SUBSPACE METHODS AND INFORMATIVE EXPERIMENTS FOR SYSTEM IDENTIFICATION

Nelson Loong Chik Chui

Pembroke College
Cambridge

A dissertation submitted for
the degree of Doctor of Philosophy

October 1997

To my parents and sisters

Abstract

This dissertation is concerned with subspace identification methods for the state-space models.

We examine the concept of informative experiments in the context of subspace methods. Informative experiments are identification experiments which provide sufficient information to discriminate between different models in an intended model set for a specific identification algorithm. Under a Hilbert space setup, it is shown that deterministic and stochastic components can incur a perfect cancelation in a subspace projection which essentially restrains subspace identification algorithms from arriving at a correct model. Criteria preventing such a cancelation and guaranteeing experiments to be informative are derived. It turns out that these criteria consist of mild conditions, mainly depending on the excitation of input signals, the observability index, and the order of the unknown system. Throughout the analysis, we employ a geometric interpretation to investigate the effect of projections in subspace methods. Such an approach also leads to a simplified derivation of a subspace projection theorem.

The study is extended to the case where *a priori* knowledge of initial Markov parameters is given. A necessary and sufficient condition for experiments to be informative is derived. We then turn our attention to the development of a subspace identification method in which initial Markov parameters are determined first. The Kalman filter state sequence is estimated by removing the effect of the input signals using the estimated Markov parameters. The advantage of this method, particularly in the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ sense, is illustrated in two case studies.

Finally, we examine several properties of eigenvalue perturbation in least-squares estimation. Sufficient conditions for least-squares solutions to be stable are also derived. These results are applied to subspace methods to acquire stable system approximants. Either asymptotic or marginal stability can be ensured with this approach.

Keywords: Modeling, System Identification, Realization, Subspace Method, Informative Experiment, Markov Parameter, Stable Realization, Eigenvalue Perturbation.

Acknowledgements

I would like to express my gratitude to my supervisor Dr Jan Maciejowski for providing me with his constant guidance, encouragement and support, and for introducing me to the life at Pembroke College. I would also like to thank Professor Keith Glover, Dr Malcolm Smith, and Dr Glenn Vinnicombe, from whom I have learned much about control theory.

I am also grateful to my colleagues in the control group for making life in Cambridge more enjoyable. In addition, I would like to thank Michael Cantoni, Olivia Kwong, and Richard Lee, who have spent days and nights proofreading various versions of the manuscript. My thanks also go to my friends, both near and far, who have made my experience in Cambridge very memorable.

I am greatly indebted to my parents and sisters for their continuing encouragement and support during my years in Cambridge.

This work has been supported financially by the Natural Sciences and Engineering Research Council of Canada, the Committee of Vice-Chancellors and Principals of the Universities of the United Kingdom in the form of an Overseas Research Student Award, and the Cambridge Commonwealth Trust. Travel grants for conference attendance have been gratefully received from the Cambridge University Engineering Department, the European Research Network in System Identification, and Pembroke College.

As required by University Statute, I hereby declare that this dissertation is not substantially the same as any that I have submitted for a degree at any other University, is the result of my own work, and includes nothing which is the outcome of work done in collaboration.

Nelson L C Chui
Pembroke College
Cambridge
October 21, 1997

Contents

List of Figures

Notation

Symbols

$\mathbb{R}$	field of real numbers
$\mathbb{C}$	field of complex numbers
$\mathbb{N}$	set of natural numbers (positive integers)
$\mathbb{Z}_+$	set of nonnegative integers
$GL_n(\mathbb{R})$	real general linear group of order n (set of $n \times n$ nonsingular real matrices)
$\in$	is an element of
$\forall$	for all
$:=$	is defined as
$=:$	is the definition of
$\subset$	is a subset of
$\cap$	the intersection of two given sets
$\bar{\sigma}(\cdot)$	the maximum singular value
$\rho(\cdot)$	the spectral radius
$\cdot^*$	the complex conjugate transpose
$\cdot^\dagger$	the general Moore-Penrose inverse
$Re\{\cdot\}$, $Im\{\cdot\}$	the real and imaginary part

Acronyms

SISO	Single-Input-Single-Output
MIMO	Multiple-Input-Multiple-Output
LTI	Linear Time-Invariant
SVD	Singular Value Decomposition
S.R.	Sufficient Richness or Sufficiently Rich
MP	Markov Parameter

1

Introduction

1.1 Overview

The purpose of system identification is to construct models of dynamical systems based on input-output observations. These models, which play an important role in many fields, such as system control, communications, econometrics, and operations research, can be used to explain the behavior of the process being identified, to make predictions in terms of the exogenous variables or input signals, and to design regulators and filters. Prior to determining a suitable model, a set of candidate models has to be specified as a first step. The model set considered in this thesis is the set of discrete-time, finite-dimensional linear time-invariant, state-space models.

Parameter estimation methods are a class of well-developed and commonly used identification schemes. These methods consist of an optimization procedure which minimizes the mismatch between the model and the data to search for the best model. One drawback of these methods is the difficulty and complexity in parametrizing the intended model set, despite extensive research in this area has been carried out. Then, from the late 1980's a new approach, known as subspace identification methods, has been developed, providing an effective and numerically efficient way to identify systems. These methods require no parametrizations and are also shown to give good initial models for parameter estimation methods.

The analysis in this dissertation starts from the subspace projection theorem presented in [VODM94], which proves that a state sequence projected onto a certain instrumental variable space is equivalent to a Kalman filter state sequence with some initial condition. By examining this subspace projection theorem and particularly the associated initial condition, a geometric interpretation, which divides the state sequence into three orthogonal components, is acquired. Apart from giving a simpler proof for the subspace projection theorem, this geometric property has provided a tool to analyze informative experiments for subspace algorithms. Under the Hilbert space setup defined in this thesis, which may also be perceived as an asymptotic analysis, it is shown that deterministic and stochastic components can incur a perfect cancelation in a subspace projection which essentially restrains the identification algorithm from arriving at a correct model. Part of the work in this thesis has been devoted to deriving conditions to prevent such a cancelation.

The study of necessary and sufficient conditions for informative experiments, which provide sufficient information for the corresponding identification algorithm to identify a model correctly, has become another motivation of the research. From some existing results on uniqueness of minimal dimensional realizations, a necessary and sufficient condition for experiments to be informative for a particular deterministic subspace identification algorithm can be procured in a straightforward manner. The study is further extended to a 'grey box' identification problem in which *a priori* knowledge of initial Markov parameters is incorporated.

Several different approaches of subspace identification methods have been proposed in the literature, such as CVA in [Lar83], MOESP in [Ver94], and N4SID in [VODM94]. Most of these methods are related to one another to a certain extent by a weighting function. To eliminate the effect of the input signals from the system equations, projections onto the orthogonal space of the input sequences are performed in most subspace methods. Effectively, the extracted state sequence will depend on the input signal used, which may in turn affect the accuracy in choosing the principal dynamics of the system. A new subspace identification approach, aiming at avoiding the projection onto the orthogonal space of the input sequence, is investigated. Such an approach requires a different algorithmic structure in that it turns out to be not related to those methods mentioned above by a weighting function.

In some cases, asymptotic or marginal stability may be a desire property of the estimated models. For this reason, stability enforcement with subspace methods has been examined in [Mac95, McK95]. Several results on this aspect are also extended towards the end of this thesis.

1.2 Outline

This dissertation has ten chapters. The remainder of this dissertation is organized as follows:

Chapter 2: Mathematical Preliminaries

In this chapter, we review some background material which will be used in later chapters. We first define a discrete-time, linear time-invariant, state-space model with state vectors in a Hilbert space. Two particular cases, an ideal probabilistic setup and a continuous time setup with piecewise constant inputs, are elaborated. Transfer functions, $\mathcal{H}_\infty$, $\mathcal{H}_2$ Hardy spaces, and positive real functions are presented. Furthermore, signal correlation functions and power spectra, and their relationship to $\mathcal{H}_\infty$, $\mathcal{H}_2$ norms and induced system gains are discussed.

Chapter 3: Review of Subspace Identification Theory

The general setup for subspace identification is stated in this chapter. Precise system definitions and identification conditions are also given. Then we present some results on subspace projection theory and several subspace identification algorithms. Both the stochastic setup and the combined deterministic-stochastic setup are discussed. Finally a closely related subject, namely model reduction theory, is briefly reviewed.

Chapter 4: A Geometric Interpretation for Subspace Projections

The main purpose of this chapter is to give a geometric interpretation for subspace projections. Van Overschee and De Moor [VODM94] showed using an algebraic approach that a state sequence projected onto a certain input-output space was actually a state sequence estimate, running through a bank of Kalman filters, with the initial condition $\hat{X}_p = \Pi_{\mathcal{U}_p+\mathcal{U}_f} X_p$. Here, we develop a framework which generalizes both purely deterministic and fully stochastic cases to an amalgamated extension. It is then shown using a geometric approach that the initial condition of the corresponding Kalman filter turns out to be $\hat{X}_p = \Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p)+\mathcal{U}_p+\mathcal{U}_f} X_p$. The advances developed in this chapter also provide the key techniques for the analysis in the next chapter.

Chapter 5: Informative Experiments for Subspace Methods

Informative experiments are identification experiments which contain sufficient information to discriminate between different models in an intended model set. However, different algorithms require different amounts of information in an experiment for a successful identification operation. In this chapter, a particular set of identification algorithms, namely subspace methods, is considered. Criteria for experiments to be informative with these methods in the deterministic setup and the combined deterministic-stochastic setup are presented. It is illustrated by means of an example that if these criteria are not satisfied, perfect cancelations of the deterministic components and the stochastic components can occur in a subspace projection. It is further shown that such cancelations can indeed be avoided under mild conditions.

Chapter 6: Informative Experiments – Markov Parameters *A Priori*

This chapter is concerned with the joint input-output criteria for informative experiments, particularly under the situation that *a priori* knowledge of initial Markov parameters is given. We first present the conditions under which a set of equivalent n-th order realizations is uniquely determined from the input-output observations. Then, we impose another constraint confining the first few Markov parameters to certain values. The conditions, under which a set of equivalent n-th order realizations with Markov parameters matching the requirements is uniquely determined from the input-output observations, are explored.

Chapter 7: Subspace Identification via Markov Parameter Estimation

In this chapter, we propose and present a different approach to subspace system identification, in which Markov parameters are first estimated, without imposing restrictions on the input signal. The prominent difference of this approach is that a three-block configuration of data matrices is used. Roughly speaking, these three data blocks can be interpreted as follows. The first block is used primarily as an instrumental variable. The second block then provides the information from which the Kalman filter state sequence is observed. Finally, the last block exploits the relationship between the Kalman filter state sequences and the input sequences acquired from the second block in order to extract the initial Markov parameters of the system. Advantages and disadvantages of this algorithm will be investigated in more detail through some case studies in Chapter 9.

Chapter 8: Stability Enforcement with Realization and Subspace Algorithms

Approximate realization algorithms and subspace methods for system identification estimate the dynamics of state-space models either by using the 'shift-invariance' property of an estimated observability or controllability matrix, or by estimating a state sequence and then solving a least-squares problem to obtain the system matrices. In either case it is possible for the estimated system to be unstable. We present algorithms to find stable approximants to a least-squares problem, which can then be applied to subspace methods to ensure stability. Either asymptotic or marginal stability can be ensured, in the latter case a pole or a pair of poles being forced to lie on the unit circle. In addition, sufficient conditions for least-squares solutions obtained by the shift invariance approach to be stable are derived.

Chapter 9: Case Studies

Various aspects regarding the performance of the subspace algorithms presented in Chapter 7 are investigated in two case studies. These case studies are based on real data extracted from two industrial systems, namely a fractional distillation column and an industrial dryer. In the first example, experimental data are set up in a rather ideal case in which $\mathcal{H}_\infty$ and $\mathcal{H}_2$ errors are investigated. One purpose of such a study is to reveal how well the algorithms perform in extracting system dynamics within a certain model order. On the other hand, performance in the second example is measured in a more classical fashion, namely in terms of simulation and prediction accuracy.

Chapter 10: Conclusions

We summarize the main contributions of this dissertation and make some recommendations for future research in this area.

2

Mathematical Preliminaries

2.1 Introduction

In this chapter we present some general background material pertinent to future chapters. As the material is not self-contained, general references are cited for a complete exposition of the topics. The first two sections mainly consist of general setups and mathematical tools for the subsequent studies. The remaining two sections center on material required for stochastic realization theory.

Section 2.2 defines a discrete-time, linear time-invariant, state-space model with state vectors in a Hilbert space. Two particular cases, an ideal probabilistic setup and a continuous time setup with piecewise constant inputs, are elaborated. Transfer functions and $\mathcal{H}_\infty$, $\mathcal{H}_2$ Hardy spaces are covered in Section 2.3. In Section 2.4, we then discuss signal correlation functions and power spectra, and their relationship to $\mathcal{H}_\infty$, $\mathcal{H}_2$ norms and induced system gains. In Section 2.5, characterizations of discrete-time positive real functions are presented.

2.2 Signals and Systems

This dissertation is mainly concerned with the identification aspects of finite-dimensional, linear time-invariant, state-space models. Such models can be represented by:

$$x(t+1) = Ax(t) + Bu(t) + w(t), \tag{2.1a}$$

$$y(t) = Cx(t) + Du(t) + v(t), \tag{2.1b}$$

where the system matrices $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$, $D \in \mathbb{R}^{p\times m}$. The input signal and the output signal are denoted by u and y, respectively. On the other hand, w and v are respectively the process noise and the output noise. For deterministic systems, we then have $w = 0$ and $v = 0$.

We will consider a general class of signals whose elements are in a Hilbert space. Such a setup allows the analysis of stochastic processes with finite covariances in an ideal probabilistic environment. This probability space will be discussed in more detail in Section 2.2.1. Continuous-time systems with piecewise constant inputs also share similar characteristics in which state signals can be put in a Hilbert space framework. Further discussion can be found in Section 2.2.2.

Now, let $\mathscr{X}$ be a Hilbert space over the field of real numbers $\mathbb{R}$ (or possibly $\mathbb{C}$) equipped with an inner product $\langle \cdot, \cdot \rangle$. A sequence of input-output measurements is then

$$\mathscr{Z}^{\infty} = (\dots, y(-1), u(-1), y(0), u(0), y(1), u(1), \dots),$$

where $y(\cdot) \in \mathscr{X}^p$, $u(\cdot) \in \mathscr{X}^m$. In addition, $\mathscr{Z}^N$ will be a finite sequence of length N:

$$\mathscr{Z}^N = (y(0), u(0), y(1), u(1), \dots, y(N-1), u(N-1)). \tag{2.2}$$

2.2.1 Hilbert Spaces of Random Variables

Modeling of stochastic processes in an ideal probabilistic setup will be considered. Let Ω be a sample space, $\mathscr{F}$ be a σ-field of subsets of Ω, and $\mathbf{P}$ be a probability measure on $(\Omega, \mathscr{F})$. We shall further assume that the probability space $(\Omega, \mathscr{F}, \mathbf{P})$ is complete. That is for any subsets $\Omega_1 \subset \Omega_2$, $\Omega_2 \in \mathscr{F}$ and $\mathbf{P}(\Omega_2) = 0$ imply $\Omega_1 \in \mathscr{F}$ (and hence $\mathbf{P}(\Omega_1) = 0$). It is noteworthy that in any probability space, it is possible to enlarge the σ-field and extend the measure so as to get a complete space [Rud87, GS93]. Let $\mathscr{F}_0$ be the collection of all subsets of null sets in $\mathscr{F}$ (where null sets are sets with zero probability measure). To complete the space, we enlarge $\mathscr{F}$ to the smallest σ-field which contains all sets in $\mathscr{F}$ and $\mathscr{F}_0$. The probability measure $\mathbf{P}$ is then extended to the new σ-field in the obvious way.

A real random variable ψ is a function $\psi : \Omega \to \mathbb{R}$ with the property that $\{\omega \in \Omega : \psi(\omega) \leq c\} \in \mathscr{F}$ for each $c \in \mathbb{R}$. Denote by $\mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P})$ the space of all real Lebesgue square integrable random variables on $(\Omega, \mathscr{F}, \mathbf{P})$:

$$\mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P}) := \left\{ \psi \,:\, \int_{\Omega} \psi^2(\omega)\, d\mathbf{P}(\omega) < \infty \right\}.$$

The inner product of the space is defined as the map:

$$\begin{aligned} \langle \cdot, \cdot \rangle \,:\, \mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P}) \times \mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P}) &\to \mathbb{R}, \\ (\psi, \phi) &\mapsto \int_{\Omega} \psi(\omega)\, \phi(\omega)\, d\mathbf{P}(\omega). \end{aligned}$$

It can be further shown that the inner product space is complete; and hence we have the following theorem (see for example [Rud87, Cai88, Bil95]).

Theorem 2.1. $\mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P})$ *is a Hilbert space.*

In addition, denote by $\mathbf{E}$ the usual expectation operator. It is straightforward that $\mathbf{E}(\psi\, \phi) = \langle \psi, \phi \rangle$. Furthermore, suppose ψ is a p-dimensional random vector and ϕ is an m-dimensional random vector, then $\mathbf{E}\,(\psi\,\phi^*)$ can be seen as taking their inner product element-wise, forming a $p \times m$ real matrix.

Finally, it can be seen without difficulty that the stochastic processes described above fit into the setup defined in the beginning of this section with $\mathscr{X} = \mathcal{L}_2(\Omega, \mathscr{F}, \mathbf{P})$. Furthermore throughout this dissertation, we shall abuse slightly the use of $\mathbf{E}\,(\cdot\,\cdot^*)$ to denote the element-wise inner product even if $\mathscr{X}$ is not a probability space. For example, let

$$\psi = \begin{bmatrix} \psi_1 \\ \vdots \\ \psi_p \end{bmatrix} \in \mathscr{X}^p, \qquad \text{and} \qquad \phi = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_m \end{bmatrix} \in \mathscr{X}^m,$$

then

$$\mathbf{E}\,(\psi\,\phi^*) = \Big[\, \langle\, \psi_i, \phi_j \,\rangle \,\Big]_{ij} \in \mathbb{R}^{p\times m}.$$

2.2.2 Continuous-Time Systems with Piecewise Constant Inputs

A finite-dimensional continuous-time system is defined as:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + w(t), \\ y(t) &= Cx(t) + Du(t) + v(t), \end{aligned}$$

with A, B, C, D of dimensions as defined in Equations (2.1). The solution of the state equation is given by

$$x(t) = e^{At}x(0) + \int_0^t e^{A(t-\tau)}(Bu(\tau) + w(\tau))d\tau.$$

Suppose the input is held constant for every interval h. That is, $u(kh) = u(kh+\tau)$ for any $0 \le \tau < h$. Now, define the 'lifted' signals $\breve{x}[k]$ and $\breve{w}[k] \in \mathcal{L}_2[0,h)$ in such a way that for $0 \le t < h$, $k \in \mathbb{Z}_+$,

$$\begin{aligned} (\breve{x}[k])(t) &= x(kh+t), \\ (\breve{w}[k])(t) &= \int_0^t e^{A(t-\tau)}w(kh+\tau)d\tau. \end{aligned}$$

Moreover, the 'lifted' signals $\breve{u}[k]$, $\breve{y}[k]$, and $\breve{v}[k]$ are defined similarly as $\breve{x}[k]$. Then, the continuous-time system can be written as

$$\begin{aligned} \breve{x}[k+1] &= \breve{A}\breve{x}[k] + \breve{B}\breve{u}[k] + \breve{w}[k], \\ \breve{y}[k] &= \breve{C}\breve{x}[k] + \breve{D}\breve{u}[k] + \breve{v}[k], \end{aligned}$$

where

$$\begin{aligned}
\breve{A} &: \mathcal{L}_2[0,h) \to \mathcal{L}_2[0,h); \quad \breve{x} \mapsto e^{Ah}\breve{x}, \\
\breve{B} &: \mathcal{L}_2[0,h) \to \mathcal{L}_2[0,h); \quad \breve{u} \mapsto \int_0^t e^{Ah-\tau} B d\tau \, \breve{u}, \\
\breve{C} &: \mathcal{L}_2[0,h) \to \mathcal{L}_2[0,h); \quad \breve{x} \mapsto C\breve{x}, \\
\breve{D} &: \mathcal{L}_2[0,h) \to \mathcal{L}_2[0,h); \quad \breve{u} \mapsto D\breve{u}.
\end{aligned}$$

It is important to note that the transformed system is again a system with system matrices of real numbers.

This procedure of transforming a continuous-time system to a discrete-time system is known as lifting [BPFT91], which has been used extensively in sampled-data control systems. It is noteworthy that the procedure described above is similar to that used in [Yam90], where state signals are lifted to an infinite-dimensional space. However, systems in [BPFT91] are lifted differently in such a way that state signals are sampled at a sampling rate h. This method produces a finite-dimensional state signal. However, $\breve{C}$ can no longer be representable by a matrix of real numbers.

Finally, note that this type of discrete-time systems is a special type of systems defined in the beginning of this section with $\mathscr{X} = \mathcal{L}_2[0,h)$.

2.3 Function Spaces

Norms are used to measure the size of signals and systems. Let $\ell_2(\mathscr{X})$ be the space of functions from $\mathbb{Z}$ to $\mathscr{X}$ that are square-summable:

$$\ell_2(\mathscr{X}) := \left\{ x \,:\, x(t) \in \mathscr{X},\ \sum_{t=-\infty}^{\infty} \|x(t)\|_{\mathscr{X}}^2 < \infty \right\}.$$

The norm of this space is defined by:

$$\|x\|_{\ell_2(\mathscr{X})}^2 := \sum_{t=-\infty}^{\infty} \|x(t)\|_{\mathscr{X}}^2.$$

Furthermore, $\ell_2(\mathscr{X})$ can be decomposed into $\ell_{2+}(\mathscr{X})$ and $\ell_{2-}(\mathscr{X})$, where $\ell_{2+}(\mathscr{X})$ is the subset of $\ell_2(\mathscr{X})$ consisting of all functions which are zero for negative t, and $\ell_{2-}(\mathscr{X})$ is the subset of $\ell_2(\mathscr{X})$ consisting of all functions which are zero for nonnegative t. Denote by z^{-1} the delay operator. Define the Z-transform of $u(t)$ as:

$$U(z) = \sum_{t=0}^{\infty} u(t)\, z^{-t}.$$

Define similarly for $Y(z)$, the Z-transform of $y(t)$. Note that if u and y are ℓ_{2+} sequences of $\mathscr{X}$-valued vectors, then $U(z)$ and $Y(z)$ are $\mathscr{X}$ vector valued functions in the complex plane z, convergent and analytic in the region $\{z \in \mathbb{C} : |z| > 1\}$. The transfer function $G(z)$ of the system (A, B, C, D) can be written as:

$$G(z) = C(zI - A)^{-1}B + D.$$

In addition, we define two useful systems – the conjugate system and an inverse system of $G(z)$.

Definition 2.1. The conjugate system $G^{\sim}(z)$ of $G(z)$ is defined as:

$$G^{\sim}(z) = B^*(z^{-1}I - A^*)^{-1}C^* + D^*.$$

Note that for $\omega \in [0, 2\pi]$, $[G(e^{i\omega})]^* = G^{\sim}(e^{i\omega})$.

Definition 2.2. $G^{\dagger}(z)$ is a left (or right) inverse system of $G(z)$ if $G^{\dagger}(z)G(z) = I$ (or $G(z)G^{\dagger}(z) = I$). Suppose D has full column (or row) rank, then

$$G^{\dagger}(z) = -D^{\dagger}C(zI - A + BD^{\dagger}C)^{-1}BD^{\dagger} + D^{\dagger}$$

is a left (or right) inverse system. If D is invertible, $G^{\dagger}(z)$ is unique and is denoted by $G^{-1}(z)$.

Suppose the initial state vector $x(0) = 0$, then we have $Y(z) = G(z)U(z)$. Furthermore, two useful classes of functions, the Hardy spaces $\mathcal{H}_\infty$ and $\mathcal{H}_2$, are considered.

Definition 2.3. The Hardy space $\mathcal{H}_\infty$ is the space of matrix valued functions that are analytic and (essentially) bounded outside the unit disc. The $\mathcal{H}_\infty$ norm is defined as

$$\|G(z)\|_\infty := \operatorname*{ess\,sup}_{|z|>1} \bar{\sigma}(G(z)).$$

Note that in order to be consistent with the standard Z-transform, the Hardy space is defined to be analytic and bounded outside the unit disc, whereas the conventional definition is defined inside the unit disc.

The $\mathcal{H}_\infty$ norm of an LTI system turns out to be the induced 2-norm of the system (see for example [SNF70]):

$$\|G(z)\|_\infty = \sup_{0 \neq u \in \ell_{2+}^m(\mathscr{X})} \frac{\|Gu\|_{\ell_{2+}^p(\mathscr{X})}}{\|u\|_{\ell_{2+}^m(\mathscr{X})}}.$$

A transfer function $G(z)$ is said to be asymptotically stable if its induced 2-norm is finite, i.e., the above supremum is finite. Thus, $G(z) \in \mathcal{H}_\infty$ if and only if $G(z)$ is asymptotically stable.

Definition 2.4. The Hardy space $\mathcal{H}_2$ is the space of matrix valued functions $G(z)$ analytic outside the unit disc such that

$$\langle\, G(z), G(z)\,\rangle < \infty,$$

where

$$\langle\, G(z), H(z)\,\rangle := \frac{1}{2\pi}\int_0^{2\pi} \operatorname{tr}\{G^*(e^{i\omega})H(e^{i\omega})\}\, d\omega < \infty.$$

The $\mathcal{H}_2$ norm is defined as $\|G(z)\|_2 = \sqrt{\langle\, G(z), G(z)\,\rangle}$.

Similarly, since a standard Z-transform is used, the $\mathcal{H}_2$ space is defined to be analytic outside the unit disc. The $\mathcal{H}_2$ norm of a rational system can also be calculated from the observability or controllability Gramians. Suppose $G(z)$ has a realization (A, B, C, D). Recall that for asymptotically stable systems, the observability Gramian L_o and the controllability Gramian L_c are defined as

$$L_o := \sum_{t=0}^{\infty} A^* C^* C A \geq 0,$$

$$L_c := \sum_{t=0}^{\infty} A B B^* A^* \geq 0,$$

which are also the solutions of the Lyapunov equations:

$$A^* L_o A - L_o + C^* C = 0, \tag{2.3a}$$

$$A L_c A^* - L_c + B B^* = 0. \tag{2.3b}$$

The $\mathcal{H}_2$ norm of $G(z)$ can be computed as

$$\|G(z)\|_2^2 = \operatorname{tr}\{D^* D + B^* L_o B\} = \operatorname{tr}\{D D^* + C L_c C^*\}.$$

2.4 Correlations and Power Spectra

Correlations and power spectra are used to identify systems in the frequency domain. Approaches of this kind are known as correlation and spectral analysis [Lju87]. It is a very natural way to model an unknown system as the analysis of the dynamics (or the hidden layer of the state equation) is avoided.

Definition 2.5. The autocorrelation $R_{uu}(\tau)$ and the cross-correlation $R_{yu}(\tau)$ of signals u and y are defined as:

$$R_{uu}(\tau) := \lim_{T\to\infty} \frac{1}{2T} \sum_{t=-T}^{T-1} \mathbf{E}\left(u(t+\tau)\, u(t)^*\right),$$

$$R_{yu}(\tau) := \lim_{T\to\infty} \frac{1}{2T} \sum_{t=-T}^{T-1} \mathbf{E}\left(y(t+\tau)\, u(t)^*\right),$$

if the limits exist for all $\tau \in \mathbb{Z}$.

Here we extend the use of $\mathbf{E}$ in the definition of correlation functions to denote the element-wise inner product of signals that are in an arbitrary Hilbert space, as mentioned at the end of Section 2.2.1.

Definition 2.6. The power spectrum $\Phi_{uu}(e^{i\omega})$ and the cross-power spectrum $\Phi_{yu}(e^{i\omega})$ of signals u and y are the discrete Fourier transforms of $R_{uu}(\tau)$ and $R_{yu}(\tau)$, respectively:

$$\Phi_{uu}(e^{i\omega}) := \sum_{\tau=-\infty}^{\infty} R_{uu}(\tau)e^{-i\tau\omega},$$
$$\Phi_{yu}(e^{i\omega}) := \sum_{\tau=-\infty}^{\infty} R_{yu}(\tau)e^{-i\tau\omega},$$

if the series exist (and may contain impulses) for all $\omega \in [0, 2\pi]$.

We will call a signal with finite autocorrelation a power signal, where the power is defined as:

$$\|u\|_{\mathcal{P}}^2 = \operatorname{tr}[R_{uu}(0)] = \frac{1}{2\pi}\int_0^{2\pi} \operatorname{tr}[\Phi_{uu}(e^{i\omega})]d\omega.$$

On the other hand, we will call a signal with finite spectral density norm a spectral signal, where the spectral density norm is defined as:

$$\|u\|_S^2 := \|\Phi_{uu}(e^{i\omega})\|_\infty.$$

Suppose $G(z)$, with convolution kernel $g(t)$, has input $u(t)$ and output $y(t)$. Then, $R_{yu}(\tau) = g(\tau) * R_{uu}$. Furthermore, $R_{yy}(\tau) = g(\tau) * R_{uu} * g^*(\tau)$. Taking Fourier transform, we have

$$\Phi_{yu}(e^{i\omega}) = G(e^{i\omega})\Phi_{uu}(e^{i\omega}), \tag{2.4a}$$
$$\Phi_{yy}(e^{i\omega}) = G(e^{i\omega})\Phi_{uu}(e^{i\omega})G^*(e^{i\omega}). \tag{2.4b}$$

The $\mathcal{H}_\infty$ and $\mathcal{H}_2$ norms can be interpreted as the induced system gains between these classes of signals. $\|G(z)\|_\infty$ can be regarded as the induced system gain from spectral signals to spectral signals, as well as from power signals to power signals. On the other hand, $\|G(z)\|_2$ can be regarded as the induced system gain from power signals to spectral signals (see [Mac89, ZDG96]).

2.5 Positive Real Functions

Positive real systems form an important class of functions in stochastic realization theory. The autocorrelation of a process is measured and used to construct a stochastic realization. The fact that the autocorrelation sequence is a positive real sequence implies that the stochastic realization must be positive real, where positive real is defined as follows:

Definition 2.7. A square asymptotically stable rational matrix function $G(z)$ is said to be positive real (or strictly positive real) if

$$G(e^{i\omega}) + G^*(e^{i\omega}) \geq (or >)\, 0,$$

for all $\omega \in [0, 2\pi]$.

In the following theorem, a state-space characterization of positive real functions is given (see for example [ZDG96]).

Theorem 2.2. *Suppose (A, B, C, D) is a minimal realization of a square asymptotically stable rational matrix function $G(z)$. Then the following are equivalent:*

1. *$G(z)$ is positive real;*
2. *there exist $P \geq 0$, Q, and R such that*

$$\begin{aligned} P - APA^* &= QQ^*, \\ B - APC^* &= QR^*, \\ D + D^* - CPC^* &= RR^*. \end{aligned}$$

3. *$G(z) + G^{\sim}(z) = M(z)M^{\sim}(z)$, where $M(z)$ has a realization (A, Q, C, R).*

Furthermore, $G(z)$ is strictly positive real iff $M(e^{i\omega})$ given above has full row rank for all $\omega \in [0, 2\pi]$.

Note that there are several ways to factorize a strictly positive real function. One useful factorization for system identification is to choose $M(z)$ to be a minimum-phase function such that all its zeros are in the open unit disc.

3

Review of Subspace Identification Theory

3.1 Introduction

Subspace identification methods have been attracting a lot of attention in the field of system identification. Such methods, which are based on concepts from system theory, linear algebra, and statistics, differ from the classical identification methods in several ways. In classical identification algorithms, an initial estimate and a certain parametrization of the models are required. Choosing a good initial estimate or a parametrization with a suitable model order and structure may not be straightforward; and improper choices of these parameters may lead to numerically ill-conditioned problems. Furthermore, classical identification algorithms usually require an iterative procedure in determining a model, raising the problem of convergence as a result. With subspace methods, an explicit parametrization is avoided by using a full set of state-space models. These methods, exploiting the relationship among the spaces of input, state, and output sequences, determine the system matrices by using primarily RQ and singular value decompositions. Together with their non-iterative property, subspace methods are numerically robust and do not have any convergence problems. Also since these methods are non-iterative, faster algorithms are generally obtained. One interesting feature of subspace identification approach is that state sequences are identified first, and then the actual model; whereas in the classical approach, the actual model is determined directly from input-output data.

Present subspace identification theory is a result derived and generalized from classical realization and system theory. Closely related pioneering work, studying system theory and minimal dimension realizations based on input-output observations in a deterministic setup, was first carried out in [Gop69, Bud71, LS77]. Meanwhile, new development in stochastic system theory was reported in [Aka74, Aka75], in which canonical correlation analysis was investigated. Balanced stochastic realizations and model reduction were also introduced in [DP84, DPK85]; while further progress on stochastic identification could be found in [Aok90]. Comparison of several identification approaches, known as principal component analysis, canonical correlation analysis, and unweighted principal component analysis, was made in [AK90].

A subspace identification method, using canonical variate analysis (CVA), an analysis of canonical correlations and variables, was first advocated in [Lar83] for combined deterministic-stochastic systems. Since then, research in this field was developing rapidly, at a particularly fast pace in the past few years. Recent influential work on deterministic subspace identification has been done in [MDMVV89, MDMRT92, VD92a, VD92b, MR93]. Other progress on development and analysis was made in [Lar90, Lju91, VD91, VODM91, VOWL91, VOWL93]. Instrumental variable based algorithms for identifying stochastic and combined deterministic-stochastic systems were also introduced in [Ver93b, Ver94], whose algorithms were known as MOESP algorithms, and in [VODM93, VODM94], whose algorithms were known as N4SID algorithms. Moreover, related ideas using subspace fitting techniques and applications to sensor array signal processing were investigated in [VO91, OVK92, SROK95]. While these methods were mostly designed for identification from data in the time domain, subspace identification methods from frequency response data were also developed in [MAL96a, MAL96b].

Analysis and improvements, from both algorithmic and numerical perspectives, have also been made. These include computational improvements in [CXK94a, CXK94b, CK95], pole estimation analyses in [OV94, WJ94, VWO95, JW96], weighting functions selections in [JW95, VODM95a, VODM95b], statistical analyses regarding consistency and efficiency in [DPS95, PSD96], and identification from closed-loop data in [Ver93a, LM96]. Extensions of subspace methods to identification of systems other than discrete-time linear time-invariant (LTI) systems were given in a number of papers, for instance [VY95, FDMVO96, VODM96a, VW96, WV96], in which identification methods for several different types of systems, including time-varying, continuous-time, LTI with static nonlinearity, and bilinear ones, were developed. To this end, a good overview on the subject could be found in [Vib95].

In this chapter, a review of subspace identification theory is presented. In Sections 3.2–3.4, system definitions and identification conditions are given. One basic assumption is that the system can be decomposed into two uncorrelated subsystems, namely a deterministic subsystem and a stochastic subsystem (see the following sections for detail). Then in Sections 3.5 and 3.6, subspace projection theory and identification algorithms are presented in the framework of [VODM93, VODM94]. Finally, Section 3.7 contains a brief review on a closely related subject, namely model reduction theory.

3.2 System Description

In this thesis, standard system definitions and identification conditions, similar to those in [VODM96b], are adopted. First of all, this system is assumed to be separable into two subsystems as follows:

$$\begin{aligned} x(t) &= x^d(t) + x^s(t), \\ y(t) &= y^d(t) + y^s(t). \end{aligned}$$

These two subsystems will be called the deterministic subsystem, denoted with a superscript d and the stochastic subsystem, denoted with a superscript s. The deterministic subsystem, describing the behavior due to the input u, has the state-space equations:

$$x^d(t+1) = Ax^d(t) + Bu(t), \tag{3.1a}$$
$$y^d(t) = Cx^d(t) + Du(t); \tag{3.1b}$$

whereas the stochastic subsystem, describing the behavior due to the process and output noises w and v, has the state-space equations:

$$x^s(t+1) = Ax^s(t) + w(t), \tag{3.2a}$$
$$y^s(t) = Cx^s(t) + v(t). \tag{3.2b}$$

It can be checked without difficulty that these two subsystems do sum up to the system (2.1). In a deterministic setup, Equations (3.1) alone are adequate to describe the dynamics of the system, while in a stochastic setup, Equations (3.2) become the only dynamics of the system.

3.2.1 A Stationary Stochastic Subsystem

Further assumptions on the stochastic subsystem, regarding its stationariness, are adopted. The process and output noises w and v are assumed to be stationary in the following sense:

$$\mathbf{E}\left[\begin{pmatrix} w(t) \\ v(t) \end{pmatrix} \begin{pmatrix} w(\tau) \\ v(\tau) \end{pmatrix}^*\right] =: \begin{pmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{pmatrix} \delta_{t\tau}, \tag{3.3}$$

for all $t, \tau \in \mathbb{Z}_+$, where δ denotes the Kronecker delta. Here, Σ^w, Σ^v, Σ^{wv}, and Σ^{vw} are real constant matrices with obvious dimensions. The stochastic process x^s is also assumed to be stationary such that

$$\begin{array}{rcl} \mathbf{E}\big(x^s(t)\,[x^s(t)]^*\big) & =: & \Sigma^s, \\ \mathbf{E}\big(x^s(t)\;[w(\tau)]^*\big) & = & 0, \\ \mathbf{E}\big(x^s(t)\;[v(\tau)]^*\,\big) & = & 0, \end{array}$$

for all $\tau \geq t$, $t, \tau, \in \mathbb{Z}_+$, where Σ^s is also a real constant matrix. With these stationary properties, simple derivation gives

$$\begin{aligned} \Sigma^s &:= \mathbf{E}\big(x^s(t)\,[x^s(t)]^*\big) \\ &= \mathbf{E}\big(\,[Ax^s(t) + w(t)]\;[Ax^s(t) + w(t)]^*\,\big) \\ &= A\mathbf{E}\big(x^s(t)[x^s(t)]^*\big)A^* + \mathbf{E}\big(w(t)[w(t)]^*\big) \\ &= A\Sigma^s A^* + \Sigma^w. \end{aligned} \tag{3.4}$$

In addition, define the output covariance matrices as:

$$\Lambda_i := \mathbf{E}\big(y^s(t+i)\,[y^s(t)]^*\big). \tag{3.5}$$

In particular, Λ_0 can be written as:

$$\begin{aligned}\Lambda_0 &= \mathbf{E}\big(y^s(t)\,[y^s(t)]^*\big)\\ &= \mathbf{E}\big(\,[Cx^s(t)+v(t)]\;[Cx^s(t)+v(t)]^*\,\big)\\ &= C\mathbf{E}\big(x^s(t)\,[x^s(t)]^*\big)C^* + \mathbf{E}\big(v(t)[v(t)]^*\big)\\ &= C\Sigma^s C^* + \Sigma^v.\end{aligned}$$

Moreover, define G as:

$$\begin{aligned}G &:= \mathbf{E}\big(x^s(t+1)\,[y^s(t)]^*\big)\\ &= \mathbf{E}\big(\,[Ax^s(t)+w(t)]\;[Cx^s(t)+v(t)]^*\,\big)\\ &= A\mathbf{E}\big(x^s(t)\,[x^s(t)]^*\big)C^* + \mathbf{E}\big(w(t)[v(t)]^*\big)\\ &= A\Sigma^s C^* + \Sigma^{wv}.\end{aligned} \tag{3.6}$$

Then, it is easy to see that

$$\Lambda_i = CA^{i-1}G, \tag{3.7a}$$

$$\Lambda_{-i} = G^*(A^{i-1})^*C^*. \tag{3.7b}$$

These output covariance matrices are in fact analogous to the Markov parameters of a deterministic system with system matrices (A, G, C, Λ_0). Finally, we shall assume that the stochastic subsystem is uncorrelated with the deterministic subsystem. More precisely, we have

$$\mathbf{E}\left[\begin{pmatrix} x^d(0) \\ u(t)\end{pmatrix}\begin{pmatrix}[x^s(0)]^* & [w(\tau)]^* & [v(\tau)]^*\end{pmatrix}\right] = 0,$$

for all $t, \tau \in \mathbb{Z}_+$. This guarantees the deterministic states and outputs are uncorrelated with the stochastic states and outputs at any time. To this end, it should be pointed out that the system (2.1) is required to be minimal in the sense that the pair (A, C) is observable and the pair $(A\,,\,[B\ \ (\Sigma^w)^{\frac{1}{2}}])$ is controllable.

3.2.2 Positive Realness

One property of a stochastic system is its positive realness, which is defined in Chapter 2. Define the Toeplitz matrix L_i as:

$$L_i := \begin{bmatrix} \Lambda_0 & \Lambda_{-1} & \cdots & \Lambda_{1-i} \\ \Lambda_1 & \Lambda_0 & \ddots & \Lambda_{2-i} \\ \vdots & \ddots & \ddots & \vdots \\ \Lambda_{i-1} & \Lambda_{i-2} & \cdots & \Lambda_0 \end{bmatrix},$$

where each Λ_j is defined as in Equations (3.7). It is well known that the system (A, G, C, Λ_0) is positive real, i.e.,

$$C(zI - A)^{-1}G + \Lambda_0 + G^*(z^{-1}I - A^*)^{-1}C^* \geq 0, \qquad \forall |z| = 1,$$

if and only if L_i is positive semidefinite for all $i \in \mathbb{N}$ (see for example [LP96]). To facilitate the analysis, we shall further assume that the covariance of the process and output noises is positive definite, namely,

$$\begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} > 0. \tag{3.8}$$

One nice property of the above condition is that it can ensure the positive definiteness of L_i for all $i \in \mathbb{N}$ (see for example [Fau76]). Moreover, this condition also guarantees that the system (A, G, C, Λ_0) is strictly positive real. In such a case, the Markov parameter sequence of this system can also be generated as the output covariance sequence of the following stationary stochastic system:

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + Ke(t), \\ y(t) &= C\hat{x}(t) + e(t), \end{aligned}$$

where

$$\begin{aligned} K &= (G - APC^*)(\Lambda_0 - CPC^*)^{-1}, \\ P &= APA^* + (G - APC^*)(\Lambda_0 - CPC^*)^{-1}(G - APC^*)^*, \end{aligned}$$

and for all $t, \tau \in \mathbb{Z}_+$,

$$\begin{aligned} \mathbf{E}\big(e(t)\,[e(\tau)]^*\big) &= (\Lambda_0 - CPC^*)\delta_{t\tau}, \\ \mathbf{E}\big(\hat{x}(t)\,[\hat{x}(t)]^*\big) &= P. \end{aligned}$$

This stochastic system is minimum-phase, i.e. all zeros of $M(z) := C(zI - A)^{-1}K + I$ are inside the unit circle, or equivalently $M^{-1}(z)$ is stable. Here $\hat{x}(t)$ is called the Kalman filter state, K the Kalman filter gain and P the forward state covariance matrix.

3.3 Block Data Equations

In subspace identification methods, input and output observations are stored in matrices in a Hankel form. Given an input-output signal pair $\mathscr{Z}^N$ as in Equations (2.2), the $\mathscr{X}$-valued block Hankel matrix U^l for the input signal u is defined as:

$$U^l := \begin{bmatrix} u(0) & u(1) & \cdots & u(q-1) \\ u(1) & u(2) & \cdots & u(q) \\ \iddots & \iddots & \iddots & \iddots \\ u(l-1) & u(l) & \cdots & u(l+q-2) \end{bmatrix}.$$

Note that the superscript l denotes the number of block rows in U^l whereas q denotes the number of columns in U^l. Since the input signal u has a length of N, this implies $q = N-l+1$. Similarly, we shall denote by Y^l, X^l, W^l, and V^l the block Hankel matrices constructed by signals y, x, w and v in an analogous way. Furthermore, an additional subscript i is used to denote the i-th block row as follows:

$$U_i^l := \begin{bmatrix} u(i) & u(i+1) & \cdots & u(i+q-1) \end{bmatrix}. \tag{3.9}$$

To indicate multiple block rows, we write

$$U_{i,j}^l := \begin{bmatrix} u(i) & u(i+1) & \cdots & u(i+q-1) \\ u(i+1) & u(i+2) & \cdots & u(i+q) \\ \iddots & \iddots & \iddots & \iddots \\ u(j) & u(j+1) & \cdots & u(j+q-1) \end{bmatrix}.$$

Thus, it is easy to see that $U_{0,l-1}^l = U^l$. In many occasions in this thesis, the number of block rows l used to construct U is not of particular importance. Under such circumstances, we will drop the superscript l and write U instead of U^l in order to simplify the notation without causing any confusion. For instance, U_i^l in Equation (3.9) will be denoted by U_i when no ambiguity will be caused. Similar remarks also apply to block Hankel matrices Y^l, X^l, W^l, and V^l.

We introduce a few more matrices. Define the observability matrix $\mathcal{O}_i$ and the controllability matrix $\mathcal{C}_i^d$ as:

$$\mathcal{O}_i := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{i-1} \end{bmatrix}, \qquad \mathcal{C}_i^d := \begin{bmatrix} A^{i-1}B & A^{i-2}B & \cdots & B \end{bmatrix}.$$

Furthermore, define the Toeplitz matrices $\mathcal{T}_i$ and $\mathcal{N}_i$ as:

$$\mathcal{T}_i := \begin{bmatrix} D & & & 0 \\ CB & D & & \\ \vdots & \ddots & \ddots & \\ CA^{i-2}B & \cdots & CB & D \end{bmatrix}, \qquad \mathcal{N}_i := \begin{bmatrix} 0 & & & 0 \\ C & 0 & & \\ \vdots & \ddots & \ddots & \\ CA^{i-2} & \cdots & C & 0 \end{bmatrix}.$$

Finally, define $\mathcal{A}_i$ and $\mathcal{C}_i^s$ as

$$\mathcal{A}_i := \begin{bmatrix} A^{i-1} & A^{i-2} & \cdots & I \end{bmatrix}, \qquad \mathcal{C}_i^s := \begin{bmatrix} A^{i-1}G & A^{i-2}G & \cdots & G \end{bmatrix}.$$

With these matrices and the block Hankel data matrices defined above, it is easy to obtain the following blick data equation:

$$Y^l = \mathcal{O}_l X_0^l + \mathcal{T}_l U^l + \mathcal{N}_l W^l + V^l.$$

Here all matrices have l block rows, except the state signal X_0^l which has only 1 block row. Furthermore, the state sequences X_0^l and X_i^l can be related in the following state sequence equation:

$$X_i^l = A^i X_0^l + \mathcal{C}_i^d U_{0,i-1}^l + \mathcal{A}_i W_{0,i-1}^l.$$

Note that so long as all of these data matrices are defined, the two equations above hold regardless of the value of l. This explains why the superscript l is not of particular importance in some cases.

3.4 Spaces and Projections of Signal Sequences

As hinted by its name, the key elements in subspace identification theory are subspaces generated by signal sequences. In this thesis an ideal setup (identical to that in [PK96a, PK96b], for instance) is adopted in such a way that these subspaces are elements in a Hilbert space. In most cases, this Hilbert space will be interpreted without loss of generality as an ideal probability space, spanned by finite variance random variables.

To begin with, let $+$ and $\oplus$ be the sum and the direct sum of vector subspaces in an ambient vector space, respectively. Define the space $\mathcal{U}_i^l$ spanned by the m rows of U_i^l as:

$$\mathcal{U}_i^l := \operatorname{span}\{ \alpha^* U_i^l \mid \alpha \in \mathbb{R}^m \} \subset \mathscr{X}^q,$$

where $q = N - l + 1$. Note that since each row of U_i^l is a sequence of length q, it is easy to see that the space spanned by these sequences will then be a subset in $\mathscr{X}^q$. Furthermore, the notation $\mathcal{U}_{i,j}^l$ is used to denote the space spanned by multiple block rows; that is,

$$\mathcal{U}_{i,j}^l := \mathcal{U}_i^l + \mathcal{U}_{i+1}^l + \cdots + \mathcal{U}_j^l \subset \mathscr{X}^q.$$

In addition, we define $\mathcal{U}^l$ in an obvious way, namely $\mathcal{U}^l := \mathcal{U}_{0,l-1}^l$. Similarly, the spaces spanned by the row sequences of Y^l, X^l, W^l, and V^l are respectively denoted by $\mathcal{Y}^l$, $\mathcal{X}^l$, $\mathcal{W}^l$, and $\mathcal{V}^l$ with appropriate subscripts specifying the particular rows involved in constructing the spaces. Consider that if a specific number of block rows l is chosen, the number of columns $q = N - l + 1$ will then be fixed; in which case, it is well understood that the ambient space concerned is always $\mathscr{X}^q$. Thus, when the value l plays no important role, the superscript l will be dropped to simplify the notation without causing any confusion.

In this thesis, a commonly used notation, Π, is adopted to denote the orthogonal projection operator. More specifically, we write $\Pi_{\mathcal{U}}\, \nu$ to denote the orthogonal projection of ν onto $\mathcal{U}$. Moreover, $\Pi_{\mathcal{U}}\mathcal{Y}$ will be the orthogonal projection of $\mathcal{Y}$ onto $\mathcal{U}$, satisfying

$$\Pi_{\mathcal{U}}\mathcal{Y} \subset \mathcal{U} \qquad \text{and} \qquad \langle \nu - \Pi_{\mathcal{U}}\, \nu \,, \mu \rangle = 0,$$

for any $\mu \in \mathcal{U}$ and $\nu \in \mathcal{Y}$. In particular, we will use a more compact form to denote the projection of multiple vectors, stacked in a column matrix form, on a space in the following way.

For instance, we will write $\Pi_{\mathcal{U}}Y$ to denote the projection of each row of Y onto $\mathcal{U} \subset \mathscr{X}^q$, that is,

$$\Pi_{\mathcal{U}}Y := \begin{bmatrix} \Pi_{\mathcal{U}}Y_0 \\ \Pi_{\mathcal{U}}Y_1 \\ \vdots \\ \Pi_{\mathcal{U}}Y_{l-1} \end{bmatrix}, \qquad \text{where} \qquad Y = \begin{bmatrix} Y_0 \\ Y_1 \\ \vdots \\ Y_{l-1} \end{bmatrix}.$$

Note that $\mathcal{U} \subset \mathscr{X}^q$ and each row of Y can be envisaged as a vector in $\mathscr{X}^q$. Thus, $\Pi_{\mathcal{U}}Y$ will then be a column vector containing lp $\mathscr{X}^q$-valued entries.

To this end, it should be pointed out that there are several other commonly used notations for orthogonal projection in the identification literature. For example, in [AK90, VODM96b] the above projection of Y onto $\mathcal{U}$ is written as Y/U, in [PK96b] as $\mathbf{E}[\,Y\,|\,\mathcal{U}\,]$, and in [McK95, Vib95] as $Y\Pi_U$. In the last case, the term Π_U indeed has a particular meaning as follows. Consider the case where $\mathscr{X}^q = \mathbb{R}^q$, then $Y\Pi_U$ actually equals $YU^*(UU^*)^{\dagger}U$. In other words, Π_U has a matrix representation $\Pi_U = U^*(UU^*)^{\dagger}U$.

Furthermore, consider the expression $\mathbf{E}[\,\mathcal{O}_l X_0\,|\,\mathcal{U}\,]$, where $\mathcal{O}_l$ is a real valued matrix. From this notation, it is easy to see that the multiplication commutes with the orthogonal projection, i.e. $\mathbf{E}[\,\mathcal{O}_l X_0\,|\,\mathcal{U}\,] = \mathcal{O}_l\,\mathbf{E}[\,X_0\,|\,\mathcal{U}\,]$. That is also to say that $\Pi_{\mathcal{U}}\mathcal{O}_l X_0 = \mathcal{O}_l\,\Pi_{\mathcal{U}}X_0$. In other words, the projection operator commutes with a real matrix multiplication as shown above. Finally, the symbol $\mathcal{U}^{\perp}$ will denote the orthogonal complement of the subspace $\mathcal{U}$ with respect to some predefined ambient space, usually $\mathscr{X}^q$.

3.5 Projection Theorems

In subspace identification methods, Kalman filter state sequences are directly identified from input-output observations as a first step, from which system matrices are then determined. To estimate Kalman filter state sequences, projections of certain input and output sequences are evaluated. One important feature of this identification scheme is that the stochastic system matrices can be estimated in an elegant way from these Kalman filter state sequences. To illustrate this, we first partition the data matrices U^{k+l}, as defined in Section 3.3, into two block matrices as follows:

$$\left[\begin{array}{c} U_p \\ \hdashline U_f \end{array}\right] := \left[\begin{array}{cccc} u(0) & u(1) & \cdots & u(q-1) \\ \iddots & \iddots & \iddots & \iddots \\ u(k-1) & u(k) & \cdots & u(k+q-2) \\ \hdashline u(k) & u(k+1) & \cdots & u(k+q-1) \\ \iddots & \iddots & \iddots & \iddots \\ u(k+l-1) & u(k+l) & \cdots & u(k+l+q-2) \end{array}\right].$$

Here, the suffices p and f are supposed to be mnemonic, representing 'past' and 'future', respectively. Define Y_p, Y_f, W_p, W_f, V_p, and V_f in a similar way. Meanwhile, define X_p and X_f

slightly differently as:

$$X_p := X_0^{k+l}, \qquad \text{and} \qquad X_f := X_k^{k+l},$$

such that each is a sequence of one block row. Moreover, note that U_p has k block rows and U_f has l block rows. As noted in [WJ94, JW95], k can be interpreted as the observability horizon and l as the prediction horizon. On the other hand, in [DPS95, PSD96] this value k is called the truncation index. In most subspace algorithms, k and l are restricted to be equal, i.e. $k = l$.

As the system can be decomposed into a deterministic subsystem and a stochastic subsystem, we denote by X_p^d and X_p^s the past state sequence of the deterministic subsystem and that of the stochastic subsystem, respectively. Furthermore, we apply this notation analogously to X_f, Y_p, and Y_f. Therefore, we have

$$\begin{aligned} X_f^d &= A^k X_p^d + \mathcal{C}_k^d U_p, & X_f^s &= A^k X_p^s + \mathcal{A}_k W_p; \\ Y_p^d &= \mathcal{O}_k X_p^d + \mathcal{T}_k U_p, & Y_p^s &= \mathcal{O}_k X_p^s + \mathcal{N}_k W_p + V_p; \\ Y_f^d &= \mathcal{O}_l X_f^d + \mathcal{T}_l U_f, & Y_f^s &= \mathcal{O}_l X_f^s + \mathcal{N}_l W_f + V_f. \end{aligned}$$

In the following subsections, we cover a few projection theorems for a deterministic system as defined in Equations (3.1), a stochastic system as defined in Equations (3.2), and then for a combined deterministic-stochastic system defined as in Equations (2.1).

3.5.1 Deterministic Systems

For deterministic systems, we have $\Sigma^w = 0$ and $\Sigma^v = 0$. Thus, the state sequence and the block data equations become:

$$X_f = A^k X_p + \mathcal{C}_k^d U_p, \tag{3.10}$$

$$Y_p = \mathcal{O}_k X_p + \mathcal{T}_k U_p, \tag{3.11}$$

$$Y_f = \mathcal{O}_l X_f + \mathcal{T}_l U_f; \tag{3.12}$$

which imply

$$\mathcal{X}_f \subset \mathcal{X}_p + \mathcal{U}_p, \qquad \mathcal{Y}_p \subset \mathcal{X}_p + \mathcal{U}_p, \qquad \mathcal{Y}_f \subset \mathcal{X}_f + \mathcal{U}_f.$$

From these inclusions, it is easy to see that $\mathcal{Y}_f \subset \mathcal{X}_p + \mathcal{U}_p + \mathcal{U}_f$. Observe that $\mathcal{O}_k X_p = Y_p - \mathcal{T}_k U_p$. Suppose further that $\mathcal{O}_k$ is injective, then $\mathcal{X}_p \subset \mathcal{Y}_p + \mathcal{U}_p$. Adding $\mathcal{U}_p$ to both sides of $\mathcal{Y}_p \subset \mathcal{X}_p + \mathcal{U}_p$ and $\mathcal{X}_p \subset \mathcal{Y}_p + \mathcal{U}_p$ yields

$$\mathcal{Y}_p + \mathcal{U}_p = \mathcal{X}_p + \mathcal{U}_p. \tag{3.13}$$

Now, let $\mathcal{S} := \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$. With the above equation, we then have $\mathcal{S} = \mathcal{X}_p + \mathcal{U}_p + \mathcal{U}_f$, which implies $\mathcal{Y}_f \subset \mathcal{S}$. In other words, for $\mathcal{O}_l$ injective this implies

$$\Pi_{\mathcal{S}} Y_f = \mathcal{O}_l X_f + \mathcal{T}_l U_f. \tag{3.14}$$

Putting this into a Kalman filter framework, it can be said that $\hat{X}_f = \Pi_{\mathcal{S}} X_f = X_f$ is the Kalman filter state sequence of the following Kalman filter equations:

$$\begin{aligned}\hat{X}_{t+1} &= A\hat{X}_t + BU_t + K_t(Y_t - C\hat{X}_t - DU_t);\\ Y_t &= C\hat{X}_t + DU_t,\end{aligned}$$

with the Kalman filter gain $K_t := 0$ and the initial condition $\hat{X}_0 := X_p$.

3.5.2 Stochastic Systems

For a stochastic system (3.2), the only observation is the output signal y. Thus, the problem is to find the system matrices (A, G, C, Λ_0) such that the second order statistics (the covariance matrices) of the output of the model are identical to those of the output signal y. First of all, we derive an explicit formula for a bank of t-step ahead non-steady state Kalman filters. Note that as discussed in Section 3.2.2, Inequality (3.8) can guarantee the positive definiteness of L_t for all $t \in \mathbb{N}$. Thus, in the following theorem we adopt this condition to ensure the existence of L_t^{-1}.

Theorem 3.1. ([VODM93]) *Given that Inequality* (3.8) *holds, and given*

1. *the initial state estimate* $\hat{x}(0) = 0$*;*
2. *the initial covariance of the state estimate* $P_0 := \mathbf{E}\big(\hat{x}(0)[\hat{x}(0)]^*\big) = 0$*;*

then the non-steady state Kalman filter state estimate $\hat{x}(t)$ *defined by the following recursive formulas*

$$\hat{x}(t) = A\hat{x}(t-1) + K_{t-1}(y(t-1) - C\hat{x}(t-1)), \tag{3.15a}$$

$$K_{t-1} = (G - AP_{t-1}C^*)(\Lambda_0 - CP_{t-1}C^*)^{-1}, \tag{3.15b}$$

$$P_t = AP_{t-1}A^* + (G - AP_{t-1}C^*)(\Lambda_0 - CP_{t-1}C^*)^{-1}(G - AP_{t-1}C^*)^*, \tag{3.15c}$$

can be written explicitly as:

$$\hat{x}(t) = \mathcal{C}_t^s L_t^{-1} \begin{pmatrix} y(0) \\ y(1) \\ \vdots \\ y(t-1) \end{pmatrix}.$$

The explicit solution of the covariance matrix P_t *is equal to*

$$P_t = \mathcal{C}_t^s L_t^{-1} (\mathcal{C}_t^s)^*.$$

From the above theorem, it can be seen that a t-step ahead Kalman filter state can be written explicitly as a linear combination of the output signal y, provided that $\hat{x}(0) = 0$ and $P_0 = 0$. Now, suppose that we run q Kalman filters in parallel, each starting with a zero initial state estimate:

$$\hat{X}_p := \begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Here we shall define

$$P_0 := \frac{1}{q}\mathbf{E}\big(\hat{X}_p \hat{X}_p^*\big) = 0.$$

With $\hat{X}_p := \hat{X}_0$, Equations (3.15) then become

$$\hat{X}_t = A\hat{X}_{t-1} + K_{t-1}(Y_{t-1} - C\hat{X}_{t-1}), \tag{3.16a}$$

$$K_{t-1} = (G - AP_{t-1}C^*)(\Lambda_0 - CP_{t-1}C^*)^{-1}, \tag{3.16b}$$

$$P_t = AP_{t-1}A^* + (G - AP_{t-1}C^*)(\Lambda_0 - CP_{t-1}C^*)^{-1}(G - AP_{t-1}C^*)^*, \tag{3.16c}$$

where X_t is the sequence of q t-step ahead Kalman filter state estimates. Furthermore, by Theorem 3.1 the k-step ahead Kalman filter state sequence $\hat{X}_f := \hat{X}_k$ can be written explicitly as:

$$\hat{X}_f = \mathcal{C}_k^s L_k^{-1} Y_p.$$

In effect, $\hat{X}_f$ can be perceived as a Kalman filter state sequence generated by a bank of q non-steady state Kalman filters with zero initial condition and covariance. The next theorem shows further that this sequence of Kalman filter state estimates can actually be obtained by a subspace projection.

Theorem 3.2. ([VODM93]) *Given that Inequality* (3.8) *holds, then*

$$\Pi_{\mathcal{Y}_p} Y_f = \mathcal{O}_l \hat{X}_f, \tag{3.17}$$

where $\hat{X}_f$ is the sequence of q k-step ahead Kalman filter state estimates defined in Equations (3.16) *with initial conditions:*

$$\hat{X}_p = 0, \qquad P_0 = 0.$$

3.5.3 Combined Deterministic-Stochastic Systems

Similar to the previous subsection, theorems on non-steady Kalman filters and subspace projections are presented for combined deterministic-stochastic systems. Note that with Inequality (3.8), L_t can be ensured to be positive definite for all $t \in \mathbb{N}$. In other words, the quantity $(L_t + \mathcal{O}_t \tilde{P}_0 \mathcal{O}_t^*)$ is always invertible, which also implies that the following theorem is valid for all $t \in \mathbb{N}$. It is important to observe that due to the external input signal u, the non-steady Kalman filter equations are modified slightly, in which $\tilde{P}_0$ is defined as the covariance of the error $x(0) - \hat{x}(0) - x^s(0)$.

Theorem 3.3. ([VODM94]) *Given that Inequality* (3.8) *holds, and given*

1. *the initial state estimate* $\hat{x}_0$*;*
2. *the initial estimate of the matrix* $\tilde{P}_0 := \mathbf{E}\big([x(0) - \hat{x}(0) - x^s(0)]\,[x(0) - \hat{x}(0) - x^s(0)]^*\big)$*;*

then the non-steady state Kalman filter state estimate $\hat{x}(t)$ *defined by the following recursive formulas:*

$$\hat{x}(t) = A\hat{x}(t-1) + Bu(t) + K_{t-1}(y(t-1) - C\hat{x}(t-1) - Du(t-1)), \tag{3.18a}$$

$$K_{t-1} = (G + A\tilde{P}_{t-1}C^*)(\Lambda_0 + C\tilde{P}_{t-1}C^*)^{-1}, \tag{3.18b}$$

$$\tilde{P}_t = A\tilde{P}_{t-1}A^* - (G + A\tilde{P}_{t-1}C^*)(\Lambda_0 + C\tilde{P}_{t-1}C^*)^{-1}(G + A\tilde{P}_{t-1}C^*)^*, \tag{3.18c}$$

can be written explicitly as:

$$\hat{x}(t) = \begin{pmatrix} A^t - \mathcal{K}_t\mathcal{O}_t & \vdots & \mathcal{C}_t^d - \mathcal{K}_t\mathcal{T}_t & \vdots & \mathcal{K}_t \end{pmatrix} \begin{pmatrix} \hat{x}(0) \\ \hline u(0) \\ \vdots \\ u(t-1) \\ \hline y(0) \\ \vdots \\ y(t-1) \end{pmatrix},$$

where

$$\mathcal{K}_t := (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^{-1}.$$

The explicit solution of the matrix $\tilde{P}_t$ *is equal to*

$$\tilde{P}_t = A^t\tilde{P}_0(A^t)^* - (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^{-1}(\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)^*.$$

The above theorem gives an explicit solution to the t-step ahead recursive Kalman filter formulas. Now, suppose that we run q Kalman filters in parallel, in a similar way as in Equations (3.16), with $\hat{X}_p := \hat{X}_0$ containing a sequence of initial conditions. Then Equations (3.18) become

$$\hat{X}_t = A\hat{X}_{t-1} + BU_t + K_{t-1}(Y_{t-1} - C\hat{X}_{t-1} - DU_{t-1}), \tag{3.19a}$$

$$K_{t-1} = (G + A\tilde{P}_{t-1}C^*)(\Lambda_0 + C\tilde{P}_{t-1}C^*)^{-1}, \tag{3.19b}$$

$$\tilde{P}_t = A\tilde{P}_{t-1}A^* - (G + A\tilde{P}_{t-1}C^*)(\Lambda_0 + C\tilde{P}_{t-1}C^*)^{-1}(G + A\tilde{P}_{t-1}C^*)^*, \tag{3.19c}$$

where $\tilde{P}_0 := \mathbf{E}\big([X_p - \hat{X}_p - X_p^s]\,[X_p - \hat{X}_p - X_p^s]^*\big)$. Furthermore, the k-step ahead Kalman filter state sequence $\hat{X}_f := \hat{X}_k$ can be written explicitly as:

$$\hat{X}_f = \begin{pmatrix} A^k - \mathcal{K}_k\mathcal{O}_k & \vdots & \mathcal{C}_k^d - \mathcal{K}_k\mathcal{T}_k & \vdots & \mathcal{K}_k \end{pmatrix} \begin{pmatrix} \hat{X}_p \\ \hline U_p \\ \hline Y_p \end{pmatrix}.$$

In effect, $\hat{X}_f$ can be perceived as a Kalman filter state sequence generated by a bank of q non-steady state Kalman filters, working in parallel on each of the columns of $\hat{X}_p$, U_p, and Y_p. In the following, we shall prove that from a subspace projection, a Kalman filter state sequence is actually obtained.

Theorem 3.4. ([VODM94]) *Given that Inequality* (3.8) *holds, and* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$, *then*

$$\Pi_{\mathcal{S}} Y_f = \mathcal{O}_l \hat{X}_f + \mathcal{T}_l U_f, \tag{3.20}$$

where $\hat{X}_f$ *is the sequence of* q *k-step ahead Kalman filter state estimates defined in Equations* (3.19), *with initial conditions:*

$$\hat{X}_p = \Pi_{\mathcal{U}_p + \mathcal{U}_f} X_p, \qquad \tilde{P}_0 = \frac{1}{q}\mathbf{E}\big([X_p - \hat{X}_p - X_p^s]\,[X_p - \hat{X}_p - X_p^s]^*\big).$$

The above theorem is the heart of subspace identification theory. To extract the Kalman filter state sequence $\hat{X}_f$, one would project Equation (3.20) on $\mathcal{U}_f^{\perp}$ and perform an SVD. This is elaborated further in the next section.

3.6 Identification Algorithms

In this section, two identification algorithms, Algorithms 3.10 and 3.11, are presented. Algorithm 3.10 is a stochastic identification algorithm, which can be found on page 30; whereas Algorithm 3.11 is a combined deterministic-stochastic identification algorithm, which can be found on page 31. As will be seen, the key steps used in these algorithms are the subspace projections discussed in the previous section. Attention should be drawn to the fact that stochastic models $(A, G, C, \frac{\Lambda_0}{2})$ produced by these algorithms are all positive real. In the following, an ideal probabilistic setup is considered such that noises are canceled perfectly. Note that both algorithms are taken from [VODM96b]; thus, readers can refer to this reference for more detail.

First, consider a stochastic system. Suppose that $\Pi_{\mathcal{Y}_p} Y_f$ has an SVD

$$\Pi_{\mathcal{Y}_p} Y_f = \Gamma \Lambda \Omega^* = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

Here, we consider a rather abstract type of SVD's, which satisfies the following properties:

1. $\Gamma \in \mathbb{R}^{kp \times kp}$ is unitary;
2. $\Lambda \in \mathbb{R}^{kp \times q}$ is diagonal with nonincreasing, nonnegative diagonal entries;
3. $\Omega \in \mathscr{X}^{q \times q}$ satisfies $\mathbf{E}[\Omega^* \Omega] = I$.

To perform such an SVD, one can

1. find an $\tilde{\Omega} \in \mathscr{X}^{q\times q}$ such that the q row vectors ($\in \mathscr{X}^q$) of $\tilde{\Omega}^*$ are orthonormal and span the row space of $\Pi_{\mathcal{Y}_p} Y_f$ (these q vectors do not span $\mathscr{X}^q$ in general);

2. compute the real matrix T such that $\Pi_{\mathcal{Y}_p} Y_f = T\tilde{\Omega}^*$;

3. perform an SVD on T as $T = \Gamma \Lambda V^*$, then set $\Omega^* = V^* \tilde{\Omega}^*$.

In the case where real data are measured, we can consider Y_f and Y_p as two column vectors containing kp $\mathscr{X}$-valued entries (i.e. with $q = 1$). Each $\mathscr{X}$-valued element is then replaced by a sample path of an ergodic process or a set of repeated experimental data. Furthermore, we introduce a different operator $\mathbf{E}_j$ to compute the correlation in the following way

$$\mathbf{E}[Y_t Y_\tau] = \mathbf{E}_j \left[\textstyle\sum_{i=1}^{j} y^i(t)(y^i(\tau))^* \right] := \lim_{j\to\infty} \left[\frac{1}{j} \textstyle\sum_{i=1}^{j} y^i(t)(y^i(\tau))^* \right],$$

where $y^i(t)$ represents data either from time $t + i - 1$ or from the i-th experiment.

Now, by Equation (3.17) it can be seen that $\mathcal{O}_k$ and $\hat{X}_f$ can be decomposed as

$$\mathcal{O}_l = \Gamma_1 \Sigma_1^{\frac{1}{2}} \qquad \text{and} \qquad \hat{X}_k = \Sigma_1^{\frac{1}{2}} \Omega_1^* = \mathcal{O}_l^\dagger \Pi_{\mathcal{Y}_p} Y_f.$$

On the other hand, $\hat{X}_{k+1}$ can be acquired similarly by partitioning Y differently as $Y_p^+ := Y_{0,k}$ and $Y_f^- := Y_{k+1,2k-1}$. Now, since $\hat{X}_k$ and $\hat{X}_{k+1}$ are the Kalman filter state sequences, we have

$$\begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \hat{X}_k + \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix},$$

where $\hat{X}_k \perp \varepsilon_w$ and $\hat{X}_k \perp \varepsilon_v$. Consequently, solving the above equation in the least-squares sense gives a consistent estimate of A and C. Moreover, G and Λ_0 can be estimated from ε_w and ε_v. Furthermore, to obtain the Kalman filter gain K, the following equations can be solved:

$$\begin{aligned} P &= APA^* + (G - APC^*)(\Lambda_0 - CPC^*)^{-1}(G - APC^*)^*, \\ K &= (\Lambda_0 - CPC^*)^{-1}(G - APC^*). \end{aligned}$$

Again, a more complete description of the algorithm can be found in Algorithm 3.10 on page 30.

For a combined deterministic-stochastic system, a projection of $Z_k := \Pi_{\mathcal{S}} Y_f$ onto $\mathcal{U}_f^\perp$ is used to eliminate the effect of U_f. Suppose that $\Pi_{\mathcal{U}_f^\perp} Z_k$ has an SVD

$$\Pi_{\mathcal{U}_f^\perp} Z_k = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

By Equation (3.20), it can be seen that $\mathcal{O}_l$ can be decomposed as $\mathcal{O}_l = \Gamma_1 \Sigma_1^{\frac{1}{2}}$. However, the determination of the system matrices is less straightforward in this case. Let $Z_{k+1} := \Pi_{\mathcal{S}+\mathcal{Y}_k} Y_{k+1,2k-1}$ and assume $\mathcal{O}_l$ is injective. Then by Equation (3.20), we have

$$\mathcal{O}_l^\dagger Z_k = \hat{X}_k + \mathcal{O}_l^\dagger \mathcal{T}_l U_{k,k+l-1}, \qquad \text{and} \qquad \mathcal{O}_{l-1}^\dagger Z_{k+1} = \hat{X}_{k+1} + \mathcal{O}_{l-1}^\dagger \mathcal{T}_{l-1} U_{k+1,k+l-1}.$$

Grouping these terms then gives

$$\begin{bmatrix} \mathcal{O}_{l-1}^{\dagger} Z_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathcal{O}_l^{\dagger} Z_k + F(B,D) U_f + \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix},$$

where

$$F(B,D) := \begin{bmatrix} B & \mathcal{O}_{l-1}^{\dagger} \mathcal{T}_{l-1} \\ D & 0 \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \mathcal{O}_l^{\dagger} \mathcal{T}_l.$$

Since the Kalman filter state sequence $\hat{X}_k$ and the input U_f are uncorrelated with the noise ε_w and ε_v, the system matrices can be determined consistently by solving the above equation in the least-squares sense. In addition, G and Λ_0 can be estimated from ε_w and ε_v. Furthermore, to obtain the Kalman filter gain K, the following equations can be solved:

$$\begin{aligned} \tilde{P} &= A\tilde{P}A^* - (G + A\tilde{P}C^*)(\Lambda_0 + C\tilde{P}C^*)^{-1}(G + A\tilde{P}C^*)^*, \\ K &= (\Lambda_0 + C\tilde{P}C^*)^{-1}(G + A\tilde{P}C^*). \end{aligned}$$

Again, a more complete description of the algorithm can be found in Algorithm 3.11 on page 31.

Finally, it is noteworthy that since $\hat{X}_k$ and $\hat{X}_{k+1}$ are non-steady state Kalman filter state sequences, the resulting stochastic system matrices G and Λ_0 are only approximations. However, these state sequences will converge to a steady state as k increases, and hence estimates closer to the true values will be procured.

3.7 Model Reduction

After a model is determined from an identification algorithm, a reduction of the complexity of the model is the next step to be considered. A common model reduction method is by balanced truncation. In this section, concepts of balanced realizations and some of their properties are briefly reviewed.

Definition 3.1. ([Moo81]) The minimal state space realization (A, B, C, D) is said to be balanced if $L_o = L_c = diag(\sigma_1 I_{r_1}, \sigma_1 I_{r_2}, \ldots, \sigma_s I_{r_s})$, where L_o and L_c are defined in Equations (2.3).

Note that every asymptotically stable, minimal state-space system has a balanced realization. The quantities $\sigma_1, \ldots, \sigma_s$ are known as the Hankel singular values of the system, where each σ_i is assumed to have a multiplicity r_i. Thus, for an n-order system, we have $r_1 + r_2 + \cdots + r_s = n$. Suppose further that $\sigma_1 > \sigma_2 > \cdots > \sigma_s > 0$ and let

$$\Sigma := \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix},$$

where $\Sigma_1 = diag(\sigma_1 I_{r_1}, \ldots, \sigma_j I_{r_j})$ and $\Sigma_2 = diag(\sigma_{j+1} I_{r_{j+1}}, \ldots, \sigma_s I_{r_s})$. Moreover, partition A, B, C conformally with Σ as:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \qquad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, \qquad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}. \tag{3.21}$$

Then, we have the following results regarding the truncated model (A_{11}, B_1, C_1, D).

Theorem 3.5. ([PS82]) *Let (A, B, C, D) be a minimal, asymptotically stable balanced realization, then the realization (A_{11}, B_1, C_1, D) is also minimal and asymptotically stable.*

Theorem 3.6. ([HP90]) *Let (A, B, C, D) be a minimal, asymptotically stable balanced realization, and let*

$$G(z) = C(zI - A)^{-1}B + D,$$
$$\hat{G}(z) = C_1(zI - A_{11})^{-1}B_1 + D,$$

where (A_{11}, B_1, C_1) are defined by Equations (3.21), then

$$\|G(z) - \hat{G}(z)\|_\infty \leq 2(\sigma_{j+1} + \cdots + \sigma_s). \tag{3.22}$$

The above theorem indicates that the $\mathcal{H}_\infty$ error of the mismatch produced by the truncated model is bounded from above by twice the sum of the discarded singular values with multiple singular values counted only once. It should be noted that truncated discrete-time systems are not necessarily balanced in the sense of Definition 3.1. However, if the definition of balanced realizations is relaxed to realizations satisfying, for some $P = Q = diag(\sigma_1 I_{r_1}, \sigma_1 I_{r_2}, \ldots, \sigma_s I_{r_s})$,

$$A^*PA - P + C^*C \leq 0,$$
$$AQA^* - Q + BB^* \leq 0,$$

then such balancedness is preserved under truncation. It is interesting to note that using this definition of balancedness, Theorem 3.6 also holds (see for example [ZDG96]).

As indicated in [Glo84], it is possible to acquire a balanced reduced model by first converting the discrete-time model to a continuous-time model via a bilinear transformation, reducing the continuous-time model by truncation, and then converting the reduced system back to discrete-time. The reason that the reduced model is still balanced is that balancedness is preserved under a bilinear transformation and under truncation in continuous-time. Another way to reduce model complexity is to use optimal Hankel-norm approximations. The Hankel-norm of a system is defined as follows.

Definition 3.2. ([Glo84]) Let $G(z) = C(zI - A)^{-1}B$ be asymptotically stable and strictly proper, then the Hankel-norm of $G(z)$ is defined as

$$\|G(z)\|_H := \sqrt{\rho(L_o L_c)},$$

where L_o and L_c are defined in Equations (2.3).

We refer to [Glo84] for the state-space characterization of optimal Hankel-norm approximations. The next theorem shows that we in fact have a similar error bound for optimal Hankel-norm approximations.

Theorem 3.7. ([Glo84]) *Let $G(z)$ be asymptotically stable and strictly proper with Hankel singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_s$ where each σ_i has multiplicity r_i. Suppose $\hat{G}(z)$ is an asymptotically stable and strictly proper $(r_1 + r_2 + \cdots + r_j)$-th order optimal Hankel norm approximation to $G(z)$. Then,*

$$\|G(z) - \hat{G}(z)\|_\infty \leq 2(\sigma_{j+1} + \sigma_{j+2} + \cdots + \sigma_s).$$

It is interesting to note that the $\mathcal{H}_\infty$ error bound can be tightened to be simply the sum of the remaining distinct singular values by using the approximation $\hat{G}(z) + \hat{D}$, for some suitable constant matrix $\hat{D}$.

To this end, an approximate realization algorithm, namely Kung's algorithm [Kun78], is worth while to be mentioned. Let $\{h_0, h_1, \ldots, h_N, 0, \ldots\}$ be the finite impulse response of a system. Let H be a block Hankel matrix having an SVD as follows:

$$H := \begin{bmatrix} h_1 & h_2 & \cdots & h_N \\ h_2 & h_3 & \iddots & 0 \\ \vdots & \iddots & \iddots & \vdots \\ h_N & 0 & \cdots & 0 \end{bmatrix} := \begin{bmatrix} \Gamma_n & \bar{\Gamma}_n \end{bmatrix} \begin{bmatrix} \Sigma_n & 0 \\ 0 & \bar{\Sigma}_n \end{bmatrix} \begin{bmatrix} \Omega_n^* \\ \bar{\Omega}_n^* \end{bmatrix},$$

where $\Sigma_n \in \mathbb{R}^{n \times n}$. Moreover, let the rank of H be r. We further partition Γ_n and Ω_n into N equal block rows and define $\Gamma_n^\uparrow$ and $\Omega_n^\uparrow$ as:

$$\Gamma_n^* =: \begin{bmatrix} \gamma_{1n}^* & \gamma_{2n}^* & \cdots & \gamma_{Nn}^* \end{bmatrix}, \qquad (\Gamma_n^\uparrow)^* := \begin{bmatrix} \gamma_{2n}^* & \cdots & \gamma_{Nn}^* & 0 \end{bmatrix};$$
$$\Omega_n^* =: \begin{bmatrix} \omega_{1n}^* & \omega_{2n}^* & \cdots & \omega_{Nn}^* \end{bmatrix}, \qquad (\Omega_n^\uparrow)^* := \begin{bmatrix} \omega_{2n}^* & \cdots & \omega_{Nn}^* & 0 \end{bmatrix}.$$

Then, an n-th order Kung's estimate is given by

$$A_n = \hat{A}_n \text{ or } \tilde{A}_n, \qquad B_n = \Sigma_n^{\frac{1}{2}} \omega_{1n}^*, \qquad C_n = \gamma_{1n} \Sigma_n^{\frac{1}{2}}, \qquad D_n = h_0,$$

where $\hat{A}_n := \Sigma_n^{-\frac{1}{2}} \Gamma_n^* \Gamma_n^\uparrow \Sigma_n^{\frac{1}{2}}$ and $\tilde{A}_n^* := \Sigma_n^{-\frac{1}{2}} \Omega_n^* \Omega_n^\uparrow \Sigma_n^{\frac{1}{2}}$. Note that $\hat{A}_n$ is actually estimated from the observability matrix and $\tilde{A}_n$ from the controllability matrix. The following two theorems give some properties of the above Kung's estimate.

Theorem 3.8. ([Cho94]) *For all $n \leq r$, $\hat{A}_n = \tilde{A}_n$.*

Theorem 3.9. ([Cho94]) *The r-th order Kung's estimate (A_r, B_r, C_r, D) is balanced and has a finite impulse response $\{h_0, h_1, \ldots, h_N, 0, \ldots\}$. Furthermore, the n-th order $(n < r)$ Kung's estimate (A_n, B_n, C_n, D) is a discrete-time balanced truncation of (A_r, B_r, C_r, D).*

Algorithm 3.10. A Stochastic Identification Algorithm

1. *Calculate the SVD of the projection:*
$$\Pi_{\mathcal{Y}_p} Y_f = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

2. *Determine* $\mathcal{O}_l$ *as:*
$$\mathcal{O}_l = \Gamma_1 \Sigma_1^{\frac{1}{2}}.$$

3. *Determine* $\hat{X}_k$ *and* $\hat{X}_{k+1}$ *as:*
$$\hat{X}_k = \mathcal{O}_l^{\dagger}(\Pi_{\mathcal{Y}_p} Y_f), \qquad \hat{X}_{k+1} = \mathcal{O}_{l-1}^{\dagger}(\Pi_{\mathcal{Y}_p^+} Y_f^-),$$
where $Y_p^+ := Y_{0,k}$ *and* $Y_f^- := Y_{k+1,k+l-1}$.

4. *Solve A and C from:*
$$\begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \hat{X}_k + \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix}.$$

5. *Determine* Σ^w, Σ^v, Σ^{wv}, *and* Σ^{vw} *as:*
$$\begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} = \frac{1}{q}\mathbf{E}\left(\begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix} \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix}^* \right).$$

6. *Determine* Σ^s, G, *and* Λ_0 *from:*
$$\begin{aligned} \Sigma^s &= A\Sigma^s A^* + \Sigma^w, \\ G &= A\Sigma^s C^* + \Sigma^{wv}, \\ \Lambda_0 &= C\Sigma^s C^* + \Sigma^v. \end{aligned}$$

Figure 3.1 *A schematic overview of a stochastic identification algorithm [VODM96b].*

Algorithm 3.11. A Robust Combined Identification Algorithm

1. *Calculate the projections:*

$$Z_k = \Pi_{\mathcal{S}} Y_f, \qquad Z_{k+1} = \Pi_{\mathcal{S}+\mathcal{Y}_k} Y_f^-,$$

   where $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$ *and* $Y_f^- = Y_{k+1,k+l-1}$.

2. *Calculate the SVD of* $\Pi_{\mathcal{U}_f^\perp} Z_k$*:*

$$\Pi_{\mathcal{U}_f^\perp} Z_k = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

3. *Determine* $\mathcal{O}_l$ *as:*

$$\mathcal{O}_l = \Gamma_1 \Sigma_1^{\frac{1}{2}}.$$

4. *Solve* A *and* C *from:*

$$\begin{bmatrix} \mathcal{O}_{l-1}^\dagger Z_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} A \\ C \end{bmatrix} \mathcal{O}_l^\dagger Z_k + F(B,D) U_f + \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix},$$

   where

$$F(B,D) := \begin{bmatrix} B & \mathcal{O}_{l-1}^\dagger \mathcal{T}_{l-1} \\ D & 0 \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \mathcal{O}_l^\dagger \mathcal{T}_l.$$

5. *Recompute* $\mathcal{O}_l$. *Solve* B *and* D *from:*

$$B, D = arg \min_{B,D} \left\| \begin{bmatrix} \mathcal{O}_{l-1}^\dagger Z_{k+1} \\ Y_k \end{bmatrix} - \begin{bmatrix} A \\ C \end{bmatrix} \mathcal{O}_l^\dagger Z_k - F(B,D) U_f \right\|_F^2.$$

6. *Solve for* Σ^w, Σ^v, Σ^{wv}, *and* Σ^{vw} *as:*

$$\begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} = \frac{1}{q} \mathbf{E} \left(\begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix} \begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix}^* \right).$$

7. *Determine* Σ^s, G, *and* Λ_0 *as in Step 6 in Algorithm 3.10.*

Figure 3.2 *A schematic overview of a robust combined identification algorithm [VODM96b].*

4

A Geometric Interpretation for Subspace Projections

4.1 Introduction

The main purpose of this chapter is to give a geometric interpretation for subspace projections. A subspace projection is a key step in a subspace algorithm in estimating state sequences. In [VODM94], Van Overschee and De Moor gave an interpretation for subspace projections in terms of Kalman filter state estimation, using an algebraic approach, as seen in Chapter 3. They showed that a state sequence projected on a certain input-output space $(\mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f)$ was actually a state sequence estimate, running through a bank of Kalman filters, with some initial condition $(\hat{X}_p = \Pi_{\mathcal{U}_p + \mathcal{U}_f} X_p)$. In other words, this projected state sequence can be conceived as an optimal (minimum variance) estimate given this specific initial condition.

In this chapter, we consider a slightly different setup, in which some stochastic disturbances may be zero, i.e. $\Sigma^w \geq 0$, $\Sigma^v \geq 0$. The motivation for such a consideration stems from the identification of deterministic systems. In a deterministic setup, we know that the projected state sequence is the original sequence itself $(\Pi_{\mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f} X_f = X_f)$, which can also be perceived as a state sequence running through a bank of Kalman filters from the original initial state sequence $(\hat{X}_p = X_p)$ with a zero Kalman filter gain (as indicated in Equation (3.14)). Owing to this consideration, we develop a framework which generalizes both purely deterministic and fully stochastic cases to an amalgamated extension. It is then shown in Section 4.4 that the space spanned by the initial condition $\hat{X}_p$ turns out to be a combination of a subspace of $\Pi_{\mathcal{U}_p + \mathcal{U}_f} \mathcal{X}_p$, which corresponds to the stochastically related modes, and a subspace of $\mathcal{X}_p$, which corresponds to the purely deterministic modes. To be more precise, we have $\hat{X}_p = \Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p) + \mathcal{U}_p + \mathcal{U}_f} X_p$.

In Section 4.2, we perform a state and output coordinate transformation to a system according to its noise structure. The state coordinates which correspond to the purely deterministic modes in a subspace projection are clearly identified under such a transformation. In Section 4.3, we then show, with the use of Theorem 3.3, that a t-step ahead Kalman filter estimate is equivalent to a 1-step ahead t-block Kalman filter estimate. In addition, the Kalman filter corresponding to the partial stochastic excitation case, which will be used for the interpretation of a subspace projection, is presented. Finally, we state and prove, using a geometric approach, a subspace projection theorem under partial stochastic excitation in Section 4.4. Furthermore, this approach provides a much shorter and simpler proof for the theorem than that provided by the algebraic approach. Some insight acquired from this geometric approach is given in the summary.

4.2 Noise Segregation under Partial Stochastic Excitation

The objective here is to identify the purely deterministic modes and construct a Kalman filter which only updates the remaining stochastically related modes. We first consider a particular realization $(\bar{A}, \bar{B}, \bar{C}, \bar{D})$, with signals $\bar{x}(t)$, $\bar{u}(t)$, $\bar{y}(t)$ and covariances $\bar{\Sigma}^w$, $\bar{\Sigma}^v$, $\bar{\Sigma}^s$ defined in the obvious way. We assume this realization has coordinates that satisfy the following:

$$\bar{x}(t) = \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \\ \bar{x}_3(t) \end{bmatrix}, \qquad \mathbf{E}\left(\begin{bmatrix} \bar{x}_1^s(t) \\ \bar{x}_2^s(t) \\ \bar{x}_3^s(t) \end{bmatrix} \begin{bmatrix} \bar{x}_1^s(t) \\ \bar{x}_2^s(t) \\ \bar{x}_3^s(t) \end{bmatrix}^* \right) = \begin{bmatrix} \bar{\Sigma}_{11}^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \bar{\Sigma}_{11}^s > 0. \tag{4.1}$$

Thus, it can be seen that $\bar{x}_1(t)$ corresponds to the stochastically excited modes. The distinction between $\bar{x}_2(t)$ and $\bar{x}_3(t)$ will be clarified later. Now, we have the following straightforward lemma.

Lemma 4.1. *Given Equation* (4.1), *then* $\bar{A}$ *and* $\bar{\Sigma}_w$ *have the forms*

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & \bar{A}_{32} & \bar{A}_{33} \end{bmatrix}, \qquad \bar{\Sigma}^w = \begin{bmatrix} \bar{\Sigma}_{11}^w & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The proof of Lemma 4.1 can be found in Appendix A. Now, suppose further that $\bar{y}(t)$ has the coordinates and $\bar{C}$ has the form:

$$\bar{y}(t) = \begin{bmatrix} \bar{y}_1(t) \\ \bar{y}_2(t) \\ \bar{y}_3(t) \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ 0 & \bar{C}_{22} & \bar{C}_{23} \\ 0 & 0 & \bar{C}_{33} \end{bmatrix}, \qquad \bar{C}_{11}\bar{C}_{11}^* > 0, \tag{4.2}$$

such that with obvious dimensions

$$\bar{\Lambda}_0 := \mathbf{E}\big(\bar{y}^s(t)\, [\bar{y}^s(t)]^* \big) = \bar{C}\bar{\Sigma}^s\bar{C}^* + \begin{bmatrix} \bar{\Sigma}_{11}^v & \bar{\Sigma}_{12}^v & 0 \\ \bar{\Sigma}_{21}^v & \bar{\Sigma}_{22}^v & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \bar{\Sigma}_{22}^v > 0. \tag{4.3}$$

Note that $\bar{C}_{11}\bar{C}_{11}^* > 0$ implies $\bar{C}_{11}$ is of full row rank. Together with $\bar{\Sigma}_{11}^s > 0$, we have $\bar{C}_{11}\bar{\Sigma}_{11}^s\bar{C}_{11}^* > 0$. Therefore,

$$(\bar{\Lambda}_0)_{11} := \mathbf{E}\left(\begin{bmatrix} \bar{y}_1^s(t) \\ \bar{y}_2^s(t) \end{bmatrix} \begin{bmatrix} \bar{y}_1^s(t) \\ \bar{y}_2^s(t) \end{bmatrix}^* \right) = \begin{bmatrix} \bar{C}_{11}\bar{\Sigma}_{11}^s\bar{C}_{11}^* & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} \bar{\Sigma}_{11}^v & \bar{\Sigma}_{12}^v \\ \bar{\Sigma}_{21}^v & \bar{\Sigma}_{22}^v \end{bmatrix} > 0. \tag{4.4}$$

Now, we shall assume that $\bar{x}_3(t)$ corresponds to all the modes that are observable from $\bar{y}_3(t)$. The following lemma is a straightforward verification using Lemma 4.1 and Equation (4.2).

Lemma 4.2. *$\bar{x}_1(t)$ corresponds to modes that are not observable from* $\left(\begin{bmatrix}0 & 0 & \bar{C}_{33}\end{bmatrix}, \bar{A}\right)$.

Thus, $\bar{x}_1(t)$ and $\bar{x}_3(t)$ describe distinct dynamics. Moreover, $\bar{x}_2(t)$ will be assigned to be the remaining modes, defined as in Equation (4.1). Note that the proof of Lemma 4.3 can be found in Appendix A.

Lemma 4.3. *Given Equations (4.1) and (4.2), and that $\bar{x}_3(t)$ corresponds to all the modes that are observable from $\left(\begin{bmatrix}0 & 0 & \bar{C}_{33}\end{bmatrix}, \bar{A}\right)$. Then, $\bar{A}$ has the form*

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix},$$

and $(\bar{C}_{33},\ \bar{A}_{33})$ is observable.

In the following lemma, we try to reconstruct Equations (4.1) and (4.3) from an arbitrary state-space realization by state and output transformations. The proof of the lemma can be found in Appendix A.

Lemma 4.4. *For any observable state-space realization of Equations (2.1), there always exist an invertible T_s and a unitary T_o satisfying*[1]

$$\bar{x}(t) := T_s x(t) = \begin{bmatrix} T_{s1} \\ T_{s2} \\ T_{s3} \end{bmatrix} x(t) =: \begin{bmatrix} \bar{x}_1(t) \\ \bar{x}_2(t) \\ \bar{x}_3(t) \end{bmatrix}, \qquad \bar{y}(t) := T_o y(t) = \begin{bmatrix} T_{o1} \\ T_{o2} \\ T_{o3} \end{bmatrix} y(t) =: \begin{bmatrix} \bar{y}_1(t) \\ \bar{y}_2(t) \\ \bar{y}_3(t) \end{bmatrix}, \tag{4.5a}$$

and

$$\bar{\Sigma}^s := T_s \Sigma^s T_s^* = \begin{bmatrix} \bar{\Sigma}^s_{11} & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \bar{\Sigma}^v := T_o \Sigma^v T_o^* = \begin{bmatrix} \bar{\Sigma}^v_{11} & \bar{\Sigma}^v_{12} & 0 \\ \bar{\Sigma}^v_{21} & \bar{\Sigma}^v_{22} & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{4.5b}$$

and

$$\bar{A} := T_s A T_s^{-1} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix}, \qquad \bar{C} := T_o C T_s^{-1} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ 0 & \bar{C}_{22} & \bar{C}_{23} \\ 0 & 0 & \bar{C}_{33} \end{bmatrix}, \tag{4.5c}$$

where

$$\bar{\Sigma}^s_{11} > 0, \qquad \bar{\Sigma}^v_{22} > 0, \qquad \bar{C}_{11}\bar{C}_{11}^* > 0, \qquad (\bar{C}_{11}, \bar{A}_{11}) \text{ and } (\bar{C}_{33}, \bar{A}_{33}) \text{ observable.} \tag{4.5d}$$

Remark 4.1. It is easy to see that $\bar{x}_1(t) := T_{s1}x(t)$ corresponds to the stochastically excited modes; $\bar{x}_2(t) := T_{s2}x(t)$ corresponds to the output disturbance modes; and $\bar{x}_3(t) := T_{s3}x(t)$ corresponds to the purely deterministic modes. △

[1] To be more precise, a matrix, say P, can represent a zero-dimensional matrix, i.e. $P \in \mathbb{R}^{0\times 0}$. Furthermore, the associating constraints on P, like those in Constraints (4.5d), can be relaxed. See Remark 4.2 for some examples.

Remark 4.2. Suppose there does not exist any purely deterministic modes or input channels which can be observed without any form of stochastic disturbances, i.e., $\Lambda_0 > 0$. Then, the realization in Lemma 4.4 becomes

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} \\ 0 & \bar{A}_{22} \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} \\ 0 & \bar{C}_{22} \end{bmatrix},$$

which can be envisaged as $\bar{x}_3(t)$ begin empty. On the other hand, suppose there is no output disturbance, i.e., $\bar{\Sigma}^v = 0$. Then, the realization becomes

$$\bar{A} = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ 0 & 0 & \bar{C}_{33} \end{bmatrix}.$$

Note that $\bar{x}_2(t)$, which still exists, corresponds to the modes disturbed by $\bar{\Sigma}^w_{11}$ at the output stage. Finally, consider the case where there is no process disturbance, i.e., $\bar{\Sigma}^w = 0$. Then, we have

$$\bar{A} = \begin{bmatrix} \bar{A}_{22} & \bar{A}_{23} \\ 0 & \bar{A}_{33} \end{bmatrix}, \qquad \bar{C} = \begin{bmatrix} \bar{C}_{22} & \bar{C}_{23} \\ 0 & \bar{C}_{33} \end{bmatrix},$$

in which case $(\bar{A}_{22}, \bar{C}_{22})$ is observable. Other situations can now be derived similarly by the above three structures. △

In the following proposition, we construct a Kalman filter which only updates $\bar{x}_1(t)$ and $\bar{x}_2(t)$. Since stochastic excitation and disturbance are not involved in $\bar{x}_3(t)$, we shall start the Kalman filter state $\hat{\bar{x}}_3(t)$ from the original value $\bar{x}_3(0)$, or equivalently, $T_{s3}\hat{x}(0) := T_{s3}x(0)$.

Proposition 4.5. *Given Equations* (2.1)*, and*

1. *an invertible T_s and a unitary T_o satisfying Equations* (4.5)*;*
2. $T_{s3}\hat{x}(0) := T_{s3}x(0)$*;*
3. $\tilde{P}_0 + \Sigma^s := \mathbf{E}\left([x(0) - \hat{x}(0)]\,[x(0) - \hat{x}(0)]^*\right)$*;*

then the non-steady state Kalman filter which only updates the stochastically related modes $T_{s1}x(t)$ and $T_{s2}x(t)$ is given by:

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + Bu(t) + K_t(y(t) - C\hat{x}(t) - Du(t)), \\ K_t &= (G + A\tilde{P}_tC^*)(\Lambda_0 + C\tilde{P}_tC^*)^\dagger, \\ \tilde{P}_{t+1} &= A\tilde{P}_tA^* - (G + A\tilde{P}_tC^*)(\Lambda_0 + C\tilde{P}_tC^*)^\dagger(G + A\tilde{P}_tC^*)^*. \end{aligned}$$

Proof. First note that T_o and T_s can be constructed as in Lemma 4.4. Pre-multiplying the Kalman filter gain by T_s and post-multiplying by T_o^* gives

$$\begin{aligned} T_sK_tT_o^* &= T_s(G + A\tilde{P}_tC^*)T_o^*T_o(\Lambda_0 + C\tilde{P}_tC^*)^\dagger T_o^* \\ &= (T_sGT_o^* + T_sA\tilde{P}_tC^*T_o^*)(T_o\Lambda_0T_o^* + T_oC\tilde{P}_tC^*T_o^*)^\dagger. \end{aligned}$$

With $T_{s3}\hat{x}(0) = T_{s3}x(0)$, we have

$$\begin{aligned} T_s G T_o^* + T_s A \tilde{P}_t C^* T_o^* &= \begin{bmatrix} \bar{A}_{11}\bar{\Sigma}_{11}^s \bar{C}_{11}^* + \bar{\Sigma}_{11}^{wv} & \bar{\Sigma}_{12}^{wv} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \bar{A} \left[\begin{array}{cc|c} \multicolumn{2}{c|}{(T_s \tilde{P}_t T_s^{-1})_{11}} & 0 \\ & & 0 \\ \hline 0 & 0 & 0 \end{array}\right] \bar{C}^* \\ T_o \Lambda_0 T_o^* + T_o C \tilde{P}_t C^* T_o^* &= \begin{bmatrix} \bar{C}_{11}\bar{\Sigma}_{11}^s \bar{C}_{11}^* + \bar{\Sigma}_{11}^{v} & \bar{\Sigma}_{12}^{v} & 0 \\ \bar{\Sigma}_{21}^{v} & \bar{\Sigma}_{22}^{v} & 0 \\ 0 & 0 & 0 \end{bmatrix} + \bar{C} \left[\begin{array}{cc|c} \multicolumn{2}{c|}{(T_s \tilde{P}_t T_s^{-1})_{11}} & 0 \\ & & 0 \\ \hline 0 & 0 & 0 \end{array}\right] \bar{C}^* \end{aligned}$$

with obvious partitions. With $\bar{A}$ and $\bar{C}$ defined as in Equations 4.5, it can be seen that the purely deterministic modes and the stochastically related modes are fully decoupled for K_t (and similarly for $\tilde{P}_{t+1}$). In effect, this is a standard Kalman filter for $T_{s1}x(t)$ and $T_{s2}x(t)$, whereas the purely deterministic error $(T_{s3}[x(t) - \hat{x}(t)])$ constantly equals zero. □

Remark 4.3. The above proposition suggests that by replacing an inverse by a Moore-Penrose general inverse, the Kalman filter can be regarded as a Kalman filter for stochastic related modes $T_{s1}x(t)$ and $T_{s2}x(t)$ if an appropriate $\hat{x}(0)$ is chosen. This Kalman filter will give an interpretation for a subspace projection. △

4.3 Block Kalman Filters

As indicated in Theorem 3.3, a t-step ahead Kalman filter state estimate $\hat{x}(t)$ can be written explicitly as an expression of $\hat{x}(0)$, $u(0), \ldots, u(t-1)$, and $y(0), \ldots, y(t-1)$. In this section, we will show that this is also a 1-step ahead Kalman filter state estimate of a t-block system. First of all, we consider the positive realness of a stochastic system. In the full stochastic excitation case, we assume the covariance of the process and output noises is positive definite in order to obtain strict positive realness:

$$\begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} > 0.$$

Equivalently, in the partial stochastic excitation case we assume the positive definiteness of the following covariance matrix, in the state and output coordinates defined in Equations (4.5):

$$\left[\begin{array}{c|cc} \bar{\Sigma}_{11}^w & \bar{\Sigma}_{11}^{wv} & \bar{\Sigma}_{12}^{wv} \\ \hline \bar{\Sigma}_{11}^{vw} & \bar{\Sigma}_{11}^{v} & \bar{\Sigma}_{12}^{v} \\ \bar{\Sigma}_{12}^{vw} & \bar{\Sigma}_{21}^{v} & \bar{\Sigma}_{22}^{v} \end{array}\right] > 0. \tag{4.6}$$

Here, we have ignored the stochastic modes $\bar{x}_2^s(t)$, $\bar{x}_3^s(t)$ and the stochastic output channels $\bar{y}_3^s(t)$, since they are essentially zero. With the positive definiteness of the above covariance matrix, it is then easy to see that the following reduced stochastic subsystem, in the state and output coordinates defined in Equations (4.5),

$$\begin{aligned} \bar{x}_1^s(t+1) &= \bar{A}_{11}\ \bar{x}_1^s(t) + \ \bar{w}_1(t)\ , \\ \begin{bmatrix} \bar{y}_1^s(t) \\ \bar{y}_2^s(t) \end{bmatrix} &= \begin{bmatrix} \bar{C}_{11} \\ 0 \end{bmatrix} \bar{x}_1^s(t) + \begin{bmatrix} \bar{v}_1(t) \\ \bar{v}_2(t) \end{bmatrix}, \end{aligned}$$

is strictly positive real. Define the output covariance matrix $\bar{\Lambda}_i$ of the above equations as:

$$\bar{\Lambda}_i = \mathbf{E}\left(\begin{bmatrix}\bar{y}_1^s(t+i)\\ \bar{y}_2^s(t+i)\end{bmatrix}\begin{bmatrix}\bar{y}_1^s(t)\\ \bar{y}_2^s(t)\end{bmatrix}^*\right).$$

Then, the Toeplitz matrix $\bar{L}_i$ constructed as

$$\bar{L}_i := \begin{bmatrix}\bar{\Lambda}_0 & \bar{\Lambda}_{-1} & \cdots & \bar{\Lambda}_{1-i}\\ \bar{\Lambda}_1 & \bar{\Lambda}_0 & \ddots & \bar{\Lambda}_{2-i}\\ \vdots & \ddots & \ddots & \vdots\\ \bar{\Lambda}_{i-1} & \bar{\Lambda}_{i-2} & \cdots & \bar{\Lambda}_0\end{bmatrix}, \tag{4.7}$$

will also be positive definite, $\bar{L}_i > 0$, for all $i \in \mathbb{N}$, as guaranteed by Inequality (4.6) (for example see [Fau76]). With $\bar{L}_i > 0$, we can then extend Theorems 3.3 and 3.4 to Proposition 4.7 and Theorem 4.9 in a straightforward manner.

Now, consider a t-block system. The state and output equations of a t-block system have the form:

$$x(t) = A^t x(0) + \mathcal{C}_t^d \underline{u}(0) + \mathcal{A}_t \underline{w}(0), \tag{4.8a}$$
$$\underline{y}(0) = \mathcal{O}_t x(0) + \mathcal{T}_t \underline{u}(0) + \mathcal{N}_t \underline{w}(0) + \underline{v}(0), \tag{4.8b}$$

where

$$\underline{u}(0) := \begin{bmatrix}u(0)\\ \vdots\\ u(t-1)\end{bmatrix}, \qquad \underline{w}(0) := \begin{bmatrix}w(0)\\ \vdots\\ w(t-1)\end{bmatrix},$$

and similarly for $\underline{y}$ and $\underline{v}$. Moreover, denote respectively by $\underline{y}^d$ and $\underline{y}^s$ the output signals of the deterministic subsystem and stochastic subsystem of the t-block system. Again, $\mathcal{A}_t$ and $\mathcal{N}_t$ are defined as

$$\mathcal{A}_t = \begin{bmatrix}A^{t-1} & A^{t-2} & \cdots & I\end{bmatrix}, \qquad \mathcal{N}_t = \begin{bmatrix}0 & 0 & 0 & 0\\ C & 0 & 0 & 0\\ \vdots & \ddots & \ddots & 0\\ CA^{t-2} & \cdots & C & 0\end{bmatrix}.$$

Now, the deterministic subsystem has system matrices $(A^t, \mathcal{C}_t^d, \mathcal{O}_t, \mathcal{T}_t)$. In the following lemma, we point out that a t-step ahead Kalman filter state estimate can be computed as a 1-step ahead t-block Kalman filter state estimate. Note that for the sake of simplicity, we shall consider the case where the stochastic subsystem is strictly positive real such that $L_t > 0$. Nonetheless, this result can be extended to the partial stochastic excitation case by carefully isolating the purely deterministic modes, which will not be discussed here.

Lemma 4.6. *Suppose $L_t > 0$. Then, a t-step ahead Kalman filter state estimate is equivalent to a 1-step ahead t-block Kalman filter state estimate.*

The proof of Lemma 4.6 can be found in Appendix A. The following proposition, which is an extended version of Theorem 3.3, considers the partial stochastic excitation case.

Proposition 4.7. *Given that Inequality* (4.6) *holds, and given*

1. *an invertible* T_s *and a unitary* T_o *satisfying Equations* (4.5)*;*
2. $T_{s3}\hat{x}(0) := T_{s3}x(0)$*;*
3. $\tilde{P}_0 + \Sigma^s := \mathbf{E}\left([x(0) - \hat{x}(0)]\,[x(0) - \hat{x}(0)]^*\right)$*;*

then the non-steady state Kalman filter which only updates the stochastically related modes $T_{s1}x(t)$ *and* $T_{s2}x(t)$*:*

$$\begin{aligned} \hat{x}(t+1) &= A\hat{x}(t) + Bu(t) + K_t(y(t) - C\hat{x}(t) - Du(t)), \\ K_t &= (G + A\tilde{P}_tC^*)(\Lambda_0 + C\tilde{P}_tC^*)^\dagger, \\ \tilde{P}_{t+1} &= A\tilde{P}_tA^* - (G + A\tilde{P}_tC^*)(\Lambda_0 + C\tilde{P}_tC^*)^\dagger(G + A\tilde{P}_tC^*)^*, \end{aligned}$$

can be written explicitly as:

$$\hat{x}(t) = A^t\hat{x}(0) + \mathcal{C}_t^d\underline{u}(0) + \mathcal{K}_t(\underline{y}(0) - \mathcal{O}_k\hat{x}(0) - \mathcal{T}_t\underline{u}(0)),$$

where

$$\mathcal{K}_t = (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^\dagger.$$

The explicit solution of the matrix $\tilde{P}_t$ *is equal to:*

$$\tilde{P}_t = A^t\tilde{P}_0(A^t)^* - (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^\dagger(\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)^*.$$

Proof. As in the proof of Proposition 4.5, pre-multiplying T_s and post-multiplying T_o^* to $\mathcal{K}_t$ clearly decouple the purely deterministic modes and the stochastically related modes (and similarly for $\tilde{P}_t$). Thus, by Theorem 3.3 the stochastically related part of the Kalman filter is proven. The statement follows as the purely deterministic error $(T_{s3}[\hat{x}(t) - x(t)])$ is identical to zero. □

4.4 A Geometric Viewpoint

In this section, we decompose the initial state $x(0)$ into three mutually orthogonal components. As we will see, this orthogonal property is the key point of obtaining the subspace projection theorems (Theorems 3.4 and 4.9). We shall assume a state sequence is fed through a bank of q Kalman filters, as covered in Chapter 3. We construct U^{k+l}, Y^{k+l}, and X^{k+l}, each of $k+l$ block rows. To simplify the notation, the superscript $k+l$ will be dropped. For convenience, we define $U_p := U_{0,k-1}$ and $U_f := U_{k,k+l-1}$. Furthermore, define Y_p and Y_f similarly. On the other hand, we define $X_p := X_0$ and $X_f := X_k$. The spaces spanned by these vectors are denoted by $\mathcal{X}_p$, $\mathcal{X}_f$, $\mathcal{U}_p$, $\mathcal{U}_f$, $\mathcal{Y}_p$, $\mathcal{Y}_f$.

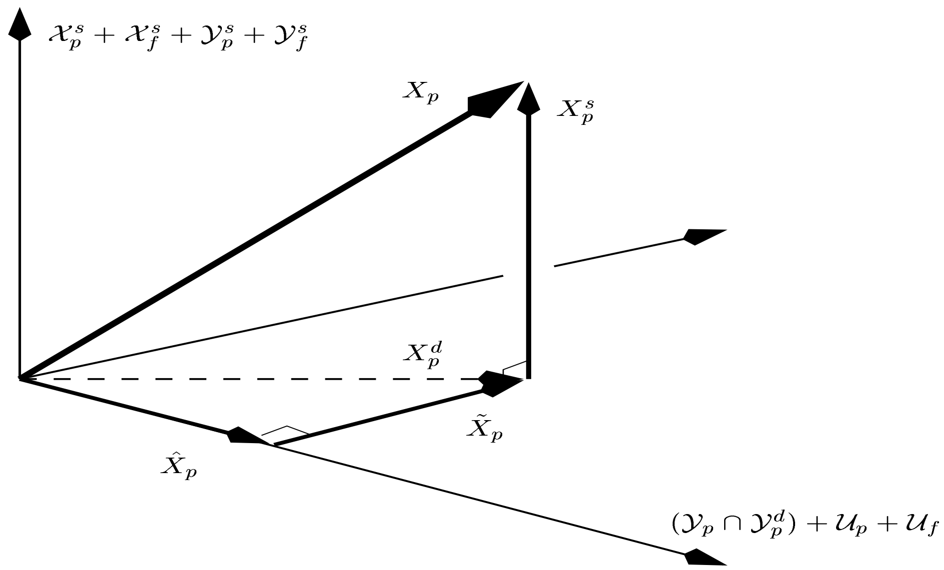

Figure 4.1 *Geometric components of initial state* X_p

Consider Equations (2.1), with an invertible T_s and a unitary T_o satisfying Equations (4.5). Suppose k is no less than the observability index of the system, then define $\hat{\mathcal{X}}_p^d$, $\hat{\mathcal{Y}}_p^d$ as:

$$\hat{\mathcal{X}}_p^d := \operatorname{span}\{\, T_{s3} X_p \,\},$$
$$\hat{\mathcal{Y}}_p^d := \operatorname{span}\{\, T_{o3} Y_j \mid j = 0, 1, \ldots, k-1 \}.$$

Remark 4.4. Comparing with Equations (4.5), it is not hard to see that $\hat{\mathcal{X}}_p^d$ is the space spanned by the purely deterministic modes of the system. △

Defining

$$\mathcal{S} := \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f; \qquad \hat{X}_p := \Pi_{\hat{\mathcal{Y}}_p^d + \mathcal{U}_p + \mathcal{U}_f} X_p; \qquad \tilde{X}_p := X_p - \hat{X}_p - X_p^s = X_p^d - \hat{X}_p, \tag{4.9}$$

we have the following properties, whose proofs can be found in Appendix A.

Lemma 4.8. *Suppose* $\hat{X}_p$ *and* $\tilde{X}_p$ *are given in Equations* (4.9), *then*

1. $\mathcal{X}_p^s \perp \hat{\mathcal{X}}_p$; $\hat{\mathcal{X}}_p \perp \tilde{\mathcal{X}}_p$; $\tilde{\mathcal{X}}_p \perp \mathcal{X}_p^s$; [2]
2. $\hat{\mathcal{Y}}_p^d = \mathcal{Y}_p^d \cap \mathcal{Y}_p$;
3. $(\mathcal{Y}_p^d \cap \mathcal{Y}_p) + \mathcal{U}_p + \mathcal{U}_f = \hat{\mathcal{X}}_p^d + \mathcal{U}_p + \mathcal{U}_f$;
4. $\tilde{P}_0 = \frac{1}{q}\mathbf{E}\big(\tilde{X}_p \tilde{X}_p^*\big)$, *where* $\tilde{P}_0 := \frac{1}{q}\mathbf{E}\big[(X_p - \hat{X}_p)(X_p - \hat{X}_p)^*\big] - \Sigma^s$.

[2] That is, $\mathbf{E}[\mathcal{X}_p^s \hat{\mathcal{X}}_p^*] = 0$, $\mathbf{E}[\hat{\mathcal{X}}_p \tilde{\mathcal{X}}_p^*] = 0$, $\mathbf{E}[\tilde{\mathcal{X}}_p (\mathcal{X}_p^s)^*] = 0$.

The following theorem is an extension of Theorem 3.4. Here, the initial condition is $\hat{X}_p = \Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p)+\mathcal{U}_p+\mathcal{U}_f} X_p$, where the extra term $\mathcal{Y}_p^d \cap \mathcal{Y}_p$ takes care of the purely deterministic modes.

Theorem 4.9. *Given that Inequality* (4.6) *holds and k is no less than the observability index of the system. Then,*

$$\Pi_{\mathcal{S}} Y_f = \mathcal{O}_l \hat{X}_f + \mathcal{T}_l U_f,$$

where $\hat{X}_f$ is the sequence of q k-step ahead Kalman filter state estimates, each defined in Proposition 4.5, with the sequence of initial conditions:

$$\hat{X}_p = \Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p)+\mathcal{U}_p+\mathcal{U}_f} X_p, \qquad \tilde{P}_0 = \frac{1}{q}\mathbf{E}(\tilde{X}_p \tilde{X}_p^*).$$

Proof. Let $\tilde{X}_f := X_f^d - \hat{X}_f = A^k \tilde{X}_p - \mathcal{K}_k(Y_p^s + \mathcal{O}_k \tilde{X}_p)$. Thus, we can write

$$Y_f = \mathcal{O}_l(\hat{X}_f + \tilde{X}_f + X_f^s) + \mathcal{T}_l U_f.$$

From Lemma 4.8 part 2, we have $\hat{\mathcal{X}}_p \subset \mathcal{S}$. Also, as shown in Proposition 4.7, $\hat{X}_f$ is a linear combination of elements in $\hat{\mathcal{X}}_p$ and $\mathcal{S}$. Thus, $\Pi_{\mathcal{S}} \hat{X}_f = \hat{X}_f$. Now, projecting Y_f onto $\mathcal{S}$ gives

$$\Pi_{\mathcal{S}} Y_f = \mathcal{O}_l(\hat{X}_f + \Pi_{\mathcal{S}}(\tilde{X}_f + X_f^s)) + \mathcal{T}_l U_f.$$

Thus, it remains to show $\Pi_{\mathcal{S}}(\tilde{X}_f + X_f^s) = 0$. It is obvious that $\Pi_{\mathcal{U}_p+\mathcal{U}_f}(\tilde{X}_f + X_f^s) = 0$. Note that $Y_p = Y_p^s + \mathcal{O}_k \tilde{X}_p + \mathcal{O}_k \hat{X}_p + \mathcal{T}_k U_p$. Therefore,

$$\begin{aligned}
\frac{1}{q}\mathbf{E}\left[(X_f^s + \tilde{X}_f) Y_p^*\right] &= \frac{1}{q}\mathbf{E}\left[\left(X_f^s + A^k \tilde{X}_p - \mathcal{K}_k(Y_p^s + \mathcal{O}_k \tilde{X}_p)\right) \cdot \left(Y_p^s + \mathcal{O}_k \tilde{X}_p\right)^*\right] \\
&= \mathcal{C}_k^s + A^k \tilde{P}_0 \mathcal{O}_k^* - \mathcal{K}_k(L_k + \mathcal{O}_k \tilde{P}_0 \mathcal{O}_k^*) \\
&= 0.
\end{aligned} \tag{4.10}$$

Note that the last equality partly stems from the fact that the purely deterministic error is identical to zero. The stochastically related part is a straightforward verification by writing $\mathcal{K}_k$ explicitly. Thus, $\Pi_{\mathcal{Y}_p}(\tilde{X}_f + X_f^s) = 0$, and hence $\Pi_{\mathcal{S}}(\tilde{X}_f + X_f^s) = 0$. □

Remark 4.5. Theorem 3.4 is a special case with Σ^w, $\Sigma^v > 0$. On the other hand, suppose $\Sigma^s = \Sigma^v = 0$. We can then see from Theorem 4.9 that $\hat{\mathcal{X}}_p = \mathcal{X}_p^d = \mathcal{X}_p$ and $\Pi_{\mathcal{S}} \mathcal{X}_f = \mathcal{X}_f^d = \mathcal{X}_f$, which are expected in deterministic identification (as shown in Equation (3.14)). △

4.5 Summary

In this chapter, we generalize the purely deterministic and fully stochastic cases to an amalgamated extension. The purely deterministic modes of a state sequence obtained in a subspace projection are identified using a particular realization defined in Lemma 4.4. Then, a generalized subspace projection theorem (Theorem 4.9) is proven in a geometric framework, which gives a more insightful understanding and a much simpler approach to interpret a subspace projection.

As seen in the proof of Theorem 4.9, the key point is to find $\hat{X}_f$ such that $\Pi_{\mathcal{S}}\hat{X}_f = \hat{X}_f$ and $\Pi_{\mathcal{S}}(\tilde{X}_f + X_f^s) = 0$. In the latter case, it is equivalent to ensure that

$$\mathbf{E}\left[(\tilde{X}_f + X_f^s)U_p^*\right] = 0, \qquad \mathbf{E}\left[(\tilde{X}_f + X_f^s)U_f^*\right] = 0, \qquad \mathbf{E}\left[(\tilde{X}_f + X_f^s)Y_p^*\right] = 0.$$

For the first two equalities, it is not hard to see that they can be written as:

$$\mathbf{E}\left[\tilde{X}_f U_p^*\right] = 0, \qquad \text{and} \qquad \mathbf{E}\left[\tilde{X}_f U_f^*\right] = 0. \tag{4.11}$$

Consider the third equality:

$$\mathbf{E}\left[(X_f^s + \tilde{X}_f)Y_p^*\right] = 0. \tag{4.12}$$

First, note that the error between the observation and the prediction can be expressed as:

$$Y_p - \mathcal{O}_k\hat{X}_p - \mathcal{T}_k U_p = Y_p^s + \mathcal{O}_k\tilde{X}_p. \tag{4.13}$$

On the other hand, the error between the state estimate and the actual state is given by:

$$\begin{aligned} X_f - \hat{X}_f &= X_f^s + \tilde{X}_f \\ &= X_f^s + A^k\tilde{X}_p - \mathcal{K}_k(Y_p - \mathcal{O}_k\hat{X}_p - \mathcal{T}_k U_p) \\ &= X_f^s + A^k\tilde{X}_p - \mathcal{K}_k(Y_p^s + \mathcal{O}_k\tilde{X}_p). \end{aligned} \tag{4.14}$$

From linear estimation theory, we know that $\mathcal{K}_k$ is the best constant to match $X_f^s + A^k\tilde{X}_p$ with $Y_p^s + \mathcal{O}_k\tilde{X}_p$, so as to minimize the error in the state signal. Thus, by the principle of orthogonality the error should be orthogonal to the estimator:

$$\left(X_f^s + A^k\tilde{X}_p - \mathcal{K}_k(Y_p^s + \mathcal{O}_k\tilde{X}_p)\right) \perp \left(Y_p^s + \mathcal{O}_k\tilde{X}_p\right),$$

i.e., their correlation is zero. In other words, we have

$$\mathbf{E}\left[(X_f^s + \tilde{X}_f)(Y_p^s + \mathcal{O}_k\tilde{X}_p)^*\right] = 0. \tag{4.15}$$

Now, Equation (4.12) can also be written as

$$\mathbf{E}\left[\left(X_f^s + A^k\tilde{X}_p - \mathcal{K}_k(Y_p^s + \mathcal{O}_k\tilde{X}_p)\right)\left(\mathcal{O}_k\hat{X}_p + \mathcal{T}_k U_p\right)^*\right] = 0,$$

which is obtained by expanding the equation with Equations (4.13) and (4.14) and simplifying the resulting expression using Equation (4.15). One way to satisfy the above equality is to find $\hat{X}_p$ and $\tilde{X}_p$ such that

$$\mathbf{E}\left[(X_f^s + A^k\tilde{X}_p)(\mathcal{O}_k\hat{X}_p + \mathcal{T}_k U_p)^*\right] = 0 \qquad \text{and} \qquad \mathbf{E}\left[(Y_p^s + \mathcal{O}_k\tilde{X}_p)(\mathcal{O}_k\hat{X}_p + \mathcal{T}_k U_p)^*\right] = 0.$$

With the use of Equations (4.11), the above equations are then simplified to

$$\mathbf{E}\left[(X_f^s + A^k\tilde{X}_p)(\mathcal{O}_k\hat{X}_p)^*\right] = 0 \qquad \text{and} \qquad \mathbf{E}\left[(Y_p^s + \mathcal{O}_k\tilde{X}_p)(\mathcal{O}_k\hat{X}_p)^*\right] = 0.$$

To ensure the above equalities, it is sufficient to find a decomposition $X_p = \hat{X}_p + \tilde{X}_p + X_p^s$ such that all three components are mutually orthogonal. This also suggests that $\hat{X}_p$ and $\tilde{X}_p$ are some deterministic components such that $\hat{X}_p + \tilde{X}_p = X_p^d$. Finally, returning to Equations (4.11) it can be seen that a natural choice of $\hat{X}_p$ is the projection of X_p onto $(\mathcal{Y}_p^d \cap \mathcal{Y}_p) + \mathcal{U}_p + \mathcal{U}_f$, in which case we also have $\Pi_{\mathcal{S}}\hat{X}_p = \hat{X}_p$. As $\hat{\mathcal{X}}_f$ is a subset of $\mathcal{S} + \hat{\mathcal{X}}_p$, it can be seen that $\Pi_{\mathcal{S}}\hat{X}_f = \hat{X}_f$ is also ensured.

5

Informative Experiments for Subspace Methods

5.1 Introduction

A number of subspace algorithms in the existing literature (e.g. see [MDMVV89, VD92a, VD92b, Ver94, VODM94, VODM96b, WV96]) are based on an assumption that particular row spaces constructed from the state sequence and the input sequence have empty intersection. The phenomenon in which input and state sequences violate this empty intersection assumption is known as rank cancelation [DM88, p. 270]. A relatively restrictive condition, which requires input signals to be periodic with infinite length, is given to prevent such a cancelation in Theorem 2 of [VD92a]. However, as remarked in [Vib95, JW96, CV97] general conditions, particularly in the combined deterministic-stochastic case, for preventing this phenomenon without referring to the state sequences which are to be identified are not explicitly formulated. Rank cancelation in turn leads to poor models being obtained with subspace methods in some cases. Motivated by this matter, informative experiments for subspace methods are studied in this chapter.

Informative experiments are identification experiments which contain sufficient information to discriminate between different models in an intended model set [Lju87]. However, different algorithms require different amounts of information in an experiment for a successful identification operation. In this chapter, a particular set of identification algorithms, namely subspace methods, is considered. Criteria for experiments to be informative with these methods in the deterministic setup and the combined deterministic-stochastic setup are presented. As it will be seen, the key factor for the construction of informative experiments in deterministic identification is the emptiness of the intersection of certain input and state sequences. In contrast, the key factor in combined deterministic-stochastic identification is the emptiness of the intersection of certain input and projected state sequences. Conditions for such emptiness are fully explored in this chapter.

In any identification algorithm, it is important to select a proper set of models and experimental data. In the light of this, aspects of proper settings of an identification algorithm are succinctly reviewed in Section 5.2. Furthermore, the motivation and the notion of informative experiments for subspace methods, scrutinized in the following sections, are also explained. In Section 5.3, we present two subspace algorithms for which informative experiments are investigated. Then this chapter is divided into two separate parts, in which Sections 5.4 and 5.5 deal with the deterministic case; whereas Sections 5.7 and 5.8 are devoted to the combined deterministic-stochastic case. To facilitate the study of the combined deterministic-stochastic case, Section 5.6 presents two useful subsystem decompositions, in which one dismantles a system in a non-stationary stochastic fashion, and the other in a Kalman filter fashion.

5.2 Model Sets and Data Sets

System identification can be divided into two main categories: parametric methods and nonparametric methods. On the one hand, parametric methods explicitly parametrize the set of models in the search for a best description. Such methods include prediction-error identification methods and instrumental-variable methods. On the other hand, nonparametric methods determine a transfer function without referring to any particular parametrization or any canonical form. Empirical transfer-function estimation methods and subspace methods are examples of this kind.

In this section, we will concentrate on the deterministic part of a system, namely $\Sigma^w = \Sigma^v = 0$. For the case where noise is present, the discussion will still hold if one substitutes $G(z)$ by $[G(z)\ H(z)]$, where $H(z)$ is the transfer function for the noise. Furthermore, we assume the order of the model is known and thus confine the model set to be of a particular order. To pursue the discussion, we define two classes of model sets. Note that notations in this section are adopted from [Mac85, Lju87, Cho94].

Definition 5.1. $TM_n^{p,m}$ is the set of all m-input, p-output discrete-time, linear time-invariant, transfer functions of McMillan degree n.

Definition 5.2. $SM_n^{p,m}$ is the set of all m-input, p-output discrete-time linear time-invariant, minimal state-space realizations (A, B, C, D) of order n.

Note that the space $SM_n^{p,m}$ can be embedded in $\mathbb{R}^{n^2+np+nm+pm}$ with the subspace topology (see for example [Cho94]). In addition, we recall that a homeomorphism is a continuous bijective map whose inverse is also continuous. Such a map preserves topological properties (see for example [Cho94]).

Now, consider two models, (A_1, B_1, C_1, D_1), $(A_2, B_2, C_2, D_2) \in SM_n^{p,m}$. The two models are said to be equivalent if

$$C_1(zI - A_1)^{-1}B_1 + D_1 = C_2(zI - A_2)^{-1}B_2 + D_2 \tag{5.1}$$

for almost all $z \in \mathbb{C}$. Equation (5.1) holds if and only if

$$TA_2 = A_1T, \quad TB_2 = B_1, \quad C_2 = C_1T, \quad D_2 = D_1, \tag{5.2}$$

for some $T \in GL_n(\mathbb{R})$, where $GL_n(\mathbb{R})$ is the set of all invertible $n \times n$ real matrices.

5.2.1 Parametric Modeling

Consider the parametric identification approach. Let the model set $\mathcal{M}^*$ be the collection of models $\mathcal{M}(\theta)$ parametrized using the parameter vector $\theta \in \mathcal{D}_{\mathcal{M}} \subset \mathbb{R}^d$, i.e.,

$$\mathcal{M}^* = \{\mathcal{M}(\theta) : \theta \in \mathcal{D}_{\mathcal{M}}\},$$

where $\mathcal{D}_{\mathcal{M}}$ is the domain of the parameter vector θ. Let $\mathscr{Z}^N$ (as defined in Equation (2.2)) be a set of input-output observations from a system. Then a parametric-based identification algorithm $\mathcal{F}$ can be associated with the map:

$$\mathcal{F} : \mathscr{Z}^N \mapsto \theta \in \mathcal{D}_{\mathcal{M}}.$$

This mapping is usually defined to be the minimizing solution of an identification objective function $V_N(\theta, \mathscr{Z}^N)$:

$$\mathcal{F} : \mathscr{Z}^N \mapsto \arg \min_{\theta \in \mathcal{D}_{\mathcal{M}}} V_N(\theta, \mathscr{Z}^N). \tag{5.3}$$

Therefore, if $\mathcal{M}^*$ is chosen to be the collection of models in $SM_n^{p,m}$, ambiguity will arise as there may exist $\theta_1 \neq \theta_2$ such that $\mathcal{M}(\theta_1)$ is equivalent to $\mathcal{M}(\theta_2)$. This problem is known as the (global structural) identifiability problem of the model set. (On the other hand, local structural identifiability considers the identifiability of a local neighborhood at a particular value of θ. Bellman and Åström raised the problem of structural identifiability in [BÅ70]. Results on local/global identifiability using Implicit Function Theorem were given in [GW74]. Some other results can be found in [Wal82] and references therein.)

Owing to this problem, we consider the quotient space $SM_n^{p,m}/GL_n(\mathbb{R})$ with the quotient topology (see for example [Cho94]). Note that there exists a bijection between this quotient space $SM_n^{p,m}/GL_n(\mathbb{R})$ and $TM_n^{p,m}$. In order to define the identification map, a model structure $\mathcal{M}$ has to be selected. Selecting the model structure involves the parametrization of the space $TM_n^{p,m}$. Suppose the model structure $\mathcal{M}$ is an bijective map

$$\mathcal{M} : \mathcal{D}_{\mathcal{M}} \to TM_n^{p,m},$$

such that $\theta_1 = \theta_2$ whenever $\mathcal{M}(\theta_1)$ is equivalent to $\mathcal{M}(\theta_2)$. Such a model structure is said to be a set of canonical forms. (In general, several canonical forms of different structures are required in parametrizing $TM_n^{p,m}$.) Using a set of canonical forms, models are then identifiable.

The use of pseudo-canonical forms in parameter estimation is advocated in [GW74]. In this approach, a model set

$$\mathcal{M}^* = \{\mathcal{M}_j(\theta_j) : \theta_j \in \mathcal{D}_{\mathcal{M}_j}, j \in J\},$$

with a collection of homeomorphic model structures $\mathcal{M}_j$, which constitutes a finite open cover of $TM_n^{p,m}$,

$$\bigcup_{j \in J} \mathcal{M}_j(\mathcal{D}_{\mathcal{M}_j}) = TM_n^{p,m},$$

is employed (see for instance [Mac85]). With $\mathcal{M}_j : \mathcal{D}_{\mathcal{M}_j} \to \mathcal{M}_j(\mathcal{D}_{\mathcal{M}_j})$ homeomorphic, topological properties are preserved. The major attribute of a set of pseudo-canonical forms, as suggested by its name, is the overlapping model structures:

$$\bigcap_{j \in J_1} \mathcal{M}_j(\mathcal{D}_{\mathcal{M}_j}) \neq \{0\},$$

for some $J_1 \subset J$.

A particular set of canonical forms using balanced parametrizations was also investigated in [Mac85]. Balanced canonical forms are less sensitive to over parametrization and singularity of Gramians, and possess nice model reduction properties [MO88]. Explicit parametrizations were given for several sets of minimal systems, including asymptotically stable systems, all-pass systems, bounded real systems, positive real systems, minimum-phase systems, and the class of all minimal systems [Obe91]. By using a particular parametrization specific properties of the identified model can then be ensured. Implementation of a prediction-error based identification algorithm with these parametrizations was given in [CM97].

In parametric identification, an initial model is usually required. Gradients of the objective function are then computed to find a best description from the model set. Nonetheless, searching for a good initial model is important in the optimization procedure. To this end, nonparametric modeling can be used for this purpose. In particular, it has been shown that recently developed subspace methods provide a good way in acquiring an initial model.

5.2.2 Nonparametric Modeling

Consider the nonparametric identification approach. Again, suppose $\mathscr{Z}^N$ is the set of input-output observations from a system. Then a nonparametric identification algorithm $\mathcal{F}$ can be associated with the map:

$$\mathcal{F} : \mathscr{Z}^N \mapsto G(z) \in TM_n^{p,m}. \tag{5.4}$$

In this approach, no parametrization is explicitly defined. For example, in the correlation and spectral analysis the transfer function can be determined as the ratio of two power spectra (c.f. Equation (2.4)),

$$G(e^{i\omega}) = \Phi_{yu}(e^{i\omega})\Phi_{uu}^{-1}(e^{i\omega}),$$

followed by any necessary model reduction or reduced model approximation.

Other examples are subspace methods. These methods are capable of identifying systems in the state-space form $(A, B, C, D) \in SM_n^{p,m}$ up to a similarity transformation. A subspace identification algorithm $\mathcal{F}$ can also be regarded as a map defined in (5.4). To be more precise, it can be perceived as a composition of the map $\mathcal{F}_1$ mapping from a data set $\mathscr{Z}^N$ to matrices

(A, B, C, D), and the map $\mathcal{F}_2$ mapping from the matrices to the corresponding element in the quotient space $SM_n^{p,m}/GL_n(\mathbb{R})$:

$$\mathcal{F}: \mathscr{X}^N \overset{\mathcal{F}_1}{\mapsto} (A, B, C, D) \overset{\mathcal{F}_2}{\mapsto} \left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\}.$$

Effectively, $\mathcal{F}_1$ represents the actual identification procedures, whereas $\mathcal{F}_2$ signifies that systems in the state-space form $(A, B, C, D) \in SM_n^{p,m}$ can only be identified up to a similarity transformation.

It can be seen that the parametrization of the model set is not something to worry about as in Mapping (5.3), in contrast to the parametric identification approach. However, this does not imply that the identification algorithm $\mathcal{F}$ will necessarily provide a meaningful model in all cases. Instead, the algorithm $\mathcal{F}$ must also be used with care. Take a simple example, suppose we are given a pair of signals, $\mathscr{X}^N$, with the input signal identical to zero ($u = 0$). If the output signal is also zero ($y = 0$), it would be something expected. On the other hand, if the output signal is something other than zero, i.e. something driven by the past input or by noises, it still comes as no surprise. Nonetheless, both cases will not provide much information about how a signal is transformed from an input signal to an output signal. To ensure a proper use of the algorithm in this respect, a restriction on the domain of the data set is required. That is,

$$\mathcal{F}: \mathcal{D}_{\mathscr{X}} \to TM_n^{p,m}, \tag{5.5}$$

where $\mathcal{D}_{\mathscr{X}}$ is the domain of the data set. The study of this domain, or sets of informative experiments, is the theme of the subsequent sections in this chapter.

5.3 Subspace Algorithms

In this section we present two subspace algorithms, whose informative experiments are investigated in subsequent sections. These two subspace algorithms are closely related to those in the existing literature. More precisely, the first algorithm (Algorithm 5.1) is analogous to the ordinary MOESP algorithm [VD92a, VD92b, Ver93b] in which no instrumental variable is used. On the contrary, the second algorithm (Algorithm 5.2) is analogous to the MOESP PO scheme [Ver94], which employs instrumental variable techniques. We refer to [VODM95b] for the connection between the MOESP PO scheme and the N4SID scheme [VODM94]. Finally, we recall that U^l, constructed from a signal u of length N, is a $\mathscr{X}$-valued Hankel block matrix defined as:

$$U^l := \begin{bmatrix} u(0) & u(1) & \cdots & u(N-l) \\ u(1) & u(2) & \cdots & u(N-l+1) \\ \iddots & \iddots & \iddots & \iddots \\ u(l-1) & u(l) & \cdots & u(N-1) \end{bmatrix}. \tag{5.6}$$

In addition, denote by $q := N - l + 1$ the number of filters in the Kalman filter bank, or simply the number of columns in U^l.

5.3.1 A One-Block Subspace Algorithm

In this subsection, an algorithm (Algorithm 5.1) which can identify systems in a deterministic setup but not a stochastic setup, is presented. Its requirements for informative experiments will be discussed in Section 5.5. Moreover, a summary of this algorithm can be found on page 49.

Remark 5.1. To see the connection between Algorithm 5.1 and the ordinary MOESP algorithm, we have the following analysis regarding the extraction of $\mathcal{O}_l$. In [VD92a, VD92b, Ver93b], the following RQ factorization is computed:

$$\begin{bmatrix} U^l \\ Y^l \end{bmatrix} = \begin{bmatrix} R_{11} & 0 \\ R_{21} & R_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = RQ.$$

Then, the SVD of R_{22} is computed:

$$R_{22} = \tilde{\Gamma}\tilde{\Sigma}\tilde{\Omega}^*,$$

where $\mathcal{O}_l$ is extracted from $\tilde{\Gamma}$. To begin our analysis, we first see that

$$\Pi_{\mathcal{U}^{l\perp}} Y^l = \begin{bmatrix} 0 & R_{22} \end{bmatrix} Q.$$

Then, it can be easily checked that $\Pi_{\mathcal{U}^{l\perp}} Y^l$ has the SVD

$$\Pi_{\mathcal{U}^{l\perp}} Y^l = \begin{bmatrix} 0 & R_{22} \end{bmatrix} Q = \tilde{\Gamma} \begin{bmatrix} \tilde{\Sigma} & 0 \end{bmatrix} \begin{bmatrix} 0 & \tilde{\Omega}^* \\ I & 0 \end{bmatrix} Q.$$

Now, it is evident that $\tilde{\Gamma} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix}$, where $\begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix}$ is defined in Equation (5.8) in Algorithm 5.1. △

5.3.2 A Two-block Subspace Algorithm

In this subsection, another subspace algorithm (Algorithm 5.2) is presented. Note that this algorithm can give consistent estimates with the presence of process and output noises under mild conditions, though noise cancelation properties will not be discussed here. Requirements for experiments to be informative with this algorithm will be presented for the deterministic case in Section 5.5 and the combined deterministic-stochastic case in Section 5.8.

In the following, we define the 'past' and 'future' data blocks. Here, the definition has a more general structure, in which the number of block rows in the 'past' data and the 'future' data can be different. This unequal splitting of data blocks has been addressed in [WJ94, JW95] and also plays an important role in Markov parameter estimation discussed in Chapter 7.

$$\begin{bmatrix} U_p \\ \hline U_f \end{bmatrix} := \left[\begin{array}{cccc} u(0) & u(1) & \cdots & u(N-k-l) \\ \iddots & \iddots & \iddots & \iddots \\ u(k-1) & u(k) & \cdots & u(N-l-1) \\ \hline u(k) & u(k+1) & \cdots & u(N-l) \\ \iddots & \iddots & \iddots & \iddots \\ u(k+l-1) & u(k+l) & \cdots & u(N-1) \end{array}\right]. \tag{5.7}$$

Algorithm 5.1. One-Block Subspace Algorithm

1. *Construct* U^l *and* Y^l *as in Equation* (5.6).
2. *Compute the SVD of* $\Pi_{\mathcal{U}^{l\perp}} Y^l$ *as (with obvious partitions)*

$$\Pi_{\mathcal{U}^{l\perp}} Y^l = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}, \quad \textit{where} \quad \Gamma_1 = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_l \end{bmatrix}. \tag{5.8}$$

3. *Compute* A *and* C *from*

$$\begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{l-1} \end{bmatrix} A = \begin{bmatrix} \gamma_2 \\ \vdots \\ \gamma_l \end{bmatrix} \qquad \textit{and} \qquad C = \gamma_1.$$

4. *Compute* $S_0, \ldots, S_{l-1}$ *from*

$$\mathcal{O}_l^{\perp} Y^l := S_0 U_0^l + \cdots + S_{l-1} U_{l-1}^l, \tag{5.9}$$

where

$$\mathcal{O}_l^{\perp} := \begin{bmatrix} P_0 & \cdots P_{l-2} & I \end{bmatrix} \qquad \textit{and} \qquad \mathcal{O}_l^{\perp} \mathcal{O}_l = 0.$$

5. *Determine* B *and* D *from*

$$\begin{bmatrix} S_0 \\ S_1 \\ \vdots \\ S_{l-1} \end{bmatrix} = \begin{bmatrix} P_0 & \cdots & P_{l-2} & I \\ P_1 & \cdots & I & 0 \\ \vdots & \iddots & \iddots & \vdots \\ I & \cdots & 0 & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \mathcal{O}_{l-1} \end{bmatrix} \begin{bmatrix} D \\ B \end{bmatrix}. \tag{5.10}$$

6. *Obtain the set of all state-space models*

$$\left\{ (TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R}) \right\} \in SM_n^{p,m} / GL_n(\mathbb{R}).$$

Figure 5.1 *A schematic overview of the one-block subspace algorithm.*

Here, U_p has k block rows while U_f has l block rows. In addition, Y_p and Y_f are defined similarly. On the other hand, we define two vectors, each consisting of one block row, $X_p = X_0^{k+l}$ and $X_f = X_k^{k+l}$. The suffices p and f are mnemonic, representing 'past' and 'future', respectively. Finally, the algorithm can be found on page 51.

Remark 5.2. To see the connection between Algorithm 5.2 and the MOESP PO scheme, we have the following analysis regarding the extraction of $\mathcal{O}_l$, which is less straightforward. In [Ver94], the following RQ factorization is computed:

$$\begin{bmatrix} U_f \\ U_p \\ Y_p \\ Y_f \end{bmatrix} = \begin{bmatrix} R_{11} & & & 0 \\ R_{21} & R_{22} & & \\ R_{31} & R_{32} & R_{33} & \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} = RQ.$$

Then, the SVD of $\begin{bmatrix} R_{42} & R_{43} \end{bmatrix}$ is computed:

$$\begin{bmatrix} R_{42} & R_{43} \end{bmatrix} = \tilde{\Gamma}\tilde{\Sigma}\tilde{\Omega}^*,$$

where $\mathcal{O}_l$ is extracted from $\tilde{\Gamma}$. To begin our analysis, we first see that

$$\Pi_{\mathcal{S}} Y_f = R_{41}Q_1 + R_{42}Q_2 + R_{43}Q_3,$$

and

$$\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} Y_f = R_{42}Q_2 + R_{43}Q_3.$$

Then , it can easily be checked that $\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} Y_f$ has the SVD

$$\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} Y_f = \begin{bmatrix} 0 & R_{42} & R_{43} & 0 \end{bmatrix} Q = \tilde{\Gamma} \begin{bmatrix} \tilde{\Sigma} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & \tilde{\Omega}^* & 0 \\ I & 0 & 0 \\ 0 & 0 & I \end{bmatrix} Q.$$

Now, it is evident that $\tilde{\Gamma} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix}$, where $\begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix}$ is defined in Equation (5.11). △

5.4 Spaces of Input and State Sequences

This section presents the central material needed for the construction of informative experiments, for which requirements in the deterministic setup are covered in Section 5.5 and the stochastic output disturbance setup in Section 5.8.

Persistent excitation and sufficient richness have been used extensively in adaptive control and identification, for example, see [Moo83, Mar84, BS85, Lju87, SS89] and references therein. Here, we define sufficient richness as the following, with U^l defined as in Equation (5.6).

Definition 5.3. A signal u is said to be sufficiently rich (S.R.) of order l, if $l \leq N$ is the largest value such that $\mathbf{E}\left(U^l\, U^{l*}\right) > 0$.

Algorithm 5.2. Two-Block Subspace Algorithm

1. *Construct* U_p, U_f *and* Y_p, Y_f *as in Equation* (5.7).
2. *Compute the SVD of* $\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}Y_f$ *for* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$ *as*

$$\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}Y_f = \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}, \quad \text{where} \quad \Gamma_1 = \begin{bmatrix}\gamma_1\\ \vdots\\ \gamma_l\end{bmatrix}. \tag{5.11}$$

3. *Compute* A *and* C *from*

$$\begin{bmatrix}\gamma_1\\ \vdots\\ \gamma_{l-1}\end{bmatrix} A = \begin{bmatrix}\gamma_2\\ \vdots\\ \gamma_l\end{bmatrix} \qquad \text{and} \qquad C = \gamma_1.$$

4. *Compute* $S_0, \ldots, S_{l-1}$ *from*

$$\mathcal{O}_l^\perp \Pi_{\mathcal{S}} Y_f = S_0 U_k^{k+l} + \cdots + S_{l-1} U_{k+l-1}^{k+l},$$

where

$$\mathcal{O}_l^\perp := \begin{bmatrix}P_0 & \cdots P_{l-2} & I\end{bmatrix} \qquad \text{and} \qquad \mathcal{O}_l^\perp \mathcal{O}_l = 0.$$

5. *Determine* B *and* D *from*

$$\begin{bmatrix}S_0\\ S_1\\ \vdots\\ S_{l-1}\end{bmatrix} = \begin{bmatrix}P_0 & \cdots & P_{l-2} & I\\ P_1 & \cdots & I & 0\\ \vdots & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \vdots\\ I & \cdots & 0 & 0\end{bmatrix}\begin{bmatrix}I & 0\\ 0 & \mathcal{O}_{l-1}\end{bmatrix}\begin{bmatrix}D\\ B\end{bmatrix}.$$

6. *Obtain the set of all state-space models*

$$\left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\} \in SM_n^{p,m}/GL_n(\mathbb{R}).$$

Figure 5.2 *A schematic overview of the two-block subspace algorithm.*

Remark 5.3. The definition of sufficient richness in Definition 5.3 is essentially equivalent to the standard definition of persistent excitation (c.f. [VODM96b, p. 39], Definition 5). The major difference is that the order of such excitation is clearly defined in Definition 5.3. △

In [Mar84, BS85], it has been shown that for any controllable system of order n, if input signals are S.R. of order $n+1$, the corresponding state signals will be S.R. of order 1. In the following, we will analyze the intersection of input signals and their corresponding state signals under all S.R. conditions of the input signals (for S.R. conditions both greater and less than $n+1$).

Note that $A \in \mathbb{R}^{n\times n}$, there exists a principal ideal of all polynomials which annihilates A (for example see [HK71]). (An annihilating polynomial is a polynomial $p(z)$ over the scalar field such that $p(A) = 0$.) The minimal polynomial $p_{\min}(z)$ (or the A-annihilator) is the monic generator of this ideal. Furthermore, let $\mathcal{R}$ be a subspace of $\mathbb{R}^n$ invariant under A. Then, the A-conductor of $\mathbb{R}^n$ into $\mathcal{R}$ is the monic polynomial $\tilde{p}_{\min}(z)$ of least degree such that $\tilde{p}_{\min}(A)\mathbb{R}^n \subset \mathcal{R}$ (see [HK71] for detail). The following lemma states a fact about A-conductor, whose proof is given in Appendix B.

Lemma 5.3. *Let $\mathcal{Q}$ be a subspace of $\mathbb{R}^n$ invariant under A. Then, the A-conductor of $\mathbb{R}^n$ into $\mathcal{Q}$ is also the $\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}$-conductor of $\mathbb{R}^n$ into $\mathcal{Q}$.*

Remark 5.4. Note that $\Pi_{\mathcal{Q}}$ is an operator mapping $\mathbb{R}^n$ to $\mathbb{R}^n$. Suppose Q_b is a matrix whose columns constitute a basis for $\mathcal{Q}$, then $\Pi_{\mathcal{Q}}$ has a $\mathbb{R}^{n\times n}$ matrix representation $Q := Q_b(Q_b^* Q_b)^{-1}Q_b^*$. Thus, the projection of a vector $\nu \in \mathbb{R}^n$ onto $\mathcal{Q}$ becomes $\Pi_{\mathcal{Q}}\nu = Q\nu$. Moreover, we write QX_0^l to represent a usual real matrix multiplication by Q of X_0^l. △

The following give two closely related theorems. The first one covers the case where $\Sigma^w = 0$ which will be used to prove Theorems 5.6 and 5.7 in the deterministic case, whereas the second one covers a more general case which will be used to prove Theorem 5.18 in the combined deterministic-stochastic case in a later section.

Theorem 5.4. *Given an input signal u and some state signal x satisfying Equation (2.1) with $\Sigma^w = 0$, where (A, B) is a reachable pair, with the A-annihilator of degree r. Suppose u is S.R. of order $l+h$, for some $h > 0$. Then*

1. $\dim\left\{\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l\right\} = n$ *if* $h \geq r$*;*
2. $\dim\left\{\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l\right\} \geq n - \tilde{n}$ *if* $h < r$*;*

where $\tilde{n}$ is the dimension of some A-invariant subspace $\mathcal{Q} \subset \mathbb{R}^n$, with the degree of the A-conductor of $\mathbb{R}^n$ into $\mathcal{Q}$ being h.

The proof of the above theorem can be found in Appendix B. The next theorem covers the case where $(A\,,\,[B \quad (\Sigma^w)^{\frac{1}{2}}\,])$ is reachable, whose proof can also be found in Appendix B.

Theorem 5.5. *Given an input signal u and some state signal x satisfying Equations* (2.1)*, where* $(A\,, [B \quad (\Sigma^w)^{\frac{1}{2}}])$ *is a reachable pair, with the A-annihilator of degree r. Suppose u is S.R. of order $l+h$, for some $h>0$. Then*

1. $\dim\{\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l\} = n$ *if* $h \geq r$*;*
2. $\dim\{\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l\} \geq n - \tilde{n}$ *if* $h < r$*;*

where $\tilde{n}$ is the dimension of some A-invariant subspace $\mathcal{Q} \subset \mathbb{R}^n$, with the degree of the A-conductor of $\mathbb{R}^n$ into $\mathcal{Q}$ being h.

Remark 5.5. In fact, it can be seen that if u is S.R. of order no smaller than $l+r$, then U^l and X_0^l will have an empty intersection. On the other hand, if u is S.R. of order smaller than $l+r$, then U^l and $Q^\perp X_0^l$ will have an empty intersection, where $Q^\perp$ is a matrix respresentation of $\Pi_{\mathcal{Q}^\perp}$. In mathematical terms, these can be written as:

1. $\mathrm{span}\{\, \alpha^* U^l \;\mid \alpha \in \mathbb{R}^{ml}\,\} \cap \mathrm{span}\{\, \alpha^* X_0^l \;\mid \alpha \in \mathbb{R}^n\,\} \;=\; \{0\} \quad \text{for } h \geq r;$
2. $\mathrm{span}\{\, \alpha^* U^l \;\mid \alpha \in \mathbb{R}^{ml}\,\} \cap \mathrm{span}\{\, \alpha^* Q^\perp X_0^l \;\mid \alpha \in \mathbb{R}^n\,\} \;=\; \{0\} \quad \text{for } h < r;$

△

5.5 Deterministic Requirements for Informative Experiments

Consider an n-order single-input single-output (SISO) linear system. It is known that such a system can be represented by an n-th order proper rational function, or an n-order ARMAX model. Thus, it is easy to see that a SISO system has $2n+1$ parameters. As shown in [KC83], an n-th order rational function can be determined by any input signal S.R. of order $2n+1$. Accordingly, it is possible to identify a multi-input multi-output (MIMO) linear system by first identifying all its input-output channels using an input signal S.R. of order $2n+1$. This MIMO system can then be recovered by congregating all the channels together. This identification procedure was also applied in [MA95].

In this section, sufficient conditions for informative experiments for subspace methods are presented for the two subspace algorithms presented in Section 5.3. To conduct the analysis, we shall confine the domain $\mathcal{D}_{\mathscr{X}}$ of data sets (c.f. Mapping (5.5)) to all data sets which are generated by systems in $SM_n^{p,m}$ with some initial condition $x(0)$. Obviously, for deterministic identification we have $\Sigma^s = \Sigma^v = 0$. Moreover, we recall that the observability index is defined to be the minimal value of h such that $\mathcal{O}_h$ is injective [Kai80].

The exact input conditions for Algorithm 5.1 are stated in the following theorem. It can be seen that for the deterministic case, the excitation of the input signal is the key point for acquiring an informative experiment.

Theorem 5.6. *Let* $\mathscr{Z}^N$ *be an input-output signal pair of a minimal system* (2.1), *with*

1. $\Sigma^s = \Sigma^v = 0$;
2. *the degree of the* A*-annihilator no greater than* r;
3. *the observability index of the system no greater than* $l-1$;
4. u *S.R. of order* $l+r$.

Then, Algorithm 5.1 ($\mathcal{F}$*) in Section 5.3.1 maps*

$$\mathcal{F} : \mathscr{Z}^N \mapsto \{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}.$$

The proof of this theorem can be found in Appendix B.

Remark 5.6. From the above theorem, we see that S.R. of order $2n+1$ is not the smallest requirement which uniquely specifies an n-th order MIMO system. In general, S.R. of order $l+r$ will be sufficient. △

The exact input conditions for Algorithm 5.2 are stated in the following theorem. Again, the excitation of the input signal is the key point for acquiring an informative experiment. However, since an instrumental variable $\mathcal{S}$ is used, the input excitation is increased to account for the extra past data block. It is also noteworthy that the number of block rows in the past data block must be no smaller than the observability index, while the number of block rows in the future data block must be greater than the observability index.

Theorem 5.7. *Let* $\mathscr{Z}^N$ *be an input-output signal pair of a minimal system* (2.1), *with*

1. $\Sigma^s = \Sigma^v = 0$;
2. *the degree of the* A*-annihilator no greater than* r;
3. *the observability index of the system no greater than* $\min\{k, l-1\}$;
4. u *S.R. of order* $k+l+r$.

Then, Algorithm 5.2 ($\mathcal{F}$*) in Section 5.3.2 maps*

$$\mathcal{F} : \mathscr{Z}^N \mapsto \{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}.$$

Proof. Notice that if $\Pi_{\mathcal{S}} Y_f$ and U_f in Algorithm 5.2 are replaced by Y^l and U^l, the algorithm will then be equivalent to Algorithm 5.1. Thus, in this proof we shall show that Conditions 1-4 in this theorem do imply those in Theorem 5.6 and $\Pi_{\mathcal{S}} Y_f = Y_f$.

First of all, Conditions 1 and 2 in this theorem are identical to those in Theorem 5.6. Moreover, Condition 3 in this theorem implies that in Theorem 5.6. For Condition 4, one can check that U_f is S.R. of order $l + r$ by a straightforward verification of Definition 5.3. Now, it remains to prove $\Pi_{\mathcal{S}} Y_f = Y_f$, or equivalently, $\mathcal{Y}_f \subset \mathcal{S}$.

From Equation (3.12), we can see $\mathcal{Y}_f \subset \mathcal{X}_f + \mathcal{U}_f$. Note that $X_f = A^k X_p + \mathcal{C}_k^d U_p$, which in turn implies $\mathcal{X}_f \subset \mathcal{X}_p + \mathcal{U}_p$. Combining the two inclusions gives $\mathcal{Y}_f \subset \mathcal{X}_p + \mathcal{U}_p + \mathcal{U}_f$. Noticing $Y_p = \mathcal{O}_k X_p + \mathcal{T}_k U_p$ and $\mathcal{O}_k X_p = Y_p - \mathcal{T}_k U_p$, we have

$$\mathcal{Y}_p \subset \mathcal{X}_p + \mathcal{U}_p \qquad \text{and} \qquad \mathcal{X}_p \subset \mathcal{Y}_p + \mathcal{U}_p,$$

where the latter inclusion stems from the fact that $\mathcal{O}_k$ is injective. Adding $\mathcal{U}_p$ to both sides gives

$$\mathcal{Y}_p + \mathcal{U}_p = \mathcal{X}_p + \mathcal{U}_p. \tag{5.12}$$

That is, $\mathcal{Y}_f \subset \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f = \mathcal{S}$, which completes the proof. □

5.6 Decomposition into Deterministic/Stochastic Subsystems

Before we present the combined deterministic-stochastic requirements for informative experiments, we first draw our attention to two system decompositions. Such decompositions, which make use of the concept developed in Chapter 4, provide two approaches for the analysis of the spaces of input and projected state sequences in the next section.

Suppose the input-output signals $u(t)$ and $y(t)$ are given for all $t \in \mathbb{Z}_+$. (Extension to the case where signals are of finite length is obvious.) We divide $x(0)$, as in Chapter 4, into three components: $\hat{x}(0)$, $\tilde{x}(0)$, and $x^s(0)$. Here, $\hat{x}(0)$ and $\tilde{x}(0)$ are defined as:

$$\hat{x}(0) := \Pi_{\mathcal{R}} x(0), \qquad \text{and} \qquad \tilde{x}(0) := x^d(0) - \hat{x}(0),$$

where

$$\mathcal{R} := \operatorname{span}\{u(t)\,,\, y^d(t) \cap y(t)\}_{t \in \mathbb{Z}_+}.$$

Note that for the case where $(\, A\,,\, [\Sigma^w]^{\frac{1}{2}}\,)$ is reachable or $\Sigma^v > 0$, we have $y^d(t) \cap y(t) = \{0\}$ as derived in Chapter 4. With such a decomposition we have

$$\hat{x}(0),\ \tilde{x}(0),\ x^s(0),\ w(t) \text{ and } v(t) \text{ are mutually orthogonal;} \tag{5.13a}$$

$$\tilde{x}(0) \text{ and } x^s(0) \text{ are orthogonal to } u(t) \text{ and } (y^d(t) \cap y(t)). \tag{5.13b}$$

This decomposition of the initial state gives rise to two different system decompositions, as will be discussed in the following subsections.

5.6.1 Decomposition with a Non-Stationary Stochastic Subsystem

In this subsection, we treat $\tilde{x}(0)$, which originally belongs to the deterministic initial condition $x^d(0) = \hat{x}(0) + \tilde{x}(0)$, as part of the initial condition in the stochastic subsystem, creating a non-stationary stochastic subsystem with $x^s(0)$.

Define the initial state of the non-stationary stochastic subsystem to be $x^{ns}(0) := x^s(0) + \tilde{x}(0)$, and the initial state of the complementary deterministic subsystem to be $x^{nd}(0) := \hat{x}(0)$. Note that the additional superscript n signifies the non-stationary property of the stochastic subsystem. In this way, Equations (2.1) can be separated into two subsystems:

$$\begin{aligned} x(t) &= x^{nd}(t) + x^{ns}(t), \\ y(t) &= y^{nd}(t) + y^{ns}(t). \end{aligned}$$

The non-stationary stochastic subsystem has the state-space equations:

$$\begin{aligned} x^{ns}(t+1) &= Ax^{ns}(t) + w(t), \\ y^{ns}(t) &= Cx^{ns}(t) + v(t); \end{aligned}$$

whereas the complementary deterministic subsystem has the state-space equations:

$$\begin{aligned} x^{nd}(t+1) &= Ax^{nd}(t) + Bu(t), \\ y^{nd}(t) &= Cx^{nd}(t) + Du(t). \end{aligned}$$

Using Statements (5.13), we have $x^{nd}(t_1) \perp x^{ns}(t_2)$, $t_1, t_2 \in \mathbb{Z}_+$. In other words, the non-stationary stochastic subsystem will be uncorrelated with its deterministic complement. Furthermore, this non-stationary stochastic subsystem can be envisaged as the original stochastic subsystem except that the initial condition is changed from $x^s(0)$ to $x^{ns}(0) := x^s(0) + \tilde{x}(0)$.

With Statement (5.13a), it is easy to see that $\tilde{x}(0) \perp \{x^s(t), y^s(t)\}$. We introduce a new matrix $\tilde{G}_t$, for $t \in \mathbb{N}$, as follows:

$$\begin{aligned} \tilde{G}_t &:= \mathbf{E}\big(\big[x^{ns}(t+1)\big] \big[y^{ns}(t)\big]^* \big) \\ &= \mathbf{E}\big(\big[x^s(t+1) + A^{t+1}\tilde{x}(0)\big] \big[y^s(t) + CA^t\tilde{x}(0)\big]^* \big). \end{aligned}$$

Now, define $\tilde{P}_0 := \mathbf{E}\big[\,\tilde{x}(0)\,\tilde{x}^*(0)\,\big]$. Then,

$$\begin{aligned} \tilde{G}_t &= A\Sigma^s C^* + \Sigma^{wv} + A^{t+1}\tilde{P}_0(A^t)^*C^* \\ &= A\big[\Sigma^s + A^t\tilde{P}_0(A^t)^*\big]C^* + \Sigma^{wv} \\ &= A\tilde{\Sigma}^s_t C^* + \Sigma^{wv}, \end{aligned} \tag{5.15}$$

where $\tilde{\Sigma}^s_t = \Sigma^s + A^t\tilde{P}_0(A^t)^*$ is the non-stationary stochastic state covariance matrix. Note that $\tilde{G}_t$ plays a similar role as G, which, with A, forms the non-stationary stochastic controllability pair $(A, \tilde{G}_t)$.

5.6.2 Decomposition with a Kalman Filter Deterministic Subsystem

Again, we decompose $x(0)$ into $\hat{x}(0) + \tilde{x}(0) + x^s(0)$, as in Statements (5.13). Define a new state variable $\tilde{x}^s(t) := \tilde{x}(t) + x^s(t) := x(t) - \hat{x}(t)$. Thus, we have the following system decomposition:

$$\begin{aligned} x(t) &= \hat{x}(t) + \tilde{x}^s(t), \\ y(t) &= \hat{y}(t) + \tilde{y}^s(t). \end{aligned}$$

The Kalman filter deterministic subsystem has the state-space equations:

$$\begin{aligned} \hat{x}(t+1) &= (A - K_t C)\hat{x}(t) + (B - K_t D)u(t) + K_t y(t), \\ \hat{y}(t) &= C\hat{x}(t) + Du(t), \end{aligned}$$

where K_t and $\tilde{P}_0$ are defined in Proposition 4.5. On the other hand, the complementary error subsystem has the state-space equations:

$$\begin{aligned} \tilde{x}^s(t+1) &= (A + K_t C)\tilde{x}^s(t) + w(t) + v(t); \\ \tilde{y}^s(t) &= C\tilde{x}^s(t) + v(t). \end{aligned}$$

For convenience, we adopt, for $t_2 > t_1$, the notation:

$$\Phi(t_2, t_1) := (A - K_{t_2-1}C)(A - K_{t_2-2}C)\dots(A - K_{t_1}C). \tag{5.16}$$

5.7 Spaces of Input and Projected State Sequences

This section presents a combined deterministic-stochastic version of Section 5.4. Due to the presence of stochastic disturbances, system equations are projected onto some instrumental variable space $\mathcal{S}$, so as to eliminate the effect of stochastic noises. It has been seen that this instrumental variable space is usually taken to be $\mathcal{Y}_f + \mathcal{U}_p + \mathcal{U}_f$ in subspace methods. Owing to this fact, the emptiness of the intersection of $\Pi_{\mathcal{S}} X_f$ and U_f is desired. More precise statements of this condition are given in the following proposition in various forms.

Proposition 5.8. *For $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$, the following are equivalent:*

1. $\mathbf{E}\left[U_f U_f^*\right] > 0,$ *and* $\dim\left\{\Pi_{\mathcal{S}}\Pi_{\mathcal{U}_f^\perp} X_f\right\} = n.$
2. $\mathbf{E}\left[U_f U_f^*\right] > 0,$ *and* $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}} X_f\right\} = n.$
3. $\mathbf{E}\left[\begin{pmatrix}\Pi_{\mathcal{S}} X_f \\ U_f\end{pmatrix}\begin{pmatrix}\Pi_{\mathcal{S}} X_f \\ U_f\end{pmatrix}^*\right] > 0.$
4. $\mathbf{E}\left[\begin{pmatrix}X_f \\ U_f\end{pmatrix}\begin{pmatrix}Y_p^* & U_p^* & U_f^*\end{pmatrix}\right]$ *is of full row rank.*

The proof of this proposition can be found in Appendix B. This section is organized into three subsections as follows. Section 5.7.1 covers the case where $\Sigma^w = 0$, while Section 5.7.2 covers the case where $(\,A\,,\,[\Sigma^w]^{\frac{1}{2}}\,)$ is reachable. Finally, Section 5.7.3 presents the generalized case where $(\,A\,,\,[\,B \quad (\Sigma^w)^{\frac{1}{2}}\,]\,)$ is reachable.

5.7.1 Stochastic Output Disturbance

Systems with only stochastic output disturbance are also known as output error systems. For such systems, we simply have $\Sigma^w = 0$. Since internal noise excitation is absent from the system, we rely on the external input excitation; in which case, we require (A, B) to be reachable. First, we have this straightforward lemma.

Lemma 5.9. *Let (A, B) be reachable. Then, $\begin{bmatrix} A^k & A^{k-1}B & \cdots & B \end{bmatrix}$ is of full row rank.*

The proof of this lemma can be found in Appendix B. The next theorem gives an input excitation requirement such that the conditions in Proposition 5.8 are satisfied. Note that these requirements are exactly the same as those in Theorem 5.7 for the deterministic case. In other words, input excitation is the key point for acquiring informative experiments.

Theorem 5.10. *Let $\mathscr{Z}^N$ be an input-output signal pair of a system* (2.1), *with*

1. *$\Sigma^w = 0$ and (A, B) reachable;*
2. *r no smaller than the degree of the A-annihilator;*
3. *k no smaller than the observability index of the system;*
4. *the input signal u S.R. of order $k + l + r$.*

Then, $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}X_f\right\} = n$, *where* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.

Proof. Note that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$ is equivalent to the space spanned by U_p, U_f and $(Y_p - \mathcal{T}_k U_p) = (\mathcal{O}_k X_p + V_p)$. Thus, by Proposition 5.8 it is equivalent to show that

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \begin{pmatrix} [\mathcal{O}_k X_p + V_p]^* & U_p^* & U_f^* \end{pmatrix}\right] = \mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \begin{pmatrix} X_p^*\mathcal{O}_k^* & U_p^* & U_f^* \end{pmatrix}\right]$$

is of full row rank. Note that the above equality stems from the fact that V_p is orthogonal to X_f and U_f. Finally, realizing that $X_f = A^k X_p + \mathcal{C}_k^d U_p$ gives

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \begin{pmatrix} X_p^*\mathcal{O}_k^* & U_p^* & U_f^* \end{pmatrix}\right] = \begin{bmatrix} A^k & \mathcal{C}_k^d & 0 \\ 0 & 0 & I \end{bmatrix} \mathbf{E}\left[\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix} \begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}^*\right] \begin{bmatrix} \mathcal{O}_k^* & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix}.$$

Note that the covariance matrix above is positive definite by Theorem 5.4. In addition, the first term is of full row rank by Lemma 5.9. Thus, the above expression, as a product of full row rank matrices, is also of full row rank. □

5.7.2 Stochastic State Excitation

This subsection covers the case where $(\,A\,,\,[\Sigma^w]^{\frac{1}{2}}\,)$ is reachable. In other words, this corresponds to systems whose states are fully stochastically excited. For such systems, we require the stochastic controllability pair (A, G) to be reachable. In effect, we rely on the internal stochastic noises to excite the state sequences. First of all, we shall show that under the assumption that the stochastic subsystem is stationary, the system matrix A must be asymptotically stable. Although it is a rather well-known result, we still give a proof for the lemma here.

Lemma 5.11. *Suppose $(\,A\,,\,[\Sigma^w]^{\frac{1}{2}}\,)$ reachable. Then, A is asymptotically stable if and only if*

$$0 < \Sigma^s = A\Sigma^s A^* + \Sigma^w < \infty.$$

Proof. Note that we can write

$$\Sigma^s = A^n \Sigma^s (A^n)^* + \sum_{j=0}^{n-1} A^j \Sigma^w (A^j)^*.$$

Since $(\,A\,,\,[\Sigma^w]^{\frac{1}{2}}\,)$ is reachable, we have $\sum_{j=0}^{n-1} A^j \Sigma^w (A^j)^* > 0$. Now, by standard properties of Lyapunov equations, $\Sigma^s > 0$ and exists if and only if A^n is asymptotically stable, or equivalently, A is asymptotically stable. □

Before proving Theorem 5.13 regarding the controllability of some non-steady stochastic pairs, we first present the following lemma.

Lemma 5.12. *Let $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ be the A-annihilator, where A is asymptotically stable. Then, the $n \times n$ matrix $I + a_{r-1}A + \cdots + a_0 A^r$ is non-singular.*

Proof. Suppose $p_{\min}(z)$ can be factorized as $p_{\min}(z) = (z - s_1)(z - s_2)\cdots(z - s_r)$, where $s_1, \ldots, s_r$ are eigenvalues of A. Then, we have

$$I + a_{r-1}A + \cdots + a_0 A^r = (I - s_1 A)(I - s_2 A)\cdots(I - s_r A).$$

Therefore, $I + a_{r-1}A + \cdots + a_0 A^r$ is singular if one of the $(I - s_j A)$ is singular. Now suppose $s_j = 0$, then certainly $(I - s_j A) = I$ is not singular. On the other hand, suppose $s_j \neq 0$. Then, $(I - s_j A)$ is singular if and only if $(s_j^{-1} I - A)$ is singular, or equivalently, s_j^{-1} is an eigenvalue of A. However, since A is asymptotically stable, we have $|s_j| < 1$. Therefore, s_j^{-1}, where $|s_j^{-1}| > 1$, cannot be an eigenvalue of A. In other words, each of $(I - s_j A)$ is non-singular; and hence, the result follows. □

In the following, the controllability of the non-steady pairs $(A, \tilde{G}_t)$ and $(\Phi(t_2, t_1), K_{t_1})$, defined in Section 5.6, is investigated. These pairs, changing with respect to time, can be seen as time-varying stochastic controllability pairs. The following theorem shows that if a longer time period is allowed, these pairs are in fact reachable.

Theorem 5.13. *Suppose each of the following holds:*

1. $(A, [\Sigma^w]^{\frac{1}{2}})$ *and* (A, G) *are reachable;*
2. k *is no smaller than two times the degree of the* A*-annihilator.*

Then for $\tilde{G}.$ *and* $\Phi(k,\cdot)$ *respectively defined in Equations* (5.15) *and* (5.16)*,*

1. $\begin{bmatrix} A^{k-1}\tilde{G}_0 & A^{k-2}\tilde{G}_1 & \cdots & \tilde{G}_{k-1} \end{bmatrix}$ *is of full row rank;*
2. $\begin{bmatrix} \Phi(k,0)K_0 & \Phi(k,1)K_1 & \cdots & K_{k-1} \end{bmatrix}$ *is of full row rank.*

Proof. Let r be the degree of the A-annihilator and $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ the A-annihilator. So, we have $k \geq 2r$. Define the Toeplitz matrix ψ as:

$$\psi := \begin{pmatrix} A^{k-r-1}G + A^k\tilde{P}_0(A^r)^*C^* & A^{k-r-2}G + A^k\tilde{P}_0(A^{r+1})^*C^* & \cdots & G + A^k\tilde{P}_0(A^{k-1})^*C^* \\ A^{k-r}G + A^k\tilde{P}_0(A^{r-1})^*C^* & A^{k-r-1}G + A^k\tilde{P}_0(A^r)^*C^* & \cdots & AG + A^k\tilde{P}_0(A^{k-2})^*C^* \\ \ddots & \ddots & \ddots & \ddots \\ A^{k-1}G + A^k\tilde{P}_0C^* & A^{k-2}G + A^k\tilde{P}_0A^*C^* & \cdots & A^rG + A^k\tilde{P}_0(A^{k-r-1})^*C^* \end{pmatrix}.$$

Since $p_{\min}(A^*) = 0$, we have

$$A^k\tilde{P}_0 \cdot p_{\min}(A^*) \cdot (A^h)^*C^* = 0,$$

for any $h \in \mathbb{Z}_+$. Thus,

$$\begin{aligned} \begin{bmatrix} I & a_{r-1}I & \cdots & a_0I \end{bmatrix} \psi &= \begin{bmatrix} I & a_{r-1}I & \cdots & a_0I \end{bmatrix} \begin{pmatrix} A^{k-r-1}G & A^{k-r-2}G & \cdots & G \\ A^{k-r}G & A^{k-r-1}G & \cdots & AG \\ \ddots & \ddots & \ddots & \ddots \\ A^{k-1}G & A^{k-2}G & \cdots & A^rG \end{pmatrix} \\ &= (I + a_{r-1}A + \cdots + a_0A^r)\begin{bmatrix} A^{k-r-1}G & A^{k-r-2}G & \cdots & G \end{bmatrix}. \end{aligned}$$

As shown in Lemma 5.11, A must be asymptotically stable for any stationary stochastic process. Consequently, we have $(I + a_{r-1}A + \cdots + a_0A^r)$ non-singular by Lemma 5.12. Finally, since $k - r - 1 \geq r - 1$ and (A, G) is reachable, it implies

$$\operatorname{rank}\left\{ \begin{bmatrix} A^{k-r-1}G & A^{k-r-2} & \cdots & G \end{bmatrix} \right\} = n.$$

Now, it is easy to see that

$$\operatorname{rank}\left\{ \begin{bmatrix} I & a_{r-1}I & \cdots & a_0I \end{bmatrix} \psi \right\} = n. \tag{5.17}$$

(1): Statement 1 will be proven by contradiction. Suppose there exists $\nu \neq 0$ such that

$$\nu \begin{bmatrix} A^{k-1}\tilde{G}_0 & A^{k-2}\tilde{G}_1 & \dots & \tilde{G}_{k-1} \end{bmatrix} = 0.$$

This implies $\nu\big(A^{k-j}G + A^k\tilde{P}_0(A^{j-1})^*C^*\big) = 0$, for $j = 1, \dots, k$. Thus,

$$\nu \begin{bmatrix} I & a_{r-1}I & \cdots & a_0 I \end{bmatrix} \psi = \begin{bmatrix} \nu & a_{r-1}\nu & \cdots & a_0\nu \end{bmatrix} \psi = 0,$$

which is clearly a contradiction to Equation 5.17.

(2): By Theorem 3.3 we write

$$\begin{aligned} \mathcal{K}_k &= \begin{bmatrix} (A - K_{k-1}C)\mathcal{K}_{k-1} & K_{k-1} \end{bmatrix} \\ &= \begin{bmatrix} (A - K_{k-1}C)(A - K_{k-2}C)\mathcal{K}_{k-2} & (A - K_{k-1}C)K_{k-2} & K_{k-1} \end{bmatrix} \\ &= \begin{bmatrix} \Phi(k,0)K_0 & \Phi(k,1)K_1 & \cdots & K_{k-1} \end{bmatrix}. \end{aligned}$$

Moreover, we have

$$\begin{aligned} \mathcal{K}_k &= (A^k\tilde{P}_0\mathcal{O}_k^* + \mathcal{C}_k^s)(L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1} \\ &= \big[A^k(\Sigma^s + \tilde{P}_0)C^* + \Sigma^{wv} \quad A^{k-1}(\Sigma^s + A\tilde{P}_0A^*)C^* + \Sigma^{wv} \quad \cdots \\ &\qquad A(\Sigma^s + A^{k-1}\tilde{P}_0A^{k-1})C^* + \Sigma^{wv}\big] \cdot (L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1} \\ &= \begin{bmatrix} A^{k-1}\tilde{G}_0 & A^{k-2}\tilde{G}_1 & \cdots & \tilde{G}_{k-1} \end{bmatrix} (L_k + \mathcal{O}_k\tilde{P}_0\mathcal{O}_k^*)^{-1}. \end{aligned}$$

With Inequality (3.8), we have $L_k > 0$. Thus, Statement 1 and Statement 2 are equivalent.

This completes the proof. □

The next theorem provides the conditions under which the statements in Proposition 5.8 will be satisfied in the case of full stochastic state excitation. Unlike the deterministic case in which we only require a certain degree of input excitation, we require in this case a longer past data block, which can avoid rank cancelation in a similar way to that shown in Theorem 5.13. This ensures the emptiness of the intersection of the input and projected state sequences.

Theorem 5.14. *Let $\mathscr{Z}^N$ be an input-output signal pair of a system* (2.1)*, with*

1. *$\big(A\,,\,(\Sigma^w)^{\frac{1}{2}}\big)$ and (A, G) reachable;*
2. *k no smaller than two times the degree of the A-annihilator;*
3. *the input signal u S.R. of order $k + l$.*

Then, $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}X_f\right\} = n$, *where* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.

The proof of the theorem can be found in Appendix B. From the above theorem, we can see that due to the internal state excitation by the process noise, a lower degree of input excitation is required.

5.7.3 General Excitation and Disturbance

Finally, this subsection will cover a general excitation and disturbance case in which Σ^v is not necessarily positive definite and $\left(A\,,\,(\Sigma^w)^{\frac{1}{2}}\right)$ is not necessarily reachable. It can be seen that requirements for such a setup are a combination of those in Theorem 5.10 for the output disturbance case and those in Theorem 5.14 for the stochastic state excitation case.

Theorem 5.15. *Let $\mathscr{Z}^N$ be an input-output signal pair of a system* (2.1)*, with*

1. *A asymptotically stable;*
2. *$(A\,,[B \quad (\Sigma^w)^{\frac{1}{2}}])$ reachable;*
3. *$(A\,,[\Sigma^w]^{\frac{1}{2}})$ and (A,G) having common reachable modes;*[1]
4. *r no smaller than the degree of the A-annihilator;*
5. *k no smaller than three times the degree of the A-annihilator;*
6. *the input signal u S.R. of order $k+l+r$.*

Then, $\dim\left\{\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}X_f\right\} = n$, *where* $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$.

The proof is tedious, but is in fact nothing more than extending the other two cases, and can be found in Appendix B.

5.8 General Requirements for Informative Experiments

In this section, we will concentrate on Algorithm 5.2 which is capable of identifying systems in a combined deterministic-stochastic case. The following three theorems give sufficient requirements for experiments to be informative with Algorithm 5.2, covering the three cases discussed in the previous section. Similar to Section 5.5, we shall first confine the domain $\mathcal{D}_{\mathscr{Z}}$ of data sets (c.f. Mapping (5.5)) to all data sets which are generated by n-th order systems described by Equations (2.1) with noises $w(t)$ and $v(t)$, and some initial condition $x(0)$.

In Algorithm 5.2, a combined deterministic-stochastic data equation (3.12) is converted to a deterministic-like data equation by projecting it onto the subspace $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$:

$$\Pi_{\mathcal{S}}Y_f = \mathcal{O}_k\Pi_{\mathcal{S}}X_f + \mathcal{T}_kU_f. \tag{5.18}$$

In this way, Algorithm 5.2 is essentially identical to Algorithm 5.1 if $\Pi_{\mathcal{S}}Y_f$ and $\Pi_{\mathcal{S}}X_f$ are replaced by Y^l and X_0^l respectively. Thus, as seen in Theorems 5.6 and 5.7 for the deterministic case,

[1] That is, for any $\lambda\in\mathbb{C}$, $\nu\in\mathbb{C}^n$, we have $\nu\,[\,A-\lambda I \quad (\Sigma^w)^{\frac{1}{2}}\,]=0$ if and only if $\nu\,[\,A-\lambda I \quad G\,]=0$.

the remaining key point for experiments to be informative is to find a condition such that

$$\dim\left\{\Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}}X_f\right\} = n. \tag{5.19}$$

First, the criteria for informative experiments for the case where $\Sigma^w = 0$ are presented in the following theorem.

Theorem 5.16. *Let $\mathscr{Z}^N$ be an input-output signal pair of an observable system* (2.1), *with*

1. $\Sigma^w = 0$ *and* (A, B) *reachable;*
2. r *no smaller than the degree of the A-annihilator;*
3. $\min\{k, l-1\}$ *no smaller than the observability index of the system;*
4. *the input signal u S.R. of order $k + l + r$.*

Then, Algorithm 5.2 ($\mathcal{F}$) in Section 5.3.2 maps

$$\mathcal{F} : \mathscr{Z}^N \mapsto \left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\}.$$

Proof. From Equation (5.18) it is easy to see that $\mathcal{O}_l\ \Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}}X_f = \Gamma_1\Sigma_1\Omega_1^*$, where Γ_1, Σ_1 and Ω_1 are defined in Equation (5.11). With the rank requirement in Equation (5.19) guaranteed by Theorem 5.10, we have

$$\mathcal{O}_l = \Gamma_1 \qquad \text{and} \qquad \Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}}X_f = \Sigma_1\Omega_1^*$$

in some particular state coordinates. Together with $\mathbf{E}\left[U_f\, U_f^*\right] > 0$, the proof is then completed by following that of Theorem 5.6. □

Next, the case where $(\,A\,,\,(\Sigma^w)^{\frac{1}{2}}\,)$ is reachable is considered. In this case, the rank requirement in Equation (5.19) is guaranteed by the conditions in Theorem 5.14. Thus, this theorem can easily be proven by modifying the proof of Theorem 5.16; and hence the proof is omitted here.

Theorem 5.17. *Let $\mathscr{Z}^N$ be an input-output signal pair of an observable system* (2.1), *with*

1. $(\,A\,,\,(\Sigma^w)^{\frac{1}{2}}\,)$ *and* (A, G) *reachable;*
2. k *no smaller than two times the degree of the A-annihilator;*
3. $l - 1$ *no smaller than the observability index of the system;*
4. *the input signal u S.R. of order $k + l$.*

Then, Algorithm 5.2 ($\mathcal{F}$) in Section 5.3.2 maps

$$\mathcal{F} : \mathscr{Z}^N \mapsto \left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\}.$$

Finally, we present the general excitation and disturbance case. As expected, the following theorem can also be derived from the conditions stated in Theorem 5.15; and hence the proof is again omitted here.

Theorem 5.18. *Let $\mathscr{Z}^N$ be an input-output signal pair of an observable system* (2.1)*, with*

1. *A asymptotically stable;*
2. *$(A, [B \quad (\Sigma^w)^{\frac{1}{2}}])$ reachable;*
3. *$(A, [\Sigma^w]^{\frac{1}{2}})$ and (A, G) having common reachable modes;*
4. *r no smaller than the degree of the A-annihilator;*
5. *k no smaller than three times the degree of the A-annihilator;*
6. *$l-1$ no smaller than the observability index of the system;*
7. *the input signal u S.R. of order $k+l+r$.*

Then, Algorithm 5.2 ($\mathcal{F}$) in Section 5.3.2 maps

$$\mathcal{F} : \mathscr{Z}^N \mapsto \left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\}.$$

Remark 5.7. At this point, attention should be drawn to the fact that the criteria presented in this chapter are only sufficient conditions for experiments to be informative. That is, in some cases subspace methods can still identify a system even though these criteria are not satisfied. $\triangle$

5.9 Some Examples

In this section, three examples are presented. The first example illustrates the case where Algorithm 5.2 may fail to identify a system if conditions in Theorem 5.7 are not met. On the other hand, the second example presents a case where conditions in Theorem 5.7 are all met, allowing Algorithm 5.2 to identify the system successfully. Finally, the third example illustrates a deterministic-stochastic cancelation,

$$\dim\left\{\Pi_{\mathcal{U}_f^{\perp}} \Pi_{\mathcal{S}} X_f\right\} < n,$$

in an ideal probabilistic setup, while conditions in Theorem 5.14 are not fully satisfied. It will be seen that such a deterministic-stochastic cancelation usually involves complicated projection relationships and exact annihilation of stochastic excitation. For the sake of simplicity, we have chosen a 1-dimensional system to demonstrate this deterministic-stochastic cancelation.

Example 5.1. Consider the deterministic system ($\Sigma^s = \Sigma^v = 0$)

$$(A, B, C, D) = \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \end{bmatrix}, 1 \right), \qquad \text{with} \quad x(0) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Let the input-output signals be

$$\begin{aligned} \{\, u(j) \,\}_{j=0}^{12} &= \{ \; -1, \; 1, \; 0, \; -1, \; 1, \; 0, \; -1, \; 0, \; 0, \; 0, \; 0, \; 0, \; 0 \; \}, \\ \{\, y(j) \,\}_{j=0}^{12} &= \{ \; -1, \; 1, \; 2, \; -1, \; 1, \; 2, \; -1, \; 0, \; 0, \; 0, \; 0, \; 0, \; 0 \; \}. \end{aligned}$$

We construct U_p, U_f as follows:

$$\begin{bmatrix} U_p \\ \hline U_f \end{bmatrix} := \left[\begin{array}{cccccccc} -1 & 1 & 0 & -1 & 1 & 0 & -1 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 \\ \hline -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right],$$

and Y_p, Y_f as follows:

$$\begin{bmatrix} Y_p \\ \hline Y_f \end{bmatrix} := \left[\begin{array}{cccccccc} -1 & 1 & 2 & -1 & 1 & 2 & -1 & 0 \\ 1 & 2 & -1 & 1 & 2 & -1 & 0 & 0 \\ 2 & -1 & 1 & 2 & -1 & 0 & 0 & 0 \\ \hline -1 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

Some calculations reveal that, for $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$,

$$\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} Y_f = \frac{1}{3} \begin{bmatrix} 2 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \end{bmatrix},$$

which is clearly of rank 1, which is less than $n = 2$. Following Algorithm 5.2, we obtain an incorrect 1-dimensional model:

$$(\hat{A}, \hat{B}, \hat{C}, \hat{D}) = \frac{1}{19} \left(0, \, 6\sqrt{2}, \, 19\sqrt{2}, \, 13 \right).$$

The reason for obtaining a poor model is that the conditions in Theorem 5.7 are not met. Theorem 5.7 requires the input signal u to be S.R. of order 8 in this case. However, the input signal used above is S.R. of order 7. △

Remark 5.8. In [VO95, VODM96b], it is suggested that in deterministic identification, an empty intersection of $\mathcal{U}_f$ and $\mathcal{X}_p$ is sufficient for experiments to be informative when Algorithm 5.2 is employed. However, in Example 5.1 we have

$$\begin{bmatrix} X_p \\ \hline U_f \end{bmatrix} = \left[\begin{array}{cccccccc} 1 & 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 \\ \hline -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

which have empty intersection. In other words, this shows that such a condition is not sufficient in all cases. △

Example 5.2. Consider the deterministic system ($\Sigma^s = \Sigma^v = 0$)

$$(A, B, C, D) = \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \end{bmatrix}, 1 \right), \qquad \text{with} \quad x(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

Let the input-output signals be

$$\begin{aligned} \{ u(j) \}_{j=0}^{14} &= \{ \ 0 \ \ -1, \ 1, \ 0, \ -1, \ 1, \ 0, \ -1, \ 0, \ 0, \ 0, \ 0, \ 0, \ 0, \ 0 \ \}, \\ \{ y(j) \}_{j=0}^{14} &= \{ \ 2 \ \ -1, \ 1, \ 2, \ -1, \ 1, \ 2, \ -1, \ 0, \ 0, \ 0, \ 0, \ 0, \ 0, \ 0 \ \}. \end{aligned}$$

We construct U_p, U_f as follows:

$$\begin{bmatrix} U_p \\ \hline U_f \end{bmatrix} := \left[\begin{array}{cccccccccc} 0 & -1 & 1 & 0 & -1 & 1 & 0 & -1 & 0 & 0 \\ -1 & 1 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ \hline 0 & -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right],$$

and Y_p, Y_f as follows:

$$\begin{bmatrix} Y_p \\ \hline Y_f \end{bmatrix} := \left[\begin{array}{cccccccccc} 2 & -1 & 1 & 2 & -1 & 1 & 2 & -1 & 0 & 0 \\ -1 & 1 & 2 & -1 & 1 & 2 & -1 & 0 & 0 & 0 \\ 1 & 2 & -1 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ \hline 2 & -1 & 1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

Some calculations reveal that, for $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$,

$$\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} Y_f = \frac{1}{4} \begin{bmatrix} 2 & 4 & 2 & 2 & -2 & 0 & 0 & 0 & 0 & 0 \\ 3 & 2 & 3 & -1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 2 & 4 & 2 & 2 & -2 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

which is of rank $n = 2$. Note that u is clearly S.R. of order 8. In other words, the conditions in Theorem 5.7 are fully met. Following Algorithm 5.2, it can easily be checked that a correct 2-dimensional model is obtained. $\triangle$

In the following, we give an example in an ideal probabilistic setup where Condition 4 (k no smaller than two times the degree of the A-annihilator) of Theorem 5.14 is not met, yielding $\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}X_f = 0$ as a result. Note that a 1-dimensional system is chosen for a simpler and clearer exposition. Applying Algorithm 5.2 with $k = 1$ and $l = 2$, we then have $\Pi_{\mathcal{U}_f^\perp}\Pi_{\mathcal{S}}Y_f = 0$, in which case a deficient model is certainly obtained. Once again, to avoid this deterministic-stochastic cancelation, one may ensure the conditions in Section 5.8 are satisfied.

Example 5.3. Consider a system (2.1):

$$(A, B, C, D) = (\,0.8,\ -0.8,\ 0.4,\ 0\,),$$

with

$$\mathbf{E}\left[\begin{pmatrix} w(t) \\ v(t) \end{pmatrix}\begin{pmatrix} w(\tau) \\ v(\tau) \end{pmatrix}^*\right] = \begin{bmatrix} 0.36 & -0.48 \\ -0.48 & 1 \end{bmatrix}\delta_{t\tau} = \begin{bmatrix} 0.6 & 0 \\ -0.8 & 0.6 \end{bmatrix}\begin{bmatrix} 0.6 & -0.8 \\ 0 & 0.6 \end{bmatrix}\delta_{t\tau}.$$

Assume that $\{e_j\} = \{e_d, e_s, e_0, e_1, e_2, e_3, e_4, e_5\}$ is a white noise process with all elements mutually orthogonal to one another, i.e. $\mathbf{E}\,(e_i\, e_j) = \delta_{ij}$. Let

$$\begin{array}{lll} x^d(0) = \sqrt{3}e_d, & x^s(0) = e_s, & x(0) = \sqrt{3}e_d + e_s; \\ u(0) = 0.5\sqrt{3}e_d, & w(0) = 0.6e_0, & v(0) = -0.8e_0 + 0.6e_3; \\ u(1) = 0.4\sqrt{3}e_d, & w(1) = 0.6e_1, & v(1) = -0.8e_1 + 0.6e_4; \\ u(2) = 0.5\sqrt{3}e_d, & w(2) = 0.6e_2, & v(2) = -0.8e_2 + 0.6e_5. \end{array}$$

Note that the stochastic subsystem satisfies

$$\Sigma^s = \mathbf{E}[x^s(t)x^s(t)] = 1, \qquad \Sigma^s - A\Sigma^s A^* = 1 - 0.64 = 0.36 = \Sigma^w.$$

Simple calculations show

$$\begin{array}{rcll} x(0) & = & \sqrt{3}e_d & + \ e_s, \\ x(1) & = & 0.4\sqrt{3}e_d & + \ 0.8e_s + 0.6e_0, \\ x(2) & = & 0 & + \ 0.64e_s + 0.48e_0 + 0.6e_1, \\ x(3) & = & -0.4\sqrt{3}e_d & + \ 0.512e_s + 0.384e_0 + 0.48e_1 + 0.6e_2; \end{array}$$

$$\begin{array}{rcll} y(0) & = & 0.4\sqrt{3}e_d & + \ 0.4e_s - 0.8e_0 + 0.6e_3, \\ y(1) & = & 0.16\sqrt{3}e_d & + \ 0.32e_s + 0.24e_0 - 0.8e_1 + 0.6e_4, \\ y(2) & = & 0 & + \ 0.256e_s + 0.192e_0 + 0.24e_1 - 0.8e_2 + 0.6e_5. \end{array}$$

Let

$$\begin{bmatrix} X_p \\ X_f \end{bmatrix} = \begin{bmatrix} x(0) & x(1) & x(2) \\ x(1) & x(2) & x(3) \end{bmatrix},$$

and

$$\begin{bmatrix} U_p \\ \hdashline U_f \end{bmatrix} = \begin{bmatrix} u(0) & u(1) & u(2) \\ \hdashline u(1) & u(2) & u(3) \\ u(2) & u(3) & u(4) \end{bmatrix}, \qquad \begin{bmatrix} Y_p \\ \hdashline Y_f \end{bmatrix} = \begin{bmatrix} y(0) & y(1) & y(2) \\ \hdashline y(1) & y(2) & y(3) \\ y(2) & y(3) & y(4) \end{bmatrix}.$$

With some calculations, it is easily seen that

$$\begin{aligned} \mathbf{E}\left[X_f U_p^* \right] &= \mathbf{E}[x(1)u(0)] + \mathbf{E}[x(2)u(1)] + \mathbf{E}[x(3)u(2)] \\ &= (0.6) + (0) + (-0.6) \\ &= 0; \\ \mathbf{E}\left[X_f Y_p^* \right] &= \mathbf{E}[x(1)y(0)] + \mathbf{E}[x(2)y(1)] + \mathbf{E}[x(3)y(2)] \\ &= (3(0.16) - 0.16) + (-0.16) + (-0.16) \\ &= 0. \end{aligned}$$

Consequently, $\mathcal{X}_f$ is orthogonal to $\mathcal{U}_p$ and $\mathcal{Y}_p$. This also implies that

$$\Pi_{\mathcal{S}} \mathcal{X}_f \subset \mathcal{U}_f. \tag{5.20}$$

Therefore,

$$\Pi_{\mathcal{U}_f^\perp} \Pi_{\mathcal{S}} X_f = 0. \tag{5.21}$$

Now, we use the geometric approach, as discussed in Chapter 4 and this chapter, to analyze this deterministic-stochastic cancelation. First, note that

$$G = A\Sigma^s C^* + \Sigma^{wv} = 0.32 - 0.48 = -0.16.$$

Observe that

$$\begin{bmatrix} X_p^d \\ U_p^d \end{bmatrix} = \begin{bmatrix} \sqrt{3}e_d & 0.4\sqrt{3}e_d & 0 \\ 0.5\sqrt{3}e_d & 0.4\sqrt{3}e_d & 0.5\sqrt{3}e_d \end{bmatrix}.$$

Suppose we decompose $X_p = \hat{X}_p + \tilde{X}_p + X_p^s$ as follows:

$$\begin{bmatrix} \hat{X}_p \\ \tilde{X}_p \\ X_p^s \end{bmatrix} := \begin{bmatrix} 0.5\sqrt{3}e_d & 0.4\sqrt{3}e_d & 0.5\sqrt{3}e_d \\ 0.5\sqrt{3}e_d & 0 & -0.5\sqrt{3}e_d \\ e_s & 0.8e_s + 0.6e_0 & 0.64e_s + 0.48e_0 + 0.6e_1 \end{bmatrix}.$$

It is noteworthy that $\hat{X}_p$, $\tilde{X}_p$ and X_p^s are mutually orthogonal, and $\hat{X}_p = U_p$. In order to show Inclusion (5.20), we show $\mathbf{E}[\,X_f U_p^*\,] = 0$ and $\mathbf{E}[\,X_f Y_p^*\,] = 0$. First, we have

$$\begin{aligned}\mathbf{E}[\,X_f U_p^*\,] &= A\,\mathbf{E}[\,\hat{X}_p U_p\,] + B\,\mathbf{E}[\,U_p U_p^*] \\ &= 0.8\,\mathbf{E}[\,U_p U_p^*] - 0.8\,\mathbf{E}[\,U_p U_p^*] \\ &= 0.\end{aligned}$$

Since $\mathbf{E}[\,X_f \hat{X}_p^*\,] = \mathbf{E}[\,X_f U_p^*\,] = 0$, we have

$$\begin{aligned}\mathbf{E}[\,X_f Y_p^*\,] &= \mathbf{E}[\,X_f (Y_p^s + C\tilde{X}_p + C\hat{X}_p)^*\,] \\ &= \mathbf{E}[\,X_f (Y_p^s + C\tilde{X}_p)^*\,] \\ &= 3G + A\mathbf{E}[\,\tilde{X}_p \tilde{X}_p^*\,]C^* \\ &= 3(G + A\tilde{P}_0 C^*) \\ &= 3(-0.16 + (0.8)(0.5)(0.4)) = 0.\end{aligned}$$

With these, we conclude Equation (5.21). △

Remark 5.9. In Example 5.3, an ideal probabilistic setup is considered. To approximate this ideal probabilistic setup, one can imagine that a large number of experiments (M) are performed on the same system. Suppose for each experiment, say the i-th experiment, we have

$$x^i(0) = \sqrt{3}, \qquad u^i(0) = 0.5\sqrt{3}, \qquad u^i(1) = 0.4\sqrt{3}, \qquad u^i(2) = 0.5\sqrt{3}.$$

We then collect the state measurements $\{x^i(1), x^i(2), x^i(3)\}$ and the output measurements $\{y^i(0), y^i(1), y^i(2)\}$. Here, the superscript i indicates the data are from the i-th experiment. Now, construct $x(0)$ as

$$x(0) = \begin{bmatrix} x^1(0) & x^2(0) & \cdots & x^M(0) \end{bmatrix},$$

and similarly for $x(1)$, $x(2)$, $x(3)$, $u(0)$, $u(1)$, $u(2)$, $y(0)$, $y(1)$, $y(2)$. Suppose further that

$$\begin{aligned} w(\cdot) &= \begin{bmatrix} w^1(\cdot) & w^2(\cdot) & \cdots & w^M(\cdot) \end{bmatrix}, \\ v(\cdot) &= \begin{bmatrix} v^1(\cdot) & v^2(\cdot) & \cdots & v^M(\cdot) \end{bmatrix} \end{aligned}$$

are the collections of the corresponding zero-mean process and output noises satisfying

$$\lim_{M\to\infty} \frac{1}{M} \left[\begin{pmatrix} w(t) \\ v(t) \end{pmatrix} \begin{pmatrix} w(\tau) \\ v(\tau) \end{pmatrix}^* \right] = \begin{bmatrix} 0.36 & -0.48 \\ -0.48 & 1 \end{bmatrix} \delta_{t\tau}.$$

It should be clear that these collections of measurements will resemble an ideal probabilistic setup as M tends to infinity. △

5.10 Summary

In this chapter, we have shown that the rank condition, or the emptiness of the intersection of certain input and projected state sequences is the key point for constructing informative experiments for subspace algorithms. The requirements for achieving this rank condition are presented

- in Section 5.4 for the the deterministic case, where $\Sigma^w = \Sigma^v = 0$;
- in Section 5.7.1 for the output disturbance case, where $\Sigma^w = 0$;
- in Section 5.7.2 for the stochastic state excitation case, where $(\, A\, ,\, [\Sigma^w]^{\frac{1}{2}}\,)$ is reachable;
- in Section 5.7.3 for the general excitation and disturbance case, where $(\, A\, ,\, [\, B \quad (\Sigma^w)^{\frac{1}{2}}\,]\,)$ is reachable.

From this analysis, the criteria for informative experiments for two subspace algorithms given in Section 5.3 are derived in Section 5.5 for the deterministic case, and in Section 5.8 for the other three cases. Moreover, several examples are given to illustrate a few points discussed in the chapter. At this point, we draw attention to the fact that these criteria are not necessary conditions for informative experiments.

The requirements presented in this chapter may be considered mild in the sense that they are easily satisfied in typical identification practice. For instance, structural conditions, like the reachability of (A, B) or (A, G), hold in most cases, except in some extreme rank deficient cases. Another requirement is the persistent excitation of the input signal, which has to be slightly larger (by the degree of the A-annihilator) than the total number of Hankel block rows. Unless the input signal has certain repeating patterns, its excitation is normally sufficient. In the general excitation and disturbance case, the truncation index k is suggested to be larger than three times the degree of the A-annihilator. Such a requirement is mild since as indicated in [DPS95, PSD96], one should increase the truncation index k at a rate equal to some function involving $\log N$ in order to minimize the BIC criterion. Therefore, we conclude that it is reasonable to assume the rank n condition holds by default in a typical subspace identification operation, since otherwise not only are these mild requirements violated but also the input-state signals coincidentally create an exact cancelation effect.

6

Informative Experiments – Markov Parameters A Priori

6.1 Introduction

This chapter is concerned with the criteria for informative experiments for Algorithm 5.1. As pointed out in Chapter 5, Algorithm 5.1 is an identification algorithm operating under the conditions where $\Sigma^w = 0$, $\Sigma^v = 0$. This is to say that the setup considered in this chapter will be deterministic. One major difference between this chapter and the last chapter is that joint input-output conditions are investigated here. As a consequence stronger results, which are both necessary and sufficient, are acquired.

The study of the uniqueness of realizations (excluding the ones obtained by similarity transformations) determined from input-output observations in a deterministic setup was carried out extensively in the 1970's [Gop69, Bud71, ESG77, LS77]. Here, we will concentrate on the input-output signal conditions given in [LS77], under which a set of equivalent realizations (those having an identical transfer function) can be determined uniquely, as well as the related generalizations given in the recent work [vdH96]. First and foremost, we shall state the precise objective of the study in this chapter. Suppose that we are given a set of input-output signals $\mathscr{Z}^N$. Then consider the following two sets of realizations:

I. the set of all n-th order realizations which can produce $\mathscr{Z}^N$;

II. the set of all n-th order realizations whose first $\hbar$ Markov parameters $h_0, \ldots, h_{\hbar-1}$ equal

$$\{h_0, h_1, \ldots, h_{\hbar-1}\} = \{D, CB, \ldots, CA^{\hbar-2}B\},$$

for some fixed A, B, C, and D of compatible sizes.

In [LS77, vdH96], the conditions on $\mathscr{Z}^N$ under which Set I is given by a set of equivalent realizations are investigated. In this chapter, we impose another constraint confining the first $\hbar$ Markov parameters to certain values. (For $\hbar = 0$, we get back the original problem.) In effect, we are interested in the set of all realizations which belongs to both Sets I and II, as well as conditions on $\mathscr{Z}^N$ under which this set of realizations is given by a set of equivalent ones. Also since these conditions are necessary and sufficient, it is clear that any input-output signal failing to satisfy the conditions cannot be an informative experiment for any identification algorithm. On the other hand, it turns out that Algorithm 5.1 (and Algorithm 6.6 for $\hbar \neq 0$) uses no more information than the amount necessary to uniquely specify a set of equivalent realizations, if the number of block rows $l-1$ in the algorithm matches the observability index of the system to be identified. Finally, recall that $\mathscr{X}^N$ is a signal pair with elements in a Hilbert space, as defined in Chapter 2.

6.2 Necessary Conditions

To begin with, we first present a necessary condition for uniqueness. Note that the set of realizations $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$ obviously belongs to Set II.

Theorem 6.1. ([LS77]) *Suppose the input-output signal pair $\mathscr{X}^N$ admits an n-th order minimal realization (A, B, C, D) with observability index k. Then u being S.R. of order at least $k+1$ is necessary for*

$$\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$$

to be the only set belonging to Set I.

Now, consider the case that we are given the input-output signals as well as $\hbar$ Markov parameters $h_0, h_1, \ldots, h_{\hbar-1}$. The following theorem gives a necessary condition for uniqueness in such a case. Note that Theorem 6.1 is the special case when no Markov parameter is available (i.e., $\hbar = 0$).

Theorem 6.2. *Suppose the input-output signal pair $\mathscr{X}^N$ admits an n-th order minimal realization (A, B, C, D) with observability index k. Then u being S.R. of order at least $k - \hbar + 1$, where $0 \leq \hbar \leq k+1$, is necessary for*

$$\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$$

to be the only set belonging to both Sets I and II.

Proof. The proof is deferred to Appendix C, where it is shown that under the assumption that u is S.R. of order less than $k - \hbar + 1$, a set of realization can be constructed such that

1. there exist two realizations in this set which are not equivalent;
2. this set of realizations has the same $\hbar$ Markov parameters;
3. $\mathscr{X}^N$ admits the output equation of this set of realizations;
4. $\mathscr{X}^N$ admits the state equation of this set of realizations.

It is clear that if all elements in both Sets I and II are equivalent, then u is necessarily S.R. of order at least $k - \hbar + 1$. □

Lemma 6.3. *Suppose the input-output signal pair $\mathscr{Z}^N$ admits an n-th order minimal realization (A, B, C, D), with an observability index k. Then the space of the state sequence $\mathcal{X}_0^{k+1}$ being n-dimensional is necessary for*

$$\left\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\right\}$$

to be the only set belonging to both Sets I and II.

Proof. The lemma is proven by contradiction. Note the superscript $k+1$ will be dropped in the proof without causing confusion. Suppose $\mathcal{X}_0$ is of dimension less than n. Partition X_0 as

$$X_0^* =: \begin{bmatrix}\chi_1^* & \chi_2^* & \cdots & \chi_n^*\end{bmatrix}.$$

We may assume without loss of generality that $\chi_n = 0$ as we can always find a linear state transformation to obtain this. Define $\phi \in \mathbb{R}^n$ such that $\phi \neq 0$ and $\mathcal{O}_{k-1}\phi = 0$. Since $\mathcal{O}_k$ is injective by definition, it is evident that $CA^{k-1}\phi \neq 0$.

Now, let $\tilde{A} := \begin{bmatrix}0 & \cdots & 0 & \phi\end{bmatrix}$ and $\hat{A} := A + \tilde{A}$. Then, for $0 \le i \le j \le k-1$, we have

$$C\hat{A}^j = CA^{i-1}(A + \tilde{A})\hat{A}^{j-i} = CA^j. \tag{6.1}$$

Furthermore, define $\hat{\mathcal{O}}_{k+1}^* := \begin{bmatrix}C^* & (C\hat{A})^* & \cdots & (C\hat{A}^k)^*\end{bmatrix}$. Then, from Equation (6.1) it is easy to see that

$$\hat{\mathcal{O}}_{k+1}X_0 = \begin{bmatrix}\mathcal{O}_k \\ CA^k + CA^{k-1}\tilde{A}\end{bmatrix} X_0 = \mathcal{O}_{k+1}X_0.$$

The last equality stems from the fact that $\chi_n = 0$ and hence $\tilde{A}\,X_0 = 0$. Again, by Equation (6.1) we have $\hat{\mathcal{O}}_k = \mathcal{O}_k$, which implies that the first $k+1$ Markov parameters of $(\hat{A}, B, C, D)$ are unchanged:

$$\begin{aligned}\begin{bmatrix}D^* & (CB)^* & (C\hat{A}B)^* & \cdots & (C\hat{A}^{k-1}B)^*\end{bmatrix} &= \begin{bmatrix}D^* & (\mathcal{O}_k B)^*\end{bmatrix} \\ &= \begin{bmatrix}D^* & (CB)^* & (CAB)^* & \cdots & (CA^{k-1}B)^*\end{bmatrix}.\end{aligned}$$

Finally, using similar definitions as in Equations (C.9), it is not hard to see

$$Y_{0,k} = \hat{\mathcal{O}}_{k+1}X_0 + \hat{\mathcal{T}}_{k+1}U_{0,k}.$$

Following Steps 4 and 5 in the proof of Theorem 6.2, it can be verified that $\mathscr{Z}^N$ satisfies

$$\begin{aligned} x(t+1) &= \hat{A}x(t) + Bu(t), \\ y(t) &= Cx(t) + Du(t). \end{aligned}$$

That is, $(\hat{A}, B, C, D)$ which is not equivalent to (A, B, C, D) also belongs to Sets I and II. This is a contradiction, which completes the proof. □

6.3 Necessary and Sufficient Conditions

An elegant result, which was both necessary and sufficient, was first given in [LS77]. However, this result depends on the rank of a sequential selector matrix. (We refer to [Gop69] for the definition of the sequential selector matrix.) Recently, conditions directly based on the input-output signals were derived in [vdH96]. First, this necessary and sufficient result is presented.

Theorem 6.4. ([LS77, vdH96]) *Suppose the input-output signal pair $\mathscr{X}^N$ admits an n-th order minimal realization (A, B, C, D) with an observability index k. The following two cases are equivalent:*

1. $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$ *is the only set belonging to Set I;*
2. (a) *u is S.R. of order at least $k+1$, and*

   (b) $\Pi_{(\mathcal{U}^{k+1})^{\perp}} \mathcal{Y}^{k+1}$ *is an n-dimensional subspace.*

Remark 6.1. This theorem also says that (A, B, C, D) is the realization of minimal dimension since Conditions 2 above imply that the dimension of $\mathcal{X}_0^{k+1}$ must be n, which is also to say that any realization of order less than n cannot produce $\mathscr{X}^N$. $\triangle$

Similar to Section 6.2, we consider the case where the input-output signals and $\hbar$ Markov parameters $h_0, h_1, \ldots, h_{\hbar-1}$ are given. To identify systems with *a priori* knowledge of Markov parameters, we shall use a modified version of Algorithm 5.1, which is shown on page 78 as Algorithm 6.6. The following theorem gives a necessary and sufficient condition under such constraints. Note that Theorem 6.4 is the special case when no Markov parameter is available (i.e., $\hbar = 0$).

Theorem 6.5. *Suppose the input-output signal pair $\mathscr{X}^N$ admits an n-th order minimal realization (A, B, C, D) with an observability index k. Suppose further that $0 \leq \hbar \leq k+1$. Then the following two cases are equivalent:*

1. $\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$ *is the only set belonging to both Sets I and II;*
2. (a) *u is S.R. of order at least $k - \hbar + 1$, and*

   (b) $\Pi_{\left(\mathcal{U}_{0,k-\hbar}^{k+1}\right)^{\perp}} \mathcal{Z}^{k+1}$ *is an n-dimensional subspace*[1]*, where*

$$Z^{k+1} = \begin{cases} Y^{k+1} & \text{if } \hbar = 0, \\ Y^{k+1} - \Upsilon_{\hbar} U^{k+1} & \text{if } \hbar > 0, \end{cases} \quad \text{and} \quad \Upsilon_{\hbar} = \begin{bmatrix} h_0 & & & & 0 \\ \vdots & \ddots & & & \\ h_{\hbar-1} & & \ddots & & \\ & \ddots & & \ddots & \\ 0 & & h_{\hbar-1} & \cdots & h_0 \end{bmatrix}. \tag{6.2}$$

[1] $\mathcal{Z}^{k+1}$ is the space spanned by the rows of Z^{k+1}. That is, $\mathcal{Z}^{k+1} := \operatorname{span}\{\, \alpha^* Z^{k+1} \mid \alpha \in \mathbb{R}^{(k+1)p} \,\}$.

Proof. For (2) $\Rightarrow$ (1), we will follow Algorithm 6.6. we shall prove (1) $\Rightarrow$ (2) by constructing a set of realizations with the first $\hbar$ Markov parameters being the same, if one of Conditions 2 is not satisfied. Note that the case where u is S.R. of order less than $k - \hbar + 1$ is proven by Theorem 6.1. Therefore, we will show the remaining case where $\Pi_{\mathcal{U}_{0,k-\hbar}^{\perp}} \mathcal{Z}_{0,k}$ is less than n-dimensional. However, since the case where $\dim\{\mathcal{X}_0\} < n$ is shown in Lemma 6.3, it remains to show the case where $\dim\{\mathcal{X}_0\} = n$ but $\mathcal{X}_0 \cap \mathcal{U}_{0,k-\hbar} \neq \{0\}$. The rest of the proof is then devoted to construct a set of realization under this condition such that

a. this set of realizations has the same $\hbar$ Markov parameters;

b. there exist two realizations in this set which are not equivalent;

c. $\mathscr{X}^N$ admits this set of realizations.

The complete proof is deferred to Appendix C. $\square$

Remark 6.2. At this point, it can be seen from the proofs that the key technical point for Lemma 6.3, Theorems 6.2, and 6.5 is the construction of $\phi \neq 0$ such that $C\phi = CA\phi = \cdots = CA^{k-2}\phi = 0$ and $CA^{k-1}\phi \neq 0$. Although the proofs are long and tedious, they use nothing more than simple linear algebra. $\triangle$

Remark 6.3. Note that Conditions 2 in Theorem 6.4 are actually necessary and sufficient for Algorithm 5.1 to identify a system, provided that the number of block rows l in the algorithm is set to $k + 1$, where k is the observability index of the system. To see this, we can follow Theorem 5.6 for sufficiency. For necessity, we shall note that if Condition 2a is not satisfied, $\mathcal{O}_l \Pi_{U^{l\perp}} X_0$ will be rank deficient, and hence an incorrect A and C will be procured. Regarding Conditions 2 in Theorem 6.5, similar remarks also apply to Algorithm 6.6. $\triangle$

6.4 An Example

As Theorems 6.2 and 6.5 are proven in a constructive way, we can in fact use the proofs to find other realizations belonging to both Sets I and II. In the following example, input-output data and constraints of Markov parameters are set up deliberately in such a way that Conditions 2 in Theorem 6.5 are violated. Using this simple example, we will then verify the theorem by applying Equations (C.13)–(C.16) to find other valid realizations.

Example 6.1. Suppose we are given the input-output signal pair $\mathscr{X}^7$:

$$\begin{aligned} \{\, u(j)\,\}_{j=0}^{6} &= \{\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1\ \}, \\ \{\, y(j)\,\}_{j=0}^{6} &= \{\ 1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7\ \}; \end{aligned}$$

and the first 2 Markov parameters: $h_0 = 1$, $h_1 = 0$. (Note that a first order system cannot satisfy the above requirements.) Suppose further that we are given the system

$$(A,B,C,D) = \left(\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix}, 1 \right), \qquad \text{with} \quad x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

which produces $\mathscr{Z}^7$ and has Markov parameters $h_0 = 1$ and $h_1 = 0$. Then, we are asked to find out if this system and its equivalences are the unique set of realizations satisfying the requirements.

Using Theorem 6.5, it can be verified that Condition 2a is satisfied, since u is S.R. of order $k - \hbar + 1 = 1$ (note that $k = 2$ and $\hbar = 2$ in this case). Now, observe that

$$Z^3 = Y^3 - \Upsilon_2 U^3 = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 2 & 3 & 4 & 5 \\ 2 & 3 & 4 & 5 & 6 \end{bmatrix}.$$

Projecting Z^3 on $(\mathcal{U}_0^3)^\perp$ gives

$$\Pi_{(\mathcal{U}_0^3)^\perp} Z^3 = \begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \\ -2 & -1 & 0 & 1 & 2 \end{bmatrix},$$

whose row space is obviously only 1-dimensional, which violates Condition 2b consequently. Therefore, (A,B,C,D) and its equivalences are not the only realizations satisfying the requirements.

We can in fact find other realizations which satisfy the requirements using Equations (C.13)–(C.16). First of all, we have to determine the values of ψ, φ, and ϕ defined in the proof. To get these values, we first observe that the state sequence X_0^3 is given by

$$X_0^3 = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

With the above equation, we then have $\psi^* = \begin{bmatrix} \psi_0^* & \psi_1^* & \psi_2^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$; and $\varphi^* = 0$; and $\phi^* = \begin{bmatrix} 0 & \alpha \end{bmatrix}^*$, where $\alpha \in \mathbb{R}$. Note that ϕ also satisfies $\mathcal{O}_{k-1}\phi = C\phi = 0$ and $CA\phi = \alpha$. Following Equations (C.13)–(C.16), the following realizations are obtained:

$$(\hat{A},\hat{B},\hat{C},\hat{D}) = \left(\begin{bmatrix} 1 & 1 \\ 0 & -\alpha \end{bmatrix}, \begin{bmatrix} 0 \\ 1+\alpha \end{bmatrix}, \begin{bmatrix} 1 & 0 \end{bmatrix}, 1 \right), \qquad \text{with} \quad x(0) = \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Interested readers can check that these realizations indeed produce $\mathscr{Z}^7$ and have Markov parameters $h_0 = 1$ and $h_1 = 0$.

Finally, it is pointed out that if the input-output data are changed to

$$\begin{aligned} \{\, u(j) \,\}_{j=0}^{6} &= \{\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1\ \}, \\ \{\, y(j) \,\}_{j=0}^{6} &= \{\ 1,\ 1,\ 2,\ 3,\ 4,\ 5,\ 6\ \}, \end{aligned}$$

then the set of realizations

$$\{(TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R})\}$$

is the unique set of second order realizations producing the input-output data and having $h_0 = 1$ and $h_1 = 0$. This set of realizations can also be identified using Algorithm 6.6. $\triangle$

6.5 Summary

In this chapter, a necessary condition and a necessary and sufficient condition, for a set of equivalent realizations to be uniquely specified by an input-output signal pair $\mathscr{Z}^N$, are reviewed and presented in Theorems 6.1 and 6.4. Furthermore, these results are extended to the case where Markov parameters are given *a priori*. It is shown in Theorems 6.2 and 6.5 that the necessary S.R. requirement of u can be reduced when some initial Markov parameters are known and fixed. Finally, these ideas are illustrated by an example.

Algorithm 6.6. One-Block Subspace Algorithm with $\hbar$ MP *A Priori*

1. *Construct U^l and Z^l as in Equation* (6.2).
2. *Compute the SVD of $\Pi_{\mathcal{U}^{\perp}_{0,l-\hbar-1}} Z_{0,l-1}$ as (with obvious partitions)*

$$\Pi_{\mathcal{U}^{\perp}_{0,l-\hbar-1}} Z_{0,l-1} = \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}, \quad \textit{where} \quad \Gamma_1 = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_l \end{bmatrix}.$$

3. *Compute A and C from*

$$\begin{bmatrix} \gamma_1^* & \cdots & \gamma_{l-1}^* \end{bmatrix}^* A = \begin{bmatrix} \gamma_2^* & \cdots & \gamma_l^* \end{bmatrix}^* \quad \textit{and} \quad C = \gamma_1.$$

4. *Compute $S_0, \ldots, S_{l-\hbar-1}$ from*

$$\mathcal{O}_l^{\perp} Z_{0,l-1} := S_0 U_0 + \cdots + S_{l-\hbar-1} U_{l-\hbar-1}, \tag{6.3}$$

where

$$\mathcal{O}_l^{\perp} := \begin{bmatrix} P_0 & \cdots P_{l-2} & I \end{bmatrix} \quad \textit{and} \quad \mathcal{O}_l^{\perp} \mathcal{O}_l = 0.$$

5. *Determine $h_\hbar, \ldots, h_{l-1}$ from*

$$\begin{bmatrix} S_0 \\ \vdots \\ S_{l-\hbar-1} \end{bmatrix} - \begin{bmatrix} P_0 & \cdots & P_{\hbar-1} \\ \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} & \cdot^{\cdot^{\cdot}} \\ P_{l-\hbar-1} & \cdots & P_{l-2} \end{bmatrix} \begin{bmatrix} h_0 \\ \vdots \\ h_{\hbar-1} \end{bmatrix} = \begin{bmatrix} P_\hbar & \cdots & I \\ \vdots & \cdot^{\cdot^{\cdot}} & \\ I & & 0 \end{bmatrix} \begin{bmatrix} h_\hbar \\ \vdots \\ h_{l-1} \end{bmatrix}. \tag{6.4}$$

6. *Determine B and D from*

$$(\mathcal{O}_{l-1} B)^* = \begin{bmatrix} h_1^* & h_2^* & \cdots & h_{l-1}^* \end{bmatrix} \quad \textit{and} \quad D = h_0.$$

7. *Obtain the set of all state-space models*

$$\left\{ (TAT^{-1}, TB, CT^{-1}, D) \mid T \in GL_n(\mathbb{R}) \right\} \in SM_n^{p,m} / GL_n(\mathbb{R}).$$

Figure 6.1 *A schematic overview of the one-block subspace algorithm with $\hbar$ Markov parameters a priori.*

7

Subspace Identification via Markov Parameter Estimation

7.1 Introduction

Various versions of subspace identification methods have been derived in the literature. However, as remarked in [VODM96b, p. 123], it is not clear at present which version of subspace identification schemes will lead to the "best" algorithm. One version of subspace methods, which has been proposed in [Ver94], estimates the initial Markov parameters and uses those Markov parameters to determine the system matrices B and D. Nonetheless, this method is limited by the condition that the input signal is white noise.

To eliminate the effect of the input signals from the system equations, projections onto the orthogonal space of the input sequences are performed in most subspace methods. Effectively, the extracted state sequence will depend on the input signal used, which may in turn affect the accuracy in choosing the principal dynamics of the system. In this chapter, we propose and present a different approach to subspace system identification, aiming at avoiding the projection onto the orthogonal space of the input sequence. In effect, Markov parameters need to be estimated as a first step under a general input signal condition. The prominent difference of this approach is that a three-block configuration of data matrices is used. Roughly speaking, these three data blocks can be interpreted as follows. The first block is used primarily as an instrumental variable. The second block then provides the information from which the Kalman filter state sequence is observed (see Remarks 7.1 and 7.2). Finally, the last block exploits the relationship between the Kalman filter state sequences and the input sequences acquired from the second block in order to extract the initial Markov parameters of the system. Advantages and disadvantages of this algorithm will be investigated in more detail through some case studies in Chapter 9.

In Section 7.2 a simpler case, in which deterministic systems are considered, is covered. Section 7.3 and 7.4 then present the use of a three-block configuration and the procedure of estimating initial Markov parameters in a combined deterministic-stochastic setup. From Section 7.5 to 7.7, identification procedures for the deterministic and stochastic system matrices are presented. Finally, numerically efficient implementation is developed in Section 7.8.

7.2 Deterministic Identification

First, the deterministic identification via Markov parameter (MP) estimation is covered. In the deterministic setup, it will be seen that a two-block configuration, similar to those algorithms described in previous chapters, is adequate. To begin with, we first define the shorthands for the following block Hankel matrices in the usual way:

$$U_p := U_{0,k-1}^{2k}, \quad U_f := U_{k,2k-1}^{2k}, \quad Y_p := Y_{0,k-1}^{2k}, \quad Y_f := Y_{k,2k-1}^{2k}.$$

In addition, define $X_p := X_0^{2k}$ and $X_f := X_k^{2k}$. Again, the superscript $2k$ indicates that the complete Hankel matrix consists of $2k$ block rows. Furthermore, we shall assume k is greater than the observability index of the system. Now, it is not hard to obtain the following data equations:

$$Y_p = \mathcal{O}_k X_p + \mathcal{T}_k U_p; \tag{7.1}$$
$$Y_f = \mathcal{O}_k X_f + \mathcal{T}_k U_f. \tag{7.2}$$

In addition, the state equation linking the past and future data equations can be written as:

$$X_f = A^k X_p + \mathcal{C}_k^d U_p. \tag{7.3}$$

Lemma 7.1. *Suppose that the conditions in Theorem 5.7 are satisfied. Then,*

$$\begin{aligned} \mathcal{X}_f &\subset \mathcal{Y}_p + \mathcal{U}_p, \\ X_f + \mathcal{U}_f = \mathcal{X}_f \oplus \mathcal{U}_f &\subset (\mathcal{Y}_p + \mathcal{U}_p) \oplus \mathcal{U}_f. \end{aligned}$$

Proof. From Equation (7.1), it is easy to see that $\mathcal{Y}_p \subset \mathcal{X}_p + \mathcal{U}_p$. On the other hand, rewriting Equation (7.1) as $\mathcal{O}_k X_p = Y_p - \mathcal{T}_k U_p$ gives $\mathcal{X}_p \subset \mathcal{Y}_p + \mathcal{U}_p$, since $\mathcal{O}_k$ is injective. Now, adding $\mathcal{U}_p$ to both sides of these inclusions gives

$$\mathcal{Y}_p + \mathcal{U}_p \subset \mathcal{X}_p + \mathcal{U}_p, \quad \text{and} \quad \mathcal{X}_p + \mathcal{U}_p \subset \mathcal{Y}_p + \mathcal{U}_p.$$

Thus, clearly $\mathcal{Y}_p + \mathcal{U}_p = \mathcal{X}_p + \mathcal{U}_p$. In addition, with Equation 7.3 we then have

$$\mathcal{X}_f \;\subset\; \mathcal{X}_p + \mathcal{U}_p = \mathcal{Y}_p + \mathcal{U}_p.$$

Finally, Theorem 5.4 guarantees the direct sum property that $\mathcal{X}_f + \mathcal{U}_f = \mathcal{X}_f \oplus \mathcal{U}_f$. Thus, the lemma follows. □

Lemma 7.1 states that the space $\mathcal{X}_f$ can be observed from $\mathcal{Y}_p + \mathcal{U}_p$ and has empty intersection with $\mathcal{U}_f$. Consequently, Y_f in Equation (7.2) has a unique decomposition into $\mathcal{O}_k X_f$ and $\mathcal{T}_k U_f$ by the direct sum property. Consequently, $\mathcal{T}_k$ can be determined by removing U_f from $\mathcal{T}_k U_f$.

Assume $\mathcal{T}_k$ is determined. We introduce two Toeplitz matrices:

$$\Upsilon_U := \begin{bmatrix} 0 & h_{k-1} & \cdots & h_1 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & h_{k-1} \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \qquad \Upsilon_L := \begin{bmatrix} h_k & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ h_{2k-1} & \cdots & \cdots & h_k \end{bmatrix}, \tag{7.4}$$

such that $\mathcal{O}_k \mathcal{C}_k^d = \Upsilon_U + \Upsilon_L$. It can bee seen that Υ_U depends only on $h_1, \ldots, h_{k-1}$ whereas Υ_L depends only on $h_k, \ldots, h_{2k-1}$. Substituting Equation (7.3) into Equation (7.2) gives

$$Y_f - \mathcal{T}_k U_f - \Upsilon_U U_p = \mathcal{O}_k A^k X_p + \Upsilon_L U_p, \tag{7.5}$$

where the terms in the left hand side are all known. Note that $\mathcal{X}_p$ can be found from the equation $\mathcal{O}_k X_p = Y_p - \mathcal{T}_k U_p$. Furthermore, by Theorem 5.4 we have $\mathcal{X}_p + \mathcal{U}_p = \mathcal{X}_p \oplus \mathcal{U}_p$. As a result, we can uniquely decompose the left hand side of Equation (7.5) into $\mathcal{O}_k A^k X_p$ and $\Upsilon_L U_f$. In this way, Υ_L is determined.

Thus, the Markov parameters $h_0, \ldots, h_{2k-1}$, or equivalently, D, CB, $\ldots$, $CA^{2k-2}B$, are acquired. Finally, by Kalman's fundamental realization criterion [Kal71], a unique realization (A, B, C, D) can be guaranteed in the deterministic case.

Algorithm 7.2. *Deterministic Subspace Algorithm via MP Determination*

1. *Decompose* Y_f *into* $\mathcal{O}_k X_f$ *and* $\mathcal{T}_k U_f$ *using* $\mathcal{X}_f \oplus \mathcal{U}_f \subset (\mathcal{Y}_p + \mathcal{U}_p) \oplus \mathcal{U}_f$. *Determine* $\mathcal{T}_k$.
2. *Compute* $\mathcal{X}_p = span\{Y_p - \mathcal{T}_k U_p\}$ *and construct* Υ_U.
3. *Decompose* $Y_f - \mathcal{T}_k U_f - \Upsilon_U U_p$ *into* $\mathcal{O}_k A^k X_p$ *and* $\Upsilon_L U_p$ *using* $\mathcal{X}_p + U_p = X_p \oplus U_p$. *Then determine* Υ_L.
4. *Determine* (A, B, C, D) *from the* $2k$ *Markov parameters using Kung's algorithm.*

7.3 Deterministic-Stochastic Identification: A Three-Block Configuration

As shown in the previous section, a two-block configuration, past (U_p) and future (U_f), is adequate to identify Markov parameters in the deterministic case. However, when it comes to the combined deterministic-stochastic case, a two-block configuration cannot determine the Markov parameters of a system. This is due to the fact that Markov parameters cannot be estimated consistently in the statistical sense, just from two data blocks. In the remaining part of this chapter, we demonstrate the use of a three-block configuration to estimate initial Markov parameters in a noisy environment. Such a method will result in consistent estimates.

We shall define the three-block configuration. Let

$$\begin{bmatrix} U_p \\ \hline U_f \\ \hline U_r \end{bmatrix} := \begin{bmatrix} u(0) & u(1) & \cdots & u(N-3k) \\ \iddots & \iddots & \iddots & \iddots \\ u(k-1) & u(k) & \cdots & u(N-2k-1) \\ \hline u(k) & u(k+1) & \cdots & u(N-2k) \\ \iddots & \iddots & \iddots & \iddots \\ u(2k-1) & u(2k) & \cdots & u(N-k-1) \\ \hline u(2k) & u(2k+1) & \cdots & u(N-k) \\ \iddots & \iddots & \iddots & \iddots \\ u(3k-1) & u(3k) & \cdots & u(N-1) \end{bmatrix}. \tag{7.6}$$

The suffices p, f and r are mnemonic, representing 'past', 'future' and 'remote future', respectively. In addition, we define Y_p, Y_f, and Y_r similarly. Also, the state sequences are defined as:

$$X_p := X_0^{3k}, \quad X_f := X_k^{3k}, \quad X_r := X_{2k}^{3k}.$$

The superscript $3k$, which refers to the number of block rows, will be dropped hereafter without causing any confusion, as the symbols U, Y, and X will all be denoting $3k$-block Hankel matrices in the rest of the chapter. For convenience, we recall that $\mathcal{A}_k$ and $\mathcal{N}_k$ are defined as

$$\mathcal{A}_k = \begin{bmatrix} A^{k-1} & A^{k-2} & \cdots & I \end{bmatrix}, \qquad \mathcal{N}_k = \begin{bmatrix} 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & 0 \\ CA^{k-2} & \cdots & C & 0 \end{bmatrix}.$$

It can be seen without difficulty that the data equations can be written as:

$$Y_p = \mathcal{O}_k X_p + \mathcal{T}_k U_p + \mathcal{N}_k W_p + V_p, \tag{7.7}$$
$$Y_f = \mathcal{O}_k X_f + \mathcal{T}_k U_f + \mathcal{N}_k W_f + V_f, \tag{7.8}$$
$$Y_r = \mathcal{O}_k X_r + \mathcal{T}_k U_r + \mathcal{N}_k W_r + V_r. \tag{7.9}$$

In addition, the state equations linking the state sequences have the relationships:

$$X_f = A^k X_p + \mathcal{C}_k^d U_p + \mathcal{A}_k W_p, \tag{7.10}$$
$$X_r = A^k X_f + \mathcal{C}_k^d U_f + \mathcal{A}_k W_f. \tag{7.11}$$

Similar to Section 7.2, we divide $\mathcal{O}_k \mathcal{C}_k^d$ into two Toeplitz matrices, one in the upper triangular form Υ_U and one in the lower triangular form Υ_L, as in Equations (7.4). In short, we have $\mathcal{O}_k \mathcal{C}_k^d = \Upsilon_U + \Upsilon_L$.

Now, we define two new variables Z_f and Z_r as follows:

$$\begin{bmatrix} Z_f \\ Z_r \end{bmatrix} := \begin{bmatrix} Y_f \\ Y_r \end{bmatrix} - \begin{bmatrix} \mathcal{T}_k & 0 \\ \Upsilon_U & \mathcal{T}_k \end{bmatrix} \begin{bmatrix} U_f \\ U_r \end{bmatrix}. \tag{7.12}$$

Using Equations (7.8), (7.9) and (7.11), it is easy to see that Z_f and Z_r can also be written as:

$$Z_f = \mathcal{O}_k X_f + \mathcal{N}_k W_f + V_f, \tag{7.13}$$

$$Z_r = \mathcal{O}_k A^k X_f + \Upsilon_L U_f + \mathcal{O}_k \mathcal{A}_k W_f + \mathcal{N}_k W_r + V_r. \tag{7.14}$$

The purpose of introducing Z_r and Z_f is to eliminate any known information relating to the Markov parameters $h_0, \ldots, h_{k-1}$, so that the subsequent Markov parameters $h_k, \ldots, h_{2k-1}$ can be determined. As in the usual notation, the spaces spanned by the rows of Z_f and Z_r will be denoted as $\mathcal{Z}_f$ and $\mathcal{Z}_r$, respectively.

In this three-block configuration, the projection space will be

$$\mathcal{S} := \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r.$$

Again, k will be assumed to be greater than the observability index, which in turn implies $\mathcal{O}_{k-1}$ is of full column rank. It is also noteworthy that since the number of block rows is increased to $3k$, an increase in the degree of input excitation is required.

Theorem 7.3. *Suppose that the conditions in Theorem 5.15 are satisfied, with l being set to $2k$. Then,*

$$\begin{aligned} \Pi_{\mathcal{S}} \mathcal{X}_r \;&\subset\; \Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f, \\ \Pi_{\mathcal{S}} \mathcal{X}_r + \mathcal{U}_r = \Pi_{\mathcal{S}} \mathcal{X}_r \oplus \mathcal{U}_r \;&\subset\; (\Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f) \oplus \mathcal{U}_r, \end{aligned}$$

where $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$.

Proof. Due to the uncorrelated property of past, future and remote future noises, we have $\Pi_{\mathcal{S}} W_f = \Pi_{\mathcal{S}} V_f = \Pi_{\mathcal{S}} W_r = \Pi_{\mathcal{S}} V_r = 0$. Projecting Equation (7.8) onto $\mathcal{S}$ gives $\Pi_{\mathcal{S}} Y_f = \mathcal{O}_k \Pi_{\mathcal{S}} X_f + \mathcal{T}_k U_f$. Thus, we have

$$\begin{aligned} \Pi_{\mathcal{S}} \mathcal{Y}_f &\subset \Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}_f, \\ \Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f &\subset \Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}_f, \end{aligned} \tag{7.15}$$

where the last inclusion is obtained by adding $\mathcal{U}_f$ to both sides of the first inclusion. On the other hand, it is also easy to see that $\mathcal{O}_k \Pi_{\mathcal{S}} X_f = \Pi_{\mathcal{S}} Y_f - \mathcal{T}_k U_f$. Thus, we have

$$\begin{aligned} \Pi_{\mathcal{S}} \mathcal{X}_f &\subset \Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f, \\ \Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}_f &\subset \Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f. \end{aligned} \tag{7.16}$$

where the last inclusion is obtained by adding $\mathcal{U}_f$ to both sides of the first inclusion. In effect, Inclusions (7.15) and (7.16) together imply

$$\Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}_f = \Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f.$$

Using Theorem 5.14, it can be seen that $\Pi_{\mathcal{S}} \mathcal{X}_f + \mathcal{U}_f + \mathcal{U}_r = \Pi_{\mathcal{S}} \mathcal{X}_f \oplus \mathcal{U}_f \oplus \mathcal{U}_r$. Moreover, we have $\Pi_{\mathcal{S}} \mathcal{X}_r \subset \Pi_{\mathcal{S}} \mathcal{X}_f \oplus \mathcal{U}_f$, which can be seen from projecting Equation (7.11) onto $\mathcal{S}$. Thus,

$$\Pi_{\mathcal{S}} \mathcal{X}_r + \mathcal{U}_r = \Pi_{\mathcal{S}} \mathcal{X}_r \oplus \mathcal{U}_r \;\subset\; \Pi_{\mathcal{S}} \mathcal{X}_f \oplus \mathcal{U}_f \oplus \mathcal{U}_r = (\Pi_{\mathcal{S}} \mathcal{Y}_f + \mathcal{U}_f) \oplus \mathcal{U}_r,$$

which completes the proof. □

Remark 7.1. From Theorem 7.3, it can be seen that $\Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f$ provides a subspace which contains the space of the Kalman filter state sequence $\Pi_{\mathcal{S}}\mathcal{X}_r$. △

Theorem 7.4. *Suppose that the conditions in Theorem 5.15 are satisfied, with l being set to $2k$. Then,*

$$\begin{aligned} \Pi_{\mathcal{S}}\mathcal{X}_f &= \Pi_{\mathcal{S}}\mathcal{Z}_f, \\ \Pi_{\mathcal{S}}\mathcal{X}_f + \mathcal{U}_f = \Pi_{\mathcal{S}}\mathcal{X}_f \oplus \mathcal{U}_f &= \Pi_{\mathcal{S}}\mathcal{Z}_f \oplus \mathcal{U}_f, \end{aligned}$$

where $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$.

Proof. Due to the uncorrelated property of past, future and remote future noises, we have $\Pi_{\mathcal{S}}W_f = \Pi_{\mathcal{S}}V_f = \Pi_{\mathcal{S}}W_r = \Pi_{\mathcal{S}}V_r = 0$. Thus, projecting Equation (7.13) onto $\mathcal{S}$ gives

$$\Pi_{\mathcal{S}}Z_f = \mathcal{O}_k\Pi_{\mathcal{S}}X_f.$$

By Theorem 5.17, we have $\Pi_{\mathcal{S}}\mathcal{X}_f + \mathcal{U}_f + \mathcal{U}_r = \Pi_{\mathcal{S}}\mathcal{X}_f \oplus \mathcal{U}_f \oplus \mathcal{U}_r$. With $\mathcal{O}_k$ being injective, the statement follows. □

Remark 7.2. From Theorem 7.4, it can be seen that $\Pi_{\mathcal{S}}\mathcal{Z}_f$ can be used to observe the space of the Kalman filter state sequence $\Pi_{\mathcal{S}}\mathcal{X}_f$. △

7.4 Markov Parameter Estimation

From Theorems 7.3 and 7.4, a corollary can immediately be obtained. In this corollary, we eliminate by orthogonal projections the state components $\Pi_{\mathcal{S}}X_r$ in Equation (7.21) and $\Pi_{\mathcal{S}}X_f$ in Equation (7.22) in order to isolate the input components. It turns out that such an isolation can be done without reducing the richness of the input sequences, as shown in Inequalities (7.19), and (7.20).

Corollary 7.5. *Suppose that the conditions in Theorem 5.15 are satisfied, with l being set to $2k$. Denote $\mathcal{R} := \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f$ and $\mathcal{Q} := \Pi_{\mathcal{S}}\mathcal{Z}_f$. Then,*

$$\Pi_{\mathcal{R}^\perp}\,\Pi_{\mathcal{S}}Y_r = \mathcal{T}_k\,\Pi_{\mathcal{R}^\perp}U_r, \tag{7.17}$$

$$\Pi_{\mathcal{Q}^\perp}\,\Pi_{\mathcal{S}}Z_r = \Upsilon_L\,\Pi_{\mathcal{Q}^\perp}U_f, \tag{7.18}$$

where $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$. Furthermore,

$$\mathbf{E}\left[\,(\Pi_{\mathcal{R}^\perp}U_r)\,\left(\Pi_{\mathcal{R}^\perp}U_r\right)^*\,\right] > 0, \tag{7.19}$$

$$\mathbf{E}\left[\,(\Pi_{\mathcal{Q}^\perp}U_f)\,\left(\Pi_{\mathcal{Q}^\perp}U_f\right)^*\,\right] > 0. \tag{7.20}$$

Proof. Projecting Equations (7.9) and (7.14) onto $\mathcal{S}$ gives

$$\Pi_{\mathcal{S}} Y_r = \mathcal{O}_k \Pi_{\mathcal{S}} X_r + \mathcal{T}_k U_r, \tag{7.21}$$
$$\Pi_{\mathcal{S}} Z_r = \mathcal{O}_k A^k \Pi_{\mathcal{S}} X_f + \Upsilon_L U_f. \tag{7.22}$$

From Theorems 7.3 and 7.4, we have $\Pi_{\mathcal{S}}\mathcal{X}_r \subset \mathcal{R}$ and $\Pi_{\mathcal{S}}\mathcal{X}_f \subset \mathcal{Q}$. Thus, we can isolate $\mathcal{T}_k$ and Υ_L by the corresponding orthogonal projection. Furthermore, the direct sum property stated in Theorems 7.3 and 7.4 guarantees Inequalities (7.19) and (7.20). □

It can been seen that the matrices $\mathcal{T}_k$ and Υ_L can directly be determined from Equations (7.17) and (7.18) in Corollary 7.5. However, recall that $\mathcal{T}_k$ is a lower triangular Toeplitz matrix composed of $h_0, \ldots, h_{k-1}$, and Υ_L a lower triangular Toeplitz matrix composed of $h_k, \ldots, h_{2k-1}$. In practice such a $\mathcal{T}_k$ and a Υ_L may not be a perfect lower triangular Toeplitz matrix. As a result, a direct extraction of the Markov parameters may not be possible since several values of a single Markov parameter may be obtained in the corresponding entries in the Toeplitz matrix. Owing to this consideration, another method of extracting the Markov parameters is proposed.

First, consider Equation (7.17). Partition the following matrices:

$$\Pi_{(\Pi_{\mathcal{S}}\mathcal{Y}_f+\mathcal{U}_f)^\perp}\Pi_{\mathcal{S}} Y_r =: \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_k \end{bmatrix} \quad \text{and} \quad \Pi_{(\Pi_{\mathcal{S}}\mathcal{Y}_f+\mathcal{U}_f)^\perp} U_r =: \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_k \end{bmatrix},$$

with obvious dimensions. From Inequality (7.19), we have $\Pi_{(\Pi_{\mathcal{S}}\mathcal{Y}_f+\mathcal{U}_f)^\perp} U_r$ of full row rank. Define S and P as:

$$S := \begin{bmatrix} S_1 & S_2 & \cdots & S_k \end{bmatrix} \quad \text{and} \quad P := \begin{bmatrix} P_1 & P_2 & \cdots & P_k \\ & P_1 & \cdots & P_{k-1} \\ & & \ddots & \vdots \\ 0 & & & P_1 \end{bmatrix},$$

where P is of full row rank as well. Furthermore, it can be seen without difficulty that

$$S = HP,$$

where $H := \begin{bmatrix} h_0 & h_1 & \cdots & h_{k-1} \end{bmatrix}$. Thus, the least-squares solutions for the first k Markov parameters equal

$$H = SP^*(PP^*)^{-1}.$$

To solve the next k Markov parameters $h_k, \ldots, h_{2k-1}$, we can simply repeat the procedure with Equation (7.18).

7.5 Kalman Filter State Estimation

In this section, a new method is derived to estimate Kalman filter state sequences $\Pi_{\mathcal{S}} X_k$ and $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$. This method first determines the initial Markov parameters, then isolates the terms $\mathcal{O}_{2k-1}\Pi_{\mathcal{S}} X_k$ and $\mathcal{O}_{2k-1}\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$ from the data equations, and finally extracts $\mathcal{O}_{2k-1}$ from a joint matrix as follows:

$$\begin{bmatrix}\mathcal{O}_{2k-1}\Pi_{\mathcal{S}} X_k & \mathcal{O}_{2k-1}\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}\end{bmatrix} = \mathcal{O}_{2k-1}\begin{bmatrix}\Pi_{\mathcal{S}} X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}\end{bmatrix}.$$

One special feature of this new method is that two data blocks (Y_f and Y_r) are used for the estimation of the observability matrix and the Kalman filter states, while only one data block (Y_p) is used as an instrumental variable.

As shown in previous sections, initial Markov parameters can be determined from a three-block configuration. First, we consider a simpler case. Suppose the first $2k$ Markov parameters are determined. Then, it is easy to see that

$$\begin{bmatrix}Y_f\\Y_r\end{bmatrix} - \mathcal{T}_{2k}\begin{bmatrix}U_f\\U_r\end{bmatrix} = \mathcal{O}_{2k}X_f + \mathcal{N}_{2k}\begin{bmatrix}W_f\\W_r\end{bmatrix} + \begin{bmatrix}V_f\\V_r\end{bmatrix}.$$

Projecting the above equation onto $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$ gives

$$\Pi_{\mathcal{S}}\begin{bmatrix}Y_f\\Y_r\end{bmatrix} - \mathcal{T}_{2k}\begin{bmatrix}U_f\\U_r\end{bmatrix} = \mathcal{O}_{2k}\Pi_{\mathcal{S}} X_f. \tag{7.23}$$

The method of joint decomposition is similar to the above-mentioned. The main idea is to isolate the terms $\mathcal{O}_{2k-1}\Pi_{\mathcal{S}} X_k$ and $\mathcal{O}_{2k-1}\Pi_{\mathcal{S}} X_{k+1}$ from the data equations by removing the terms of the input sequences. To this end, we introduce two more matrices U_U and U_L, each of which with one block row (U_k or U_{3k-1}) deleted, in the following way:

$$U_U = \begin{bmatrix}U_k\\ \vdots \\ U_{3k-2}\end{bmatrix} \quad \text{and} \quad U_L = \begin{bmatrix}U_{k+1}\\ \vdots \\ U_{3k-1}\end{bmatrix}. \tag{7.24}$$

Illustratively, these matrices can also be envisaged as

$$\begin{bmatrix}U_f\\U_r\end{bmatrix} = \begin{bmatrix}U_k\\ \vdots \\ U_{3k-1}\end{bmatrix} = \begin{bmatrix}U_U\\ \hdashline U_{3k-1}\end{bmatrix} = \begin{bmatrix}U_k\\ \hdashline U_L\end{bmatrix}.$$

Moreover, we define similarly for Y_U and Y_L. Thus, as in Equation (7.23) we have

$$\Pi_{\mathcal{S}} Y_U - \mathcal{T}_{2k-1}U_U = \mathcal{O}_{2k-1}\Pi_{\mathcal{S}} X_k, \tag{7.25}$$

$$\Pi_{\mathcal{S}+\mathcal{Y}_k} Y_L - \mathcal{T}_{2k-1}U_L = \mathcal{O}_{2k-1}\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}. \tag{7.26}$$

Finally, we separate the observability matrix and the Kalman filter states using SVD, namely

$$\begin{bmatrix}\Pi_{\mathcal{S}} Y_U & \Pi_{\mathcal{S}+\mathcal{Y}_k} Y_L\end{bmatrix} - \mathcal{T}_{2k-1}\begin{bmatrix}U_U & U_L\end{bmatrix} =: \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}. \tag{7.27}$$

Then, in some particular state coordinates we have

$$\mathcal{O}_{2k-1} = \Gamma_1 \Sigma_1^{\frac{1}{2}}, \quad \text{and} \quad \begin{bmatrix}\Pi_{\mathcal{S}} X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}\end{bmatrix} = \Sigma_1^{\frac{1}{2}}\Omega_1^*.$$

Remark 7.3. In Equation (7.25), the projection space $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$ is used. Recalling from Chapter 4, $\Pi_{\mathcal{S}} X_k$ is the Kalman filter state sequence with the initial condition $\Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p)+\mathcal{U}_p+\mathcal{U}_f} X_0$. Moreover, from Lemma 4.8 we have

$$(\mathcal{Y}_p^d \cap \mathcal{Y}_p) + \mathcal{U}_p + \mathcal{U}_f = \hat{\mathcal{X}}_0^d + \mathcal{U}_p + \mathcal{U}_f,$$

where $\hat{\mathcal{X}}_0^d$ represents the purely deterministic modes. In other words, the initial condition of $\Pi_{\mathcal{S}} X_k$ is $\Pi_{\hat{\mathcal{X}}_0^d+\mathcal{U}_p+\mathcal{U}_f} X_0$.

On the contrary, an extended projection space $\mathcal{S}+\mathcal{Y}_k$ is employed in Equation (7.26). This additional term $\mathcal{Y}_k$ provides the additional output information required for estimating $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$ from $\Pi_{\mathcal{S}} X_k$. Analogously, we observe that

$$\big((\mathcal{Y}_p^d + \mathcal{Y}_k^d) \cap (\mathcal{Y}_p + \mathcal{Y}_k)\big) + \mathcal{U}_p + \mathcal{U}_f = \hat{\mathcal{X}}_0^d + \mathcal{U}_p + \mathcal{U}_f.$$

Thus, it turns out that the initial condition of $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$ is also $\Pi_{\hat{\mathcal{X}}_0^d+\mathcal{U}_p+\mathcal{U}_f} X_0$. △

7.6 Modeling of the Deterministic Part

In this section, three different approaches are presented for procuring system matrices. One method is to extract the system matrices directly from the initial Markov parameters in a similar fashion as in Kung's algorithm. The second method determines system matrices using the shift invariance approach. Finally, the third method determines system matrices using the state sequence approach.

7.6.1 A Markov Parameter Approach

In this chapter, subspace identification via Markov parameter estimation is advocated. Intuitively, system matrices can readily be determined from the identified Markov parameters. In fact, it is well known that a system with known order can be identified from a set of initial Markov parameters if the set is informative enough. One effective way of finding system matrices from Markov parameters is to use Kung's algorithm. The corresponding procedure will be as follows. We first determine the initial $2k$ Markov parameters as in Section 7.4. Then, construct the block Hankel matrix H as:

$$H := \begin{bmatrix} h_1 & h_2 & \cdots & h_k \\ \iddots & \iddots & \iddots & \iddots \\ h_k & h_{k+1} & \cdots & h_{2k-1} \end{bmatrix}.$$

Decompose the block Hankel matrix using SVD, and follow Kung's algorithm as discussed in Chapter 3.

One advantage of this direct identification is that a balanced (and hence asymptotically stable) approximation can be procured in an overwhelmingly simple fashion for rapidly decaying systems. By the meaning of rapidly decaying systems, we assume all eigenvalues of the systems are very close to zero, or their Markov parameters tend to zero quickly, such that the first $2k$ Markov parameters contain all predominant dynamics of such systems. To obtain a balanced approximation, we construct the block Hankel matrix

$$H := \begin{bmatrix} h_1 & \cdots & h_{2k-1} \\ \vdots & \iddots & 0 \\ h_{2k-1} & 0 & 0 \end{bmatrix}$$

instead and then apply Kung's algorithm. Such a technique has also been shown in Theorem 3.9.

However, to determine the stochastic part of the system, Kalman filter state sequences will be needed. In the light of this, a joint decomposition can be used to determine A, B, and C, as well as the Kalman filter state sequences $\Pi_{\mathcal{S}} X_k$ and $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$. Thus, instead of calculating the SVD of Equation (7.27), we compute the SVD of the following:

$$\left[\begin{array}{c:c:c} h_1 & & \\ \vdots & \Pi_{\mathcal{S}} Y_U - \mathcal{T}_{2k-1} U_U & \Pi_{\mathcal{S}+\mathcal{Y}_k} Y_L - \mathcal{T}_{2k-1} U_L \\ h_{2k-1} & & \end{array}\right] =: \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

It can be seen from Equations (7.25) and (7.26) that the above can also be written as:

$$\mathcal{O}_{2k-1} \left[\begin{array}{c:c:c} B & \Pi_{\mathcal{S}} X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1} \end{array}\right] = \Gamma_1 \Sigma_1 \Omega_1^*.$$

In some particular state coordinates, we have

$$\mathcal{O}_{2k-1} = \Gamma_1 \Sigma_1^{\frac{1}{2}}, \qquad \left[\begin{array}{c:c:c} B & \Pi_{\mathcal{S}} X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1} \end{array}\right] = \Sigma_1^{\frac{1}{2}} \Omega_1.$$

Clearly, B, $\Pi_{\mathcal{S}} X_k$, and $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$ are acquired in the above equation, whereas D is simply equal to h_0. A and C can be obtained from $\mathcal{O}_k$ with the use of the shift invariance property as shown in Equation (7.28) discussed in the next subsection.

7.6.2 A Shift Invariance Approach

The identification method using the shift invariance property is more complicated and requires a greater computational effort compared with the other two methods mentioned in this section. However, these shortcomings are usually considered outweighed by its ability to provide a better fitting with the input-output signals. Consequently, it is more commonly used in subspace identification.

In this method, we first determine $\mathcal{O}_{2k-1}$ and $\Pi_{\mathcal{S}} X_f$ as in Section 7.5. In practice, there may not exist matrices A and C which fit $\mathcal{O}_{2k-1}$ perfectly, due to imperfect noise cancellation. A commonly used remedy is to find the best A and C in the least-squares sense with the use of the shift invariance property:

$$C = \gamma_1 \quad \text{and} \quad A = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{2k-2} \end{bmatrix}^{\dagger} \begin{bmatrix} \gamma_2 \\ \vdots \\ \gamma_{2k-1} \end{bmatrix}, \quad \text{where} \quad \mathcal{O}_{2k-1} = \begin{bmatrix} \gamma_1 \\ \vdots \\ \gamma_{2k-1} \end{bmatrix}. \tag{7.28}$$

Reconstruct $\mathcal{O}_{2k}$ from A and C. Now, from Equation (7.23) it is not hard to see

$$\Pi_{\mathcal{S}} \begin{bmatrix} Y_f \\ Y_r \end{bmatrix} - \mathcal{O}_{2k} \Pi_{\mathcal{S}} X_f = \mathcal{T}_{2k} \begin{bmatrix} U_f \\ U_r \end{bmatrix}.$$

Again, in practice there may not exist matrices B and D which fit the above equation perfectly, due to imperfect noise cancellation. To solve B and D from the above, we first note that

$$\begin{aligned}
\mathcal{T}_{2k} \begin{bmatrix} U_f \\ U_r \end{bmatrix} = {} & \left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & \\ 0 & \mathcal{O}_{2k-1} \end{array}\right] \begin{bmatrix} D \\ B \end{bmatrix} U_k + \left[\begin{array}{c:c} 0 & 0 \\ \hdashline I & 0 \\ 0 & \mathcal{O}_{2k-2} \end{array}\right] \begin{bmatrix} D \\ B \end{bmatrix} U_{k+1} + \cdots \\
& + \left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & 0 \\ I & 0 \end{array}\right] \begin{bmatrix} D \\ B \end{bmatrix} U_{3k-1}.
\end{aligned}$$

Furthermore, note that B and D appear linearly in the equation. Using the Kronecker product (see for instance [HJ94]), we have

$$\begin{aligned}
\operatorname{vec}\left(\mathcal{T}_{2k} \begin{bmatrix} U_f \\ U_r \end{bmatrix}\right) = {} & \left(\left[\begin{array}{c:c} I & 0 \\ \hdashline 0 & \\ 0 & \mathcal{O}_{2k-1} \end{array}\right] \otimes U_k\right) \operatorname{vec}\left(\begin{bmatrix} D \\ B \end{bmatrix}\right) \\
& + \left(\left[\begin{array}{c:c} 0 & 0 \\ \hdashline I & 0 \\ 0 & \mathcal{O}_{2k-2} \end{array}\right] \otimes U_{k+1}\right) \operatorname{vec}\left(\begin{bmatrix} D \\ B \end{bmatrix}\right) + \cdots \\
& + \left(\left[\begin{array}{c:c} 0 & 0 \\ \hdashline 0 & 0 \\ I & 0 \end{array}\right] \otimes U_{3k-1}\right) \operatorname{vec}\left(\begin{bmatrix} D \\ B \end{bmatrix}\right).
\end{aligned} \tag{7.29}$$

Here, $\otimes$ denotes the Kronecker product and *vec* denotes the vector operation, i.e. stacking the columns of the matrix on top of each other in a vector. Thus, from the above equation the least-squares solutions for B and D can easily be acquired by taking the appropriate pseudo-inverse. Finally, note that discussion on stability enforcement can be found in Chapter 8.

7.6.3 A State Sequence Approach

In the shift invariance approach, we have already seen that the matrices A and C are first determined, and then the matrices B and D. However, in determining B and D, Kronecker products are employed to convert the data equation into a simpler matrix-times-vector form. For this reason, numerical complexity can increase dramatically if a huge amount of data is used. In [VODM94], an alternative method, which is known as the state sequence approach, is proposed. Such a method has also been used for the deterministic identification in [MDMVV89]. In a combined deterministic-stochastic setup, this method reduces the complexity of the numerical calculation, at the expense of introducing a bias in the estimation. Loosely speaking, this bias can be regarded as resulting from the fact that initial Markov parameters are not available. Consequently, inconsistent Kalman filter state sequences, in the sense that their initial conditions are not consistent, are obtained for different time instants. Further detail regarding this issue is omitted here but is covered in more depth in Chapter 3, or can be found in [VODM94].

With the three-block configuration, unbiased estimates can be obtained using the state sequence approach. Unlike the two-block subspace algorithm, the three-block subspace algorithm first estimates the initial Markov parameters. Using these Markov parameters, Kalman filter state sequences with the same initial condition can be acquired, which in turn produce unbiased system matrices.

As seen in Remark 7.3, $\Pi_{\mathcal{S}} X_k$ and $\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$ are the Kalman filter state sequences with the same initial condition. Denote $\hat{X}_k := \Pi_{\mathcal{S}} X_k$ and $\hat{X}_{k+1} := \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1}$. Thus, we have

$$\hat{X}_{k+1} = A\hat{X}_k + BU_k + K_k(Y_k - C\hat{X}_k - DU_k), \tag{7.30a}$$

$$Y_k = C\hat{X}_k + DU_k + (Y_k - C\hat{X}_k - DU_k). \tag{7.30b}$$

The above equations can also be written as:

$$\begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ U_k \end{bmatrix} + \begin{bmatrix} K_k \\ I \end{bmatrix} (Y_k - C\hat{X}_k - DU_k).$$

In [VODM94], it is pointed out that $Y_k - C\hat{X}_k - DU_k$ is orthogonal to $\hat{X}_k$ and U_k. Such orthogonality is the result of some basic projection properties. More precisely, we have

$$\mathbf{E}\left[\, (Y_k - C\hat{X}_k - DU_k)(\hat{X}_k)^* \,\right] = \mathbf{E}\left[\, (Y_k - C\hat{X}_k - DU_k)(U_k)^* \,\right] = 0.$$

Thus, it can in fact be seen that unbiased estimates can be obtained by estimating the matrices A, B, C, and D in the least-squares sense:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \mathbf{E}\left[\begin{pmatrix} \hat{X}_{k+1} \\ Y_k \end{pmatrix} \begin{pmatrix} \hat{X}_k \\ U_k \end{pmatrix}^* \right] \mathbf{E}\left[\begin{pmatrix} \hat{X}_k \\ U_k \end{pmatrix} \begin{pmatrix} \hat{X}_k \\ U_k \end{pmatrix}^* \right]^{-1}.$$

Note that the conditions for the above covariance matrix to be invertible is covered in Chapter 5. In practice when real data are concerned, this unbiasedness requires an infinite amount of data, i.e., an asymptotic unbiasedness with $N \to \infty$. Here, we shall defer the discussion on stability enforcement to Chapter 8.

7.7 Modeling of the Stochastic Part

In this section, we will follow the method proposed in [VODM93, VODM94] to model the steady state stochastic subsystem. With A, B, C, D, $\hat{X}_k$ and $\hat{X}_{k+1}$ determined, this method uses Equations (7.30) to estimate the steady state errors: Σ^w, Σ^{wv} and Σ^v. However, due to the fact that Equations (7.30) are not in steady state, approximation error results. Nonetheless, as pointed out in [VODM96b] this approximation error will grow smaller, as the number of blocks k grows larger.

The stochastic identification procedure is summarized as follows. First determine A, B, C, D, $\hat{X}_k$ and $\hat{X}_{k+1}$ as in the previous sections. Now, let

$$\begin{bmatrix} \varepsilon_w \\ \varepsilon_v \end{bmatrix} = \begin{bmatrix} \hat{X}_{k+1} \\ Y_k \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{X}_k \\ U_k \end{bmatrix}.$$

Then, obtain the approximation:

$$\begin{bmatrix} \Sigma^w & \Sigma^{wv} \\ \Sigma^{vw} & \Sigma^v \end{bmatrix} \approx \mathbf{E}\left[\begin{pmatrix} \varepsilon_w \\ \varepsilon_v \end{pmatrix} \begin{pmatrix} \varepsilon_w \\ \varepsilon_v \end{pmatrix}^* \right]. \tag{7.31}$$

Solve Σ^s, G, and Λ_0 from:

$$\Sigma^s = A\Sigma^s A^* + \Sigma^w; \tag{7.32a}$$
$$G = A\Sigma^s C^* + \Sigma^{wv}; \tag{7.32b}$$
$$\Lambda_0 = C\Sigma^s C^* + \Sigma^v. \tag{7.32c}$$

Finally, the Kalman filter gain K and the associated covariance matrix P can be solved from:

$$K = (G + APC^*)(\Lambda_0 + CPC^*)^{-1}, \tag{7.33a}$$
$$P = APA^* - (G + APC^*)(\Lambda_0 + CPC^*)^{-1}(G + APC^*)^*. \tag{7.33b}$$

A final remark will be drawn about the positive realness of the resulting stochastic subsystem. From (7.31), it can be seen that covariance of the process and output noises is nonnegative definite. Thus, as shown in [Fau76] positive realness of the resulting stochastic subsystem is guaranteed.

7.8 Numerically Efficient Implementation by RQ Factorization

In practice, RQ factorization has been used extensively in subspace based algorithms. For instance, both the MOESP and N4SID algorithms use RQ factorization to speed up the numerical computation. To implement an efficient algorithm for the identification scheme developed in this chapter, the following RQ factorization is computed:

$$\begin{bmatrix} U_f \\ U_r \\ U_p \\ Y_p \\ Y_f \\ Y_r \end{bmatrix} =: \begin{bmatrix} R_{11} & & & & & 0 \\ R_{21} & R_{22} & & & & \\ R_{31} & R_{32} & R_{33} & & & \\ R_{41} & R_{42} & R_{43} & R_{44} & & \\ R_{51} & R_{52} & R_{53} & R_{54} & R_{55} & \\ R_{61} & R_{62} & R_{63} & R_{64} & R_{65} & R_{66} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \\ Q_5 \\ Q_6 \end{bmatrix} =: RQ. \tag{7.34}$$

Note that the above arrangement of the data blocks shares a similar structure with that used in the MOESP PO scheme [Ver94], both having the past data blocks placed in the center.

The major advantage of this factorization is that sequences projected onto $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f + \mathcal{U}_r$ can easily be acquired from above by eliminating the Q_5 and Q_6 components. That is,

$$\Pi_{\mathcal{S}} \begin{bmatrix} Y_f \\ Y_r \end{bmatrix} = \begin{bmatrix} R_{51} & R_{52} & R_{53} & R_{54} & 0 & 0 \\ R_{61} & R_{62} & R_{63} & R_{64} & 0 & 0 \end{bmatrix} Q.$$

To see how this factorization facilitates the computation further, we have the following discussion concerning the computation of Markov parameters and Kalman filter state sequences. Any other part of the implementation will be more or less straightforward.

7.8.1 Determination of the First k Markov Parameters

From the RQ factorization in Equation (7.34), a smaller RQ factorization is computed:

$$\begin{bmatrix} R_{52} & R_{53} & R_{54} \\ R_{22} & 0 & 0 \\ R_{62} & R_{63} & R_{64} \end{bmatrix} =: \begin{bmatrix} \hat{R}_{52} & 0 & 0 \\ \hat{R}_{22} & \hat{R}_{23} & 0 \\ \hat{R}_{62} & \hat{R}_{63} & \hat{R}_{64} \end{bmatrix} \begin{bmatrix} \hat{Q}_2 \\ \hat{Q}_3 \\ \hat{Q}_4 \end{bmatrix}. \tag{7.35}$$

It will been seen in Proposition 7.6 that $\mathcal{T}_k$ can in fact be determined from this factorization. Also, with this smaller factorization we immediately obtain the following factorization:

$$\Pi_{\mathcal{S}} \begin{bmatrix} U_f \\ Y_f \\ U_r \\ Y_r \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{51} & R_{52} & R_{53} & R_{54} \\ R_{21} & R_{22} & 0 & 0 \\ R_{61} & R_{62} & R_{63} & R_{64} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{51} & \hat{R}_{52} & 0 & 0 \\ R_{21} & \hat{R}_{22} & \hat{R}_{23} & 0 \\ R_{61} & \hat{R}_{62} & \hat{R}_{63} & \hat{R}_{64} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \hat{Q}_2 \\ 0 & \hat{Q}_3 \\ 0 & \hat{Q}_4 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix} =: \hat{R}\hat{Q}. \tag{7.36}$$

Proposition 7.6. *Suppose that the conditions in Theorem 5.15 are satisfied, with l being set to $2k$. Then, for $\hat{R}_{23}$ and $\hat{R}_{63}$ defined in Equation* (7.35)*, we have*

$$\hat{R}_{63} = \mathcal{T}_k \hat{R}_{23}.$$

Proof. By Theorem 7.3 we have $\Pi_{\mathcal{S}}\mathcal{Y}_r \subset \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f + \mathcal{U}_r$. This implies $\hat{R}_{64} = 0$. Now, let $\mathcal{R} := \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f$. From Equation (7.36), it is easy to see that

$$\begin{aligned} \Pi_{\mathcal{R}^\perp} U_r &= \begin{bmatrix} 0 & 0 & \hat{R}_{23} & 0 \end{bmatrix} \hat{Q}, \\ \Pi_{\mathcal{R}^\perp} \Pi_{\mathcal{S}} Y_r &= \begin{bmatrix} 0 & 0 & \hat{R}_{63} & 0 \end{bmatrix} \hat{Q}. \end{aligned}$$

Thus, by Corollary 7.5 the result follows. □

Using the RQ factorization in (7.35) and Proposition 7.6, we can determine $h_0, \ldots, h_{k-1}$ by following the procedure described in Section 7.4.

Remark 7.4. It is noteworthy that the RQ factorization in Equation (7.35) requires much less numerical calculation than that in Equation (7.34). △

7.8.2 Determination of the Next k Markov Parameters

With Markov parameters $h_0, \ldots, h_{k-1}$, we can construct Υ_U as in Equation (7.4). The next step is to determine $\Pi_{\mathcal{S}} Z_f$ and $\Pi_{\mathcal{S}} Z_r$ as in Equation (7.12). From Equation (7.36), we have

$$\begin{aligned} \Pi_{\mathcal{S}} Z_f &= \Pi_{\mathcal{S}} Y_f - \mathcal{T}_k U_f \\ &= \begin{bmatrix} R_{51} & \hat{R}_{52} & 0 & 0 \end{bmatrix} \hat{Q} - \mathcal{T}_k \begin{bmatrix} R_{11} & 0 & 0 & 0 \end{bmatrix} \hat{Q} \\ &=: \begin{bmatrix} \bar{R}_{51} & \hat{R}_{52} & 0 & 0 \end{bmatrix} \hat{Q}; \end{aligned}$$

and

$$\begin{aligned} \Pi_{\mathcal{S}} Z_r &= \Pi_{\mathcal{S}} Y_r - \begin{bmatrix} \Upsilon_U & \mathcal{T}_k \end{bmatrix} \begin{bmatrix} U_f \\ U_r \end{bmatrix} \\ &= \begin{bmatrix} R_{61} & \hat{R}_{62} & \hat{R}_{63} & \hat{R}_{64} \end{bmatrix} \hat{Q} - \begin{bmatrix} \Upsilon_U & \mathcal{T}_k \end{bmatrix} \begin{bmatrix} R_{11} & 0 & 0 & 0 \\ R_{21} & \hat{R}_{22} & \hat{R}_{23} & 0 \end{bmatrix} \hat{Q} \\ &=: \begin{bmatrix} \bar{R}_{61} & \bar{R}_{62} & \bar{R}_{63} & \hat{R}_{64} \end{bmatrix} \hat{Q}. \end{aligned}$$

To obtain the next k Markov parameters, another small RQ factorization, similar to that in Equation (7.35), is computed:

$$\begin{bmatrix} \bar{R}_{51} & \hat{R}_{52} \\ R_{11} & 0 \\ \bar{R}_{61} & \bar{R}_{62} \end{bmatrix} =: \begin{bmatrix} \tilde{R}_{51} & 0 \\ \tilde{R}_{11} & \tilde{R}_{12} \\ \tilde{R}_{61} & \tilde{R}_{62} \end{bmatrix} \begin{bmatrix} \tilde{Q}_1 \\ \tilde{Q}_2 \end{bmatrix}. \tag{7.37}$$

Note that Υ_L can be determined from the above RQ factorization. To see how, we first observe the following factorization which can be obtained immediately from the above RQ factorization:

$$\Pi_{\mathcal{S}} \begin{bmatrix} Z_f \\ U_f \\ Z_r \end{bmatrix} = \begin{bmatrix} \bar{R}_{51} & \hat{R}_{52} & 0 & 0 \\ R_{11} & 0 & 0 & 0 \\ \bar{R}_{61} & \bar{R}_{62} & \bar{R}_{63} & \hat{R}_{64} \end{bmatrix} \hat{Q} = \begin{bmatrix} \tilde{R}_{51} & 0 & 0 & 0 \\ \tilde{R}_{11} & \tilde{R}_{12} & 0 & 0 \\ \tilde{R}_{61} & \tilde{R}_{62} & \bar{R}_{63} & \hat{R}_{64} \end{bmatrix} \begin{bmatrix} \tilde{Q}_1 & 0 \\ \tilde{Q}_2 & 0 \\ 0 & I \end{bmatrix} \hat{Q} =: \tilde{R}\tilde{Q}. \quad (7.38)$$

Proposition 7.7. *Suppose that the conditions in Theorem 5.15 are satisfied, with l being set to $2k$. Then, for $\tilde{R}_{12}$ and $\tilde{R}_{62}$ defined in Equation (7.37), we have*

$$\tilde{R}_{62} = \Upsilon_L \tilde{R}_{12}.$$

Proof. By Theorem 7.4 we have $\Pi_{\mathcal{S}}\mathcal{Z}_r \subset \Pi_{\mathcal{S}}\mathcal{Z}_f + \mathcal{U}_f$. This implies $\bar{R}_{63} = \hat{R}_{64} = 0$. Now, let $\mathcal{Q} := \Pi_{\mathcal{S}}\mathcal{Z}_f$. From Equation (7.38), it is easy to see that

$$\begin{aligned} \Pi_{\mathcal{Q}^\perp} U_f &= \begin{bmatrix} 0 & \tilde{R}_{12} & 0 & 0 \end{bmatrix} \tilde{Q}, \\ \Pi_{\mathcal{Q}^\perp} \Pi_{\mathcal{S}} Z_r &= \begin{bmatrix} 0 & \tilde{R}_{62} & 0 & 0 \end{bmatrix} \tilde{Q}. \end{aligned}$$

Thus, by Corollary 7.5 the result follows. □

Using the RQ factorization in (7.37) and Proposition 7.7, we can determine $h_k, \ldots, h_{2k-1}$ by following the procedure described in Section 7.4.

7.8.3 Determination of Kalman Filter State Sequences

The determination of Kalman filter state sequences is relatively straightforward. The first step is to extract U_U, U_L, Y_U and Y_L, which are defined in Equation (7.24). To obtain these block Hankel matrices, we can simply repartition the RQ factorization in Equation (7.34) as:

$$\left[\begin{array}{c} U_U \\ U_{3k-1} \\ \hline U_p \\ Y_p \\ \hline Y_U \\ Y_{3k-1} \end{array}\right] =: \left[\begin{array}{cccccc} R_{11}^{+} & & & & & \\ R_{21}^{-} & R_{22}^{-} & & & & \\ \hline R_{31} & R_{32} & R_{33} & & & \\ R_{41} & R_{42} & R_{43} & R_{44} & & \\ \hline R_{51}^{+} & R_{52}^{+} & R_{53}^{+} & R_{54}^{+} & R_{55}^{+} & \\ R_{61}^{-} & R_{62}^{-} & R_{63}^{-} & R_{64}^{-} & R_{65}^{-} & R_{66}^{-} \end{array}\right] Q := RQ, \quad (7.39)$$

and

$$\left[\begin{array}{c} U_k \\ U_L \\ \hline U_p \\ Y_p \\ \hline Y_k \\ Y_L \end{array}\right] =: \left[\begin{array}{cccccc} R_{11}^{-} & & & & & \\ R_{21}^{+} & R_{22}^{+} & & & & \\ \hline R_{31} & R_{32} & R_{33} & & & \\ R_{41} & R_{42} & R_{43} & R_{44} & & \\ \hline R_{51}^{-} & R_{52}^{-} & R_{53}^{-} & R_{54}^{-} & R_{55}^{-} & \\ R_{61}^{+} & R_{62}^{+} & R_{63}^{+} & R_{64}^{+} & R_{65}^{+} & R_{66}^{+} \end{array}\right] Q := RQ. \quad (7.40)$$

Using the first $2k$ Markov parameters determined in the previous subsections, we reconstruct $\mathcal{T}_{2k-1}$. Consequently, Equations (7.25) and (7.26) can be rewritten as:

$$\begin{aligned}
\mathcal{O}_{2k-1}\,\Pi_{\mathcal{S}} X_k &= \Pi_{\mathcal{S}} Y_U - \mathcal{T}_{2k-1} U_U \\
&= \begin{bmatrix} R_{51}^{+} & R_{52}^{+} & R_{53}^{+} & R_{54}^{+} & 0 & 0 \end{bmatrix} Q - \mathcal{T}_{2k-1} \begin{bmatrix} R_{11}^{+} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} Q \\
&=: \begin{bmatrix} \bar{R}_{51}^{+} & R_{52}^{+} & R_{53}^{+} & R_{54}^{+} & 0 & 0 \end{bmatrix} Q,
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{O}_{2k-1}\,\Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1} &= \Pi_{\mathcal{S}+\mathcal{Y}_k} Y_L - \mathcal{T}_{2k-1} U_L \\
&= \begin{bmatrix} R_{61}^{+} & R_{62}^{+} & R_{63}^{+} & R_{64}^{+} & R_{65}^{+} & 0 \end{bmatrix} Q - \mathcal{T}_{2k-1} \begin{bmatrix} R_{21}^{+} & R_{22}^{+} & 0 & 0 & 0 & 0 \end{bmatrix} Q \\
&=: \begin{bmatrix} \bar{R}_{61}^{+} & \bar{R}_{62}^{+} & R_{63}^{+} & R_{64}^{+} & R_{65}^{+} & 0 \end{bmatrix} Q.
\end{aligned}$$

Finally, the following SVD is computed:

$$[\,\bar{R}_{51}^{+} \;\; R_{52}^{+} \;\; R_{53}^{+} \;\; R_{54}^{+} \;\; 0 \;\; 0 \;\; \bar{R}_{61}^{+} \;\; \bar{R}_{62}^{+} \;\; R_{63}^{+} \;\; R_{64}^{+} \;\; R_{65}^{+} \;\; 0\,] =: \begin{bmatrix} \Gamma_1 & \Gamma_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \Omega_1^* \\ \Omega_2^* \end{bmatrix}.$$

Then, we have

$$\mathcal{O}_{2k-1} = \Gamma_1 \Sigma_1^{\frac{1}{2}}, \quad \text{and} \quad \begin{bmatrix} \Pi_{\mathcal{S}} X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k} X_{k+1} \end{bmatrix} = \Sigma_1^{\frac{1}{2}} \Omega_1^* Q.$$

7.9 Summary

In this chapter, a subspace identification method via Markov parameter estimation is presented. A new approach, using a three-block configuration, is advocated. In this three-block configuration, each block serves a different function — either an instrumental variable, a Kalman filter state observer, or a computational data set. Numerically efficient implementation using RQ factorization is also developed.

One immediate advantage of the proposed method is its compatibility with Kung's algorithm, and the simplicity of acquiring balanced approximations. Another advantage is its unbiasedness in estimating the system matrices with the state sequence approach. Note also that in estimating the observability matrix and the Kalman filter state sequence, all components relating to the input sequences are first removed, as in Equations (7.25) and (7.26). In such a way, any negative effect due to ill conditioning of the input signals can be minimized.

One drawback of the proposed method is that it has a more complicated overall procedure, in which two additional RQ factorizations (7.35) and (7.37) are required. Nonetheless, one can argue that these RQ factorizations should not adversely affect the performance of the algorithms, as they are numerically robust and of relatively small size. On the other hand, one should instead query whether the Markov parameters are accurate or the algorithms are effective. To this end, further investigation, by means of a few case studies, is carried out in Chapter 9. Finally, schematic overviews of the algorithms developed in this chapter can be found in pages 96–98.

Algorithm 7.8. ID via MP Estimation — A Markov Parameter Appr.

1. *Construct* U_p, U_f, U_r *and* Y_p, Y_f, Y_r *as in Equation* (7.6)*.*
2. *Determine* $h_0, \ldots, h_{k-1}$ *from:*

$$\Pi_{\mathcal{R}^\perp}\Pi_{\mathcal{S}}Y_r = \mathcal{T}_k\,\Pi_{\mathcal{R}^\perp}U_r, \qquad \textit{for } \mathcal{R} := \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f.$$

3. *Compute* Z_f *and* Z_r *as in* (7.12)*. Then determine* $h_k, \ldots, h_{2k-1}$ *from:*

$$\Pi_{\mathcal{Q}^\perp}\Pi_{\mathcal{S}}Z_r = \Upsilon_L\,\Pi_{\mathcal{Q}^\perp}U_f, \qquad \textit{for } \mathcal{Q} := \Pi_{\mathcal{S}}\mathcal{Z}_f.$$

4. *Extract* U_U, U_L *and* Y_U *and* Y_L *as in* (7.24)*. Calculate the following SVD and partition accordingly by selecting an model order:*

$$\left[\begin{array}{c:c:c} h_1 & & \\ \vdots & \Pi_{\mathcal{S}}Y_U - \mathcal{T}_{2k-1}U_U & \Pi_{\mathcal{S}+\mathcal{Y}_k}Y_L - \mathcal{T}_{2k-1}U_L \\ h_{2k-1} & & \end{array}\right] =: \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}.$$

5. *Assign* $D = h_0$*. Determine* $\mathcal{O}_{2k-1}$, B, $\Pi_{\mathcal{S}}X_k$ *and* $\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}$ *as:*

$$\mathcal{O}_{2k-1} = \Gamma_1\Sigma_1^{\frac{1}{2}}, \qquad \textit{and} \qquad \begin{bmatrix}B & \Pi_{\mathcal{S}}X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\end{bmatrix} = \Sigma_1^{\frac{1}{2}}\Omega_1^*.$$

6. *Determine* A *and* C *from* $\mathcal{O}_{2k-1}$ *using the shift invariance property, as in* (7.28)*.*
7. *Determine* Σ^w, Σ^v, *and* Σ^{wv} *by taking the covariance of:*

$$\begin{bmatrix}\varepsilon_w\\ \varepsilon_v\end{bmatrix} := \begin{bmatrix}\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\\ Y_k\end{bmatrix} - \begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\Pi_{\mathcal{S}}X_k\\ U_k\end{bmatrix}.$$

8. *Determine Kalman filter gain* K *and the error covariance matrix* P *using* (7.32) *and* (7.33)*.*

Figure 7.1 *A schematic overview of the subspace identification via Markov parameter estimation — a Markov parameter approach.*

Algorithm 7.9. ID via MP Estimation — A Shift Invariance Approach

1. *Construct* U_p, U_f, U_r *and* Y_p, Y_f, Y_r *as in Equation* (7.6).
2. *Determine* $h_0, \ldots, h_{k-1}$ *from:*
$$\Pi_{\mathcal{R}^\perp}\Pi_{\mathcal{S}}Y_r = \mathcal{T}_k\, \Pi_{\mathcal{R}^\perp}U_r, \qquad \textit{for } \mathcal{R} := \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f.$$
3. *Compute* Z_f *and* Z_r *as in* (7.12). *Then determine* $h_k, \ldots, h_{2k-1}$ *from:*
$$\Pi_{\mathcal{Q}^\perp}\Pi_{\mathcal{S}}Z_r = \Upsilon_L\, \Pi_{\mathcal{Q}^\perp}U_f, \qquad \textit{for } \mathcal{Q} := \Pi_{\mathcal{S}}\mathcal{Z}_f.$$
4. *Extract* U_U, U_L *and* Y_U *and* Y_L *as in* (7.24). *Calculate the following SVD and partition accordingly by selecting an model order:*
$$\begin{bmatrix}\Pi_{\mathcal{S}}Y_U & \Pi_{\mathcal{S}+\mathcal{Y}_k}Y_L\end{bmatrix} - \mathcal{T}_{2k-1}\begin{bmatrix}U_U & U_L\end{bmatrix} =: \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}.$$
5. *Determine* $\mathcal{O}_{2k-1}$, $\Pi_{\mathcal{S}}X_k$ *and* $\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}$ *as:*
$$\mathcal{O}_{2k-1} = \Gamma_1\Sigma_1^{\frac{1}{2}}, \quad \textit{and} \quad \begin{bmatrix}\Pi_{\mathcal{S}}X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\end{bmatrix} = \Sigma_1^{\frac{1}{2}}\Omega_1^*.$$
6. *Determine* A *and* C *from* $\mathcal{O}_{2k-1}$ *using the shift invariance property, as in* (7.28).
7. *Reconstruct* $\mathcal{O}_{2k}$. *Solve for* B *and* D *as in* (7.29) *from:*
$$B, D = \arg\min_{B,D}\left\| \Pi_{\mathcal{S}}\begin{bmatrix}Y_f\\ Y_r\end{bmatrix} - \mathcal{O}_{2k}\Pi_{\mathcal{S}}X_k - \mathcal{T}_{2k}(B,D)\begin{bmatrix}U_f\\ U_r\end{bmatrix}\right\|_F.$$
8. *Determine* Σ^w, Σ^v, *and* Σ^{wv} *by taking the covariance of:*
$$\begin{bmatrix}\varepsilon_w\\ \varepsilon_v\end{bmatrix} := \begin{bmatrix}\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\\ Y_k\end{bmatrix} - \begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\Pi_{\mathcal{S}}X_k\\ U_k\end{bmatrix}.$$
9. *Determine Kalman filter gain* K *and the error covariance matrix* P *using* (7.32) *and* (7.33).

Figure 7.2 *A schematic overview of the subspace identification via Markov parameter estimation — a shift invariance approach.*

Algorithm 7.10. ID via MP Estimation — A State Sequence Approach

1. *Construct* U_p, U_f, U_r *and* Y_p, Y_f, Y_r *as in Equation* (7.6)*.*
2. *Determine* $h_0, \ldots, h_{k-1}$ *from:*

$$\Pi_{\mathcal{R}^\perp}\Pi_{\mathcal{S}}Y_r = \mathcal{T}_k\ \Pi_{\mathcal{R}^\perp}U_r, \qquad \text{for } \mathcal{R} := \Pi_{\mathcal{S}}\mathcal{Y}_f + \mathcal{U}_f.$$

3. *Compute* Z_f *and* Z_r *as in* (7.12)*. Then determine* $h_k, \ldots, h_{2k-1}$ *from:*

$$\Pi_{\mathcal{Q}^\perp}\Pi_{\mathcal{S}}Z_r = \Upsilon_L\ \Pi_{\mathcal{Q}^\perp}U_f, \qquad \text{for } \mathcal{Q} := \Pi_{\mathcal{S}}\mathcal{Z}_f.$$

4. *Extract* U_U, U_L *and* Y_U *and* Y_L *as in* (7.24)*. Calculate the following SVD and partition accordingly by selecting an model order:*

$$\begin{bmatrix}\Pi_{\mathcal{S}}Y_U & \Pi_{\mathcal{S}+\mathcal{Y}_k}Y_L\end{bmatrix} - \mathcal{T}_{2k-1}\begin{bmatrix}U_U & U_L\end{bmatrix} =: \begin{bmatrix}\Gamma_1 & \Gamma_2\end{bmatrix}\begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix}\begin{bmatrix}\Omega_1^*\\ \Omega_2^*\end{bmatrix}.$$

5. *Determine* $\Pi_{\mathcal{S}}X_k$ *and* $\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}$ *as:*

$$\begin{bmatrix}\Pi_{\mathcal{S}}X_k & \Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\end{bmatrix} = \Sigma_1^{\frac{1}{2}}\Omega_1^*.$$

6. *Determine* A, B, C *and* D *from:*

$$A,B,C,D = \arg\min_{A,B,C,D}\left\| \begin{bmatrix}\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\\ Y_k\end{bmatrix} - \begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\Pi_{\mathcal{S}}X_k\\ U_k\end{bmatrix}\right\|_F.$$

7. *Determine* Σ^w, Σ^v, *and* Σ^{wv} *by taking the covariance of:*

$$\begin{bmatrix}\varepsilon_w\\ \varepsilon_v\end{bmatrix} := \begin{bmatrix}\Pi_{\mathcal{S}+\mathcal{Y}_k}X_{k+1}\\ Y_k\end{bmatrix} - \begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\Pi_{\mathcal{S}}X_k\\ U_k\end{bmatrix}.$$

8. *Determine Kalman filter gain* K *and the error covariance matrix* P *using* (7.32) *and* (7.33)*.*

Figure 7.3 *A schematic overview of the subspace identification via Markov parameter estimation — a state sequence approach.*

8

Enforcing Stability with Realization and Subspace Algorithms

8.1 Introduction

Approximate realization algorithms [ZM74, Kun78, DPK85] and subspace algorithms [Lar83, MDMVV89, Ver93b, VODM94, Ver94] have been developed in the field of system identification. These methods have also been proven to be extremely effective, particularly for the identification and approximation of multivariable systems. Closely related algorithms are also used for signal detection. In all these methods the range space of the extended observability matrix $\mathcal{O}_k$ is estimated by low-rank approximate factorization of a Hankel matrix built up from data, and an approximate extended observability matrix Γ is obtained by choosing a basis for this space.[1] Some approximate realization algorithms and subspace algorithms then use the shift invariance property of the approximate observability matrix, namely

$$\Gamma^{\uparrow} \approx \Gamma A,$$

to estimate system matrices. Here, $\Gamma^{\uparrow}$ denotes Γ with each row shifted upwards by p rows, where p is the number of outputs. Solving this overdetermined equation in a least-squares sense can yield an unstable system.

In some applications, such as some adaptive control schemes containing explicit internal models, it is essential to ensure that models obtained by automatic procedures are stable. In other applications, like identification of systems containing periodic dynamics, marginally stable models are then desired. In all these cases the requirement of stability or marginal stability can justify a slight loss of model accuracy in certain respects. It was shown in [Mac95] that asymptotic stability of the estimated system can be ensured by filling the lowest p rows of $\Gamma^{\uparrow}$ with zeros. This procedure inevitably distorts the results, however, because setting these rows to zero may involve considerable distortion of the estimated observability matrix. In this chapter we first show that one can ensure stability with smaller distortions, by adopting a new approach. Moreover, with this new approach, particular eigenvalues can also be forced to lie on the unit circle, creating certain periodic behavior of the model.

[1] Here, we denote by Γ the approximant of $\mathcal{O}_k$; that is, $\Gamma \approx \mathcal{O}_k$. Note that unlike $\mathcal{O}_k$, Γ in general does not satisfy the shift invariance property exactly.

Some subspace algorithms use the estimated observability matrix to determine a sequence of state approximates $X = \begin{bmatrix} x_1 & \cdots & x_q \end{bmatrix}$, then estimate all the matrices of a state-space model by solving the relation

$$\begin{bmatrix} X^{\leftarrow} \\ Y \end{bmatrix} \approx \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix}$$

in the least-squares sense, where U and Y contain input and output signals stacked side-by-side, and $X^{\leftarrow}$ is the estimated state sequence of the next time instant; i.e. $X^{\leftarrow} = \begin{bmatrix} x_2^{\leftarrow} & \cdots & x_{q+1}^{\leftarrow} \end{bmatrix}$.[2] Again, this state sequence method can yield an unstable model. It seems to be difficult to extend the approach taken in [Mac95] to the state sequence method. In this chapter, however, we show that the new approach developed for the shift invariance method extends naturally to the state sequence method.

Towards the end of this chapter, some quantitative estimates of the amount of distortion introduced by the new approach and other existing algorithms are provided in both the shift invariance case and the state sequence case. Note that a matrix is said to be stable if it has a spectral radius strictly less than one, semistable if less than or equal to one, marginally stable if equal to one with eigenvalues on the unit circle having geometric multiplicity 1.

8.2 The Shift Invariance Approach

When the approximate extended observability matrix Γ or the approximate extended controllability matrix Ω is estimated, the system matrix A is usually recovered in the least-squares sense:

$$\hat{A} = \arg\min_A \left\| \Gamma^{\uparrow} - \Gamma A \right\|_F, \qquad \text{and} \qquad \hat{A} = \arg\min_A \left\| \Omega^{\leftarrow} - A\Omega \right\|_F.$$

The solutions of these equations are given by $\Gamma^{\dagger}\Gamma^{\uparrow}$ and $\Omega^{\leftarrow}\Omega^{\dagger}$, respectively. We shall assume that Γ has full column rank and Ω has full row rank so that $\Gamma^{\dagger} = (\Gamma^*\Gamma)^{-1}\Gamma^*$ and $\Omega^{\dagger} = \Omega^*(\Omega\Omega^*)^{-1}$.

8.2.1 Stability: Sufficient Condition

Here we re-derive and generalize the main result of [Mac95] in a different manner which highlights a connection with autoregressive modeling. Let $R_0 := \Gamma^*\Gamma \in \mathbb{R}^{n\times n}$ and $R_1 := \Gamma^*\Gamma^{\uparrow} \in \mathbb{R}^{n\times n}$. Then, the least-squares solution can also be written as $\hat{A} = R_0^{-1}R_1$.

Lemma 8.1. *Suppose $R_0 > 0$ and*

$$S := \begin{bmatrix} R_0 & R_1 \\ R_1^* & R_0 \end{bmatrix},$$

then $\hat{A}$ is stable if $S > 0$, or $\hat{A}$ is semistable if $S \geq 0$.

[2] Note that $X^{\leftarrow}$ may not be a left-shifted version of X. That is, x_τ and $x_\tau^{\leftarrow}$ may not be the same, for $\tau = 2, \ldots, q$.

Proof. Note that the Schur complement of the Hermitian matrix S is $R_0 - R_1^* R_0^{-1} R_1$. Thus, S is positive definite (semidefinite) if and only if $R_0 - R_1^* R_0^{-1} R_1$ is positive definite (semidefinite), since the inertia of S is equal to the sum of the inertia of R_0 and its Schur complement (for example, see [HJ94]). Substituting $\hat{A}$ into the Schur complement, we have $R_0 - \hat{A}^* R_0 \hat{A} > 0\ (\geq 0)$, a Lyapunov inequality, which ensures the stability (semistability) of $\hat{A}$. $\square$

Note that the converse of the above theorem is in general not true. One reason is that the condition $S > 0$ or $S \geq 0$ with a fixed $R_0 > 0$ characterizes only a convex set of stable or semistable $\hat{A} \in \mathbb{R}^{n \times n}$. However, the set of all $n \times n$ stable or semistable matrices is non-convex for $n > 1$.

Remark 8.1. Lemma 8.1 can also be obtained as a consequence of the stability results for the Whittle-Wiggins-Robinson algorithm for autoregressive model identification (for example, see [SS89]). $\triangle$

Partition $\Gamma \in \mathbb{R}^{(k-1)p \times n}$ into $k-1$ blocks, each of size $p \times n$

$$\Gamma = \begin{bmatrix} \gamma_1^* & \gamma_2^* & \cdots & \gamma_{k-1}^* \end{bmatrix}^*.$$

Then $\Gamma^{\uparrow} \in \mathbb{R}^{(k-1)p \times n}$ is a matrix with each block shifted upwards

$$\Gamma^{\uparrow} = \begin{bmatrix} \gamma_2^* & \cdots & \gamma_{k-1}^* & \gamma_k^* \end{bmatrix}^*.$$

Theorem 8.2. *Let $\hat{A} = \Gamma^{\dagger} \Gamma^{\uparrow}$. If $\gamma_k = T\gamma_1$ where $T \in \mathbb{R}^{p \times p}$, $\bar{\sigma}(T) \leq 1$, then*

$$\rho(\hat{A}) \leq 1.$$

Proof. Note that by letting $R_0 = \Gamma^* \Gamma$ and $R_1 = \Gamma^* \Gamma^{\uparrow}$, S can be written as:

$$S = \begin{bmatrix} \Gamma^* & 0 \\ \Gamma^{\uparrow *} & \gamma_1^* (I - T^* T)^{\frac{1}{2}} \end{bmatrix} \begin{bmatrix} \Gamma & \Gamma^{\uparrow} \\ 0 & (I - T^* T)^{\frac{1}{2}} \gamma_1 \end{bmatrix},$$

where $(I - T^* T)^{\frac{1}{2}}$ is defined in $\mathbb{R}^{p \times p}$ if $\bar{\sigma}(T) \leq 1$. It is now easy to see that S is positive semidefinite which implies $\hat{A}$ is semistable. $\square$

Obviously, the following also holds, with $\Omega, \Omega^{\leftarrow} \in \mathbb{R}^{n \times (k-1)m}$

$$\Omega = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_{k-1} \end{bmatrix},$$
$$\Omega^{\leftarrow} = \begin{bmatrix} \omega_2 & \cdots & \omega_{k-1} & \omega_k \end{bmatrix}.$$

Theorem 8.3. *Let $\hat{A} = \Omega^{\leftarrow} \Omega^{\dagger}$. If $\omega_k = \omega_1 T$ where $T \in \mathbb{R}^{m \times m}$, $\bar{\sigma}(T) \leq 1$, then*

$$\rho(\hat{A}) \leq 1.$$

Proof. As for Theorem 8.2, taking transposes of everything. □

Theorems 8.2 and 8.3 give a sufficient stability condition stating that if the last piece of data equals the first piece of data under a linear transformation with an operator norm no greater than 1, then the least-squares solution must have all its eigenvalues in the closed unit disc. Furthermore, we can infer that by adding an extra piece of data which complies with the stated condition, we can always obtain a stable system. Taking $\gamma_k = 0$ gives the result of [Mac95], which is thus obtained as a special case. It is apparently well known in autoregressive modeling that forming sample covariances as $R_0 = \frac{1}{k-1}\Gamma^*\Gamma$ and $R_1 = \frac{1}{k-1}\Gamma^*\Gamma^\uparrow$, with $\gamma_k = 0$, leads to a semistable first-order AR model.

8.2.2 Augmenting the Data

Our objective is to obtain an estimated system which is not only stable, but also matches the data reasonably well. The idea is to add a pair of correcting blocks γ_i and γ_o, each of appropriate size, to Γ and $\Gamma^\uparrow$ respectively,

$$\Gamma_D = \begin{bmatrix} \gamma_1^* & \cdots & \gamma_{k-1}^* & \gamma_i^* \end{bmatrix}^*, \tag{8.1}$$

$$\Gamma_D^\uparrow = \begin{bmatrix} \gamma_2^* & \cdots & \gamma_k^* & \gamma_o^* \end{bmatrix}^*. \tag{8.2}$$

Let $\hat{A} = \Gamma^\dagger\Gamma^\uparrow$ be the original estimate and $\hat{A}_D = \Gamma_D^\dagger\Gamma_D^\uparrow$ be the corrected estimate. The difference between the two matrices is given by

$$\begin{aligned}
\hat{A}_D - \hat{A} &= (\Gamma^*\Gamma + \gamma_i^*\gamma_i)^{-1}(\Gamma^*\Gamma^\uparrow + \gamma_i^*\gamma_o) - \hat{A} \\
&= (I + (\Gamma^*\Gamma)^{-1}\gamma_i^*\gamma_i)^{-1}(\Gamma^*\Gamma)^{-1}\gamma_i^*\gamma_o - \left[I - (I + (\Gamma^*\Gamma)^{-1}\gamma_i^*\gamma_i)^{-1}\right]\hat{A} \\
&= (I + (\Gamma^*\Gamma)^{-1}\gamma_i^*\gamma_i)^{-1}(\Gamma^*\Gamma)^{-1}\left[\gamma_i^*\gamma_o - \gamma_i^*\gamma_i\hat{A}\right] \\
&= (\Gamma^*\Gamma + \gamma_i^*\gamma_i)^{-1}\gamma_i^*[\gamma_o - \gamma_i\hat{A}].
\end{aligned} \tag{8.3}$$

This is not surprising since $\gamma_i\hat{A}$ is the least-squares estimate of γ_o. Note that $\hat{A}$ is recovered either by setting $\gamma_i = 0$ or $\gamma_o = \gamma_i\hat{A}$. One simple way to obtain a semistable $\hat{A}_D$ is to set $\gamma_i = cQ$ and $\gamma_o = 0$. Here, we suppose $c > 0$ and $Q \in \mathbb{R}^{n\times n}$ is unitary. We can conclude that with c large enough, $\hat{A}_D = \Gamma_D^\dagger\Gamma_D^\uparrow$ is semistable. In fact, this can be done if c is as large as the largest singular value of γ_k:

Proposition 8.4. *Suppose $\hat{A}_D = \Gamma_D^\dagger\Gamma_D^\uparrow$, with $\gamma_i = cQ$, $\gamma_o = 0$, and $c \geq \bar{\sigma}(\gamma_k)$, where Q is unitary. Then*

$$\rho(\hat{A}_D) \leq 1. \tag{8.4}$$

Proof. Suppose

$$T = \begin{bmatrix} 0 & I & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & 0 & I & 0 \\ \vdots & & \cdots & \ddots & \frac{\gamma_k Q^*}{c} \\ 0 & \cdots & \cdots & \cdots & 0 \end{bmatrix}. \tag{8.5}$$

Then, we have $\Gamma_D^{\uparrow} = T\Gamma_D$. Note that $\bar{\sigma}(T) \leq 1$.

$$\begin{aligned} \rho(\hat{A}) &= \rho\big((\Gamma_D^*\Gamma_D)^{-1}\Gamma_D^* T\Gamma_D\big) \\ &\leq \bar{\sigma}\big(\Gamma_D(\Gamma_D^*\Gamma_D)^{-1}\Gamma_D^*\big)\bar{\sigma}(T) \\ &= \bar{\sigma}(T). \end{aligned}$$

Thus, the result follows. □

One advantage of this approach is that it can be extended to the state sequence approach, as will be seen later. It is noteworthy that in Theorems 8.2 and 8.3, if we relax the full rank condition on Γ and Ω, the results still hold. This can be seen from the approach used in the proof of Proposition 8.4. In fact, for Theorem 8.2 we can write $\Gamma^{\uparrow} = \tilde{T}\Gamma$ for some $\tilde{T}$, $\bar{\sigma}(\tilde{T}) \leq 1$. Then from the fact that $\bar{\sigma}(\Gamma\Gamma^{\dagger}) = 1$ (or 0 if $\Gamma = 0$), we arrive at the same results.

8.2.3 Eigenvalue Perturbations

There are many ways of perturbing the eigenvalues of a matrix. Before developing the data augmentation method further, we first propose some other feasible methods of perturbing the eigenvalues which can be applied with subspace methods.

One simple method of perturbing the eigenvalues is to multiply by a scaling factor. For example, to get a stable matrix with spectral radius no greater than $M < 1$ from the original matrix $\hat{A}$, we can divide the matrix by a factor as follows,

$$\hat{A}_S := \hat{A}/\rho(\hat{A}) * M. \tag{8.6}$$

Hereafter we refer to the matrix obtained by this scaling method as $\hat{A}_S$.

Other methods of perturbing eigenvalues involve complex Schur decomposition (which has been suggested in [McK95]) or Jordan decomposition. These methods basically decompose $\hat{A}$ into $T\Lambda T^*$ or $T(\Lambda + \aleph)T^{-1}$, where $\aleph$ is a nilpotent matrix (see for example [HJ93]), and then modify the eigenvalues in Λ to the desired values. For instance, to perturb eigenvalue λ_1 to $\frac{\lambda_1}{M}$, we have

$$T\left(\begin{bmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \Lambda_1 \end{bmatrix} + \aleph\right) T^{-1} \quad \rightarrow \quad T\left(\begin{bmatrix} \frac{\lambda_1}{M} & & * \\ & \ddots & \\ 0 & & \Lambda_1 \end{bmatrix} + \aleph\right) T^{-1}. \tag{8.7}$$

(For Schur decomposition, $\aleph = 0$, $T^{-1} = T^*$, whereas for Jordan decomposition, $* = 0$.) We refer to the matrix obtained by the Schur decomposition technique as $\hat{A}_U$, and that obtained from the Jordan decomposition technique as $\hat{A}_J$. Note that the Jordan decomposition approach preserves the eigenspace structure but is not numerically stable.

The above two methods use no information about the available data series to perform corrections. As described in the previous section, one can always pad some terms to the data series to guarantee semistability. We refer to the matrix modified by padding a block of zeros to the end of the data series (i.e., as in Theorem 8.2, with $T = 0$) as $\hat{A}_Z$.

We have seen in Proposition 8.4 that γ_o of size $n \times n$ can adequately semi-stabilize the matrix $\hat{A}$. Nevertheless, padding with such γ_o may overwhelm the characteristics of the other modes of the system, and thus introduce considerable distortion. The following theorems, whose proofs can be found in Appendix D, provide some methods of adjusting one eigenvalue, or a complex-conjugate pair of eigenvalues at one time.

Theorem 8.5. *Suppose $\nu \in \mathbb{C}^{1\times n}$ is a normalized left eigenvector of $\hat{A}$ corresponding to the non-zero eigenvalue λ. If $\gamma_i = c\nu$, $\gamma_o = b\nu\hat{A}$, $b, c \in \mathbb{R}$, and $\hat{A}_D = \Gamma_D^{\dagger}\Gamma_D^{\uparrow}$ as defined in Equations* (8.1) *and* (8.2), *then*

1. *every eigenvalue of $\hat{A}$ is also an eigenvalue of $\hat{A}_D$ with an exception of the eigenvalue λ;*
2. *the perturbed eigenvalue λ_D of $\hat{A}_D$ is*
$$\lambda_D = \frac{1 + bc\nu(\Gamma^*\Gamma)^{-1}\nu^*}{1 + c^2\nu(\Gamma^*\Gamma)^{-1}\nu^*}\,\lambda; \tag{8.8}$$
3. *$\hat{A}_D$ is independent of the choice of the pair b and c in the set which gives the same λ_D;*
4. *the induced 2-norm of the difference of $\hat{A}$ and $\hat{A}_D$ is bounded from above by*
$$\|\hat{A}_D - \hat{A}\|_2 \leq \frac{|\lambda_D - \lambda|}{|\lambda|}\sqrt{\kappa(\Gamma^*\Gamma)}\;\|\hat{A}\|_2. \tag{8.9}$$

If the eigenvalue concerned is real, one can always choose a real valued b, c and a corresponding real eigenvector. So, the resulting system $\hat{A}_D$ will also be real. However, if a complex eigenvalue is chosen, a complex eigenvector is then inevitable; and hence, the system will no longer be real. One way to solve this problem is to adjust both the complex eigenvalue and its complex conjugate simultaneously. Observe that if ν is an eigenvector of a real system corresponding to a complex eigenvalue λ, then $\bar{\nu}$ is an eigenvector corresponding to $\bar{\lambda}$, where $\bar{\nu}$ and $\bar{\lambda}$ are the complex conjugates of ν and λ respectively.

Theorem 8.6. *Suppose* $\gamma_i = c\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}$, $\gamma_o = b\begin{bmatrix} \nu\hat{A} \\ \bar{\nu}\hat{A} \end{bmatrix}$, *where* $b, c \in \mathbb{R}$ *and* $\nu, \bar{\nu} \in \mathbb{C}^{1\times n}$ *are normalized left eigenvectors of* $\hat{A}$ *corresponding to* λ *and its complex conjugate* $\bar{\lambda}$ *respectively. If* $\hat{A}_D = \Gamma_D^{\dagger}\Gamma_D^{\uparrow}$ *as defined in Equations* (8.1) *and* (8.2), *then*

1. $\hat{A}_D$ *is real and shares every eigenvalue with* $\hat{A}$ *except* λ *and* $\bar{\lambda}$;
2. *the perturbed eigenvalues* λ_{c1} *and* λ_{c2} *are the eigenvalues of both matrices*

$$\Lambda_D := \left(I + c^2\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}(\Gamma^*\Gamma)^{-1}\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}^*\right)^{-1} \left(I + bc\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}(\Gamma^*\Gamma)^{-1}\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}^*\right)\begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}, \tag{8.10}$$

and

$$\tilde{\Lambda}_D := \left(I + c^2\begin{bmatrix} \nu \\ \mu \end{bmatrix}(\Gamma^*\Gamma)^{-1}\begin{bmatrix} \nu \\ \mu \end{bmatrix}^*\Theta^*\Theta\right)^{-1} \left(I + bc\begin{bmatrix} \nu \\ \mu \end{bmatrix}(\Gamma^*\Gamma)^{-1}\begin{bmatrix} \nu \\ \mu \end{bmatrix}^*\Theta^*\Theta\right)\begin{bmatrix} \lambda & 0 \\ \delta & \bar{\lambda} \end{bmatrix}, \tag{8.11}$$

where

$$\alpha = \bar{\nu}\nu^*, \qquad \beta = \sqrt{1-|\alpha|^2}, \qquad \delta = \frac{\alpha}{\beta}(\bar{\lambda}-\lambda), \qquad \mu = \frac{\bar{\nu}-\alpha\nu}{\beta}, \qquad \Theta = \begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix};$$

3. $\hat{A}_D$ *is independent of the choice of the pair* b *and* c *in the set which gives the same* Λ_D;
4. *the induced 2-norm of the difference of* $\hat{A}$ *and* $\hat{A}_D$ *is bounded from above by*

$$\|\hat{A}_D - \hat{A}\|_2 \le \frac{\bar{\sigma}\left(\Lambda_D - \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}\right)}{|\lambda|}\sqrt{\kappa(\Gamma^*\Gamma)}\;\|\hat{A}\|_2. \tag{8.12}$$

For numerical stability, eigenvalues of $\hat{A}$ in Theorems 8.5 and 8.6 can be determined by a complex Schur decomposition in an upper triangular Schur form. To find a left eigenvector corresponding to an eigenvalue, one can re-order the eigenvalues in the complex Schur form in such a way that the corresponding eigenvalue is in the lowest diagonal entry. Therefore, its normalized left eigenvector can be obtained from the unitary matrix of the Schur decomposition. For ordering eigenvalues in the Schur form, see [GVL89] for details.

Remark 8.2. From Equations (D.3) and (D.4), we see that the size of each Jordan block associated with each eigenvalue λ_i, $\lambda_i \neq \lambda$ and $\lambda_i \neq \bar{\lambda}$, is preserved for almost every b and c. The size of the Jordan block of λ (and $\bar{\lambda}$) is reduced by one after the perturbation. One should be aware that if the perturbed eigenvalue has the same value as one of the other eigenvalues, it may increase the dimension of the corresponding eigenspace by one, or may even link two eigenspaces of the same eigenvalue together. It is also interesting to note from Equation (D.3) that the eigenvector ν is also an eigenvector corresponding to the perturbed eigenvalue in the real valued case. $\triangle$

8.2.4 Acquiring Stable Systems

Theorems 8.5 and 8.6 provide a data augmentation method for perturbing the eigenvalues of a least-squares solution. This method can perturb eigenvalues either towards the origin (attenuating the magnitudes) or away from the origin (magnifying the magnitudes). Note that when $b = c$, the augmentation induces no effect on $\hat{A}$. To modify an unstable estimate, we want to attenuate unstable eigenvalues to a stable location. In this case, we will set $b = 0$ and assume $c > 0$ for the sake of simplicity. As a result, Equations (8.8) and (8.10) respectively simplify to

$$\lambda_D = \frac{\lambda}{1 + c^2 \nu (\Gamma^* \Gamma)^{-1} \nu^*}, \tag{8.13}$$

and

$$\Lambda_D = \left(I + c^2 \begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix} (\Gamma^* \Gamma)^{-1} \begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix}^* \right)^{-1} \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}. \tag{8.14}$$

For the real eigenvalue case, we can use Equation (8.13) to find a value c so that $|\lambda_D| = M$, where $M \leq 1$, a pre-defined magnitude. It is easy to see that the mapping $c \mapsto |\lambda_D|$, $c > 0$, is strictly decreasing since $\nu(\Gamma^*\Gamma)\nu^* > 0$. Therefore there exists a unique $c > 0$ such that the perturbed eigenvalue reaches the magnitude M when $|\lambda| > 1$. This is precisely

$$c = c_r := \sqrt{\frac{|\lambda|/M - 1}{\nu(\Gamma^*\Gamma)\nu^*}}. \tag{8.15}$$

For the complex case, we shall proceed from Equation (8.14). First let

$$\begin{bmatrix} \upsilon & \chi \\ \varphi & \zeta \end{bmatrix} := \begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix} (\Gamma^* \Gamma)^{-1} \begin{bmatrix} \nu^* & \bar{\nu}^* \end{bmatrix}.$$

Note that $\upsilon = \zeta$ since $(\Gamma^*\Gamma)^{-1} \in \mathbb{R}^{n \times n}$ and $\chi^* = \varphi$. The characteristic polynomial of Λ_D is

$$\begin{aligned} \det(sI - \Lambda_D) &= \det \begin{bmatrix} s - \frac{(1+c^2\zeta)\lambda}{\phi} & \frac{c^2\chi\bar{\lambda}}{\phi} \\ \frac{c^2\varphi\lambda}{\phi} & s - \frac{(1+c^2\upsilon)\bar{\lambda}}{\phi} \end{bmatrix} \\ &= s^2 - \left[\frac{(1+c^2\zeta)\lambda + (1+c^2\upsilon)\bar{\lambda}}{\phi} \right] s + \frac{|\lambda|^2}{\phi}, \end{aligned} \tag{8.16}$$

where $\phi = (1+c^2\upsilon)(1+c^2\zeta) - c^4\chi\varphi$. Suppose λ_{c1} and λ_{c2} are complex, then $|\lambda_{c1}|^2 = |\lambda_{c2}|^2 = \frac{|\lambda|^2}{\phi}$. We have

$$\frac{b}{bc}\frac{|\lambda|^2}{\phi} = -\frac{|\lambda|^2}{\phi^2}\big(4c^3(\upsilon\zeta - \chi\varphi) + 2c(\upsilon + \zeta)\big) < 0$$

for $c > 0$. So, the mapping $c \mapsto (|\lambda_{c1}|, |\lambda_{c2}|)$ is also strictly decreasing. Now, suppose $|\lambda| \geq 1$, then $\lambda_{c1}\lambda_{c2} = M^2$ when $\frac{|\lambda|^2}{\phi} = M^2$, or equivalently,

$$c = c_c := \left(\frac{-(\upsilon + \zeta)}{2(\upsilon\zeta - \chi\varphi)} + \frac{\big\{(\upsilon + \zeta)^2 - 4(\upsilon\zeta - \chi\varphi)(1 - |\lambda|^2/M^2)\big\}^{\frac{1}{2}}}{2(\upsilon\zeta - \chi\varphi)}\right)^{\frac{1}{2}}. \tag{8.17}$$

There are two possible cases if we employ this c: either λ_{c1} is a complex conjugate of λ_{c2} with magnitude M, or λ_{c1} and λ_{c2} are both real with product M. Note that the function $f : \mathbb{R} \to \mathbb{C}^2$; $c \mapsto (\lambda_{c1}, \lambda_{c2})$ is continuous and $\hat{A}_D$ is always real. Thus if the latter case occurs, there is another $c \in [0, c_c]$ at which the two eigenvalues equal each other. Such c can be found by setting the discriminant to 0 in Equation (8.16). Some more algebra shows that for $\lambda_{c1} = \lambda_{c2}$,

$$c = \tilde{c}_c := \sqrt{\frac{-\vartheta + (\vartheta^2 - 4\eta\xi)^{\frac{1}{2}}}{2\eta}}, \tag{8.18}$$

where

$$\begin{aligned} \eta &= \zeta^2\big(Re\{\lambda\}^2 - |\lambda|^2\big) + \chi\varphi|\lambda|^2, \\ \vartheta &= 2\zeta\big(Re\{\lambda\}^2 - |\lambda|^2\big), \\ \xi &= Re\{\lambda\}^2 - |\lambda|^2. \end{aligned}$$

In fact, it can be shown that the two perturbed eigenvalues will coincide for some $c > 0$ if and only if $\zeta^2\big(Re\{\lambda\}^2 - |\lambda|^2\big) + \chi\varphi|\lambda|^2 > 0$.

We can apply the theory developed above to modify unstable estimated models of systems which are known to be stable. The idea is to attenuate the magnitude of each of the unstable poles using the data augmentation scheme. For this purpose, the procedures are organized into the following algorithm.

Algorithm 8.7. *(Data Augmentation Method)*

1. *Choose an unstable eigenvalue λ of $\hat{A}$ and a corresponding normalized left eigenvector ν; or a pair of complex eigenvalues $\lambda, \bar{\lambda}$ and corresponding eigenvectors ν and $\bar{\nu}$. If the eigenvector is real, goto Step (2); else goto Step (3).*

2. *Set $c = c_r$ and augment Γ with $\gamma_i = c\nu$ and $\Gamma^{\uparrow}$ with $\gamma_o = 0$. Skip Step (3).*

3. *Set $c = c_c$ and append $\gamma_i = c\begin{bmatrix}\nu \\ \bar{\nu}\end{bmatrix}$ to Γ and $\gamma_o = 0$ to $\Gamma^{\uparrow}$. If the resulting $|\lambda_{c1}|$ or $|\lambda_{c2}| > M$, use $c = \tilde{c}_c$ instead of c_c.*

4. $\hat{A}_D = \Gamma_D^{\dagger}\Gamma_D^{\uparrow}$ *with* Γ_D *and* $\Gamma_D^{\uparrow}$ *defined as in* (8.1) *and* (8.2). *Stop if* $\hat{A}_D$ *is stable.*
5. *Replace* $\hat{A}$, Γ, *and* $\Gamma^{\uparrow}$ *by* $\hat{A}_D$, Γ_D, *and* $\Gamma_D^{\uparrow}$, *respectively. Repeat Step* (1).

For marginally stable systems, the magnitude of the perturbed eigenvalue(s) M can be set to 1. For stable systems, we refer to [McK95, p. 137] or the simulation example in Section 8.4.1 for further discussion.

We should mention some issues about numerical stability. Each time we calculate $\hat{A}_D$, it is better to evaluate it using the expression $(I + (\Gamma^*\Gamma)^{-1}\gamma_i^*\gamma_i)^{-1}\hat{A}$. In this way, $\gamma_i^*\gamma_i$ is suppressed by $(\Gamma^*\Gamma)^{-1}$ when c turns out to be large, even though this is unlikely to occur with data generated by a stable system. Consequently, we can evaluate the outer inverse in a more stable manner. Furthermore, the perturbed eigenvalues of the complex case should be evaluated using Equation (8.10) which is only of order 2. This will greatly reduce the effort of computing the perturbed eigenvalues when the estimated system has a large number of states.

One more fact about the complex conjugate case is that if the two eigenvalues coincide with each other, one may expect the corresponding root loci to split into two real trajectories. As a result, one eigenvalue may have an attenuation rate much faster than its counterpart. In other words, if we force both eigenvalues to a stable location, one eigenvalue may be in a small neighborhood of the origin, which may not be desired. So if this occurs, we can instead place the two eigenvalues at the position at which they coincide as a first step. Then, we can treat the two eigenvalues as two distinct real eigenvalues so as to relocate them separately onto the unit disc. This is indeed taken into consideration in Step (3) in the above algorithm.

To complete this section, it remains to consider the case with the extended controllability matrix Ω. This can be done by taking conjugate transposes everywhere.

8.3 The State Sequence Approach

As described in the introduction, some subspace algorithms estimate the system matrices by solving the problem

$$\min_{A,B,C,D} \left\| \begin{bmatrix} X^{\leftarrow} \\ Y \end{bmatrix} - \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} \right\|_F, \tag{8.19}$$

where

$$\begin{aligned} X &= \begin{bmatrix} x_1 & x_2 & \cdots & x_q \end{bmatrix}, &\text{and}\quad X^{\leftarrow} &= \begin{bmatrix} x_2^{\leftarrow} & \cdots & x_q^{\leftarrow} & x_{q+1}^{\leftarrow} \end{bmatrix}, \\ U &= \begin{bmatrix} u_1 & u_2 & \cdots & u_q \end{bmatrix}, &\text{and}\quad Y &= \begin{bmatrix} y_1 & y_2 & \cdots & y_q \end{bmatrix}. \end{aligned}$$

Note that x_τ is not necessarily equal to $x_\tau^{\leftarrow}$ for $\tau = 2, \ldots, q$. In fact, Equation (8.19) can be separated into two parallel least-squares problems

$$\min_{A,B} \left\| X^{\leftarrow} - \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} \right\|_F, \qquad \text{and} \qquad \min_{C,D} \left\| Y - \begin{bmatrix} C & D \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix} \right\|_F.$$

Again we shall assume X and U have full row rank. Accordingly, we can replace the Moore-Penrose inverse by the right inverse

$$\begin{bmatrix} X \\ U \end{bmatrix}^{\dagger} = \begin{bmatrix} X \\ U \end{bmatrix}^{*} \left(\begin{bmatrix} X \\ U \end{bmatrix} \begin{bmatrix} X \\ U \end{bmatrix}^{*} \right)^{-1}.$$

Note that if we impose a stability constraint on $\hat{A}$, the least-squares solutions for C and D are unchanged. This suggests that in this case it is sufficient to modify $\hat{A}$ and $\hat{B}$ so that $\hat{A}$ satisfies the constraint. Now, the key step at this point is to introduce a series of zeros of the same size as U below $X^{\leftarrow}$. That is,

$$Z := \begin{bmatrix} X \\ U \end{bmatrix} =: \begin{bmatrix} z_1 & z_2 & \cdots & z_q \end{bmatrix}, \qquad \text{and} \qquad Z^{\leftarrow} := \begin{bmatrix} X^{\leftarrow} \\ 0 \end{bmatrix} =: \begin{bmatrix} z_2^{\leftarrow} & \cdots & z_q^{\leftarrow} & z_{q+1}^{\leftarrow} \end{bmatrix},$$

with the obvious partition. Now, we have

$$\Psi := \begin{bmatrix} \hat{A} & \hat{B} \\ 0 & 0 \end{bmatrix} = Z^{\leftarrow} Z^{*} (Z Z^{*})^{-1}.$$

Remark 8.3. To show the connection between this section and the previous one, we recall the following simple fact. Define

$$\Psi := \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad \nu := \begin{bmatrix} \nu_1 \\ 0 \end{bmatrix},$$

where $\Psi \in \mathbb{R}^{(n+m)\times(n+m)}$, $\Psi_{11} \in \mathbb{R}^{n\times n}$, $\Psi_{12} \in \mathbb{R}^{n\times m}$, and $\nu \in \mathbb{C}^{(n+m)\times 1}$, $\nu \in \mathbb{C}^{n\times 1}$. Then the square matrix Ψ has n eigenvalues identical to those of Ψ_{11} and m additional zero eigenvalues. Furthermore, suppose ν_1 is a right eigenvector of Ψ, then ν is also a right eigenvector of Ψ, for any Ψ_{12}. △

8.3.1 Stability: Sufficient Condition

A result similar to Lemma 8.1 also holds for the state sequence approach. Suppose $R_0 := ZZ^*$ and $R_1 := Z^{\leftarrow}Z^*$. Then $\hat{A} = R_1 R_0^{-1}$ is stable (semistable) if S is positive definite (semidefinite), where

$$S := \begin{bmatrix} R_0 & R_1^* \\ R_1 & R_0 \end{bmatrix}.$$

As indicated in Theorems 8.2 and 8.3, if $\gamma_k = T\gamma_1$ or $\omega_k = \omega_1 T$, $\bar{\sigma}(T) \le 1$, the system estimate $\hat{A}$ is always semistable. However, we cannot extend these results to the state sequence method, i.e., $\hat{A}$ is not necessarily semistable if $x_{q+1} = tx_1$, $|t| < 1$. This is due to the fact that Z and $Z^{\leftarrow}$ are not shift invariant.

8.3.2 Augmenting the Data

Again, our objective is to produce a stable $\hat{A}$ and a corresponding $\hat{B}$ which also match the data well. So, we append one more term of compatible size to each of the series,

$$X_D = \begin{bmatrix} x_1 & \cdots & x_q & x_i \end{bmatrix}, \quad X_D^{\leftarrow} = \begin{bmatrix} x_2^{\leftarrow} & \cdots & x_{q+1}^{\leftarrow} & x_o \end{bmatrix}, \quad U_D = \begin{bmatrix} u_1 & \cdots & u_q & u_i \end{bmatrix},$$

and evaluate new estimates $\hat{A}_D$ and $\hat{B}_D$. Accordingly, we have

$$Z_D := \begin{bmatrix} X_D \\ U_D \end{bmatrix} =: \begin{bmatrix} z_1 & \cdots & z_q & z_i \end{bmatrix}, \quad Z_D^{\leftarrow} := \begin{bmatrix} X_D^{\leftarrow} \\ 0 \end{bmatrix} =: \begin{bmatrix} z_2^{\leftarrow} & \cdots & z_{q+1}^{\leftarrow} & z_o \end{bmatrix}. \tag{8.20}$$

Thus, the new $\hat{A}$, $\hat{B}$ estimates are obtained as the following:

$$\Psi_D = \begin{bmatrix} \hat{A}_D & \hat{B}_D \\ 0 & 0 \end{bmatrix} = Z_D^{\leftarrow} Z_D^* (Z_D Z_D^*)^{-1}.$$

Moreover, the difference of the two estimates Ψ_D and Ψ is given by

$$\Psi_D - \Psi = [z_o - \Psi z_i]\, z_i^* (ZZ^* + z_i z_i^*)^{-1} \tag{8.21}$$

which is similar to Equation (8.3).

The following theorem is an extension of Proposition 8.4. Because of the loss of the shift invariance property, Proposition 8.8 suggests a much larger c (as large as the largest singular value of the whole sequence $X^{\leftarrow}$) to guarantee semistability.

Proposition 8.8. *Suppose*

$$\hat{A}_D = \begin{bmatrix} Z^{\leftarrow} & 0 \end{bmatrix} \begin{bmatrix} Z & \begin{matrix} cQ \\ 0 \end{matrix} \end{bmatrix}^{\dagger},$$

and $c \geq \bar{\sigma}(X^{\leftarrow})$, where $Q \in \mathbb{C}^{n \times n}$ is unitary. Then, $\rho(\hat{A}_D) \leq 1$.

Proof. First let $T = \begin{bmatrix} 0 & 0 \\ (Q^* X^{\leftarrow})/c & 0 \end{bmatrix}$. Then the proof follows that of Proposition 8.4. □

8.3.3 Eigenvalue Perturbations

The results for eigenvalue perturbations in the state sequence case are analogous to those in the shift invariance case with obvious changes, i.e., $\Omega \leftrightarrow Z$ and $\Omega^{\leftarrow} \leftrightarrow Z^{\leftarrow}$. Therefore, we will briefly state the main point here and refer to the shift invariance case for details.

Suppose $\nu \in \mathbb{C}^{n\times 1}$ is a normalized right eigenvector of $\hat{A}$ corresponding to the eigenvalue λ. From Remark 8.3, we know that $\begin{bmatrix} \nu \\ 0 \end{bmatrix}$ is also an eigenvector of Ψ corresponding to λ. If $z_i = c\begin{bmatrix} \nu \\ 0 \end{bmatrix}$, and $z_o = b\begin{bmatrix} \hat{A}\nu \\ 0 \end{bmatrix}$, then the perturbed eigenvalue is

$$\lambda_D = \frac{1 + bc\begin{bmatrix} \nu^* & 0 \end{bmatrix}(ZZ^*)^{-1}\begin{bmatrix} \nu \\ 0 \end{bmatrix}}{1 + c^2\begin{bmatrix} \nu^* & 0 \end{bmatrix}(ZZ^*)^{-1}\begin{bmatrix} \nu \\ 0 \end{bmatrix}}\lambda. \tag{8.22}$$

On the other hand, suppose $z_i = c\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}$, $z_o = b\begin{bmatrix} \hat{A}\nu & \hat{A}\bar{\nu} \\ 0 & 0 \end{bmatrix}$, where $\nu, \bar{\nu} \in \mathbb{C}^{n\times 1}$ are normalized right eigenvectors of $\hat{A}$ corresponding to λ and $\bar{\lambda}$, respectively. Then the perturbed eigenvalues λ_{c1} and λ_{c2} are the eigenvalues of Λ_D,

$$\begin{aligned}\Lambda_D := &\left(I + c^2\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}^*(ZZ^*)^{-1}\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}\right)^{-1} \\ &\left(I + bc\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}^*(ZZ^*)^{-1}\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}\right)\begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}.\end{aligned} \tag{8.23}$$

8.3.4 Acquiring Stable Systems

Again, the data augmentation method in the state sequence case is equivalent to that in the shift invariance case with obvious modifications; for instance, replace $\hat{A}$ by Ψ and eigenvector ν by $\begin{bmatrix} \nu \\ 0 \end{bmatrix}$. Also, the analogous values of υ, χ, φ, ζ should be defined as

$$\begin{bmatrix} \upsilon & \chi \\ \varphi & \zeta \end{bmatrix} := \begin{bmatrix} \nu^* & 0 \\ \bar{\nu}^* & 0 \end{bmatrix}(ZZ^*)^{-1}\begin{bmatrix} \nu & \bar{\nu} \\ 0 & 0 \end{bmatrix}.$$

It is noteworthy that the data augmentation algorithm now adjusts $\hat{A}$ and $\hat{B}$ simultaneously. Moreover, $\hat{C}$ and $\hat{D}$ are kept unchanged during the modification.

For the decomposition methods and the scaling method, we still have to find the new $\hat{B}$ using the state sequence approach. One natural approach is to find the least-squares solutions using the corrected $\hat{A}_J$, $\hat{A}_U$ and $\hat{A}_S$ as

$$\hat{B}_* := \arg\min_{\hat{B}} \left\| X^{\leftarrow} - \hat{A}_* X - \hat{B}U \right\|_F,$$

where $\hat{A}_*$ can be $\hat{A}_J$, $\hat{A}_U$, $\hat{A}_S$. Finally, note also that the zero padding technique, namely setting $z_i = \begin{bmatrix} x_{q+1} \\ 0 \end{bmatrix}$ and $z_o = 0$ in Equation (8.20), does not guarantee semistability in the state sequence case, in which the shift invariance property does not hold in general.

8.4 Simulation Examples

In this section, we provide in two simulation examples some quantitative comparison of the five stability enforcement methods, namely

1. the data augmentation method ($\hat{A}_D$) – see Section 8.2.4, 8.3.4;
2. the Schur decomposition method ($\hat{A}_U$) – see Section 8.2.3, 8.3.4 and [McK95];
3. the Jordan decomposition method ($\hat{A}_J$) – see Section 8.2.3, 8.3.4;
4. the scaling method ($\hat{A}_S$) – see Section 8.2.3, 8.3.4;
5. the zero padding method ($\hat{A}_Z$) – see Section 8.2.3 and [Mac95].

In the first example (Section 8.4.1), stable approximate realizations are determined from initial impulse responses, through the joint application of stability enforcement and Kung's algorithm (an approximate realization algorithm). In the second example (Section 8.4.2), marginal stability is enforced on state-space models obtained by N4SID algorithm (a subspace identification algorithm) from input-output observations. The performance of the stability enforcement methods is then assessed in these simulations.

8.4.1 Approximate Realizations of a Stable Process Control Model

In this example, we extract 4 impulse responses and construct the corresponding Markov parameters from 2 particular input and 2 particular output channels of a stable process control system. In the study, we suppose that only the first 64 Markov parameters $h_0, h_1, \ldots, h_{63}$ are available for approximating state-space realizations. From Figures 8.1 and 8.2, it can be seen that these 64 Markov parameters provide an excellent description of the model dynamics, except that the response of the $u2$-$y2$ channel is truncated shortly before it decays to zero. Moreover, the full model has a Hankel norm of 55.542 and the block Hankel matrix H constructed by the 63 Markov parameters,

$$H := \begin{bmatrix} h_1 & \cdots & h_{32} \\ \iddots & \iddots & \iddots \\ h_{32} & \cdots & h_{63} \end{bmatrix},$$

has a largest singular value $\bar{\sigma}(H) = 53.697$.

State-space models of order 8 and 12 are then approximated by applying Kung's algorithm to this block Hankel matrix H. However, both the approximate realizations turn out to be unstable, with the 8-th order estimate having a pole at 1.0089 and the 12-th order estimate having a pair of poles at $1.0003 \pm 0.0157i$. Denote respectively by $\hat{H}_8$ and $\hat{H}_{12}$ the block Hankel

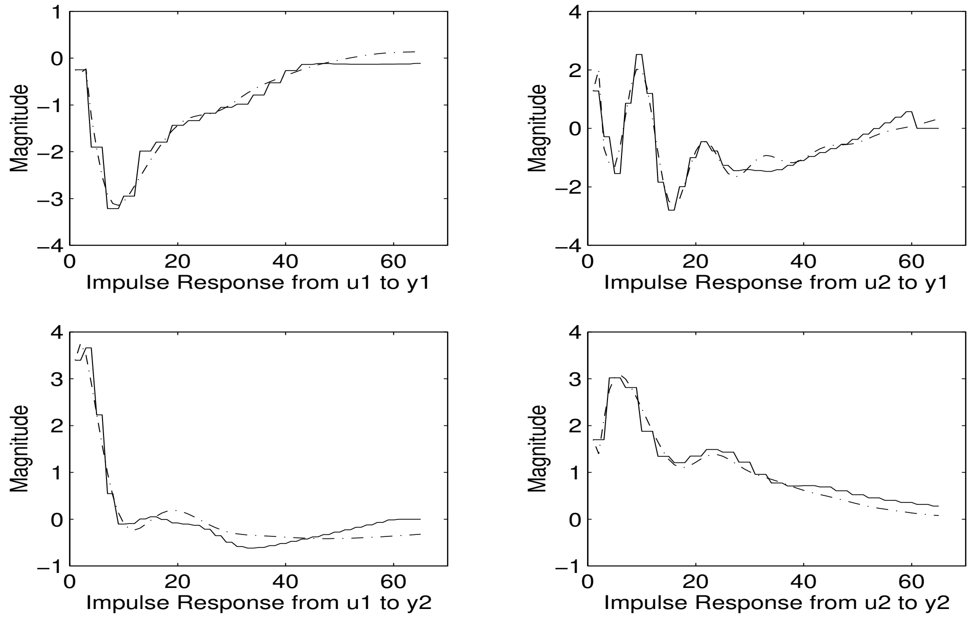

Figure 8.1 *Impulse responses of the original model (solid) and the* 8*-th order Kung's estimate (dashdot).*

matrices constructed from the 63 Markov parameters of the 8-th order model and the 12-th order model. Accordingly, we have

$$\bar{\sigma}(H - \hat{H}_8) = 5.805, \qquad \text{and} \qquad \bar{\sigma}(H - \hat{H}_{12}) = 4.320,$$

from which we can see that the 12-th order model gives a better approximation to the original model. Furthermore, their impulse responses are depicted in Figures 8.1 and 8.2 respectively.

Five stability enforcement methods are then employed together with Kung's algorithm to obtain stable estimates. The corresponding block Hankel matrices are then constructed using each of the system matrices. Their performance is measured by the 2-norm of the difference of

Errors		(1) $\hat{H}_{8,D}$	(2) $\hat{H}_{8,U}$	(3) $\hat{H}_{8,J}$	(4) $\hat{H}_{8,S}$	(5) $\hat{H}_{8,Z}$
H	$M = 0.99$	9.008	11.110	11.160	19.748	15.542
	$M = 0.995$	8.202	9.974	9.720	16.121	15.542
	$M = 0.999$	7.503	8.895	8.448	12.910	15.542
$\hat{H}_8$	$M = 0.99$	4.948	6.548	8.982	17.279	12.174
	$M = 0.995$	3.892	5.308	7.212	13.524	12.174
	$M = 0.999$	2.931	4.100	5.525	10.139	12.174

Table 8.1 *Errors of the* 8*-th order stable approximations, measured as* $\bar{\sigma}(H - \hat{H}_{8,\cdot})$ *and* $\bar{\sigma}(\hat{H}_8 - \hat{H}_{8,\cdot})$.

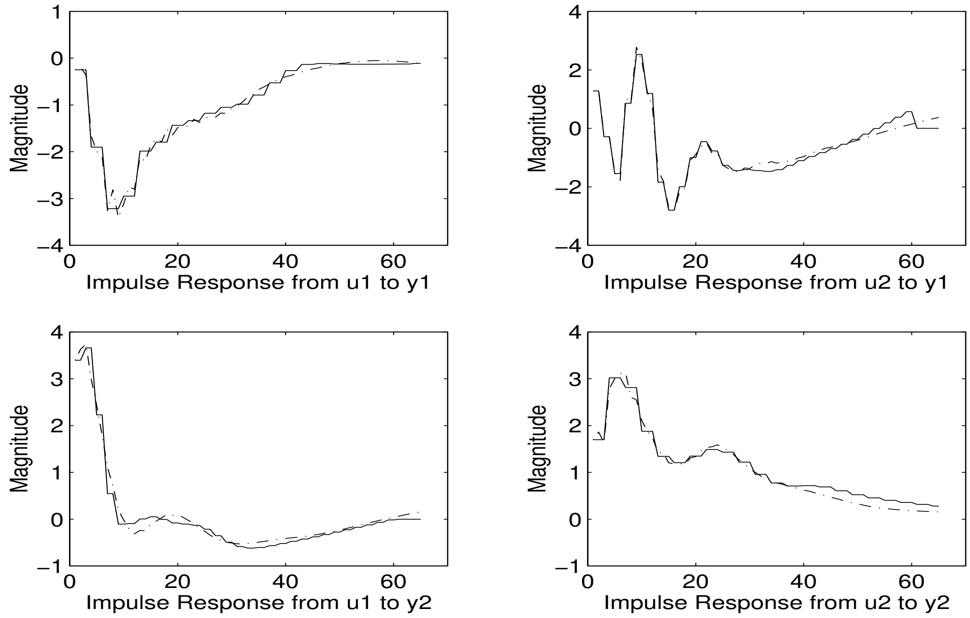

Figure 8.2 *Impulse responses of the original model (solid) and the* 12*-th order Kung's estimate (dashdot).*

the block Hankel matrices, namely

$$\bar{\sigma}(H - \hat{H}_{8,\cdot}) \qquad \text{and} \qquad \bar{\sigma}(\hat{H}_8 - \hat{H}_{8,\cdot}),$$

where $\hat{H}_{8,\cdot}$ can be either $\hat{H}_{8,D}$ (by the data augmentation method), $\hat{H}_{8,U}$ (by the Schur decomposition method), $\hat{H}_{8,J}$ (by the Jordan decomposition method), $\hat{H}_{8,S}$ (by the scaling method), or $\hat{H}_{8,Z}$ (by the zero padding method).

As proposed by McKelvey [McK95, p. 137], one may place the unstable pole λ_i to a stable position at $(2\,|\lambda_i|^{-1}{-}1)\,\lambda_i$. Thus, for the 8-th order estimate one may set $M = 0.9911$ (i.e.

Errors		(1) $\hat{H}_{12,D}$	(2) $\hat{H}_{12,U}$	(3) $\hat{H}_{12,J}$	(4) $\hat{H}_{12,S}$	(5) $\hat{H}_{12,Z}$
H	$M = 0.995$	6.045	6.686	6.578	8.267	18.487
	$M = 0.999$	4.724	4.857	4.812	4.978	18.487
	$M = 0.9995$	4.572	4.652	4.624	4.705	18.487
$\hat{H}_{12}$	$M = 0.995$	2.421	3.287	3.668	6.008	15.708
	$M = 0.999$	0.660	0.908	1.006	1.626	15.708
	$M = 0.9995$	0.423	0.584	0.646	1.043	15.708

Table 8.2 *Errors of the* 12*-th order stable approximations, measured as* $\bar{\sigma}(H - \hat{H}_{12,\cdot})$ *and* $\bar{\sigma}(\hat{H}_{12} - \hat{H}_{12,\cdot})$.

$M = 2-1.0089$). To illustrate the effect of M, we place the unstable pole to several stabilizing positions, with $M = 0.99$, 0.995 and 0.999 instead. Simulation results are depicted in Table 8.1. On the other hand, one may set $M = 0.9996$ (i.e. $M = 2-1.0004$) for the 12-th order estimate according to McKelvey's scheme. Again, we test a few stabilizing positions, with $M = 0.995$, 0.999, and 0.9995. Simulation results are depicted in Table 8.2.

In this example, it can be seen from these tables that the data augmentation method gives a better stable approximation than the other methods, though the Schur and Jordan decomposition methods do realize good estimates. In addition, it can be seen that, except for the zero padding method which does not depend on M, the closer we place the poles to the unit circle, the better are the stable estimates. Such a phenomenon is reasonable since the closer those poles are to the unit circle, the closer they will be to the unstable least-squares estimates obtained in the first place.

8.4.2 Identification of a Marginally Stable State-Space Model

A system of order 6 is generated randomly with 2 inputs and 2 outputs. The system is then rescaled so that a marginally stable system is acquired with a complex-conjugate pair of poles on the unit circle. The eigenvalues are $0.9289 \pm 0.3704j$, $-0.4106 \pm 0.1727j$, and $0.0713 \pm 0.2702j$. A sequence of output data, of length 100, is generated by a random input signal and a random initial state value. To simulate an inconsistent data set, white noises of variance 0.1 are introduced through the states and the measurements. A model is then determined using the N4SID subspace method [VODM96b] with 6 Hankel blocks. 1000 unstable estimates $\hat{A}$ are generated according to the scheme described above, each of which is restored by each of the proposed methods.

Two aspects of the performance are measured: the least-squares fitting, and the simulation errors. The least-squares fitting is assessed as follows. Suppose $E_1 = \|\hat{A}\Gamma - \Gamma^{\leftarrow}\|_F$, or $E_2 = \|\hat{A}X + \hat{B}U - X^{\leftarrow}\|_F$. We then measure the percent error as:

$$\%\ \text{Error} = \frac{\left\|\hat{A}_*\Omega - \Omega^{\leftarrow}\right\|_F - E_1}{E_1} \times 100$$

for the shift invariance case, and

$$\%\ \text{Error} = \frac{\left\|\hat{A}_*X + \hat{B}_*U - X^{\leftarrow}\right\|_F - E_2}{E_2} \times 100$$

for the state sequence case, where $\hat{A}_*$ and $\hat{B}_*$ are the modified $\hat{A}$ and $\hat{B}$ using one of the corresponding methods.

For the second assessment, the simulation errors for validation data are measured. Validation data are constructed with random input signals and random initial states in a noise free

Method \ Result	Least-Sq. fitting		Sim. Err. $N=25$		Sim. Err. $N=300$	
	Avg.	Best in	Avg.	Best in	Avg.	Best in
(1) $\hat{A}_D$	0.38%	1000	28.4268%	405	8.1322%	457
(2) $\hat{A}_U$	0.63%	0	28.4059%	275	8.1360%	229
(3) $\hat{A}_J$	0.63%	0	28.4050%	85	8.1349%	95
(4) $\hat{A}_S$	6.77%	0	28.5170%	235	8.5625%	219

Table 8.3 *Least-squares fitting errors and simulation errors using the shift invariance approach.*

environment in each of the 1000 trials. The error is then measured as:

$$\% \text{ Error} = \frac{1}{p}\sum_{i=1}^{p}\sqrt{\frac{\sum_{j=0}^{N-1}(y_i(j)-\hat{y}_i(j))^2}{\sum_{j=0}^{N-1}(y_i(j))^2}} \times 100,$$

where y_i denotes the i-th channel of the output and $\hat{y}$ denotes the simulation output. Validation data with length $N = 25$ and 300 are tested, with the case $N = 25$ emphasizing the effect on the rapidly decaying modes, and $N = 300$ emphasizing the effect on the marginally stable modes. Finally, two statistics of these errors are provided, which are the average and the number of trials in which an algorithm performed better than the alternative algorithms.

Simulation results are summarized in Tables 8.3 and 8.4. From these tables, it can be seen that the data augmentation method gives a better least-squares fitting and smaller simulation errors more often, yet the difference is less distinct in the latter measure. In addition, it is noteworthy that although it is not good in least-squares fitting, the scaling method can sometimes produce good models in terms of simulation errors. Such situations arise if the original unstable estimate is very close to being stable.

Method \ Result	Least-Sq. fitting		Sim. Err. $N=25$		Sim. Err. $N=300$	
	Avg.	Best in	Avg.	Best in	Avg.	Best in
(1) $\hat{A}_D$	0.057%	995	28.9176%	478	8.5514%	494
(2) $\hat{A}_U$	0.064%	3	28.9275%	115	8.5616%	93
(3) $\hat{A}_J$	0.066%	2	28.9229%	148	8.5600%	185
(4) $\hat{A}_S$	1.010%	0	28.8983%	259	9.0416%	228

Table 8.4 *Least-squares fitting errors and simulation errors using the state sequence approach.*

8.5 Summary

In this chapter, we have extended the stability analysis on approximation realization algorithms and subspace algorithms carried out in [Mac95, McK95]. Several methods of enhancing these algorithms for finding stable approximations to least-squares problems are examined.

As seen in the simulation examples, it is suggested that the data augmentation method provides better stable approximates more often, particularly in the least-squares sense. The decomposition methods give a fast algorithm and good estimates, though the Jordan decomposition method is not numerically stable. In addition, the scaling method can also produce good models in some cases. However, if a bigger or more unstable system is under consideration, we may expect a larger discrepancy due to the scaling of other stable modes. The zero padding method tends to attenuate the tail of the system response; yet it is very simple, and coincides with a procedure commonly used in autoregressive modeling.

Of course we do not advocate the application of these algorithms if the data indicate that the assumption of system stability may be false. They are intended for applications in which the true system is known to be stable, and in which assuring stability of the model is important. As remarked in [Mac95], in some applications it would be sensible to apply a standard subspace method first, and only use our modifications if the standard method yielded an unstable model.

9

Case Studies

9.1 Introduction

In this chapter, various aspects regarding the performance of the subspace algorithms presented in Chapter 7 are investigated in two case studies. These case studies are based on real data extracted from the following industrial systems.

- The first example, presented in Section 9.2, is based on a simulation model of a fractional distillation column. This model, which has been investigated in [CM97], is provided by SAST Ltd. and is available at DAISY database [DM97].

- The second example, presented in Section 9.3, is based on a set of process measurements recorded from an industrial dryer. This data set, which has also been investigated in [CM97], is provided by Cambridge Control Ltd. and can be obtained from DAISY database [DM97].

In the subspace identification literature, the $\mathcal{H}_\infty$ and $\mathcal{H}_2$ model errors seem not to be used commonly as a performance indicator, due to the unavailability of a perfect original model. Nevertheless, in the first example experimental data are set up in a rather ideal case in which $\mathcal{H}_\infty$ and $\mathcal{H}_2$ errors are investigated. One purpose of such a study is to reveal how well the algorithms perform in extracting system dynamics within a certain model order. Then, in the second example input-output observations drawn from a real industrial process are used for identification. For a more thorough examination, performance in this example is measured in a more classical fashion, namely in terms of simulation and prediction accuracy. Furthermore, some concluding remarks obtained in these studies are provided at the end of this chapter.

As it is difficult to judge whether the performance of the new proposed algorithms is satisfactory, throughout the case studies a rather well-known subspace identification algorithm, N4SID Robust Combined Algorithm [VODM96b, p. 131], is used as a reference method, whose performance is provided for comparison with these algorithms. Note that this reference algorithm has been encoded as a MATLAB m-file 'subid.m', which can be found in [VODM96b].

9.2 A Distillation Column

A very detailed simulation model of a fractional distillation column is studied in this section. This simulation model, which is made up of more than a thousand non-linear partial differential equations, can be run at or faster than real time. The whole model is composed of 3 input channels and 2 output channels, where the input channels control

- the input cooling temperature – $u1$;
- the reboiling temperature – $u2$;
- the pressure – $u3$;

and the output channels measure

- the top product flow rate – $y1$;
- the C4 concentration – $y2$.

After allowing the model to reach a steady operating point, six step responses are determined by applying a unit step to each of the input channels, while keeping the other two input channels at zero. The responses from the output channels are then sampled at a rate of once every second for a period of 2500 seconds, giving a total of 2501 samples.

Pretreatment of Data

As the step responses are generated from a simulation model in a noise free environment, no pre-filtering of the data is needed to be done. However, since the system is running at an operating point different from zero and hence introducing some D.C. offsets, subtraction is taken from these responses to remove these offsets. Furthermore, it is seen from a closer inspection that the model reacts to the step excitation at a rate much slower than the sampling time; in which case, a lower sampling rate can in fact be used. To strive for a balance between the accuracy and the simplicity of the model, a sampling rate of once every thirty seconds is chosen, leaving a total of 84 samples. Moreover, these step responses are converted to impulse responses by taking the successive differences.

As pointed out in [Cho94], it is important to make sure that the scales of the input-output channels are of comparable sizes, such that all input-output channels are weighted equally. A good indicator of their sizes is the energy contents in each of the input-output channels, where each channel energy content is calculated as the sum of squares of all data points of the corresponding impulse response. Thus, one final step of preprocessing the responses is to rescale the input-output channels according to the scheme developed in [Cho94]. These final step responses, depicted in Figure 9.1, will be referred to as the original model and are ready for a direct application of the proposed identification algorithms.

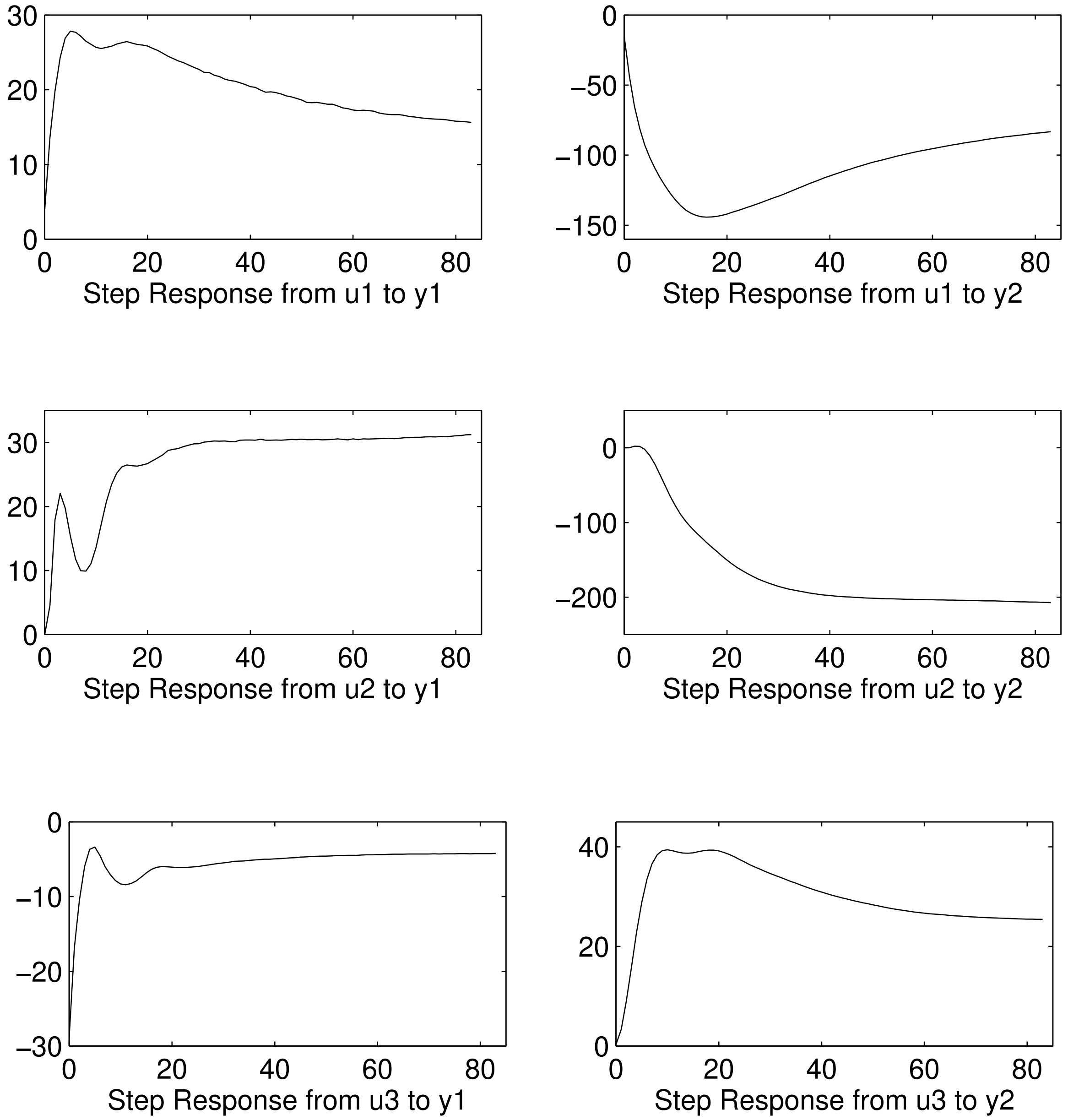

Figure 9.1 *Scaled step responses of the distillation column simulation model. Unit time interval = 30 sec.*

Order	Algorithm 7.8		Algorithm 7.9		Algorithm 7.10		N4SID Algorithm	
	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err
4	39.2	16.96	26.4	13.84	**21.8**	**13.65**	35.0	13.86
6	7.9	3.44	8.1	3.27	**7.9**	**3.25**	12.3	3.54
8	7.8	2.61	**7.3**	**2.61**	7.6	2.62	11.1	2.98
10	**8.2**	**2.25**	9.8	2.34	8.5	2.26	14.7	2.80
12	6.6	1.84	**6.3**	**1.83**	6.8	1.86	7.8	1.95

Table 9.1 *Accuracy of distillation column models in the noise free case.*

Experimental Data Sets

Simulation data used in this case study are generated from an equivalent balanced state-space model. To compute this balanced state-space model, we first construct the initial Markov parameter sequence from the scaled impulse responses. These Markov parameters, excluding the leading one h_0, are then packed into an upper triangular matrix in a Hankel form, leaving the lower triangular entries zero. With this block Hankel matrix, a perfect balanced state-space model is then obtained by Kung's algorithm.

Finally, two sets of simulation data, each of length 1000, are generated from this state-space model, with an initial condition generated by normal random variables and an input signal generated by uniform random variables distributed in $[0, 1]$. Furthermore, one set of simulation data is left as a noise free data set; whereas the other is corrupted by white process and output noises of variance 0.15. The purpose of introducing process and output noises will be clear when the selection of the truncation index is discussed below.

Performance Measurement

The proposed identification algorithms, Algorithms 7.8–7.10, are applied to these two sets of simulation data in order to find low order state-space models. To compare the performance of each algorithm, we measure the distance of the identified model to the original model in terms of two well-known system norms, the $\mathcal{H}_\infty$ norm and the $\mathcal{H}_2$ norm, as defined in Chapter 2. Note that the original model has a $\mathcal{H}_\infty$ norm of 247.6 and a $\mathcal{H}_2$ norm of 75.63. The performance of the proposed algorithms and N4SID algorithm is given in Table 9.1 for the noise free case and Table 9.2 for the noise corrupted case. Note that algorithms with the smallest $\mathcal{H}_\infty$ or $\mathcal{H}_2$ error are highlighted for each model order.

Figures 9.2 and 9.3 show the frequency responses of the original model and the 8-th order identified models in the noise free and noise corrupted case respectively. For a clear view, only models produced by Algorithm 7.10 and N4SID algorithm are displayed for comparison. On the other hand, the errors of the identified models, measured as the maximum singular value of the difference from the original model, are plotted for each frequency in Figures 9.4 and 9.5.

Order	Algorithm 7.8		Algorithm 7.9		Algorithm 7.10		N4SID Algorithm	
	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err	$\mathcal{H}_\infty$ Err	$\mathcal{H}_2$ Err
4	42.6	16.80	32.3	14.14	**28.7**	**13.92**	36.1	14.13
6	**20.1**	**5.05**	42.7	7.03	26.8	5.86	23.6	5.31
8	17.2	**4.15**	14.7	4.31	**14.6**	4.29	31.2	5.38
10	13.9	4.08	**13.6**	**3.96**	14.7	3.98	83.7	11.18
12	21.1	4.49	22.9	4.42	**17.7**	**3.96**	86.1	9.79

Table 9.2 *Accuracy of distillation column models in the noise corrupted case.*

From these figures, it is easy to verify the $\mathcal{H}_\infty$ errors calculated earlier. At this point, some observations can be drawn from the tables and figures as follows:

- the proposed algorithms seems to be less sensitive to disturbance and provide better models in most orders tested.
- though none of the proposed algorithms seems to be more superior than the others, Algorithm 7.10 seems to be more robust against the choice of the model order.

As for the computational complexity, Algorithms 7.8–7.10 consume about $135M$, $134M$, and $156M$ floating point operations respectively for the 8-th order models, while N4SID algorithm requires about $142M$ floating point operations. Finally, note that the truncation index k for Algorithms 7.8–7.10 is chosen to be 15; whereas the truncation index k for N4SID algorithm is chosen to be 23. The rationale for choosing these values is explained in the following.

Determination of the Truncation Index

One major difficulty in using subspace based methods is to find a suitable truncation index k. As the model order is determined from some matrix closely related to $\mathcal{O}_k X_f$, increasing k will provide more information about $\mathcal{O}_k$, or equivalently the dynamics of the pair (A, C), which is certainly of importance when principal components are to be extracted from this pair. Nevertheless, an increase in k obviously decreases the number of columns q in X_f, creating two shortcomings. First of all, since X_f intrinsically contains information about the controllability pair (A, B), decreasing q will reduce the information relating to the pair (A, B), giving a similar problem mentioned above. On the other hand, these columns in X_f actually have an averaging effect on the data in each row, which has a capability of eliminating small spikes in the observations, either due to the nature of the dynamics or measurement noises. Therefore, if q is small, these spikes will overwhelm the fundamental behavior of the system, producing less accurate models as a result. Further analyses of k against the accuracy of pole estimation have been made in [WJ94, JW95, JW96], while some rigorous statistical treatments on this issue can be found in [DPS95, PSD96].

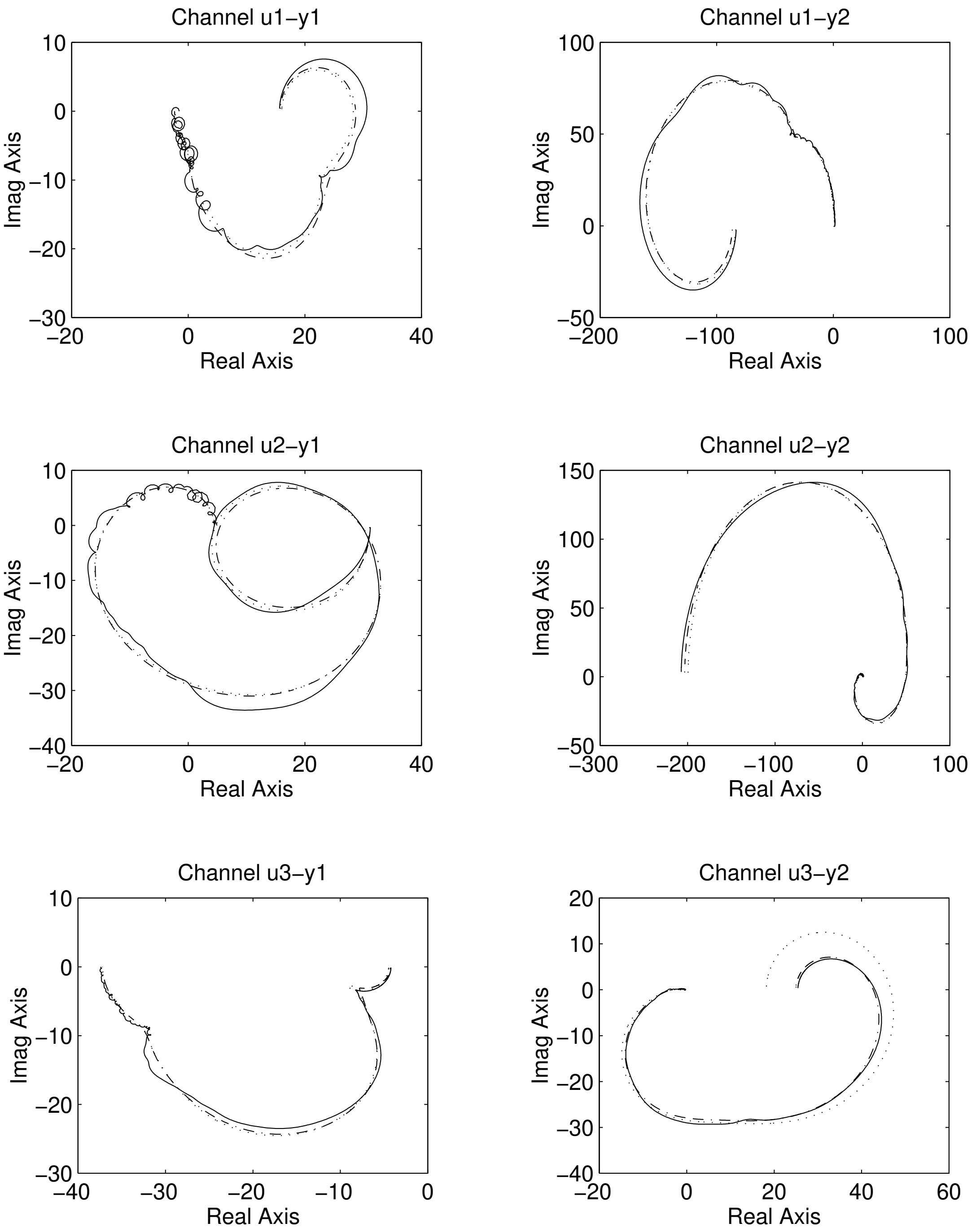

Figure 9.2 *Comparison of frequency responses of alternative* 8*-state models in the noise free case: original model (solid); model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted).*

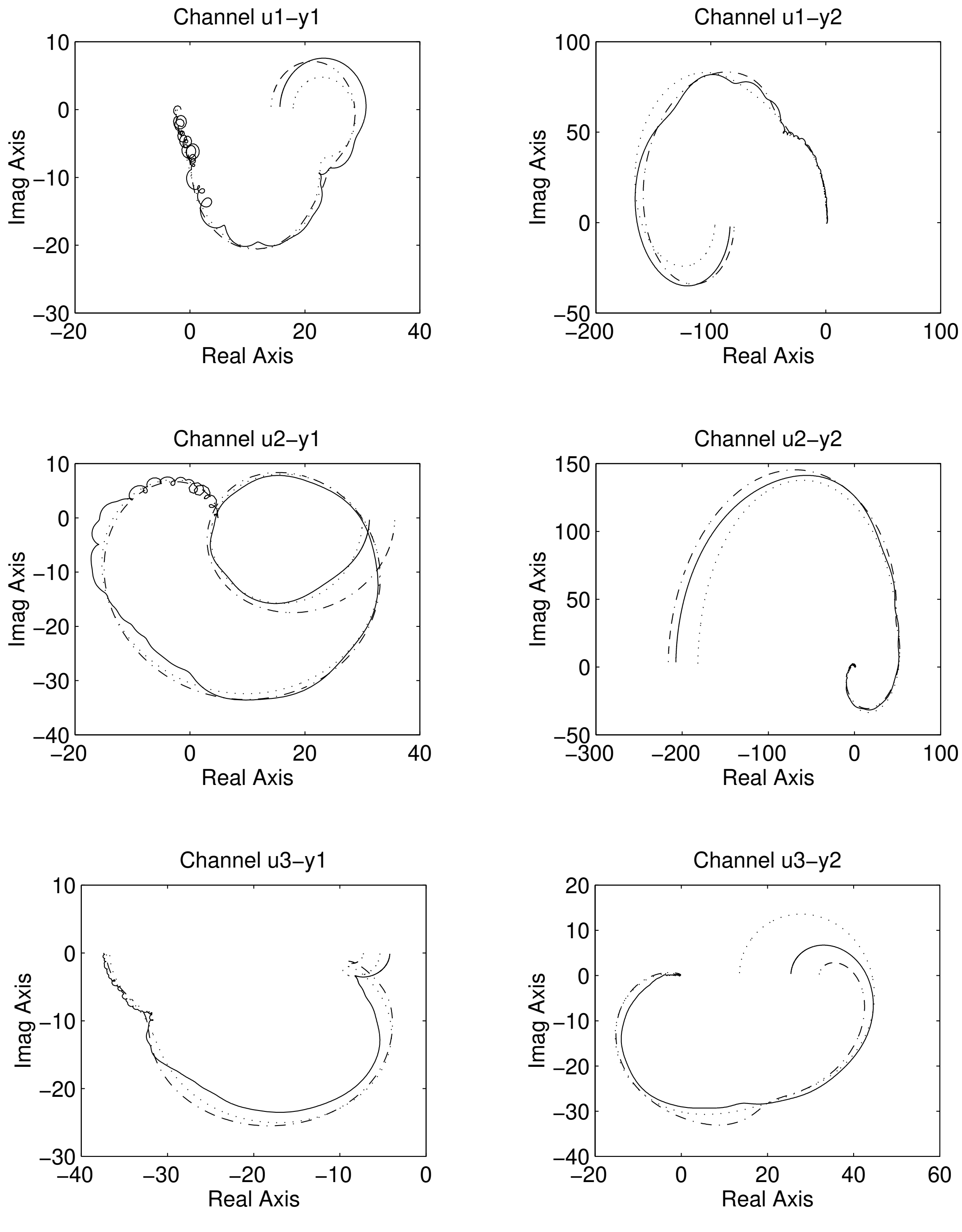

Figure 9.3 *Comparison of frequency responses of alternative 8-state models in the noise corrupted case: original model (solid); model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted).*

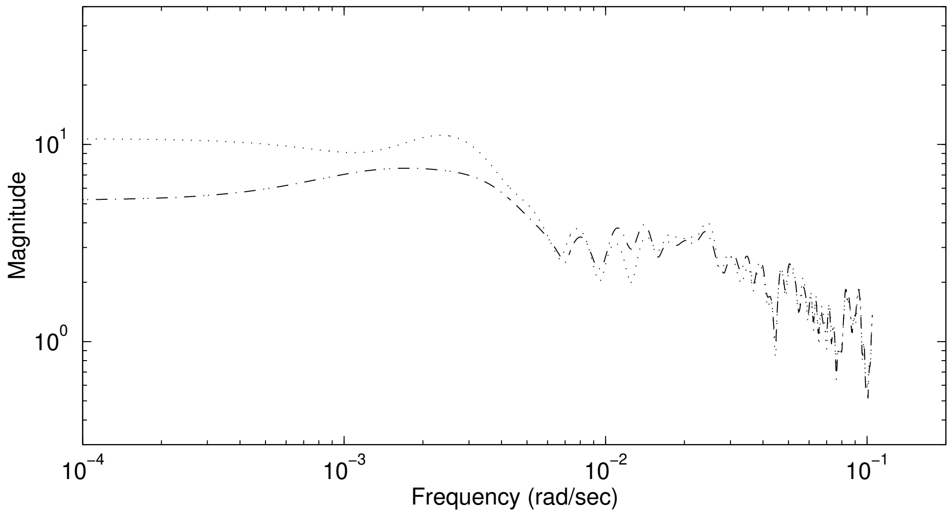

Figure 9.4 *The maximum singular value of the difference from the original model in the noise free case: model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted).*

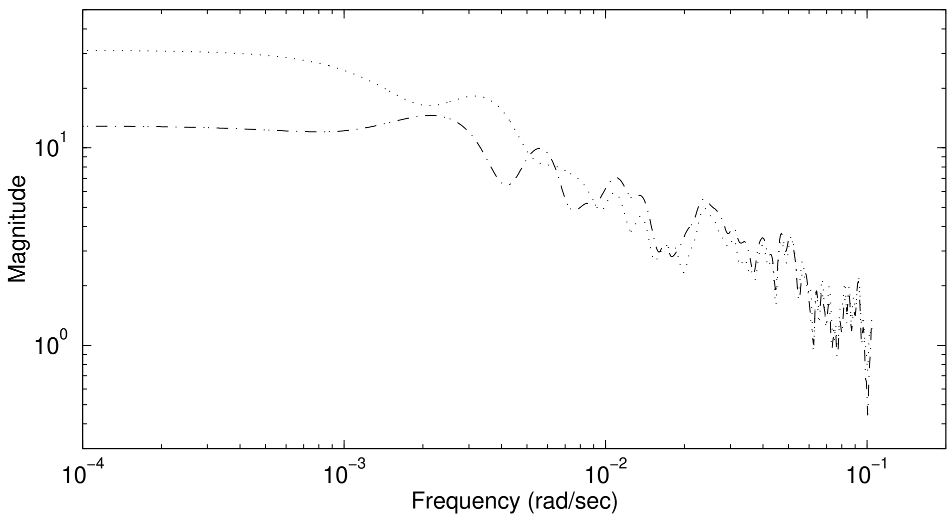

Figure 9.5 *The maximum singular value of the difference from the original model in the noise corrupted case: model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted).*

Algorithms 7.8–7.10 are applied to two sets of simulation data – the noise free set and the noise corrupted set – to find models of various orders. From the modeling of the noise free set, it is seen that as we increase k within a certain range, a better model of a fixed order is procured for each time, in terms of both the $\mathcal{H}_\infty$ norm and the $\mathcal{H}_2$ norm, though the improvement seems to diminish as k reaches a certain value. On the other hand, from the modeling of the corrupted data set it is observed that identified models of a fixed order actually start to degrade as k reaches some value within this certain range. Such degradation is more obvious in higher order models for N4SID algorithm as it can be seen from Tables 9.1 and 9.2 that higher order models obtained by this method degrade more rapidly. In this example, the values k are in fact chosen by trial and error in such a way that these values give a good balance between the model accuracy and the disturbance sensitivity.

9.3 An Industrial Dryer

An industrial drying process with three input channels and three outputs channels is considered in this example. Raw material enters the dryer continuously, and moisture is removed by passing hot gas over it. The efficiency of the process depends on both the temperature and the rate of flow of the hot gas. The input channels control

- the fuel flow rate – $u1$;
- the hot gas exhaust fan speed – $u2$;
- the rate of flow of raw material – $u3$;

and the output channels measure

- the dry bulb temperature – $y1$;
- the wet bulb temperature – $y2$;
- the moisture content of the raw material when it leaves the dryer – $y3$.

The data set used for identification in this example is taken on a real process. The first two inputs which could be manipulated are excited by pseudo-random binary signals. The third input, on the other hand, is an external disturbance and could not be manipulated, yet could be measured. Values of the input and output signals are recorded at a sampling rate of once every second to obtain 11665 samples.

Pretreatment of Data

For a better identification result, the raw data set is pre-processed beforehand. This includes discarding the first 3000 data points so as to ensure that the process has reached a steady

Order	Algorithm 7.8		Algorithm 7.9		Algorithm 7.10		N4SID Algorithm	
	Sim Err	Pre Err	Sim Err	Pre Err	Sim Err	Pre Err	Sim Err	Pre Err
2	35.26	22.18	39.30	19.77	**31.90**	16.55	32.72	**15.19**
4	31.69	10.36	32.40	9.90	**30.66**	**9.12**	33.33	9.26
6	33.96	9.22	32.21	8.86	**30.72**	8.18	31.17	**8.11**
8	48.09	8.58	31.10	8.03	**29.72**	**7.60**	31.83	7.68
10	36.77	8.00	33.62	7.62	**30.05**	**7.73**	—	9.22

Table 9.3 *Simulation errors and predication errors for the identification data set of an industrial dryer.*

operating point, and subtraction by the mean of the samples in order to remove any D.C. offsets due to the non-zero operating point. Finally, the input and output data are filtered through an anti-aliasing filter of an appropriate cut-off frequency; and the filtered data are sampled at a rate of once every 10 samples, or equivalently, once every 10 seconds. As a result, a data set of 867 data points is obtained and depicted in Figure 9.6. As in [CM97], we further divide the data set into two subsets, where the first one containing the first 600 data points is used for identification while the second one containing the rest is used for validation.

Performance Measurement

Algorithms 7.8–7.10 are applied to the identification data set to obtain state-space models of various orders. A commonly used performance indicator, which measures the discrepancy between the original output and the model output as

$$\%\ \text{Error} = \frac{1}{p}\sum_{i=1}^{p}\sqrt{\frac{\sum_{j=0}^{N-1}\left(y_i(j)-\hat{y}_i(j)\right)^2}{\sum_{j=0}^{N-1}\left(y_i(j)\right)^2}}\times 100,$$

is used to judge the fitness of each model produced by each algorithm in this example. Here, $y_i(t)$ is the i-th channel of the original output, N is the number of the data points in the data set, and p is the number of output channels, which is 3 in this case. On the other hand, $\hat{y}_i(t)$ represents either the simulated i-th output generated by the deterministic part of the model, with which simulation errors are measured, or the one-step ahead predicted i-th output, with which prediction errors are measured. As the data set is divided into the identification and validation data sets, simulation errors and prediction errors can be evaluated for each of these sets, producing four performance indices. These indices are given in Table 9.3 for the identification data set and Table 9.4 for the validation data set. Note that an entry with "—" in the tables indicates that the identified model is unstable, thus giving a large error. Moreover, algorithms with best performance in each category are highlighted for easier comparison.

To illustrate the effect of Kalman filters, Figure 9.7 shows the simulation outputs (generated by the deterministic part of the models) and the prediction outputs (generated by the full models

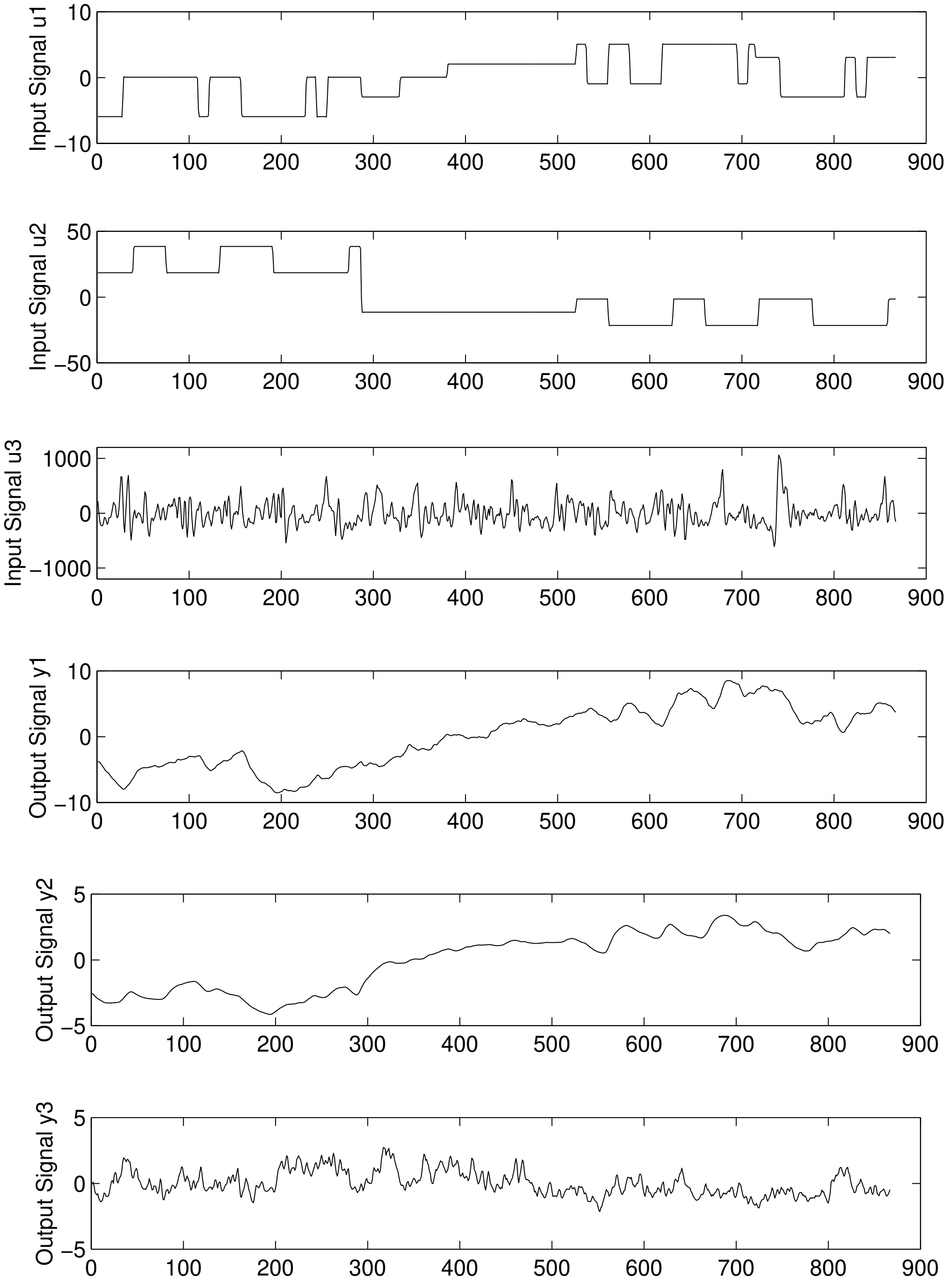

Figure 9.6 *Input and output signals of an industrial dryer. Unit time interval = 10 sec.*

Order	Algorithm 7.8		Algorithm 7.9		Algorithm 7.10		N4SID Algorithm	
	Sim Err	Pre Err	Sim Err	Pre Err	Sim Err	Pre Err	Sim Err	Pre Err
2	59.20	29.59	**57.02**	25.19	59.72	21.17	58.72	**18.52**
4	61.56	12.25	64.13	11.92	**60.00**	**11.01**	64.84	11.82
6	68.23	11.24	61.20	10.35	**60.20**	9.53	61.19	**9.27**
8	88.99	11.11	58.19	9.12	58.22	**8.81**	**56.89**	9.07
10	65.09	9.44	**53.58**	8.22	59.05	**8.13**	—	10.15

Table 9.4 *Simulation errors and predication errors for the validation data set of an industrial dryer.*

with Kalman filters) for the identification data set. From this figure, it is clear that models with Kalman filters can produce much better outputs. To avoid confusion, only models produced by Algorithm 7.10 and N4SID algorithm are displayed for comparison. On the other hand, Figure 9.8 shows simulation outputs and cumulative simulation errors for the validation data set, where the cumulative error at time t is the error computed up to time t:

$$\sum_{j=0}^{t-1} (y_i(j) - \hat{y}_i(j))^2.$$

The reason for plotting these cumulative errors is that it is difficult to distinguish between the performance of both algorithms just by inspecting the simulation outputs. Here, it can be seen that the model produced by Algorithm 7.10 is slightly better than that produced by N4SID algorithm. Finally, prediction outputs and cumulative prediction errors for the validation data set are depicted in Figure 9.9. This time, Algorithm 7.10 gives a better fitting to Outputs $y1$ and $y2$, whereas N4SID algorithm gives a better fitting to Output $y3$. At this point, some observations can also be drawn from the tables and figures as follows:

- Algorithm 7.10 seems to produce a better model in terms of simulation and prediction errors, both in identification and validation, and more distinctly in terms of simulation errors for the identification data set.
- models obtained by Algorithms 7.8–7.10 improve, in terms of prediction errors, each time the model order increases within an order of 10.
- Algorithm 7.8 does not seem to be as robust as the others, as it is not consistent in producing good estimates of various orders.

As for the computational complexity, Algorithms 7.8–7.10 consume about $12.3M$, $9.0M$, and $4.3M$ floating point operations respectively for the 6-th order models, while N4SID algorithm uses about $6.0M$ floating point operations for the same order. The number of floating point operations in fact depends heavily on the truncation index k used in the algorithms. For Algorithms 7.8–7.10, the truncation indices are set to 5, 4, and 3 respectively whereas for N4SID algorithm, this index is set to 5. To this end, it is noteworthy that similar remarks, regarding the choice of the truncation index, discussed in the previous section also apply to this case.

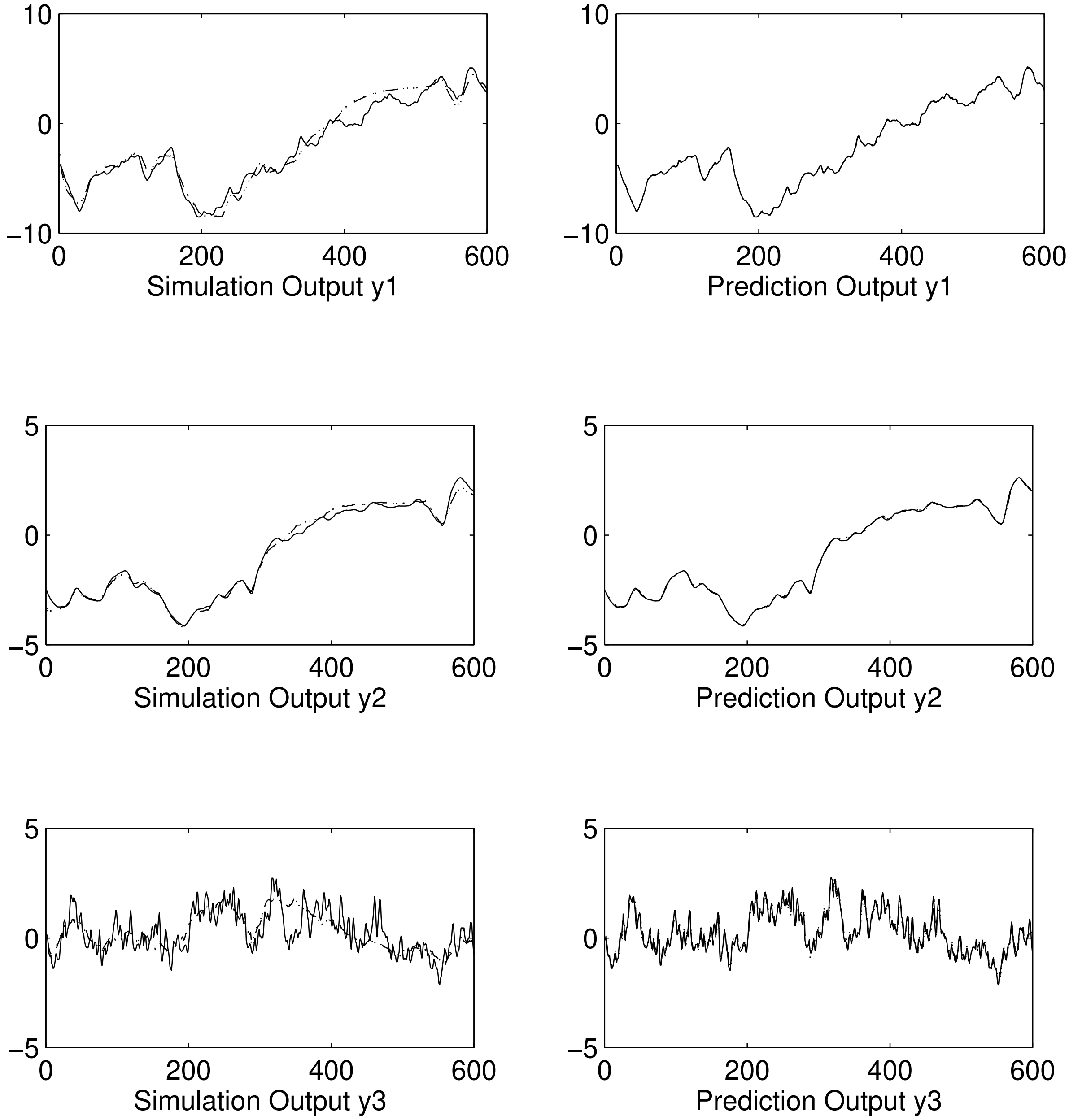

Figure 9.7 *Comparison of alternative* 6*-state models for identification data: original model (solid); model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted). Unit time interval = 10 sec.*

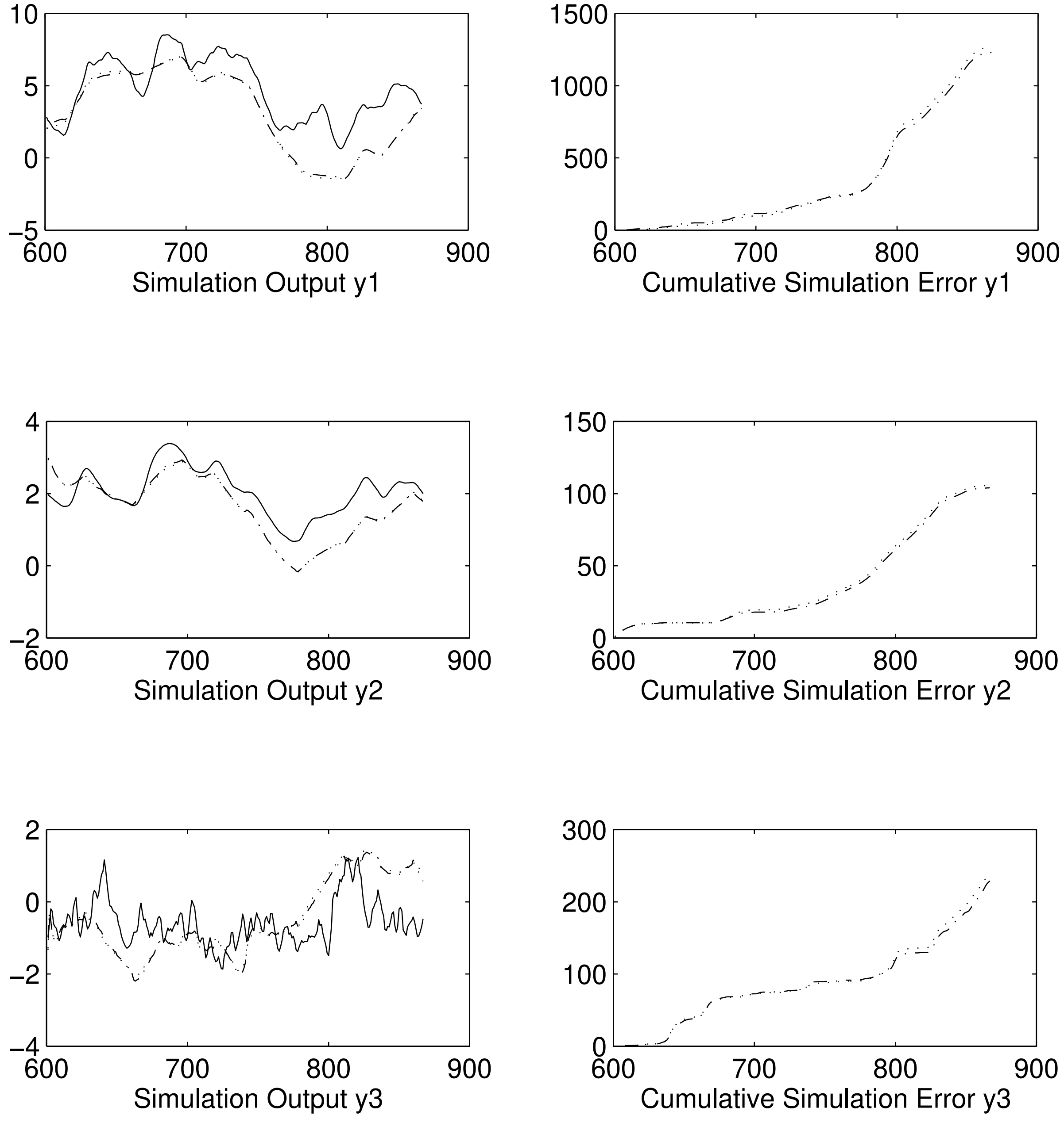

Figure 9.8 *Comparison of simulation outputs of alternative* 6*-state models for validation data: original model (solid); model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted). Unit time interval = 10 sec.*

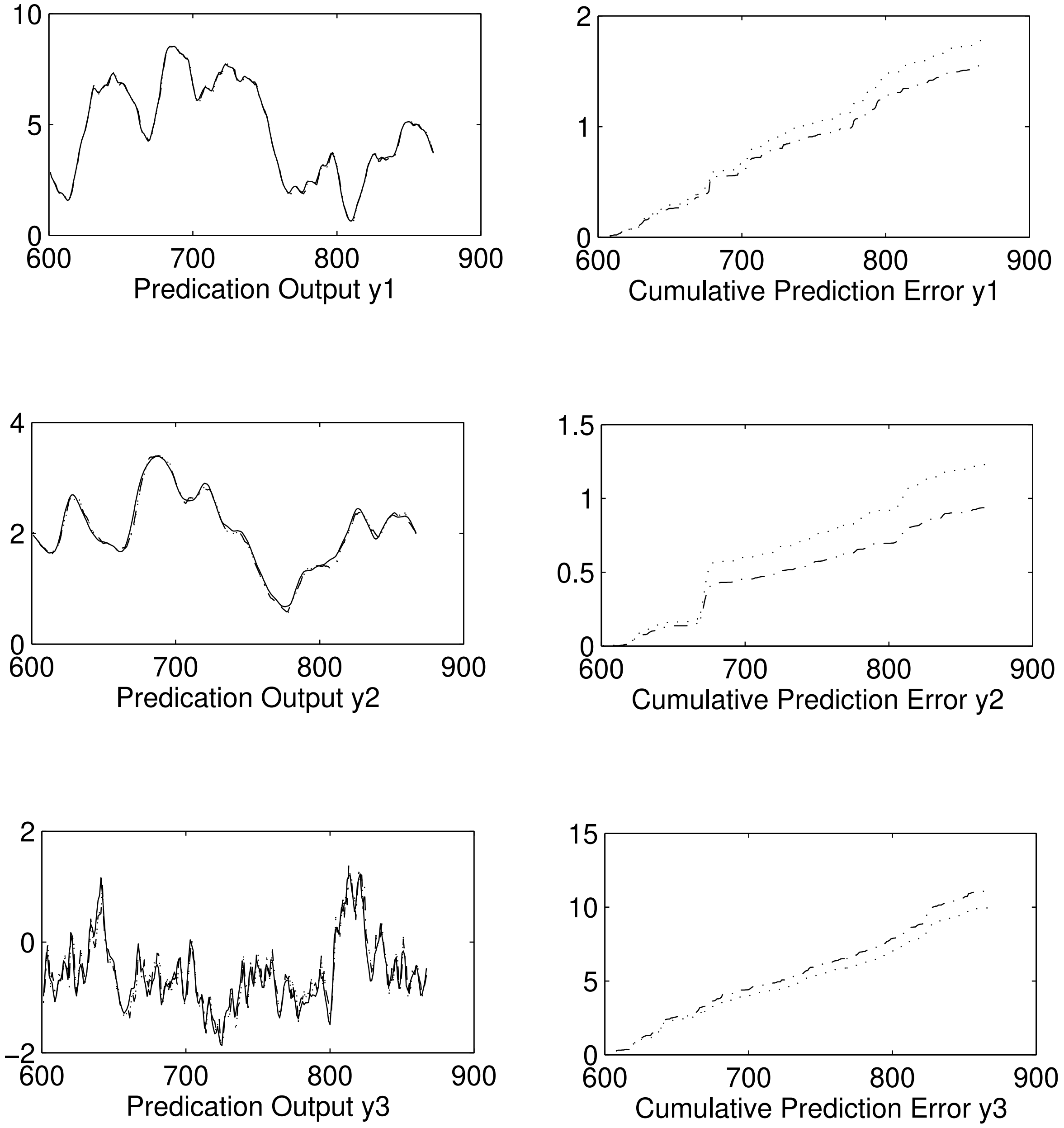

Figure 9.9 *Comparison of prediction outputs of alternative 6-state models for validation data: original model (solid); model from Algorithm 7.10 (dashdot); model from N4SID algorithm (dotted). Unit time interval = 10 sec.*

9.4 Summary

In this chapter, the three algorithms proposed in Chapter 7 are scrutinized by means of two case studies – a distillation column and an industrial dryer. From these studies, it is suggested that Algorithm 7.8 is relatively less robust compared to the other two algorithms. The reason for this is that in determining the system matrices, this algorithm computes the matrices directly from the estimated Markov parameters without referring to the input-output data for a second time. Lacking further fine adjustment with the data, such a scheme could introduce more distortion to the system matrices. Regarding Algorithm 7.9, it is demonstrated that this algorithm could identify the deterministic part of a system fairly well, while identification of the Kalman filter has yet to be improved. The studies in this chapter also show that the performance of Algorithm 7.10 is satisfactory, at a level comparable to that of N4SID algorithm. Furthermore, in quite a number of cases the algorithm indeed produces better models under various measures, namely the $\mathcal{H}_\infty$ error, the $\mathcal{H}_2$ error, the simulation error, and the prediction error.

10

Conclusions

10.1 Contributions of this Dissertation

We summarize the main contributions of this dissertation.

- The state coordinates which correspond to the purely deterministic modes in a subspace projection are clearly identified through a state and output coordinate transformation. The corresponding Kalman filter which only updates the state coordinates other than the purely deterministic ones is presented.
- A subspace projection theorem is proven using a geometric approach. It is also shown that the initial condition of the corresponding Kalman filter is $\hat{X}_p = \Pi_{(\mathcal{Y}_p^d \cap \mathcal{Y}_p)+\mathcal{U}_p+\mathcal{U}_f} X_p$, where the additional term $(\mathcal{Y}_p^d \cap \mathcal{Y}_p)$ has the function of capturing the space spanned by the purely deterministic modes.
- Sufficient conditions for experiments to be informative for two subspace identification algorithms are derived in the deterministic case.
- Using the geometric properties shown earlier, sufficient conditions for experiments to be informative for a subspace identification algorithm are derived for three cases, namely the stochastic output disturbance case, the stochastic state excitation case, and the general excitation and disturbance case.
- A phenomenon, in which deterministic components and stochastic components cancel each other perfectly in a subspace projection, is shown in an example. It is also concluded that such a cancelation can indeed be avoided under very mild conditions derived earlier.
- A necessary condition and a necessary and sufficient condition for experiments to be informative for a subspace algorithm, under *a priori* assumption that initial Markov parameters are known, are derived in the deterministic case.
- A subspace identification method which uses a 3-block data configuration is derived to estimate initial Markov parameters in an unbiased fashion.

- Three new subspace algorithms, based on the Markov parameters estimation method derived earlier, are proposed. All three algorithms, including the one using the state sequence approach, are unbiased.
- All these algorithms are applied to two real case studies, namely a fractional distillation column and an industrial dryer. It is shown that these algorithms estimate system responses at low frequencies fairly well, which in turn result in better $\mathcal{H}_\infty$ and $\mathcal{H}_2$ accuracy compared with N4SID algorithm.
- Sufficient conditions for least-squares solutions obtained by the shift invariance approach to be semistable are acquired.
- Results on eigenvalue perturbation in least-squares estimations are derived. These results are applied successfully to a subspace identification method to enforce stability. Performance of several other stability enforcing methods are also compared.

10.2 Recommendations for Future Work

We conclude with a summary of related research directions which deserve further investigation.

- The algorithms developed in Chapter 7 have been applied to two other case studies, namely a process control system and a glass furnace, proving similar performance. However, more practical applications of the algorithms may be required in order to understand thoroughly their robustness under different situations.
- Necessary and sufficient conditions for informative experiments have been derived in the deterministic case. It may be possible to extend similar necessary and sufficient results to the combined deterministic-stochastic case, probably under a Hilbert space or an ideal probability setup.
- Bilinear models form a class of systems which exhibits a number of phenomena commonly observed in real-world systems, such as biochemical processes. One research direction is to extend the 3-block identification scheme developed in Chapter 7 to identify bilinear systems. A bilinear subspace identification algorithm subject to white input signals has been proposed in [FDMVO96]. However, bilinear subspace identification methods with general input signals are still under investigation in the field.
- Informative experiments for bilinear subspace identification algorithms will need to be examined, probably in a way similar to the analyses in Chapter 5. Some techniques used in Chapter 4 may also be required.

A

Proofs for Chapter 4

A.1 Proof of Lemma 4.1

Proof. Let $\bar{A}_{21}$, $\bar{A}_{31}$ be the $(2,1)$-th and $(3,1)$-th blocks of $\bar{A}$. Then,

$$\begin{bmatrix} \bar{\Sigma}^s_{22} & \bar{\Sigma}^s_{23} \\ \bar{\Sigma}^s_{32} & \bar{\Sigma}^s_{33} \end{bmatrix} := \mathbf{E}\left(\begin{bmatrix} \bar{x}^s_2(t) \\ \bar{x}^s_3(t) \end{bmatrix} \begin{bmatrix} \bar{x}^s_2(t) \\ \bar{x}^s_3(t) \end{bmatrix}^* \right) = \begin{bmatrix} \bar{A}_{21} \\ \bar{A}_{31} \end{bmatrix} \bar{\Sigma}^s_{11} \begin{bmatrix} \bar{A}^*_{21} & \bar{A}^*_{31} \end{bmatrix} + \begin{bmatrix} \bar{\Sigma}^w_{22} & \bar{\Sigma}^w_{23} \\ \bar{\Sigma}^w_{32} & \bar{\Sigma}^w_{33} \end{bmatrix} = 0.$$

The result follows since $\bar{\Sigma}^s_{11} > 0$. □

A.2 Proof of Lemma 4.3

Proof. Note that $\begin{bmatrix} 0 & 0 & \bar{C}_{33} \end{bmatrix} \bar{A} = \begin{bmatrix} 0 & \bar{C}_{33}\bar{A}_{32} & \bar{C}_{33}\bar{A}_{33} \end{bmatrix}$. Since $\bar{x}_2(t)$ corresponds to the modes that are not observable from $(\begin{bmatrix} 0 & 0 & \bar{C}_{33} \end{bmatrix}, \bar{A})$ by definition, we have $\bar{C}_{33}\bar{A}_{32} = 0$. Repeating this procedure, we have

$$\begin{bmatrix} 0 & 0 & \bar{C}_{33} \end{bmatrix} \bar{A}^k = \begin{bmatrix} 0 & \bar{C}_{33}\bar{A}^k_{33}\bar{A}_{32} & \bar{C}_{33}\bar{A}^k_{33} \end{bmatrix}.$$

Thus, we can conclude that

$$\bar{C}_{33}\bar{A}^k_{33}\bar{A}_{32} = 0, \qquad \forall k \in \mathbb{Z}_+. \tag{A.1}$$

Now using the above statement, we can see that

$$\begin{bmatrix} 0 & 0 & \bar{C}_{33} \end{bmatrix} \bar{A}^k = \begin{bmatrix} 0 & 0 & \bar{C}_{33}\bar{A}^k_{33} \end{bmatrix}.$$

Since $\bar{x}_3(t)$ corresponds to all the observable modes, $(\bar{C}_{33}, \bar{A}_{33})$ must be observable. By the standard property of observability, Equation (A.1) implies $\bar{A}_{32} = 0$. □

A.3 Proof of Lemma 4.4

Before we prove Lemma 4.4, we need Definition A.1 and Lemma A.1.

Definition A.1. A matrix R is called a row echelon matrix if R is upper triangular and for any partition $R = \begin{bmatrix} R_l & R_r \end{bmatrix}$, there always exist further partitions of R_l and R_r in such a way that either[1]

$$R = \begin{bmatrix} 0 & R_{r2} \end{bmatrix} \qquad \text{or} \qquad R = \begin{bmatrix} R_{l1} & R_{r1} \end{bmatrix} \qquad \text{or} \qquad R = \begin{bmatrix} R_{l1} & R_{r1} \\ 0 & R_{r2} \end{bmatrix}, \tag{A.2}$$

where $R_{l1}R_{l1}^* > 0$ and R_{r2} is upper triangular.

To illustrate the definition, we shall give an example $\hat{R}$ and a counter-example $\tilde{R}$ of row echelon matrices:

$$\hat{R} = \begin{bmatrix} 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \tilde{R} = \begin{bmatrix} 0 & 1 & 2 & 3 & 0 \\ 0 & 1 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Note also that this is closely related to the reduced row echelon form in linear algebra.

Lemma A.1. *Given C is a $p \times n$ matrix. Then, there exists a factorization $C = QR$ such that Q is unitary and R is in row echelon form.*

Proof. We shall give a constructive proof. Here, we construct such a factorization by reapplying QR factorizations to the submatrix at the lower-right corner column-by-column. Let Q_1R_1 be a QR factorization of C. Partition R_1, an upper triangular matrix, into (see footnote 1):

$$R_1 = \begin{bmatrix} \overbrace{R_{11,1}}^{1} & \overbrace{R_{12,1}}^{n-1} \\ 0 & R_{22,1} \end{bmatrix},$$

with $R_{11,1}R_{11,1}^* > 0$ and $R_{22,1}$ of $n-1$ columns. Next let Q_2R_2 be a QR factorization of $R_{22,1}$. Therefore, C has a QR factorization:

$$C = Q_1 \begin{bmatrix} I & 0 \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} R_{11,1} & R_{12,1} \\ 0 & R_2 \end{bmatrix} =: \tilde{Q}_2\tilde{R}_2,$$

which is a product of a unitary matrix $\tilde{Q}_2$ and an upper triangular matrix $\tilde{R}_2$. Now, R_2 can be partitioned into

$$R_2 = \begin{bmatrix} R_{11,2} & R_{12,2} \\ 0 & R_{22,2} \end{bmatrix},$$

[1] The partition for both cases in Equations (A.2) can in fact be written more compactly as:

$$R = \begin{bmatrix} R_{l1} & R_{r1} \\ 0 & R_{r2} \end{bmatrix},$$

if submatrices are allowed to be of zero dimension. In such a case, if R_{l1} or R_{r2} is of zero dimension, then the condition $R_{l1}R_{l1}^* > 0$ or R_{r2} is upper triangular can respectively be dropped.

with $R_{11,2}R_{11,2}^* > 0$ and $R_{22,2}$ of $n-2$ columns. From this, C can be written as:

$$C = \tilde{Q}_2\tilde{R}_2 =: \tilde{Q}_2 \begin{bmatrix} \overbrace{\tilde{R}_{11,2}}^{2} & \overbrace{\tilde{R}_{12,2}}^{n-2} \\ 0 & R_{22,2} \end{bmatrix} := \tilde{Q}_2 \left[\begin{array}{cc|c} R_{11,1} & * & * \\ 0 & R_{11,2} & R_{12,2} \\ \hline 0 & 0 & R_{22,2} \end{array}\right],$$

where $R_{11,1}R_{11,1}^* > 0$ and $R_{11,2}R_{11,2}^* > 0$. It is now obvious that $\tilde{R}_{11,2}\tilde{R}_{11,2}^* > 0$. Repeating the QR factorization procedure for the remaining $n-2$ columns, we complete the proof by induction. □

Proof. (Lemma 4.4) Since Σ^s is symmetric, the matrix can always be diagonalized with an invertible matrix T_1. By Lemma 4.1, we then have

$$T_1 A T_1^{-1} = \left[\begin{array}{c|cc} \bar{A}_{11} & * & * \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$(\bar{A})_{22}$}} \\ 0 & & \end{array}\right], \qquad T_1\Sigma^s T_1^{-1} = \begin{bmatrix} \bar{\Sigma}_{11}^s & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Now by Lemma A.1, CT_1^{-1} has a QR factorization such that $C_{11}C_{11}^* > 0$ and

$$Q_1^* C T_1^{-1} = \begin{bmatrix} \bar{C}_{11} & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}, \qquad Q_1^*\Sigma^v Q_1 = \begin{bmatrix} \bar{\Sigma}_{11}^v & * & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

Again, Σ^v is symmetric; and hence $Q_1^*\Sigma^v Q_1$ always has an SVD, $Q_2\bar{\Sigma}^v Q_2^*$, such that

$$\left[\begin{array}{c|cc} I & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$(Q_2)_{22}$}} \\ 0 & & \end{array}\right]^* Q_1^*\Sigma^v Q_1 \left[\begin{array}{c|cc} I & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$(Q_2)_{22}$}} \\ 0 & & \end{array}\right] = \begin{bmatrix} \bar{\Sigma}_{11}^v & \bar{\Sigma}_{12}^v & 0 \\ \bar{\Sigma}_{21}^v & \bar{\Sigma}_{22}^v & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where Q_2 has a partition as above. Thus, T_o can be defined as $Q_2^*Q_1^*$. Now, let

$$T_o C T_1^{-1} =: \left[\begin{array}{c|cc} \bar{C}_{11} & * & * \\ \hline 0 & \multicolumn{2}{c}{(\bar{C})_{22}} \\ 0 & \multicolumn{2}{c}{(\bar{C})_{32}} \end{array}\right].$$

There always exists an invertible T_2 such that

$$T_2 := \left[\begin{array}{c|cc} I & 0 & 0 \\ \hline 0 & \multicolumn{2}{c}{\multirow{2}{*}{$(T_2)_{22}$}} \\ 0 & & \end{array}\right], \qquad \begin{bmatrix} (\bar{C})_{32} \\ (\bar{C})_{32}\,(\bar{A})_{22} \\ (\bar{C})_{32}\,(\bar{A})_{22}^2 \\ \vdots \end{bmatrix} (T_2)_{22}^{-1} =: \left[\begin{array}{c|c} 0 & \bar{C}_{33} \\ 0 & \bar{C}_{33}\,\bar{A}_{33} \\ 0 & \bar{C}_{33}\,\bar{A}_{33}^2 \\ \vdots & \vdots \end{array}\right],$$

where $(\bar{C}_{33}, \bar{A}_{33})$ is observable. Let $T_s := T_2T_1$; hence $T_oCT_s^{-1}$ has the form as in Equation (4.2). Therefore, by Lemma 4.3 we have

$$T_o C T_s^{-1} =: \begin{bmatrix} \bar{C}_{11} & \bar{C}_{12} & \bar{C}_{13} \\ 0 & \bar{C}_{22} & \bar{C}_{23} \\ 0 & 0 & \bar{C}_{33} \end{bmatrix}, \qquad T_s A T_s^{-1} =: \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix}.$$

Note that

$$T_o C A^k T_s^{-1} = \begin{bmatrix} \bar{C}_{11}\bar{A}_{11}^k & * & * \\ 0 & * & * \\ 0 & 0 & * \end{bmatrix}.$$

Now, since (C, A) is observable, $(\bar{C}_{11}, \bar{A}_{11})$ is also observable. □

A.4 Proof of Lemma 4.6

Proof. Let $\mathcal{K}_t$ be the 1-step ahead t-block Kalman filter gain (i.e., the usual Kalman filter gain for System (4.8)). Furthermore, let

$$\tilde{P}_0 := \mathbf{E}([x(0) - \hat{x}(0)]\,[x(0) - \hat{x}(0)]^*) - \Sigma^s.$$

With the use of the equation $\Sigma^s = A\Sigma^s A^* + \Sigma^w$ and $\hat{x}(0) \perp (\underline{w}(0), \underline{v}(0))$, we have

$$\Sigma^s = A^t \Sigma^s (A^t)^* + \mathcal{A}_t \mathbf{E}[\underline{w}(0)\underline{w}^*(0)]\mathcal{A}_t^*.$$

Moreover,

$$\begin{aligned} \mathbf{E}\left[x^s(t)(\underline{y}^s(0))^*\right] &= \mathbf{E}\left\{[A^t x^s(0) + \mathcal{A}_t\underline{w}(0)] \cdot [\mathcal{O}_t x^s(0) + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\right\} \\ &= A^t\Sigma^s\mathcal{O}_t^* + \mathbf{E}\big\{[\mathcal{A}_t\underline{w}(0)] \cdot [\mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\big\} \\ &= \mathcal{C}_t^s; \end{aligned}$$

and

$$\begin{aligned} \mathbf{E}\left[\underline{y}^s(0)(\underline{y}^s(0))^*\right] &= \mathbf{E}\left\{[\mathcal{O}_t x^s(0) + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)] \cdot [\mathcal{O}_t x^s(0) + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\right\} \\ &= \mathcal{O}^t\Sigma^s\mathcal{O}_t^* + \mathbf{E}\big\{[\mathcal{N}_t\underline{w}(0) + \underline{v}(0)] \cdot [\mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\big\} \\ &= L_t. \end{aligned}$$

Thus,

$$\begin{aligned} A^t(\Sigma^s + \tilde{P}_0)\mathcal{O}_t^* + \mathbf{E}\big\{[\mathcal{A}_t\underline{w}(0)] \cdot [\mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\big\} &= \mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*, \\ \mathcal{O}^t(\Sigma^s + \tilde{P}_0)\mathcal{O}_t^* + \mathbf{E}\big\{[\mathcal{N}_t\underline{w}(0) + \underline{v}(0)] \cdot [\mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\big\} &= L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*. \end{aligned}$$

Then, using standard estimation theory, we have

$$\begin{aligned} \mathcal{K}_t &= \mathbf{E}\Big\{(x(t) - A^t\hat{x}(0) - \mathcal{C}_t^d\underline{u}(0)) \cdot (\underline{y}(0) - \mathcal{O}_t\hat{x}(0) - \mathcal{T}_t\underline{u}(0))^*\Big\} \\ &\qquad \times \mathbf{E}\left\{[\underline{y}(0) - \mathcal{O}_t\hat{x}(0) - \mathcal{T}_t\underline{u}(0)] \cdot [\underline{y}(0) - \mathcal{O}_t\hat{x}(0) - \mathcal{T}_t\underline{u}(0)]^*\right\}^{-1} \\ &= \mathbf{E}\left\{[A^t[x(0) - \hat{x}(0)] + \mathcal{A}_t\underline{w}(0)] \cdot [\mathcal{O}_t[x(0) - \hat{x}(0)] + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\right\} \\ &\qquad \times \mathbf{E}\left\{[\mathcal{O}_t[x(0) - \hat{x}(0)] + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)] \cdot [\mathcal{O}_t[x(0) - \hat{x}(0)] + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]^*\right\}^{-1} \\ &= (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^{-1}. \end{aligned}$$

Moreover,

$$\begin{aligned}
\tilde{P}_t &= \mathbf{E}\left([x(t)-\hat{x}(t)][x(t)-\hat{x}(t)]^*\right) - \Sigma^s \\
&= \mathbf{E}\big[\big(A^t[x(0)-\hat{x}(0)] + \mathcal{A}_t\underline{w}(0) - \mathcal{K}_t[\mathcal{O}_t[x(0)-\hat{x}(0)] + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]\big) \\
&\qquad \times \big(A^t[x(0)-\hat{x}(0)] + \mathcal{A}_t\underline{w}(0) - \mathcal{K}_t[\mathcal{O}_t[x(0)-\hat{x}(0)] + \mathcal{N}_t\underline{w}(0) + \underline{v}(0)]\big)^*\big] - \Sigma^s \\
&= A^t(\Sigma^s + \tilde{P}_0)(A^t)^* + \mathcal{A}_t\mathbf{E}[\underline{w}(0)\underline{w}^*(0)]\mathcal{A}_t^* - \Sigma^s \\
&\qquad - (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)\mathcal{K}_t^* - \mathcal{K}_t(\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)^* + \mathcal{K}_t(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)\mathcal{K}_t^* \\
&= A^t\tilde{P}_0(A^t)^* - (\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)(L_t + \mathcal{O}_t\tilde{P}_0\mathcal{O}_t^*)^{-1}(\mathcal{C}_t^s + A^t\tilde{P}_0\mathcal{O}_t^*)^*.
\end{aligned}$$

Since $\mathcal{K}_t$ and $\tilde{P}_t$ are exactly the same as those given in Theorem 3.3, this completes the proof. □

A.5 Proof of Lemma 4.8

Proof. For part 1, $\mathcal{X}_p^s \perp \hat{\mathcal{X}}_p$ is obvious. $\hat{\mathcal{X}}_p \perp \tilde{\mathcal{X}}_p$ is a projection property. $\mathcal{X}_p^s \perp \tilde{\mathcal{X}}_p$ follows by noting $(\mathcal{X}_p^d, \hat{\mathcal{X}}_p) \perp \mathcal{X}_p^s$. For part 2, recall that

$$Y_p^d = \mathcal{O}_k X_p^d + \mathcal{T}_k U_p.$$

Now, the spaces spanned by $\{T_{o1}Y_j \,|\, j=0,\dots,k-1\}$ and $\{T_{o2}Y_j \,|\, j=0,\dots,k-1\}$ have no non-zero intersection with $\mathcal{Y}_p^d$ since it is assumed that the k-block stochastic covariance $\bar{L}_k$ is fully excited, i.e. $\bar{L}_k > 0$, where $\bar{L}_k$ is defined in Equation (4.7). However, $\{T_{o3}Y_j \,|\, j=0,\dots,k-1\}$ lies exactly in $\mathcal{Y}_p^d$. Thus, part 2 is proven. Part 3 can be seen by recognizing, with the use of Equations (4.5), that:

$$\begin{aligned}
T_{s3}X_{t+1} &= \bar{A}_{33}T_{s3}X_t + T_{s3}BU_t, \\
T_{o3}Y_t &= \bar{C}_{33}T_{s3}X_t + T_{o3}DU_t.
\end{aligned}$$

The result follows from Equations (3.13). Part 4 is an immediate consequence of part 1. □

B

Proofs for Chapter 5

B.1 Proof of Lemma 5.3

Proof. With the fact that $A\mathcal{Q} \subset \mathcal{Q}$, we have

$$A = \Pi_{\mathcal{Q}} A + \Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp} + \Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}} = \Pi_{\mathcal{Q}} A + \Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}.$$

Using the above equation, it is easy to show that $\Pi_{\mathcal{Q}^\perp} A^j = \left(\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}\right)^j$, for $j \in \mathbb{N}$. Suppose $\tilde{p}_{\min}(z) = z^r + \tilde{a}_{r-1} z^{r-1} + \cdots + \tilde{a}_0$ is the A-conductor of $\mathbb{R}^n$ into $\mathcal{Q}$. Then,

$$\Pi_{\mathcal{Q}^\perp}(A^r + \tilde{a}_{r-1} A^{r-1} + \cdots + \tilde{a}_0 I) = 0. \tag{B.1}$$

Substituting $\Pi_{\mathcal{Q}^\perp} A^j = \left(\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}\right)^j$, for $j \in \mathbb{N}$, and factorizing $\Pi_{\mathcal{Q}^\perp}$ from the left give

$$\Pi_{\mathcal{Q}^\perp}\left[(\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp})^r + \tilde{a}_{r-1}(\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp})^{r-1} + \cdots + \tilde{a}_0 I\right] = 0, \tag{B.2}$$

which clearly conducts $\mathbb{R}^n$ into $\mathcal{Q}$ under $\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}$. Since Equation (B.1) and Equation (B.2) are equivalent, that $\tilde{p}_{\min}(z)$ is the polynomial of least degree conducting $\mathbb{R}^n$ into $\mathcal{Q}$ under A implies that it is also the polynomial of least degree conducting $\mathbb{R}^n$ into $\mathcal{Q}$ under $\Pi_{\mathcal{Q}^\perp} A \Pi_{\mathcal{Q}^\perp}$. □

B.2 Proof of Theorem 5.4

Proof. First of all, we will make use of two additional Hankel matrices of signals u and x. Denote by shorthands $\bar{U}$ and $\bar{X}$ the matrices

$$\bar{U} := U^{l+h}, \qquad \bar{X} := X^{l+h},$$

where U^{l+h} and X^{l+h} are defined as in Equation (5.6). These matrices effectively have more rows but less columns. The introduction of these matrices is to fully utilize their excitations in the proof. In addition, we introduce two matrices Ξ and $\tilde{\Xi}$,

$$\Xi := \begin{bmatrix} I & & & & & 0 \\ a_{r-1}I & \ddots & & & & \\ \vdots & \ddots & I & & & \\ a_0 I & & a_{r-1}I & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & a_0 I & \cdots & a_{r-1}I & I \end{bmatrix}, \qquad \tilde{\Xi} := \begin{bmatrix} I & & & & & 0 \\ \tilde{a}_{h-1}I & \ddots & & & & \\ \vdots & \ddots & I & & & \\ \tilde{a}_0 I & & \tilde{a}_{h-1}I & \ddots & & \\ & \ddots & & \ddots & \ddots & \\ 0 & & \tilde{a}_0 I & \cdots & \tilde{a}_{h-1}I & I \end{bmatrix},$$

where Ξ will be used in Part 1 and $\tilde{\Xi}$ in Part 2 of the proof.

(1): Let $p_{\min}(z) = z^r + a_{r-1}z^{r-1} + \cdots + a_0$ be the minimal polynomial of A, $r \leq n$. Note that for $j = 0, 1, \ldots, r$, we have

$$\bar{X}_{h-j} = A^{r-j}\bar{X}_{h-r} + \sum_{i=1}^{r-j} A^{i-1}B\bar{U}_{h-i-j}.$$

Now, assume that $h \geq r$ and define ψ as

$$\begin{aligned} \psi &:= a_0 \begin{bmatrix} \bar{X}_{h-r} \\ \bar{U}_{h-r,l+h-r-1} \end{bmatrix} + \cdots + a_{r-1} \begin{bmatrix} \bar{X}_{h-1} \\ \bar{U}_{h-1,l+h-2} \end{bmatrix} + \begin{bmatrix} \bar{X}_h \\ \bar{U}_{h,l+h-1} \end{bmatrix} \\ &= a_0 \begin{bmatrix} \bar{X}_{h-r} \\ \bar{U}_{h-r,l+h-r-1} \end{bmatrix} + \cdots + \begin{bmatrix} A^r\bar{X}_{h-r} + \sum_{i=1}^{r} A^{i-1}B\bar{U}_{h-i} \\ \bar{U}_{h,l+h-1} \end{bmatrix} \\ &= \begin{bmatrix} (A^{r-1}B + a_{r-1}A^{r-2}B + \cdots + a_1 B)\bar{U}_{h-r} + \cdots + B\bar{U}_{h-1} \\ a_0\bar{U}_{h-r,l+h-r-1} + \cdots + a_{r-1}\bar{U}_{h-1,l+h-2} + \bar{U}_{h,l+h-1} \end{bmatrix}. \end{aligned}$$

Moreover, let

$$H := \begin{bmatrix} A^{r-1}B & A^{r-2}B & \cdots & B & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix} \Xi,$$

where Ξ has been defined in the beginning of the proof. It can easily be seen that H is of full row rank and $\psi = H\bar{U}_{h-r,l+h-1}$. Since it is given that

$$\mathbf{E}\left(U^{l+h}\,(U^{l+h})^*\right) = \mathbf{E}\left(\bar{U}_{0,l+h-1}\,\bar{U}^*_{0,l+h-1}\right) > 0,$$

its principal submatrix is also positive definite, i.e. $\mathbf{E}\left(\bar{U}_{h-r,l+h-1}\,\bar{U}^*_{h-r,l+h-1}\right) > 0$. Thus, we have

$$\begin{aligned} \mathbf{E}\,(\psi\,\psi^*) &= \mathbf{E}\left(H\bar{U}_{h-r,l+h-1}\,\bar{U}^*_{h-r,l+h-1}H^*\right) \\ &= H\mathbf{E}\left(\bar{U}_{h-r,l+h-1}\,\bar{U}^*_{h-r,l+h-1}\right)H^* \\ &> 0. \end{aligned} \tag{B.3}$$

The first statement will be proven by contradiction. Suppose $\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l$ is of dimension less than n. There exist constants $\alpha = \begin{bmatrix}\alpha_1 & \cdots & \alpha_{n+lm}\end{bmatrix}^* \in \mathbb{R}^{n+lm}$, $\alpha \neq 0$, such that

$$\alpha^* \begin{bmatrix}x(t)^* & u(t)^* & \cdots & u(t+l-1)^*\end{bmatrix}^* = 0,$$

for all $t = 0, 1, \ldots, N-l$. Then it is easy to see that

$$\alpha^*\psi = a_0\alpha^* \begin{bmatrix}\bar{X}_{h-r} \\ \bar{U}_{h-r,l+h-r-1}\end{bmatrix} + \cdots + a_{r-1}\alpha^* \begin{bmatrix}\bar{X}_{h-1} \\ \bar{U}_{h-1,l+h-2}\end{bmatrix} + \alpha^* \begin{bmatrix}\bar{X}_h \\ \bar{U}_{h,l+h-1}\end{bmatrix} = 0;$$

or equivalently, $\alpha^*\mathbf{E}\left(\psi\,\psi^*\right)\alpha = 0$, which is a contradiction to Equation B.3.

(2): It remains to consider the case $h < r$. Suppose that $\tilde{p}_{\min}(z) = z^h + \tilde{a}_{h-1}z^{h-1} + \cdots + \tilde{a}_0$ is the A-conductor of $\mathbb{R}^n$ into $\mathcal{Q} \subset \mathbb{R}^n$, where $A\mathcal{Q} \subset \mathcal{Q}$. Let $Q^\perp \in \mathbb{R}^{n\times n}$ be the matrix presentation of $\Pi_{\mathcal{Q}^\perp} : \mathbb{R}^n \to \mathbb{R}^n$ as discussed in Remark 5.4. Thus, we have

$$Q^\perp(A^h + \tilde{a}_{h-1}A^{h-1} + \cdots + \tilde{a}_0 I) = 0.$$

Moreover, for $j = 0, 1, \ldots, h$, we have

$$\bar{X}_{h-j} = A^{r-j}\bar{X}_{h-r} + \sum_{i=1}^{r-j} A^{i-1}B\bar{U}_{h-i-j}.$$

Therefore, we can write

$$\begin{aligned}
\tilde{\psi} &:= \tilde{a}_0 \begin{bmatrix}Q^\perp\bar{X}_0 \\ \bar{U}_{0,l-1}\end{bmatrix} + \cdots + \tilde{a}_{h-1} \begin{bmatrix}Q^\perp\bar{X}_{h-1} \\ \bar{U}_{h-1,l+h-2}\end{bmatrix} + \begin{bmatrix}Q^\perp\bar{X}_h \\ \bar{U}_{h,l+h-1}\end{bmatrix} \\
&= \tilde{a}_0 \begin{bmatrix}Q^\perp\bar{X}_0 \\ \bar{U}_{0,l-1}\end{bmatrix} + \cdots + \begin{bmatrix}Q^\perp A^h\bar{X}_0 + Q^\perp\sum_{j=1}^h A^{j-1}B\bar{U}_{h-j} \\ \bar{U}_{h,l+h-1}\end{bmatrix} \\
&= \begin{bmatrix}Q^\perp & 0 \\ 0 & I\end{bmatrix} \begin{bmatrix}(A^{h-1}B + \tilde{a}_{h-1}A^{h-2}B + \cdots + \tilde{a}_1 B)\bar{U}_0 + \cdots + B\bar{U}_{h-1} \\ \tilde{a}_0\bar{U}_{0,l-1} + \cdots + \tilde{a}_{h-1}\bar{U}_{h-1,l+h-2} + \bar{U}_{h,l+h-1}\end{bmatrix}.
\end{aligned}$$

Now, let

$$\tilde{H} := \begin{bmatrix}Q^\perp & 0 \\ 0 & I\end{bmatrix} \begin{bmatrix}A^{h-1}B & A^{h-2}B & \cdots & B & 0 \\ 0 & 0 & 0 & 0 & I\end{bmatrix} \tilde{\Xi}, \tag{B.4}$$

where $\tilde{\Xi}$ has been defined in the beginning of the proof. We then have $\tilde{\psi} = \tilde{H}\bar{U}_{0,l+h-1} = \tilde{H}U^{l+h}$.

The second statement will be proven by contradiction. Suppose $\Pi_{\mathcal{U}^{l\perp}}\mathcal{X}_0^l$ is of dimension less than $\tilde{n}$. There exist constants $\tilde{\alpha} = \begin{bmatrix}\tilde{\alpha}_1 & \cdots & \tilde{\alpha}_{n+lm}\end{bmatrix}^* \in \mathbb{R}^{n+lm}$, $\tilde{\alpha} \neq 0$ such that

$$\tilde{\alpha}^* \begin{bmatrix}Q^\perp & 0 \\ 0 & I\end{bmatrix} \neq 0, \qquad \tilde{\alpha}^* \begin{bmatrix}Q^\perp & 0 \\ 0 & I\end{bmatrix} \begin{bmatrix}x(t)^* & u(t)^* & \cdots & u(t+l-1)^*\end{bmatrix}^* = 0, \tag{B.5}$$

for all $t = 0, 1, \ldots, N-l$. The first inequality above, together with the definition of $\tilde{H}$ in Equation (B.4), assures that $\tilde{\alpha}^* \tilde{H} \neq 0$, which in turn implies

$$\tilde{\alpha}^* \mathbf{E}\left[\,\tilde{\psi}\tilde{\psi}^*\,\right]\tilde{\alpha} = \tilde{\alpha}^* \tilde{H}\, \mathbf{E}\left[\,(U^{l+h})(U^{l+h})^*\,\right] \tilde{H}^* \tilde{\alpha} > 0.$$

However, the second equality in Equation (B.5) implies

$$\tilde{\alpha}^* \tilde{\psi} = \tilde{a}_0 \tilde{\alpha}^* \begin{bmatrix} Q^\perp \bar{X}_0 \\ \bar{U}_{0,l-1} \end{bmatrix} + \cdots + \tilde{a}_{h-1} \tilde{\alpha}^* \begin{bmatrix} Q^\perp \bar{X}_{h-1} \\ \bar{U}_{h-1,l+h-2} \end{bmatrix} + \tilde{\alpha}^* \begin{bmatrix} Q^\perp \bar{X}_h \\ \bar{U}_{h,l+h-1} \end{bmatrix} = 0.$$

Now, taking the covariance of the above equation will result in $\tilde{\alpha}^* \mathbf{E}\left[\,\tilde{\psi}\tilde{\psi}^*\,\right]\tilde{\alpha} = 0$, which is clearly a contradiction.

This completes the proof. $\square$

B.3 Proof of Theorem 5.5

Proof. First, let e be a noise process such that e is uncorrelated with u, and

$$\mathbf{E}\left(e(t)\, e^*(\tau)\right) = I\delta_{t\tau}, \qquad w(t) = (\Sigma^w)^{\frac{1}{2}} e(t).$$

Furthermore, we define the notations

$$\bar{B} = \begin{bmatrix} B & (\Sigma^w)^{\frac{1}{2}} \end{bmatrix}, \qquad \bar{u}(t) := \begin{bmatrix} u(t)^* & e(t)^* \end{bmatrix}^*.$$

Now, the theorem can be proven using Theorem 5.4 by replacing B by $\bar{B}$ and $u(t)$ by $\bar{u}(t)$. $\square$

B.4 Proof of Theorem 5.6

Proof. The proof will be based on Theorem 5.4, which guarantees $\Pi_{\mathcal{U}^{l\perp}} \mathcal{Y}^l$ is of dimension n. Consider Equation (3.12). Projecting each row onto the space $\mathcal{U}^{l\perp}$ yields

$$\Pi_{\mathcal{U}^{l\perp}} Y^l = \mathcal{O}_l \Pi_{\mathcal{U}^{l\perp}} X_0^l,$$

which cancels the effect of the input Hankel block U^l.

With the fact that $\mathcal{O}_l$ is injective, the above equation spans an n-dimensional subspace by Theorem 5.4. Suppose we choose certain state coordinates (note that we allow a similarity transformation $GL_n(\mathbb{R})$) such that

$$\mathbf{E}\left[\left(\Pi_{\mathcal{U}^{l\perp}} X_0^l\right)\left(\Pi_{\mathcal{U}^{l\perp}} X_0^l\right)^*\right] = \Sigma_1 \Sigma_1^*,$$

with Σ_1 as defined in Equation (5.8). Then, we have

$$\mathcal{O}_l = \Gamma_1, \qquad \text{and} \qquad \Pi_{\mathcal{U}^{l\perp}} X_0^l = \Sigma_1 \Omega_1^*,$$

with Γ_1 and Ω_1^* as defined in Equation (5.8). Thus, with the injectivity of $\mathcal{O}_{l-1}$, A is uniquely determined by the shift invariant property and C is uniquely determined from the first p rows of $\mathcal{O}_l$, as shown in Step 3 of Algorithm 5.1.

Note that since $\mathcal{O}_{l-1}$ is injective, there always exists P such that

$$P\mathcal{O}_{l-1} + CA^{l-1} = 0.$$

Define $\mathcal{O}_l^{\perp} := \begin{bmatrix} P & I \end{bmatrix}$. It is obvious that $\mathcal{O}_l^{\perp}\mathcal{O}_l = 0$. Furthermore, by the definition of sufficient richness we have $U_0^l, \ldots, U_{l-1}^k$ mutually linearly independent. Thus, $S_0, \ldots, S_{l-1}$ in Equation (5.9) are uniquely determined. Partition $\mathcal{O}_l^{\perp}$ into

$$\mathcal{O}_l^{\perp} = \begin{bmatrix} P_0 & \cdots & P_{l-2} & I \end{bmatrix}.$$

It is then easy to see that Equation (5.10) holds, which uniquely determines B and D. This completes the proof. □

B.5 Proof of Proposition 5.8

Proof. (1) ⇔ (2): We shall prove $\Pi_{\mathcal{U}_f^{\perp}}$ and $\Pi_{\mathcal{S}}$ commute. Since $\mathcal{U}_f \subset \mathcal{S}$, we have $\mathcal{S}^{\perp} \subset \mathcal{U}_f^{\perp}$. This gives

$$\Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}^{\perp}} = \Pi_{\mathcal{S}^{\perp}} = \Pi_{\mathcal{S}^{\perp}}\Pi_{\mathcal{U}_f^{\perp}}. \tag{B.6}$$

Since $\Pi_{\mathcal{S}} + \Pi_{\mathcal{S}^{\perp}} = I$, it implies $(\Pi_{\mathcal{S}} + \Pi_{\mathcal{S}^{\perp}})\Pi_{\mathcal{U}_f^{\perp}} = \Pi_{\mathcal{U}_f^{\perp}}(\Pi_{\mathcal{S}} + \Pi_{\mathcal{S}^{\perp}})$. Using Equation (B.6), we have $\Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}} = \Pi_{\mathcal{S}}\Pi_{\mathcal{U}_f^{\perp}}$.

(2) ⇒ (3): We shall prove by contradiction. Suppose Statement 3 is not true, then there exist column vectors $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^{km}$, not both zero, such that $\alpha^* \Pi_{\mathcal{S}} X_f + \beta^* U_f = 0$. This implies the $\mathcal{U}_f^{\perp}$ component must satisfy:

$$\alpha^* \Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}}\mathcal{X}_f = 0.$$

Thus, if $\alpha \neq 0$, this violates $\dim\left\{\Pi_{\mathcal{U}_f^{\perp}}\Pi_{\mathcal{S}}X_f\right\} = n$. On the other hand, suppose $\alpha = 0$. Then, $\beta \neq 0$ and $\beta^* U_f = 0$ violate $\mathbf{E}\left[U_f U_f^*\right] > 0$.

(3) ⇒ (4): Note that since $\mathbf{E}\left[\left(\Pi_{\mathcal{S}^{\perp}}X_f\right)\begin{pmatrix} Y_p^* & U_p^* & U_f^* \end{pmatrix}\right] = \mathbf{E}\left[\left(\Pi_{\mathcal{S}^{\perp}}U_f\right)\begin{pmatrix} Y_p^* & U_p^* & U_f^* \end{pmatrix}\right] = 0$, we have

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\begin{pmatrix} Y_p^* & U_p^* & U_f^* \end{pmatrix}\right] = \mathbf{E}\left[\begin{pmatrix} \Pi_{\mathcal{S}}X_f \\ \Pi_{\mathcal{S}}U_f \end{pmatrix}\begin{pmatrix} Y_p^* & U_p^* & U_f^* \end{pmatrix}\right]. \tag{B.7}$$

Now, we shall prove by contradiction. Recall that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$. Thus, Equation (B.7) is not of full row rank if and only if there exist $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}^{km}$, not both zero, such that $\alpha^*\Pi_{\mathcal{S}}X_f + \beta^*U_f = 0$. This implies that Statement 3 is not valid.

(4) $\Rightarrow$ (1): Let $S^* = \begin{bmatrix} Y_p^* & U_p^* & U_f^* \end{bmatrix}$. Note that

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \begin{pmatrix} Y_p^* & U_P^* & U_f^* \end{pmatrix}\right] = \mathbf{E}\left[\begin{pmatrix} \Pi_{\mathcal{U}_f} X_f \\ U_f \end{pmatrix} S^*\right] + \mathbf{E}\left[\begin{pmatrix} \Pi_{\mathcal{U}_f^\perp} X_f \\ 0 \end{pmatrix} S^*\right].$$

Since the above equation is of full row rank, the first term implies $\mathbf{E}\left[U_f\, S^*\right]$ is of full row rank, or equivalently, $\mathbf{E}\left[U_f U_f^*\right] > 0$. On the other hand, since $\Pi_{\mathcal{U}_f} \mathcal{X}_f \subset \mathcal{U}_f$, the second term in turn implies $\mathbf{E}\left[(\Pi_{\mathcal{U}_f^\perp} X_f)\, S^*\right]$ is of full row rank. This is also to say that $\mathbf{E}\left[(\Pi_{\mathcal{S}} \Pi_{\mathcal{U}_f^\perp} X_f)\, S^*\right]$ is of full row rank, which guarantees $\Pi_{\mathcal{S}} \Pi_{\mathcal{U}_f^\perp} \mathcal{X}_f$ to be of dimension n.

This completes the proof. □

B.6 Proof of Lemma 5.9

Proof. Let $\nu \neq 0$ such that

$$\nu \begin{bmatrix} A^k & A^{k-1}B & \cdots & B \end{bmatrix} = 0.$$

Since $\nu A^k = 0$, we have $\nu A^{k+i} B = 0$ for all $i \in \mathbb{N}$. Thus,

$$\nu \begin{bmatrix} A^i B & A^{i-1}B & \cdots & B \end{bmatrix} = 0,$$

for all $i \in \mathbb{N}$, which is a contradiction. □

B.7 Proof of Theorem 5.14

Proof. Let $\hat{X}_p := \Pi_{\mathcal{U}_p + \mathcal{U}_f} X_p$ and $\tilde{X}_p := X_p^d - \hat{X}_p$. Thus,

$$\Pi_{(\mathcal{U}_p + \mathcal{U}_f)^\perp} X_p = \tilde{X}_p + X_p^s.$$

It is easy to verify that

$$\begin{aligned} \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp} Y_p &= \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp} (\mathcal{O}_k X_p^d + \mathcal{C}_k^d U_p + Y_p^s) \\ &= \mathcal{O}_k \tilde{X}_p + Y_p^s. \end{aligned}$$

Observe that $\mathcal{S} = \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f = \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp} \mathcal{Y}_p + \mathcal{U}_p + \mathcal{U}_f$. Thus, by Lemma 5.8, we shall prove

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix} \left([\mathcal{O}_k \tilde{X}_p + Y_p^s]^* \quad U_p^* \quad U_f^* \right)\right]$$

is of full row rank. It is obvious that $\mathbf{E}\left[\,(U_f)\,\begin{pmatrix} U_p^* & U_f^* \end{pmatrix}\right]$ is of full row rank by the definition of sufficient richness. Thus, it is sufficient to show $\mathbf{E}\left[\,(X_f)\,(\mathcal{O}_k \tilde{X}_p + Y_p^s)^*\,\right]$ is of full row rank.

Since $\hat{X}_p$ and U_p are orthogonal to $\tilde{X}_p$ and Y_p^s, we can write

$$\begin{aligned}\frac{1}{q}\mathbf{E}\big[\,(X_f)\,(\mathcal{O}_k\tilde{X}_p+Y_p^s)^*\,\big] &= \frac{1}{q}\mathbf{E}\big[\,(A^kX_p^d+\mathcal{C}_k^dU_p+X_f^s)\,(\mathcal{O}_k\tilde{X}_p+Y_p^s)^*\big]\\ &= \frac{1}{q}\mathbf{E}\big[(A^k\tilde{X}_p+X_f^s)(\mathcal{O}_k\tilde{X}_p+Y_p^s)^*\big]\\ &= A^k\tilde{P}_0\mathcal{O}_k^*+\mathcal{C}_k^s,\end{aligned}$$

where $\tilde{P}_0=\frac{1}{q}\mathbf{E}[\tilde{X}_p\tilde{X}_p^*]$ and $q:=N-k-l+1$ is the number of filters in the Kalman filter bank. However, we also have

$$\begin{aligned}A^k\tilde{P}_0\mathcal{O}_k^*+\mathcal{C}_k^s &= \begin{bmatrix}A^k(\Sigma^s+\tilde{P}_0)C^*+\Sigma^{wv} & \cdots & A(\Sigma^s+A^{k-1}\tilde{P}_0A^{k-1})C^*+\Sigma^{wv}\end{bmatrix}\\ &= \begin{bmatrix}A^{k-1}\tilde{G}_0 & A^{k-2}\tilde{G}_1 & \cdots & \tilde{G}_{k-1}\end{bmatrix}.\end{aligned}$$

Finally, the above is of full row rank by Theorem 5.13. □

B.8 Proof of Theorem 5.15

Proof. In this proof, we shall drop the superscript $k+l$ when referring to data Hankel matrices of $k+l$ block rows; for instance, we shall write U instead of U^{k+l}. Moreover, without loss of generality, we may assume the system has a realization satisfying Equations (4.5), such that $\bar{x}_1(t)$ corresponds to the stochastically excited modes, $\bar{x}_2(t)$ to the output disturbance modes, and $\bar{x}_3(t)$ to the purely deterministic modes (see Lemma 4.4 for detail). Now, partition X_p and X_f as:

$$X_p=:\begin{bmatrix}X_{p1}\\X_{p2}\\X_{p3}\end{bmatrix}:=\begin{bmatrix}\bar{x}_1(0) & \bar{x}_1(1) & \cdots & \bar{x}_1(N-k-l)\\ \bar{x}_2(0) & \bar{x}_2(1) & \cdots & \bar{x}_2(N-k-l)\\ \bar{x}_3(0) & \bar{x}_3(1) & \cdots & \bar{x}_3(N-k-l)\end{bmatrix},$$

and

$$X_f=:\begin{bmatrix}X_{f1}\\X_{f2}\\X_{f3}\end{bmatrix}:=\begin{bmatrix}\bar{x}_1(k) & \bar{x}_1(k+1) & \cdots & \bar{x}_1(N-l)\\ \bar{x}_2(k) & \bar{x}_2(k+1) & \cdots & \bar{x}_2(N-l)\\ \bar{x}_3(k) & \bar{x}_3(k+1) & \cdots & \bar{x}_3(N-l)\end{bmatrix}.$$

Furthermore, we may assume without loss of generality that $k=3j$, where j is no smaller than the degree of the A-annihilator. Noticing that $Y_p^*=\begin{bmatrix}Y_{0,j-1}^* & Y_{j,3j-1}^*\end{bmatrix}$, it is equivalent to prove as indicated by Lemma 5.8 that

$$\mathbf{E}\left[\begin{pmatrix}X_f\\U_f\end{pmatrix}\begin{pmatrix}Y_p^* & U_p^* & U_f^*\end{pmatrix}\right]=\mathbf{E}\left[\begin{pmatrix}X_f\\U_f\end{pmatrix}\begin{pmatrix}Y_{0,j-1}^* & Y_{j,3j-1}^* & U_p^* & U_f^*\end{pmatrix}\right] \tag{B.8}$$

is of full row rank. Now let $\hat{X}_j:=\Pi_{(\mathcal{U}_p+\mathcal{U}_f)}X_j$ and $\tilde{X}_j:=X_j^d-\hat{X}_j$ such that

$$\Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}X_j=\tilde{X}_j+X_j^s \qquad \text{and} \qquad \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}Y_{j,3j-1}=\mathcal{O}_{2j}\tilde{X}_j+Y_{j,3j-1}^s.$$

Note further that $\mathcal{S}$ can be written as:

$$\begin{aligned}\mathcal{S} &:= \mathcal{U}_p + \mathcal{U}_f + \mathcal{Y}_{0,j-1} + \mathcal{Y}_{j,3j-1} \\ &= \mathcal{U}_p + \mathcal{U}_f + \mathcal{Y}_{0,j-1} + \Pi_{(\mathcal{U}_p+\mathcal{U}_f)^\perp}\mathcal{Y}_{j,3j-1}.\end{aligned}$$

As a result, the rank of Equation (B.8) is equivalent to that of

$$\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\begin{pmatrix} Y_{0,j-1}^* & U_p^* & U_f^* & [\mathcal{O}_{2j}\tilde{X}_j + Y_{j,3j-1}^s]^* \end{pmatrix}\right].$$

Now, we consider that the following two expressions:

$$\mathbf{E}\left[\begin{pmatrix} X_{f2} \\ X_{f3} \\ U_f \end{pmatrix}\begin{pmatrix} Y_{0,j-1}^* & U_p^* & U_f^* \end{pmatrix}\right], \qquad \text{and} \qquad \mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\left(\mathcal{O}_{2j}\tilde{X}_j + Y_{j,3j-1}^s\right)^*\right]. \tag{B.9}$$

(1): To show the first expression in (B.9) is of full row rank, we follow the proof of Theorem 5.10. Note that $Y_{0,j-1} - \mathcal{T}_k U_{0,j-1} = \mathcal{O}_k X_p + \mathcal{N}_k W_{0,j-1} + V_{0,j-1}$ and that $\mathcal{W}_{0,j-1}$ and $\mathcal{V}_{0,j-1}$ are orthogonal to $\mathcal{X}_{f2}$ and $\mathcal{X}_{f3}$. Thus, it is equivalent to show

$$\mathbf{E}\left[\begin{pmatrix} X_{f2} \\ X_{f3} \\ U_f \end{pmatrix}\begin{pmatrix} X_p^*\mathcal{O}_k^* & U_p^* & U_f^* \end{pmatrix}\right] = \begin{bmatrix} A_{OD}^k & \mathcal{C}_{OD,k}^d & 0 \\ 0 & 0 & I \end{bmatrix}\mathbf{E}\left[\begin{pmatrix} X_{p2} \\ X_{p3} \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}^*\right]\begin{bmatrix} \mathcal{O}_k^* & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \tag{B.10}$$

is of full row rank, where OD stands for 'Output Disturbance' and

$$A_{OD} = \begin{bmatrix} \bar{A}_{22} & \bar{A}_{23} \\ 0 & \bar{A}_{33} \end{bmatrix}, \quad B_{OD} = \begin{bmatrix} \bar{B}_2 \\ \bar{B}_3 \end{bmatrix}, \quad \mathcal{C}_{OD,k}^d = \begin{bmatrix} \bar{A}_{OD}^{k-1}\bar{B}_{OD} & \bar{A}_{OD}^{k-2}\bar{B}_{OD} & \cdots & \bar{B}_{OD} \end{bmatrix},$$

where B has an obvious partition $B^* = \begin{bmatrix} \bar{B}_1^* & \bar{B}_2^* & \bar{B}_3^* \end{bmatrix}$. Observe by Condition 2 that (A_{OD}, B_{OD}) is reachable; and hence by Theorem 5.5 we have

$$\mathbf{E}\left[\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_p \\ U_p \\ U_f \end{pmatrix}^*\right] > 0,$$

which implies that its principal submatrix is also positive definite:

$$\mathbf{E}\left[\begin{pmatrix} X_{p2} \\ X_{p3} \\ U_p \\ U_f \end{pmatrix}\begin{pmatrix} X_{p2} \\ X_{p3} \\ U_p \\ U_f \end{pmatrix}^*\right] > 0.$$

Now, it can readily be seen that Equation B.10 is of full row rank since the expression is a product of full row rank matrices.

(2): Next, we consider the second expression in (B.9). Following the proof of Theorem 5.14, we get

$$\frac{1}{q}\,\mathbf{E}\left[\begin{pmatrix} X_f \\ U_f \end{pmatrix}\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right] = \begin{bmatrix} A^{2j-1}\tilde{G}_0 & A^{2j-2}\tilde{G}_1 & \cdots & \tilde{G}_{2j-1} \\ 0 & 0 & \cdots & 0 \end{bmatrix}, \tag{B.11}$$

where $q := N-k-l+1$ is the number of filters in the Kalman filter bank. Now, let $p_{\min}(z) = z^j + a_{j-1}z^{j-1} + \cdots + a_0$ be an annihilating polynomial and

$$\Xi := \begin{bmatrix} a_0 I & & 0 \\ \vdots & \ddots & \\ a_{j-1}I & & a_0 I \\ I & \ddots & \vdots \\ & \ddots & a_{j-1}I \\ 0 & & I \end{bmatrix}.$$

Following the proof of Theorem 5.13, we obtain

$$\begin{aligned} &\begin{bmatrix} A^{2j-1}\tilde{G}_0 & A^{2j-2}\tilde{G}_1 & \cdots & \tilde{G}_{2j-1} \end{bmatrix}\Xi \\ &\qquad = (I + a_{j-1}A + \cdots + a_0 A^j)\begin{bmatrix} A^{j-1}G & A^{j-2}G & \cdots & G \end{bmatrix}. \end{aligned} \tag{B.12}$$

Furthermore, by Lemma 5.12 we have $(I+a_{j-1}A+\cdots+a_0A^j)$ non-singular. Thus, together with Condition 3 we have

$$\mathbf{E}\left[\left(X_{f1}\right)\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]\Xi$$

of full row rank. On the other hand, note that

$$A = \begin{bmatrix} \bar{A}_{11} & \bar{A}_{12} & \bar{A}_{13} \\ 0 & \bar{A}_{22} & \bar{A}_{23} \\ 0 & 0 & \bar{A}_{33} \end{bmatrix}, \qquad G = A\Sigma^s C^* + \Sigma^{wv} = \begin{bmatrix} \bar{A}_{11}\bar{\Sigma}^s_{11}\bar{C}^*_{11} + \bar{\Sigma}^{wv}_{11} & \bar{\Sigma}^{wv}_{12} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Using the above equations, it is easy to derive from Equation (B.11) and (B.12) that

$$\mathbf{E}\left[\begin{pmatrix} X_{f2} \\ X_{f3} \\ U_f \end{pmatrix}\left(\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}\right)^*\right]\Xi = 0.$$

Finally, we complete the proof by observing that

$$\begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & 0 \end{bmatrix} := \mathbf{E}\left[\begin{pmatrix} X_{f1} \\ X_{f2} \\ X_{f3} \\ U_f \end{pmatrix}\left(Y^*_{0,j-1} \quad U^*_p \quad U^*_f \quad [\mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1}]^*\right)\right]\begin{bmatrix} I & 0 \\ 0 & \Xi \end{bmatrix}$$

is of full row rank since

$$\Delta_{21} := \mathbf{E}\left[\begin{pmatrix} X_{f2} \\ X_{f3} \\ U_f \end{pmatrix} \begin{pmatrix} Y^*_{0,j-1} & U^*_p & U^*_f \end{pmatrix}\right]$$

is of full row rank from part (1) and

$$\Delta_{12} := \mathbf{E}\left[\begin{pmatrix} X_{f1} \end{pmatrix} \begin{pmatrix} \mathcal{O}_{2j}\tilde{X}_j + Y^s_{j,3j-1} \end{pmatrix}^*\right] \Xi$$

is of full row rank from part (2). $\square$

C

Proofs for Chapter 6

C.1 Proof of Theorem 6.2

Proof. We will prove the statement by constructing a set of realizations producing $\mathscr{X}^N$ with the first $\hbar$ Markov parameters being the same, if u is S.R. of order less than $k-\hbar+1$. The proof will be structured as follows:

1. define a set of realizations;
2. show that there exist two realizations in this set which are not equivalent;
3. show that this set of realizations has the same $\hbar$ Markov parameters;
4. show that $\mathscr{X}^N$ admits the output equation of this set of realizations;
5. show that $\mathscr{X}^N$ admits the state equation of this set of realizations.

Construct U^{k+1} and Y^{k+1} from the signals u and y, respectively. Here, we shall drop the superscript $k+1$ without cause of confusion.

(1): Since u is S.R. of order less than $k-\hbar+1$, there exist by definition $\psi_0, \psi_1, \ldots, \psi_{k-\hbar} \in \mathbb{R}^m$, such that $\begin{bmatrix}\psi_0^* & \psi_1^* & \cdots & \psi_{k-\hbar}^*\end{bmatrix} \neq 0$ and

$$\psi_0^* U_0 + \psi_1^* U_1 + \cdots + \psi_{k-\hbar}^* U_{k-\hbar} = 0.$$

Define $\psi \in \mathbb{R}^{(k+1)m}$ such that $\psi^* := \begin{bmatrix}\psi_0^* & \psi_1^* & \cdots & \psi_k^*\end{bmatrix} \neq 0$ and

$$\psi_i^* = 0, \qquad \text{for} \quad i > k-\hbar. \tag{C.1}$$

Thus, it is easy to see that $\psi^* U = 0$. Furthermore, define $\phi \in \mathbb{R}^{n\times m}$, $\phi \neq 0$ such that

$$\mathcal{O}_{k-1}\phi = 0, \qquad \text{and} \qquad CA^{k-1}\phi \neq 0. \tag{C.2}$$

(It is easily seen that such a ϕ exists since $\mathcal{O}_{k-1}$ is not injective by definition. In addition, since $\mathcal{O}_k$ is injective by definition, we have $\mathcal{O}_k\phi \neq 0$ if and only if $\phi \neq 0$. In other words, $\mathcal{O}_{k-1}\phi = 0$ and $\phi \neq 0$ imply that $CA^{k-1}\phi \neq 0$.)

The set of realizations is given by $(A, \hat{B}, C, \hat{D})$ with initial condition $\hat{X}_0 := X_0 - \tilde{X}_0$, where $\hat{B} := B + \tilde{B}$, $\hat{D} := D + \tilde{D}$, and

$$\tilde{B} := A^k \,\phi\psi_k^* + A^{k-1}\,\phi\psi_{k-1}^* + \cdots + \phi\,\psi_0^*, \tag{C.3}$$

$$\tilde{D} := CA^{k-1}\,\phi\psi_k^*, \tag{C.4}$$

$$\tilde{X}_0 := \sum_{j=0}^{k-1} \sum_{i=0}^{k-1-j} A^{k-1-i-j}\,\phi\psi_{k-i}^*\,U_j. \tag{C.5}$$

(2): To show there are more than one element in the set defined above, it is sufficient to show the differences in the first $k+1$ Markov parameters are not identical to zero:

$$\begin{bmatrix}\tilde{h}_0^* & \tilde{h}_1^* & \cdots & \tilde{h}_k^*\end{bmatrix} = \begin{bmatrix}\tilde{D}^* & (C\tilde{B})^* & \cdots & (CA^{k-1}\tilde{B})^*\end{bmatrix} \neq 0.$$

Define $\tilde{\phi} := A^{k-1}\phi$ such that $C\tilde{\phi} \neq 0$. Since $C\phi = CA\phi = \cdots = CA^{k-2}\phi = 0$ by (C.2), it is easy to see that

$$\begin{bmatrix}\tilde{D}\\ C\tilde{B}\\ \vdots\\ CA^{k-1}\tilde{B}\end{bmatrix} = \begin{bmatrix} C\tilde{\phi} & 0 & \cdots & 0\\ CA\tilde{\phi} & \ddots & \ddots & \vdots\\ \vdots & \ddots & C\tilde{\phi} & 0\\ CA^k\tilde{\phi} & \cdots & CA\tilde{\phi} & C\tilde{\phi}\end{bmatrix} \begin{bmatrix}\psi_k^*\\ \vdots\\ \psi_1^*\\ \psi_0^*\end{bmatrix}. \tag{C.6}$$

Therefore, we shall show the above matrix is not identical to zero.

Now, define $P_0, P_1, \ldots, P_{k-1} \in \mathbb{R}^{p\times p}$ such that

$$P_0C + P_1CA + \cdots + P_{k-1}CA^{k-1} + CA^k = 0. \tag{C.7}$$

Note that this is possible as each row of CA^k can be represented by the combinations of the rows of $C, CA, \ldots, CA^{k-1}$ since $\mathcal{O}_k$ is injective by definition. Define H as:

$$H := \begin{bmatrix}H_0\\ H_1\\ \vdots\\ H_k\end{bmatrix} := \begin{bmatrix} P_0 & P_1 & \cdots & P_{k-1} & I\\ P_1 & P_2 & \cdots & I & 0\\ \vdots & \vdots & ⋰ & \vdots & \vdots\\ I & 0 & \cdots & 0 & 0\end{bmatrix}.$$

From this, we have

$$H\mathcal{O}_{k+1}\tilde{\phi} = H\begin{bmatrix} C\tilde{\phi} \\ CA\tilde{\phi} \\ \vdots \\ CA^k\tilde{\phi} \end{bmatrix} = \begin{bmatrix} P_0C\tilde{\phi} + P_1CA\tilde{\phi} + \cdots + CA^k\tilde{\phi} \\ P_1C\tilde{\phi} + P_2CA\tilde{\phi} + \cdots + CA^{k-1}\tilde{\phi} \\ \vdots \\ C\tilde{\phi} \end{bmatrix},$$

such that for $j = 0, \ldots, k$,

$$H_j\mathcal{O}_{k+1}\tilde{\phi} = P_jC\tilde{\phi} + P_{j+1}CA\tilde{\phi} + \cdots + CA^{k-j}\tilde{\phi}.$$

We will then show $H_j\mathcal{O}_{k+1}\tilde{\phi} = 0$ for $j = 0, \ldots, k-1$. First of all, it is easy to see that $H_0\mathcal{O}_{k+1}\tilde{\phi} = 0$ by Equation (C.7). Now, for $j = 1, \ldots, k-1$, we first note that

$$P_0CA^{k-1-j}\phi + P_1CA^{k-j}\phi + \cdots + P_{j-1}CA^{k-2}\phi = 0,$$

since $C\phi = \cdots = CA^{k-2}\phi = 0$. Adding these terms to $H_j\mathcal{O}_{k+1}\tilde{\phi}$ reveals

$$\begin{aligned} H_j\mathcal{O}_{k+1}\tilde{\phi} &= (P_0CA^{k-1-j}\phi + \cdots + P_{j-1}CA^{k-2}\phi) + (P_jCA^{k-1}\phi + \cdots + CA^{2k-1-j}\phi) \\ &= (P_0C + P_1CA + \cdots + CA^k)\ A^{k-1-j}\phi \\ &= 0 \end{aligned}$$

by Equation (C.7). Since H is in block Hankel form, it can be seen that

$$H\begin{bmatrix} C\tilde{\phi} & 0 & \cdots & 0 \\ CA\tilde{\phi} & \ddots & \ddots & \vdots \\ \vdots & \ddots & C\tilde{\phi} & 0 \\ CA^k\tilde{\phi} & \cdots & CA\tilde{\phi} & C\tilde{\phi} \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 & C\tilde{\phi} \\ 0 & \cdots & C\tilde{\phi} & 0 \\ \vdots & \iddots & \iddots & \vdots \\ C\tilde{\phi} & 0 & \cdots & 0 \end{bmatrix}.$$

Now, post-multiplying the equation by $\begin{bmatrix}\psi_k & \cdots & \psi_0\end{bmatrix}^*$ gives

$$H\begin{bmatrix} C\tilde{\phi} & 0 & \cdots & 0 \\ CA\tilde{\phi} & \ddots & \ddots & \vdots \\ \vdots & \ddots & C\tilde{\phi} & 0 \\ CA^k\tilde{\phi} & \cdots & CA\tilde{\phi} & C\tilde{\phi} \end{bmatrix}\begin{bmatrix} \psi_k^* \\ \vdots \\ \psi_1^* \\ \psi_0^* \end{bmatrix} = \begin{bmatrix} C\tilde{\phi}\psi_0^* \\ C\tilde{\phi}\psi_1^* \\ \vdots \\ C\tilde{\phi}\psi_k^* \end{bmatrix} \neq 0. \tag{C.8}$$

The above equation is not zero since $C\tilde{\phi} = CA^{k-1}\phi \neq 0$ and $\psi \neq 0$ by assumption. Finally, substituting Equation (C.6) into (C.8) gives

$$H\begin{bmatrix}\tilde{D}^* & (C\tilde{B})^* & \cdots & (CA^{k-1}\tilde{B})^*\end{bmatrix} \neq 0,$$

which implies

$$\begin{bmatrix}\tilde{D}^* & (C\tilde{B})^* & \cdots & (CA^{k-1}\tilde{B})^*\end{bmatrix} \neq 0.$$

(3): Note that

$$\begin{bmatrix} \tilde{h}_0 \\ \tilde{h}_1 \\ \vdots \\ \tilde{h}_k \end{bmatrix} = \begin{bmatrix} \tilde{D} \\ C\tilde{B} \\ \vdots \\ CA^{k-1}\tilde{B} \end{bmatrix} = \begin{bmatrix} CA^{k-1}\,\phi\psi_k^* \\ CA^k\,\phi\psi_k^* + CA^{k-1}\,\phi\psi_{k-1}^* + \cdots + C\,\phi\psi_0^* \\ \vdots \\ CA^{2k-1}\,\phi\psi_k^* + CA^{2k-2}\,\phi\psi_{k-1}^* + \cdots + CA^{k-1}\,\phi\psi_0^* \end{bmatrix}$$

$$= \begin{bmatrix} C\,\tilde{\phi}\psi_k^* \\ CA\,\tilde{\phi}\psi_k^* + C\,\tilde{\phi}\psi_{k-1}^* \\ \vdots \\ CA^k\,\tilde{\phi}\psi_k^* + \cdots + C\,\tilde{\phi}\psi_0^* \end{bmatrix},$$

where the last equality stems from the fact that $C\phi = \cdots = CA^{k-2}\phi = 0$. Furthermore, since as in Equation (C.1),

$$\psi_i^* = 0, \qquad \text{for} \quad i > k - \hbar,$$

it is now easy to see that

$$\tilde{h}_0 = \cdots = \tilde{h}_{\hbar-1} = 0.$$

(4): Let

$$\tilde{\mathcal{T}}_{k+1} = \begin{bmatrix} \tilde{D} & 0 & 0 \\ \vdots & \ddots & 0 \\ CA^{k-1}\tilde{B} & \cdots & \tilde{D} \end{bmatrix}, \qquad \hat{\mathcal{T}}_{k+1} = \begin{bmatrix} \hat{D} & 0 & 0 \\ \vdots & \ddots & 0 \\ CA^{k-1}\hat{B} & \cdots & \hat{D} \end{bmatrix}. \tag{C.9}$$

Multiplying CA^s, $s = 0, \ldots, k$ to $\tilde{X}_0$ gives

$$CA^s\tilde{X}_0 = \sum_{j=0}^{s}\sum_{i=0}^{s-j} CA^{s-i-j}\,\tilde{\phi}\,\psi_{k-i}^*\,U_j$$

where terms beginning with $C\phi, CA\phi, \ldots CA^{k-2}\phi$ vanish. Finally, realizing that

$$\sum_{i=0}^{s-j} CA^{s-i-j}\,\tilde{\phi}\,\psi_{k-i}^* = \begin{cases} \tilde{D} & s-j=0, \\ CA^{s-j-1}\tilde{B} & s-j = 1,\ldots,k; \end{cases}$$

we have

$$\mathcal{O}_{k+1}\tilde{X}_0 = \tilde{\mathcal{T}}_{k+1}U_{0,k}.$$

With this, we then obtain

$$\begin{aligned} Y_{0,k} &= \mathcal{O}_{k+1}X_0 + \mathcal{T}_{k+1}U_{0,k} \\ &= \mathcal{O}_{k+1}\,(X_0 - \tilde{X}_0) + (\mathcal{T}_{k+1} + \tilde{\mathcal{T}}_{k+1})\,U_{0,k} \\ &= \mathcal{O}_{k+1}\hat{X}_0 + \hat{\mathcal{T}}_{k+1}U_{0,k}. \end{aligned}$$

Suppose that $\hat{X}_0$ is a valid state sequence, i.e., $\hat{x}(t+1) = A\hat{x}(t) + \hat{B}u(t)$, then it is easy to see that $y(t) = C\hat{x}(t) + \hat{D}u(t)$.

(5): It remains to check the validity of $\hat{X}_0$ as a state sequence. Now, by considering $Y_0, \ldots, Y_k$ entry-wise, it is evident that the first column of $Y_{0,k}$ is equal to the second column with each entry shifted up by one block row. This gives

$$\mathcal{O}_k A\hat{x}(t) + \begin{bmatrix} C\hat{B} & \hat{D} & & 0 \\ \vdots & & \ddots & \\ CA^{k-1}\hat{B} & \ldots & \ldots & \hat{D} \end{bmatrix} \begin{bmatrix} u(t) \\ \vdots \\ u(t+k) \end{bmatrix} = \mathcal{O}_k \hat{x}(t+1) + \hat{\mathcal{T}}_k \begin{bmatrix} u(t+1) \\ \vdots \\ u(t+k) \end{bmatrix},$$

which implies $\mathcal{O}_k\big(A\hat{x}(t) + \hat{B}u(t)\big) = \mathcal{O}_k\hat{x}(t+1)$, or equivalently, $\hat{x}(t+1) = A\hat{x}(t) + \hat{B}u(t)$.

This completes the proof. □

C.2 Proof of Theorem 6.5

Proof. (2) $\Rightarrow$ (1): We will follow Algorithm 6.6 to find A, B, C, and D with $l = k+1$. To simplify the notation, the superscript $k+1$ is dropped in the proof. First of all, it is easy to derive

$$\Pi_{\mathcal{U}^{\perp}_{0,k-\hbar}} Z_{0,k} = \mathcal{O}_{k+1} \Pi_{\mathcal{U}^{\perp}_{0,k-\hbar}} X_0.$$

Since the above equation is guaranteed to be n-dimensional by Conditions 2, A and C are uniquely determined (up to a similarity transformation). Moreover, since $Z_{0,k}$ depends only on $U_{0,k-\hbar}$, Equation 6.3 is obtained. Then, Equation 6.4 can easily be verified by noting that it is the first l–$\hbar$–1 block rows of Equation 5.10 of Algorithm 5.1, which can also be used to determine the Markov parameters $h_{\hbar}, \ldots, h_k$. Note that since $\mathcal{O}_k$ is injective, a unique B is determined. Obviously, D is given by h_0.

(1) $\Rightarrow$ (2): This part is proven by constructing a set of realizations with the first $\hbar$ Markov parameters being the same, if one of Conditions 2 is not satisfied. Note that the case where u is S.R. of order less than $k - \hbar + 1$ is proven by Theorem 6.1. Therefore, we will show the remaining case where $\Pi_{\mathcal{U}^{\perp}_{0,k-\hbar}} \mathcal{Z}_{0,k}$ is less than n-dimensional. Moreover, the case where $\dim\{\mathcal{X}_0\} < n$ is shown in Lemma 6.1. Thus, it remains to show the case where $\dim\{\mathcal{X}_0\} = n$ but $\mathcal{X}_0 \cap \mathcal{U}_{0,k-\hbar} \neq \{0\}$. The rest of the proof is structured as follows:

a. define a set of realizations;

b. show that this set of realizations has the same $\hbar$ Markov parameters;

c. show that there exist two realizations in this set which are not equivalent;

d. show that $\mathscr{Z}^N$ admits this set of realizations.

(a): First of all, partition X_0 as

$$X_0^* =: \begin{bmatrix}\chi_1^* & \chi_2^* & \cdots & \chi_n^*\end{bmatrix} =: \begin{bmatrix}\chi_{1,n-1}^* & \chi_n^*\end{bmatrix}.$$

As it is assumed that $\mathcal{X}_0 \cap \mathcal{U}_{0,k-\hbar} \neq \{0\}$, we may assume further without loss of generality that there exist constants $\psi_0, \psi_1, \ldots, \psi_{k-\hbar} \in \mathbb{R}^m$ and $\varphi \in \mathbb{R}^{n-1}$ such that

$$\chi_n = \begin{bmatrix}\psi_0^* & \psi_1^* & \cdots & \psi_{k-\hbar}^*\end{bmatrix} U_{0,k-\hbar} + \varphi^* \chi_{1,n-1}.$$

(Note that this is valid since there always exists a state transformation to obtain the above.)

Moreover, let $\psi^* := \begin{bmatrix}\psi_0^* & \psi_1^* & \cdots & \psi_k^*\end{bmatrix}$, where $\psi_{k-\hbar+1} = \psi_{k-\hbar+2} = \psi_k = 0 \in \mathbb{R}^m$. With this definition of ψ, it is easy to see that

$$\begin{bmatrix}\chi_{1,n-1} \\ \chi_n\end{bmatrix} = \left[\begin{array}{ccc|c} 0 & \cdots & 0 & I \\ \psi_0^* & \cdots & \psi_{k-\hbar}^* & \varphi^* \end{array}\right] \left[\begin{array}{c} U_{0,k-\hbar} \\ \hline \chi_{1,n-1} \end{array}\right] = \begin{bmatrix} 0 & I \\ \psi^* & \varphi^* \end{bmatrix} \begin{bmatrix} U_{0,k} \\ \chi_{1,n-1} \end{bmatrix}. \tag{C.10}$$

Note that since it is assumed that $\dim\{\mathcal{X}_0\}{=}n$, this implies $\psi \neq 0$. Denote by s, $0 \leq s \leq k-\hbar$, the position of the element of ψ such that

$$\psi_s \neq 0 \qquad \text{and} \qquad \psi_{s+1} = \psi_{s+2} = \cdots = \psi_k = 0.$$

Define $\phi \in \mathbb{R}^n$, $\phi \neq 0$, such that $\mathcal{O}_{k-1}\phi = 0$. Note that such a ϕ exists since $\mathcal{O}_{k-1}$ is not injective by definition. As a result, we have

$$CA^{k-1}\phi \neq 0 \qquad \text{and} \qquad CA^{k-1}\,\phi\,\psi_s^* \neq 0. \tag{C.11}$$

In addition, define the shorthands, for $0 \leq i \leq k$,

$$\Psi_i := \begin{bmatrix}0 & \cdots & 0 & \psi_i\end{bmatrix} \in \mathbb{R}^{m\times n},$$

and $\Psi^* := \begin{bmatrix}\Psi_1^* & \Psi_2^* & \cdots & \Psi_k^*\end{bmatrix} = \begin{bmatrix}0 & \psi\end{bmatrix}^*$. Therefore, Equation (C.10) can be written as:

$$\begin{bmatrix}\chi_{1,n-1} \\ \chi_n\end{bmatrix} = \left[\begin{array}{c|c} \Psi^* & \begin{matrix} I \\ \varphi^* \end{matrix} \end{array}\right] \left[\begin{array}{c} U_{0,k} \\ \hline \chi_{1,n-1} \end{array}\right]. \tag{C.12}$$

The set of realizations is given by $(\hat{A}, \hat{B}, C, \hat{D})$ with initial condition $\hat{X}_0$, where $\hat{A} := A+\tilde{A}$,

$\hat{B} := B - \tilde{B}$, $\hat{D} := D - \tilde{D}$, and

$$\tilde{A} := \begin{bmatrix} -\phi\,\varphi^* & \phi \end{bmatrix}, \tag{C.13}$$

$$\tilde{B} := \hat{A}^s\,\phi\,\psi_s^* + \hat{A}^{s-1}\,\phi\,\psi_{s-1}^* + \cdots + \phi\,\psi_0^*, \tag{C.14}$$

$$\tilde{D} := \begin{cases} 0 & \text{if } 0 \le s \le k-1, \\ C\hat{A}^{k-1}\,\phi\,\psi_k^* & \text{if } s = k, \end{cases} \tag{C.15}$$

$$\hat{X}_0 := \left[\begin{array}{c|c} \tilde{\Psi}^* & \begin{matrix} I \\ \varphi^* \end{matrix} \end{array}\right] \begin{bmatrix} U_{0,k} \\ \hline \chi_{1,n-1} \end{bmatrix}, \tag{C.16}$$

with $\tilde{\Psi}^* = \begin{bmatrix} \tilde{\Psi}_0^* & \tilde{\Psi}_1^* & \cdots & \tilde{\Psi}_k^* \end{bmatrix}$ defined as

$$\tilde{\Psi}_i^* = \begin{cases} \hat{A}^{s-i-1}\tilde{A}\,\Psi_s^* + \cdots + \tilde{A}\,\Psi_{i+1}^* + \Psi_i^* & \text{if } 0 \le i \le s-1, \\ \Psi_i^* & \text{if } s \le i \le k. \end{cases}$$

(b): Owing to the fact that $\mathcal{O}_{k-1}\phi = 0$, we have

$$C\hat{A}^j = CA^{i-1}(A + \tilde{A})\hat{A}^{j-i} = CA^j, \qquad \text{for} \quad 0 \le i \le j \le k-1. \tag{C.17}$$

Now, construct $\tilde{h}_i$ from $\hat{A}$, $\tilde{B}$, C, and $\tilde{D}$ and Equation (C.17). It can be seen that

$$\tilde{h}_i = \begin{cases} 0 & \text{if } i < k-s, \\ C\hat{A}^{s+i-1}\,\phi\,\psi_s^* + C\hat{A}^{s+i-2}\,\phi\,\psi_{s-1}^* + \cdots + C\hat{A}^{k-1}\,\phi\,\psi_{k-i}^* & \text{if } i \ge k-s. \end{cases}$$

Recall that $0 \le s \le k - \hbar$, which ensures that

$$\tilde{h}_0 = \tilde{h}_1 = \cdots = \tilde{h}_{\hbar-1} = 0.$$

(c): To show the realization $(\hat{A}, \hat{B}, C, \hat{D})$ is not equivalent to (A, B, C, D) under a similarity transformation, we will show $h_{k-s} \neq 0$. Note that

$$h_{k-s} = C\hat{A}^{k-1}\,\phi\,\psi_s^* = CA^{k-1}\,\phi\,\psi_s^* \neq 0,$$

by Equations (C.17) and (C.11).

(d): Here, we will first show $\mathcal{O}_{k+1}X_0 + \tilde{\mathcal{T}}_{k+1}U_{0,k} = \hat{\mathcal{O}}_{k+1}\hat{X}_0$, where

$$\hat{\mathcal{O}}_{k+1} := \begin{bmatrix} C \\ \vdots \\ C\hat{A}^k \end{bmatrix} \qquad \text{and} \qquad \tilde{\mathcal{T}}_{k+1} := \begin{bmatrix} \tilde{h}_0 & 0 & 0 \\ \vdots & \ddots & 0 \\ \tilde{h}_k & \cdots & \tilde{h}_0 \end{bmatrix} := \begin{bmatrix} \tilde{D} & 0 & 0 \\ \vdots & \ddots & 0 \\ C\hat{A}^{k-1}\tilde{B} & \cdots & \tilde{D} \end{bmatrix}.$$

With the use of Equation (C.12), it is sufficient to show

$$\mathcal{O}_{k+1}\left[\begin{array}{c:c} \Psi^* & \begin{array}{c} I \\ \varphi^* \end{array} \end{array}\right] + \left[\begin{array}{c:c} \tilde{\mathcal{T}}_{k+1} & 0 \end{array}\right] = \hat{\mathcal{O}}_{k+1}\left[\begin{array}{c:c} \tilde{\Psi}^* & \begin{array}{c} I \\ \varphi^* \end{array} \end{array}\right].$$

Note that we have $\mathcal{O}_k = \hat{\mathcal{O}}_k$ by Equation (C.17) and $\tilde{A}\begin{bmatrix} I & \varphi \end{bmatrix}^* = -\phi\varphi^* + \phi\varphi^* = 0$ by Equation (C.13). It is then easy to see that $\mathcal{O}_k \begin{bmatrix} I & \varphi \end{bmatrix}^* = \hat{\mathcal{O}}_{k+1}\begin{bmatrix} I & \varphi \end{bmatrix}^*$. Thus, we will only need to show

$$\mathcal{O}_{k+1}\Psi^* + \tilde{\mathcal{T}}_{k+1} = \hat{\mathcal{O}}_{k+1}\tilde{\Psi}^*. \tag{C.18}$$

To show Equation (C.18), we will show the equation column by column. First for $s+1 \leq i \leq k$, observe that $\Psi_i = \tilde{\Psi}_i = 0$ and $\tilde{h}_0 = \cdots = \tilde{h}_{k-s-1} = 0$. Thus,

$$\mathcal{O}_{k+1}\Psi_i^* + \begin{bmatrix} 0 & \tilde{h}_0^* & \cdots & \tilde{h}_{k-i}^* \end{bmatrix}^* = \hat{\mathcal{O}}_{k+1}\tilde{\Psi}_i^*.$$

For $i = s$, observe that $\Psi_s = \tilde{\Psi}_s$ and $\tilde{h}_{k-s} = C\hat{A}^{k-1}\,\phi\,\psi_s^* = C\hat{A}^{k-1}\tilde{A}\;\Psi_s^*$. Thus,

$$\mathcal{O}_{k+1}\Psi_i^* + \begin{bmatrix} 0 & \tilde{h}_{k-s}^* \end{bmatrix}^* = \hat{\mathcal{O}}_{k+1}\tilde{\Psi}_i^*.$$

Now for $i = 0, 1, \ldots, s-1$, observe that $CA^{k-1} = C\hat{A}^{k-1}$ and $C\phi = \cdots = CA^{k-2}\phi = 0$. Thus,

$$\begin{aligned}
\mathcal{O}_{k+1}\Psi_i^* + \begin{bmatrix} 0 \\ h_{k-s} \\ h_{k-s+1} \\ \vdots \\ h_{k-i} \end{bmatrix} &= \hat{\mathcal{O}}_{k+1}\Psi_i^* + \begin{bmatrix} & 0 & \\ C\hat{A}^{k-1}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k-s+i-1}\,\phi\,\psi_i^* \\ C\hat{A}^{k}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k-s+i}\,\phi\,\psi_i^* \\ & \vdots & \\ C\hat{A}^{k+s-i-1}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k-1}\,\phi\,\psi_i^* \end{bmatrix} \\
&= \hat{\mathcal{O}}_{k+1}\Psi_i^* + \begin{bmatrix} \hat{\mathcal{O}}_{k-s+i}\hat{A}^{s-i-1}\,\phi\,\psi_s^* & +\cdots+ & \hat{\mathcal{O}}_{k-s+i}\,\phi\,\psi_{i+1}^* \\ C\hat{A}^{k-1}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k-s+i}\,\phi\,\psi_{i+1}^* \\ C\hat{A}^{k}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k-s+i+1}\,\phi\,\psi_{i+1}^* \\ & \vdots & \\ C\hat{A}^{k+s-i-1}\,\phi\,\psi_s^* & +\cdots+ & C\hat{A}^{k}\,\phi\,\psi_{i+1}^* \end{bmatrix} \\
&= \hat{\mathcal{O}}_{k+1}\left(\hat{A}^{s-i-1}\,\tilde{A}\,\Psi_s^* + \cdots + \tilde{A}\,\Psi_{i+1}^* + \Psi_i^*\right) \\
&= \hat{\mathcal{O}}_{k+1}\tilde{\Psi}_i^*.
\end{aligned}$$

Therefore, we have proven $\mathcal{O}_{k+1}X_0 + \tilde{\mathcal{T}}_{k+1}U_{0,k} = \hat{\mathcal{O}}_{k+1}\hat{X}_0$, which gives

$$\begin{aligned}
Y_{0,k} &= \mathcal{O}_{k+1}X(0) + \mathcal{T}_{k+1}U_{0,k} \\
&= \hat{\mathcal{O}}_{k+1}\hat{X}(0) + \left(\mathcal{T}_{k+1} - \tilde{\mathcal{T}}_{k+1}\right)U_{0,k}.
\end{aligned}$$

Following Steps 4 and 5 in the proof of Theorem 6.1, it can be verified that $\mathscr{X}^N$ satisfies

$$\begin{aligned} \hat{x}(t+1) &= \hat{A}\hat{x}(t+1) + \hat{B}u(t+1), \\ y(t) &= C\hat{x}(t+1) + \hat{D}u(t+1). \end{aligned}$$

This completes the proof. $\square$

D

Proofs for Chapter 8

D.1 Proof of Theorem 8.5

Before we prove Theorem 8.5, we shall prove the following lemma.

Lemma D.1. *Let $P, S \in \mathbb{R}^{n\times n}$, $P = P^T > 0$, $S = S^T \geq 0$. Then,*

$$\frac{\rho(P^{-1}S)}{1+\rho(P^{-1}S)} \leq \bar{\sigma}\left(I-(I+P^{-1}S)^{-1}\right) \leq \frac{\rho(P^{-1}S)}{1+\rho(P^{-1}S)}\sqrt{\kappa(P)}, \tag{D.1}$$

where $\kappa(P)$ is the 2-norm condition number of P. Furthermore, the first equality holds if P^{-1} and S commute.

Proof. (Theorem 8.5) Note that $\lambda(P^{-1}S) = \lambda(P^{-\frac{1}{2}}SP^{-\frac{1}{2}}) \geq 0$. Therefore, the first inequality is straight forward.

$$\begin{aligned}
\bar{\sigma}[I-(I+P^{-1}S)^{-1}] &= \bar{\sigma}\left[P^{-\frac{1}{2}}\left[I-(I+P^{-\frac{1}{2}}SP^{-\frac{1}{2}})^{-1}\right]P^{\frac{1}{2}}\right] \\
&\leq \bar{\sigma}\left[I-(I+P^{-\frac{1}{2}}SP^{-\frac{1}{2}})^{-1}\right]\rho(P^{-\frac{1}{2}})\rho(P^{\frac{1}{2}}) \\
&= \rho\left[I-(I+P^{-\frac{1}{2}}SP^{-\frac{1}{2}})^{-1}\right]\sqrt{\frac{\overline{\lambda}(P)}{\underline{\lambda}(P)}} \\
&= \frac{\rho(P^{-1}S)}{1+\rho(P^{-1}S)}\sqrt{\kappa(P)}.
\end{aligned}$$

Equality holds if $P^{-1}S$ is Hermitian, in which case its singular values equal the absolute values of its eigenvalues. □

Proof. To reveal the eigenspace structure of the perturbed matrix, first convert $\hat{A}$ into Jordan canonical form by a similarity transformation. Let $T = \begin{bmatrix} \nu \\ * \end{bmatrix}$ be a non-singular matrix containing the left eigenvectors and generalized eigenvectors of $\hat{A}$ such that

$$T\hat{A}T^{-1} = \begin{bmatrix} \lambda & 0 \\ 0 & \Lambda \end{bmatrix} + \aleph,$$

where Λ is a diagonal matrix of eigenvalues and $\aleph$ is a matrix with all zero entries except some possible 1's immediately below the diagonal entries. This is just the same as transforming Γ to ΓT^{-1} and $\Gamma^{\uparrow}$ to $\Gamma^{\uparrow} T^{-1}$. So, $\gamma_i T^{-1} = c \begin{bmatrix} 1 & 0 \end{bmatrix}$. Let $\Delta_{11} \in \mathbb{C}$ and $\Delta_{21} \in \mathbb{C}^{n-1 \times 1}$ such that

$$(T^{-*}\Gamma^*\Gamma T^{-1})^{-1} = \begin{bmatrix} \Delta_{11} & * \\ \Delta_{21} & * \end{bmatrix},$$

where Δ_{11} can also be written as $\Delta_{11} = \nu(\Gamma^T\Gamma)^{-1}\nu^*$.

(1 and 2): The eigenvalues of $\hat{A}_D$ can be obtained from

$$\begin{aligned}
T\hat{A}_D T^{-1} &= T(\Gamma^*\Gamma + c^2\nu^*\nu)^{-1}(\Gamma^*\Gamma^{\uparrow} + bc\nu^*\nu\hat{A})T^{-1} \\
&= \left(I + c^2 T(\Gamma^*\Gamma)^{-1}T^* \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right)^{-1} \left(I + bcT(\Gamma^*\Gamma)^{-1}T^* \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) T\hat{A}T^{-1} \\
&= \begin{bmatrix} 1 + c^2\Delta_{11} & 0 \\ c^2\Delta_{21} & I \end{bmatrix}^{-1} \begin{bmatrix} 1 + bc\Delta_{11} & 0 \\ bc\Delta_{21} & I \end{bmatrix} \left(\begin{bmatrix} \lambda & 0 \\ 0 & \Lambda \end{bmatrix} + \aleph \right) \qquad \text{(D.2)} \\
&= \begin{bmatrix} \frac{1+bc\Delta_{11}}{1+c^2\Delta_{11}}\lambda & 0 \\ \frac{c(b-c)\Delta_{21}}{1+c^2\Delta_{11}}\lambda & \Lambda \end{bmatrix} + \aleph. \qquad \text{(D.3)}
\end{aligned}$$

Note that T is not a real matrix if some eigenvalues of $\hat{A}$ are complex. Also, the inverse in Equation (D.2) exists because $(T^{-*}\Gamma^*\Gamma T^{-1})^{-1}$ is positive definite, which implies $\Delta_{11} > 0$.

(3): From Equation (8.8), we can derive b in terms of λ, λ_D, and c as

$$b = \frac{(1 + c^2\Delta_{11})\lambda_D - \lambda}{c\Delta_{11}\lambda}.$$

Now, we have

$$\frac{c(b-c)\Delta_{21}}{1 + c^2\Delta_{11}} = \frac{\Delta_{21}(\lambda_D - \lambda)}{\Delta_{11}\lambda}.$$

Substituting this into Equation (D.3) gives us

$$T\hat{A}_D T^{-1} = \begin{bmatrix} \lambda_D & 0 \\ \frac{\Delta_{21}(\lambda_D - \lambda)}{\Delta_{11}\lambda} & \Lambda \end{bmatrix} + \aleph,$$

which is independent of b and c.

(4): By Lemma D.1, we have

$$\begin{aligned}\|\hat{A}_D - \hat{A}\|_2 &= \|c(b-c)(\Gamma^*\Gamma + c^2\nu^*\nu)^{-1}\nu^*\nu\hat{A}\|_2. \\ &\le |c(b-c)|\frac{\rho(\Delta_{11})}{1+c^2\rho(\Delta_{11})}\sqrt{\kappa(\Gamma^*\Gamma)}\ \|\hat{A}\|_2.\end{aligned}$$

Then the statement can be seen by noting that

$$c(b-c) = \frac{(1+c^2\Delta_{11})(\lambda_D - \lambda)}{\Delta_{11}\lambda},$$

where $\Delta_{11} = \nu(\Gamma^*\Gamma)^{-1}\nu^*$.

This completes the proof. □

D.2 Proof of Theorem 8.6

Proof. First note that $\nu^*\nu + \bar{\nu}^*\bar{\nu}$, and thus, $(\Gamma_D^*\Gamma_D)^{-1} = (\Gamma^*\Gamma + c^2(\nu^*\nu + \bar{\nu}^*\bar{\nu}))^{-1}$ and $\Gamma_D^*\Gamma_D^\uparrow = \Gamma^*\Gamma^\uparrow + bc(\nu^*\nu + \bar{\nu}^*\bar{\nu})\hat{A}$ are real. That is, $\hat{A}_D$ is also real. Let T be the non-singular matrix of eigenvectors and generalized eigenvectors of $\hat{A}$ with ν and $\bar{\nu}$ in the first two rows. To reveal the eigenspace structure of $\hat{A}_D$, we decompose $\hat{A}$ into

$$T\hat{A}T^{-1} = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \bar{\lambda} & 0 \\ 0 & 0 & \Lambda \end{bmatrix} + \aleph,$$

where $\aleph$ has only some 1's in the strictly lower triangular entries. Let $\Delta_{11} \in \mathbb{C}^{2\times 2}$ and $\Delta_{21} \in \mathbb{C}^{(n-2)\times 2}$,

$$(T^{-*}\Gamma^*\Gamma T^{-1})^{-1} = \begin{bmatrix} \Delta_{11} & * \\ \Delta_{21} & * \end{bmatrix}.$$

Here,

$$\Delta_{11} = \begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix} (\Gamma^*\Gamma)^{-1} \begin{bmatrix} \nu^* & \bar{\nu}^* \end{bmatrix}$$

is positive definite.

(1 and 2): The eigenvalues of $\hat{A}_D$ can be obtained from

$$\begin{aligned}T\hat{A}_D T^{-1} &= T(\Gamma^*\Gamma + c^2\nu^*\nu + c^2\bar{\nu}^*\bar{\nu})^{-1}(\Gamma^*\Gamma^\uparrow + bc\nu^*\nu\hat{A} + bc\bar{\nu}^*\bar{\nu}\hat{A})T^{-1} \\ &= \begin{bmatrix} (I+c^2\Delta_{11})^{-1} & 0 \\ -c^2\Delta_{21}(I+c^2\Delta_{11})^{-1} & I \end{bmatrix} \begin{bmatrix} I + bc\Delta_{11} & 0 \\ bc\Delta_{21} & I \end{bmatrix} \left(\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \bar{\lambda} & 0 \\ 0 & 0 & \Lambda \end{bmatrix} + \aleph \right) \\ &= \begin{bmatrix} (I+c^2\Delta_{11})^{-1}(I+bc\Delta_{11}) & 0 \\ c(b-c)\Delta_{21}(I+c^2\Delta_{11})^{-1} & I \end{bmatrix} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \bar{\lambda} & 0 \\ 0 & 0 & \Lambda \end{bmatrix} + \aleph. \qquad \text{(D.4)}\end{aligned}$$

The last equation stems from the fact that the second principal submatrix of $\aleph$ must be zero since λ and $\bar{\lambda}$ are distinct. Note that the perturbed eigenvalues are the same as those of

$$\Lambda_D = (I + c^2\Delta_{11})^{-1}(I + bc\Delta_{11}) \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}.$$

Now use the Gram-Schmidt process to produce an orthonormal basis for ν and $\bar{\nu}$,

$$\begin{bmatrix} \nu \\ \bar{\nu} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ \alpha & \beta \end{bmatrix} \begin{bmatrix} \nu \\ \mu \end{bmatrix} = \Theta \begin{bmatrix} \nu \\ \mu \end{bmatrix},$$

where $\alpha = \bar{\nu}\nu^*$ and $\beta = \sqrt{1 - |\alpha|^2}$. Let Q be a unitary matrix with the first two rows being ν and μ. A similarity transformation of $\hat{A}$ gives

$$Q\hat{A}Q^* = \left[\begin{array}{cc:c} \lambda & 0 & \\ \delta & \bar{\lambda} & 0 \\ \hdashline \multicolumn{2}{c:}{*} & \tilde{\Lambda} \end{array}\right],$$

where $\delta = \frac{\alpha}{\beta}(\bar{\lambda} - \lambda)$ and $\tilde{\Lambda}$ contains the other eigenvalues of $\hat{A}$. Next, let

$$(Q\Gamma^*\Gamma Q^*)^{-1} = \begin{bmatrix} \tilde{\Delta}_{11} & * \\ \tilde{\Delta}_{21} & * \end{bmatrix},$$

where

$$\tilde{\Delta}_{11} = \begin{bmatrix} \nu \\ \mu \end{bmatrix} (\Gamma^*\Gamma)^{-1} \begin{bmatrix} \nu^* & \mu^* \end{bmatrix}.$$

Then

$$Q\hat{A}_DQ^* = \begin{bmatrix} (I + c^2\tilde{\Delta}_{11}\Theta^*\Theta)^{-1}(I + bc\tilde{\Delta}_{11}\Theta^*\Theta) & 0 \\ c(b-c)\tilde{\Delta}_{21}\Theta^*\Theta(I + c^2\tilde{\Delta}_{11}\Theta^*\Theta)^{-1} & I \end{bmatrix} \left[\begin{array}{cc:c} \lambda & 0 & \\ \delta & \bar{\lambda} & 0 \\ \hdashline \multicolumn{2}{c:}{*} & \tilde{\Lambda} \end{array}\right]$$

and

$$Q\hat{A}_DQ^* = \begin{bmatrix} \tilde{\Lambda}_D & 0 \\ * & \tilde{\Lambda} \end{bmatrix}.$$

It is obvious that $\tilde{\Lambda}_D$ contains the perturbed eigenvalues and all other eigenvalues remain the same.

(3): Now, suppose the pair b and c satisfy

$$\begin{aligned} & (I + c^2\Delta_{11})^{-1}(I + bc\Delta_{11}) \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix} = \Lambda_D \\ \Leftrightarrow \quad & (I + bc\Delta_{11}) = (I + c^2\Delta_{11})\Lambda_D \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}^{-1} \\ \Leftrightarrow \quad & bI = \frac{\Delta_{11}^{-1}}{c} \left((I + c^2\Delta_{11})\Lambda_D \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}^{-1} - I \right). \end{aligned}$$

Substituting the above equation into Equation (D.4) gives

$$T\hat{A}_D T^{-1} = \left[\begin{array}{c:c} \Lambda_D & 0 \\ \hdashline \Delta_{21}\Delta_{11}^{-1}\left(\Lambda_D - \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}\right) & \Lambda \end{array}\right] + \aleph,$$

which is independent of b and c.

(4): By Lemma D.1,

$$\|\hat{A}_D - \hat{A}\|_2 \leq |c(b-c)| \frac{\rho(\Delta_{11})}{1 + c^2\rho(\Delta_{11})} \sqrt{\kappa(\Gamma^*\Gamma)}\, \|\hat{A}\|_2.$$

Then the statement can be seen by noting that

$$\begin{aligned} \|c(b-c)I\|_2 &= \left\| (I + c^2\Delta_{11})\Delta_{11}^{-1} \left(\Lambda_D - \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}\right) \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}^{-1} \right\|_2 \\ &\leq \frac{\bar{\sigma}(I + c^2\Delta_{11})}{\bar{\sigma}(\Delta_{11})|\lambda|} \bar{\sigma}\left(\Lambda_D - \begin{bmatrix} \lambda & 0 \\ 0 & \bar{\lambda} \end{bmatrix}\right), \end{aligned}$$

where $\Delta_{11} = \nu(\Gamma^*\Gamma)^{-1}\nu^* + \bar{\nu}(\Gamma^*\Gamma)^{-1}\bar{\nu}^*$.

This completes the proof. □

Bibliography

[AK90] K.S. Arun and S.Y. Kung. Balanced approximation of stochastic systems. *SIAM J. Matrix Anal. Appl.*, 11:42–68, 1990.

[Aka74] H. Akaike. Stochastic theory of minimal realization. *IEEE Trans. on Automatic Control*, 19:667–674, 1974.

[Aka75] H. Akaike. Markovian representation of stochastic processes by canonical variables. *SIAM J. Control*, 13:162–173, 1975.

[Aok90] M. Aoki. *State Space Modeling of Time Series.* Springer-Verlag, Berlin, 1990.

[BÅ70] R. Bellman and K.J. Åström. On structural identifiability. *Math. Biosciences*, 7:329–339, 1970.

[Bil95] P. Billingsley. *Probability and Measure.* John Wiley & Sons, New York, U.S.A., 1995.

[BPFT91] B. Bamieh, J.B. Pearson, B.A. Francis, and A. Tannenbaum. A lifting technique for linear periodic systems with application to sampled-data control. *Syst. and Contr. Letters*, 17:79–88, 1991.

[BS85] E.W. Bai and S.S. Sastry. Persistency of excitation, sufficient richness and parameter convergence in discrete time adaptive control. *Syst. and Contr. Letters*, 6:153–163, 1985.

[Bud71] M.A. Budin. Minimal realization of discrete linear systems from input-output observations. *IEEE Trans. on Automatic Control*, 5:395–401, 1971.

[Cai88] P.E. Caines. *Linear Stochastic Systems.* John Wiley & Sons, New York, U.S.A., 1988.

[Cho94] C.T. Chou. *Geometry of Linear Systems and Identification.* PhD thesis, Cambridge University, U.K., 1994.

[CK95] Y.M. Cho and T. Kailath. Fast subspace-based system identification: An instrumental variable approach. *Automatica*, 31:903–905, 1995.

[CM97] C.T. Chou and J.M. Maciejowski. System identification using balanced parameterizations. *IEEE Trans. on Automatic Control*, 42:956–974, 1997.

[CV97] C.T. Chou and M.H. Verhaegen. Subspace algorithms for the identification of multivariable dynamic errors-in-variables models. *Automatica*, 33:1857–1869, 1997.

[CXK94a] Y.M. Cho, G. Xu, and T. Kailath. Fast identification of state-space models via exploitation of displacement structure. *IEEE Trans. on Automatic Control*, 39:2004–2017, 1994.

[CXK94b] Y.M. Cho, G. Xu, and T. Kailath. Fast recursive identification of state-space models via exploitation of displacement structure. *Automatica*, 30:45–59, 1994.

[DM88] B. De Moor. *Mathematical Concepts and Techniques for Modeling of Static and Dynamic Systems.* PhD thesis, Katholieke Universiteit Leuven, Belgium, 1988.

[DM97] B.L.R. De Moor. *DAISY: Database for the Identification of Systems.* Department of Electrical Engineering, ESAT/SISTA, K.U.Leuven, Belgium, URL: "http://www.esat.kuleuven.ac.be/sista/daisy/", 1997.

[DP84] U.B. Desai and D. Pal. A transformation approach to stochastic model reduction. *IEEE Trans. on Automatic Control*, 29:1097–1100, 1984.

[DPK85] U.B. Desai, D. Pal, and R.D. Kirkpatrick. A realization approach to stochastic model reduction. *Int. J. Control*, 42:821–838, 1985.

[DPS95] M. Deistler, K. Peternell, and W. Scherrer. Consistency and relative efficiency of subspace methods. *Automatica*, 31:1865–1875, 1995.

[ESG77] E. Emre, L.M. Silverman, and K. Glover. Generalized dynamic covers for linear systems with applications to deterministic identification and realization problems. *IEEE Trans. on Automatic Control*, 22:26–35, 1977.

[Fau76] P.L. Faurre. Stochastic realization algorithms. In R. Mehra and D. Lainiotis, editors, *System Identification: Advances and Case Studies.* Academic Press, New York, 1976.

[FDMVO96] W. Favoreel, B. De Moor, and P. Van Overschee. Subspace identification of bilinear systems subject to white inputs. Technical Report ESAT-SISTA/TR 1996-53I, Katholieke Universiteit Leuven, Belgium, 1996.

[Glo84] K. Glover. All optimal Hankel-norm approximation of linear multivariable systems and their L^∞-error bounds. *Int. J. Control*, 39:1115–1193, 1984.

[Gop69] B. Gopinath. On the identification of linear time-invariant systems from input-output data. *The Bell Sys. Tech. Journal*, 48:1101–1113, 1969.

[GS93] G.R. Grimmett and D.R. Stirzaker. *Probability and Random Processes.* Oxford University Press, Oxford, U.K., 1993.

[GVL89] G.H. Golub and C.F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, Baltimore, U.S.A., 1989.

[GW74] K. Glover and J.C. Willems. Parametrizations of linear dynamical systems: Canonical forms and identifiability. *IEEE Trans. on Automatic Control*, 19:640–646, 1974.

[HJ93] R.A. Horn and C.R. Johnson. *Matrix Analysis.* Cambridge University Press, Cambridge, U.K., 1993.

[HJ94] R.A. Horn and C.R. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, U.K., 1994.

[HK71] K. Hoffman and R. Kunze. *Linear Algebra.* Prentice-Hall, Inc., N.J., U.S.A., 1971.

[HP90] D. Hinrichsen and A.J. Pritchard. An improved error estimate for reduced-order models of discrete-time systems. *IEEE Trans. on Automatic Control*, 35:317–320, 1990.

[JW95] M. Jansson and B. Wahlberg. On weighting in state-space subspace system identification. In *3rd European Control Conf.*, volume 1, pages 435–440, Rome, Italy, 1995.

[JW96] M. Jansson and B. Wahlberg. A linear regression approach to state-space subspace system identification. *Signal Processing*, 52:103–129, 1996.

[Kai80] T. Kailath. *Linear Systems.* Prentice-Hall, Inc., N.J., U.S.A., 1980.

[Kal71] R.E. Kalman. On minimal partial realizations of a linear input/output map. In R.E. Kalman and N. DeClaris, editors, *Aspects of Network and System Theory.* Holt, Reinhart and Winston, New York, 1971.

[KC83] C.E. Kim and J.T. Cain. An *a priori* identifiability condition and order determination algorithm for MIMO systems. *IEEE Trans. on Systems, Man, and Cybernetics*, 13:1119–1128, 1983.

[Kun78] S.Y. Kung. A new low-order approximation algorithm via singular value decomposition. In *12th Asilomar Conf. on Circuits, Systems and Computers*, pages 705–714, Asilomar, 1978.

[Lar83] W.E. Larimore. System identification, reduced-order filtering and modelling via canonical variate analysis. In *American Control Conf.*, pages 445–451, San Francisco, 1983.

[Lar90] W.E. Larimore. Canonical variate analysis in identification, filtering, and adaptive control. In *29th IEEE Conf. on Decision and Control*, pages 596–604, Honolulu, Hawaii, 1990.

[Lju87] L. Ljung. *System Identification - Theory for the Users.* Prentice-Hall, Inc., N.J., U.S.A., 1987.

[Lju91] L. Ljung. A simple start-up procedure for canonical form state space identification, based on subspace approximation. In *30th IEEE Conf. on Decision and Control*, pages 1333–1336, Brighton, England, 1991.

[LM96] L. Ljung and T. McKelvey. Subspace identification from closed loop data. *Signal Processing*, 52:209–215, 1996.

[LP96] A. Lindquist and G. Picci. Canonical correlation-analysis, approximate covariance extension, and identification of stationary time-series. *Automatica*, 32:709–733, 1996.

[LS77] R. Liu and L.C. Suen. Minimal dimension realization and identifiability of input-output sequences. *IEEE Trans. on Automatic Control*, 22:227–232, 1977.

[MA95] T. McKelvey and H. Akçay. Subspace based system identification with periodic excitation signals. *Syst. and Contr. Letters*, 26:349–361, 1995.

[Mac85] J.M. Maciejowski. Balanced realizations in system identification. In *7th IFAC Symposium on Identification & System Parameter Estimation*, York, U.K., 1985.

[Mac89] J.M. Maciejowski. *Multivariable Feedback Design.* Addison-Wesley Publishers, Wokingham, England, 1989.

[Mac95] J.M. Maciejowski. Guaranteed stability with subspace methods. *Syst. and Contr. Letters*, 26:153–156, 1995.

[MAL96a] T. McKelvey, H. Akçay, and L. Ljung. Subspace-based identification of infinite-dimensional multivariable systems from frequency-response data. *Automatica*, 32:885–902, 1996.

[MAL96b] T. McKelvey, H. Akçay, and L. Ljung. Subspace-based multivariable system identification from frequency response data. *IEEE Trans. on Automatic Control*, 41:960–978, 1996.

[Mar84] I. Mareels. Sufficiency of excitation. *Syst. and Contr. Letters*, 5:159–163, 1984.

[McK95] T. McKelvey. *Identification of State-Space Models from Time and Frequency Domain.* PhD thesis, Linköping University, Sweden, 1995.

[MDMRT92] M. Moonen, B. De Moor, J. Ramos, and S. Tan. A subspace identification algorithm for descripter systems. *Syst. and Contr. Letters*, 19:47–52, 1992.

[MDMVV89] M. Moonen, B. De Moor, L. Vandenberghe, and J. Vandewalle. On- and off-line identification of linear state-space models. *Int. J. Control*, 49:219–232, 1989.

[MO88] J.M. Maciejowski and R.J. Ober. Balanced parametrizations and canonical forms for system identification. In *8th IFAC Symposium on Identification & System Parameter Estimation*, Beijing, China, 1988.

[Moo81] B.C. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE Trans. on Automatic Control*, 26:17–32, 1981.

[Moo83] J.B. Moore. Persistence of excitation in extended least squares. *IEEE Trans. on Automatic Control*, 28:60–68, 1983.

[MR93] M. Moonen and J. Ramos. A subspace algorithm for balanced state space system identification. *IEEE Trans. on Automatic Control*, 38:1727–1729, 1993.

[Obe91] R.J. Ober. Balanced parametrization of classes of linear systems. *SIAM J. Control and Optimization*, 29:1251–1287, 1991.

[OV94] B. Ottersten and M. Viberg. A subspace based instrumental variable method for state space system identification. In *SYSID '94*, pages 139–144, Copenhagen, Denmark, 1994.

[OVK92] B. Ottersten, M. Viberg, and T. Kailath. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Trans. on Signal Proc.*, 40:590–600, 1992.

[PK96a] G. Picci and T. Katayama. A simple "subspace" identification method with exogenous inputs. In *13th World Congress of IFAC*, volume I, pages 175–180, San Francisco, 1996.

[PK96b] G. Picci and T. Katayama. Stochastic realization with exogenous inputs and "subspace-methods" identification. *Signal Processing*, 52:145–160, 1996.

[PS82] L. Pernebo and L.M. Silverman. Model reduction via balanced state space reduction. *IEEE Trans. on Automatic Control*, 27:382–387, 1982.

[PSD96] K. Peternell, W. Scherrer, and M. Deistler. Statistical analysis of novel subspace identification methods. *Signal Processing*, 52:161–177, 1996.

[Rud87] W. Rudin. *Real and Complex Analysis.* McGraw-Hill Book Co., Singapore, 1987.

[SNF70] B. Sz.-Nagy and C. Foiaş. *Harmonic Analysis of Operators on Hilbert Space.* North-Holland Publishing Company, Amsterdam, 1970.

[SROK95] A. Swindlehurst, R. Roy, B. Ottersten, and T. Kailath. A subspace fitting method for identification of linear state-space models. *IEEE Trans. on Automatic Control*, 40:311–316, 1995.

[SS89] P. Stoica and T. Söderström. *System Identification.* Prentice-Hall, Inc., N.J., U.S.A., 1989.

[VD91] M. Verhaegen and E. Deprettere. A fast, recursive MIMO state space model identification algorithm. In *30th IEEE Conf. on Decision and Control*, pages 1349–1354, Brighton, England, 1991.

[VD92a] M. Verhaegen and P. Dewilde. Subspace model identification – Part 1. *Int. J. Control*, 56:1187–1210, 1992.

[VD92b] M. Verhaegen and P. Dewilde. Subspace model identification – Part 2. *Int. J. Control*, 56:1211–1241, 1992.

[vdH96] J.M. van den Hof. *System Theory and System Identification of Compartmental Systems.* PhD thesis, Centrum voor Wiskunde en Informatica, The Netherlands, 1996.

[Ver93a] M. Verhaegen. Application of a subspace model identification technique to identify LTI systems operating in closed-loop. *Automatica*, 29:1027–1040, 1993.

[Ver93b] M. Verhaegen. Subspace model identification – Part 3. *Int. J. Control*, 58:555–586, 1993.

[Ver94] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30:61–74, 1994.

[Vib95] M. Viberg. Subspace-based methods for the identification of linear time-invariant systems. *Automatica*, 31:1835–1851, 1995.

[VO91] M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Trans. on Signal Proc.*, 39:1110–1121, 1991.

[VO95] P. Van Overschee. *Subspace Identification, Theory - Implementation - Application.* PhD thesis, Katholieke Universiteit Leuven, Belgium, 1995.

[VODM91] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. In *30th IEEE Conf. on Decision and Control*, pages 1321–1326, Brighton, England, 1991.

[VODM93] P. Van Overschee and B. De Moor. Subspace algorithms for the stochastic identification problem. *Automatica*, 29:649–660, 1993.

[VODM94] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic stochastic systems. *Automatica*, 30:75–93, 1994.

[VODM95a] P. Van Overschee and B. De Moor. Choice of state-space basis in combined deterministic-stochastic subspace identification. *Automatica*, 31:1877–1883, 1995.

[VODM95b] P. Van Overschee and B. De Moor. A unifying theorem for three subspace system-identification algorithms. *Automatica*, 31:1853–1864, 1995.

[VODM96a] P. Van Overschee and B. De Moor. Continuous-time frequency domain subspace system identification. *Signal Processing*, 52:179–194, 1996.

[VODM96b] P. Van Overschee and B. De Moor. *Subspace Identification for Linear Systems: Theory, Implementation, Applications.* Kluwer Academic Publishers, The Netherlands, 1996.

[VOWL91] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung. A statistical perspective on state-space modeling using subspace methods. In *30th IEEE Conf. on Decision and Control*, pages 1337–1342, Brighton, England, 1991.

[VOWL93] M. Viberg, B. Ottersten, B. Wahlberg, and L. Ljung. Performance of subspace based state space system identification methods. In *12th IFAC World Congress*, pages 369–372, Sydney, Australia, 1993.

[VW96] M. Verhaegen and D. Westwick. Identifying MIMO Hammerstein systems in the context of subspace model identification methods. *Int. J. Control*, 63:331–349, 1996.

[VWO95] M. Viberg, B. Wahlberg, and B. Ottersten. Analysis of state space system identification methods based on instrumental variables and subspace fitting. Technical Report CTH-TE-34, Chalmers University of Technology, Sweden, 1995.

[VY95] M. Verhaegen and X. Yu. A class of subspace model identification algorithms to identify periodically and arbitrarily time-varying systems. *Automatica*, 31:201–216, 1995.

[Wal82] E. Walter. *Identifiability of State Space Models.* Lecture Notes in Biomathematics, 46. Springer-Verlag, Berlin, 1982.

[WJ94] B. Wahlberg and M. Jansson. 4SID linear regression. In *33rd IEEE Conf. on Decision and Control*, pages 2858–2863, Florida, 1994.

[WV96] D. Westwick and M. Verhaegen. Identifying MIMO Wiener systems using subspace model identification methods. *Signal Processing*, 52:253–258, 1996.

[Yam90] Y. Yamamoto. New approach to sampled-data control systems – A function space method. In *29th IEEE Conf. on Decision and Control*, pages 1882–1887, Honolulu, Hawaii, 1990.

[ZDG96] K. Zhou, J.C. Doyle, and K. Glover. *Robust and Optimal Control.* Prentice-Hall, Inc., N.J., U.S.A., 1996.

[ZM74] H.P. Zeiger and A.J. McEwen. Approximate linear realization of given dimension via Ho's algorithm. *IEEE Trans. on Automatic Control*, 19:153, 1974.