\Sigma_{ij}\Sigma_{kl}+\Sigma_{ik}\Sigma_{jl}+\Sigma_{il}\Sigma_{jk}+\Sigma_{ij}\eta_k\eta_l+\Sigma_{ik}\eta_j\eta_l+\Sigma_{il}\eta_j\eta_k \\
&\quad +\Sigma_{jk}\eta_j\eta_k+\Sigma_{jl}\eta_i\eta_k+\Sigma_{kl}\eta_i\eta_j+\eta_i\eta_j\eta_k\eta_l)\phi_Z(\lambda) \quad (2.106)
\end{aligned}
$$

where

$$\eta_\alpha = \sum_{\beta=1}^{n}\Sigma_{\beta\alpha}\lambda_\beta + m_\alpha \quad ; \ \alpha = i, j, k, l\,. \tag{2.107}$$

If the above partial derivatives are evaluated at $\lambda = 0$, the corresponding moments 2.99 are obtained.

Note that as would be expected, the higher moments of the Gaussian distribution depend only on the elements of the mean vector and the covariance matrix.

Now using Lemma 2.3, the terms on the right-hand side of equation (2.98) can be evaluated in terms of the elements of the mean vector $\hat{s}_{k|k}$ and the covariance matrix $P_{k|k}$ as follows:

For simplicity, for systems with $d = 2$ and $d = 3$ it will be shown that the optimal d-step-ahead predictor can be obtained using Lemma 2.3 ; the extension to general d is straightforward but lengthy.

Case $d = 2$: For a system with $d = 2$ the optimal d-step-ahead predictor $E\{y_{k+d}|\mathcal{F}_k\}$ can be evaluated as follows. From (2.98)

$$\hat{y}_{k+2|k} = E\{x_k^{(2)}|\mathcal{F}_k\} - E\{a_2 y_k|\mathcal{F}_k\} + E\{b_0 u_k|\mathcal{F}_k\} - E\{a_1 x_k^{(1)}|\mathcal{F}_k\} + E\{a_1^2 y_k|\mathcal{F}_k\}\,. \tag{2.108}$$

Since y_k and u_k are $\mathcal{F}_k$ measurable. (2.108) using the conditional expectation property can be simplified as follows:

$$\hat{y}_{k+2|k} = E\{x_k^{(2)}|\,\mathcal{F}_k\} - E\{a_2|\,\mathcal{F}_k\}y_k + E\{b_0|\,\mathcal{F}_k\}u_k - E\{a_1 x_k^{(1)}|\,\mathcal{F}_k\} + E\{a_1^2|\,\mathcal{F}_k\}y_k\,. \tag{2.109}$$

Hence, from (2.99b),

$$\hat{y}_{k+2|k} = \hat{x}_{k|k}^{(2)} - \hat{a}_{2,k|k}y_k + \hat{b}_{0,k|k}u_k - (\hat{a}_{1,k|k}\hat{x}_{k|k}^{(1)} + P_{a_1 x_k^{(1)}|k}) + (\hat{a}_{1,k|k}^2 + P_{a_1^2|k})y_k\,. \tag{2.110}$$

Once again it is emphasised that the above result (2.110) is for a system with an order higher than two. If the order is $n = 2$. the first term on the right-hand side of (2.110), $\hat{x}_{k|k}^{(2)}$. will be omitted.

Using the conventional method explained in Section 2.6.1, the two-step-ahead predictor of a second order system with $d = 2$ and white noise would have the following form:

$$E\{y_{k+d}|\,\mathcal{F}_k\} \approx \alpha(q^{-1}.\hat{\theta}_{k|k})\,y_k + \beta(q^{-1}.\hat{\theta}_{k|k})\,u_k\,. \tag{2.111}$$

where

$$\begin{aligned}\alpha(q^{-1}) &= \alpha_0 + \alpha_1 q^{-1} = (-a_2 + a_1^2) + (a_1 a_2)q^{-1}\\ \beta(q^{-1}) &= \beta_0 + \beta_1 q^{-1} = b_0 - a_1 b_0 q^{-1}\end{aligned}$$

and

$$\begin{aligned}\alpha(q^{-1},\hat{\theta}_{k|k}) &= (-\hat{a}_{2,k|k} + \hat{a}_{1,k|k}^2) + \hat{a}_{1,k|k}\hat{a}_{2,k|k}q^{-1}\\ \beta(q^{-1},\hat{\theta}_{k|k}) &= \hat{b}_{0,k|k} - \hat{a}_{1,k|k}\hat{b}_{0,k|k}q^{-1}\end{aligned}$$

and finally

$$E\{y_{k+d}|\,\mathcal{F}_k\} \approx (-\hat{a}_{2,k|k} + \hat{a}_{1,k|k}^2)y_k + \hat{a}_{1,k|k}\hat{a}_{2,k|k}y_{k-1} + \hat{b}_{0,k|k}u_k - \hat{a}_{1,k|k}\hat{b}_{0,k|k}u_{k-1}\,. \tag{2.112}$$

As expected there are some similarities between the two predictors. (2.110) and (2.112), but the main difference is that the real optimal predictor (2.110) includes the uncertainty associated with the parameter and state estimates whereas the optimal steady-state predictor (2.112) ignores the uncertainty in the parameter estimates.

Case $d = 3$: Using (2.98), the optimal d-step-ahead predictor for a system with $d = 3$ has the following form:

$$\hat{y}_{k+3|k} = E\{(x_k^{(3)} - a_3 y_k + b_0 u_k) - a_1(x_k^{(2)} - a_2 y_k) + (-a_2 + a_1^2)(x_k^{(1)} - a_1 y_k)|\, \mathcal{F}_k\}\,. \tag{2.113}$$

If the same procedure as employed for case $d = 2$ is applied for a system with $d = 3$ the optimal predictor can be evaluated as follows:

$$\begin{aligned}\hat{y}_{k+3|k} \; = \; & E\{x_k^{(3)}|\, \mathcal{F}_k\} - E\{a_3 y_k|\, \mathcal{F}_k\} + E\{b_0 u_k|\, \mathcal{F}_k\} - E\{a_1 x_k^{(2)}|\, \mathcal{F}_k\} + E\{a_1 a_2 y_k|\, \mathcal{F}_k\} \\ & - E\{a_2 x_k^{(1)}|\, \mathcal{F}_k\} + E\{a_2 a_1 y_k|\, \mathcal{F}_k\} + E\{a_1^2 x_k^{(1)}|\, \mathcal{F}_k\} - E\{a_1^3 y_k|\, \mathcal{F}_k\}\,. \end{aligned} \tag{2.114}$$

Using the conditional expectation property

$$\begin{aligned}\hat{y}_{k+3|k} \; = \; & E\{x_k^{(3)}|\, \mathcal{F}_k\} - E\{a_3|\, \mathcal{F}_k\}y_k + E\{b_0|\, \mathcal{F}_k\}u_k - E\{a_1 x_k^{(2)}|\, \mathcal{F}_k\} + E\{a_1 a_2|\, \mathcal{F}_k\}y_k \\ & - E\{a_2 x_k^{(1)}|\, \mathcal{F}_k\} + E\{a_2 a_1|\, \mathcal{F}_k\}y_k + E\{a_1^2 x_k^{(1)}|\, \mathcal{F}_k\} - E\{a_1^3|\, \mathcal{F}_k\}y_k\,. \end{aligned} \tag{2.115}$$

Now from the results (2.99) the conditional expectation terms on the right-hand side of the equation above (2.115) are simplified. The last term of (2.115), $E\{a_1^3|\, \mathcal{F}_k\}$, is selected as an example to show how the results (2.99) can be used to simplify (2.115). According to the result (2.99c)

$$E\{z_i z_j z_k\} = \Sigma_{ij} m_k + \Sigma_{ik} m_j + \Sigma_{jk} m_i + m_i m_j m_k\,. \tag{2.116}$$

Now if it is assumed that

$$z_i = z_j = z_k = a_1 \tag{2.117}$$

and, accordingly, it can be deduced that

$$\begin{aligned} \Sigma_{ij} = \Sigma_{ik} = \Sigma_{jk} &= P_{a_1^2|k} \\ m_i = m_j = m_k &= \hat{a}_{1,k|k} \,. \end{aligned}$$

Substituting into (2.99c) leads to

$$E\{a_1^3 | \mathcal{F}_k\} = 3P_{a_1^2|k}\hat{a}_{1,k|k} + \hat{a}_{1,k|k}^3 \,. \qquad (2.118)$$

Using the same procedure as previously, the optimal d-step-ahead predictor (2.115) can be rewritten in terms of the elements of the mean vector and the covariance matrix as follows:

$$\begin{aligned} \hat{y}_{k+3|k} &= \hat{x}_{k|k}^{(3)} - \hat{a}_{3,k|k}y_k + \hat{b}_{0,k|k}u_k - (\hat{a}_{1,k|k}\hat{x}_{k|k}^{(2)} + P_{a_1x_k^{(2)}|k}) + 2(\hat{a}_{1,k|k}\hat{a}_{2,k|k} + P_{a_1a_2|k})y_k \\ &\quad -(\hat{a}_{2,k|k}\hat{x}_1^{(1)} + P_{a_2x_k^{(1)}|k}) + (P_{a_1^2|k}\hat{x}_{k|k}^{(1)} + 2P_{a_1x_k^{(1)}|k}\hat{a}_{1,k|k} + \hat{a}_{1,k|k}^2\hat{x}_{k|k}^{(1)}) \\ &\quad -(3P_{a_1^2|k}\hat{a}_{1,k|k} + \hat{a}_{1,k|k}^3)y_k \,. \end{aligned} \qquad (2.119)$$

Using the conventional method, the 3-step-ahead predictor of a third order system with $d = 3$ and white noise would have the following form:

$$E\{y_{k+d} | \mathcal{F}_k\} \approx \alpha(q^{-1}, \hat{\theta}_{k|k})\, y_k + \beta(q^{-1}, \hat{\theta}_{k|k})\, u_k \,, \qquad (2.120)$$

where

$$\begin{aligned} \alpha(q^{-1}) &= \alpha_0 + \alpha_1 q^{-1} + \alpha_2 q^{-1} \\ &= (-a_3 + 2a_1a_2 - a_1^3) + (a_1a_3 + a_2^2 - a_1^2a_2)q^{-1} + (a_2a_3 - a_1^2a_3)q^{-1} \\ \beta(q^{-1}) &= \beta_0 + \beta_1 q^{-1} + \beta_2 q^{-2} \\ &= b_0 + (-a_1b_0)q^{-1} + (-a_2b_0 + a_1^2b_0)q^{-2} \end{aligned}$$

and

$$\begin{aligned} \alpha(q^{-1}, \hat{\theta}_{k|k}) &= (-\hat{a}_{3,k|k} + 2\hat{a}_{1,k|k}\hat{a}_{2,k|k} - \hat{a}_{1,k|k}^3) + (\hat{a}_{1,k|k}\hat{a}_{3,k|k} + \hat{a}_{1,k|k}^2 - \hat{a}_{1,k|k}^2\hat{a}_{2,k|k})q^{-1} \\ &\quad + (\hat{a}_{2,k|k}\hat{a}_{3,k|k} - \hat{a}_{1,k|k}^2\hat{a}_{3,k|k})q^{-2} \\ \beta(q^{-1}, \hat{\theta}_{k|k}) &= \hat{b}_{0,k|k} + (-\hat{a}_{1,k|k}\hat{b}_{0,k|k})q^{-1} + (-\hat{a}_{2,k|k}\hat{b}_{0,k|k} + \hat{a}_{1,k|k}^2\hat{b}_{0,k|k})q^{-2} \end{aligned}$$

and finally

$$\begin{aligned}
E\{y_{k+d}|\mathcal{F}_k\} \approx\ & (-\hat{a}_{3,k|k} + 2\hat{a}_{1,k|k}\hat{a}_{2,k|k} - \hat{a}^3_{1,k|k})y_k + (\hat{a}_{1,k|k}\hat{a}_{3,k|k} + \hat{a}^2_{1,k|k} - \hat{a}^2_{1,k|k}\hat{a}_{2,k|k})y_{k-1} \\
& + (\hat{a}_{2,k|k}\hat{a}_{3,k|k} - \hat{a}^2_{1,k|k}\hat{a}_{3,k|k})y_{k-2} + \hat{b}_{0,k|k}u_k + (-\hat{a}_{1,k|k}\hat{b}_{0,k|k})u_{k-1} \\
& + (-\hat{a}_{2,k|k}\hat{b}_{0,k|k} + \hat{a}^2_{1,k|k}\hat{b}_{0,k|k})u_{k-2}\,. \qquad (2.121)
\end{aligned}$$

2.7 Summary

This chapter considered an indirect optimal approach to prediction . An exact optimal solution was obtained whereas conventional approaches use an approximation method to solve the problem. In the proposed approach, first the plant parameters and states are estimated with a Kalman filter. Second the adaptive optimal predictor is designed based on their uncertainty. Such a predictor is optimal at each step in a minimum variance sense. This predictor can handle cases with unknown initial conditions. On the other hand, the conventional predictor uses the estimates as if they were the true parameters for the purpose of design and ignores any uncertainty in the parameter estimates. This conventional predictor is based on the certainty equivalence principle and is asymptotically optimal. As a result, both results have the same steady-state performance. However, the proposed predictor can have a better transient performance.

In the next chapter it will be explained why an indirect scheme is the only way to solve the optimal problem considered in this chapter.

Chapter 3

SISO Cautious control

3.1 Introduction

This chapter is concerned with indirect SISO stochastic adaptive control. The purpose of the control is to minimize the variance of the output around a desired value one step ahead for SISO linear time-invariant discrete-time systems having general delay and white noise perturbation. Minimization over only one step leads to the one-step ahead or cautious controller. This controller takes the parameter uncertainties into account, in contrast to the certainty equivalence controller.

The SISO cautious control problem has been discussed in Åström and Wittenmark [8], Wieslander and Wittenmark [51], and Nahorski and Vidal [40]. The method proposed in this research and the method used by Åström and Wittenmark [8] are the same except that the proposed method can handle cases with uncertainty in initial conditions. Wieslander and Wittenmark [51] used a very complicated way to solve the adaptive problem for a unity delay system. Nahorski and Vidal [40] handle the case of general delay by using a special model.

This chapter is organized as follows. In Section 3.2 the control criterion of interest is addressed. Section 3.3 contains the main results. In Sections 3.3.1 and 3.3.2 using the theory of stochastic optimal control, the general one-step ahead optimization problem is solved in both nonadaptive and adaptive cases, respectively. It is proved that the optimal solution can be expressed analytically in a closed form. In Section 3.4 it is shown how the general optimal solution obtained in this research can be applied to some special cases discussed in the literature. Because of the initial condition x_0, there are two different cases for the estimation process. Hence the proposed indirect cautious control with unknown and known initial condition are presented in Sections 3.5 and 3.6, respectively. In Section 3.7 in order to explain why the indirect approach is the only way to solve the minimization problem above, the direct cautious control proposed by Nahorski and Vidal [40] is considered. In Section 3.8 simulation results are presented to show the performance of the proposed cautious control compared with that of the certainty equivalence controller.

3.2 The control criterion

The purpose of the control is to keep the output of the system (2.1) as close as possible to a known reference value trajectory $\{y_k^*\}$. In other words, the objective is to choose the input u_k so as to minimize the sample mean-square error between the output and the desired value y_{k+d}. Thus consider the following cost function:

$$J_{k+d} = E\{(y_{k+d} - y_{k+d}^*)^2\} \, . \tag{3.1}$$

The cost function should be minimized with respect to u_k. Then a feature of the cost function above is that it is only concerned with the situation one step ahead. A control strategy is admissible if u_k is a function of all outputs observed up to and including time k, i.e. , $y_k, y_{k-1}, \dots, y_0$, all applied control signals $u_{k-1}, \dots, u_0$ and the a priori data. Using the smoothing property of conditional means, we have

$$J_{k+d} = E\{E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\}\} . \tag{3.2}$$

Hence the optimal cost is

$$J^*_{k+d} = \min_{u_k} E\{E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\}\} . \tag{3.3}$$

It follows from a fundamental lemma of stochastic control theory, see Åström [7], that

$$J^*_{k+d} = E\{\min_{u_k} E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\}\} \tag{3.4}$$

where u_k is $\mathcal{F}_k$ measurable. In other words, the control objective will be to minimize the sample mean of the sequence $E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\}$, and the new criterion may be written as

$$J_{k+d|k} = E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\} . \tag{3.5}$$

Once again using the property of conditional means the index becomes

$$J_{k+d} = E\{E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k, x_0, \theta\}\} \tag{3.6}$$

where x_0 is the initial condition and θ is the parameter vector. A new index when x_0 and θ are known will be obtained as follows:

$$J_{k+d|k,x_0,\theta} = E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k, x_0, \theta\} . \tag{3.7}$$

3.3 One-step-ahead optimization

In this section the one-step-ahead optimization problems above are solved using the theory of stochastic optimal control following the philosophy of Åström [5, 6], and Wieslander and Wittenmark [51]. In order to minimize the two cost functions above, (3.5) and (3.7), the results obtained in Chapter 2 are employed in this section.

As explained in Chapter 2, Lemma 2.1, the ARX model (2.1) can be written in one of the following predictor forms:

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{3.8}$$

or

$$y_{k+d} = E_0(E_* - A_* E_0)^{d-1} (E_*^{k+1} x_0 - \sum_{j=0}^{k} E_*^{k-j} A_* y_j + \sum_{j=0}^{k} E_*^{k-j} B_* u_j) + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{3.9}$$

for $k \geq 0$ and $d \geq 1$.

In order to show which term is a function of u_k, (3.9) is rewritten as follows:

$$y_{k+d} = b_0 u_k + E_0(E_* - A_* E_0)^{d-1} (E_*^{k+1} x_0 - \sum_{j=0}^{k} E_*^{k-j} A_* y_j + \sum_{j=0}^{k-1} E_*^{k-j} B_* u_j) + \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{3.10}$$

or y_{k+d} can be expressed in a compact form as

$$y_{k+d} = b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) + \eta(\theta, \mathcal{E}_{k+1}^{k+d}) \tag{3.11}$$

where $\gamma(.)$ and $\eta(.)$ are nonlinear functions in terms of the parameter vector θ and

$$\mathcal{Y}_0^k = \{y_k, \ldots, y_0\} \tag{3.12}$$

$$\mathcal{U}_0^{k-1} = \{u_{k-1}, \ldots, u_0\} \tag{3.13}$$

$$\mathcal{E}_{k+1}^{k+d} = \{e_{k+d}, \ldots, e_{k+1}\}\,. \tag{3.14}$$

The independent process e_k denotes a white noise which has the following properties

$$E\{e_{k+l} | \mathcal{F}_k\} = 0 \quad \text{a.s.} \tag{3.15a}$$

$$E\{e_{k+l} e_{k+m} | \mathcal{F}_k\} = R_{k+l}\delta(l-m) \quad \text{a.s.} \tag{3.15b}$$

for $l \geq 1$, $m \geq 1$, and $k \geq 0$.

Using (3.15a), it is clear that conditional mean of the noise seen by the predictor (3.9) or (3.11), $\sum_{j=0}^{d-1} f_j e_{k+d-j}$, is zero as follows:

$$E\{\sum_{j=0}^{d-1} f_j e_{k+d-j} | \mathcal{F}_k\} = 0 \quad \text{or} \quad E\{\eta(\theta, \mathcal{E}_{k+1}^{k+d}) | \mathcal{F}_k\} = 0 \tag{3.16}$$

for $k \geq 0$ and $d \geq 1$.

At this point, it can be shown how an optimal d-step-head predictor in a minimum variance sense given $\mathcal{F}_k$, the initial condition x_0, and the parameter vector θ can be obtained. Taking conditional mean of both sides of (3.11) with respect to $\mathcal{F}_k$, x_0, and θ gives

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) + E\{\eta(\theta, \mathcal{E}_{k+1}^{k+d}) | \mathcal{F}_k, x_0, \theta\} \,. \tag{3.17}$$

Using (3.16),

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) \tag{3.18}$$

or, using (3.8)

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} \tag{3.19}$$

or

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = E_0(E_* - A_* E_0)^{d-1}(E_*^{k+1} x_0 - \sum_{j=0}^{k} E_*^{k-j} A_* y_j + \sum_{j=0}^{k} E_*^{k-j} B_* u_j) \tag{3.20}$$

from (3.9), for $k \geq 0$ and $d \geq 1$. Note that this case is a special case or a nonadaptive case of the predictor discussed in Chapter 2.

In the next two sections it will be shown how the criteria $J_{k+d|k,x_0,\theta}$ and $J_{k+d|k}$ can be minimized without using approximation methods.

3.3.1 Minimization of $J_{k+d|k,x_0,\theta}$

If the parameters and the initial condition of the system are known it is easy to determine the optimal feedback or to minimize the cost function $J_{k+d|k,x_0,\theta}$, (3.7). In the following theorem a stochastic controller based on an optimal predictor is derived. The result shows that putting the predicted output equal to the desired output achieves optimal control.

Theorem 3.1 *For the system (2.1) the value of u_k which minimizes the criterion*

$$J_{k+d|k,x_0,\theta} = E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k, x_0, \theta\} \tag{3.21}$$

can be determined as follows:

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = y^*_{k+d} \tag{3.22}$$

for $k \geq 0$ and $d \geq 1$.
(Note that it is necessary to assume that $b_0 \neq 0$ and that the system is minimum-phase at every instant of time. The control signal may otherwise be unbounded.)

Proof: Substituting (3.11) into (3.21) gives

$$J_{k+d|k,x_0,\theta} = E\{[b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) + \eta(\theta, \mathcal{E}_{k+1}^{k+d}) - y^*_{k+d}]^2 | \mathcal{F}_k, x_0, \theta\} \tag{3.23}$$

or

$$\begin{aligned} J_{k+d|k,x_0,\theta} &= E\{[b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) - y^*_{k+d}]^2 + \eta(\theta, \mathcal{E}_{k+1}^{k+d})^2 \\ &\quad + 2[b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) - y^*_{k+d}]\eta(\theta, \mathcal{E}_{k+1}^{k+d}) | \mathcal{F}_k, x_0, \theta\} . \end{aligned} \tag{3.24}$$

Since e_k is an independent noise which satisfies (3.16), the third term on the right side of (3.24) is zero. From

$$\eta(\theta, \mathcal{E}_{k+1}^{k+d}) = \sum_{j=0}^{d-1} f_j e_{k+d-j} \tag{3.25}$$

and (3.15b), the second term of the right side of (3.24) is determined as follows:

$$E\{\eta(\theta, \mathcal{E}_{k+1}^{k+d})^2 | \mathcal{F}_k, x_0, \theta\} = \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} \tag{3.26}$$

which is independent of u_k. Therefore (3.24) is simplified as follows:

$$J_{k+d|k,x_0,\theta} = [b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) - y_{k+d}^*]^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} \,. \tag{3.27}$$

The necessary condition for a minimum is

$$\frac{\partial J_{k+d|k,x_0,\theta}}{\partial u_k} = 0 \,. \tag{3.28}$$

Since only the first term on the right side of (3.27) is a function of u_k, differentiating (3.27) with respect to u_k gives the following equation:

$$\begin{aligned} \frac{\partial J_{k+d|k,x_0,\theta}}{\partial u_k} &= 2b_0[b_0 u_k + \gamma(\theta. x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) - y_{k+d}^*] = 0 \\ &= 2b_0[E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} - y_{k+d}^*] = 0 \end{aligned} \tag{3.29}$$

using (3.18).

According to the initial assumptions $b_0 \neq 0$, so $J_{k+d|k,x_0,\theta}$ can be minimized when:

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} - y_{k+d}^* = 0 \tag{3.30}$$

or

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} = y_{k+d}^* \,. \tag{3.31}$$

This establishes (3.22).

This result had already been obtained in the literature for the case when the parameters and the initial condition are known . In words, the minimum variance control can be obtained by equating the predicted output to the desired output given the parameters and the initial condition.

The minimum value of the cost function is then given by

$$J_{k+d|k,x_0,\theta}^* = \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} \,. \tag{3.32}$$

3.3.2 Minimization of $J_{k+d|k}$

Here the optimization of the design criterion $J_{k+d|k}$ given in (3.5) is discussed. Using the same approach as used in Section 3.3.1, it is proved that the optimal solution can be expressed analytically in a closed form. The result is summarized in the following theorem.

Theorem 3.2 *For the system (2.1) the value of u_k which minimizes the criterion*

$$J_{k+d|k} = E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k\} \tag{3.33}$$

can be determined as follows:

$$E\{b_0 y_{k+d} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+d} \tag{3.34}$$

for $k \geq 0$ and $d \geq 1$.

Proof: Using the property of conditional means

$$J_{k+d|k} = E\{E\{(y_{k+d} - y^*_{k+d})^2 | \mathcal{F}_k, x_0, \theta\} | \mathcal{F}_k\} \tag{3.35}$$

or

$$J_{k+d|k} = E\{J_{k+d|k,x_0,\theta} | \mathcal{F}_k\} \tag{3.36}$$

which shows the relationship between the two cost functions.

Hence, substituting (3.27) into (3.36) gives

$$J_{k+d|k} = E\{[b_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) - y^*_{k+d}]^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} | \mathcal{F}_k\} \tag{3.37}$$

or from (3.18)

$$J_{k+d|k} = E\{(E\{y_{k+d} | \mathcal{F}_k, x_0, \theta\} - y^*_{k+d})^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} | \mathcal{F}_k\} \tag{3.38}$$

for $k \geq 0$ and $d \geq 1$.

From (3.19)

$$E\{y_{k+d}|\,\mathcal{F}_k, x_0, \theta\} = E_0(E_* - A_* E_0)^{d-1} x_{k+1} \;. \tag{3.39}$$

Using the method used in Chapter 2

$$\begin{aligned} E\{y_{k+d}|\,\mathcal{F}_k, x_0, \theta\} &= [0, \cdots, 0, f_0, \cdots, f_{d-1}] x_{k+1} \\ &= f_0 x_{k+1}^{(d-1)} + \cdots + f_{d-1} x_{k+1}^{(0)} \end{aligned} \tag{3.40}$$

where $f_0, \ldots, f_{d-1}$ are obtained from (2.30) of Chapter 2.
Then from (3.40)

$$E\{y_{k+d}|\,\mathcal{F}_k, x_0, \theta\} = b_0 u_k + [f_0, \cdots, f_{d-1}] \left(\begin{bmatrix} x_k^{(d)} \\ \vdots \\ x_k^{(1)} \end{bmatrix} - \begin{bmatrix} a_d \\ \vdots \\ a_1 \end{bmatrix} y_k \right) \tag{3.41}$$

In order to facilitate the writing, (3.41) is written in the following compact form:

$$E\{y_{k+d}|\,\mathcal{F}_k, x_0, \theta\} = b_0 u_k + \rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) \tag{3.42}$$

where $\rho(.)$ is a nonlinear function, and also it is clear that

$$\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) = \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) \;. \tag{3.43}$$

Substituting (3.42) into (3.38) leads to

$$J_{k+d|k} = E\{[b_0 u_k + \rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y_{k+d}^*]^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j}|\,\mathcal{F}_k\} \tag{3.44}$$

or

$$\begin{aligned} J_{k+d|k} &= E\{b_0^2 u_k^2 + 2 b_0 u_k [\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y_{k+d}^*]|\,\mathcal{F}_k\} \\ &\quad + E\{[\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y_{k+d}^*]^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j}|\,\mathcal{F}_k\} \;. \end{aligned} \tag{3.45}$$

Using the conditional mean properties

$$
\begin{aligned}
J_{k+d|k} &= u_k^2 E\{b_0^2 | \mathcal{F}_k\} + 2u_k E\{b_0[\rho(x_k^{(d)}, \dots, x_k^{(1)}, y_k, \theta) - y^*_{k+d}] | \mathcal{F}_k\} \\
&\quad + E\{[\rho(x_k^{(d)}, \dots, x_k^{(1)}, y_k, \theta) - y^*_{k+d}]^2 + \sum_{j=0}^{d-1} f_j^2 R_{k+d-j} | \mathcal{F}_k\} \, . \quad (3.46)
\end{aligned}
$$

In Chapter 2, it was proved that the conditional distribution of $s_k = [x_k^T \; \theta_k^T]^T$ given $\mathcal{F}_k$ is normal with mean $\hat{s}_{k|k}$ and covariance $P_{k|k}$, where the conditional mean and covariance are generated by the Kalman filter using the following difference equations:

$$
\begin{aligned}
\hat{s}_{k|k} &= \hat{s}_{k|k-1} + P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1} (y_k - E_0 \hat{x}_{k|k-1}) && (3.47a) \\
\hat{s}_{k+1} &= F_k \hat{s}_{k|k} && (3.47b) \\
P_{k|k} &= P_{k|k-1} - P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1} H P_{k|k-1} && (3.47c) \\
P_{k+1|k} &= F_k P_{k|k} F_k^T && (3.47d)
\end{aligned}
$$

where

$$
\hat{s}_{k|k} = \begin{bmatrix} \hat{x}_{k|k} \\ \hat{\theta}_{k|k} \end{bmatrix} = \begin{bmatrix} E\{x_k | \mathcal{F}_k\} \\ E\{\theta_k | \mathcal{F}_k\} \end{bmatrix} . \quad (3.48)
$$

Note that the conditional means and covariances of x_k and θ given $\mathcal{F}_k$ are independent of u_k. Therefore, the conditional mean of any nonlinear function of x_k and θ given $\mathcal{F}_k$, which is a function of the conditional means and covariances of x_k and θ, are also independent of u_k.

Thus the last two terms on the right side of (3.46) are not functions of u_k,

$$
\frac{\partial J_{k+d|k}}{\partial u_k} = 2u_k E\{b_0^2 | \mathcal{F}_k\} + 2E\{b_0[\rho(x_k^{(d)}, \dots, x_k^{(1)}, y_k, \theta) - y^*_{k+d}] | \mathcal{F}_k\} = 0 \quad (3.49)
$$

or

$$
u_k E\{b_0^2 | \mathcal{F}_k\} + E\{b_0[\rho(x_k^{(d)}, \dots, x_k^{(1)}, y_k, \theta) - y^*_{k+d}] | \mathcal{F}_k\} = 0 \, . \quad (3.50)
$$

In order to reach a general conclusion, (3.50) is manipulated as follows:

$$E\{b_0^2 u_k + b_0[\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y_{k+d}^*]|\mathcal{F}_k\} = 0 \tag{3.51}$$

or

$$E\{b_0[b_0 u_k + \rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y_{k+d}^*]|\mathcal{F}_k\} = 0 \,. \tag{3.52}$$

From (3.42),

$$E\{b_0(E\{y_{k+d}|\mathcal{F}_k, x_0, \theta\} - y_{k+d}^*)|\mathcal{F}_k\} = 0 \tag{3.53}$$

or

$$E\{b_0 E\{y_{k+d}|\mathcal{F}_k, x_0, \theta\}|\mathcal{F}_k\} = E\{b_0|\mathcal{F}_k\}y_{k+d}^* \,. \tag{3.54}$$

Now, noting that $b_0 E\{y_{k+d}|\mathcal{F}_k, x_0, \theta\} = E\{b_0 y_{k+d}|\mathcal{F}_k, x_0, \theta\}$,

$$E\{E\{b_0 y_{k+d}|\mathcal{F}_k, x_0, \theta\}|\mathcal{F}_k\} = E\{b_0|\mathcal{F}_k\}y_{k+d}^* \,. \tag{3.55}$$

Using the smoothing property of conditional means again gives

$$E\{b_0 y_{k+d}|\mathcal{F}_k\} = E\{b_0|\mathcal{F}_k\}y_{k+d}^* \,. \tag{3.56}$$

This establishes (3.34).

3.4 Some important special cases

Using the general optimal solution (3.34), the following important special cases can be explained:

1. From (3.34) it is seen that if the parameters θ and the initial condition x_0 in the system are known and also $b_0 \neq 0$, (3.34) can be simplified as follows:

$$E\{y_{k+d}|\mathcal{F}_k, x_0, \theta\} = y_{k+d}^* \,. \tag{3.57}$$

In words, the input $\{u_k\}$ is obtained by equating the predicted output of the system with its desired value. Actually (3.34) can be considered as a generalization of (3.57); thus (3.34) will be equal to (3.57) when the parameters and states and their estimates are equal, and the covariances are equal to zero.

2. If $b_0 \neq 0$ and known then

$$E\{y_{k+d} | \mathcal{F}_k\} = y^*_{k+d} \tag{3.58}$$

where in Chapter 2 it was shown how $E\{y_{k+d} | \mathcal{F}_k\}$ can be determined even if b_0 is unknown.

3. Using the certainty equivalence approximation, as shown in Chapter 2, (3.56) can be approximated as follows:

$$E\{b_0(y_{k+d} - y^*_{k+d}) | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = 0 \tag{3.59}$$

or

$$E\{b_0 | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} E\{(y_{k+d} - y^*_{k+d}) | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = 0\,. \tag{3.60}$$

Now if $E\{b_0 | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} \neq 0$,

$$E\{y_{k+d} | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = y^*_{k+d}\,. \tag{3.61}$$

The comparison of (3.61) and (3.57) shows that an approximation of the optimal control law can be obtained by solving the control problem in the case of known parameters and substituting the known parameters with their estimates. The controller can be interpreted as a certainty-equivalence controller.

Because of the initial condition x_0, there could be the following two different cases for the estimation process:

- The initial condition x_0 is known or zero.
- There is uncertainty in x_0, or x_0 is unknown.

These two cases are analyzed in the following two sections.

3.5 Cautious control with unknown x_0

In this section using the general optimal solution obtained in Theorem 3.2, an indirect cautious control with unknown initial condition is developed. It is shown that if there is uncertainty in the initial condition x_0, or x_0 is unknown, not only the parameters of the system but also the states of the system have to be estimated. Therefore, the identification process is the same as employed in Chapter 2. Finally, using the same property of Gaussian distributions employed in Chapter 2, the indirect method proposed leads to control laws which not only use the estimates of the parameters and states but also the variances of the estimation errors.

As in Chapter 2, using state variables the ARX model (2.1) can be written in the following canonical form:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k \tag{3.62a}$$

$$y_k = E_0 x_k + e_k \,. \tag{3.62b}$$

If θ is defined as follows:

$$\theta = [a_1, \ldots, a_n, b_0, \ldots, b_{n-d}]^T \tag{3.63}$$

then the model (3.62) can be rearranged in the following form:

$$x_{k+1} = E_* x_k + G_* \theta \tag{3.64a}$$

$$y_k = E_0 x_k + e_k \,. \tag{3.64b}$$

where the matrix G_* is constructed according to the order of the parameters in θ. Since for a time invariant system $\theta_{k+1} = \theta_k = \theta$, by appending the parameter vector θ_k to the state vector x_k and using the model (3.64), the following dynamical model can be used for the estimation process:

$$s_{k+1} = F_k s_k \tag{3.65a}$$

$$y_k = H s_k + e_k \tag{3.65b}$$

where

$$s_k = \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} \quad F_k = \begin{bmatrix} E_* & G_* \\ 0 & I \end{bmatrix} \quad H = [\, E_0 \quad 0 \,] \,. \tag{3.66}$$

It is assumed that s_0 and $\{e_k\}$ are independent and gaussian: s_0 is $N(\bar{s}_0, \Sigma_0)$; $\{e_k\}$ is zero mean, with covariance $R_k \delta_{k-l}$. Then a Kalman filter may be used to find $\hat{s}_{k|k} = E\{s_k | \mathcal{F}_k\}$, the conditional mean of s_k given $\mathcal{F}_k$, as follows:

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k(y_k - H\hat{s}_{k|k-1}) \,; \qquad \hat{s}_{0|-1} = \bar{s}_0 \tag{3.67a}$$

$$K_k = P_{k|k-1} H^T (H P_{k|k-1} H^T + R_k)^{-1} \tag{3.67b}$$

$$\hat{s}_{k+1|k} = F_k \hat{s}_{k|k} \tag{3.67c}$$

$$P_{k|k} = P_{k|k-1} - K_k H P_{k|k-1} \,; \qquad P_{0|-1} = \Sigma_0 \geq 0 \tag{3.67d}$$

$$P_{k+1|k} = F_k P_{k|k} F_k^T \,. \tag{3.67e}$$

Finally, the input signal is generated by solving

$$E\{b_0 y_{k+d} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+d} \tag{3.68}$$

or, from (3.50), the feedback law (3.68) is explicitly given by

$$u_k = -\frac{E\{b_0[\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) - y^*_{k+d}] | \mathcal{F}_k\}}{E\{b_0^2 | \mathcal{F}_k\}} \,, \qquad \text{for } k \geq 0 \tag{3.69a}$$

where

$$\rho(x_k^{(d)}, \ldots, x_k^{(1)}, y_k, \theta) = [f_0, \cdots, f_{d-1}] \left(\begin{bmatrix} x_k^{(d)} \\ \vdots \\ x_k^{(1)} \end{bmatrix} - \begin{bmatrix} a_d \\ \vdots \\ a_1 \end{bmatrix} y_k \right) \tag{3.69b}$$

and $f_0, \ldots, f_{d-1}$ are obtained from (2.30) of Chapter 2.

Initially the system delay will be taken to be $d = 1$. Later it will be shown how this can be readily extended to the case $d \geq 1$.

Case $d = 1$: Using (3.69), the control law for the case $d = 1$ is:

$$\begin{aligned} u_k &= -\frac{E\{b_0(x_k^{(1)} - a_1 y_k - y_{k+1}^*) | \mathcal{F}_k\}}{E\{b_0^2 | \mathcal{F}_k\}} \\ &= -\frac{E\{b_0 x_k^{(1)} | \mathcal{F}_k\} - E\{b_0 a_1 | \mathcal{F}_k\} y_k - E\{b_0 | \mathcal{F}_k\} y_{k+1}^*}{E\{b_0^2 | \mathcal{F}_k\}} \end{aligned} \tag{3.70}$$

for $k \geq 0$. Using Lemma 2.3 in Chapter 2, (2.99), the conditional mean terms in (3.70) can be written in terms of the elements of the mean vector and the covariance matrix as follows:

$$u_k = -\frac{(\hat{b}_{0,k|k} \hat{x}_{k|k}^{(1)} + P_{b_0 x_k^{(1)}|k}) - (\hat{b}_{0,k|k} \hat{a}_{1,k|k} + P_{b_0 a_1|k}) y_k - \hat{b}_{0,k|k} y_{k+1}^*}{\hat{b}_{0,k|k}^2 + P_{b_0^2|k}} \tag{3.71}$$

for $k \geq 0$.

For $k \geq n-1$ the control law can be written in the conventional form discussed by Åström and Wittenmark [8, 10]. It can be seen from (2.4) of Chapter 2 that

$$x_k^{(1)} = -a_2 y_{k-1} - \cdots - a_n y_{k-n+1} + b_1 u_{k-1} + \cdots + b_{n-1} u_{k-n+1} \tag{3.72}$$

for $k \geq n-1$. Substituting (3.72) into (3.70) gives

$$u_k = -\frac{E\{b_0(-a_2 y_{k-1} - \cdots - a_n y_{k-n+1} + b_1 u_{k-1} + \cdots + b_{n-1} u_{k-n+1} - a_1 y_k - y_{k+1}^*) | \mathcal{F}_k\}}{E\{b_0^2 | \mathcal{F}_k\}} . \tag{3.73}$$

Once again using Lemma 2.3 in Chapter 2, (2.99),

$$u_k = -\frac{-\sum_{j=1}^{n}(\hat{b}_{0,k|k}\hat{a}_{j,k|k} + P_{b_0a_j|k})y_{k-j+1} + \sum_{j=1}^{n-1}(\hat{b}_{0,k|k}\hat{b}_{j,k|k} + P_{b_0b_j|k})u_{k-j} - \hat{b}_{0,k|k}y^*_{k+1}}{\hat{b}^2_{0,k|k} + P_{b_0^2|k}} \tag{3.74}$$

for $k \geq n-1$.

As seen, for the case $d = 1$, there is actually no difference between the direct approach and the indirect approach, since in this case the ARX model coincides with the one-step-ahead predictor. As the number of time delays increases the control law becomes more complex.

Case $d = 2$: Now consider a system with $d = 2$. Using the some procedure as used for the case $d = 1$, the control law will be determined. From (3.69), the control law for the case $d = 2$ is

$$\begin{aligned} u_k &= -\frac{E\{b_0 \begin{bmatrix} f_0 & f_1 \end{bmatrix} \left(\begin{bmatrix} x_k^{(2)} \\ x_k^{(1)} \end{bmatrix} - \begin{bmatrix} a_2 \\ a_1 \end{bmatrix} y_k \right) | \mathcal{F}_k\} - E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \\ &= -\frac{E\{b_0(f_0 x_k^{(2)} + f_1 x_k^{(1)} - f_0 a_2 y_k - f_1 a_1 y_k) | \mathcal{F}_k\} - E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \end{aligned} \tag{3.75}$$

where $f_0 = 1$ and $f_1 = -a_1$, from (2.30) of Chapter 2. Hence

$$\begin{aligned} u_k &= -\frac{E\{b_0(x_k^{(2)} - a_1 x_k^{(1)} - a_2 y_k + a_1^2 y_k) | \mathcal{F}_k\} - E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \\ &= -\frac{E\{b_0 x_k^{(2)} - b_0 a_1 x_k^{(1)} - b_0 a_2 y_k + b_0 a_1^2 y_k | \mathcal{F}_k\} - E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \end{aligned} \tag{3.76}$$

for $k \geq 0$.

Using the results obtained in Chapter 2, u_k, (3.76), can be determined in terms of the mean values of the parameters and states and the covariance matrix elements as follows:

$$u_k = -\left(\frac{1}{\hat{b}^2_{0,k|k} + P_{b_0^2|k}}\right)[(\hat{b}_{0,k|k}\hat{x}^{(2)}_{k|k} + P_{b_0 x_k^{(2)}|k})$$

$$
\begin{aligned}
&-(\hat{b}_{0,k|k}P_{a_1x_k^{(1)}|k} + \hat{a}_{1,k|k}P_{b_0x_k^{(1)}|k} + \hat{x}^{(1)}_{k|k}P_{b_0a_1|k} + \hat{b}_{0,k|k}\hat{a}_{1,k|k}\hat{x}^{(1)}_{k|k}) \\
&-(\hat{b}_{0,k|k}\hat{a}_{2,k|k} + P_{b_0a_1|k})y_k + (\hat{b}_{0,k|k}P_{a_1^2|k} + 2\hat{a}_{1,k|k}P_{b_0a_1|k} + \hat{b}_{0,k|k}\hat{a}^2_{1,k|k})y_k \\
&-\hat{b}_{0,k|k}y^*_{k+2}]
\end{aligned}
\tag{3.77}
$$

for $k \geq 0$.

The control law (3.77) clearly shows the influence of the uncertainties of the parameter and state estimates. The covariance matrix will help the controller to make a more cautious control action when the estimates are poor.

3.6 Cautious control with known initial condition

In this section an indirect cautious control with known initial condition is discussed, and it is shown that if the initial condition is known or zero, it is not required to estimate the states of the system. Therefore, using the same method employed in the previous section, Section 3.5, only the optimal parameter estimation, provided by a Kalman filter, is used to obtain the optimal control law. As a result, the algorithm proposed in this section can be considered as a special case of that discussed in Section 3.5.

The ARX model (2.1) can be written in the following form:

$$y_k = -a_1y_{k-1} - \cdots - a_ny_{k-n} + b_0u_{k-d} + \cdots + b_{n-d}u_{k-n+d} + e_k \ . \tag{3.78}$$

Then (3.78) can be expressed in the following regression form:

$$y_k = \phi^T_{k-1}\theta + e_k \tag{3.79}$$

where

$$
\begin{aligned}
\theta &= [a_1, \ldots, a_n, b_0, \ldots, b_{n-d}]^T \\
\phi^T_{k-1} &= [-y_{k-1}, \ldots, -y_{k-n}, u_{k-d}, \ldots, u_{k-n+d}]
\end{aligned}
$$

with initial values

$$[y_{-1}, \ldots, y_{-d-n+1}, u_{-1}, \ldots, u_{-d-n+1}] \; . \tag{3.80}$$

Since for a time invariant system $\theta_{k+1} = \theta_k = \theta$, using (3.79) the following dynamical system can be used to estimate the system parameters:

$$\theta_{k+1} = \theta_k \tag{3.81a}$$

$$y_k = \phi_{k-1}^T \theta_k + e_k \; . \tag{3.81b}$$

If it is assumed that θ_0 and $\{e_k\}$ are jointly gaussian and mutually independent: θ_0 is $N(\bar{\theta}_0, \Sigma_0)$; $\{e_k\}$ is zero mean, covariance $R_k \delta_{k-l}$. Then $\hat{\theta}_{k|k} = E\{\theta_k | \mathcal{F}_k\}$, the conditional mean of θ_k given $\mathcal{F}_k$, satisfies the following recursions (the Kalman filter):

$$\hat{\theta}_{k|k} = \hat{\theta}_{k|k-1} + K_k(y_k - \phi_{k-1}^T \hat{\theta}_{k|k-1}) \; ; \quad \hat{\theta}_{0|-1} = \bar{\theta}_0 \tag{3.82a}$$

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k} \tag{3.82b}$$

where K_k is the filter gain given by

$$K_k = P_{k|k-1} \phi_{k-1} (\phi_{k-1}^T P_{k|k-1} \phi_{k-1} + R_k)^{-1} \tag{3.82c}$$

$P_{k|k-1}$ is the error covariance matrix, that is,

$$P_{k|k-1} = E\{(\theta - \hat{\theta}_{k|k-1})(\theta - \hat{\theta}_{k|k-1})^T | \mathcal{F}_{k-1}\} \tag{3.82d}$$

$P_{k|k-1}$ satisfies the following equations:

$$P_{k|k} = P_{k|k-1} - K_k \phi_{k-1}^T P_{k|k-1} \; ; \quad P_{0|-1} = \Sigma_0 \tag{3.82e}$$

$$P_{k+1} = P_{k|k} \; . \tag{3.82f}$$

Finally the input signal u_k is generated from the following feedback control law:

$$E\{b_0 y_{k+d} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y_{k+d}^* \tag{3.83}$$

Initially it will be shown below how the algorithm above can be applied for the model (2.1) with $d = 1$. Later it will be shown how this can be used for the general case, a system with $d \geq 1$.

Case $d = 1$: For the case $d = 1$, the ARX model (??) can be readily expressed in a predictor form as follows:

$$y_{k+1} = -a_1 y_k - \cdots - a_n y_{k-n+1} + b_0 u_k + \cdots + b_{n-1} u_{k-n+1} + e_{k+1} \tag{3.84}$$

Substituting (3.84) into (3.83) gives

$$E\{b_0(-a_1 y_k - \cdots - a_n y_{k-n+1} + b_0 u_k + \cdots + b_{n-1} u_{k-n+1} + e_{k+1}) | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+1} \tag{3.85}$$

Using $E\{b_0 e_{k+1} | \mathcal{F}_k\} = 0$, and after some manipulations

$$u_k = \frac{-E\{b_0(-a_1 y_k - \cdots - a_n y_{k-n+1} + b_1 u_{k-1} + \cdots + b_{n-1} u_{k-n+1} | \mathcal{F}_k\} + E\{b_0 | \mathcal{F}_k\} y^*_{k+1}}{E\{b_0^2 | \mathcal{F}_k\}} \tag{3.86}$$

Using Lemma 2.3,

$$u_k = \frac{\sum_{j=1}^{n} (\hat{b}_{0,k|k} \hat{a}_{j,k|k} + P_{b_0 a_j|k}) y_{k-j+1} - \sum_{j=1}^{n-1} (\hat{b}_{0,k|k} \hat{b}_{j,k|k} + P_{b_0 b_j|k}) u_{k-j} + \hat{b}_{0,k|k} y^*_{k+1}}{\hat{b}^2_{0,k|k} + P_{b_0^2|k}} \tag{3.87}$$

where $\hat{b}_{0,k|k}$ and $P_{b_0^2|k}$ are the conditional mean and variance of b_0, respectively, that is,

$$\begin{aligned} \hat{b}_{0,k|k} &= E\{b_0 | \mathcal{F}_k\} \\ P_{b_0^2|k} &= E\{(b_0 - \hat{b}_{0,k|k})^2 | \mathcal{F}_k\} \end{aligned}$$

and $P_{b_0 a_j|k}$ is the conditional covariance of b_0 and a_j as well as $P_{b_0 a_j|k}$ is the conditional covariance of b_0 and b_j, i.e.,

$$\begin{aligned} P_{b_0 a_j|k} &= E\{(b_0 - \hat{b}_{0,k|k})(a_j - \hat{a}_{j,k|k}) | \mathcal{F}_k\} \\ P_{b_0 b_j|k} &= E\{(b_0 - \hat{b}_{0,k|k})(b_j - \hat{b}_{j,k|k}) | \mathcal{F}_k\} . \end{aligned}$$

For a unit delay system, the approach is identical to the one discussed in Åström and Wittenmark (1973), [9], differing only for delays greater than unity.

General case, $d \geq 1$: As explained above, the parameters of the system in its representation (3.78) are recursively identified in time. Then the control law u_k is obtained by setting

$$E\{b_0 y_{k+d} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+d} \tag{3.88}$$

from which the control law can be expressed in the following compact form:

$$u_k = -\frac{E\{\vartheta^T | \mathcal{F}_k\}\psi_k - E\{\beta_0 | \mathcal{F}_k\} y^*_{k+d}}{E\{\beta_0^2 | \mathcal{F}_k\}} \tag{3.89}$$

where

$$\begin{aligned} \psi_k &= [y_k, \ldots, y_{k-n+1}, u_{k-1}, \ldots, u_{k-n+1}]^T \\ \vartheta^T &= [\vartheta_1, \ldots, \vartheta_{2n-1}] \end{aligned}$$

and

$$\vartheta_j = \begin{cases} \beta_0 \alpha_{j-1}; & j = 1, \ldots, n \\ \beta_0 \beta_{j-n}; & j = n+1, \ldots, 2n-1 . \end{cases} \tag{3.90}$$

Using (2.59)-(2.62) of Chapter 2, the coefficients α_i and β_i for $i = 1, \ldots, n-1$ are obtained and then substituted into (3.90). After computing the conditional mean $E\{\vartheta^T | \mathcal{F}_k\}$ using the method described in Chapter 2 and then substituting into (3.89) the control law is obtained.

The method above is illustrated in the following example:

Example (Case $d = 2$): Consider the following second-order system with $d = 2$:

$$A(q^{-1}) y_k = B(q^{-1}) u_k + e_k \tag{3.91}$$

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} \tag{3.92}$$

$$B(q^{-1}) = q^{-2}(b_0) . \tag{3.93}$$

According to Lemma 2.2 of Chapter 2, the system above can be expressed in the following predictor form:

$$y_{k+2} = (\alpha_0 + \alpha_1 q^{-1}) y_k + (\beta_0 + \beta_1 q^{-1}) u_k + (f_0 q^2 + f_1 q) e_k \tag{3.94}$$

where the coefficients $\alpha_0, \alpha_1, \beta_0, \beta_1, f_0, f_1$ are:

$$\begin{array}{lll} \alpha_0 = -a_2 + a_1^2 & \beta_0 = b_0 & f_0 = 1 \\ \alpha_1 = -a_2 & \beta_1 = -a_1 b_0 & f_1 = -a_1 \end{array} \tag{3.95}$$

using (2.59)-(2.61). Substituting (3.95) into (3.94) gives

$$y_{k+2} = (-a_2 + a_1^2) y_k + (-a_2) y_{k-1} + b_0 u_k + (-a_1 b_0) u_{k-1} + e_{k+2} - a_1 e_{k+1} . \tag{3.96}$$

Returning to the cautious controller, recall that the control law is achieved by setting $E\{b_0 y_{k+d} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+d}$. Using (3.96), it follows immediately that

$$E\{b_0[(-a_2 + a_1^2) y_k + (-a_2) y_{k-1} + b_0 u_k + (-a_1 b_0) u_{k-1} | \mathcal{F}_k\} = E\{b_0 | \mathcal{F}_k\} y^*_{k+2} \tag{3.97}$$

using $E\{b_0(e_{k+2} - a_1 e_{k+1}) | \mathcal{F}_k\} = 0$. Hence the control law can be written as

$$u_k = \frac{-E\{b_0[(-a_2 + a_1^2) y_k + (-a_2) y_{k-1} + (-a_1 b_0) u_{k-1}] | \mathcal{F}_k\} + E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \tag{3.98}$$

or in a compact form as follows:

$$u_k = -\frac{E\{\vartheta^T | \mathcal{F}_k\} \psi_k - E\{b_0 | \mathcal{F}_k\} y^*_{k+2}}{E\{b_0^2 | \mathcal{F}_k\}} \tag{3.99}$$

where

$$\begin{aligned}\psi_k &= [y_k, y_{k-1}, u_{k-1}]^T \\ \vartheta^T &= [\vartheta_1, \vartheta_2, \vartheta_3]\end{aligned}$$

and

$$\vartheta_1 = b_0(-a_2 + a_1^2)\ , \quad \vartheta_2 = b_0(-a_2)\ , \quad \vartheta_3 = b_0(-a_1 b_1)\ . \tag{3.100}$$

The only term in (3.99) which requires further explanation is $E\{\vartheta^T | \mathcal{F}_k\}$. Using the property of random vectors, we have

$$\begin{aligned}E\{\vartheta^T | \mathcal{F}_k\} &= E\{[\vartheta_1, \vartheta_2, \vartheta_3| \mathcal{F}_k\} & (3.101)\\ &= [E\{\vartheta_1| \mathcal{F}_k\}, E\{\vartheta_2| \mathcal{F}_k\}, E\{\vartheta_3| \mathcal{F}_k\}] & (3.102)\end{aligned}$$

and then using the results obtained in Chapter 2, the conditional means above are computed as follows:

$$\begin{aligned}E\{\vartheta_1| \mathcal{F}_k\} &= E\{b_0(-a_2 + a_1^2)| \mathcal{F}_k\} \\ &= -E\{b_0 a_2| \mathcal{F}_k\} + E\{b_0 a_1^2| \mathcal{F}_k\} & (3.103a)\\ &= -(\hat{b}_{0|k}\hat{a}_{2|k} + P_{b_0 a_2|k}) + (P_{a_1^2|k}\hat{b}_{0|k} + 2P_{b_0 a_1|k}\hat{a}_{1|k} + \hat{a}_{1|k}^2\hat{b}_{0|k})\end{aligned}$$

$$\begin{aligned}E\{\vartheta_2| \mathcal{F}_k\} &= E\{b_0(-a_2)| \mathcal{F}_k\} \\ &= -(\hat{b}_{0|k}\hat{a}_{2|k} + P_{b_0 a_2|k}) & (3.103b)\end{aligned}$$

$$\begin{aligned}E\{\vartheta_3| \mathcal{F}_k\} &= E\{b_0(-a_1 b_0)| \mathcal{F}_k\} \\ &= -(P_{b_0^2|k}\hat{a}_{1|k} + 2P_{b_0 a_1|k}\hat{b}_{0|k} + \hat{b}_{0|k}^2\hat{a}_{1|k})\ . & (3.103c)\end{aligned}$$

Substituting (3.103) into (3.99) results in the following control law:

$$u_k = -(\frac{1}{\hat{b}_{0,k|k}^2 + P_{b_0^2|k}})[-[(\hat{b}_{0,k|k}\hat{a}_{2,k|k} + P_{b_0 a_2|k})$$

$$+(P_{a_1^2|k}\hat{b}_{0,k|k} + 2P_{b_0 a_1|k}\hat{a}_{1,k|k} + \hat{a}_{1|k}^2\hat{b}_{0,k|k})]y_k - (\hat{b}_{0,k|k}\hat{a}_{2,k|k} + P_{b_0 a_2|k})y_{k-1}$$
$$-(P_{b_0^2|k}\hat{a}_{1,k|k} + 2P_{b_0 a_1|k}\hat{b}_{0,k|k} + \hat{b}_{0,k|k}^2\hat{a}_{1|k})u_{k-1} - \hat{b}_{0,k|k}y_{k+2}^*]] \qquad (3.104)$$

for $k \geq 0$.

Note that in this research the minimization of $J_{k+d|k}$ is done for general systems, i.e., the general delay and white noise case. The minimization of $J_{k+d|k}$ has been discussed in:

- Åström and Wittenmark in 1971, [8], for a unit delay system.
- Wieslander and Wittenmark in 1971, [51], when $d = 2$.
- Nahorski and Vidal in 1974, [40], the case with general delay d.

As explained before, for a unit delay system there is no difference between the indirect and direct approaches. Therefore, the proposed method and the method used by Åström and Wittenmark are the same except that the proposed method can handle cases with uncertainty in the initial condition.

Wieslander and Wittenmark used the following criterion for a unit delay system:

$$J = \min_{u_k} E\{y_{k+1}^2 | \mathcal{F}_{k-1}\} \qquad (3.105)$$

which is different from what has been discussed in this research. Note that the system time delay is $d = 1$, whereas the number of steps looked ahead in the criterion (3.105), to determine the control signal u_k, is two. In this research the criterion used for a system with $d = 1$ is

$$J = \min_{u_k} E\{y_{k+1}^2 | \mathcal{F}_k\} \qquad (3.106)$$

which is more applicable than the criterion used by Wittenmark. The system delay and the number of steps looked ahead in the criterion (3.106) are the same.

According to (3.106), the control signal u_k is determined when the output signal y_k has been observed which makes sense for every system with $d \geq 1$. However, using (3.105) the control signal u_k is obtained without observing y_k which is already available. For this reason, the method proposed by Wieslander and Wittenmark solved the optimal problem for a unit delay system in a complicated way which does not seem to be necessary at all.

The method proposed by Nahorski and Vidal in 1974 to minimize $J_{k+d|k}$ for systems with general delay $d \geq 1$ is explained in the following section and it is shown how their method differs from the method proposed in this research. Also, it will be explained why an indirect scheme is the only way to solve both the adaptive optimal control discussed in this chapter and the adaptive optimal prediction discussed in Chapter 2.

3.7 Direct cautious control

As described previously, the proposed approach is an indirect algorithm because this approach estimates the parameters in an ARX model for the system and then converts this model into the required predictor format. The general advantages of this approach, discussed by Goodwin and Sin in [28], are: (1) it is likely to involve fewer parameters, and (2) it is natural to expect priori knowledge regarding physical quantities in a system to be more easily mapped into priori knowledge of parameters in an input-output model of the process than into priori knowledge regarding predictor parameters. Besides, it will be shown below that the indirect approach is the only way to solve the minimization problem above for a system with an ARX model with general delay d. An alternative approach is to estimate the parameters of the optimal predictor directly, which has the advantage that less calculation is required to determine control law. For a reason shown below, the direct approach cannot handle the case with general delay d.

Using Lemma 2.2, the ARX model (2.1) can be transformed to the following predictor form:

$$\begin{aligned} y_{k+d} &= \alpha(q^{-1})y_k + \beta(q^{-1})u_k + F(q)e_k \\ &= \sum_{j=0}^{n-1} \alpha_j y_{k-j} + \sum_{j=0}^{n-1} \beta_j u_{k-j} + \sum_{j=0}^{d-1} f_j e_{k+d-j} \,. \end{aligned} \tag{3.107}$$

Then (3.107) can be written in the following stochastic regression form

$$y_{k+d} = \phi_k^T \theta + \nu_{k+d} \tag{3.108}$$

or

$$y_k = \phi_{k-d}^T \theta + \nu_k \tag{3.109}$$

with

$$\phi_k^T = [y_k, \ldots, y_{k-n+1}, u_k, \ldots, u_{k-n+1}] \tag{3.110}$$

$$\theta^T = [\alpha_0, \ldots, \alpha_{n-1}, \beta_0, \ldots \beta_{n-1}] \tag{3.111}$$

where $\nu_k = \sum_{i=0}^{d-1} f_i e_{k-i}$. The future (unpredictable) noise ν_k is a nongaussian colored noise. The noise ν_k is colored because the value of this noise at one time is correlated with the value at another time, and is nongaussian due to the occurrence of the products between the unknown parameters $\{f_0, \ldots, f_{d-1}\}$ and the gaussian input noise $\{e_k\}$. Since the a priori knowledge of the statistical characteristics of the noise ν_k is not available, the Kalman filter cannot be used to estimate the parameter vector θ. A possible estimation method is using the following least squares procedure

$$\hat{\theta}_k = \hat{\theta}_{k-1} + P_{k-1}\phi_{k-d}(y_k - \phi_{k-d}^T \hat{\theta}_{k-1}) \tag{3.112}$$

$$P_{k-1} = P_{k-2} - \frac{P_{k-2}\phi_{k-d}\phi_{k-d}^T P_{k-2}}{1 + \phi_{k-d}^T P_{k-2}\phi_{k-d}} \,. \tag{3.113}$$

Hence the estimator does not inherit the optimality properties of the Kalman filter. In other words, since the noise ν_k is colored the unbiased estimate $\hat{\theta}_k$ generated by least squares will be nonoptimal. The estimate $\hat{\theta}_k$ is unbiased because ϕ_{k-d} and ν_k are uncorrelated.

Because of the difficulty discussed above, in 1974, to handle the case with general d without involving a noise with unknown statistical characteristics, Nahorski and Vidal [40] used the following particular model

$$y_{k+d} = \sum_{j=0}^{n-1} \alpha_j y_{k-j} + \sum_{j=0}^{n-1} \beta_j u_{k-j} + e_{k+d} \tag{3.114}$$

where e_{k+d} was assumed a stochastic variable independent of $\mathcal{F}_k$. Moreover, the disturbance e_{k+d} was assumed a normal stochastic variable with zero mean and variance σ^2. The model (3.114) can be written in the following regression form:

$$y_{k+d} = \phi_k^T \theta + e_{k+d} \tag{3.115}$$

where θ is an $2n \times 1$ column vector consisting the coefficients of $[\alpha(q^{-1}), \beta(q^{-1})]$ as follows:

$$\theta = [\alpha_0, \ldots, \alpha_{n-1}, \beta_0, \ldots, \beta_{n-1}]^T \tag{3.116}$$

and the regression vector ϕ_k is defined as

$$\phi_k^T = [y_k, \ldots, y_{k-n+1}, u_k, \ldots, u_{k-n+1}] \; , \tag{3.117}$$

these components being available at time k.

Based on the initial assumption the system (3.114) is time invariant, i.e., $\theta_{k+1} = \theta_k = \theta$ and using (3.115), the following dynamical system is obtained to estimate the parameters:

$$\theta_{k+1} = \theta_k \tag{3.118a}$$

$$y_k = \phi_{k-d}^T \theta_k + e_k \; . \tag{3.118b}$$

It was assumed that θ_0 is a gaussian random variable of known mean $\hat{\theta}_{0|-1}$ and known covariance $P_{0|-1}$, i.e.,

$$\hat{\theta}_{0|-1} = E\{\theta | \mathcal{F}_{-1}\} \qquad P_{0|-1} = E\{(\theta_0 - \hat{\theta}_{0|-1})(\theta_0 - \hat{\theta}_{0|-1}) | \mathcal{F}_{-1}\} . \tag{3.119}$$

It was further assumed that θ_0 is independent of e_k for any k.

Based on the assumptions above and using (3.118), it can be shown that the conditional distribution of θ_k given $\mathcal{F}_k$ is normal. Hence, the unknown parameters can be estimated using a Kalman filter, which gives the estimates, $\hat{\theta}_{k|k}$, and the covariance matrix, $P_{k|k}$, based on data obtained up to and included time k, i.e. $\mathcal{F}_k$, as follows:

$$\hat{\theta}_{k|k} = \hat{\theta}_{k|k-1} + K_k(y_k - \phi_{k-d}^T \hat{\theta}_{k|k-1}) \tag{3.120a}$$

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k} \tag{3.120b}$$

$$K_k = P_{k|k-1}\phi_{k-d}(\phi_{k-d}^T P_{k|k-1}\phi_{k-d} + \sigma^2)^{-1} \tag{3.120c}$$

$$P_{k|k} = P_{k|k-1} - K_k \phi_{k-d}^T P_{k|k-1} \tag{3.120d}$$

$$P_{k+1|k} = P_{k|k} \tag{3.120e}$$

with initial values $\hat{\theta}_{0|-1}$, $P_{0|-1}$, and

$$[y_{-1}, \ldots, y_{-d-n+1}, u_{-1}, \ldots, u_{-d-n+1}] . \tag{3.121}$$

Under these assumptions and using the general formula (3.34), it will be shown below how the optimal control law can be obtained. The result will be the same as that of Nahorski and vital.

From (3.34) and using $\beta_0 = b_0$, (2.62), we have

$$E\{\beta_0 y_{k+d} | \mathcal{F}_k\} = E\{\beta_0 | \mathcal{F}_k\} y_{k+d}^* . \tag{3.122}$$

Substituting (3.114) into (3.122) gives

$$E\{\beta_0(\sum_{j=0}^{n-1}\alpha_j y_{k-j} + \sum_{j=0}^{n-1}\beta_j u_{k-j} + e_{k+d})|\,\mathcal{F}_k\} = E\{\beta_0|\,\mathcal{F}_k\}y^*_{k+d}\,. \qquad (3.123)$$

Then the control law explicitly can be obtained:

$$u_k = -\frac{\sum_{j=0}^{n-1}(\hat{\beta}_{0,k|k}\hat{\alpha}_{j,k|k} + P_{\beta_0\alpha_j|k})y_{k-j} + \sum_{j=1}^{n-1}(\hat{\beta}_{0,k|k}\hat{\beta}_{j,k|k} + P_{\beta_0\beta_j|k})u_{k-j} - \hat{\beta}_{0,k|k}y^*_{k+d}}{\hat{\beta}^2_{0,k|k} + P_{\beta_0^2|k}} \qquad (3.124)$$

where the estimates satisfy the equations (3.120). As expected, using the proposed approach we reached the same result as Nahorski and Vital. Once again, this verifies the generality of the proposed approach.

Next a simulated example showing the performance the proposed algorithm, Section 3.6, will be presented.

3.8 Simulation results

In this section, an example of a minimum-phase system will be presented to illustrate the algorithm developed in this research, Section 3.6. The performance of this algorithm will be compared to that of the adaptive minimum variance control designed based on the certainty-equivalence principle.

Consider the following SISO linear time invariant system,

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \qquad (3.125)$$

where

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1} + a_2 q^{-2} \\ B(q^{-1}) &= q^{-2}(b_0)\,. \end{aligned}$$

The system above can be written in the following form:

$$y_k = -a_1 y_{k-1} - a_2 y_{k-2} + b_0 u_{k-2} + e_k \; . \tag{3.126}$$

Then (3.126) can be written in the following regression form:

$$y_k = \phi_{k-1}^T \theta + e_k \tag{3.127}$$

where

$$\begin{aligned} \theta &= [a_1, a_2, b_0]^T \\ \phi_{k-1}^T &= [-y_{k-1}, -y_{k-2}, u_{k-2}] \; . \end{aligned}$$

The true value of the parameter vector θ is

$$\theta = [1, -2, 1.5]^T \; . \tag{3.128}$$

The minimum phase system above has one pole on the unit circle and an unstable pole at -2.0. It is assumed that the unknown parameter vector θ, generated by a random generator, is a gaussian random vector with a priori statistics having mean and covariance

$$\hat{\theta}_{0|-1} = [2.9822, 1.6839, 1.7374]^T \tag{3.129}$$

$$P_{0|-1} = 10 * I \; . \tag{3.130}$$

The initial state is assumed to be known:

$$x_0 = 0 \; . \tag{3.131}$$

The measurements noise $\{e_k\}$ is a white zero-mean noise with known covariance $R = 10^{-6}$. The signal to noise ratio is rather high to avoid running the program many times to obtain the mean values of estimation as well as the input and output signals.

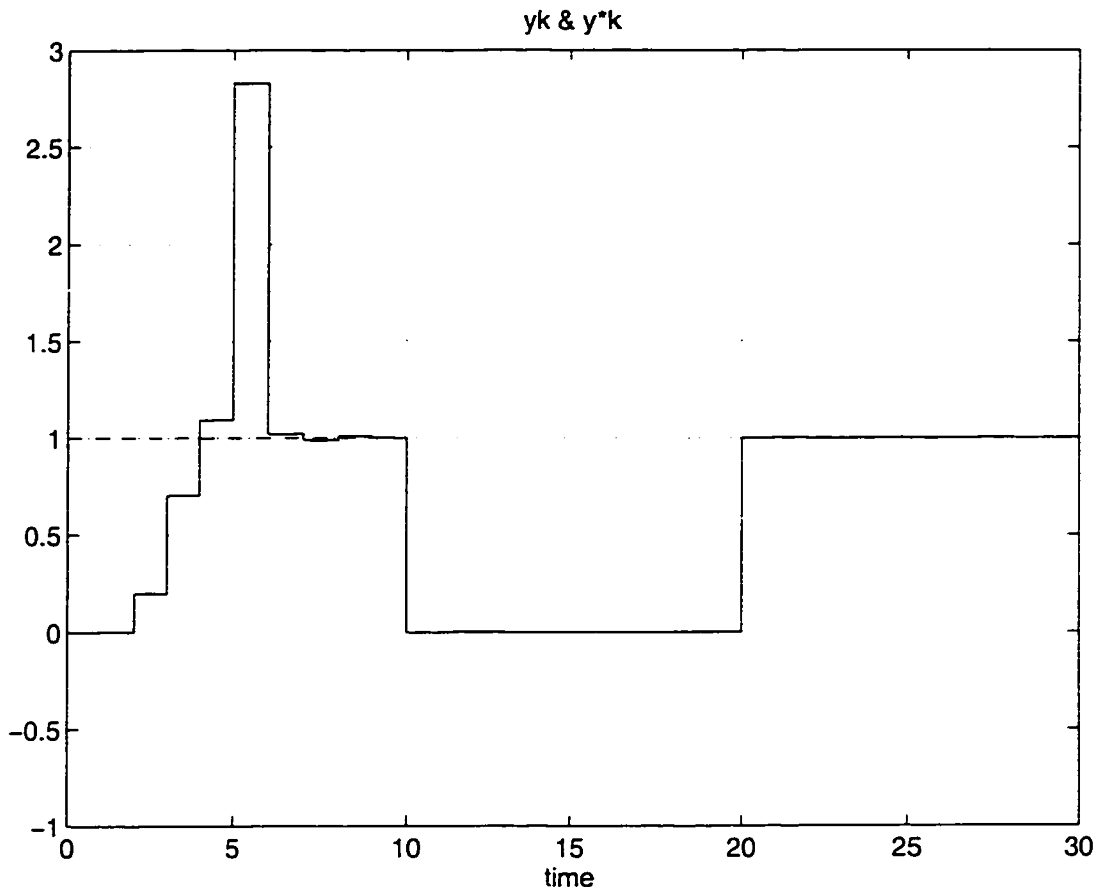


Figure 3.1: Output y_k (solid line) and reference signal y_k^* (dashed line)

The desired output sequence $\{y_k^*\}$, is taken to be a unit square wave of period 20 samples.

The performance of the indirect cautious control with known initial condition algorithm (3.82) to (3.83), proposed in this research, Section 3.6, is shown in Figs 3.1 to 3.5.

Figures 3.1 to 3.3 show the output and reference signal, input signal, and the estimated parameter, respectively. They correspond to one sample run for this example. Fig. 3.1 shows that the algorithm converges and that perfect tracking is

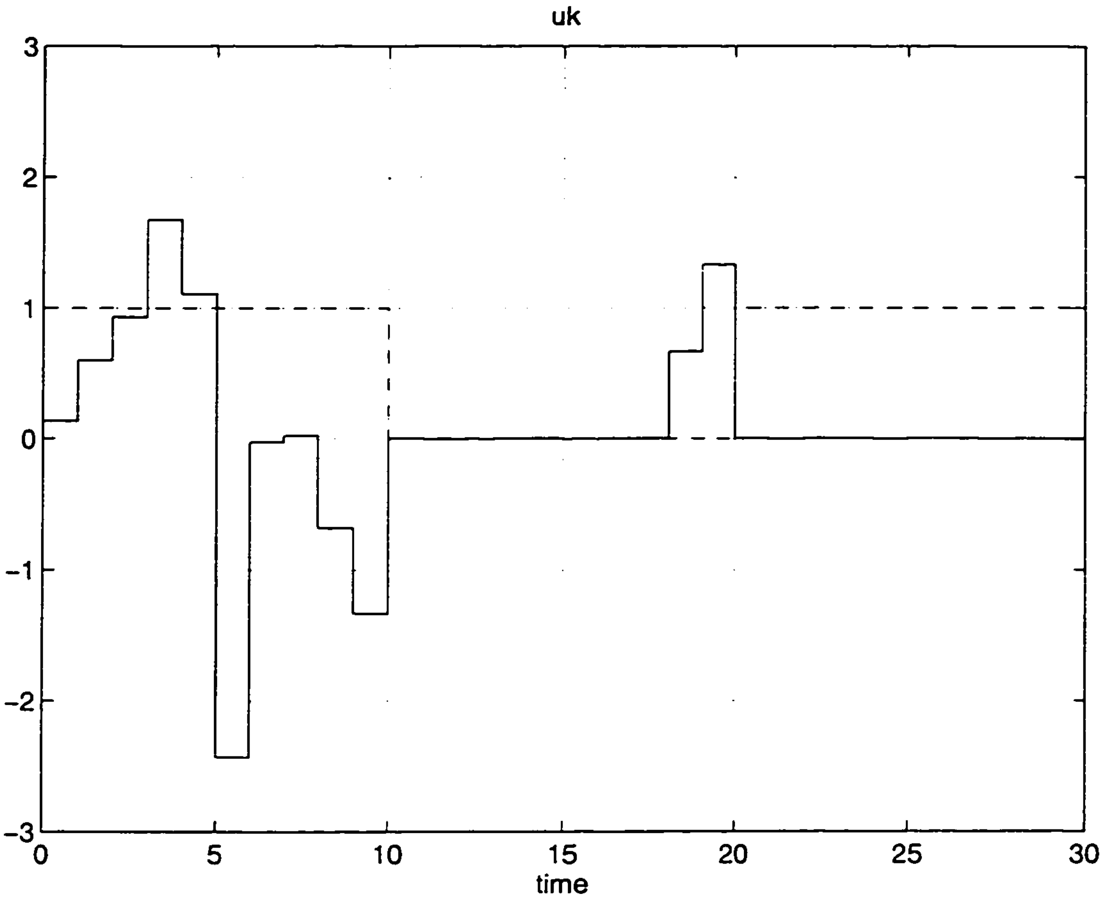


Figure 3.2: Input signal u_k

ultimately achieved. As shown by Fig. 3.3, the true parameter set is identified in less than 5 samples.

Fig. 3.4 demonstrates the superiority of the proposed adaptive algorithm over the certainty-equivalence controller. Fig. 3.5 exhibits the transient and steady-state performance of both the cautious control and the certainty-equivalence controller. As expected, both the controllers have the same steady-state behavior; However, the cautious control shows a better transient behavior.

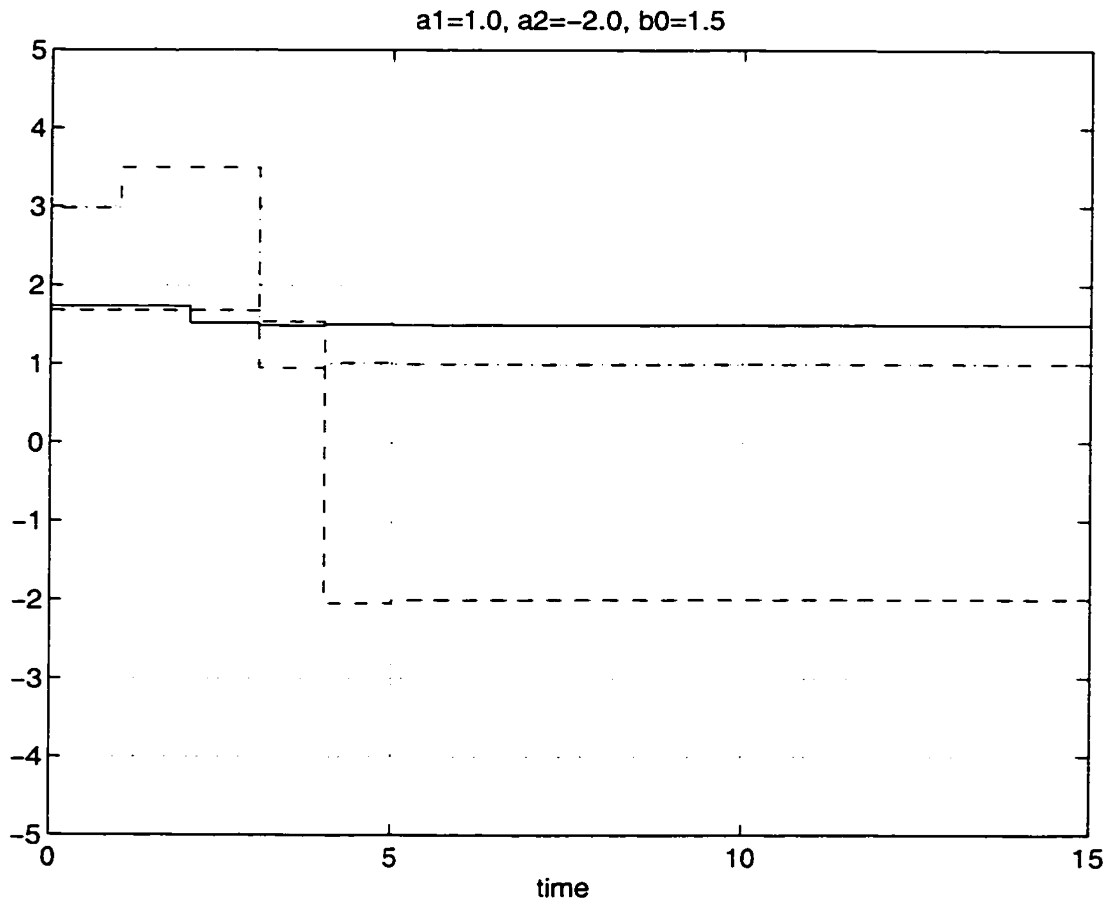


Figure 3.3: Estimated parameters; a_1 (dotted and dashed line), a_2 (dashed line), and b_0 (solid line)

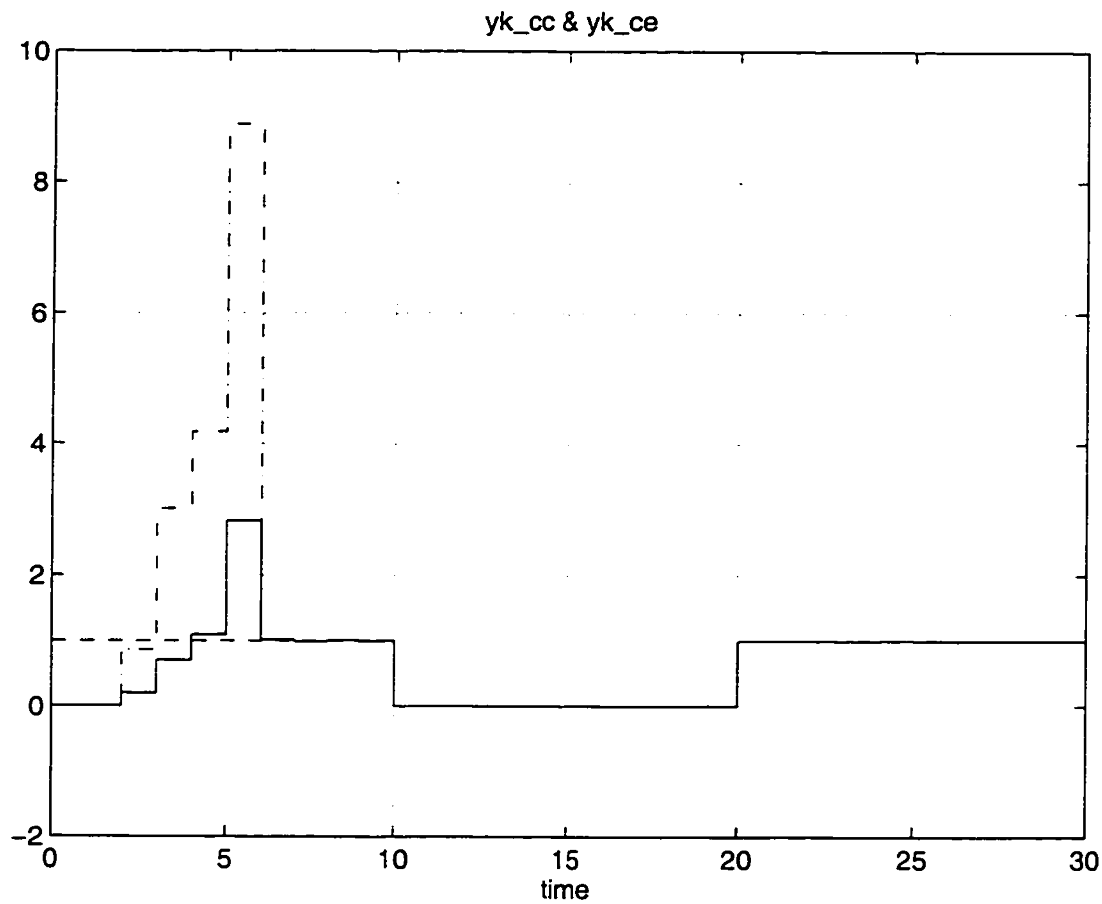


Figure 3.4: Cautious control response (solid line) and certainty-equivalence controller response (dotted and dashed line)

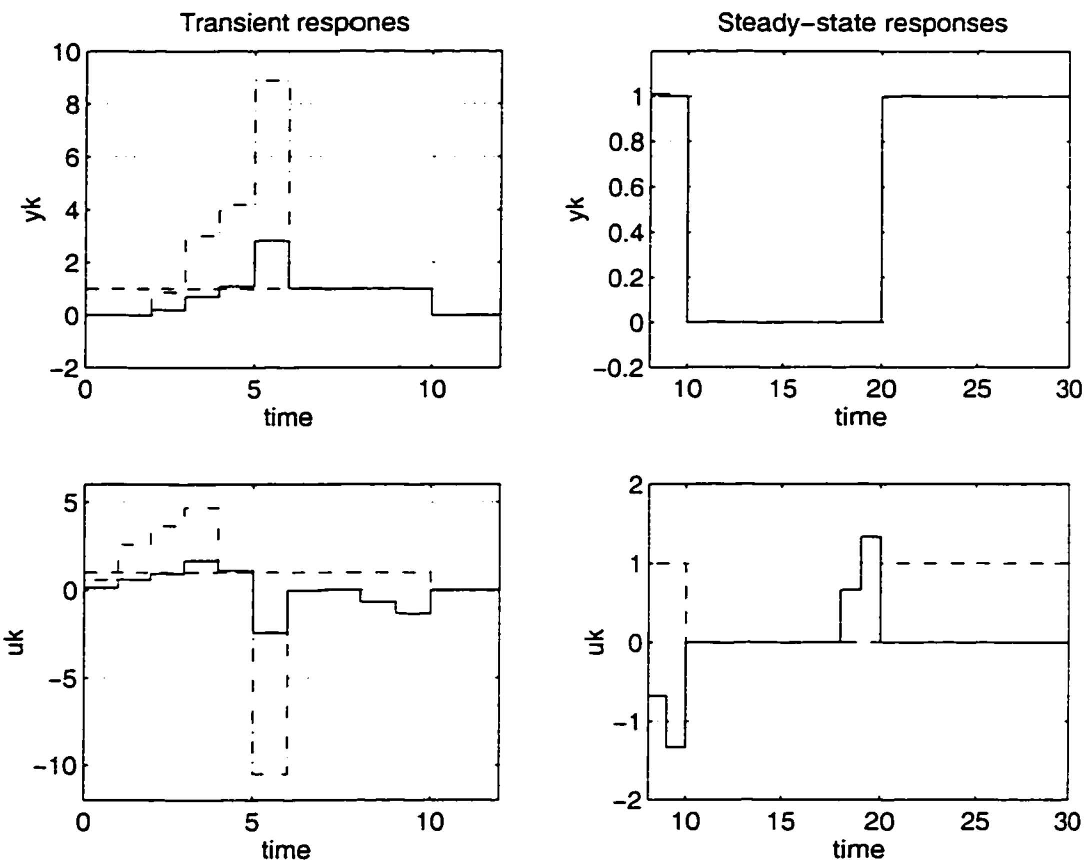


Figure 3.5: Transient and steady-state responses; the left graphs show the transient responses and the right graphs show the steady-state responses. Cautious control responses (solid line), certainty-equivalence controller responses (dotted and dashed line), and reference signal y_k^* (dashed line)

Chapter 4

MIMO adaptive control

4.1 Introduction

This chapter is concerned with indirect MIMO stochastic adaptive control. The goal is to design an adaptive minimum variance controller for MIMO linear time-invariant discrete-time systems having white noise perturbation and unknown interactor matrix.

The new indirect approach proposed in this research is based on the result obtained by Das [17]. Das (1986) [17] showed that the a priori knowledge of the degrees of the diagonal entries of the interactor matrix alone is sufficient for proving global stability. Dion, Dugard, and Carrillo (1988) [18] verified the result obtained by Das [17] using a different approach. As for the SISO case, discussed in Chapter 3, it is shown how the general optimal solution of the one-step-ahead criterion, defined for a MIMO system with unknown interactor matrix, can be obtained. Using this solution, it is proved that MIMO cautious control can be designed only if the interactor matrix is known. However, if the interactor matrix is unknown, an exact solution for the MIMO one-step-ahead optimization problem cannot be obtained. Hence, the most popular approximation method, i.e. the certainty equivalence principle, is employed. Then it will be shown that the conventional approaches, [30, 19, 17], cannot be applied in the stochastic case. However, the proposed approach is capable of dealing with stochastic cases.

This chapter is organized as follows. In Section 4.2 the system model is described. The concept of an interactor matrix which is a generalized concept of a relative degree of a scalar system is reviewed in Section 4.3. In Section 4.4 the control criterion is addressed. In Section 4.5 using state variables, it is shown how the system model can be transformed into a special model proposed by Salut [45] or a canonical singular pencil (SP). This minimal canonical model has been discussed in detail by Aplevich [4, 3]. Then using this minimal canonical model, in Section 4.6 it is shown how the system model can be transformed into various predictor forms. In Section 4.7 it is shown how the entries of the interactor matrix are related to the entries of the canonical ARX model using two different methods. In Section 4.8 as for the SISO case, the general optimal solution of the one-step-step ahead criterion, defined for a MIMO system with unknown interactor matrix, is obtained. Then in Section 4.9 it is shown how the general optimal solution obtained in this research can be applied to some special cases discussed in the literature. In Section 4.10 in order to explain why the conventional direct approach cannot be applied in the stochastic case, the direct adaptive control proposed by Dugard, Goodwin, and deSouza [19] is considered. In Section 4.11 the proposed indirect stochastic adaptive control with unknown interactor matrix is presented. At the end of this chapter, Section 4.12, simulation results show the performance of the proposed algorithm compared with that of the minimum variance controller with known parameters.

4.2 System description

A linear time-invariant MIMO system is assumed to be represented by the following ARX model:

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \tag{4.1}$$

where y_k is $p \times 1$, u_k is $m \times 1$, and e_k is $p \times 1$. $A(q^{-1})$, $B(q^{-1})$ are polynomial matrices in the unit delay operator q^{-1} which have the following forms:

$$A(q^{-1}) = A_0 + A_1 q^{-1} + \cdots + A_v q^{-v} \tag{4.2a}$$

$$B(q^{-1}) = B_0 + B_1 q^{-1} + \cdots + B_v q^{-v} \tag{4.2b}$$

where v is the maximum of the degrees of the polynomials in $A(q^{-1})$ and $B(q^{-1})$. The independent sequence $\{e_k\}$ is a gaussian zero mean white noise process with given covariance

$$E\{e_k e_l^T\} = R_k \delta_{k-l} \quad . \quad \mathrm{tr} R_k < \infty \tag{4.3}$$

where δ_{k-l} is the Kronecker delta and $tr(.)$ is the trace operator.

In the sequel the following assumptions, as used by Wolovich and Falb [54], will be employed:

Assumption 4.1 *The number of inputs, m, is equal to the number of outputs, p, and the input-output transfer function $T(z) = A(z^{-1})^{-1}B(z^{-1})$ satisfies*

$$det T(z) \neq 0 \qquad \textit{almost all } z \tag{4.4}$$

Assumption 4.2 *The transfer function $T(z)$ is strictly proper,*

$$\lim_{z \to \infty} T(z) = 0 \ . \tag{4.5}$$

Note that Assumption 4.1 ensures output tracking and Assumption 4.2 ensures that there is a delay of at least one unit between each input and each output. Hence, for a strictly proper system B_0 is zero in (4.2b).

4.3 The interactor matrix

An important concept for design of multivariable zero cancelling control strategies, such as deadbeat and model matching is the system delay structure, discussed in Elliott and Wolovich (1984) [22]. As seen in Chapter 2, for the SISO system (2.1) the delay structure was very transparent. In fact, one simply chose the delay, d, so that the leading coefficient of $q^d B(q^{-1})$ was nonzero. This can be stated slightly more formally by saying that in the SISO case, there exists a scalar function $\xi(q)$ of the form $\xi(q) = q^d$ such that

$$\lim_{z\to\infty} \xi(z)A(z^{-1})^{-1}B(z^{-1}) = k_T \tag{4.6}$$

where k_T is a nonzero scalar.

In the multivariable case it turns out that the delay structure of the transfer function matrix can also be specified in terms of a polynomial matrix $\xi(q)$. The following result proved by Wolovich and Falb [54] applies to the multivariable case.

Lemma 4.1 *Given any transfer function satisfying Assumptions 4.1 and 4.2, there exists a polynomial matrix, $\xi(q)$, known as the interactor matrix, satisfying:*

$$\lim_{z\to\infty} \xi(z)T(z) = K_T \tag{4.7}$$

where K_T is a nonsingular matrix.

In general, $\xi(z)$ can be taken to have the following structure:

$$\xi(z) = H(z)D(z) \tag{4.8}$$

where

$$D(z) = diag[z^{d_1}, \dots, z^{d_p}] , \tag{4.9}$$

$d_i \geq \min_{1\leq j\leq p} d_{ij}$ and d_{ij} is the delay between the jth input and ith output. $H(z)$ is a lower triangular, unimodular matrix, with ones on the diagonal as follows:

$$H(z) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ h_{21}(z) & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ h_{p1}(z) & h_{p2}(z) & \cdots & 1 \end{bmatrix} \tag{4.10}$$

and $h_{ij}(z)$ is divisible by z or is zero.

Proof: See Ref. [54] or Ref. [28].

In [17] Das proposed a simple formula to determine the upper bounds on the degrees of the off diagonal entries of $\xi(q)$. The main result obtained by Das is stated in the following corollary:

Corollary 4.1 *Upper bounds on the degrees of the off diagonal entries of $\xi(q)$ are given by*

$$\partial(\xi_{ij}(q)) \leq (\sum_{k=j}^{i} d_k) - (i-j) \tag{4.11}$$
$$\text{for } i > j \qquad i,j = 1,\ldots,p\,.$$

Proof: See Ref. [17].

Dion, Dugard, and Carrillo (1988) [18] verified the result obtained by Das [17] using a different approach. In the sequel it will be shown how the results obtained by Das can be employed to design an adaptive controller for a MIMO stochastic system.

4.4 The control criterion

The objective is to choose the input so as to minimize the mean-square error between the filtered output $\xi(q)y_k$ and the filtered desired output $\xi(q)y_k^*$. Thus consider the following one-step-ahead cost function

$$J_{k+d'} = E\{\|\xi(q)(y_k - y_k^*)\|^2\} \tag{4.12a}$$

$$= E\{\mathrm{tr}[\xi(q)(y_k - y_k^*)(y_k - y_k^*)^T\xi(q)^T]\} . \tag{4.12b}$$

Using the smoothing property of conditional expectations, we have

$$J_{k+d'} = E\{E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k\}\} . \tag{4.13}$$

Hence the optimal cost is

$$\begin{aligned} J_{k+d'}^* &= \min_{u_k} E\{E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k\}\} \\ &= E\{\min_{u_k} E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k\}\} \end{aligned} \tag{4.14}$$

where u_k is constrained to be $\mathcal{F}_k$ measurable. In other words. the new criterion may be written as

$$J_{k+d'|k} = E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k\} . \tag{4.15}$$

Once again using the property of conditional means the index becomes

$$J_{k+d'} = E\{E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k, x_0, \theta\}\} \tag{4.16}$$

where x_0 is the initial condition and θ is the parameter vector. A new index when x_0 and θ are known will be obtained as follows:

$$J_{k+d'|k,x_0,\theta} = E\{\|\xi(q)(y_k - y_k^*)\|^2|\mathcal{F}_k, x_0, \theta\} . \tag{4.17}$$

As in the SISO case, a key step in deriving the control signal to minimize (4.15) and (4.17) is to develop a predictor form for $\xi(q)y_k$.

4.5 Minimal canonical representations

In order to solve the optimal problem above, the same procedure as used in Chapter 3 will be used. Before beginning this discussion, a brief introduction to the class of minimal canonical representations required is given in this section, discussed in detail by Aplevich [2, 3].

Using state variables, the ARX model (4.1) is transformed into a special canonical model proposed by Salut [45]. First, by introducing auxiliary variables in a vector x_k with $n = vp$ entries, the ARX model (4.1) is written as

$$\begin{bmatrix} I_n \\ 0_{p\times n} \end{bmatrix} x_{k+1} = \begin{bmatrix} 0_{p\times p} & & 0 \\ I_p & & \\ & \ddots & \\ 0 & & I_p \end{bmatrix} x_k - \begin{bmatrix} A_v \\ \vdots \\ A_1 \\ A_0 \end{bmatrix} y_k + \begin{bmatrix} B_v \\ \vdots \\ B_1 \\ B_0 \end{bmatrix} u_k + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ I_p \end{bmatrix} e_k \quad (4.18)$$

where I_α denotes an identity matrix of order α and $0_{\alpha\times\beta}$ an $\alpha \times \beta$ zero matrix.
In the special case of a scalar system, for which $m = p = 1$, all sub-matrices $A_{(.)}$, $B_{(.)}$ of equation (4.18) are scalars, and the system parameters are obviously identical to those of equation (4.1). For both efficiency and uniqness of identification, a canonical form is required. In general, however, (4.18) must be modified slightly to be canonical. The details are given in References [45] and [2]; here only a summary of the results will be given.

It has been shown by Aplevich [2] that by premultiplying (4.18) by a non-singular matrix, and by selectively deleting rows and columns, a canonical form can be obtained which is externally equivalent to the original in the sense that the set of external vectors $\{y_k, u_k\}$ admitted by the equations is invariant. The canonical form to be described is a minor variation of that proposed by Salut in [45], and for simplicity will be written

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k \tag{4.19a}$$

$$0 = E_0 x_k - A_0 y_k + B_0 u_k + e_k \tag{4.19b}$$

A canonical model of the above form has the following properties:

- For causal discrete-time systems, the row degrees of q^{-1} in $A(q^{-1})$ are no less than the corresponding row degrees of $B(q^{-1})$, and $A(q^{-1})$ is invertible. Thus A_0 is invertible, and furthermore in the canonical form considered, A_0 is upper right triangular with unit diagonal entries.

- The integers n_i, $i = 1, \ldots, p$ are the *observability indices* of the system and $n = n_1 + n_2 + \cdots + n_p$.

- The matrix $E = \begin{bmatrix} E_* \\ E_0 \end{bmatrix}$ is zero except for a positive unit element in each column. The unit elements called pivots, in the n non-zero rows are to the right of pivot elements in superior rows. The locations of the pivots are uniquely defined by n_i such that if row i of E_0 is denoted by E_{0i}, then $E_{0i}(E_*)^{n_i}$ is the lowest zero row in the matrix

$$\tilde{E} = \begin{bmatrix} E_0(E_*)^v \\ \vdots \\ E_0 E_* \\ E_0 \end{bmatrix}. \tag{4.20}$$

  Thus E contains zero rows $E_{0i}(E_*)^{n_i}$ together with the n non-zero rows of $\tilde{E}$, so that each row of E is of the form $E_{0i}(E_*)^j$, and may be referred to as row (i, j).

- The non-identically-zero, non-unit entries in $A = \begin{bmatrix} A_* \\ A_0 \end{bmatrix}$ are designated $a_{ik}^{(j)}$ and their locations are uniquely determined by $\{n_i\}$. Each $a_{ik}^{(j)}$ is in row (i,j) of A, using the row indexing of E above, and column $k = 1, 2, \ldots, p$. There is an entry $a_{ik}^{(j)}$ for each non-zero row of the form $E_{0k}(E_*)^{n_i - j}$ below $E_{0i}(E_*)^{n_i}$ in $\tilde{E}$, that is, for each non-zero row with index $(k, n_i - j)$ below row (i, n_i) in E.

- For causal systems the matrix $B = \begin{bmatrix} B_* \\ B_0 \end{bmatrix}$ does not have a special structure and the subscripts of its entries $b_{ik}^{(j)}$ are as for $a_{ik}^{(j)}$ except that $k = 1, 2, \ldots, m$. For a strictly proper system, B_0 is zero.

It has been shown by Aplevich [2] that the parameters $\{a_{ik}^{(j)}\}$ and $\{b_{ik}^{(j)}\}$ in (4.19) are identical to those in a canonical ARX model (4.1). Specifically, the parameters $a_{ik}^{(j)}$ and $b_{ik}^{(j)}$ in (4.19) are the ik entries of the canonical coefficient matrices A_j and B_j in equations (4.2a)-(4.2b), respectively.

Equation (4.19) can be written

$$P(\lambda) \begin{bmatrix} x_k \\ y_k \\ u_k \\ e_k \end{bmatrix} = \begin{bmatrix} E_* - \lambda I & -A_* & B_* & 0 \\ E_0 & -A_0 & B_0 & I \end{bmatrix} \begin{bmatrix} x_k \\ y_k \\ u_k \\ e_k \end{bmatrix} = 0 \tag{4.21}$$

where $P(\lambda)$ can be regarded as a singular pencil of matrices. Hence the above model is sometimes called a singular pencil (SP) model, [3].

It has been proved by Aplevich [2] that with $[-A_0 \; B_0]$ in reduced row echelon form, the dynamical system above, (4.21), is minimal iff it possesses the two properties

Property 4.1 $[E_* - \lambda I \quad -A_* \quad B_*]$ *has full row rank for all complex* λ.

Property 4.2 $\begin{bmatrix} E_* - \lambda I \\ E_0 \end{bmatrix}$ *has full column rank for all complex* λ.

The following example further illustrates these points.

Example 4.1:

Consider a 2-input, 2-output strictly proper system used in [?],

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \tag{4.22a}$$

where

$$A(q^{-1}) = \begin{bmatrix} (1+q^{-1})(1+2q^{-1}) & 0 \\ 0 & (1+3q^{-1})(1+4q^{-2}) \end{bmatrix} \tag{4.22b}$$

$$B(q^{-1}) = \begin{bmatrix} q^{-1}+2q^{-2} & q^{-1}+q^{-2} \\ q^{-1}+4q^{-2} & q^{-1}+3q^{-2} \end{bmatrix} \tag{4.22c}$$

The corresponding SP model is

$$\begin{bmatrix} -\lambda & 0 & 0 & 0 & -2.0 & 0.0 & 2.0 & 1.0 \\ 0 & -\lambda & 0 & 0 & 0.0 & -12.0 & 4.0 & 3.0 \\ 1 & 0 & -\lambda & 0 & -3.0 & 0 & 1.0 & 1.0 \\ 0 & 1 & 0 & -\lambda & 0 & -7.0 & 1.0 & 1.0 \\ 0 & 0 & 1 & 0 & -1 & 0.0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_k^{(1)} \\ x_k^{(2)} \\ x_k^{(3)} \\ x_k^{(4)} \\ y_k^{(1)} \\ y_k^{(2)} \\ u_k^{(1)} \\ u_k^{(2)} \\ e_k^{(1)} \\ e_k^{(2)} \end{bmatrix} = 0 \ . \tag{4.23}$$

Note that throughout this chapter the entries 0 and ±1, as opposed to 0.0 and ±1.0, in the system pencil are used to denote *identically zero* and *identically unit* entries, which do not need to be estimated.

Since $[-A_0\ B_0]$ in (4.23) is in reduced row echelon form and the dynamical system (4.23) satisfy Properties 4.1 and 4.2, it can be deduced that the minimum dimension of the system is $n = 4$. By constructing $\tilde{E}$ in (4.20), $\{n_i\} = \{2, 2\}$ which the integers n_i satisfy $n_1 + n_2 = n$, as expected.

The structure of the canonical SP form for this system, obtained by applying the algorithms developed in References [2] and [24], is given in equation (4.19) with

$$\begin{bmatrix} E_* \\ E_0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{matrix} \text{row}(1,2) \\ \text{row}(2,2) \\ \text{row}(1,1) \\ \text{row}(2,1) \\ \text{row}(1,0) \\ \text{row}(2,0) \end{matrix} \tag{4.24a}$$

$$\begin{bmatrix} -A_* & B_* \\ -A_0 & B_0 \end{bmatrix} = \begin{bmatrix} -a_{11}^{(2)} & -a_{12}^{(2)} & b_{11}^{(2)} & b_{12}^{(2)} \\ -a_{21}^{(2)} & -a_{22}^{(2)} & b_{21}^{(2)} & b_{22}^{(2)} \\ -a_{11}^{(1)} & 0 & b_{11}^{(1)} & b_{12}^{(1)} \\ 0 & -a_{22}^{(1)} & b_{21}^{(1)} & b_{22}^{(1)} \\ -1 & -a_{11}^{(0)} & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}. \tag{4.24b}$$

The corresponding canonical ARX model can be obtained from the system pencil by inspection, with

$$A(q^{-1})y_k = B(q^{-1})u_k + e_k \tag{4.25}$$

where

$$
\begin{aligned}
A(q^{-1}) &= \begin{bmatrix} 1 & a_{11}^{(0)} \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} a_{11}^{(1)} & 0 \\ 0 & a_{22}^{(1)} \end{bmatrix} q^{-1} + \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} \\ a_{21}^{(2)} & a_{22}^{(2)} \end{bmatrix} q^{-2} \\
&= \begin{bmatrix} 1 & 0.0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 3.0 & 0 \\ 0 & 7.0 \end{bmatrix} q^{-1} + \begin{bmatrix} 2.0 & 0.0 \\ 0.0 & 12.0 \end{bmatrix} q^{-2} \\
B(q^{-1}) &= \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} q^{-1} + \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \end{bmatrix} q^{-2} \\
&= \begin{bmatrix} 1.0 & 1.0 \\ 1.0 & 1.0 \end{bmatrix} q^{-1} + \begin{bmatrix} 2.0 & 1.0 \\ 4.0 & 3.0 \end{bmatrix} q^{-2}
\end{aligned}
$$

Note that the canonical matrix polynomials $A(q^{-1})$ and $B(q^{-1})$ are obtained from (4.24) by inspection.

Using the minimal canonical model (4.19) the cost function (4.12) can be minimized including the initial conditions. As the SISO case, it is necessary that a ARX or the model (4.19) be expressed in a predictor form.

4.6 MIMO predictor forms

In this section it is shown how a MIMO system having a minimal canonical model can be transformed to a predictor form. Then it is shown the conventional predictor in the literature, discussed by Goodwin and Sin in [28], can be considered as a special case of the predictor proposed in this research.

As explained in Section 4.5, in the MIMO case an ARX model can be written in the minimal canonical form (4.19) or

$$
\begin{aligned}
x_{k+1} &= E_* x_k - A_* y_k + B_* u_k && \text{(4.26a)} \\
y_k &= A_0^{-1} E_0 x_k + A_0^{-1} B_0 u_k + A_0^{-1} e_k && \text{(4.26b)}
\end{aligned}
$$

since A_0 is required to be an invertible matrix. Substituting (4.26b) into (4.26a) gives the following standard state-space model:

$$x_{k+1} = \bar{A}x_k + \bar{B}u_k - A_* A_0^{-1} e_k \tag{4.27a}$$

$$y_k = \bar{C}x_k + \bar{D}u_k + A_0^{-1} e_k \tag{4.27b}$$

where

$$\bar{A} = E_* - A_* A_0^{-1} E_0 \tag{4.27c}$$

$$\bar{B} = B_* - A_* A_0^{-1} B_0 \tag{4.27d}$$

$$\bar{C} = A_0^{-1} E_0 \tag{4.27e}$$

$$\bar{D} = A_0^{-1} B_0 \, . \tag{4.27f}$$

As mentioned, for a strictly proper system, B_0 is zero. Hence, the two models above (4.26) and (4.27), are simplified as follow, respectively:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k \tag{4.28a}$$

$$y_k = A_0^{-1} E_0 x_k + A_0^{-1} e_k \tag{4.28b}$$

and the corresponding state-space model is

$$x_{k+1} = \bar{A}x_k + B_* u_k - A_* A_0^{-1} e_k \tag{4.29a}$$

$$y_k = \bar{C}x_k + A_0^{-1} e_k \, . \tag{4.29b}$$

The following lemma shows how a predictor form for a MIMO model can be derived or how the filtered output $\xi(q)y_k$ can be determined.

Lemma 4.2 *Subject to Assumptions 4.1 and 4.2 and using the minimal canonical model (4.28), the filtered output $\xi(q)y_k$ is related to $\{u_k\}$ and $\{y_k\}$ by a model of the following predictor form:*

$$\xi(q)y_k = \sum_{l=0}^{d'-1} \Lambda_l \bar{C} \bar{A}^{d'-l-1} x_{k+1} + \sum_{j=0}^{d'-1} F_j e_{k+d'-j} \tag{4.30}$$

where

$$x_{k+1} = E_*^{k+1}x_0 - \sum_{j=0}^{k} E_*^{k-j}A_*y_j + \sum_{j=0}^{k} E_*^{k-j}B_*u_j \tag{4.31}$$

$$\xi(q) = \Lambda_0 q^{d'} + \Lambda_1 q^{d'-1} + \cdots + \Lambda_{d'-1}q \tag{4.32}$$

$$d' = \text{maximum advance in } \xi(q) \tag{4.33}$$

for $k \geq 0$ and $d' \geq 1$. Furthermore, the coefficient matrices $F_0, \ldots, F_{d-1}$ can be computed as follows:

$$F_j = \begin{cases} \Lambda_0 A_0^{-1} & j = 0 \\ (\Lambda_j - \sum_{l=0}^{j-1} \Lambda_l \bar{C}\bar{A}^{j-l-1}A_*)A_0^{-1} & j = 1, \ldots, d' - 1 \end{cases} \tag{4.34}$$

Proof: As the SISO case, from (4.28) and using a sequential method, we reach

$$x_{k+1} = E_*^{k+1}x_0 - \sum_{j=0}^{k} E_*^{k-j}A_*y_j + \sum_{j=0}^{k} E_*^{k-j}B_*u_j \tag{4.35}$$

for $k \geq 0$. Following (4.29), i-step-ahead prediction may be developed as follows. First,

$$x_{k+2} = \bar{A}x_{k+1} + B_*u_{k+1} - A_*A_0^{-1}e_{k+1} \,. \tag{4.36}$$

Similarly,

$$\begin{aligned} x_{k+3} &= \bar{A}x_{k+2} + B_*u_{k+2} - A_*A_0^{-1}e_{k+2} \\ &= \bar{A}^2x_{k+1} + (\bar{A}B_*u_{k+1} + B_*u_{k+2}) - (\bar{A}A_*A_0^{-1}e_{k+1} + A_*A_0^{-1}e_{k+2}) \,. \end{aligned}$$

Hence

$$x_{k+i} = \bar{A}^{i-1}x_{k+1} + \sum_{j=1}^{i-1} \bar{A}^{i-j-1}B_*u_{k+j} - \sum_{j=1}^{i-1} \bar{A}^{i-j-1}A_*A_0^{-1}e_{k+j} \tag{4.37}$$

for $k \geq 0$ and $i > 1$, and also from (4.29b)

$$y_{k+1} = \bar{C}x_{k+1} + A_0^{-1}e_{k+1} \tag{4.38}$$

and

$$y_{k+i} = \bar{C}x_{k+i} + A_0^{-1}e_{k+i} \tag{4.39}$$

for $k \geq 0$ and $i \geq 1$. Substituting (4.39) into (4.37) gives

$$y_{k+i} = \bar{C}\bar{A}^{i-1}x_{k+1} + \sum_{j=1}^{i-1}\bar{C}\bar{A}^{i-j-1}B_*u_{k+j} - \sum_{j=1}^{i-1}\bar{C}\bar{A}^{i-j-1}A_*A_0^{-1}e_{k+j} + A_0^{-1}e_{k+i} \tag{4.40}$$

for $k \geq 0$ and $i > 1$.

The filtered output $\xi(q)y_k$ can be written as follows:

$$\begin{aligned} \xi(q)y_k &= (\Lambda_0 q^{d'} + \Lambda_1 q^{d'-1} + \cdots + \Lambda_{d'-1}q)\, y_k \\ &= \Lambda_0 y_{k+d'} + \Lambda_1 y_{k+d'-1} + \cdots + \Lambda_{d'-1}y_{k+1} \\ &= \sum_{l=0}^{d'-2}\Lambda_l y_{k+d'-l} + \Lambda_{d'-1}y_{k+1} \end{aligned} \tag{4.41}$$

for $d' \geq 1$. Substituting (4.40) and (4.38) into (4.41) gives

$$\begin{aligned} \xi(q)y_k \;=\; & \sum_{l=0}^{d'-2}\Lambda_l\bar{C}\bar{A}^{d'-l-1}x_{k+1} - \sum_{l=0}^{d'-2}\Lambda_l(\sum_{j=1}^{d'-l-1}\bar{C}\bar{A}^{d'-l-j-1}A_*A_0^{-1}e_{k+j} + A_0^{-1}e_{k+d'-l}) \\ & + \sum_{l=0}^{d'-2}\Lambda_l(\sum_{j=1}^{d'-l-1}\bar{C}\bar{A}^{d'-l-j-1}\bar{B}u_{k+j}) + \Lambda_{d'-l}(\bar{C}x_{k+1} + A_0^{-1}e_{k+1}) \end{aligned} \tag{4.42}$$

for $k \geq 0$ and $d' \geq 1$. After some manipulations

$$\xi(q)y_k = \sum_{l=0}^{d'-1}\Lambda_l\bar{C}\bar{A}^{d'-l-1}x_{k+1} + \sum_{j=k+1}^{k+d'-1}\beta_{k-j}u_j + \sum_{j=0}^{d'-1}F_j e_{k+d'-j} \tag{4.43}$$

for $k \geq 0$, $d' \geq 1$. Now substituting (4.35) into (4.43) gives

$$\xi(q)y_k = L_k x_0 + \sum_{j=0}^{k}\alpha_{k-j}y_j + \sum_{j=0}^{k+d'-1}\beta_{k-j}u_j + \sum_{j=0}^{d'-1}F_j e_{k+d'-j} \tag{4.44}$$

where

$$L_k \;=\; \sum_{l=0}^{d'-1}\Lambda_l\bar{C}\bar{A}^{d'-l-1}E_*^{k+1} \tag{4.45}$$

$$\alpha_{k-j} = -\sum_{l=0}^{d'-1} \Lambda_l \bar{C} \bar{A}^{d'-l-1} E_*^{k-j} A_* \qquad j = 0, \ldots, k \tag{4.46}$$

$$\beta_{k-j} = \begin{cases} \sum_{l=0}^{d'-1} \Lambda_l \bar{C} \bar{A}^{d'-l-1} E_*^{k-j} B_* & j = 0, \ldots, k \\ \sum_{l=0}^{k+d'-1-j} \Lambda_l \bar{C} \bar{A}^{k+d'-l-j-1} B_* & j = k+1, \ldots, k+d'-1 \end{cases} \tag{4.47}$$

$$F_j = \begin{cases} \Lambda_0 A_0^{-1} & j = 0 \\ (\Lambda_j - \sum_{l=0}^{j-1} \Lambda_l \bar{C} \bar{A}^{j-l-1} A_*) A_0^{-1} & j = 1, \ldots, d'-1 \end{cases} \tag{4.48}$$

for $k \geq 0$ and $d' \geq 1$.

Now the coefficients of $\xi(q)$, $\{\Lambda_0, \ldots, \Lambda_{d'-1}\}$, can be chosen so that the coefficients of $u_{k+1}, \ldots, u_{k+d'-1}$ are zero and also the coefficient of u_k is nonsingular (in the SISO case the coefficient of u_k in predictor form is nonzero).

$$\beta_{-d'+1} = \cdots = \beta_{-1} = 0 \tag{4.49a}$$

$$\beta_0 = \text{nonsingular} \tag{4.49b}$$

where from (4.47)

$$\beta_j = \sum_{l=0}^{d'-1+j} \Lambda_l \bar{C} \bar{A}^{d'-l+j-1} B_* ; \qquad j = -d'+1, \ldots, -1 \tag{4.49c}$$

$$\beta_0 = \sum_{l=0}^{d'-1} \Lambda_l \bar{C} \bar{A}^{d'-l-1} B_* . \tag{4.49d}$$

After substituting (4.49) into (4.43), equation (4.30) follows immediately.

In the following lemma the predictor proposed in Lemma 4.2 will be expressed in an alternative form that will be used in the next sections.

Lemma 4.3 *Subject to assumptions 4.1 and 4.2, the filtered output $\xi(q)y_k$ can be written as follows:*

$$\xi(q)y_k = L_k x_0 + \sum_{j=0}^{k} \alpha_{k-j} y_j + \sum_{j=0}^{k} \beta_{k-j} u_j + \sum_{j=0}^{d'-1} F_j e_{k+d'-j} \tag{4.50}$$

for $k \geq 0$ and $d' \geq 1$.

Proof: Follows by substituting (4.49a) into (4.44).

Note that as expected the filtered output $\xi(q)y_k$ is a function of the system outputs and the system inputs up to and including time k as well as the future (unpredictable) noise $F(q)e_k$.

Now in the following lemma it is shown how the predictor model (4.50) proposed in this research can be related to the conventional predictor from in the literature.

Lemma 4.4 *The predictor model (4.30) in Lemma 4.3 can be expressed in a conventional polynomial form, discussed by Goodwin and Sin [28], as follows:*

$$\xi(q)y_k = \alpha(q^{-1})y_k + \beta(q^{-1})u_k + F(q)e_k \tag{4.51a}$$

where

$$\begin{aligned}
\alpha(q^{-1}) &= \alpha_0 + \alpha_1 q^{-1} + \cdots + \alpha_{v-1} q^{-(v-1)} && (4.51b)\\
\beta(q^{-1}) &= \beta_0 + \beta_1 q^{-1} + \cdots + \beta_{v-1} q^{-(v-1)} && (4.51c)\\
F(q) &= F_0 q^{d'} + \cdots + F_{d'-1} q && (4.51d)\\
\xi(q) &= \Lambda_0 q^{d'} + \cdots + \Lambda_{d'-1} q && (4.51e)\\
d' &= \textit{maximum advance in } \xi(q) && (4.51f)
\end{aligned}$$

and v in the ARX model (4.1) is the maximum of the degrees of the polynomials in $A(q^{-1})$ and $B(q^{-1})$, or $v = \max_{i=1,\ldots,p}\{n_i\}$ in the minimum canonical model (4.28).

Proof: Because of the special structure of matrix E_*

$$E_*^l = 0_{n\times n} \qquad \text{for} \quad l \geq v = \max_{i=1,\ldots,p}\{n_i\}\ , \tag{4.52}$$

see Ref. [3], where the integers n_i, $i = 1, \ldots, p$ are the observability indices of the system, as explained in Section 4.5.

Then for $k \geq v-1$ the effect of initial condition x_0 disappears and (4.50) can be simplified as follows:

$$\xi(q)y_k = \sum_{j=k-(v-1)}^{k} \alpha_{k-j}y_j + \sum_{j=k-(v-1)}^{k} \beta_{k-j}u_j + \sum_{j=0}^{d'-1} f_j e_{k+d-j} \tag{4.53}$$

or

$$\xi(q)y_k = \sum_{j=0}^{v-1} \alpha_j y_{k-j} + \sum_{j=0}^{v-1} \beta_j u_{k-j} + \sum_{j=0}^{d'-1} f_j e_{k+d-j}\,. \tag{4.54}$$

This establishes (4.51).

As seen the predictor model above (4.51) is a special case of the predictor (4.50) proposed in this research. In a zero initial condition case there is no difference between two models. However, the proposed predictor is capable of dealing with a nonzero initial condition whereas the conventional predictor model lacks such a capability.

4.7 Determination of the parameters of $\xi(q)$

In this section two procedures for constructing the interactor matrix are considered. Using an example it is shown that the two procedure reach the same results based on different assumptions.

Procedure 1: The procedure proposed by Furuta and Kamiyama (1977) [25] will be used here to show how the entries of the interactor matrix are related to the parameters of the canonical ARX model (4.1) or the minimal canonical state-space model (4.29). For simplicity, the notations used here are the same as used by Wolovich and Falb [54].

There exist unique integers μ_i, $i = 1, \ldots, p$, such that

$$\bar{C}_i \bar{A}^{\mu_i - 1} B_* = \tau_i \tag{4.55}$$

where $\bar{C}_i\bar{A}^{\mu_i-1}B_*$ denotes the ith row of β_0, the coefficient of u_k in the predictor form (4.50), and τ_i is both finite and nonzero.

The first row of $\xi(q)$ denoted by $\xi(q)_1$ is given by

$$\xi(q)_1 = [\ q^{\mu_1} \quad 0 \quad \cdots \quad 0\ ] \tag{4.56}$$

so that

$$\bar{C}_1\bar{A}^{\mu_1-1}B_* = \xi_1 = \tau_1\ . \tag{4.57}$$

If τ_2 is linearly independent of ξ_1, then set

$$\xi(q)_2 = [\ 0 \quad q^{\mu_2} \quad 0 \quad \cdots \quad 0\ ] \tag{4.58}$$

so that the coefficient of u_k in $\xi(q)_2 y_k$

$$\bar{C}_2\bar{A}^{\mu_2-1}B_* = \xi_2 = \tau_2\ . \tag{4.59}$$

On the other hand, if τ_2 and ξ_1 are linearly dependent so that $\tau_2 = \delta_1^1\xi_1$ with $\delta_1^1 \neq 0$, then let

$$\bar{\xi}^1(q)_2 = q^{\mu_2^1}\{[\ 0 \quad q^{\mu_2} \quad 0 \quad \cdots \quad 0\ ] - \delta_1^1\xi(q)_1\}\ , \tag{4.60}$$

where d_2^1 is the unique integer for which $-\delta_1^1\bar{C}_1\bar{A}^{\mu_2^1+\mu_1-1}B_* + \bar{C}_2\bar{A}^{\mu_2^1+\mu_2-1}B_* = \bar{\xi}_2^1$ is both finite and nonzero. If $\bar{\xi}_2^1$ is linearly independent of of ξ_1, then we set

$$\xi(q)_2 = \bar{\xi}^1(q)_2\ . \tag{4.61}$$

If $\bar{\xi}_2^1$ is not linearly independent of ξ_1, then $\bar{\xi}_2^1 = \delta_1^2\xi_1$ and we let

$$\bar{\xi}^2(q)_2 = q^{\mu_2^2}[\bar{\xi}^1(q)_2 - \delta_1^2\xi(q)_1]\ , \tag{4.62}$$

where μ_2^2 is the unique integer for which $-\delta_1^2\bar{C}_1\bar{A}^{\mu_2^2+\mu_1-1}B_* + \bar{C}_2\bar{A}^{\mu_2^2+\mu_2-1}B_* = \bar{\xi}_2^2$ is both finite and nonzero.

If $\tilde{\xi}_2^2$ and ξ_1 are linearly independent, we set $\xi(q)_2 = \tilde{\xi}^2(q)_2$, and if not, we repeat the procedure until linear independence is obtained. The remaining rows of $\xi(q)$ are defined recursively in an entirely analogous manner. Finally, we see that $K = \beta_0 =$ nonsingular since the ξ_i is linearly independent and also it was proved by Furuta and Kamiyama (1977) [25] that using this method $\beta_{-d'+1} = \cdots = \beta_{-1} = 0$.

The algorithm above proposed by Furuta and Kamiyama (1977) [25] is illustrated in the following example:

Example 4.2 (Continued):

Consider the minimal canonical state-space model (4.29). For simplicity, it is assumed that $a_{12}^{(0)} = 0$ but then it will be shown how the final results can be modified when $a_{12}^{(0)} \neq 0$.

The procedure proposed by Furuta and Kamiyama results in

$$\begin{aligned}
\xi(q)_1 &= \begin{bmatrix} q & 0 \end{bmatrix} \\
\tau_1^{*} &= \bar{C}_1 \bar{A}^{1-1} B_* = \bar{C}_1 B_* \\
&= \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} = \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} \end{bmatrix} = \xi_1 \neq 0
\end{aligned}$$

If we put

$$\xi(q)_2 = \begin{bmatrix} 0 & q \end{bmatrix} \tag{4.63}$$

then

$$\tau_2 = \bar{C}_2 \bar{A}^{1-1} B_* = \bar{C}_2 B_*$$

$$= \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} = \begin{bmatrix} b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix}$$

Note that if we know that ξ_1 and τ_2 are linearly dependent

$$\tau_2 = \delta_1^1 \xi_1 \tag{4.64}$$

with

$$\delta_1^1 = \begin{cases} \frac{b_{21}^{(1)}}{b_{11}^{(1)}} & \text{if } b_{11}^{(1)} \neq 0 \\ \\ \frac{b_{22}^{(1)}}{b_{12}^{(1)}} & \text{if } b_{12}^{(1)} \neq 0 \end{cases} \tag{4.65}$$

Let

$$\begin{aligned} \tilde{\xi}^1(q)_2 &= q\{[\, 0 \quad q \,] - \delta_1^1 \xi(q)_1\} \\ &= [\, -\delta_1^1 q^2 \quad q^2 \,] \end{aligned}$$

and

$$\begin{aligned} \tilde{\xi}_2^1 &= -\delta_1^1 \bar{C}_1 \bar{A}^{2-1} B_* + \bar{C}_2 \bar{A}^{2-1} B_* \\ &= -\delta_1^1 [\, 0 \quad 0 \quad 1 \quad 0 \,] \begin{bmatrix} 0 & 0 & -a_{11}^{(2)} & -a_{12}^{(2)} \\ 0 & 0 & -a_{21}^{(2)} & -a_{22}^{(2)} \\ 1 & 0 & -a_{11}^{(1)} & 0 \\ 0 & 1 & 0 & -a_{22}^{(1)} \end{bmatrix} \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} \\ &\quad + [\, 0 \quad 0 \quad 0 \quad 1 \,] \begin{bmatrix} 0 & 0 & -a_{11}^{(2)} & -a_{12}^{(2)} \\ 0 & 0 & -a_{21}^{(2)} & -a_{22}^{(2)} \\ 1 & 0 & -a_{11}^{(1)} & 0 \\ 0 & 1 & 0 & -a_{22}^{(1)} \end{bmatrix} \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} \end{aligned}$$

$$= -\delta_1^1[\ b_{11}^{(2)} - a_{11}^{(1)}b_{11}^{(1)} \quad b_{12}^{(2)} - a_{11}^{(1)}b_{12}^{(1)}\ ] + [\ b_{21}^{(2)} - a_{22}^{(1)}b_{21}^{(1)} \quad b_{22}^{(2)} - a_{22}^{(1)}b_{22}^{(1)}\ ]$$
$$= [\ -\delta_1^1(b_{11}^{(2)} - a_{11}^{(1)}b_{11}^{(1)}) + (b_{21}^{(2)} - a_{22}^{(1)}b_{21}^{(1)}) \quad -\delta_1^1(b_{12}^{(2)} - a_{11}^{(1)}b_{12}^{(1)}) + (b_{22}^{(2)} - a_{22}^{(1)}b_{22}^{(1)})\ ]$$

If we know that $\tilde{\xi}_2^1$ and ξ_1 are not linearly independent, then $\tilde{\xi}_2^1 = \delta_1^2\xi_1$ with

$$\delta_1^2 = \begin{cases} \frac{1}{b_{11}^{(1)}}[-\frac{b_{21}^{(1)}}{b_{11}^{(1)}}(b_{11}^{(2)} - a_{11}^{(1)}b_{11}^{(1)}) + (b_{21}^{(2)} - a_{22}^{(1)}b_{21}^{(1)})] & \text{if } b_{11}^{(1)} \neq 0 \\ \\ \frac{1}{b_{12}^{(1)}}[-\frac{b_{22}^{(1)}}{b_{12}^{(1)}}(b_{12}^{(2)} - a_{11}^{(1)}b_{12}^{(1)}) + (b_{22}^{(2)} - a_{22}^{(1)}b_{22}^{(1)})] & \text{if } b_{21}^{(1)} \neq 0 \end{cases} \tag{4.66}$$

Since $\tilde{\xi}_2^1$ depends linearly on ξ_1, we continue by defining

$$\begin{aligned} \tilde{\xi}^2(q)_2 &= q\{[\ -\delta_1^1q^2 \quad q^2\ ] - \delta_1^2\xi(q)_1\} \\ &= [\ -\delta_1^1q^3 - \delta_1^2q^2 \quad q^3\ ] \end{aligned}$$

for which

$$\tilde{\xi}_2^2 = -\delta_1^1\bar{C}_1\bar{A}^{3-1}B_{*} - \delta_1^2\bar{C}_1\bar{A}^{2-1}B_{*} + \bar{C}_2\bar{A}^{3-1}B_{*}\,. \tag{4.67}$$

If we know $\tilde{\xi}_2^2$ and ξ_1 are linearly independent the procedure must terminate. Hence

$$\xi(q) = \begin{bmatrix} q & 0 \\ -\delta_1^1q^3 - \delta_1^2q^2 & q^3 \end{bmatrix}. \tag{4.68}$$

For instance, if $b_{11}^{(1)} \neq 0$ then

$$\xi(q) = \begin{bmatrix} q & 0 \\ -\frac{b_{21}^{(1)}}{b_{11}^{(1)}}q^3 - \frac{1}{b_{11}^{(1)}}[-\frac{b_{21}^{(1)}}{b_{11}^{(1)}}(b_{11}^{(2)} - a_{11}^{(1)}b_{11}^{(1)}) + (b_{21}^{(2)} - a_{22}^{(1)}b_{21}^{(1)})]q^2 & q^3 \end{bmatrix}. \tag{4.69}$$

If the true values of the parameters in the model (4.25) are substituted into (4.69)

$$\begin{aligned} \xi(q) &= \begin{bmatrix} q & 0 \\ -\frac{1.0}{1.0}q^3 - \frac{1}{1.0}[-\frac{1.0}{1.0}(2.0 - 3.0 \times 1.0) + (4.0 - 7.0 \times 1.0)]q^2 & q^3 \end{bmatrix} \\ &= \begin{bmatrix} q & 0 \\ -q^3 + 2q^2 & q^3 \end{bmatrix} \end{aligned}$$

which this is the same result as Goodwin obtained [?]. However, if $a_{12}^{(0)} \neq 0$, using the same procedure as explained in above, the coefficients δ_1^1 and δ_1^2 are modified as follows:

$$\delta_1^1 = \begin{cases} \frac{b_{21}^{(1)}}{b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)}} & \text{if } b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)} \neq 0 \\ \\ \frac{b_{22}^{(1)}}{b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)}} & \text{if } b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)} \neq 0 \end{cases} \tag{4.70}$$

and

$$\delta_1^2 = \begin{cases} \frac{1}{b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)}} M & \text{if } b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)} \neq 0 \\ \\ \frac{1}{b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)}} N & \text{if } b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)} \neq 0 \end{cases} \tag{4.71}$$

where

$$\begin{aligned} M &= -\delta_1^1[b_{11}^{(2)} - a_{11}^{(1)} b_{11}^{(1)} + b_{21}^{(1)} a_{11}^{(1)} a_{12}^{(0)} - a_{12}^{(0)}(b_{21}^{(2)} - a_{22}^{(1)} b_{21}^{(1)})] + (b_{21}^{(2)} - a_{22}^{(1)} b_{21}^{(1)}) \\ N &= -\delta_1^1[b_{12}^{(2)} - a_{11}^{(1)} b_{12}^{(1)} + b_{22}^{(1)} a_{11}^{(1)} a_{12}^{(0)} - a_{12}^{(0)}(b_{22}^{(2)} - a_{22}^{(1)} b_{22}^{(1)})] + (b_{22}^{(2)} - a_{22}^{(1)} b_{22}^{(1)})\,. \end{aligned}$$

Procedure 2: The procedure above, Procedure 1, was based on the algorithm described in Furuta and Kamiyama (1977) [25] and involves nontrivial computations including tests for linear dependence. Whereas the procedure proposed in this research, Procedure 2, is based on the result obtained by Das [17] and a priori knowledge of the degrees of the diagonal entries of the interactor matrix and involves solution of a set of linear algebraic equations.

In the proposed procedure, the first step is to determine the structure of the interactor matrix $\xi(q)$. Because of the special structure of $\xi(q)$, $\xi_{ij}(q) = 0$ for $i < j$ and $\xi_{ij}(q) = q^{d_i}$ for $i = j$ and $i, j = 1, \ldots, p$. The major difficulty is to determine the polynomials of the nondiagonal entries of $\xi(q)$. In the following lemma it is shown how the polynomials of the nondiagonal entries of $\xi(q)$ are determined.

Lemma 4.5 *The polynomials of the nondiagonal entries of $\xi(q)$ are given by*

$$\xi_{ij}(q) = \begin{cases} \xi_{0,ij}q^{w_{ij}} + \cdots + \xi_{w_{ij}-d_j-1,ij}q^{d_j+1} & \text{if} \quad w_{ij} \geq d_j + 1 \\ 0 & \text{if} \quad w_{ij} < d_j + 1 \end{cases} \tag{4.72}$$

where

$$w_{ij} = \partial(\xi_{ij}(q)) \leq \sum_{k=j}^{i} d_k - (i-j) \tag{4.73}$$

$$\text{for } i > j \qquad i, j = 1, \ldots, p \, .$$

Proof: From Corollary 4.1 proved by Das [17], the upper bounds on the degrees of the entries of $\xi_{ij}(q)$ are given by (4.73). Also from Lemma 4.1

$$\xi_{ij}(q) = h_{ij}(q)q^{d_j} \tag{4.74}$$

$$\text{for } i > j \qquad i, j = 1, \ldots, p \, .$$

Since $h_{ij}(q)$ is divisible by q (or is zero), the least powers of q in the polynomials of $\xi_{ij}(q)$ are $d_j + 1$ if $w_{ij} \geq d_j + 1$; otherwise, the polynomials of $\xi_{ij}(q)$ are equal to zero if $w_{ij} < d_j + 1$ for $i > j$ and $i, j = 1, \ldots, p$. This establishes (4.72).

For adaptive control Das [17] showed that the inequality sign in (4.73) can be replaced by an equality as follows:

$$w_{ij} = \partial(\xi_{ij}(q)) = \sum_{k=j}^{i} d_k - (i-j) \tag{4.75}$$

$$\text{for } i > j \qquad i, j = 1, \ldots, p \, .$$

Now using (4.75) and Lemma 4.5, the structure of $\xi(q)$ can be specified. In other words, the structure of the the matrices Λ_l in (4.41),

$$\xi(q) = \Lambda_0 q^{d'} + \Lambda_1 q^{d'-1} + \cdots + \Lambda_{d'-1} q = \sum_{l=0}^{d'-1} \Lambda_l q^{d'-l} \tag{4.76}$$

where d', the maximum advance in $\xi(q)$, is obtained as follows:

$$d' = \max_{i,j} \partial \xi_{ij}(q) \tag{4.77}$$

$$\text{for } i \geq j \qquad i,j = 1,\ldots,p\,.$$

Finally, the unknown matrices Λ_l are determined by solving the matrix equations (4.49a),

$$\beta_{-d'+1} = \cdots = \beta_{-1} = 0\,, \tag{4.78}$$

where

$$\beta_j = \sum_{l=0}^{d'-1+j} \Lambda_l \bar{C} \bar{A}^{d'-l+j-1} B_* \,; \qquad j = -d'+1,\ldots,-1\,. \tag{4.79}$$

Finally, the following example is presented to illustrate Procedure 2.

Example 4.3 (Continued):

Consider again the system of Example 4.1 but with the priori knowledge of the degrees of the diagonal entries of $\xi(q)$. Assume that the degrees of the diagonal entries of the interactor matrix are 1 and 3, i.e. ,$d_1 = 1$ and $d_2 = 3$.

From (4.75)

$$w_{21} = \partial(\xi_{21}(q)) = \sum_{k=1}^{2} d_k - (2-1) = 3\,. \tag{4.80}$$

Using Lemma 4.5

$$\xi_{21}(q) = \xi_{0,21} q^3 + \xi_{1,21} q^2\,. \tag{4.81}$$

Then the interactor matrix has the following structure:

$$\xi(q) = \begin{bmatrix} q & 0 \\ \Lambda_{21}^{(0)} q^3 + \Lambda_{21}^{(1)} q^2 & q^3 \end{bmatrix} \tag{4.82}$$

where $\Lambda_{21}^{(0)} = \xi_{0,21}$ and $\Lambda_{21}^{(1)} = \xi_{1,21}$. Hence, the coefficient matrices of $\xi(q)$ are

$$\Lambda_0 = \begin{bmatrix} 0 & 0 \\ \Lambda_{21}^{(0)} & 1 \end{bmatrix}, \quad \Lambda_1 = \begin{bmatrix} 0 & 0 \\ \Lambda_{21}^{(1)} & 0 \end{bmatrix}, \quad \Lambda_2 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}. \tag{4.83}$$

Using the equations (4.49), we have

$$\Lambda_0 \bar{C}\bar{A}B_* + \Lambda_1 \bar{C}B_* = 0 \quad (4.84a)$$

$$\Lambda_0 \bar{C}B_* = 0\,. \quad (4.84b)$$

From (4.24) and (4.27c)-(4.27f),

$$\bar{A} = E_* - A_* A_0^{-1} E_0$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} \\ a_{21}^{(2)} & a_{22}^{(2)} \\ a_{11}^{(1)} & 0 \\ 0 & a_{22}^{(1)} \end{bmatrix} \begin{bmatrix} 1 & a_{11}^{(0)} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -a_{11}^{(2)} & a_{11}^{(2)}a_{11}^{(0)} - a_{12}^{(2)} \\ 0 & 0 & -a_{21}^{(2)} & a_{21}^{(2)}a_{11}^{(0)} - a_{22}^{(2)} \\ 1 & 0 & -a_{11}^{(1)} & a_{11}^{(2)}a_{11}^{(0)} \\ 0 & 1 & 0 & -a_{22}^{(1)} \end{bmatrix} \quad (4.85a)$$

$$B_* = \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{12}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} \quad (4.85b)$$

$$\bar{C} = A_0^{-1}E_0 = \begin{bmatrix} 1 & a_{11}^{(0)} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -a_{11}^{(0)} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (4.85c)$$

Substituting (4.83) and (4.85) into (4.84b) gives

$$\begin{bmatrix} 0 & 0 \\ \Lambda_{21}^{(0)} & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & -a_{11}^{(0)} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{12}^{(2)} & b_{22}^{(2)} \\ b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix} = 0 \quad (4.86)$$

which $\Lambda_{21}^{(0)}$ is obtained as follows:

$$\Lambda_{21}^{(0)} = \begin{cases} \frac{-b_{21}^{(1)}}{b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)}} & \text{if } b_{11}^{(1)} - a_{12}^{(0)} b_{21}^{(1)} \neq 0 \\ \\ \frac{-b_{22}^{(1)}}{b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)}} & \text{if } b_{12}^{(1)} - a_{12}^{(0)} b_{22}^{(1)} \neq 0 \end{cases} \tag{4.87}$$

The comparison of (4.82) and (4.68) shows that $\Lambda_{21}^{(0)} = -\delta_1^1$ and $\Lambda_{21}^{(1)} = -\delta_1^2$. The result obtained in (4.87) verifies that $\Lambda_{21}^{(0)} = -\delta_1^1$. Then it can be verified that $\Lambda_{21}^{(1)}$ is equal to $-\delta_1^2$ by solving the equation (4.84a).

The two procedure above reach the same results but based on different assumptions. Also, note that Examples 4.2 and 4.3 show how the entries of the interactor matrix are related to the entries of the canonical ARX model. This relation can simplify the problem of adaptive control of multivariable stochastic systems. This can be interpreted as a technique for eliminating the need for knowing the noninteger valued variables in the interactor matrix.

4.8 General optimal solution

Here the optimization of the control criterion $J_{k+d'|k}$ given in (4.15) is discussed. Using the same approach as used in the SISO case in Chapter 3, it will be proved that the optimal solution can be expressed analytically. The result is summarized in the following theorem.

Theorem 4.1 *For the system (4.1) having predictor form (4.50), the control law minimizing the criterion*

$$J_{k+d'|k} = E\{\|\xi(q)(y_k - y_k^*)\|^2 | \mathcal{F}_k\} \tag{4.88}$$

is given by

$$E\{\beta_0^T \xi(q)(y_k - y_k^*) | \mathcal{F}_k\} = 0 \tag{4.89}$$

for $k \geq 0$ *and* $d' \geq 1$.

Note that it is necessary to assume that the system is minimum-phase at every instant of time. The control signal may otherwise be unbounded.

Proof: Using (4.50), the term $\xi(q)y_k - \xi(q)y_k^*$ can be written as follows:

$$\xi(q)y_k - \xi(q)y_k^* = L_k x_0 + \sum_{j=0}^{k} \alpha_{k-j} y_j + \sum_{j=0}^{k} \beta_{k-j} u_j + \sum_{j=0}^{d'-1} F_j e_{k+d'-j} - \sum_{l=0}^{d'-1} \Lambda_l y_k^* \tag{4.90}$$

for $k \geq 0$ and $d' \geq 1$. In order to show which term is a function of u_k, the equation above (4.90) is rearranged as follows:

$$\xi(q)y_k - \xi(q)y_k^* = \beta_0 u_k + \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) + \eta(\theta, \mathcal{E}_{k+1}^{k+d'}) - \mu(\theta, \mathcal{Y}^{*}{}_{k+1}^{k+d'}) \tag{4.91}$$

where $\gamma(.)$, $\eta(.)$, and $\mu(.)$ are nonlinear functions and

$$\mathcal{Y}_0^k = \{y_k, \ldots, y_0\} \tag{4.92}$$

$$\mathcal{U}_0^{k-1} = \{u_{k-1}, \ldots, u_0\} \tag{4.93}$$

$$\mathcal{E}_{k+1}^{k+d'} = \{e_{k+d'}, \ldots, e_{k+1}\} \tag{4.94}$$

$$\mathcal{Y}^{*}{}_{k+1}^{k+d'} = \{y_{k+d'}^*, \ldots, y_{k+1}^*\} \,. \tag{4.95}$$

For simplicity, (4.91) is written in the following compact form:

$$\xi(q)y_k - \xi(q)y_k^* = \beta_0 u_k + \Delta_k \tag{4.96}$$

where

$$\Delta_k = \gamma(\theta, x_0, \mathcal{Y}_0^k, \mathcal{U}_0^{k-1}) + \eta(\theta, \mathcal{E}_{k+1}^{k+d'}) - \mu(\theta, \mathcal{Y}^{*}{}_{k+1}^{k+d'}) \tag{4.97}$$

and Δ_k is independent of u_k.

Substituting (4.96) into (4.88) gives

$$J_{k+d'|k} = E\{\|\beta_0 u_k + \Delta_k\|^2 | \mathcal{F}_k\} \quad (4.98)$$

$$= E\{\mathrm{tr}\,[\,(\beta_0 u_k + \Delta_k)(\beta_0 u_k + \Delta_k)^T\,]\,|\,\mathcal{F}_k\} \quad (4.99)$$

or

$$\begin{aligned} J_{k+d'|k} &= E\{\mathrm{tr}(\beta_0 u_k u_k^T \beta_0^T + \beta_0 u_k \Delta_k^T + \Delta_k u_k^T \beta_0^T + \Delta_k \Delta_k^T) | \mathcal{F}_k\} \\ &= E\{\mathrm{tr}(\beta_0 u_k u_k^T \beta_0^T) + \mathrm{tr}(\beta_0 u_k \Delta_k^T) + \mathrm{tr}(\Delta_k u_k^T \beta_0^T) + \mathrm{tr}(\Delta_k \Delta_k^T) | \mathcal{F}_k\}\ . \end{aligned} \quad (4.100)$$

Using one of the properties of the trace operator, we have

$$\mathrm{tr}(\beta_0 u_k u_k^T \beta_0^T) = \mathrm{tr}(u_k^T \beta_0^T \beta_0 u_k) \quad (4.101a)$$

$$\mathrm{tr}(\beta_0 u_k \Delta_k^T) = \mathrm{tr}(u_k \Delta_k^T \beta_0) \quad (4.101b)$$

$$\mathrm{tr}(\Delta_k u_k^T \beta_0^T) = \mathrm{tr}(\beta_0^T \Delta_k u_k^T)\ . \quad (4.101c)$$

Substituting (4.101) into (4.100) yields

$$J_{k+d'|k} = E\{\mathrm{tr}(u_k^T \beta_0^T \beta_0 u_k) + \mathrm{tr}(u_k \Delta_k^T \beta_0) + \mathrm{tr}(\beta_0^T \Delta_k u_k^T) + \mathrm{tr}(\Delta_k \Delta_k^T) | \mathcal{F}_k\}\ . \quad (4.102)$$

The expectation and trace operations can be interchanged, and as u_k is constrained to be $\mathcal{F}_k$ measurable, this becomes

$$\begin{aligned} J_{k+d'|k} &= \mathrm{tr}(u_k^T E\{\beta_0^T \beta_0 | \mathcal{F}_k\} u_k) + \mathrm{tr}(u_k E\{\Delta_k^T \beta_0 | \mathcal{F}_k\}) + \mathrm{tr}(E\{\beta_0^T \Delta_k | \mathcal{F}_k\} u_k^T) \\ &\quad + \mathrm{tr}(E\{\Delta_k \Delta_k^T | \mathcal{F}_k\})\ . \end{aligned} \quad (4.103)$$

As in the SISO case, all the conditional expectations in (4.103) are independent of u_k, and minimization of the cost function (4.88) can be achieved by setting the partial derivative to zero:

$$\begin{aligned} \frac{\partial J_{k+d'|k}}{\partial u_k} &= 2E\{\beta_0^T \beta_0 | \mathcal{F}_k\} u_k + (E\{\Delta_k^T \beta_0 | \mathcal{F}_k\})^T + E\{\beta_0^T \Delta_k | \mathcal{F}_k\} = 0 \\ &= 2E\{\beta_0^T \beta_0 | \mathcal{F}_k\} u_k + 2E\{\beta_0^T \Delta_k | \mathcal{F}_k\} = 0 \end{aligned} \quad (4.104)$$

or

$$u_k = -(E\{\beta_0^T\beta_0 | \mathcal{F}_k\})^{-1}E\{\beta_0^T\Delta_k | \mathcal{F}_k\} . \tag{4.105}$$

In order to obtain (4.89), (4.104) is rewritten as follows:

$$\begin{aligned} E\{\beta_0^T\beta_0 u_k | \mathcal{F}_k\} + E\{\beta_0^T\Delta_k | \mathcal{F}_k\} &= 0 \\ E\{\beta_0^T(\beta_0 u_k + \Delta_k) | \mathcal{F}_k\} &= 0 . \end{aligned} \tag{4.106}$$

Using (4.96)

$$E\{\beta_0^T\xi(q)(y_k - y_k^*) | \mathcal{F}_k\} = 0 . \tag{4.107}$$

This establishes (4.89).

4.9 Some important special cases

The general optimal solution (4.107) is one of the main contributions in this research. This solution can be applied to any system with or without uncertainties in the model. In other words, (4.107) can be applied in both adaptive and nonadaptive cases. In different cases, the only problem is to determine the conditional expectation in (4.107). Now, it will be shown how the general solution presented above can be applied to the following special cases:

1. *Nonadaptive case:* If the parameters θ, the initial condition x_0, and the interactor matrix $\xi(q)$ are known and also β_0 is nonsingular, (4.107) can be simplified as follows:

$$E\{\xi(q)y_k | \mathcal{F}_k, x_0, \theta\} = \xi(q)y_k^* . \tag{4.108}$$

   In words, the input $\{u_k\}$ is obtained by equating the optimal prediction $E\{\xi(q)y_k | \mathcal{F}_k, x_0, \theta\}$ with the filtered desired outputs $\xi(q)y_k^*$.

2. If β_0 is known and nonsingular, and as well the interactor matrix is known, then

$$E\{\xi(q)y_k|\,\mathcal{F}_k,\beta_0\} = \xi(q)y_k^*. \tag{4.109}$$

3. *MIMO cautious control*: If the interactor matrix is assumed known, equation (4.107) is simplified as follows:

$$E\{\beta_0^T\xi(q)y_k|\,\mathcal{F}_k\} = E\{\beta_0^T|\,\mathcal{F}_k\}\xi(q)y_k^* \tag{4.110}$$

and the control law which satisfies the equation above is called the MIMO cautious control.

4. *Certainty equivalence adaptive control:* Using the certainty equivalence approximation, as shown in Chapter 2, (4.107) can be approximated as follows:

$$E\{\beta_0^T\xi(q)(y_k - y_k^*)|\,\mathcal{F}_k,x_0,\theta = \hat{\theta}_{k|k}\} = 0 \tag{4.111}$$

or

$$E\{\beta_0^T|\,\mathcal{F}_k,x_0,\theta = \hat{\theta}_{k|k}\}E\{\xi(q)(y_k - y_k^*)|\,\mathcal{F}_k,x_0,\theta = \hat{\theta}_{k|k}\} = 0\,. \tag{4.112}$$

Now if $E\{\beta_0^T|\,\mathcal{F}_k,x_0,\theta = \hat{\theta}_{k|k}\}$ is nonsingular,

$$E\{\xi(q)(y_k - y_k^*)|\,\mathcal{F}_k,x_0,\theta = \hat{\theta}_{k|k}\} = 0\,. \tag{4.113}$$

The equation above (4.113) shows that an approximation of the optimal control law can be obtained by solving the control problem in the case of known parameters and substituting the known parameters with their estimates. The controller can be interpreted as a certainty-equivalence controller.

Hence, there would be two different cases with known $\xi(q)$ and unknown $\xi(q)$ as follows:

- *With known* $\xi(q)$*:* If the interactor matrix $\xi(q)$ is known, then (4.113) is simplified as follows:

$$E\{\xi(q)y_k | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = \xi(q)y_k^* . \tag{4.114}$$

- *With unknown* $\xi(q)$*:* If $\xi(q)$ is unknown, then (4.113) is written as follows:

$$E\{\xi(q)y_k | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = E\{\xi(q)y_k^* | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} . \tag{4.115}$$

4.10 Direct adaptive control with unknown $\xi(q)$

In this section it will be explained why the conventional adaptive control with unknown $\xi(q)$ in the literature, which is based on a direct scheme, cannot be applied for a stochastic case. However, the indirect approach proposed in this research, which will be explained in the section, is capable of dealing with stochastic case.

In the following algorithm proposed by Dugard, Goodwin and deSouza (1983) [19] the real variables appearing in the interactor matrix are estimated along with the other parameters appearing in the system model in a deterministic case. This approach follows a suggestion originally made by Johansson 1982 [30].

Consider the predictor model (4.51) in a deterministic case, $e_k = 0$,

$$\xi(q)y_k = \alpha(q^{-1})y_k + \beta(q^{-1})u_k . \tag{4.116}$$

Using (4.8)

$$H(q)D(q)y_k = \alpha(q^{-1})y_k + \beta(q^{-1})u_k \tag{4.117}$$

where $D(q) = \text{diag}[q^{d_1} \cdots q^{d_p}]$. Using the structure of $H(q)$, Lemma 4.1,

$$D(q)y_k = L(q)D(q)y_k + \alpha(q^{-1})y_k + \beta(q^{-1})u_k \tag{4.118}$$

where

$$L(q) = [I - H(q)] = \begin{bmatrix} 0 & & & \\ -h_{21}(q) & 0 & & \\ \vdots & & \ddots & \\ -h_{p1}(q) & \cdots & -h_{pp-1}(q) & 0 \end{bmatrix} . \tag{4.119}$$

Equation (4.118) can be written in regression form as:

$$i\text{th element of } [\,D(q)y_k] = y_{k+d_i}^{(i)} = \phi_{k+l_i}^{(i)T}\theta^{(i)} \ ; \quad i = 1, \dots, p . \tag{4.120}$$

In (4.120), $\phi_{k+l_i}^{(i)T}$ is a vector of values of the system outputs up to time $(k + l_i)$ and of the system inputs up to time k where l_i is the maximum forward shift in the ith row of $L(q)D(q)$. $\theta^{(i)}$ is a vector of coefficients in the ith row of $L(q)$, $\alpha(q^{-1})$, and $\beta(q^{-1})$. (Note $d_i > l_i$; $i = 1, \cdots, p$.)

The estimated vectors $\{\hat{\theta}_k^{(i)}\}$, obtained by the least squares algorithm, are now regrouped to form

$$\hat{\alpha}(k, q^{-1}),\ \hat{\beta}(k, q^{-1}),\ \text{and}\ \hat{\xi}(k, q^{-1}) = (I - \hat{L}(k, q^{-1}))D(q) . \tag{4.121}$$

Finally, using (4.115), the feedback control signal u_k based on the certainty equivalence principle is generated by solving

$$\hat{\beta}(k, q^{-1})u_k + \hat{\alpha}(k, q^{-1})y_k = \hat{\xi}(k, q^{-1})y_k^* . \tag{4.122}$$

As mentioned before, the adaptive algorithm above with unknown $\xi(q)$ is a direct scheme since the design is based on estimation of the parameters of the predictor model. The difficulty is that this algorithm cannot be applied for a stochastic case. To show this disadvantage, once again consider the predictor model (4.51) including noise which can be written as follows:

$$D(q)y_k = L(q)y_k + \alpha(q^{-1})y_k + \beta(q^{-1})u_k + F(q)e_k . \tag{4.123}$$

Then (4.123) can be written in the following stochastic regression form

$$i\text{th element of } [\,D(q)y_k] = y^{(i)}_{k+d_i} = \phi^{(i)^T}_{k+l_i}\theta^{(i)} + \nu^{(i)}_{k+d_i} \ ; \quad i = 1, \ldots, p \tag{4.124}$$

where $\nu^{(i)}_{k+d_i}$ is a colored noise with unknown statistical characteristics. The important point that has to be stressed in this method is that the parameter estimator will be biased because of the correlation between the noise and the regression vectors. This is the reason why the algorithm above is not used in a stochastic case.

However, it will be shown how the difficulties in the direct algorithm above can be resolved by using the proposed indirect scheme.

4.11 Indirect adaptive control with unknown $\xi(q)$

In this section an indirect adaptive control with unknown $\xi(q)$ is developed. It is shown the proposed approach is capable of dealing with stochastic cases.

In the following discussion, it is assumed that the degrees of the diagonal entries of the interactor matrix are known. In order words, the structure of the interactor matrix is known based on the results obtained in Section 4.7, in which it was shown how the entries of the interactor matrix are related to the entries of the canonical ARX model. As shown in Examples 4.2 and 4.3, the evaluation of the interactor matrix requires real divisions. Hence, we will not be able to evaluate the conditional expectation in Equation (4.107). However, there are some approximation methods to solve the problem. The most common approximation method, i.e. the certainty equivalence principle, is employed, Equation (4.115).

The indirect approach involves first estimating the parameters and states in a standard model (e.g., a one-step-ahead predictor), such as the canonical ARX model (4.1) or the canonical SP model (4.21) and then evaluating the control law by on-line calculations.

Because of the initial condition x_0, there could be two different cases for the estimation process:

- The initial condition x_0 is known or zero,
- There is uncertainty in x_0, or x_0 is unknown.

Both cases will be considered below.

Case 1 : If the initial condition x_0 is known or zero, it is not required to estimate the states of the system. Hence, the canonical ARX model (4.1) can be written in the following form:

$$y_k = -(A_0 - I)y_k - A_1 y_{k-1} - \cdots - A_v y_{k-v} + B_1 u_{k-1} + \cdots + B_v u_{k-v} + e_k. \quad (4.125)$$

Then (4.125) is written in regression form as

$$y_k = \phi_k^T \theta + e_k \quad (4.126)$$

where θ is a column vector consisting of nonidentically-zero and nonidentically-unit entries in $[A_0 - I, A_1, \ldots, A_v, B_1, \ldots, B_v]$ and ϕ_k is constructed according to the known order of the parameters in θ.

Then using (4.126), the following dynamical system is obtained to estimate the parameters:

$$\theta_{k+1} = \theta_k \quad (4.127a)$$

$$y_k = \phi_k^T \theta_k + e_k\,. \quad (4.127b)$$

Hence, the unknown parameters will be chosen from the Gaussian family and their estimate $\hat{\theta}_{k|k}$ and associated error covariance $P_{k|k}$ are sufficient statistics. The parameter vector estimate $\hat{\theta}_{k|k}$ and the associated covariance $P_{k|k}$ are obtained from a Kalman filter according to

$$\hat{\theta}_{k|k} = \hat{\theta}_{k|k-1} + K_k(y_k - \phi_k^T \hat{\theta}_{k|k-1}) \tag{4.128a}$$

$$\hat{\theta}_{k+1|k} = \hat{\theta}_{k|k} \tag{4.128b}$$

$$K_k = P_{k|k-1}\phi_k^T(\phi_k P_{k|k-1}\phi_k^T + R_k)^{-1} \tag{4.128c}$$

$$P_{k|k} = P_{k|k-1} - K_k \phi_k P_{k|k-1} \tag{4.128d}$$

$$P_{k+1|k} = P_{k|k} \,. \tag{4.128e}$$

Finally, the input u_k is generated by solving (4.115),

$$E\{\xi(q)y_k | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = E\{\xi(q)y_k^* | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} \,. \tag{4.129}$$

Case 2: If the initial condition x_0 is unknown, the canonical ARX model (4.1) or the canonical SP model (4.21) can be put into the following equivalent form for the simultaneous estimation of the state variable x_k and the system parameters θ:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k \tag{4.130a}$$

$$y_k = E_0 x_k - (A_0 - I)y_k + e_k \,. \tag{4.130b}$$

Then it is not difficult to see that the model above (4.130) can be written as follows:

$$x_{k+1} = E_* x_k + G_* \theta \tag{4.131a}$$

$$y_k = E_0 x_k + G_0 \theta + e_k \tag{4.131b}$$

where G_* and G_0 are matrices constructed from the measurements y_k and u_k according to the ordering of the system parameters $a_{ik}^{(j)}$ and $b_{ik}^{(j)}$ in θ, such that

$$G_* \theta = [-A_* \; B_*] \begin{bmatrix} y_k \\ u_k \end{bmatrix}, \quad \text{and} \quad G_0 \theta = [-(A_0 - I) \;\; 0] \begin{bmatrix} y_k \\ u_k \end{bmatrix} \,. \tag{4.132}$$

As the system is assumed to be time-invariant, then $\theta = \theta_k = \theta_{k+1}$ and eqn (4.131) can be written as follows:

$$s_{k+1} = F_k s_k \tag{4.133a}$$

$$y_k = H_k s_k + e_k \tag{4.133b}$$

where

$$s_k = [x_k^T \; \theta_k^T]^T, \qquad H_k = [E_0 \; G_0]$$

$$F_k = \begin{bmatrix} E_* & G_* \\ 0 & I \end{bmatrix}.$$

The simultaneous state and parameter estimation algorithm is therefore given by the following recursive equations [45]:

$$\hat{s}_{k|k} = \hat{s}_{k|k-1} + K_k(y_k - H_k \hat{s}_{k|k-1}) \tag{4.134a}$$

$$\hat{s}_{k+1|k} = F_k \hat{s}_{k|k} \tag{4.134b}$$

$$K_k = P_{k|k-1} H_k^T (H_k P_{k|k-1} H_k^T + R_k)^{-1} \tag{4.134c}$$

$$P_{k|k} = P_{k|k-1} - K_k H_k P_{k|k-1} \tag{4.134d}$$

$$P_{k+1|k} = F_k P_{k|k} F_k^T \tag{4.134e}$$

with initial conditions

$$\hat{s}_{0|-1} = E\{s_0 | \mathcal{F}_{-1}\}, \quad P_{0|-1} = E\{(s_0 - \hat{s}_{0|-1})(s_0 - \hat{s}_{0|-1})^T | \mathcal{F}_{-1}\}. \tag{4.135}$$

Finally, the input u_k is generated by solving (4.115),

$$E\{\xi(q) y_k | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = E\{\xi(q) y_k^* | \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\}. \tag{4.136}$$

4.12 Simulation results

In this section, an example of a two-input, two-output system with fifteen unknown parameters will be presented to illustrate the performance of the algorithm developed above. Also, the performance of this algorithm will be compared to that of the minimum variance control with known parameters.

Consider the system given in the previous examples, Examples 4.1-4.3. The model can be expressed in the following form:

$$y_k = -(A_0 - I)y_k - A_1 y_{k-1} - A_2 y_{k-2} + B_1 u_{k-1} + B_2 u_{k-2} + e_k \qquad (4.137)$$

or

$$\begin{bmatrix} y_k^{(1)} \\ y_k^{(2)} \end{bmatrix} = -\begin{bmatrix} 0 & a_{11}^{(0)} \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_k^{(1)} \\ y_k^{(2)} \end{bmatrix} - \begin{bmatrix} a_{11}^{(1)} & 0 \\ 0 & a_{22}^{(1)} \end{bmatrix}\begin{bmatrix} y_{k-1}^{(1)} \\ y_{k-1}^{(2)} \end{bmatrix} - \begin{bmatrix} a_{11}^{(2)} & a_{12}^{(2)} \\ a_{21}^{(2)} & a_{22}^{(2)} \end{bmatrix}\begin{bmatrix} y_{k-2}^{(1)} \\ y_{k-2}^{(2)} \end{bmatrix}$$
$$+ \begin{bmatrix} b_{11}^{(1)} & b_{12}^{(1)} \\ b_{21}^{(1)} & b_{22}^{(1)} \end{bmatrix}\begin{bmatrix} u_{k-1}^{(1)} \\ u_{k-1}^{(2)} \end{bmatrix} + \begin{bmatrix} b_{11}^{(2)} & b_{12}^{(2)} \\ b_{21}^{(2)} & b_{22}^{(2)} \end{bmatrix}\begin{bmatrix} u_{k-2}^{(1)} \\ u_{k-2}^{(2)} \end{bmatrix} + \begin{bmatrix} e_k^{(1)} \\ e_k^{(2)} \end{bmatrix}$$

Let the parameter vector θ be

$$\theta = [a_{11}^{(0)}\ a_{11}^{(1)}\ a_{22}^{(1)}\ a_{11}^{(2)}\ a_{12}^{(2)}\ a_{21}^{(2)}\ a_{22}^{(2)}\ b_{11}^{(1)}\ b_{12}^{(1)}\ b_{21}^{(1)}\ b_{22}^{(1)}\ b_{11}^{(2)}\ b_{12}^{(2)}\ b_{21}^{(2)}\ b_{22}^{(2)}]^T$$

which contains 15 unknown parameters. The exact value of θ is

$$\theta = [0, 3, 7, 2, 0, 0, 12, 1, 1, 1, 1, 2, 1, 4, 3]^T$$

As seen, the minimum phase system above has one pole on the unit circle and three unstable poles at -2, -3, and -4 and four zeros at infinity.

It is assumed that the unknown parameter vector θ, generated by a random generator, is a gaussian random vector with a priori statistics having mean and covariance

$$\hat{\theta}_{0|-1} = [0.0577, 3.1165, 7.0627, 2.0075, 0.0352, -0.00697, 12.1696,$$

$$
\begin{aligned}
& 1.0059, 1.1797, 1.0264, 1.0872, 1.8554, 0.9299, 4.1246, 2.9361]^T \\
P_{0|-1} &= 0.01 * I .
\end{aligned}
$$

The initial state is assumed to be known:

$$x_0 = 0 .$$

The $\{e_k\}$ is a white zero-mean process with known covariance

$$R = \begin{bmatrix} 10^{-6} & 0 \\ 0 & 10^{-6} \end{bmatrix} .$$

The signal to noise ratio is rather high to avoid running the program many times to obtain the expected values of estimation as well as the input and output signals.

The desired output sequences $\{y_k^{*(1)}\}$ and $\{y_k^{*(2)}\}$ are unit square waves of 20 and 30 samples, respectively.

The performance of the proposed algorithm (4.128) to (4.129), the indirect adaptive control with unknown $\xi(q)$, is shown in Figs 4.1 to 4.3. Fig. 4.1 exhibits the general response of the system. The steady-state responses and the transient responses are demonstrated in Figs 4.2 and 4.3, respectively.

The algorithm converges and perfect tracking is ultimately achieved. It is evident from Fig. 4.3 that the first output $y_k^{(1)}$ converges faster than the second output $y_k^{(2)}$. Because the interactor matrix is nondiagonal, an unconditional minimum variance is achieved for $y_k^{(1)}$ but for $y_k^{(2)}$ the optimization is achieved subject to constraints.

The conclusion is that the control law does not simultaneously minimize the individual output variance when the interactor matrix is nondiagonal.

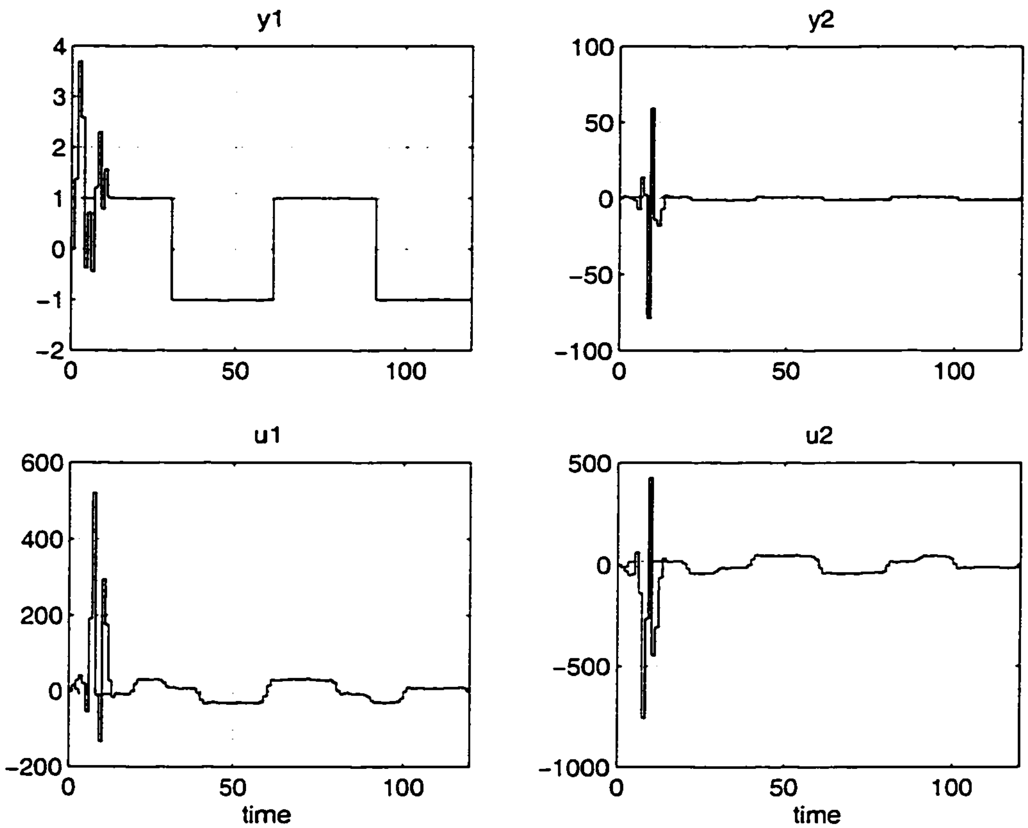


Figure 4.1: General response

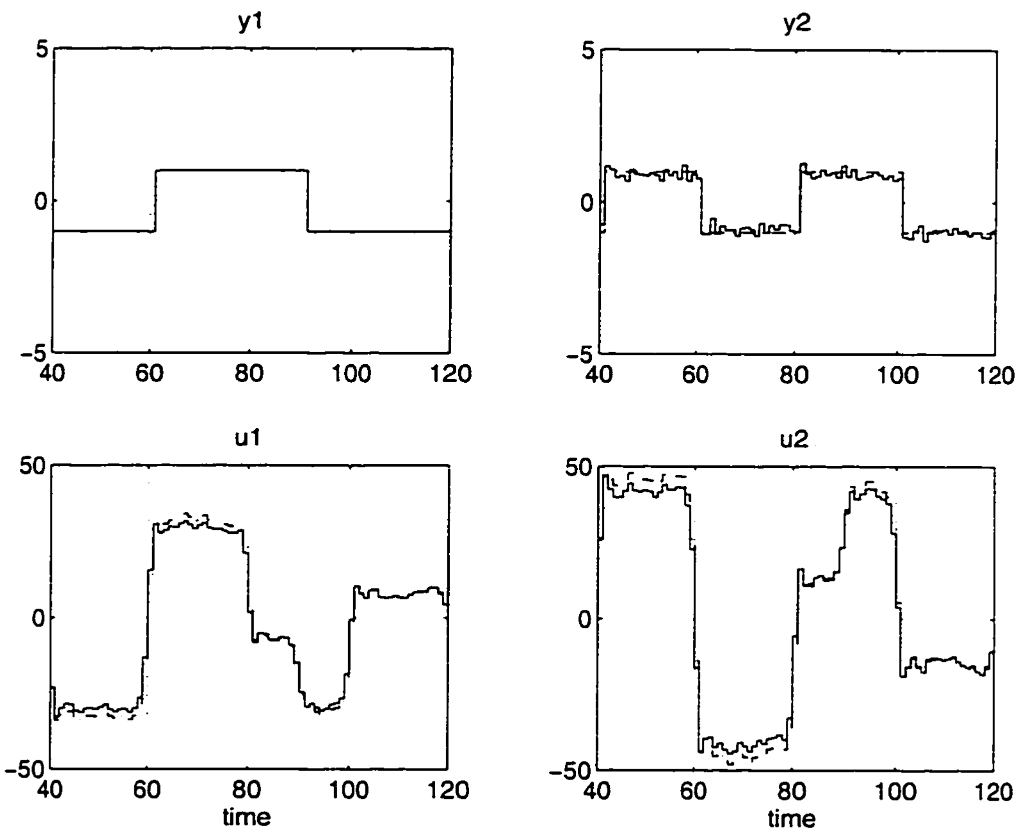


Figure 4.2: Steady-state response

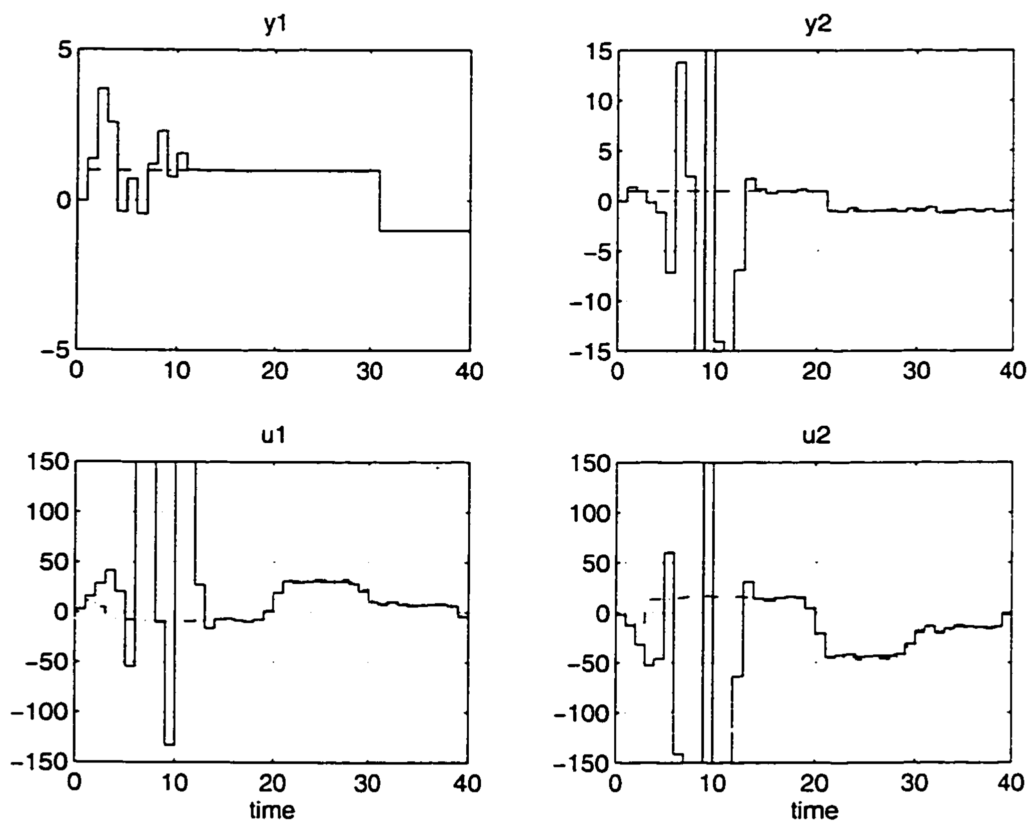


Figure 4.3: Transient response

Chapter 5

Conclusion and further research directions

5.1 Overview

In this thesis, a number of open problems in SISO adaptive prediction and cautious control as well as in MIMO adaptive control have been solved.

In Chapter 2, a problem in SISO adaptive prediction related to transient performance was discussed. Using an indirect adaptive scheme, it was shown in a general case how the optimal d-step-ahead prediction of y_k at time k in a minimum variance sense, $E\{y_{k+d}|\mathcal{F}_k\}$, can be determined. First using state variables, the ARX model is transformed into a state space form which allows to us to express the model into a predictor form. Then using the estimated parameters and states provided by a Kalman filter, the optimal d-step-ahead predictor is constructed. It was shown that the optimal prediction is obtained based on both the plant parameter-state estimates and estimates of their uncertainty. As a result, the conventional adaptive predictors discussed in the literature are asymptotically optimal whereas the proposed predictor is optimal in a minimum variance sense at each step resulting in good transient performance as well as asymptotic performance.

In Chapter 3, SISO cautious control was solved for the general delay-white noise case. Using the theory of stochastic optimal control, it was proved that the value of u_k which minimizes the following criterion:

$$J_{k+d|k} = E\{(y_{k+d} - y^{*}_{k+d})^2 | \mathcal{F}_k\} \tag{5.1}$$

satisfies the following equation:

$$E\{\beta_0(y_{k+d} - y^{*}_{k+d}) | \mathcal{F}_k\} = 0 \ . \tag{5.2}$$

Then using an indirect approach, the same method as used in Chapter 2, the equation above was solved to determine the control law u_k. Note that the solution does not depend on the model used and the approach applied, direct or indirect, provided that the conditional mean of the noise seen by the predictor given $\mathcal{F}_k$ is zero.

It was shown that the proposed method and the method used by Åström and Wittenmark [8] are the same except that the proposed method can handle cases with uncertainty in the initial conditions. Wieslander and Wittenmark [51] used a very complicated way to solve the adaptive problem for a unit delay system. Nahorski and Vidal in 1974 [40] handle the case with general delay by using a special model and employing a direct approach. In order to verify the generality of the solution above, it was shown that if the solution above is applied to the model used by Nahorski and Vidal, their result is obtained.

As expected, simulations indicate that the proposed adaptive controller has a superior performance compared with the certainty equivalence controller.

In Chapter 4, it was shown how a MIMO system having a minimal canonical model can be transformed to a predictor form. Then it was shown conventional predictor in the literature, discussed by Goodwin and Sin in [28], can be considered as a special case of the predictor proposed in this research. Hence using the same method as applied to the SISO case, it was proved that the one-step-ahead control law minimizing a criterion of the form

$$J_{k+d'|k} = E\{\|\xi(q)(y_k - y_k^*)\|^2 | \mathcal{F}_k\} \tag{5.3}$$

is given by

$$E\{\beta_0^T \xi(q)(y_k - y_k^*) | \mathcal{F}_k\} = 0 \ . \tag{5.4}$$

As in the SISO case, it was shown how the general optimal solution above can be applied to some special cases discussed in the literature. Also the solution does not depend on the model used and the approach applied, direct or indirect, provided that the conditional mean of the noise seen by the predictor given $\mathcal{F}_k$ is zero.

Furthermore, it was shown that if the interactor matrix is assumed known, equation (5.4) is simplified as follows:

$$E\{\beta_0^T \xi(q) y_k | \mathcal{F}_k\} = E\{\beta_0^T | \mathcal{F}_k\} \xi(q) y_k^* \tag{5.5}$$

and the control law which satisfies the equation above is called MIMO cautious control. However, requiring knowledge of the interactor matrix is a severe practical limitation. In general, this matrix contains noninteger-valued real variables and one can argue that knowledge of these variables is tantamount to requiring knowledge of the complete system transfer function. Therefore there has been strong motivation in the literature to investigate ways in which the requirement of knowing the system interactor matrix might be removed.

It was shown that if $\xi(q)$ is assumed unknown, an exact solution for equation (5.4) cannot be obtained. Hence, in this research the most popular approximation method, i.e. the certainty equivalence principle, was employed based on the result obtained by Das [17] and a priori knowledge of the degrees of the diagonal entries of the interactor matrix. As a result, equation (5.4) is simplified as follows:

$$E\{\xi(q)y_k \mid \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} = E\{\xi(q)y_k^* \mid \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\} \tag{5.6}$$

if $E\{\beta_0^T \mid \mathcal{F}_k, x_0, \theta = \hat{\theta}_{k|k}\}$ is nonsingular.

It was shown that the conventional approaches discussed in the literature cannot be applied to a stochastic case. However, the indirect approach proposed in this research is capable of dealing with stochastic cases. At the end of Chapter 4, simulation results show the performance of the proposed algorithm compared with that of the minimum variance controller with known parameters.

5.2 Suggestions for further studies

The results of this thesis can serve as the basis for the following further research.

- *Convergence and stability analysis*: Convergence and stability analysis of the algorithms proposed in this research were not given here. It would be of theoretical and practical importance to establish conditions that will guarantee the convergence and the stability of these algorithms. It is believed that the results in Middelton *et. al.* [35] and Kumar [33] can be applied to these algorithms.

- *Nonminimum phase and nonsquare systems*: In Chapter 4 it was assumed that the system is minimum phase and the number of inputs is equal to the number of outputs. As discussed in the literature [28], using different criteria from that was used in this research, the adaptive control problem for nonminimum phase and nonsquare systems have been tackled in deterministic cases. It could be an interesting research problem to solve these optimal problems using the approach proposed in this research for stochastic cases.

- *Suboptimal dual control*: Most adaptive controllers are based on a separation between the estimation of unknown parameters and state variables from the determination of the control signal. The methods discussed so far have not considered the interaction between identification and control. This means that the control laws have not been designed to facilitate the identification. The control strategies have been designed in order to minimize some control error. sometimes under the assumptions that some parameters are uncertain. The learning procedure has not been active but accidental. In a dual controller there is, however. an interaction between identification and control in the sense that the controller must compromise between a control action and a probing action, [23].

  The ultimate purpose is to minimize the following multistep-ahead criteria:

$$J = \frac{1}{N} E\{ \sum_{t=k}^{k+N-1} (y_{t+d} - y^*_{t+d})^2 | \mathcal{F}_k \} \tag{5.7}$$

  for the SISO case, and

$$J = \frac{1}{N} E\{ \sum_{t=k}^{k+N-1} \|\xi(q)(y_t - y^*_t)\|^2 | \mathcal{F}_k \} \tag{5.8}$$

  for the MIMO case, with respect to $u_k, \ldots, u_{k+N-1}$.

  The formal solution of the dual control problem has been known for a long time, [23, 1]. The solution leads, however, to a functional equation which in most cases is difficult to solve.

  There are many ways to solve the dual control problem in an approximate way. In 1989, Mookerjee and Bar-Shalom, [38], proposed a suboptimal solution for a MIMO system with $\xi(q) = qI$. Now it seems to be feasible to combine the approach proposed by Mookerjee and Bar-Shalom and the approach used in this research to find a suboptimal solution for both the SISO case with the general delay-white noise and the MIMO case with a general form for $\xi(q)$.

Appendix A

Optimal Filtering

This appendix is concerned with estimating the state variables and parameters of discrete-time, linear stochastic, SISO and MIMO time-varying systems. This topic has been discussed in detail by Salut [45]. Nevertheless, using a different method from that given by Salut, it is shown how the simultaneous estimation of state variables and system parameters with known noise statistics can be solved as an optimal linear filtering problem.

A.1 SISO case

The input-output characteristics of a general stochastic linear system can be described by an ARMAX model of the form

$$A(q^{-1})y_k = B(q^{-1})u_k + C(q^{-1})e_k \tag{A.1}$$

where $\{e_k\}$ is a zero mean nonstationary white gaussian process with covariance given by

$$E\{e_k e_l^T\} = R_k \delta(k-l) \tag{A.2}$$

for all k and l, where $\delta(k-l)$ is the Kronecker delta, which is 1 for $k = l$ and 0 otherwise. The polynomials $A(q^{-1})$, $B(q^{-1})$, and $C(q^{-1})$ are defined as

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n} \tag{A.3}$$

$$B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_n q^{-n} \tag{A.4}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_n q^{-n}. \tag{A.5}$$

In order to facilitate the estimation process the model above can be rewritten in the following way:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k + C_* e_k \tag{A.6a}$$

$$y_k = E_0 x_k + B_0 u_k + e_k \tag{A.6b}$$

where

$$E_* = \begin{bmatrix} 0 & \cdots & 0 & 0 \\ 1 & & & \vdots \\ & \ddots & & 0 \\ & & 1 & 0 \end{bmatrix}, \; A_* = \begin{bmatrix} a_n \\ \vdots \\ a_2 \\ a_1 \end{bmatrix}, \; B_* = \begin{bmatrix} b_n \\ \vdots \\ b_2 \\ b_1 \end{bmatrix}, \; C_* = \begin{bmatrix} c_n \\ \vdots \\ c_2 \\ c_1 \end{bmatrix} \tag{A.7}$$

$$E_0 = [\, 0 \;\cdots\; 0 \;\; 1 \,].$$

If θ is defined as follows:

$$\theta^T = [A_*^T \; B_*^T \; b_0] \tag{A.8}$$

then the model (A.6) can be rearranged in the following form:

$$x_{k+1} = E_* x_k + G_* \theta + C_* e_k \tag{A.9a}$$

$$y_k = E_0 x_k + G_0 \theta + e_k \tag{A.9b}$$

where the matrix G_* and G_0 are constructed according to the order of the parameters in θ. The identification problem is more realistic if θ is subject to random perturbations, so we suppose that

$$\theta_{k+1} = \theta_k + \pi_k \tag{A.10}$$

where π_k is an independent, zero mean, gaussian white noise possibly correlated with e_k. By appending the parameter vector θ_{k+1} to the state vector x_{k+1} and using the model (A.9) , we obtain the following model:

$$\begin{bmatrix} x_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} E_* & G_* \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} C_* e_k \\ \pi_k \end{bmatrix} \tag{A.11a}$$

$$y_k = \begin{bmatrix} E_0 & G_0 \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + e_k \tag{A.11b}$$

or

$$s_{k+1} = F_k s_k + w_k \tag{A.12a}$$

$$y_k = H_k s_k + e_k \tag{A.12b}$$

where $s_k = \begin{bmatrix} x_k \\ \theta_k \end{bmatrix}$. It is assumed that both noises $\{w_k\}$ and $\{e_k\}$ are independent, zero mean, gaussian white processes with known covariances

$$E\left\{ \begin{bmatrix} \begin{bmatrix} C_* e_k \\ \pi_k \end{bmatrix} \\ e_k \end{bmatrix} \begin{bmatrix} \begin{bmatrix} C_* e_l \\ \pi_l \end{bmatrix} \\ e_l \end{bmatrix}^T \right\} = \begin{bmatrix} Q_k & S_k \\ S_k^T & R_k \end{bmatrix} \delta(k-l). \tag{A.13}$$

Now we can redefine the one-step prediction problem to be one of requiring computation of the sequence $E\{s_{k+1}|y_0, \ldots, y_k\}$ for $k = 0, 1, 2, \ldots$. We shall denote this quantity by $\hat{s}_{k+1|k}$ and shall use the symbol $\mathcal{F}_k$ to denote the set $\{y_0, \ldots, y_k\}$. Here, we use the fact that $\{u_k\}$ is known. So far, we have not specified an initial condition for the difference equation (A.12). Under normal circumstances, one might expect to be told that at the initial time $k = 0$, the state vector s_0 was some prescribed vector. Here, however, we prefer to leave our options more open. From a practical point of view, if it is impossible to measure s_k exactly for arbitrary k, it is unlikely that s_0 will be available. This leads to the adoption of a random initial condition for the system. In particular, assume that the initial state vector s_0 is a gaussian random variable with mean $\hat{s}_0$ and covariance P_0, i.e.,

$$E[s_0] = \hat{s}_0 \qquad E\{(s_0 - \hat{s}_0)(s_0 - \hat{s}_0)^T\} = P_0 . \tag{A.14}$$

Further, assume that s_0 is independent of $\{w_k\}$ and $\{e_k\}$. Now we wish to determine the estimates

$$\hat{s}_{k+1|k} = E\{s_{k+1} | \mathcal{F}_k\} \tag{A.15}$$

and the associated error covariance matrix $P_{k+1|k}$, where

$$P_{k+1|k} = E\{(s_{k+1} - \hat{s}_{k+1})(s_{k+1} - \hat{s}_{k+1})^T | \mathcal{F}_k\}. \tag{A.16}$$

The following lemmas will be frequently used in this appendix.

Lemma A.1 *If X is conditionally gaussian given Z with mean $\hat{X}$ and covariance P, and $A(\cdot)$ and $b(\cdot)$ are measurable functions with $\|A(Z)\| < +\infty$ and $\|b(Z)\| < +\infty$, then $A(Z)X + b(Z)$ is also conditionally gaussian given Z with conditional mean $A(Z)\hat{X} + b(Z)$ and conditional covariance $A(Z)PA(Z)^T$.*

Proof: See Ref. [15].

Lemma A.2 *Suppose that $\begin{bmatrix} X \\ Y \end{bmatrix}$ is conditionally gaussian given Z with conditional covariance $\begin{bmatrix} \Sigma_{xx|z} & \Sigma_{xy|z} \\ \Sigma_{xy|z}^T & \Sigma_{yy|z} \end{bmatrix}$. Then, given $\{Y, Z\}$, X is conditionally gaussian with conditional mean*

$$E\{X | Y, Z\} = E\{X | Z\} + \Sigma_{xy|z}\Sigma_{yy|z}^{-1}(Y - E\{Y | Z\}), \tag{A.17}$$

and conditional covariance

$$\Sigma_{xx|y,z} = \Sigma_{xx|z} - \Sigma_{xy|z}\Sigma_{yy|z}^{-1}\Sigma_{xy|z}^{T} . \tag{A.18}$$

Proof: See Ref. [15].

Note that s_0 and $\begin{bmatrix} w_0 \\ e_0 \end{bmatrix}$ are jointly gaussian as a result of their being individually gaussian and independent. Therefore the joint distribution of $\begin{bmatrix} s_0 \\ w_0 \\ e_0 \end{bmatrix}$ is gaussian with mean and covariance

$$\begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_0 & S_0 \\ 0 & S_0^T & R_0 \end{bmatrix} . \tag{A.19}$$

In order to prove that $\begin{bmatrix} s_0 \\ w_0 \\ e_0 \\ y_0 \end{bmatrix}$ is gaussian, use (A.12b) and $s_0 = s_0$, $w_0 = w_0$, and $e_0 = e_0$ to form the following matrix equation:

$$\begin{bmatrix} s_0 \\ w_0 \\ e_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_0 & 0 & I \end{bmatrix} \begin{bmatrix} s_0 \\ w_0 \\ e_0 \end{bmatrix} . \tag{A.20}$$

Now since linear transformations of gaussian random variables preserve their gaussian character, it follows that $\begin{bmatrix} s_0 \\ w_0 \\ e_0 \\ y_0 \end{bmatrix}$ is gaussian with mean and covariance

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_0 & 0 & I \end{bmatrix} \begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_0 & 0 & I \end{bmatrix} \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_0 & S_0 \\ 0 & S_0^T & R_0 \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_0 & 0 & I \end{bmatrix}^T \tag{A.21}$$

or, equivalently, with mean and covariance

$$\begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \\ H_0\hat{s}_0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_0 & 0 & 0 & P_0H_0^T \\ 0 & Q_0 & S_0 & S_0 \\ 0 & S_0^T & R_0 & R_0 \\ H_0P_0 & S_0^T & R_0 & H_0P_0H_0^T + R_0 \end{bmatrix}. \tag{A.22}$$

Hence $\begin{bmatrix} s_0 \\ w_0 \\ e_0 \end{bmatrix}$ is conditionally gaussian given y_0 with conditional mean

$$m_{0|0} = \begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} P_0H_0^T \\ S_0 \\ R_0 \end{bmatrix} (H_0P_0H_0^T + R_0)^{-1}(y_0 - H_0\hat{s}_0) \tag{A.23}$$

and conditional covariance

$$\Sigma_{0|0} = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & Q_0 & S_0 \\ 0 & S_0^T & R_0 \end{bmatrix} - \begin{bmatrix} P_0H_0^T \\ S_0 \\ R_0 \end{bmatrix} (H_0P_0H_0^T + R_0)^{-1} \begin{bmatrix} H_0P_0 & S_0^T & R_0 \end{bmatrix}. \tag{A.24}$$

Then proving that the conditional distribution for $\begin{bmatrix} s_0 \\ w_0 \\ e_0 \end{bmatrix}$ given y_0 is gaussian with the mean and covariance above, we determine the conditional distribution for s_1 given y_0. From (A.12a)

$$s_1 = [F_0 \quad I \quad 0] \begin{bmatrix} s_0 \\ w_0 \\ e_0 \end{bmatrix} . \tag{A.25}$$

Hence using the result for transformations of gaussian random variables, the distribution of s_1 given y_0 is gaussian with mean

$$\begin{aligned} \hat{s}_{1|0} &= E\{s_1 \,|\, y_0\} = [F_0 \;\; I \;\; 0] m_{0|0} \\ &= F_0 \hat{s}_0 + K_0(y_0 - H_0 \hat{s}_0) \end{aligned} \tag{A.26}$$

and covariance

$$\begin{aligned} P_{1|0} &= E\{(s_1 - \hat{s}_{1|0})(s_1 - \hat{s}_{1|0})^T \,|\, y_0\} \\ &= [F_0 \; I \; 0] \Sigma_{0|0} [F_0 \; I \; 0]^T \\ &= F_0 P_0 F_0^T + Q_0 - K_0(H_0 P_0 H_0^T + R_0) K_0^T \end{aligned} \tag{A.27}$$

where

$$K_0 = (F_0 P_0 H_0^T + S_0)(H_0 P_0 H_0^T + R_0)^{-1} . \tag{A.28}$$

It is also straightforward to obtain the conditional distribution for s_2 given $\{y_1, y_0\}$. Let us proceed formally, in the same manner as used in deriving the conditional distribution for s_1 given y_0. Since s_1 given y_0 and $\begin{bmatrix} w_1 \\ e_1 \end{bmatrix}$ are gaussian and independent we can deduce that $\begin{bmatrix} s_1 \\ w_1 \\ e_1 \end{bmatrix}$ is conditionally gaussian given y_0. From (A.12b),

$s_1 = s_1$, $w_1 = w_1$, and $e_1 = e_1$, the following matrix equation is obtained,

$$\begin{bmatrix} s_1 \\ w_1 \\ e_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_1 & 0 & I \end{bmatrix} \begin{bmatrix} s_1 \\ w_1 \\ e_1 \end{bmatrix} . \tag{A.29}$$

The linear transformation above proves that the conditional joint distribution of $\begin{bmatrix} s_1 \\ w_1 \\ e_1 \\ y_1 \end{bmatrix}$ given y_0 is gaussian with mean and covariance

$$\begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_1 & 0 & I \end{bmatrix} \begin{bmatrix} \hat{s}_{1|0} \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_1 & 0 & I \end{bmatrix} \begin{bmatrix} P_{1|0} & 0 & 0 \\ 0 & Q_1 & S_1 \\ 0 & S_1^T & R_1 \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \\ H_1 & 0 & I \end{bmatrix}^T \tag{A.30}$$

or, equivalently, with mean and covariance

$$\begin{bmatrix} \hat{s}_{1|0} \\ 0 \\ 0 \\ H_1 \hat{s}_{1|0} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} P_{1|0} & 0 & 0 & P_{1|0} H_1^T \\ 0 & Q_1 & S_1 & S_1 \\ 0 & S_1^T & R_1 & R_1 \\ H_1 P_{1|0} & S_1^T & R_1 & H_1 P_{1|0} H_1^T + R_1 \end{bmatrix} . \tag{A.31}$$

Now the vector $\begin{bmatrix} s_1 \\ w_1 \\ e_1 \end{bmatrix}$ conditioned on y_0 and y_1 has mean

$$m_{1|1} = \begin{bmatrix} \hat{s}_{1|0} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} P_{1|0} H_1^T \\ S_1 \\ R_1 \end{bmatrix} (H_1 P_{1|0} H_1^T + R_1)^{-1} (y_1 - H_1 \hat{s}_{1|0}) \tag{A.32}$$

and covariance

$$\Sigma_{1|1} = \begin{bmatrix} P_{1|0} & 0 & 0 \\ 0 & Q_1 & S_1 \\ 0 & S_1^T & R_1 \end{bmatrix} - \begin{bmatrix} P_{1|0}H_1^T \\ S_1 \\ R_1 \end{bmatrix} (H_1 P_{1|0} H_1^T + R_1)^{-1} \begin{bmatrix} H_1 P_{1|0} & S_1^T & R_1 \end{bmatrix} . \tag{A.33}$$

From (A.12a), we have

$$s_2 = [F_1 \;\; I \;\; 0] \begin{bmatrix} s_1 \\ w_1 \\ e_1 \end{bmatrix} \tag{A.34}$$

By the linear transformation principle, s_2 given y_1 and y_0 is gaussian with conditional mean

$$\begin{aligned} \hat{s}_{2|1} = E\{s_2 | y_1, y_0\} &= [F_1 \; I \; 0] m_{1|1} \\ &= F_1 \hat{s}_{1|0} + K_1 (y_1 - H_1 \hat{s}_{1|0}) \end{aligned} \tag{A.35}$$

and conditional covariance

$$\begin{aligned} P_{2|1} &= E\{(s_2 - \hat{s}_{2|1})(s_2 - \hat{s}_{2|1})^T | y_1, y_0\} \\ &= [F_1 \; I \; 0] \Sigma_{1|1} [F_1 \; I \; 0]^T \\ &= F_1 P_{1|0} F_1^T + Q_1 - K_1 (H_1 P_{1|0} H_1^T + R_1) K_1^T \end{aligned} \tag{A.36}$$

where

$$K_1 = (F_1 P_{1|0} H_1^T + S_1)(H_1 P_{1|0} H_1^T + R_1)^{-1} . \tag{A.37}$$

More generally, repetition of the steps above yields that the conditional distribution for s_{k+1} given $y_k, \ldots, y_1, y_0$ is gaussian with mean $\hat{s}_{k+1|k}$ and covariance $P_{k+1|k}$ as follows:

$$\begin{aligned} \hat{s}_{k+1|k} &= E\{s_{k+1} | y_k, \ldots, y_1, y_0\} \\ &= F_k \hat{s}_{k|k-1} + K_k (y_k - H_k \hat{s}_{k|k-1}) \end{aligned} \tag{A.38}$$

and

$$\begin{aligned} P_{k+1|k} &= E\{(s_{k+1} - \hat{s}_{K+1|k})(s_{k+1} - \hat{s}_{K+1|k})^T | y_k, \ldots, y_1, y_0\} \\ &= F_k P_{k|k-1} F_k^T + Q_k - K_k (H_k P_{k|k-1} H_k^T + R_k) K_k^T \end{aligned} \tag{A.39}$$

in which K_k is the filter gain given by

$$K_k = (F_k P_{k|k-1} H_k^T + S_k)(H_k P_{k|k-1} H_k^T + R_k)^{-1} . \tag{A.40}$$

A.2 MIMO case

In this section it is assumed that a linear stochastic MIMO system can be described by an ARMAX model of the form

$$A(q^{-1}) y_k = B(q^{-1}) u_k + C(q^{-1}) e_k \tag{A.41}$$

where the polynomial matrices $A(q^{-1})$, $B(q^{-1})$, and $C(q^{-1})$ are defined as follows:

$$A(q^{-1}) = A_0 + A_1 q^{-1} + \ldots + A_r q^{-v} \tag{A.42}$$

$$B(q^{-1}) = B_0 + B_1 q^{-1} + \ldots + B_r q^{-v} \tag{A.43}$$

$$C(q^{-1}) = C_0 + C_1 q^{-1} + \ldots + C_r q^{-v} \tag{A.44}$$

and y_k, u_k, and e_k denote the $p \times 1$ output vector, $m \times 1$ input vector , and $p \times 1$ noise vector, respectively. Thus $A(q^{-1})$, $B(q^{-1})$, and $C(q^{-1})$ are $p \times p$, $p \times m$, and $p \times p$ polynomial matrices. Equation (A.41) is also called a left matrix fraction description of the system. For the estimation process, the model above can be written as follows:

$$x_{k+1} = E_* x_k - A_* y_k + B_* u_k + C_* e_k \tag{A.45a}$$

$$y_k = E_0 x_k + (I - A_0) y_k + B_0 u_k + C_0 e_k \tag{A.45b}$$

It must be emphasized that in the canonical form under consideration. A_0 is upper-right or lower-left unit triangular. We will take advantage of the triangular form in the MIMO case, as to be explained. The model (A.45) in canonical form can simply be used to estimate the states and parameters of the system by extending the state vector x_k with parameter vector $\theta = \theta_k$:

$$s_k = \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} . \tag{A.46}$$

Then the following augmented state model can be immediately obtained:

$$\begin{bmatrix} x_{k+1} \\ \theta_{k+1} \end{bmatrix} = \begin{bmatrix} E_* & G_*(y_k\,, u_k) \\ 0 & I \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + \begin{bmatrix} C_* e_k \\ \pi_k \end{bmatrix} \tag{A.47a}$$

$$y_k = \begin{bmatrix} E_0 & G_0(y_k\,, u_k) \end{bmatrix} \begin{bmatrix} x_k \\ \theta_k \end{bmatrix} + C_0 e_k\,. \tag{A.47b}$$

It is immediately noted, looking at equations (A.47), that in this case all unknown parameters or state variables appear linearly multiplied by either external variables that appear in the record, or by matrices that are only composed of zeros and ones. This very important property enables us to construct a linear estimator for all unknown variables and parameters of the system. The composite model above can be written in compact form as

$$s_{k+1} = F_k s_k + w_k \tag{A.48a}$$

$$y_k = H_k s_k + \nu_k \tag{A.48b}$$

Let us call $H_k^{(i)}$ for $i = 1, \ldots, p$, the rows of the matrix H_k in (A.48); they appear in the following form:

$$H_k^{(i)} = H_k^{(i)}(y_k^{(i-1)}, \ldots, y_k^{(1)}, u_k) \tag{A.49}$$

Rewrite equation (A.48b) as

$$y_k^{(1)} = H_k^{(1)} s_k + \nu_k^{(1)} \tag{A.50}$$

$$y_k^{(2)} = H_k^{(2)} s_k + \nu_k^{(2)} \tag{A.51}$$

$$\vdots = \vdots \tag{A.52}$$

$$y_k^{(p)} = H_k^{(p)} s_k + \nu_k^{(p)} . \tag{A.53}$$

In this section, the aim is to derive the Kalman filter equations. for the one-step prediction state estimate and associated error covariance. In order to simplify the problem, assume that the system (A.48) has just two outputs as follows:

$$s_{k+1} = F_k s_k + w_k \tag{A.54a}$$

$$y_k^{(1)} = H_k^{(1)} s_k + \nu_k^{(1)} \tag{A.54b}$$

$$y_k^{(2)} = H_k^{(2)} s_k + \nu_k^{(2)} . \tag{A.54c}$$

Here, $\{u_k\}$ is a known input sequence, s_0 has mean $\hat{s}_0$ and covariance P_0 ,$\{w_k\}$ and $\{\nu_k\}$ are zero mean. independent gaussian processes with covariance

$$E\left\{ \begin{bmatrix} w_k \\ \nu_k^{(1)} \\ \nu_k^{(2)} \end{bmatrix} \begin{bmatrix} w_l \\ \nu_l^{(1)} \\ \nu_l^{(2)} \end{bmatrix}^T \right\} = \begin{bmatrix} Q_k & S_k^{(1)} & S_k^{(2)} \\ S_k^{(1)^T} & R_k^{(1)} & R_k^{(1,2)} \\ S_k^{(2)^T} & R_k^{(1,2)^T} & R_k^{(2)} \end{bmatrix} \delta(k-l) . \tag{A.55}$$

According to the initial assumptions, the random variable $[s_0^T \; w_0^T \; \nu_0^{(1)^T} \; \nu_0^{(2)^T}]^T$ is gaussian with mean and covariance

$$\begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \Sigma_0 = \begin{bmatrix} P_0 & 0 & 0 & 0 \\ 0 & Q_0 & S_0^{(1)} & S_0^{(2)} \\ 0 & S_0^{(1)^T} & R_0^{(1)} & R_0^{(1,2)} \\ 0 & S_0^{(2)^T} & R_0^{(1,2)^T} & R_0^{(2)} \end{bmatrix} . \tag{A.56}$$

From (A.54b)

$$\begin{bmatrix} s_0 \\ w_0 \\ \nu_0^{(1)} \\ \nu_0^{(2)} \\ y_0^{(1)} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 & 0 \\ 0 & I & 0 & 0 \\ 0 & 0 & I & 0 \\ 0 & 0 & 0 & I \\ H_0^{(1)} & 0 & I & 0 \end{bmatrix} \begin{bmatrix} s_0 \\ w_0 \\ \nu_0^{(1)} \\ \nu_0^{(2)} \end{bmatrix} . \tag{A.57}$$

Hence using Lemma A.1, the joint distribution of $[s_0^T \;\; w_0^T \;\; \nu_0^{(1)^T} \;\; \nu_0^{(2)^T} \;\; y_0^{(1)^T}]^T$ is gaussian with mean and covariance

$$\begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \\ 0 \\ H_0^{(1)}\hat{s}_0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \Sigma_0 & \begin{bmatrix} P_0 H_0^{(1)^T} \\ S_0^{(1)} \\ R_0^{(1)} \\ R_0^{(1,2)^T} \end{bmatrix} \\ \begin{bmatrix} H_0^{(1)}P_0 & S_0^{(1)^T} & R_0^{(1)} & R_0^{(1,2)} \end{bmatrix} & (H_0^{(1)}P_0H_0^{(1)^T} + R_0^{(1)}) \end{bmatrix} . \tag{A.58}$$

From Lemma A.2, it follows that $[s_0^T \;\; w_0^T \;\; \nu_0^{(1)^T} \;\; \nu_0^{(2)^T}]^T$ conditioned on $y_0^{(1)}$ is gaussian and has mean

$$m_{0|y_0^{(1)}} = \begin{bmatrix} \hat{s}_0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} P_0 H_0^{(1)^T} \\ S_0^{(1)} \\ R_0^{(1)} \\ R_0^{(1,2)^T} \end{bmatrix} (H_0^{(1)}P_0H_0^{(1)^T} + R_0^{(1)})^{-1}(y_0^{(1)} - H_0^{(1)}\hat{s}_0) \tag{A.59}$$

and covariance

$$\Sigma_{0|y_0^{(1)}} = \Sigma_0 - \begin{bmatrix} P_0 H_0^{(1)^T} \\ S_0^{(1)} \\ R_0^{(1)} \\ R_0^{(1,2)^T} \end{bmatrix} (H_0^{(1)}P_0H_0^{(1)^T} + R_0^{(1)})^{-1} \begin{bmatrix} H_0^{(1)}P_0 & S_0^{(1)^T} & R_0^{(1)} & R_0^{(1,2)} \end{bmatrix} . \tag{A.60}$$

From (A.54c), form the following matrix equation:

$$\begin{bmatrix} s_0 \\ w_0 \\ \nu_0^{(1)} \\ \nu_0^{(2)} \\ y_0^{(2)} \end{bmatrix} = \begin{bmatrix} I \\ \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix} \end{bmatrix} \begin{bmatrix} s_0 \\ w_0 \\ \nu_0^{(1)} \\ \nu_0^{(2)} \end{bmatrix}. \tag{A.61}$$

Then using Lemma A.1, the joint distribution of $[s_0^T \ w_0^T \ \nu_0^{(1)^T} \ \nu_0^{(2)^T} \ y_0^{(2)}]^T$ given $y_0^{(1)}$ is gaussian with mean

$$m_{0,y_0^{(2)}|y_0^{(1)}} = \begin{bmatrix} I \\ \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix} \end{bmatrix} m_{0|y_0^{(1)}} \tag{A.62}$$

and covariance

$$\begin{aligned} \Sigma_{0,y_0^{(2)}|y_0^{(1)}} &= \begin{bmatrix} \Sigma^{(11)}_{0,y_0^{(2)}|y_0^{(1)}} & \Sigma^{(12)}_{0,y_0^{(2)}|y_0^{(1)}} \\ \Sigma^{(12)^T}_{0,y_0^{(2)}|y_0^{(1)}} & \Sigma^{(22)}_{0,y_0^{(2)}|y_0^{(1)}} \end{bmatrix} \\ &= \begin{bmatrix} I \\ \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix} \end{bmatrix} \Sigma_{0|y_0^{(1)}} \begin{bmatrix} I \\ \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix} \end{bmatrix}^T \end{aligned} \tag{A.63}$$

where

$$\Sigma^{(11)}_{0,y_0^{(2)}|y_0^{(1)}} = \Sigma_{0|y_0^{(1)}} \tag{A.64}$$

$$\Sigma^{(12)}_{0,y_0^{(2)}|y_0^{(1)}} = \Sigma_{0|y_0^{(1)}} \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix}^T \tag{A.65}$$

$$\Sigma^{(22)}_{0,y_0^{(2)}|y_0^{(1)}} = \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix} \Sigma_{0|y_0^{(1)}} \begin{bmatrix} H_0^{(2)} & 0 & 0 & I \end{bmatrix}^T. \tag{A.66}$$

From Lemma A.2, it follows that $[s_0^T \ w_0^T \ \nu_0^{(1)^T} \ \nu_0^{(2)^T}]$ given $\{y_0^{(2)}, y_0^{(1)}\}$ is gaussian with mean

$$m_{0|y_0^{(2)},y_0^{(1)}} = m_{0|y_0^{(1)}} + \Sigma^{(12)}_{0,y_0^{(2)}|y_0^{(1)}} (\Sigma^{(22)}_{0,y_0^{(2)}|y_0^{(1)}})^{-1} (y_0^{(2)} - E\{y_0^{(2)}|\, y_0^{(1)}\}) \tag{A.67}$$

where $E\{y_0^{(2)}|\, y_0^{(1)}\}$ is obtained from (A.62) :

$$E\{y_0^{(2)}|\, y_0^{(1)}\} = H_0^{(2)}\hat{s}_0 + (H_0^{(2)}P_0H_0^{(1)^T} + R_0^{(1,2)^T})(H_0^{(1)}P_0H_0^{(1)^T} + R_0^{(1)})^{-1}(y_0^{(1)} - H_0^{(1)}\hat{s}_0) \tag{A.68}$$

and the conditional covariance of $[s_0^T \;\; w_0^T \;\; \nu_0^{(1)^T} \;\; \nu_0^{(2)^T}]$ given $\{y_0^{(2)}, y_0^{(1)}\}$ is

$$\Sigma_{0|y_0^{(2)},y_0^{(1)}} = \Sigma^{(11)}_{0,y_0^{(2)}|y_0^{(1)}} - \Sigma^{(12)}_{0,y_0^{(2)}|y_0^{(1)}}(\Sigma^{(22)}_{0,y_0^{(2)}|y_0^{(1)}})^{-1}\Sigma^{(12)^T}_{0,y_0^{(2)}|y_0^{(1)}} \,. \tag{A.69}$$

It remains only to find the formulas for the conditional mean and conditional covariance of s_1 given $\{y_0^{(2)}, y_0^{(1)}\}$. Hence, we write (A.54a) in the following matrix form for $k = 0$:

$$s_1 = [\; F_0 \quad I \quad 0 \quad 0 \;]\begin{bmatrix} s_0 \\ w_0 \\ \nu_0^{(1)} \\ \nu_0^{(2)} \end{bmatrix} . \tag{A.70}$$

The linear transformation above shows that s_1 given $\{y_0^{(2)}, y_0^{(1)}\}$ is gaussian. The conditional mean and covariance of s_1 given $\{y_0^{(2)}, y_0^{(1)}\}$ follows immediately from Lemma A.1;

$$\hat{s}_{1|0} = E\{s_1|\, y_0^{(2)}, y_0^{(1)}\} = [\; F_0 \quad I \quad 0 \quad 0 \;]\, m_{0|y_0^{(2)},y_0^{(1)}} \tag{A.71a}$$

$$\begin{aligned} P_{1|0} &= E\{(s_1 - \hat{s}_{1|0})(s_1 - \hat{s}_{1|0})^T|\, y_0^{(2)}, y_0^{(1)}\} \\ &= [\; F_0 \quad I \quad 0 \quad 0 \;]\, \Sigma_{0|y_0^{(2)},y_0^{(1)}}\, [\; F_0 \quad I \quad 0 \quad 0 \;]^T . \end{aligned} \tag{A.71b}$$

Substituting (A.67) and (A.69) into (A.71a) and (A.71b), respectively, give

$$\hat{s}_{1|0} = F_0\hat{s}_0 + K_0(y_0 - H_0\hat{s}_0) \tag{A.72}$$

$$P_{1|0} = F_0P_0F_0^T + Q_0 - K_0(H_0P_0H_0^T + R_0)K_0^T \tag{A.73}$$

where

$$K_0 = (F_0P_0H_0 + S_0)(H_0P_0H_0^T + R_0)^{-1} . \tag{A.74}$$

More generally, repetition of the steps above yields

$$\hat{s}_{k+1|k} = F_k\hat{s}_k + K_k(y_k - H_k\hat{s}_{k|k-1}) \tag{A.75}$$

$$P_{k+1|k} = F_kP_{k|k-1}F_k^T + Q_k - K_k(H_kP_{k|,k-1}H_k^T + R_k)K_k^T \tag{A.76}$$

where

$$K_k = (F_kP_{k|k-1}H_k + S_k)(H_kP_{k|k-1}H_k^T + R_k)^{-1}. \tag{A.77}$$

Bibliography

[1] M. Aoki. "Optimization of Stochastic Systems,". *New York: Academic*, 1967.

[2] J. D. Aplevich. "Time-domain Input-Output Representations of Linear Systems,". *Automatica*, Vol. 17(No. 3):pp. 509–522, 1981.

[3] J. D. Aplevich. "Implicit linear systems,". *Springer-Verlag*, 1991.

[4] J. D. Aplevich and Y. L. Shi. "Simultaneous estimation of linear multivariable systems with constrained parameters,". *Int. J. Control and Computers*, Vol. 15:pp. 87–91, 1987.

[5] K. J. Åström. "Optimal control of markov processes with incomplete state information, Part I,". *J. math. Anal. Applic.*, Vol. 10:pp. 174–205, 1965.

[6] K. J. Åström. "Optimal control of markov processes with incomplete state information, Part II,". *J. math. Anal. Applic.*, Vol. 36:pp. 403–406, 1969.

[7] K. J. Åström. "Introduction to Stochastic Control Theory,". *New York: Academic*, 1970.

[8] K. J. Åström and B. Wittenmark. "Problems of identification and control,". *J. math. Analysis Applic.*, Vol. 34:pp. 90–113, 1971.

[9] K. J. Åström and B. Wittenmark. "On self-tuning regulators,". *Automatica*, Vol. 9:pp. 195–199, 1973.

[10] K. J. Åström and B. Wittenmark. "Adaptive control,". *Addison-Wesley*, 1989.

[11] U. Borrison. "Self tuning regulator for a class of multivariable systems,". *Automatica*, Vol. 15:pp. 209–217, 1979.

[12] P. E. Caines. "Linear Stochastic Systems,". *Wiley, New York*, 1988.

[13] F. R. Chang and L. C. Wang. "Algorithms for the interactor matrix and the row -column-reduced form,". *Control-Theory and Advanced Technology*, 6:pp. 483–491, 1990.

[14] H. F. Chen and L. Guo. "Asymptotically optimal adaptive control with consistent parameter estimates,". *SIAM J. Control and Optimization*, Vol. 25(No. 3):pp. 558–575, 1987.

[15] Han-Fu Chen, P. R. Kumar, and J. H. van Schuppen. "On Kalman filtering for conditionally Gaussian systems with random matrices,". *Systems and Control Letters*, Vol. 13:pp. 397–404, 1989.

[16] Y.C. Chen, J. D. Aplevich, and W.J. Wilson. "Simultaneous estimation of state variables and parameters for multivariable linear systems with singular pencil models,". *IEE Proceedings*, Vol. 133, Pt.D(2):pp. 65–72, 1986.

[17] M. Das. "Multivariable adaptive model matching using less a priori information,". *J. Dynam. Syst. Measurement Contr.*, Vol. 108:pp. 151–153, 1986.

[18] J. M. Dion, L. Dugard, and J. Carrillo. "Interactor and multivariable adaptive model matching,". *IEEE Trans. Automat. Contr.*, Vol. 33(No. 4):pp. 399–401, 1988.

[19] L. Dugard, G. C. Goodwin, and C. de Souza. "Prior knowledge in model reference adaptive control of multi-input multi-output systems,". *22nd IEEE Decision and Control Conference, San Antonio, USA*, pages pp. 455–459, 1983.

[20] L. Dugard, G. C. Goodwin, and X. Xianya. "The role of the interactor matrix in multivariable stochastic adaptive control,". *Automatica*, (No. 5):pp. 701–709, 1984.

[21] H. Elliot and W. A. Wolovich. "A parameter adaptive control structure for linear multivariable systems,". *IEEE Trans. Automat. Contr.*, Vol. AC-27:pp. 340–352, 1982.

[22] H. Elliott and W. A. Wolovich. "Parameterization Issues in Multivariable Adaptive Control,". *Automatica*, Vol. 20(No. 5):pp. 533–545, 1984.

[23] A. A. Feldbaum. "Optimal Control Systems,". *New York: Academic*, 1965.

[24] S. I. Fingerote. "Multivariable frequency domain approximation using singular pencil models,". *M.A. Sc. thesis, Dept. of Electrical Engineering University of Waterloo, Waterloo, Ontario, Canada*, 1982.

[25] K. Furuta and S. Kamiyama. "State feedback and inverse system,". *International Journal of Control*, 25:pp. 229–241, 1977.

[26] C. G. Goodwin and L. Dugard. "Stochastic adaptive control with known and unknown interactor matrices,". *IFAC workshop on adaptive Control, San Francisco, USA*, pages pp. 351–355, 1983.

[27] G. C. Goodwin, P. J. Ramadge, and P. E. Caines. "Discrete time stochastic adaptive control,". *SIAM J. Control and Optimization*, Vol. 19(No. 6):pp. 829–853, 1981.

[28] G. C. Goodwin and K. S. Sin. "Adaptive filtering prediction and control,". *Englewood Cliffs, New Jersey, Prentice-Hall*, 1984.

[29] L. Guo and Han-Fu Chen. "The Astrom-Wittenmark self-tuning regulator revisited and ELSE-based adaptive trackers,". *IEEE Trans. Automat. Contr.*, Vol. 36(No. 7):pp. 802–812, 1991.

[30] R. Johansson. "Parametric models of linear multivariable systems for adaptive control,". *Proc. IEEE conf. Decision and Control, Orlando, FL*, pages pp. 989–993, 1982.

[31] H. N. Koivo. "A multivariable self tuning controller,". *Automatica*, Vol. 16:pp. 351–366, 1980.

[32] P. R. Kumar. "A survey of some results in stochastic adaptive control,". *SIAM J. Control and Optimization*, Vol. 23:pp. 329–380, 1985.

[33] P. R. Kumar. "Convergence of Adaptive Control Schemes Using Least-squares Parameter Estimates,". *IEEE Trans. Automat. Contr.*, Vol. 35(No. 4):pp. 416–424, 1990.

[34] L. Ljung and T. Söderström. "Theory and practice of recursive identification,". *MIT Press, Cambridge*, 1985.

[35] R. H. Middelton, G. C. Goodwin, D. J. Hill, and D. Q. Mayne. "Design Issues in Adaptive Control,". *IEEE Trans. Automat. Contr.*, Vol. 33(No. 1):pp. 50–58, 1988.

[36] K. S. Miller. "Multidimensional Gaussian Distributions,". *The SIAM Series in Applied Mathematics, John Wiley and Sons Inc.*, 1964.

[37] L. Mo and M. M. Bayoumi. "Adaptive pole placement control of MIMO systems,". *IEEE Trans. Automat. Contr.*, Vol. 38(No. 6):pp. 967–970, 1993.

[38] P. Mookerjee and Y. Bar-Shalom. "An Adaptive Dual Control for a MIMO-ARMA System,". *IEEE Trans. Automat. Contr.*, Vol. 34(No. 7):pp. 795–800, 1989.

[39] A. S. Morse. "Parametrization for multivariable adaptive control,". *Proc. IEEE Conf. Decision and Control, San Diego, Calif.*, 1981.

[40] Z. Nahorski and R. V. V. Vidal. "Stablization of a linear discrete time system using simultaneous identification and control,". *Int. J. control*, Vol. 19(No. 2):pp. 353–364, 1974.

[41] K. S. Narendra, R. Ortega, and P. Dorato. "Advances in adaptive control,". *IEEE Press*, 1991.

[42] V. Panuska. "A new form of the extended Kalman filter for parameter estimation in linear systems with correlated noise,". *IEEE Trans. Automat. Contr.*, Vol. 25:pp. 229–235, 1980.

[43] B. Pasik-Duncan. "Stochastic adaptive control: an historical prospective,". *Proceedings of the 33rd conference on Decision and Control, Lake Buena Vista. FL.*, pages pp. 2126–2130, 1994.

[44] W. Ren and P. R. Kumar. "Stochastic adaptive prediction and model reference control,". *IEEE Trans. Automat. Contr.*, Vol. 39(No. 10):pp. 2047–2060, 1994.

[45] G. Salut. "Identification optimale des systèmes linéaires stochastiques,". *PhD thesis, Université Paul Sabtier, Toulouse, France,*, 1976.

[46] G. Salut, J. Aguilar-Martin, and S. Lefebvre. "New results on optimal joint parameter and state estimation of linear stochastic systems,". *Transactions of the ASME*, Vol. 102:pp. 28–34, 1980.

[47] K. S. Sin, G. C. Goodwin, and R. R. Bitmead. "An adaptive d-step ahead predictor based on least squares,". *IEEE Trans. Automat. Contr.*, Vol. AC-25:pp. 1161–1165, 1980.

[48] R. P. Singh and K.S. Narendra. "Prior information in the design of multivariable adaptive controllers,". *IEEE Trans. Automat. Contr.*, Vol. AC-29:pp. 1108–1111, 1984.

[49] V. Solo. "The convergence of AML,". *IEEE Trans. Automat. Contr.*, Vol. 24(No. 6):pp. 958–962, 1979.

[50] H. Elliot W., A. Wolovich, and M. Das. "Arbitrary adaptive pole-placement for linear multivariable systems,". *IEEE Trans. Automat. Contr.*, Vol. AC-29:pp. 221–229, 1984.

[51] J. Wieslander and B. Wittenmark. "An approach to adaptive control using real-time identification,". *Automatica*, Vol. 7:pp. 211–217, 1971.

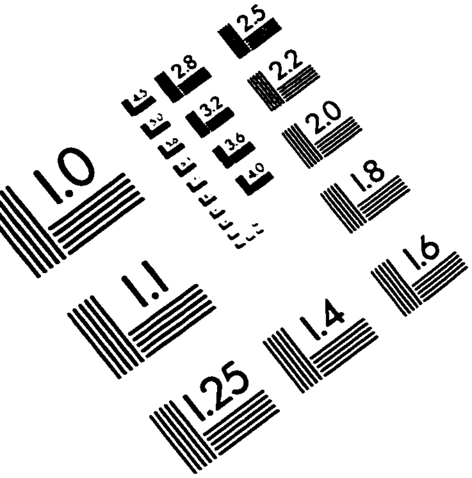
[52] B. Wittenmark. "A self-tuning predictor,". *IEEE Trans. Automat. Contr.*, Vol. 19:pp. 848–851, 1974.

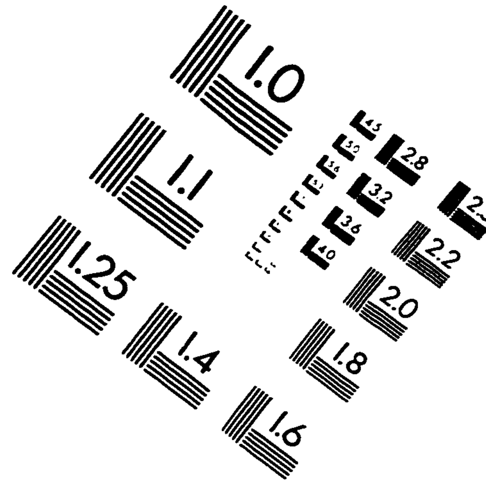
[53] B. Wittenmark. "Stochastic adaptive control methods: a survey,". *Int. J. Control*, Vol. 21(No. 5):pp. 705–730, 1975.

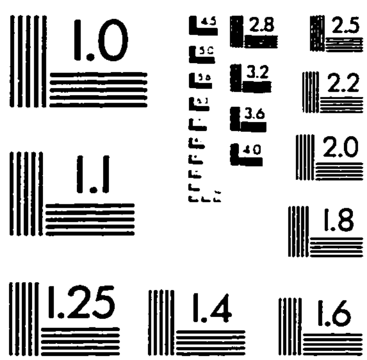
[54] W. A. Wolovich and P. L. Falb. "Invariants and canonical forms under dynamic compensation,". *SIAM J. Control and Optimization*, Vol. 14(No. 6):pp. 996–1008, 1976.

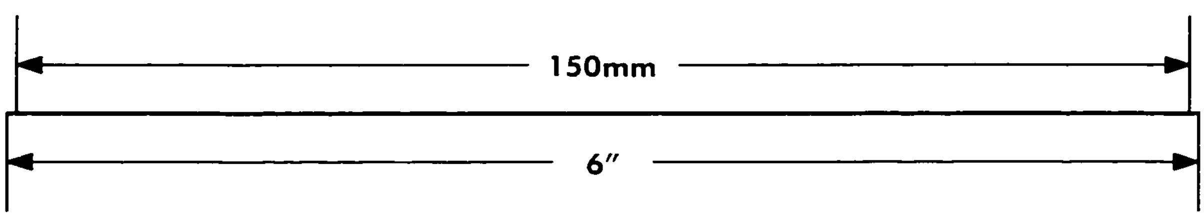
[55] J. Zhang and S. Lang. "Adaptive control of a class of multivariable nonlinear systems and convergence analysis,". *IEEE Trans. Automat. Contr.*, Vol. 34(No. 7):pp. 787–791, 1989.

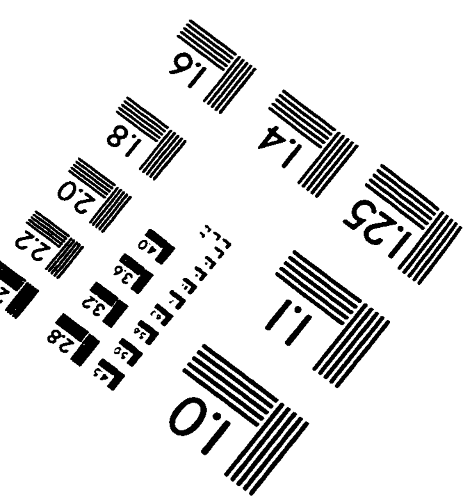
IMAGE EVALUATION TEST TARGET (QA-3)

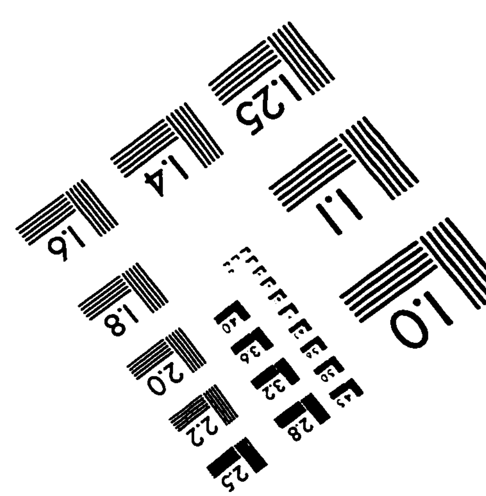












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989