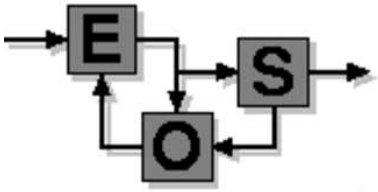

D I S S E R T A T I O N

SOME ASYMPTOTIC THEORY FOR THE ESTIMATION OF LINEAR SYSTEMS USING MAXIMUM LIKELIHOOD METHODS OR SUBSPACE ALGORITHMS

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines Doktors der technischen Wissenschaften unter der Leitung von

Hr. o. Prof. Dipl. Ing. Dr. techn. Manfred Deistler

E1191
Institut für Ökonometrie, Operations Research und Systemtheorie

eingereicht an der Technischen Universität Wien
Technisch Naturwissenschaftliche Fakultät

von
Dietmar Bauer
9025508
Landsbergerstr. 26/3, A-3100 St.Pölten

Wien, am 11. November 1998

Kurzfassung

Diese Dissertation befaßt sich mit der Schätzung von linearen, zeit-invarianten, endlich dimensionalen Zustandsraummodellen. Die Daten hierbei bestehen aus einer Anzahl von quantitativen Beobachtungen möglicherweise mehrerer Größen, wobei die Beobachtungen in zeitlich konstanten Abständen gemessen werden. Für diese Klasse von Daten stellen die linearen Zustandsraummodelle eine häufig verwendete Modellklasse dar, die sowohl in den Ingenieurwissenschaften als auch zum Beispiel in der Ökonometrie, der medizinischen Statistik und der Ökologie Anwendungen findet.

Die Standardmethode zur Schätzung solcher Zustandsraummodelle stellt die Maximum-Likelihood (ML) Methode dar (siehe u.a. Hannan and Deistler, 1988). Die asymptotischen Eigenschaften der ML Schätzer wurden erschöpfend erforscht: Sie sind konsistent und asymptotisch effizient für die gewöhnlichen Annahmen an den Prozeß. Für die Implementierung der ML Methode muß eine Parametrisierung aller Systeme gegebener Ordnung benutzt werden. Die topologischen Eigenschaften dieser Parametrisierung beeinflussen die Eigenschaften der Schätzer wesentlich.

Der erste Teil dieser Dissertation beschäftigt sich mit der Analyse der topologischen Eigenschaften von sogenannten balanzierten Parametrisierungen. Für diese Art von Parametrisierungen wird eine Zerlegung der Menge aller Systeme der Ordnung n präsentiert, sodaß jedes Teilstück dieser Zerlegung stetig parametrisiert werden kann. Die topologischen Eigenschaften dieser Stücke sowie deren Abschlusses werden untersucht, und schließlich auch die Struktur der Parameterräume sowie deren Ränder. Abschließend werden die balanzierten Parametrisierungen mit den Echelon Parametrisierungen (vergl. Hannan and Deistler, 1988) verglichen. Einige dieser Ergebnisse werden auf den Fall von 'strictly-minimum-phase' Systemen übertragen: In dieser Hinsicht sind vor allem die Parametrisierung, sowie die Eigenschaften der Abschlüße der Teilstücke der Menge aller Systeme der Ordnung n zu nennen. In diesem Kapitel werden auch 'frequenz-gewichtet balanzierte' Systeme untersucht und deren Modellreduktions- Eigenschaften vorgestellt. Das Kapitel beenden einige bekannte Resultate bezüglich der asymptotischen Eigenschaften der ML Schätzer, sowie Hinweise zur tatsächlichen Implementierung der ML Prozedur.

Der zweite Teil der Dissertation ist gleichzeitig der Haupteil der Arbeit. Kapitel 3 liefert eine überblicksmäßige Einführung in die sogenannten 'Subspace'-Algorithmen. Diese Algorithmen können grob in 3 Schritte untergliedert werden:

1. Schätzung eines hochdimensionalen Zustandes
2. Modellreduktion durch Komprimierung der Information im Zustandsvektor auf n Komponenten.
3. Schätzung der Systemmatrizen

Hierbei wird eine Reihe von Algorithmen besprochen, die allesamt in der Literatur unter 'Subspace'-Algorithmen geführt werden, allerdings relativ unterschiedliche Ansätze benutzen. Das Hauptaugenmerk hierbei liegt im Aufzeigen der Zusammenhänge und Unterschiede der einzelnen Prozeduren. Am Ende des Kapitels wird dann jene Klasse von Algorithmen beschrieben, welche im 4.Kapitel genau analysiert wird.

Das 4. Kapitel präsentiert das Hauptresultat dieser Dissertation, asymptotische Normalverteilung der geschätzten Systemmatrizen. Dieses Resultat wird für eine Reihe von Situationen hergeleitet, wobei der Fall ohne exogene Inputs genauso behandelt wird wie der Fall der Berücksichtigung exogener Inputs. Auch jene Klasse von Algorithmen, welche die Struktur der den Zustandsraumsystemen zugrundeliegenden Rekursionen noch stärker für die Schätzung benutzt (siehe Peternell *et al.*, 1996), wird behandelt.

Kapitel 5 beschäftigt sich dann mit den Auswirkungen der Wahl von bestimmten Designvariablen, die der Benutzer im Algorithmus wählen muß. Dabei werden vor allem Ordnungsschätzprozeduren entwickelt, die sich aus den Ergebnissen von Kapitel 4 als naheliegend erweisen. Für alle neu entwickelten Ordnungsschätzer wird Konsistenz bewiesen. Anschließend werden die neuen Prozeduren mit den bereits vorhandenen in Simulationsstudien verglichen. Weiters wird die Wahl von Gewichtsmatrizen untersucht. So wird die Verteilung der Systemmatrizenschätzer hergeleitet in dem Fall, daß die Ordnung, die für die Schätzung benutzt wird niedriger

ist als die wahre Ordnung des Systemes. Dieses Resultat beinhaltet auch ein Konsistenzresultat, wobei das für das Grenzsystem, gegen welches die Schätzer für Anzahl der Beobachtungen gegen unendlich konvergieren, explizite Formeln angegeben werden können. Berechnungen für eine Anzahl von Systemen zeigen den Effekt von unterschiedlichen Wahlen der Gewichtsmatrizen auf die asymptotische Varianz der Schätzer und den asymptotischen Approximationsfehler bei zu geringer Wahl der Ordnung des geschätzten Systemes. Schließlich werden Möglichkeiten aufgezeigt, um die Stabilität beziehungsweise die 'Minimum-phase' Eigenschaft des geschätzten Systemes zu garantieren.

Kapitel 6 benutzt die Ergebnisse aus Kapitel 4, um die asymptotische Verteilung für einige andere 'Subspace' Algorithmen zu klären. Erneut bildet der Beweis der asymptotischen Normalverteilung das zentrale Resultat.

Danksagung

An dieser Stelle möchte ich mich bedanken bei allen Leuten, die diese Dissertation ermöglicht oder zur Qualität beigetragen haben: An erster Stelle bedanke ich mich bei meinen Doktorvater, Prof. Manfred Deistler, für viele richtungsweisende Ideen und für enorm viel Vertrauen in meine Arbeit und die damit verbundenen Chancen zur Präsentation meiner Ergebnisse. Die Arbeit wäre wohl unmöglich gewesen, ohne die tatkräftige Mithilfe von ao. Prof. Wolfgang Scherrer, der nicht nur die soft- und hardwaremäßigen Problemchen meinerseits immer im Griff hatte, sondern vor allem durch viele Anregungen und Kritikpunkte mich immer wieder zu neuen Überlegungen trieb, sowie Fehler gnadenlos entlarvte. Dann gilt mein Dank natürlich auch allen Leuten im Umfeld des Instituts, die den Alltag erträglich gestalteten. Ohne die tägliche Kaffeepause und so manches Schwätzchen wäre ich wohl an der Uni vereinsamt. Das ist gleichzeitig ein Stichwort, um auch meine Freunde nicht unerwähnt zu lassen, die sich von meinen mürrischen Launen und abwesenden Blicken nicht davon abhalten ließen, mit mir ein Gläschen im Kastl oder sonstwo zu kippen. Zu guter Letzt möchte ich noch herzlich meiner Familie danken, die mich in der gesamten Zeit nicht nur finanziell über Wasser hielt, sondern auch moralisch unterstützte. Ich hoffe, meine Eltern, mein Bruder und meine Freundin sind ein wenig stolz, zu dieser Dissertation etwas beigetragen zu haben.

Ein Großteil der Arbeit wurde unter anderem finanziert durch den österreichischen FWF, Projekt Nummer P11213-MAT.

Abstract

This thesis is concerned with the estimation of linear, time invariant, finite dimensional state space systems. The data is assumed to consist of quantitative measurements of possibly several variables, where the samples are chosen using a fixed sampling rate. For this class of data, state space systems are a commonly used model class, equally in engineering sciences as well as in other fields of science, such as e.g. econometrics, medical statistics or ecology.

The standard methodology to estimate state space models at present consists in the maximum likelihood (ML) approach (see Hannan and Deistler, 1988). The asymptotic properties of ML estimates have been studied extensively in the past (comp. Hannan and Deistler, 1988): The estimates are known to be consistent and asymptotically normally distributed under the usual assumptions on the process. For the implimentation of the ML approach we have to use a parametrization of the set of all systems of order n. The topological properties of this parametrization influence the statistical properties of the estimates of the parameters and thus of the estimates of the system decisively.

The first part of the thesis investigates the topological properties of so called balanced parametrizations. For this kind of parametrization a partitioning of the set of all systems of order n is presented, where each piece in this partitioning can be parametrized continuously. The topological properties of these pieces, as well as their boundaries are examined. Also the structure of the parameter spaces and their closures is investigated. A comparison of the balanced parametrizations to the echelon parametrizations (see e.g. Hannan and Deistler, 1988) is provided. Some of the result, which have been achieved for the balanced parametrizations are used to obtain similar results for the case of strictly minimum-phase systems: In particular the partitioning of the set of all strictly minimum-phase systems of order n into pieces, which allow for a continuous parametrization, is presented and the results corresponding to the structure of the closures of the pieces are transferred from the corresponding results for the balanced parametrizations. Finally frequency weighted balancing is presented and the model reduction properties of these realizations demonstrated. The chapter is concluded with a summary of known results corresponding to the asymptotic distribution of the ML estimates and some comments concerned with the actual implimentation of the ML approach.

The second part of the thesis contains the main results. Chapter 3 provides a survey of subspace methods. Basically, subspace methods can be decomposed roughly into three main steps:

1. Estimation of a high dimensional state
2. Model reduction leading to an estimate of the n dimensional state
3. Estimation of the system matrices

Here the emphasis is put on a comparison of several methods, which have all been termed 'subspace' methods, whereas they use quite different ideas. The discussion centers on the similarities and differences between the various proposed methods. The chapter is concluded with a presentation of the main algorithm investigated in this thesis.

Chapter 4 then presentes the main result of the thesis, i.e. asymptotic normality of the estimates of the system matrices, in several situations, including the case of no observed inputs as well as the case of additional observed inputs. Also for the algorithms, which use more of the structure of the recursions defining the state space models for the estimation, proposed in (Peternell *et al.*, 1996), similar results are obtained.

Chapter 5 deals with the effects of user choices on the performance of the subspace algorithms. Especially the choice of the order is examined in more depth and the asymptotic theory developed in Chapter 4 is used to define two new order estimation procedures, which are shown to be consistent under the general assumptions of the thesis. These two new procedures are compared in simulation studies to the procedure proposed by (Peternell, 1995) and the order estimation procedure implemented in `MATLAB` (see Ljung, 1991). Furthermore, the effect of the weighting matrices on the asymptotic accuracy is investigated. In particular the asymptotic distribution of the estimates of the system matrices in the case, where the order used for estimation is lower than the

true order, is obtained. This result also contains a consistency result, where explicit expressions for the approximating lower order system are given. For a number of systems, these results are applied to show both, the asymptotic as well as the finite sample behaviour of algorithms using different weighting matrices. These simulations show the effects of the choice of the weighting matrices on the approximation in the case, where the estimation order is chosen too low as well as the asymptotic variance in the case, where the order is specified correctly. Finally we present possibilities in order to ensure, that the estimated system is stable or minimum-phase respectively.

Chapter 6 concludes the discussion with showing some results on the asymptotic distribution for other commonly used subspace procedures, which are easily obtained using the tools provided in Chapter 4. Again the discussion is focussed on the asymptotic normality of the system matrix estimates.

Contents

Chapter 1

Introduction

In many fields of science, a reliable model for the representation of the major features of some measured data is needed. Such a model may be sought for diverse reasons as e.g. for the control of a plant, for an improvement of the understanding of a process in economy, for prediction or for the simulation of some process. In many circumstances the structure of the process is either not known or too complicated, in order to obtain a good model from a priori knowledge, and the scientist has to ressort to black box modelling, which means deciding about some major assumptions (e.g. using linear or nonlinear model structures) describing the model class and then letting the data decide about the complexity of the model and the actually best model for the given purpose. This is rather obvious for the econometrist, but also for the engineer modelling using physical principals only e.g. of movements of a human body will be much too demanding to obtain good results from that (comp. Akaike, 1997). In this thesis we restrict the data to be quantitative, since this corresponds to the methods investigated. As has been stated already, the reason for building a model can be inter alii prediction, simulation of the process, control or discrimination between different models. The actual reason is of no importance for the theory presented in this thesis. We will assume, that no continuous time model is wanted, but rather a model explaining the data only at certain time instants, i.e. in discrete time. Thus we assume, that we are dealing with data $z_t \in \mathbb{R}^{s+m}$ for some integer $s+m$ (the reason for using the sum $s+m$ instead of only s will be clear from the following). Here the time index is restricted to $t \in \mathbb{Z}$, i.e. we deal exclusively with equally spaced (in time) observations.

In this thesis we will investigate two different setups: The main part of the thesis will deal with what is sometimes called time series analysis, i.e. the attempt to explain the future of the measurements of the process only by the past measurements. The alternative to this approach would be to divide the measured variables into two groups and to call the data in one group the *input* (denoted throughout the thesis as $u_t \in \mathbb{R}^m$) and the data in the other group *output* (usually denoted with $y_t \in \mathbb{R}^s$). The reason for this notion probably lies in the viewpoint of the (control) engineer, who applies some controls (e.g. positioning a valve or steering a rudder), and then measures the outcome corresponding to his actions. This viewpoint also motivates different assumptions on the measurements, as introduced later on. It will be assumed, that the user has performed this partitioning into inputs and outputs, and the algorithms do not interfere with this decomposition. The analysis will be mostly confined to the case $m = 0$.

Under practical circumstances the process can only be measured over a certain amount of time i.e. only $(y_t, u_t), 1 \leq t \leq T$ for some integer T will be known. One main line of research in the field of time series analysis is to develop automatic procedures to estimate models within a specified class, i.e. to find a model out of the class, which explains the data 'best' in some sense. The need for such automatic procedures stems from mainly three sources: The first one is probably the wish to provide 'easy to handle' tools, which can be applied by a large community and not only by some specialists, to make more sophisticated methods available for nonspecialists. This point is even stronger today, since the use of computers and graphical interfaces leads to automatic procedures, which can be applied with very few training of the user and still should lead to reasonable results.

The second reason for automatic procedures lies in the fact, that many of the methods developed in the terminology of (Box and Jenkins, 1970) already have been pushed towards a degree, that a typical analysis reduces to routine work. And finally the third point lies in the fact, that there might be simply too many time series to be treated at a time, such that a very thorough analysis of each data set by its own is simply not possible. As an example we cite the case of an economist, building separate models for many stocks (say 1000) in paralell.

In this thesis we will restrict the model set to the case of time invariant, finite dimensional, discrete time, linear state space systems of the form:

$$\begin{array}{rcl} x(t+1) & = & Ax(t) + Bu(t) + K\varepsilon(t) \\ y(t) & = & Cx(t) + Du(t) + E\varepsilon(t) \end{array} \tag{1.1}$$

where $(y(t))_{t\in\mathbb{Z}}$ denotes the s-dimensional output series, $(u(t))_{t\in\mathbb{Z}}$ the m-dimensional (measured) input series, $(\varepsilon(t))_{t\in\mathbb{Z}}$ s-dimensional white noise with covariance matrix equal to the identity (for a definition of *white noise* see Appendix A) and $x(t)$ the n-dimensional state. $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, K \in \mathbb{R}^{n\times s}, C \in \mathbb{R}^{s\times n}, D \in \mathbb{R}^{s\times m}, E \in \mathbb{R}^{s\times s}$ are real matrices. We will always indicate random variables by appending the time index in bracketts (e.g. $y(t)$), whereas the observations of the random variables will be marked by adding the time index as a subscript (e.g. y_t). Note that this system description also includes the case of $m = 0$, where no exogenous inputs are present. The restrictions implied by this model set are rather obvious: First all systems in the model set are linear in the input-to-output behaviour. Of course, this corresponds to the methods analyzed in the main part, which are applicable to this model class. The fact, that many applications especially in the engineering literature use these assumptions, may be seen as a justification for this restriction. The second main restriction lies in the time invariance. This assumption is vital to our analysis, since we are interested in the asymptotic behaviour of the estimates, which only make sense, if the model does not change in time. The remaining assumptions of the model set do not seem to be very restrictive. The finiteness of the model order is a common requirement and the discrete time results from the time scale of the measurements.

Throughout the thesis it will be assumed, that the system (1.1) is stable (i.e. $|\lambda_{max}(A)| < 1$, where $\lambda_{max}(A)$ denotes an eigenvalue of maximum modulus of the square $n \times n$ matrix A). In the following we are only interested in the steady state solution of the equations (1.1) i.e. we assume, that the system had been started a long time ago such that the influences of the initial state $x(t_0)$ have died out due to the stability of the matrix A.

Thus $y(t)$ can formally be written down as follows:

$$y(t) = E\varepsilon(t) + \sum_{j=1}^{\infty} CA^{j-1}K\varepsilon(t-j) + Du(t) + \sum_{j=1}^{\infty} CA^{j-1}Bu(t-j) \tag{1.2}$$

where the limits exist under general assumptions on $\varepsilon(t)$ and $u(t)$ (to be given below) and are to be understood in the mean square sense (see Appendix A). We can define the two *transfer functions* $k(z) = E + \sum_{j=1}^{\infty} CA^{j-1}Kz^{-j} = \sum_{j=0}^{\infty} K(j)z^{-j}$ and $l(z) = D + \sum_{j=1}^{\infty} CA^{j-1}Bz^{-j} = \sum_{j=0}^{\infty} L(j)z^{-j}$. Assuming $[y(t)^T, u(t)^T]^T$ to be a stationary process (see Appendix A), we can define the forward shift operator z as $z(y(t)) = y(t+1), z(u(t)) = u(t+1)$. Interpreting powers of z as iterations of the shiftoperator (e.g. $z^2(y(t)) = y(t+2)$ and so on), we may interpret the transfer functions as power series in the inverse of the forwardshift operator, the backwardshift operator z^{-1}. It is well known, that there exists an isomorphism between power series in the backwardshift and power series in the complex variable z. Therefore all questions of convergence as well as multiplication and inversion of power series in the backward shift can be regarded as questions concerning complex power series and vice versa. The stability assumption on the matrix A implies the assumption, that all poles of the complex function $k(z)$ lie strictly inside the unit circle. This formalism makes it possible to shorten the notation considerably: $y(t) = l(z)u(t) + k(z)\varepsilon(t)$. This results in an additive decomposition of $y(t)$ into a part, which is due to the (measured) inputs (i.e. $l(z)u(t)$) and another part that stems from the white noise $\varepsilon(t)$. Since we do not measure the noise, we may without restriction of generality choose an

arbitrary representation of the (stationary) process $k(z)\varepsilon(t)$ (see Appendix A). Therefore we will throughout the thesis assume, that the transfer function $k(z)$ corresponding to the noise contribution is stable and minimum-phase i.e. that all poles of the complex function $k(z)$ lie strictly inside the unit circle and that $\det(k(z)) \neq 0, |z| > 1$. However sometimes we will also need the strict minimum-phase assumption (i.e. $\det(k(z)) \neq 0, |z| \geq 1$). The strict minimum-phase assumption excludes stationary processes with spectra having zeros on the unit circle. We will also assume the matrix E to be nonsingular. Then we may w.r.o.g. assume E to be lower triangular with positive entries on the diagonal, since $\mathbb{E}\varepsilon(t)\varepsilon(t)^T = I$.

Prior to identification, there are many things to be thought of (data acquisition, grouping into inputs and outputs, data pre-treatment, the choice of appropriate model classes and identification methods depending on the quality of the data, to name just a few). Despite of this fact, we will always consider the situation, where we are given data, which is grouped into outputs and possibly inputs, which is equally spaced in time, and which has been checked not to contradict the basic assumptions. Then the problem of identification can be formulated as follows:

> Given measurements $y_t, u_t, 1 \leq t \leq T$ of one realization of the processes $(y(t), u(t))_{t\in\mathbb{Z}}$ find the model in the model set defined by the equations (1.1), which 'gives the best description of the process'.

The exact meaning of 'description of the process' is specified by the identification objectives. Throughout the thesis we will assume, that the user invests no a priori information on the process in the identification or the model class, i.e. we are confined to black-box identification. Also we assume, that the objective of the identification is either prediction or not classified. This serves as a justification to use general model sets and general criterion functions. The analysis will concentrate on the asymptotic behaviour.

Throughout the thesis, we will use some standard assumptions on the process. In the following we will state these assumptions and give a justification.

DEFINITION (STANDARD ASSUMPTIONS) The ergodic noise $\varepsilon(t)$ fulfills the following four conditions:

$$\begin{array}{lcl} \mathbb{E}\{\varepsilon(t)|\mathcal{F}_{t-1}\} & = & 0 \\ \mathbb{E}\{\varepsilon(t)\varepsilon(t)^T|\mathcal{F}_{t-1}\} = \mathbb{E}\{\varepsilon(t)\varepsilon(t)^T\} & = & I \\ \mathbb{E}\{\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)|\mathcal{F}_{t-1}\} & = & \omega_{a,b,c} \\ \mathbb{E}\{\varepsilon_i(t)^4\} & < & \infty \end{array} \tag{1.3}$$

and where $\mathbb{E}$ denotes expectation. The transfer functions (k, l) are rational functions of finite degree n. k is stable and strictly minimum-phase, l is stable. The constant term of k is lower triangular with strictly positive diagonal entries.

Note, that we assume the system to be finite dimensional. This seems to be very restrictive at first sight. However, it is well known, that any pair of transfer functions (k, l), which have coefficients decreasing sufficiently fast ($\sum_{j=0}^{\infty} j^{1/2}||(K_j, L_j)|| < \infty$) can be approximated arbitrarily close (in a proper sense) by rational transfer functions of finite degree (see Hannan and Deistler, 1988).

The assumption on the conditional first moment of the noise is motivated by the use of linear models: This condition can be shown to be sufficient for the best linear predictor in the mean square sense to be the best predictor (see Appendix A, section A.2). When using linear models, this seems to be a reasonable requirement.

The condition on the second order moments includes two statements: The first statement is, that the conditional variance of the noise is equal to the unconditional variance, meaning that the knowledge of the whole past of the noise does not lead to a more accurate (in terms of prediction error variance) prediction of the noise. The second statement contains a mere technical point and a conceptual point: For convenience of statement of the main results, we decided to set the noise covariance to be equal to unity, while including a matrix E into the description of the system in equation (1.1). Thus the noise sequence $\varepsilon(t)$ will be different from the innovation sequence $\nu(t) = E\varepsilon(t)$. The innovation variance is equal to EE^T. This can be done, if the innovations variance is nonsingular, meaning that no linear combination of the process can be predicted perfectly from the past. For many applications this seems to be no important restriction. From a technical point of view, this condition is needed to ensure the uniqueness of the factorization of the spectrum (see Hannan and Deistler, 1988).

These two conditions on the noise are related to the intuitive interpretation of the noise sequence: The knowledge of the past does not tell us anything of the future of the noise, in the sense that neither does it help us to predict the mean (first order conditions), nor does it reduce the prediction variance (second order conditions). The remaining conditions are of rather technical nature: The condition on the third moments was included to make some arguments simpler. It is not claimed, that this condition is necessary for our results, it is merely more convenient to work with. In particular some calculations corresponding to the asymptotic variance simplify significantly (see REMARK 4.1.1). Also the merit of imposing less restricting assumptions on the third order moments may be doubted, since these conditions are hard to verify anyway. The last condition on the moments of the noise is the finiteness of the unconditional fourth order moments. This condition is needed, since we use a central limit theorem for the covariance estimates. In our framework, the finiteness of the fourth order moment is used to ensure the finiteness of the variance of the covariance estimates.

Concerning the input $(u(t))_{t\in\mathbb{Z}}$ we will use different concepts and thus different assumptions. This is the reason for not including the assumptions on the inputs in the standard assumptions. For a discussion on the different scenarios for the input see Appendix A. However, we will always assume the input $u(t)$ to be 'uncorrelated' with the noise $\varepsilon(s)$ for all $s, t \in \mathbb{Z}$. The exact definition of uncorrelatedness differs with the assumptions on the inputs. This restriction excludes szenarios, where the input is chosen to depend on the output, which is referred to as the *closed loop* case in the engineering literature.

The thesis is divided into two parts: First we will discuss the maximum likelihood approach. Since much of the material presented thereby is standard (especially the analysis of the asymptotic properties), the exposition will be short in some respect. However, this part serves as a motivation for the development of other tools such as the subspace algorithms and it sets the benchmarks for the subspace algorithms. Additionally it gives us the chance to present some realization theory, which will be of importance in the second part of the thesis, which is indeed the main part: Subspace identification methods. The second part is organized as follows: First the various proposed algorithms are presented in some depth. This is done - although not a complete analysis for all algorithms will be given - since there has been much confusion in the last few years about the term 'subspace methods'. Maybe this kind of survey leads to a better understanding, that there is no such thing as 'a subspace method', but that there are dozens of possibilities and proposed algorithms. In Chapter 4 then the main algorithm will be analyzed. In Chapter 5 we will deal with some questions concerning the main class of algorithms, which will be investigated partly analytically and partly by means of simulation studies. This discussion will also give a justification for picking this class of algorithms for the analysis. Chapter 6 then uses the tools presented in Chapter 4 to derive some results about other subspace algorithms, whenever they are obvious and easily accessible in our framework. Appendix A clarifies the stochastic setting and Appendix B reviews some facts of operator theory, that are needed in the body of the thesis. Finally Appendix C gives an overview of the notation used throughout the thesis.

The contributions of this thesis are the following: For the maximum likelihood approach some

topological results concerning the balanced parametrization have been derived, which are needed in order to obtain results concerning the properties of the parameter estimates in this context and in order to motivate and construct algorithms using this structure. Concerning the subspace algorithms we have derived asymptotic distributions for the estimates for a certain class of algorithms and we have applied these results also to various other algorithms. The main contribution of this thesis is to prove the asymptotic normality of the estimates in several situations and to reaffirm the simulation evidence obtained e.g. in (Peternell, 1995). In particular it is observed, that in the case of no observed inputs one algorithm, which is called `CCA` and is due to (Larimore, 1983), outperforms the other algorithms and comes (at least) close to being optimal in the presented examples.

Chapter 2

Maximum Likelihood Estimation

This chapter deals with the identification of systems using the maximum likelihood (ML) approach as described in (Hannan and Deistler, 1988). We will heavily draw from their presentation. For most of the results, we will not give any proofs but refer to the original source. The reasons for the inclusion of this chapter into the thesis are the following:

- First of all, it provides a justification for the introduction of subspace algorithms by showing the problems inheritant to the ML approach.
- Secondly it also provides a benchmark for the subspace algorithms, because Theorem 2.4.2 tells us, that asymptotically the ML approach gives estimates, which are optimal with respect to the asymptotic variance under standard assumptions.
- The third reason is the presentation of certain parametrizations, which are needed in order to perform the optimization involved in the ML approach. In the discussion of the parametrizations we will present a realization theory, which will prove useful in the main part of this thesis.

2.1 The ML approach

The ML approach is a universal principle of statistics: The estimates of the parameters of a model are obtained by maximizing the likelihood over the parameter space, given the observations. Consider the process $(y(t))_{t\in\mathbb{Z}}$: From the system equations (1.1) it follows, that $y(t) = l(z)u(t) + k(z)\varepsilon(t)$, where $(k(z), l(z))$ denote rational transfer functions. l describes the relation of the inputs $u(t)$ to the outputs and k describes the relation of the noise to the output respectively (comp. Chapter 1). Note that the representation of $y(t)$ used in equations (1.1) is not unique. Since we do not observe the state, we are free to choose a basis in the state space. Transforming the state $x(t)$ to a new state $x'(t) = Tx(t)$ with a nonsingular matrix $T \in \mathbb{R}^{n\times n}$ leads to a new description (A', B', C', D', E', K'), where

$$A' = TAT^{-1}, B' = TB, C' = CT^{-1}, D' = D, E' = E, K' = TK \tag{2.1}$$

However all possible realizations obtained by different choices of the basis in the state space correspond to the same pair of transfer functions (k, l). Denote the mapping attaching pairs of transfer functions (k, l) to system matrices (A, B, C, D, E, K) with π, i.e. $\pi(A, B, C, D, E, K) = (k, l)$. Let S_n denote the set of system matrices (A, B, C, D, E, K), where $A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, K \in \mathbb{R}^{n\times s}, C \in \mathbb{R}^{s\times n}, D \in \mathbb{R}^{s\times m}, E \in \mathbb{R}^{s\times s}$ have real entries and fulfill the stability and the minimum-phase assumption, the matrix E being lower diagonal with strictly positive entries on the diagonal. The set $\pi^{-1}(k, l)$ of all system representations $(A, B, C, D, E, K) \in S_{n'}$ for some integer n' corresponding to the same pair of transfer functions is called *equivalence class*. Note, that in this definition of the equivalence class, we allow realizations of different state dimension to lie in the

same equivalence class. However it can be shown (see e.g. Hannan and Deistler, 1988), that there always exists a smallest state dimension.

DEFINITION (MINIMALITY) A realization $(A, B, C, D, E, K) \in S_n$ of $(k, l) = \pi(A, B, C, D, E, K)$ is called minimal, if for all $(A', B', C', D', E', K') \in S_{n'}, \pi(A', B', C', D', E', K') = (k, l)$ it follows, that $n \leq n'$. In this case the integer n is called the order of the system.

Two realizations $(A, B, C, D, E, K) \in S_n, (A', B', C', D', E', K') \in S_n$, where n is the order of (k, l) are equivalent, if and only if there exists a transformation T, such that equations (2.1) hold. π maps S_n onto some set of transfer functions $\overline{M_n^j} = \pi(S_n)$. Here $\overline{M_n^j}$ (The reason for the notation $\overline{M_n^j}$ will be clear in the following) is the set of all pairs (k, l) of rational transfer functions of 'joint' order smaller or equal to n, where k is stable and minimum-phase with lower diagonal constant term having positive diagonal entries and l is stable. Here 'joint' order means the order needed to combine the dynamics of k and l in the matrix A. Note, that this order will differ from the maximum degree of the two transfer functions in general. Another possibility is to treat k and l seperately. In this case, we will use the notation M_n to denote the set of all rational stable transfer functions of degree smaller or equal to n and the symbol $U_n^{(m)}$ to denote the set of all rational, stable, strictly minimum-phase transfer functions having lower triangular constant term with positive entries on the diagonal. Thus we neglect the dimension of the transfer functions (k, l) in this notation. Whenever there might be confusions due to this, we will note this.

Now assume for the moment, that the white noise $\varepsilon(t)$ is Gaussian of zero mean and unit covariance and that $\varepsilon(t)$ and $u(s)$ are independent for all s and t (of course this is only done for convenience of notation and will not be required for the results). Assume further, thast $u(s) = 0, s < 1$. Using the stability assumption it follows, that $y(t)$ conditioned on the input $u(j), j \leq t$ is Gaussian distributed with (conditional) mean $\overline{y}(t; l) = l(z)u(t)$ and covariance $\gamma(0; k)$, where we have made the dependence of the conditional mean on the transfer function l as well as the dependence of the variance on k more explicit. Moreover, the stacked output vector $Y_{1,T}^+ = [y(1)^T, y(2)^T, \cdots, y(T)^T]^T$ conditioned on the input process is Gaussian distributed with mean equal to $\overline{Y_{1,T}^+}(l) = [\overline{y(1; l)}^T, \cdots, \overline{y(T; l)}^T]^T$ and variance matrix $\Gamma_T^+(k)$, which depends only on the transfer function k, which can be seen as follows: $y(t) - \overline{y(t)} = v(t; l) = \sum_{j=0}^{\infty} K(j)\varepsilon(t - j) = y(t) - \sum_{j=0}^{\infty} L(j)u(t - j)$, thus the covariance $\gamma(i; k) = \mathbb{E}v(t; l)v(t - i; l)^T, i \geq 0$ can be calculated as $\gamma(i; k) = \mathbb{E}(\sum_{j=0}^{\infty} K(j)\varepsilon(t - j))(\sum_{r=0}^{\infty} K(r)\varepsilon(t - r - i))^T = \sum_{j=0}^{\infty}\sum_{r=0}^{\infty} K(j)\mathbb{E}\{\varepsilon(t - j)\varepsilon(t - r - i)^T\}K(r)^T = \sum_{j=i}^{\infty} K(j)K(j - i)^T$. Then the (Gaussian) distribution function of the vector $Y_{1,T}^+ - \overline{Y_{1,T}^+}(l)$ (i.e. $Y_{1,T}^+$ conditional on the whole past of the input process $u(t), t \leq T$) given the true transfer functions (k, l) can be written as follows:

$$L(Y_{1,T}^+, (u(t))_{t\in\mathbb{Z}}; (k, l)) = \frac{\exp\{-\frac{1}{2}(Y_{1,T}^+ - \overline{Y_{1,T}^+}(l))^T(\Gamma_T^+(k))^{-1}(Y_{1,T}^+ - \overline{Y_{1,T}^+}(l))\}}{\sqrt{(2\pi)^T \det \Gamma_T^+(k)}} \tag{2.2}$$

In this equation the dependence on the transfer functions (k, l) is stressed. Let $\Theta \subset \bigcup_{i\leq n} M_i^j$ for some integer n. Then we can define the ML estimate $(\hat{k}, \hat{l})$ as follows:

DEFINITION (ML ESTIMATE) Given observations of the output process $y_t, 1 \leq t \leq T$, observations of the input process $u_t, 1 \leq t \leq T$ and a model set Θ, the ML estimate $(\hat{k}, \hat{l}) \in \overline{\Theta}$, is defined as the maximizing argument of the function $L(Y_{1,T}^+, U_{1,T}^+; k, l)$ i.e.:

$$(\hat{k}, \hat{l}) = \arg \max_{(k,l) \in \overline{\Theta}} \frac{1}{\sqrt{(2\pi)^T \det \Gamma_T^+(k)}} \exp\{-\frac{1}{2} V_{1,T}^+(l)^T (\Gamma_T^+(k))^{-1} V_{1,T}^+(l)\} \qquad (2.3)$$

where $V_{1,T}^+(l) = Y_{1,T}^+ - \overline{Y_{1,T}^+}(l)$ and $U_{1,T}^+$ is defined analogously to $Y_{1,T}^+$. Here random variables are replaced by observations in the definition of $Y_{1,T}^+$ and $U_{1,T}^+$. u_t is assumed to be zero for $t < 1$. $\overline{\Theta}$ denotes the closure of Θ in the pointwise topology T_{pt} to be defined below.

REMARK 2.1.1 Note, that in the definition 2.3 we neglected the effects of the initial conditions on the input $u(t)$, setting all observations prior to $t = 1$ equal to zero. This is no restriction for our purposes, since we are only interested in asymptotic results. This will be evident from the results in Chapter refsubspace:results. Also note, that the definition of the ML-estimate depends on the specified set Θ. Thus when referring to an ML-estimate, it will be necessary to mention the set, over which the likelihood has been optimized. Finally note, that the set, over which the parametrization has to be performed, in fact is equal to $\overline{\Theta}$ reflecting the fact, that in optimization of the likelihood certain boundary points cannot be avoided (see also section 2.4).

REMARK 2.1.2 In practice of course a suitable integer n has to be estimated for the definition of Θ and is not known a priori. This can be done e.g. using an information criterion (Akaike, 1976; Shibata, 1980; Rissanen, 1978, see also section 2.4) by optimizing the likelihood for several values of n and comparing the optimal values. However in the following, for the statistical analysis, we will be mainly concerned with the situation, where the order has already been specified.

The major justification for using the Gaussian density are the statistical properties of the resulting estimators, which can be obtained by this choice (see section 2.4) without actually assuming the white noise $\varepsilon(t)$ to be distributed Gaussian. Thus from now on, we will drop the assumption, that the process is Gaussian, but rather use the standard assumptions introduced in the introduction (see Chapter 1).

Once we have defined the maximum likelihood estimator in terms of the transfer functions (k, l), there arise several questions, which will be answered in the following sections:

- In the next section, we will describe two approaches to obtain parametrizations for the class of systems given by equations (1.1). These parametrizations are needed, since the likelihood function as a mapping attaching some value in $\mathbb{R}$ to pairs of transfer functions $(k, l) \in \Theta$ does not lead directly to tractable estimation algorithms.
- In section 2.4 the main statistical properties of the ML approach are stated.
- In section 2.5 we will deal with the actual implementation of the likelihood optimization based on approximations to the likelihood function, which decreases the computational load.

2.2 Parametrization of the pair (k, l)

The problem of maximizing the likelihood as given in equation (2.3) cannot be solved explicitely in general. It will prove to be convenient to introduce a parametrization of the set Θ, since this reduces the problem to the well understood minimization of a (scalar) function of real valued parameters. A natural starting point to do this are the system matrices.

It is obvious, that two sets of system matrices, which are mapped onto the same pair (k, l) under π, give the same value of the likelihood function. Thus by simply using all entries of the system matrices as parameters, there will be problems with multiple optima in the optimization, which would require special optimization algorithms, taking into account the structure of the problem (see McKelvey, 1995). Thus the rest of this section will deal with the question of how to define a bijection between transfer functions and system matrices in order to avoid these problems. Bijectivity, however, will not be the only interesting property of the parametrization.

REMARK 2.2.1 In the following we will assume, that k and l are parametrized as indicated by the system equations (1.1) i.e. the dynamics of the process is modelled by the matrix A. This is done simply because it fits better into the general framework of the thesis. However it is straightforward to apply the structure theory given in this section also to the case, where the two transfer functions k and l are independently parametrized. Whenever there are significant differences between these two approaches, we will remark this. In order to avoid clumsy notation we will identify the pair (k, l) with the transfer function k of appropriate dimensions from now on, i.e. $k(z) \in \mathbb{C}^{s\times(s+m)}$. The corresponding system matrices are (A, B, C, D), $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times(m+s)}$, $C \in \mathbb{R}^{s\times n}$, $D \in \mathbb{R}^{s\times(m+s)}$, where $D = [D_1, D_2]$, $D_1 \in \mathbb{R}^{s\times m}$, $D_2 \in \mathbb{R}^{s\times s}$ lower triangular with positive diagonal entries. Also we will use the notation M_n for the set of all stable rational transfer functions $k(z)$ of order n and dimension $s \times (s + m)$ and drop the minimum-phase assumption for the moment. Whenever there are problems with this notation, we will note this.

DEFINITION (CANONICAL FORM) A bijective mapping φ attaching to every $k \in M_n$ system matrices (A, B, C, D), $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times(m+s)}$, $C \in \mathbb{R}^{s\times n}$, $D \in \mathbb{R}^{s\times(m+s)}$, such that $\pi(\varphi(k)) = k$, is called *canonical form*.

The construction of a canonical form has to face several problems, due to the nonuniqueness of the system representation with system matrices: As has already been stated, to every transfer function there exist system matrices of various state dimensions. However, for every rational transfer function, there exists a minimal dimension, which can be seen as follows:

For given system matrices (A, B, C, D) introduce the following two matrices: the *controllability matrix* $\mathcal{C} = [B, AB, A^2B, \cdots]$ and the *observability matrix* $\mathcal{O} = [C^T, A^TC^T, (A^2)^TC^T, \cdots]^T$.

DEFINITION (OBSERVABILITY AND CONTROLLABILITY) A pair (A, C) is called *observable*, if the observability matrix is of full column rank. A pair (A, B) is called *controllable*, if the controllability matrix is of full row rank.

A realization (A, B, C, D) is minimal, if and only if (A, C) is observable and (A, B) is controllable. The product of these two matrices is a (block) *Hankel matrix*:

$$\bar{\mathcal{H}} = \mathcal{O}\mathcal{C} = \begin{bmatrix} CB & CAB & CA^2B & \cdots \\ CAB & CA^2B & \cdots & \\ CA^2B & & & \\ \vdots & & & \end{bmatrix} \tag{2.4}$$

This Hankel matrix will play a major role in the following. Note that the blocks in the Hankel matrix are just the coefficients of the power series corresponding to k. It can be shown, that for a minimal representation the dimension of the matrix A is equal to the rank of $\bar{\mathcal{H}}$. Once the order of a system is determined, there still exists the freedom to choose the basis of the state vector $x(t)$.

DEFINITION (OBSERVATIONAL EQUIVALENCE) The set $E(k) = \pi^{-1}(k) \cap S_n$, where n is the order of k, is called *the class of observationally equivalent systems*. Every point in $E(k)$ is called *minimal realization of* k. Two minimal realizations of a transfer function k are called *observationally equivalent*, which is denoted by $(A_1, B_1, C_1, D_1) \sim (A_2, B_2, C_2, D_2)$.

From the decomposition of the Hankel matrix into observability and controllability matrix, it follows that for two observationally equivalent minimal realizations of k, (A_1, B_1, C_1, D_1) and (A_2, B_2, C_2, D_2) say, there exists a nonsingular matrix $T \in \mathbb{R}^{n \times n}$ such that the two realizations are connected by a state space transformation (2.1). A mapping φ is a canonical form, if and only if from $(A_1, B_1, C_1, D_1) \sim (A_2, B_2, C_2, D_2)$ it follows, that $\varphi(\pi(A_1, B_1, C_1, D_1)) \sim (A_2, B_2, C_2, D_2)$. Thus canonical forms pick a representative out of every equivalence class.

DEFINITION (PARAMETRIZATION) A bijective mapping ψ attaching vectors $x \in \mathbb{R}^d$ for some integer $d < \infty$ to the set U of transfer functions $k \in M_n$ is called *parametrization of* U. The vector $x = \psi(k)$ is called *parameter vecotr*, the entries of x are called *parameters*. If the set $\psi(M_n^j)$ is open in $\mathbb{R}^d$, then the parameters are called *free*.

For the 'well-posedeness' of the problem of optimizing the likelihood, we also have to prove some topological properties of the parametrizations: In the ML approach, the maximum of the likelihood has to be calculated. Since there is no closed form solution to this problem, the maximum is calculated using numerical optimization procedures. For the analysis of the properties of the resulting estimates we have to know the connections between the topological structure of the parameter spaces and of the set of all transfer functions M_n: Sets of transfer functions will be endowed with the so called *pointwise topology* T_{pt} as follows: Identify k with the sequence $(K(j))_{j \in \mathbb{N}+}$, where $\mathbb{N}^+ = \mathbb{N} \cup \{0\}$. Sequences of matrices $(K(j)) \in \mathbb{R}^{s \times s+m}$ are endowed with the product topology in $(\mathbb{R}^{s \times s+m})^{\mathbb{N}^+}$ and the pointwise topology T_{pt} is the corresponding relative topology. A norm, which will be used frequently in the following, is the *Hankel norm* $\|.\|_{\bar{\mathcal{H}}}$, which is defined as the maximal singular value of the Hankel matrix $\bar{\mathcal{H}}$ (for a definition of the SVD for operators see Appendix B). Finite dimensional spaces are endowed with the Euclidean norm and compact subsets of finite dimensional spaces with the corresponding Hausdorff distance (see Appendix B). Since it is possible to prove consistency in the pointwise topology for the ML estimates of the transfer function independent of the parametrization (see Hannan *et al.*, 1980, comp. section 2.4), it will also be of interest to analyze the continuity of the parametrizations. It is well known, that for multi-input multi-output (MIMO) systems there does not exist a globally continuous parametrization of M_n, the set of all rational transfer functions of order exactly equal to n (see Hazewinkel and Kalman, 1976). Thus we have to break down M_n into several pieces, which are indexed by some integer valued parameters. Within the optimization, one also has to estimate this index. In this thesis we are confined to the concept of parametrization using canonical forms. Overlapping parametrizations are another alternative. The idea of overlapping parametrizations is to use various parametrizations of overlapping pieces of M_n, such that the parameter spaces have the same dimension and that for each transfer function k in M_n there exists at least one parametrization, which is continuous in the neighborhood (in M_n in the sense of the pointwise topology) of k. In the identification, one is free to choose the one, that shows the best numerical properties. Overlapping echelon parametrizations are discussed in (Hannan and Deistler, 1988), overlapping balanced parametrizations are presented in (Hanzon and Ober, 1997).

In the following we will present two different concepts leading to two different classes of parametrizations: In the next subsection we will describe the echelon canonical forms and cite the results corresponding to the items mentioned above. This approach is included because these parametrizations are widely used and a rather complete theory has been established for these parametrizations. Afterwards we will deal with (several) balanced parametrizations, where we

will establish results comparable to those, which have been obtained in the past for the echelon canonical forms. In the final subsection of this section, we then compare the two approaches corresponding to their properties, which are relevant for ML estimation. The properties of the parametrizations, which will be investigated in the following are:

- Definition of the pieces of M_n and the continuity of the parametrization of the pieces.
- Topological structure of the different pieces and the relations in between different pieces.
- The structure of the parameter sets and the properties of their boundaries.

The continuity of the parametrization will be of importance, since Theorem 2.4.1 states the consistency of the ML estimates of the transfer function independently of the parametrization. Thus the continuity of the parametrizations can be used to prove the consistency of the estimates of the parameter values. In the optimization of the likelihood, certain boundary points cannot be avoided in general, i.e. when one optimizes the likelihood over some set of transfer functions, the optimization is in fact performed on the closure of the set. Thus the topological structure of the pieces, in particular their openess and their boundaries, is of importance. Finally the optimization has to be performed in the parameter space, and thus also in this set the boundary bounds cannot be avoided and have to be considered. Thus it is of importance to investigate the topological properties of the parameter spaces.

2.2.1 Parametrization based on Echelon Canonical Forms

Recall the definition of the (block) Hankel matrix $\bar{\mathcal{H}} = \mathcal{OC}$. The (i,j) block of $\bar{\mathcal{H}}$ is equal to $CA^{i+j-2}B$. Now the so called *echelon canonical forms* are based on a selection of linearily independent rows of the Hankel matrix. Denote the i-th row of the Hankel matrix with h_i. Consider a set of r linearily dependent rows $h_{i_1}, \cdots, h_{i_r}$ of $\bar{\mathcal{H}}$. Then it follows from the Hankel structure, that also the rows $h_{i_1+s}, \cdots, h_{i_r+s}$ are linearily dependent. A basis of the row space of $\bar{\mathcal{H}}$ can be described by a set of s integers $(n_1, \cdots, n_s)$ e.g. as follows: Construct a basis of the (finite-dimensional) rowspace of $\bar{\mathcal{H}}$ by selecting rows in a certain way: Consider first the rows $h_{(l-1)s+i}, l \in \mathbb{N}$ for some $1 \leq i \leq s$. Suppose $h_i, \cdots, h_{n_1 s+i}$ is a set of linearily independent rows, whereas $h_{(n_i+1)s+i} = \sum_{j=0}^{n_i-1} a_{ii}(j) h_{js+i}$. From the Hankel structure it follows, that all rows $h_{ls+i}, l > n_i$ are linearily dependent on the first n_i rows. Continuing this procedure with the rows $h_{(l-1)s+j}, l \in \mathbb{N}, j \neq i$ and so on, we obtain a basis of the row space described by s integers $\alpha = (n_1, \cdots, n_s)$. The system order is equal to the sum of these integers, which is denoted with $|\alpha| = \sum_{j=1}^{s} n_j$. However, these indices are not unique for given transfer function, and we have to introduce another restriction. There exists exactly one set of indices $(n_1, \cdots, n_s)$, such that the rows in the basis are indeed the first n rows of the Hankel matrix, that span the row space, in the sense that if in the sequence $I(n_1, \cdots, n_s) = \{h_1, \cdots, h_{(n_1-1)s+1}, \cdots, h_{(n_s-1)s+s}\}$ we replaced h_i with a row $h_j, j < i$, then the dimension of the subspace spanned by these rows is equal to $n-1$. Indices which fulfill this restriction are called *Kronecker indices*. Note, that the Kronecker indices are unique for given transfer function and can be used to construct a canonical form. Introduce a semi-ordering on the set of all multi-indices $(n_1, \cdots, n_s)$ by the following rule: $(n_1, \cdots, n_s) \leq (m_1, \cdots, m_s)$ iff $n_i \leq m_i, i = 1, \cdots, s$. This ordering coincides with the structure of the parametrization, as will be apparent in Theorem 2.2.1.

Let the Kronecker indices of the transfer function k be specified as $\alpha = (n_1, \cdots, n_s), |\alpha| = n$ and assume for notational convenience, that the Kronecker indices are such, that the first n rows of the Hankel matrix form the basis of the row span. Let the matrix $\mathcal{C}$ contain the n basis rows according to the Kronecker index α in the order used in the definition of $I(n_1, \cdots, n_s)$ and consider the corresponding decomposition of the Hankel matrix $\bar{\mathcal{H}}$ as $\bar{\mathcal{H}} = \mathcal{OC}$. Let $s(i)$ denote the index of the i-th row of $\mathcal{C}$ within the Hankel matrix $\bar{\mathcal{H}}$. Then in $\mathcal{O}$ the $s(i)$-th row is zero, with the exception of the i-th column, where the entry is equal to 1. The remaining elements of $\mathcal{O}$ follow from the fact, that the rows of $\mathcal{C}$ are a basis of the row space of $\bar{\mathcal{H}}$. Once the decomposition of the Hankel matrix is fixed, it is straightforward to extract the system matrices: B is obtained as the

first block column of $\mathcal{C}$, and thus contains elements of the Markoff parameters $K(i)$ directly. C is obtained from the first s rows of $\mathcal{O}$ and thus contains only zeros and in each row one entry equal to one or in the case $n_i = 0$ for some i also the matrix C contains some free parameters. The matrix A can be obtained using the shift invariance structure of the observability matrix: $\mathcal{O}A = \mathcal{O}^{\uparrow}$, where $\mathcal{O}^{\uparrow}$ denotes the matrix obtained by omitting the first block row of the observability matrix. From this a special structure of the matrix A can be deduced (comp. Hannan and Deistler, 1988): $A = (A_{ij})_{i,j=1,\cdots,s}$, $A_{ij} \in \mathbb{R}^{n_i \times n_j}$, where

$$A_{ii} = \begin{bmatrix} 0 & & & \\ \vdots & & I & \\ & & & \\ a_{ii}(0) & \cdots & \cdots & a_{ii}(n_i-1) \end{bmatrix}, A_{ij} = \begin{bmatrix} & \mathbf{0} & & \\ a_{ij}(0) & \cdots & a_{ij}(n_{ij}) & 0 \end{bmatrix}$$

with $n_{ij} = \min(n_i + 1, n_j)$ for $j < i$ and $n_{ij} = \min(n_i, n_j)$ for $j \geq i$. The $a_{ij}(l)$ are the coordinates of the row $h_{n_i s+i}$ in the coordinate system introduced by the basis $h_1, \cdots, h_{(n_s-1)s+s}$. For given Kronecker indices α we may define the set V_α, say, of transfer functions $k \in M_n$ with the property, that the Kronecker indices of k are equal to α. For each $k \in V_\alpha$ the parameter vector τ consists of all entries in the matrices (A, B, C, D), which are not a priori restricted to 1 or 0. Denote the set of all parameter vectors corresponding to $k \in V_\alpha$ with T_α. Note that the dimension of T_α depends on α. This set is equipped with the Euclidean norm. Then we may define a mapping $\psi_\alpha : V_\alpha \to T_\alpha$ attaching to every transfer function $k \in V_\alpha$ the vector of the free entries in the matrices (A, B, C, D). In the following we will denote with a bar the closure of the sets in the corresponding spaces i.e. $\overline{V_\alpha}$ denotes the closure of V_α in $\overline{M_n}$. The set $\{\psi_\alpha : |\alpha| = n\}$ is a canonical form, which is stated in the following result (see Hannan and Deistler, 1988, Theorem 2.5.3). Since the parameters are entries in the system matrices, we will use the symbol π also for the mapping attaching transfer functions to parameter vectors in this case, with slight abuse of notation.

Theorem 2.2.1 (Echelon Parametrizations) *The state space parametrization $\psi_\alpha : V_\alpha \to T_\alpha$ has the following properties:*

(i) T_α is an open subset of $\mathbb{R}^{d_\alpha}$, where $d_\alpha = |\alpha|(s+m+1) + \sum_{i,j,j<i} n_{i,j} + n_{j,i}$.

(ii) $\psi_\alpha : V_\alpha \to T_\alpha$ is a (T_{pt})- homeomorphism

(iii) $\{V_\alpha : |\alpha| = n\}$ is a disjoint partition of M_n containing $\begin{pmatrix} n+s-1 \\ s-1 \end{pmatrix}$ sets

(iv) $\pi(\overline{T_\alpha}) = \bigcup_{\beta \leq \alpha} V_\beta$

(v) V_α is (T_{pt})-open in $\overline{V_\alpha}$

(vi) $\pi(\overline{T_\alpha}) \subset \overline{V_\alpha}$ and equality holds for $s = 1$.

For a proof (see Hannan and Deistler, 1988). The significance of this theorem can be explained as follows: (i) states, that the entries of the parameter vectors contain in some real vector space of dimension d_α are free parameters in the sense, that the parameter space is open and thus not a 'thin' set, which is a necessary requirement for numerical optimization. (ii) implies the continuity of the mapping attaching parameters to transfer functions on the various pieces, which will be of importance for the statistical results. (iii) states, that the set of all functions ψ_α constitutes a canonical form. The last three points clarify the structure of the boundary of the pieces and the corresponding parameter spaces: From (iv) we conclude that to every transfer function $k \in V_\beta, \beta \leq \alpha$ there exists (at least) one point τ on the boundary of the parameter space T_α such that $\pi(\tau) = k$. The fact, that V_α is open in its closure will be important in connection with the consistency result. On the other hand, the semi-ordering of the Kronecker indices directly corresponds to the hierarchical structure of the parameter spaces, as is concluded from (vi): Systems, which correspond to Kronecker indices, which are not comparable to α in the semi-ordering, cannot be

approximated within the parameter space. On the contrary, the corresponding parameter vectors diverge to infinity (comp. Hannan and Deistler, 1988).

One disadvantage of the echelon canonical forms is that the stability and the strict minimum-phase restriction impose nonlinear constraints on the free parameters, which have to be monitored within the search procedure. Also there will be problems with nonminimal systems (comp. the discussion in Hannan and Deistler, 1988). This is one of the reasons for the introduction of other parametrizations.

2.2.2 Lyapunov Balanced Canonical Forms

In the last years there has been an increasing interest in so called balanced canonical forms and their use in system identification. The concept of balancing was first introduced by (Moore, 1981). The canonical forms presented in this section were derived in a number of papers (Ober, 1987; Ober, 1991; Ober, 1996; Kabamba, 1985; Enns, 1985; Desai *et al.*, 1985; Verriest and Kailath, 1983), and analyzed with respect to many different properties (e.g. in McGinnie, 1993; Maciejowski, 1985; Chou and Maciejowski, 1997; Pernebo and Silverman, 1982; Verriest and Gray, 1995) to name just a few of the most important contributions. One of the most attracting properties of balanced canonical forms are their model reduction capabilities as demonstrated in (Glover, 1984) for the continuous time case and in (Al-Saggaf and Franklin, 1987) for the discrete time case.

The basic idea of these canonical forms is to use the singular value decomposition to perform the decomposition of the Hankel matrix $\bar{\mathcal{H}}$ into $\mathcal{O}$ and $\mathcal{C}$. In the following sections, we will present three types of balancing, which are the most interesting for our purposes.

The first concept of balancing, we will investigate, is the original as proposed by (Moore, 1981). The main idea of balancing in the sense of Moore is to choose a basis of the state space, such that the components of the state are decoupled in some sense and ordered according to their 'importance' (for a discussion in this respect see Moore, 1981). Define the *observability Gramian* $W_o = \mathcal{O}^T\mathcal{O}$ and the *controllability Gramian* $W_c = \mathcal{C}\mathcal{C}^T$, which thus depends on the particular realization of the transfer function. Then balancing in the Moore sense is defined as follows:

> DEFINITION (LYAPUNOV BALANCING) A minimal realization (A, B, C, D) of a transfer function k is called *balanced in the Moore sense* (or Lyapunov balanced), if $W_o = W_c = \Sigma = diag(\sigma_1, \cdots, \sigma_n), \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$.

In the following, we repeat a few facts about balancing. First note that balanced realizations always exist (see Moore, 1981). One way to construct balanced realizations is to use the singular value decomposition (SVD) of the Hankel matrix $\bar{\mathcal{H}}$ (for the definition of the SVD for operators see Appendix B). Let $\bar{\mathcal{H}} = [U(\bar{\Sigma})^{1/2}][\bar{\Sigma}^{1/2}V^T]$ be the SVD of $\bar{\mathcal{H}}$ and choose $\mathcal{O} = U_n\Sigma_n^{1/2}, \mathcal{C} = \Sigma_n^{1/2}V_n^T$, where U_n denotes the first n columns of U, V_n the first n columns of V and Σ_n the heading $n \times n$ subblock of $\bar{\Sigma}$ with n equal to the minimal order of the system. Having fixed the decomposition, the system matrices are obtained from $\mathcal{O}$ and $\mathcal{C}$ as described for the echelon canonical forms. As easily can be seen, balanced realizations are not unique for given transfer function: This corresponds to the nonuniqueness of the SVD. Every basis transformation in the state space according to a nonsingular matrix T fulfilling $T\Sigma_n T^{-1} = \Sigma_n$ leads to another balanced realization. In particular for minimal (A, B, C, D), if all diagonal entries of Σ_n are different, the set of all feasible matrices T relating observationally equivalent balanced realizations consists of all diagonal matrices having only $+1$ or -1 as diagonal entries.

Using the structure of $\mathcal{O}$ and $\mathcal{C}$ the following pair of Lyapunov equations are easily obtained:

$$\begin{aligned} AW_cA^T - W_c &= -BB^T \\ A^TW_oA - W_o &= -C^TC \end{aligned} \tag{2.5}$$

The Lyapunov equations play an important role for the construction of balanced canonical forms. Note that for given W_c and B (W_o and C respectively) these equations are quadratic in A,

which causes serious complications for the construction of a parametrization. This is the reason for introducing a homeomorphism i defined for all stable systems by:

$$\begin{array}{rcl} (A,B,C,D) = i(A_c,B_c,C_c,D_c) & = & ((I-A_c)^{-1}(I+A_c), \sqrt{2}(I-A_c)^{-1}B_c, \\ & & \sqrt{2}C_c(I-A_c)^{-1}, (D_c + C_c(I-A_c)^{-1}B_c)U_c) \\ (A_c,B_c,C_c,D_c) = i^{-1}(A,B,C,D) & = & ((I+A)^{-1}(A-1), \sqrt{2}(I+A)^{-1}B, \\ & & \sqrt{2}C(I+A)^{-1}, (D - C(I+A)^{-1}B)U) \end{array} \tag{2.6}$$

Here U and U_c are orthonormal, block diagonal matrices, $U_c = diag(I_m, U_{c,1}), U = diag(I_m, U_1)$, where $U_1^T U_1 = I_s = U_{c,1}^T U_{c,1}$, which are only needed to obtain suitable restrictions for imposing the additional structure of D and D_c respectively in the case that the strictly minimum-phase transfer function k is involved: From Chapter 1 we know, that in order to achieve identifiability of the transfer function from input/output data, we have to impose additional restrictions on the matrix E e.g. to be lower triangular with positive entries on the diagonal. This restricts also the set of feasible matrices D_c. In order to obtain convenient restrictions of the parameters, we also restrict the trailing $s \times s$ submatrix of D_c to be lower triangular with positive diagonal elements. Note, that this difficulty is only relevant for the transfer function k and thus does not appear, if we decided to parametrize l independently. In this case $U_c = U = I$ is used.

The matrices (A_c, B_c, C_c, D_c) can be given an interpretation as the realization of a continuous time system of the form:

$$\begin{array}{rcl} \dot{x}(t) & = & A_c x(t) + B_c u(t) \\ y(t) & = & C_c x(t) + D_c u(t) \end{array} \tag{2.7}$$

Here the transfer function of the discrete time system $k(z)$ and of the continuous time system $g(s) = D_c + C_c(sI - A_c)^{-1}B_c$ (where s denotes the differential operator) are related by:

$$g(s) = k(z(s))$$

where

$$z(s) = \frac{1+s}{s-1} \tag{2.8}$$

Note that the bilinear transformation (2.8) is a bijection on the compactified complex plane, which maps the open left halfplane of the complex plane onto the open unit disc and gives a one-to-one relation between the poles and the zeros respectively of k and g. Clearly the homeomorphism i preserves system equivalence i.e. maps the observational equivalence classes $E(k)$ to the continuous time observational equivalence classes $E^c(k)$, which are defined analogously to $E(k)$.

Define the observability Gramian and the controllability Gramian for continuous time systems (2.7) as:

$$\begin{array}{rcl} W_c & = & \int_0^\infty e^{tA_c} B_c B_c^T e^{tA_c^T} dt \\ W_o & = & \int_0^\infty e^{tA_c^T} C_c^T C_c e^{tA_c} dt \end{array} \tag{2.9}$$

Now balancing can be defined completely analogously to the discrete time case. Note that W_c and W_o respectively are not changed by the homeomorphism (see Ober, 1987). Therefore the homeomorphism (2.6) also preserves balancing. It is easily verified that the following continuous time Lyapunov equations hold:

$$\begin{array}{rcl} A_c W_c + W_c A_c^T & = & -B_c B_c^T \\ A_c^T W_o + W_o A_c & = & -C_c^T C_c \end{array} \tag{2.10}$$

These equations are linear in A_c for given W_c and B_c (given W_o and C_c respectively), which is an essential advantage compared to the discrete time versions. The stability condition for a

continuous time system (2.7) is that all eigenvalues of the matrix A_c have real part strictly less than zero. Clearly since the poles (and the zeros) of $k(z)$ and $g(s)$ respectively are related by the bilinear transformation (2.8), the stability (and the strict minimum-phase) property are preserved by the homeomorphism.

Next the Lyapunov equations for the continuous time system are used to construct a canonical form. Since balancing does not uniquely fix a representative i.e. the set $E_b(k)$ of all observationally equivalent balanced realizations of k is no singleton, we have to introduce further restrictions. This will be done for the continuous time representations. For notational simplicity we will in the following omit the index $_c$.

First consider the case $\Sigma_n = \sigma I$: In this case $E_b(k)$ is defined by the class of all orthogonal basis transformations (2.1). Uniqueness can be obtained as follows: Commencing from an arbitrary balanced realization (A_1, B_1, C_1, D_1), we first apply an orthogonal basis transformation, which makes B_1 'positive upper triangular', i.e. there exist integers $1 \leq t_1 < t_2 < \cdots < t_r \leq (s+m)$, where $r \leq s$ is the rank of B_1, such that (x denotes arbitrary entries):

$$B = TB_1 = \begin{bmatrix} 0 & \cdots & 0 & b_{1,t_1} & x & \cdots & \cdots & x \\ 0 & \cdots & \cdots & 0 & b_{2,t_2} & x & \cdots & x \\ & & & & \cdots & & & \\ 0 & & \cdots & & 0 & b_{r,t_r} & x & x \\ & & & & & & & \\ & & & & \mathbf{0} & & & \\ & & & & & & & \end{bmatrix} \tag{2.11}$$

where $b_{i,t_i} > 0, \forall i$. This fixes the orthogonal transformation in the first r columns. In the considered case $\Sigma_n = \sigma I$, the Lyapunov equations simplify to:

$$\begin{array}{rcl} \sigma(A + A^T) & = & -BB^T \\ \sigma(A^T + A) & = & -C^TC \end{array} \tag{2.12}$$

This shows, that the symmetric part $A^s = \frac{1}{2}(A + A^T)$ of A is uniquely determined from B and that B and C^T are square roots of the same positive semidefinite matrix (hence the restriction $r \leq s$). Therefore we have $C = [U(B_rB_r^T)^{1/2}, \mathbf{0}]$, where B_r is the matrix of the first r rows of B, $(B_rB_r^T)^{1/2}$ denotes the uniquely defined positive definite square root of $(B_rB_r^T)$, the r columns of $U \in \mathbb{R}^{s \times r}$ are a set of orthonormal vectors i.e. $U^TU = I_r$ and where $\mathbf{0}$ denotes the $s \times (n-r)$ nullmatrix.

Write A as $A = A^s + A^{sk}$, where A^{sk} is the skewsymmetric part of A. Up to now, we have fixed the orthogonal transformation in the first r columns. In order to obtain a canonical form, the orthogonal transformation has to be fixed also in the last $n-r$ columns. This can be done, using the skewsymmetric part A^{sk} of A. In order to do this, we define the *r-balanced form* (refer to Ober, 1987):

$$A^{sk} = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$A_{22} = \begin{bmatrix} 0 & \alpha_2 & & & & \\ -\alpha_2 & 0 & \alpha_3 & & & \\ & -\alpha_3 & 0 & & \mathbf{0} & \\ & & \ddots & \ddots & \ddots & \\ & \mathbf{0} & & & 0 & \alpha_{n-r} \\ & & & & -\alpha_{n-r} & 0 \end{bmatrix}$$

$$A_{12} = \begin{bmatrix} \vdots & \vdots & & & & \\ & & & a_{g_2-1,h_2-1} & a_{g_2-1,h_2} & \cdots \\ & & & a_{g_2,h_2-1} & a_{g_2,h_2} & 0 \\ & & & \vdots & 0 & 0 \\ a_{g_1-1,h_1} & a_{g_1-1,h_1+1} & \cdots & a_{g_1-1,h_2-1} & & \\ a_{g_1,h_1} & 0 & \cdots & 0 & & \\ 0 & \cdots & & & & \end{bmatrix}$$

$$A_{21} = -A_{12}^T$$

where $A_{11} \in \mathbb{R}^{r\times r}, A_{12}, A_{21}^T \in \mathbb{R}^{r\times(n-r)}, A_{22} \in \mathbb{R}^{(n-r)\times(n-r)}, a_{g_i,h_i} > 0, \alpha_i \geq 0$ and the q indices h_i correspond to the case $\alpha_i = 0$. It can be shown, that for given skewsymmetric matrix this form is determined uniquely and can be achieved by orthogonal transformations of the form $diag(I_r, Q)$, where $QQ^T = I_{n-r}, Q \in \mathbb{R}^{(n-r)\times(n-r)}$.

This result can be used to choose the representative of the equivalence class $E_b(k)$: Consider a transfer function $k \in M_n$, where the balanced Gramian is equal to σI. Then there exists exactly one realization with the property, that for the (continuous time realization) B is positive upper triangular and A^{sk} is in r-balanced form, where A^{sk} denotes the skewsymmetric part of A. For this realization $C = [U(B_r B_r^T)^{1/2}, \mathbf{0}]$.

For the case of general Σ_n, the structure of the Lyapunov equations (2.10) makes it possible to apply the above result: Partition the matrices (A, B, C, D) according to the multiplicities of the second order modes. Then the subsystems (A_{ii}, B_i, C_i, D), where A_{ii} denotes the i-th diagonal block of A, B_i is the i-th block row of B and C_i denotes the i-th block column of C, are balanced with Gramian $\sigma_i I$. Thus these subsystems can be brought into the canonical form for systems having $\Sigma_n = \sigma I$ derived in the last paragraph. The (block) offdiagonal entries of A can be calculated from the Lyapunov equations as $A_{ij} = \frac{\sigma_j B_i B_j^T - \sigma_i C_i^T C_j}{\sigma_i^2 - \sigma_j^2}$. This leads to the following result (for a proof see Ober, 1987)

Theorem 2.2.2 (Ober) *Let $k \in M_n$ have singular values $\sigma_1 > \sigma_2 > \cdots > \sigma_k$ of multiplicities $n_1, \cdots, n_k, \sum_{j=1}^k n_j = n$. Then there exists exactly one continuous time realization of the form:*

- $B = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_k \end{bmatrix}$, *where the B_i are positive upper triangular*
- $A = (A_{ij})_{i,j=1,\cdots,k}, A_{ii} = -\frac{1}{2\sigma_i} B_i B_i^T + A_{ii}^{sk}, A_{ij} = \frac{\sigma_j B_i B_j^T - \sigma_i C_i^T C_j}{\sigma_i^2 - \sigma_j^2}$ *and the A_{ii}^{sk} are in r-balanced form.*
- $C_i = [U_i(B_{i,r}B_{i,r}^T)^{1/2}, 0], U_i^T U_i = I_{r_i}, r_i \leq \min(n_j, s)$ *is the rank of B_i and B_{i,r_i} denotes the matrix of the first r_i rows of B_i.*
- $D \in \mathbb{R}^{s\times(s+m)}$.

Conversely, for every given set of integers $n, k, n_1, \cdots, n_k$ and $t_{j,i}, r_i \leq \min(n_i, s), h_{j,i}, g_{j,i}, q_j$ defining the structure of the positive upper triangular and r-balanced matrices, the free parameters are given by

- $\sigma_1 > \cdots > \sigma_k$
- *the entries in B, which are not restricted to be zero by the positive upper triangular form, $(B_i)_{j,t_{j,i}} > 0, i = 1, \cdots, k, j = 1, \cdots, r_i$*
- *the entries in the r-balanced form of the skewsymmetric part of A, $\alpha_{j,i} \geq 0$ and $[A_{ii}]_{g_{j,i},h_{j,i}} > 0, i = 1, \cdots, k, j = 1, \cdots, q_i$*

- *free parameters for the orthogonal matrices* $U_i \in \mathbb{R}^{s\times r_i}, i = 1, \cdots, k$
- $D \in \mathbb{R}^{s\times s+m}$

result in a minimal and stable system $(A, B, C, D), \pi(i(A, B, C, D)) = k \in M_n$.

REMARK 2.2.2 The structure of the matrix $D_c = [D_1, D_2]$ *in our case imposes the restrictions* $(D_2)_{i,j} = 0, i < j, (D_2)_{i,i} > 0$, *where* $(D_2)_{i,j}$ *denotes the* (i, j) *element of* $D_2 \in \mathbb{R}^{s\times s}$.

The significance of the theorem is that it defines a canonical form and suggests a parametrization for stable transfer functions. The strict minimum-phase property introduces a nonlinear constraint on the parameter space, which leaves an open subset. The rest of this subsection is confined to the analysis of this canonical form and to establish a corresponding parametrization. The main goal will be to obtain a result similar to the result for the echelon parametrizations given in Theorem 2.2.1. A multi-index in the case of balanced parametrizations is rather complicated to be described in full generality. It has to contain all the structural information needed to specify the structure of the positive upper triangular and r-balanced forms. We will not attempt to be as general as possible and we will divide the integers into three groups: Let $\gamma = (\alpha, \beta, \gamma^*)$ be the collection of all integer parameters, where $\alpha = (n_1, \cdots, n_k), \beta$ contains all the other structural information for A and B, i.e. integers to specify the structure of the positive upper triangular forms and the r-balanced form of the skewsymmetric part of A (if needed). Finally γ^* corresponds to the integer parameters, which will turn out to be necessary for the parametrization of the orthonormal matrices U_i. The reason for this decomposition will be clear from the proofs. The general idea is that α represents the structure that comes from the balancation, whereas the additional restrictions to ensure identifiability are expressed in β and γ^*. Note, that due to the bijectivity of i the canonical form for the continuous time systems defines a canonical form also for the discrete time case.

In order to investigate the continuity properties of the canonical form derived above, we will heavily use results of operator theory, which can be found in Appendix B. The main idea in the proof of continuity is to follow the construction of the canonical form as defined above: Let U_γ denote the set of all transfer functions $k \in M_n$, such that the structure index is equal to γ, and let T_γ denote the set of corresponding parameter vectors, which will be defined below. The parametrization (using the same symbols as for the echelon canonical form) can be seen as a concatenation of two different mappings: $\psi_\gamma : U_\gamma \to T_\gamma, \psi_\gamma(k) = \rho_\gamma(A, B, C, D)$, where $(A, B, C, D) = \varphi_\gamma(k)$ denotes the (discrete time) canonical form corresponding to k, which is obtained from the SVD of the Hankel matrix $\bar{\mathcal{H}}$, and ρ_γ denotes the mapping attaching the free parameters to system matrices in the canonical form. Thus the mapping can be described by the following sequence of mappings:

$$k \leftrightarrow (\bar{\mathcal{H}}, D) \leftrightarrow (\mathcal{U}_n, \Sigma_n, \mathcal{V}_n^T, D) \leftrightarrow E_b(k) \leftrightarrow (A, B, C, D) \leftrightarrow \tau$$

where τ denotes the vector of free parameters and the $\leftrightarrow$ indicate bijective functions. $E_b(k)$ denotes the equivalence class of all (discrete time) balanced minimal realizations of k, (A, B, C, D) denotes the balanced canonical form. For the mapping attaching $(\mathcal{U}_n, \Sigma_n, \mathcal{V}_n, D)$ to $(\bar{\mathcal{H}}, D)$ to be defined correctly the symbols $\mathcal{U}_n$ and $\mathcal{V}_n$ denote a set of subspaces corresponding to the SVD i.e. $\mathcal{U}_n = \{\bar{U}_i, i = 1, \cdots, k\}$, where $\bar{U}_i = span\{u_{i,1}, \cdots, u_{i,n_i}\}$. Here u_i denotes a singular vector of $\bar{\mathcal{H}}$ associated with singular value σ_i of multiplicity n_i and $u_{i,1}, \cdots, u_{i,n_i}$ denote any basis of the space of all singular vectors associated with σ_i. As a topology for k we consider the pointwise topology T_{pt}, sets of operators are endowed with the operator norm, finite dimensional spaces with the Euclidean metric. $\mathcal{U}_n$ and $\mathcal{V}_n$ are endowed with products of the so called *gap topology* (see Appendix B).

The set of all balanced equivalence classes $E_b(k)$ is also endowed with the Hausdorff metric induced by the Frobenius norm on the set of all system matrices (A, B, C, D), which is possible, since the classes $E_b(k)$ are compact due to the compactness of the set of all orthogonal matrices. It will prove necessary to break up the set M_n into pieces according to equal configuration of the multiplicities of the second order modes, i.e. according to equal indices α. Let $U_\alpha \subset M_n$ denote

the set of all transfer functions k, whose second order modes have multiplicities described by α. In the analysis of the properties of the canonical form the next lemma will be essential:

Lemma 2.2.1 *i) Let the Hankel norm $\|.\|_{\bar{\mathcal{H}}}$ for a transfer function be defined as the largest second order mode. Then the topology induced by this norm on M_n is equivalent to the pointwise topology T_{pt}.*

ii) The mapping attaching to every $k \in M_n$ its second order modes $\sigma_1 \geq \cdots \geq \sigma_n > 0$ is T_{pt}-continuous.

iii) For $k \in U_\alpha, \alpha = (n_1, \cdots, n_k)$ let $\bar{\mathcal{H}}_k = (K_{i+j-1})_{i,j \in \mathbb{N}}$ denote the corresponding block Hankel matrix and let $\bar{\mathcal{H}}_k = U_n \Sigma_n V_n^T$ be a corresponding SVD, where $\Sigma_n = diag(\sigma_1 I_{n_1}, \cdots, \sigma_k I_{n_k})$. Then the mapping attaching to $\bar{\mathcal{H}}_k$ the spaces $\mathcal{U}_n, \mathcal{V}_n$ spanned by the singular vectors corresponding to identical singular values is continuous. Thereby U_α is endowed with the pointwise topology and $\mathcal{U}_n, \mathcal{V}_n$ are endowed with products of the gap metric.

PROOF: ad i) It is straightforward to see, that on M_n the Hankel norm is topologically equivalent to the Frobenius norm of a transfer function, the latter being defined as $\|k\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$. Now consider a sequence $k_T \in M_n$, which converges to $k_0 \in M_n$ in Frobenius norm. Then clearly $K_T(i) \to K_0(i), \forall i$ and thus $k_T \to k_0$ in the pointwise topology. Conversely consider a sequence $k_T \in M_n$, which converges in the sense of the pointwise topology to $k_0 \in M_n$. Taking an overlapping parametrization (see Hannan and Deistler, 1988) we have a continuous function φ' (defined in the neighborhood of k_0) attaching system matrices (A, B, C, D) to k. Thus $\varphi'(k_T) \to \varphi'(k_0) = (A_0', B_0', C_0', D_0')$. Since $|\lambda_{max}(A_0')| < 1$ holds, there exists a neighborhood of (A_0', B_0', C_0', D_0') where $|\lambda_{max}(A)| \leq c < 1$ and thus the difference of the Frobenius norm of the infinite Hankel matrix of the transfer function and the Frobenius norm of a suitably chosen top left corner can be made arbitrarily small. Since the elements of the top left corner converge by convergence in pointwise topology the first part of the lemma has been proved.

ad ii) Consider a sequence $k_T \to k_0$, where $k_T, k_0 \in M_n$ and convergence is in the pointwise topology and thus also in Hankel norm (see part i)). Let $\bar{\mathcal{H}}_T$ and $\bar{\mathcal{H}}_0$ denote the corresponding Hankel matrices. As easily can be seen, $\bar{\mathcal{H}}_T \to \bar{\mathcal{H}}_0$ in operator norm implies $\bar{\mathcal{H}}_T \bar{\mathcal{H}}_T^T \to \bar{\mathcal{H}}_0 \bar{\mathcal{H}}_0^T$ in operator norm. Note that the eigenvalues of $\bar{\mathcal{H}}_T \bar{\mathcal{H}}_T^T$ are just the squared second order modes and the eigenvectors of $\bar{\mathcal{H}}_T \bar{\mathcal{H}}_T^T$ are the left singular vectors corresponding to $\bar{\mathcal{H}}_T$. Now the proof of the continuity of the mapping attaching the singular values to the Hankel matrices follows from Lemma B.2.1 given in Appendix B.

ad iii) From ii) and from the fact, that $k_T \in U_\alpha$ holds, we conclude the strongly stable convergence of $\bar{\mathcal{H}}_T \bar{\mathcal{H}}_T^T$ (for a definition see Appendix B). Then Lemma B.2.1 in Appendix B states the convergence of the invariant subspaces in the gap topology. □

In the next step continuity of the mapping φ_1 attaching balanced equivalence classes $E_b(k)$ to transfer functions k is considered. Recall, that the set of all balanced equivalence classes corresponding to M_n is equipped with the Hausdorff metric. Now the following theorem holds:

Theorem 2.2.3 *The mapping φ_1 attaching to every $k \in U_\alpha$ the equivalence class $E_b(k)$ is T_{pt} continuous.*

PROOF: Consider the SVD of $\bar{\mathcal{H}}_T = U_T \Sigma_T V_T^T$, where $\bar{\mathcal{H}}_T$ denotes the Hankel matrix corresponding to the transfer function k_T, and choose $\mathcal{O}_T = U_T \Sigma_T^{1/2}, \mathcal{C}_T = \Sigma_T^{1/2} V_T^T$ for some choice of U_T. Then balanced realizations can be calculated as follows: B consists of the first $m+s$ columns of $\mathcal{C}$ and C of the first s rows of $\mathcal{O}$. A can be calculated from the equation $\mathcal{O}A = \mathcal{O}^\uparrow$, where $\mathcal{O}^\uparrow$ denotes the matrix obtained by omitting the first s rows of $\mathcal{O}$. Without restriction of generality we only give the argument for the left invariant subspace, $U_{1,T}^-$ say, corresponding to the largest singular value $\sigma_{1,T}$ of $\bar{\mathcal{H}}_T$. $U_{1,T}^-$ is spanned by the corresponding columns of $\mathcal{O}_T$. From Lemma

2.2.1 iii) we see that $k_T \to k_0$ implies that $U_{1,T}^-$ converges to $U_{1,0}^-$ in the gap metric, where $U_{1,0}^-$ is the space spanned by the left singular vectors corresponding to $\sigma_{1,0}$, the largest singular value of $\bar{\mathcal{H}}_0$. Let $(u_{1,0}, \cdots, u_{n_{1,0}})$ denote an orthonormal basis for $U_{1,0}^-$. We next show, that we can define orthonormal bases $(u_{1,T}, \cdots, u_{n_1,T})$ for $U_{1,T}^-$ such that $u_{i,T} \to u_{i,0}, i = 1, \cdots, n_1$ in the ℓ^2 norm. For each $\varepsilon > 0$ there exists a T_0 such that $d_H(U_{1,T}^-, U_{1,0}^-) < \varepsilon, T > T_0$ (Here $d_H(.)$ denotes the distance defining the gap topology). Therefore there exist $u_{1,T} \in U_{1,T}^-$ such that $d_{\ell^2}(u_{1,T}, u_{1,0}) < \sqrt{2}\varepsilon$ (where d_{ℓ^2} denotes the distance induced by the ℓ^2 norm). Let us assume, that we have already fixed $u_{i,T} \in U_{1,T}^-, i = 1, \cdots, j$. Then the $(j+1)$-th basis vectors are selected as follows: in the same way as above choose $u'_{j+1,T}$ such that $d_{\ell^2}(u'_{j+1,T}, u_{j+1,0}) < \sqrt{2}\varepsilon$. Decompose $u'_{j+1,T}$ as $u'_{j+1,T} = p + p^\perp$, where $p \in span\{u_{i,T} : i = 1, \cdots, j\}$ and $p^\perp$ is orthogonal to $u_{i,T}, i = 1, \cdots, j$. It is straightforward to see, that $\|p\|_{\ell^2} < c\varepsilon$ holds for some constant c. Thus we have constructed a sequence of basis vectors such that $u_{i,T} \to u_{i,0}$ in the ℓ^2 norm.

Performing this procedure for every second order mode, we obtain a sequence $\mathcal{O}_T \to \mathcal{O}_0$ in Frobenius norm and thus also $\mathcal{C}_T \to \mathcal{C}_0$ follows. As indicated before, from $\mathcal{O}_T, \mathcal{C}_T$ and D_T we obtain balanced realizations (A_T, B_T, C_T, D_T) and as easily can be seen, the mapping attaching (A_T, B_T, C_T, D_T) to $(\mathcal{O}_T, \mathcal{C}_T, D_T)$ is continuous in Frobenius norm. Thus $(A_T, B_T, C_T, D_T) \to (A_0, B_0, C_0, D_0)$ follows. In the last step we show, that the convergence of the representatives (A_T, B_T, C_T, D_T) implies convergence of the equivalence classes $E_b(k_T)$ in the Hausdorff metric: Note that due to the orthogonal invariance of the Frobenius norm i.e. $\|A\|_{Fr} = \|QA\|_{Fr}$, where Q is an arbitrary orthogonal matrix, the distance of all elements within one balanced equivalence class to another balanced equivalence class is the same. Thus $\|(A_T, B_T, C_T, D_T) - (A_0, B_0, C_0, D_0)\|_{Fr} < \varepsilon$ for some representatives implies $\bar{d_H}(E_b(k_T), E_b(k_0)) < \varepsilon$, where $\bar{d_H}$ denotes the Hausdorff distance in the set of all balanced equivalence classes. □

However U_α cannot be parametrized in a continuous way using our canonical forms, i.e. the function φ_α attaching to every $k \in U_\alpha$ the balanced Ober form is not continuous. The discontinuities occur at transfer functions, where the corresponding continuous time canonical forms have zeros at certain entries in the matrix B_c or in the matrix A_c (note that we have restricted some elements to be positive). In order to show, that φ_α is not continuous on U_α consider the following example, which is simple but nevertheless typical. Consider $k_0 \in U_\gamma, \gamma = (\alpha, \beta, \gamma^*), \alpha = (1, \cdots, 1), \beta = (2, 1, \cdots, 1), \gamma^*$ to be defined below arbitrary i.e. $B_{0,11}$ is zero, again omitting the subscript $_c$, whereas $B_{0,12}$ has an arbitrary positive value. Now construct a sequence $k_T \in U_{\alpha,\beta'}, \beta' = (1, \cdots, 1)$ having balanced realizations (which are not necessarily in the balanced Ober form), where we change only the entry $B_{T,11}$ such that $B_{T,11} \to 0$ and $B_{T,11}$ is positive for even T and negative for odd T. Then in the corresponding Ober form (using the same symbols) $B_{T,12}$ is equal to $B_{0,12}$ for even T and equal to $-B_{0,12}$ for odd T. Thus the canonical form cannot be continuous at such a point.

Recall that β is defined as the multiindex determining the location of the zero elements in B_c and A_c in the Ober canonical form and let $U_{\alpha,\beta} \subset U_\alpha$ denote the set of all $k \in U_\alpha$, which correspond to β. For example consider the case, where all second order modes are distinct i.e. $\alpha = (1, \cdots, 1)$, then $\beta = (t_{1,1}, \cdots, t_{n,1})$, where $t_{i,1}$ is the index of the first column in the i-th row of B_c, which contains a nonzero entry.

Theorem 2.2.4 *The mapping $\varphi_{\alpha,\beta}$ attaching the Ober canonical forms to $k \in U_{\alpha,\beta}$ is T_{pt}-continuous on $U_{\alpha,\beta}$.*

PROOF: Note that $\varphi_{\alpha,\beta}$ is the restriction of $\varphi_1 \circ \varphi_2$ to $U_{\alpha,\beta}$. Here φ_2 is the mapping attaching to every balanced equivalence class its representative in the Ober canonical form. Thus it remains to show, that φ_2 restricted to $\varphi_1(U_{\alpha,\beta})$ is continuous. Since the mapping i is a homeomorphism, $E_b(k_T) \to E_b(k_0)$ is equivalent to $E_b^c(k_T) \to E_b^c(k_0)$, where $i(E_b^c(k)) = E_b(k), j = 0, T$ and where the set of equivalence classes of minimal stable balanced continuous time system matrices (A_c, B_c, C_c, D_c) again is endowed with the Hausdorff metric. For the sake of clarity of the proof,

we will start with the case, where all multiplicities n_i of the second order modes are equal to one. Remember that in this case the set of all feasible basis transformations is the set of all diagonal orthogonal matrices and thus the equivalence classes $E_b(k)$ are finite. Furthermore B_c can have no zero row due to the fact, that $n_1 = \cdots = n_n = 1$ and the assumed stability: Suppose that the i-th row of B_c is equal to zero. Then also the i-th column of C_c will be equal to zero. Now using the Lyapunov equations (2.10) it is observed, that $A_{i,j} = 0 = A_{j,i}, \forall j$ contradicting the stability assumption.
As has been pointed out already, in this case β is the vector of integers $t_{i,1}, i = 1, \cdots, n$. Here continuity of φ_2 on $\varphi_1(U_{\alpha,\beta})$ is straightforward to see: Let $E_b^c(k_T) \to E_b^c(k_0)$ for $k_0, k_T \in U_{\alpha,\beta}$, which implies that for every representative from $E_b^c(k_T)$ and thus in particular for the representative (A_0, B_0, C_0, D_0) satisfying the positivity requirements described by β, there is a convergent subsequence of representatives (A_T, B_T, C_T, D_T) from $E_b^c(k_T)$ converging to (A_0, B_0, C_0, D_0). Clearly then from a certain T_0 onwards (A_T, B_T, C_T, D_0) is in Ober canonical form.

In order to avoid clumsy notation, we will only give an outline of the proof in the general case. Since the feasible basis transformations are orthogonal block diagonal matrices, we without restriction of generality restrict the discussion to the first block. For simplicity of the argument first assume, that the first block of $B_{c,0}$ has full row rank, which implies that in the corresponding part of $A_{c,0}$ the r-balanced form results in no further restrictions. Now it is straightforward to see, that the positive upper triangularization used by Ober is continous for fixed β, since it can be represented by a product of Householder transformations, which depend continuously on the vector used to define them, at nonzero vectors (see e.g. Golub and Van-Loan, 1989). The additional structure of the r-balanced forms can also be obtained by a sequence of Householder transformations and thus similar arguments show the continuity of the mapping. □

In the next step we are concerned with the free parameters τ say for the matrices (A, B, C, D). A main difference between Ober canonical forms and echelon forms is that in the first case the matrices (A, B, C, D) and the free parameters are related in a rather complicated way, whereas for echelon forms the free parameters are just certain prescribed entries of (A, B, C, D). On the other hand, restrictions like minimality and stability have a much simpler form in the case of balanced canonical forms, as will be clear in the sequel.
For simplicity of presentation we will consider the case $k \in U_n^+$ in detail, where U_n^+ is the set of all $k \in M_n$, where $\alpha^+ = (1, \cdots, 1)$ and $\beta^+ = (1, \cdots, 1)$. We will only sketch the arguments for the other cases. Let $D_n^+ = \varphi_{\alpha^+,\beta^+}(U_n^+)$. As will be shown in the sequel, $D_n^+ \subset \mathbb{R}^{n^2+n(m+s)+sn+sm+s(s+1)/2}$ is a real analytic manifold of dimension $n(m+s) + ns + ms + s(s+1)/2$. We distinguish between the multi-output and the single-output case.

Consider the multi-output case first: Let $C_n^+ = i^{-1}(D_n^+)$. Clearly i defines a homeomorphism between D_n^+ and C_n^+. As has been stated already, the free parameters for $(A_c, B_c, C_c, D_c) \in C_n^+$ consist of $\sigma_1 > \sigma_2 > \cdots > \sigma_n > 0$, the entries in B_c, where $(B_c)_{i,1} > 0$, the entries in $D_c = [D_{c,1}, D_{c,2}]$, $(D_{c,2})_{i,j} = 0, j > i$, $(D_{c,2})_{i,i} > 0$ and the suitably chosen free parameters of the n vectors $U_i \in \mathbb{R}^s$ of length 1 corresponding to the columns of C_c. As is well known, the unit sphere in $\mathbb{R}^s$ is a real analytic manifold of dimension $s-1$, which cannot be described by one coordinate system. However removing one point (e.g. the 'southpole') from the unit sphere, we can define a diffeomorphism to $\mathbb{R}^{s-1}$. Let D_n^{++} denote the subset of D_n^+, where we omit those $(A, B, C, D) \in D_n^+$, which have columns of the corresponding matrix C_c pointing to the southpole. Taking the stereographic projection for the s-dimensional unit sphere we obtain (with some minor adjustments) vectors of free parameters $\tau = (\tau_\Sigma, \tau_{B^+}, \tau_{B^r}, \tau_C, \tau_{D^+}, \tau_{D^r})$ for $(A, B, C, D) \in D_n^{++}$, where $\tau_\Sigma = (\sigma_1 - \sigma_2, \cdots, \sigma_{n-1} - \sigma_n, \sigma_n)$, $\tau_{B^+} = ((B_c)_{1,1}, \cdots, (B_c)_{n,1})$, $\tau_{B^r} = ((B_c)_{1,2}, \cdots, (B_c)_{n,m})$, τ_C denotes the free parameters corresponding to the directions of the columns of C_c, $\tau_{D^+} = \{(D_c)_{1,m+1}, \cdots, (D_c)_{s,m+s}\}$ and $\tau_{D^r} = \{(D_c)_{1,1}, \cdots, (D_c)_{s,(m+s-1)}\}$. Again the additional structure for D is only needed for the transfer function k and not in the case, where l is parametrized independently. Thus τ has $d(n) = n+n(m+s)+ms+s(s+1)/2+(s-1)n = mn+2ns+ms+s(s+1)/2$ components. By $T_n^{++} \subset \mathbb{R}^{d(n)}$ we denote the set of all free parameters corresponding to D_n^{++} and finally let $U_n^{++} = \pi(D_n^{++}) \subset U_n^+$. Thus $T_n^{++} = \{(\tau_\Sigma, \tau_{B^+}, \tau_{B^r}, \tau_C, \tau_{D^+}, \tau_{D^r}) \in \mathbb{R}^{d(n)} : \tau_\Sigma >$

$0, \tau_{G^+} > 0, \tau_{D^+} > 0\}$. Let $\mathbb{R}_+$ denote the set of all nonnegative reals. It is straightforward that T_n^{++} is open and dense in $\mathbb{R}_+^{2n+s} \times \mathbb{R}^{nm+2n(s-1)+ms+s(s-1)/2}$. Note that in D_n^{++} no column of C_c can be zero, since the same holds for the rows of B_c. Thus it is straightforward to see, that the mapping $\rho_n^{++} : D_n^{++} \to T_n^{++}$, such that $\rho_n^{++}(A, B, C, D) = \tau$ is a homeomorphism and that D_n^{++} is open and dense in D_n^+. Using the homeomorphism between D_n^+ and U_n^+ we obtain, that also the mapping $\psi_n^{++} : U_n^{++} \to T_n^{++}$ such that $\psi_n^{++}(k) = \tau$ is a homeomorphism and that U_n^{++} is open and dense in U_n^+. More generally it is straightforward to see that D_n^+ is a real analytic manifold of dimension $d(n)$, whose charts may be derived in a straightforward manner from the usual charts of the unit sphere. This charts are indexed using the multiindex γ^*. Note that the manifold D_n^+ has a very simple structure, since for any sphere there is only one 'southpole'.

In the single-output case, the columns of the matrix C_c are just scalars, therefore the equation $B_{c,i}B_{c,i}^T = C_{c,i}^T C_{c,i}$ reduces to $C_{c,i} = \pm\|B_{c,i}\|_2$. Thus we may introduce $\overline{B_i} = B_{c,i}/\|B_{c,i}\|_2$. Using the convention, that $\overline{B_i}$ lies in the northern hemisphere, we obtain a continuous parametrization for the vectors $\overline{B_i}$ e.g. by using the stereographic projection. Thus we are able to choose the following set of $2n + m + 1 + mn$ free parameters for a system $(A, B, C, D) \in D_n^+$:

- $\sigma_1 > \cdots > \sigma_n > 0$
- n parameter vectors $\tau_{B,i} \in \mathbb{R}^m$ for the n vectors $\overline{B_i}$, $\|\overline{B_i}\|_2 = 1$.
- $C_{c,i} \neq 0, 1 \leq i \leq n$
- $D_c = [D_{c,1}, D_{c,2}], D_{c,2} > 0$

Note, that in the single-output case the mapping ρ attaching to $(A, B, C, D) \in D_n^+$ the vector $\tau = (\sigma_1 - \sigma_2, \cdots, \sigma_n, \tau_{B,1}, \cdots, \tau_{B,n}, C_{c,1}, \cdots, C_{c,n}, D_{c,1}, D_{c,2}) \in (\mathbb{R}_+ - \{0\})^{n+1} \times (\mathbb{R} - \{0\})^n \times \mathbb{R}^{nm+m} = T_n^+$ is already a homeomorphism. By choosing the parametrization of $\overline{B_i}$ suitably, the set of feasible parameter vectors T_n^+ is open and dense in $\mathbb{R}_+^{n+1} \times \mathbb{R}^{(n+1)m+n}$. We do not need to introduce D_n^{++} in the single-output case. However, in order to treat the multi-output and the single-output case with the same notation, we will in the single-output case use the symbol D_n^{++} for D_n^+ and analogously we use T_n^{++} and U_n^{++}.

The case of general (α, β) follows along the same lines: Consider the multi-output case first: For $k \in U_{\alpha,\beta}$ we define the parametrizations as described above, where the free parameters are composed of all distinct second order modes, of the free elements in B_c and D_c, the free elements corresponding to C_c and the entries in the r-balanced skewsymmetric part of A_c. Note that the parametrization of the orthonormal matrices corresponding to C_c requires the parametrization of unit spheres, which results in problems with 'southpoles' as described above. Removing these 'southpoles', we obtain a set $U_{\alpha,\beta}^{++}$ on which the mapping attaching the free parameters to transfer functions is a homeomorphism. By $T_{\alpha,\beta}^{++}$ we denote the set of all parameter vectors corresponding to $k \in U_{\alpha,\beta}^{++}$ and consequently we denote with $\rho_{\alpha,\beta}^{++}$ the homeomorphism attaching the parameter vector $\tau \in T_{\alpha,\beta}^{++}$ to a transfer function $k \in U_{\alpha,\beta}^{++}$. Note that in the case $rank(B_{c,i}) < s, i = 1, \cdots, k$, $U_{\alpha,\beta}^{++}$ is open and dense in $U_{\alpha,\beta}$, which follows from the fact, that $\varphi_{\alpha,\beta}$ is a homeomorphism (see Theorem 2.2.4) and the properties of the chosen parametrization of orthogonal matrices. However for $rank(B_{c,i}) = s$ for some i the parametrization of the orthogonal matrix $U_i \in \mathbb{R}^{s\times s}$ introduces disjoint sets of the same dimension, since after parametrizing $(s-1)$ unit vectors, the direction of the last is already determined, only the orientation may be chosen.

In the single output case again the situation is simpler: For second order modes of multiplicity larger than one, the corresponding vector $C_{c,i} = (c_{c,i}, 0, \cdots, 0)$, and the corresponding block row $B_{c,i}$ is zero with the exception of the first row. Thus there are no problems with southpoles as reported before for the multi-output case, and the parametrization as given above is valid also in this case. To be consistent with the notation however, we will also in this case use the symbols $U_{\alpha,\beta}^{++}$ and $T_{\alpha,\beta}^{++}$.

Note, that the exclusion of the 'southpoles' introduces another integer parameter and splits the set $U_{\alpha,\beta}$ into several pieces. However, the additional index is not included into β, but rather indicated by another multi-index γ^*. The case of no southpoles described above will be denoted by the superscript $^{++}$. The introduction of this new multiindex γ^* is done to distinguish between integer parameters due to discontinuities of the parametrization of the orthogonal matrices U_i from integers due to discontinuities of the mapping φ_α. The motivation for doing so lies in the fact, that the problem of parametrization of orthonormal matrices is well understood and that there are several possible solutions to it. The reader might want to use different parameters and therefore the structure indexed with γ^* will change.

The decomposition of the set of all transfer functions depends on the special parametrization chosen, which in this section is via Ober's balanced canonical forms. In order to cover the set M_n in this case a great number of pieces has to be used. Although we have no formal proof for this statement, a comparison of the necessary pieces for small n and s leads to this statement as well as the great number of integers needed to specify the parameter spaces. This means that in identification a great number (as compared to e.g. echelon forms) of integers (defining dynamic specification) has to be estimated prior to e.g. maximum likelihood estimation of the real valued parameters. However, as for the echelon case, there exists a set of integer parameters, that is 'generic' i.e. open and dense in M_n (see also Bauer and Deistler, 1996) and thus serves as a natural starting point for likelihood optimization.

Theorem 2.2.5 *U_n^{++} is open and dense in $\overline{M_n}$.*

PROOF: Let $\alpha = (1, \cdots, 1)$. We will first show, that U_α is open and dense in $\overline{M_n}$. The openness of U_α follows from the facts, that M_n is open in $\overline{M_n}$ (see Hannan and Deistler, 1988) and that the mapping attaching the second order modes to $k \in M_n$ is continuous (cf. Lemma 2.2.1, ii).

In the next step we show, that U_α is dense in $\overline{M_n}$. This part of the proof is along the lines indicated by (Kabamba, 1985). Let $S_n^+ \subset S_n$ denote the set of all minimal $(A, B, C, D) \in S_n$. Note that S_n^+ is an open and dense subset of S_n.

The squares of the second order modes are the solutions to the equation $\det(\lambda I - W_o W_c) = 0$. If the multiplicity of such a solution λ_1 is greater than one, also the equation $\frac{\mathbf{d}\det(\lambda I - W_o W_c)}{\mathbf{d}\lambda}|_{\lambda=\lambda_1} = 0$ holds and thus the matrix polynomials $a(\lambda) = \sum_{j=0}^{n} a_j \lambda^j = \det(\lambda I - W_o W_c)$ and $b(\lambda) = \sum_{j=0}^{n-1} b_j \lambda^j = \frac{\mathbf{d}a(\lambda)}{\mathbf{d}\lambda}$ have a common zero in this case. Therefore the determinant of the Sylvester matrix

$$R = \begin{bmatrix} a_0 & a_1 & \cdots & \cdots & a_n & 0 & \cdots & 0 \\ 0 & a_0 & a_1 & \cdots & \cdots & a_n & & \\ & & \ddots & & & & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & a_1 & \cdots & \cdots & a_n \\ b_0 & \cdots & \cdots & b_{n-1} & 0 & & \cdots & 0 \\ 0 & \ddots & & & & \ddots & & \\ & & \ddots & & & & & 0 \\ 0 & \cdots & 0 & b_0 & \cdots & \cdots & \cdots & b_{n-1} \end{bmatrix} \in \mathbb{R}^{(2n-1)\times(2n-1)}$$

is equal to zero in this case. Note that the function, $f : S \to \mathbb{R}$ say, attaching to every triple (A, B, C, D) the determinant of the Sylvester matrix R can be represented by an absolutely convergent power series in the entries of the triple (A, B, C, D) and thus is an analytic function. Assume that $\pi^{-1}(U_\alpha) \cap S_n$ were not dense in S_n, i.e. $S_n - [\pi^{-1}(U_\alpha) \cap S_n]$ contained a nontrivial open set. But then the analytic function f must be zero on S_n, which cannot be the case, as easily can be seen. Thus $\pi^{-1}(U_\alpha) \cap S_n$ is dense in S_n and by the continuity of the function π attaching transfer functions to the triples (A, B, C, D), the set U_α is dense in M_n and thus in $\overline{M_n}$.

Note that $U_n^{++} \subset U_\alpha$ is obtained from U_α by excluding these transfer functions, which correspond to a zero in the first column of B_c or to a southpole in the direction of the columns of C_c.

It is easy to see, that the omission of transfer functions, which correspond to a 'southpole' in the directions of the columns of C_c leaves an open subset. Thus we are left to deal with the case of zeros in the first column of B_c. Now assume, that U_n^{++} were not open in U_α. Then there exists a $k_0 \in U_n^{++}$, such that there exists a sequence $k_T \in U_\alpha - U_n^{++}$ with $k_T \to k_0$. Therefore there exists an index i and a subsequence again denoted by k_T such that $(B_{T,c})_{i,1} = 0$ holds for all T and every balanced (continuous time) realization $(A_{T,c}, B_{T,c}, C_{T,c}, D_{T,c})$ of k_T. Now by Theorem 2.2.3 the mapping φ_1 attaching to every $k \in U_\alpha$ the equivalence class $E_b(k)$ is continuous. Therefore we can select corresponding representatives (A_T, B_T, C_T, D_T) of k_T and (A_0, B_0, C_0, D_0) of k_0 such that $(A_T, B_T, C_T, D_T) \to (A_0, B_0, C_0, D_0)$ holds. But this would imply $B_{0,c,i1} = 0$, which is a contradiction to $k_0 \in U_n^{++}$.

For each $k_0 \in U_n - U_n^{++}$ it is straightforward to construct a sequence $k_T \in U_n^{++}$ converging to k_0 using the Ober canonical form. Therefore U_n^{++} is dense. □

The next theorem shows, that in the closure of D_n^{++} the same transfer functions are described as in the closure of U_n^{++} (see also Bauer and Deistler, 1996). This is a difference to e.g. echelon forms (see also section 2.3). If V_α denotes the generic neighborhood for the echelon forms and T_α the corresponding parameter space, then, for $s > 1$, $\overline{V_\alpha} - \pi(\overline{T_\alpha})$ is not empty (see Theorem2.2.1), i.e. in the echelon case there are transfer functions in $\overline{M_n}$ which can only be obtained as a limit of transfer functions in the generic neighborhood, where the corresponding parameters diverge to infinity.

Theorem 2.2.6 $\bigcup_{j \le n} M(j) = \overline{U_n^{++}} = \pi(\overline{D_n^{++}})$

PROOF: From the continuity of π we see that $\overline{U_n^{++}} \supset \pi(\overline{D_n^{++}})$. In order to prove that $\overline{U_n^{++}} \subset \pi(\overline{D_n^{++}})$ we proceed as follows: We have to show, that for every $k_0 \in \overline{U_n^{++}}$ there exists a sequence of matrices $(A_T, B_T, C_T, D_T) \in D_n^{++}$, such that $(A_T, B_T, C_T, D_T) \to (A_0, B_0, C_0, D_0)$. The continuity of π ensures, that $\pi(A_T, B_T, C_T, D_T) \to \pi(A_0, B_0, C_0, D_0) = k_0$. First let $k_0 \in \overline{U_n^{++}} \cap M_n$ and choose a sequence $(k_T)_{T \in \mathbb{N}} \in U_n^{++}$, which converges to k_0. Now again using an overlapping parametrization as in the proof of Lemma 2.2.1 we see, that there exists a sequence $(A_T, B_T, C_T, D_T) \to (A, B, C, D)$ such that $\pi(A_T, B_T, C_T, D_T) = k_T$ and $\pi(A, B, C, D) = k_0$. We now transform the sequence (A_T, B_T, C_T, D_T) to the uniquely defined Ober canonical form $(A_T^b, B_T^b, C_T^b, D_T)$ in discrete time, which defines a sequence of transformations T_T (comp. (2.1)). It is easy to see (cf. Moore, 1981) that T_T is of the form $T_T = \Sigma_T^{-1/2} U_T^T P_T$, where P_T denotes the upper triangular Cholesky factor of the observability Gramian $W_{o,T}$ and where $P_T W_{c,T} P_T^T = U_T \Sigma_T^2 U_T^T$ corresponds to the eigenvalue decomposition. Note that P_T and Σ_T continuously depend on $W_{o,T}$ and $W_{c,T}$ (for the latter see Appendix B). Thus the singular values of P_T and of Σ_T are bounded and also the sequence T_T is bounded. Therefore it is possible to choose a convergent subseries of T_T. From $\det T_T = \det P_T \det \Sigma_T^{-1/2}$ we conclude, that the limiting element of the subsequence is nonsingular. Thus there exists a sequence $(A_T^b, B_T^b, C_T^b, D_T) \in D_n^{++}$ (using T also as the index for the subsequence) which is convergent and where $\pi(A_T^b, B_T^b, C_T^b, D_T) \to k_0$.

For the case, that $k_0 \in \overline{M_n} - M_n$ i.e. k_0 corresponds to a system of order $n_1 < n$, we proceed as follows: First construct a sequence of matrices $(A_T^b, B_T^b, C_T^b, D_T) \in D_{n_1}^+$ converging to $(A_0, B_0, C_0, D_0) \in \mathbb{R}^{n_1^2 + n_1(m+s) + s(n_1+m) + s^2}$, $\pi(A_0, B_0, C_0, D_0) = k_0$ using the construction described in the first part of this proof. Next we augment this system using the structure of the balanced canonical forms: Using the homeomorphism i, we obtain a convergent sequence of matrices of continuous time systems $(A_{c,T}^b, B_{c,T}^b, C_{c,T}^b, D_{c,T}) \to (A_{c,0}, B_{c,0}, C_{c,0}, D_{c,0})$. Now the block structure of the Ober canonical form (cf. Theorem 2.2.2) makes it possible to augment the sequence of systems of order n_1 to a sequence of minimal systems $(A_{c,T}^{b,a}, B_{c,T}^{b,a}, C_{c,T}^{b,a}, D_{c,0})$ of order n in $C_n^{++} = i^{-1}(D_n^{++})$, without changing the limiting transfer function. This can be done e.g. by choosing the last $n - n_1$ second order modes smaller than the smallest second order mode of k_0, tending to zero at a certain rate, the last $n - n_1$ rows of $B_{c,T}^{b,a}$ such that the first entry in each row is positive and the length of the rows tends to zero at the same rate as the corresponding singular value, and finally choosing the last $n - n_1$ columns of $C_{c,T}^{b,a}$ as the transpose of the last rows of $B_{c,T}^{b,a}$.

Since for systems in C_n^{++} the matrix A is completely specified by Σ, B and C, and since the Ober canonical form assures stability and minimality, it follows that $(A_{c,T}^{b,a}, B_{c,T}^{b,a}, C_{c,T}^{b,a}, D_{c,T}) \in C_n^{++}$. Now the limit of this sequence is equal to

$$\left(\begin{bmatrix} A_{c,0} & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} B_{c,0} \\ 0 \end{bmatrix}, [C_{c,0}, 0], D_{c,0} \right)$$

which is a (nonminimal) realization of $g_0(s) = k_0(s(z))$. Using the homeomorphism i again, we obtain a sequence of system matrices $(A_T^{b,a}, B_T^{b,a}, C_T^{b,a}, D_T) \in D_n^{++}$, which converges to a matrix triple $(A_0, B_0, C_0, D_0), \pi(A_0, B_0, C_0, D_0) = k_0$. □

This clarifies the topological structure of the boundaries of U_α for $\alpha = (1, \cdots, 1)$. For the other indices α there will be many possible multiindices β. However, one is of particular importance.

Theorem 2.2.7 *For every α there exists a β, β^* say, such that U_{α,β^*} is open and dense in U_α.*

PROOF: For $\alpha = (1, \cdots, 1)$ the choice is $\beta^* = (1, \cdots, 1)$. The proof for this case has been given in the last part of the proof of Theorem 2.2.5. The proof for general α is in a certain sense analogous. Here, for given α, β^* is chosen in a way to allow for a maximum number of elements of B_c and A_c, which are not a priori restricted to zero. To be more detailed, such a generic β^* is described as follows: The entries of β^* corresponding to the positive upper triangular form of the i-th block of B_c, denoted with B_i, are of the following form (compare (2.11)): $t(j) = j, j = 1, \cdots, \min(n_i, s)$. The entries of β^* corresponding to the bidiagonal form of the skewsymmetric part of the diagonal blocks A_{ii} indicate (e.g. by a 1), that all entries in the two diagonals are nonzero (see Ober, 1991). Denseness of U_{α,β^*} in U_α can easily be shown. Openness is demonstrated for the special case, where $n_1 \geq (s + m)$. In this case for $k \in U_{\alpha,\beta^*}$ the corresponding block, $B_{c,1}$ say, in B_c is of rank $(s + m)$. By an analogous argument as in the proof of Theorem 2.2.5 it can be shown that for every $k_0 \in U_{\alpha,\beta^*}$ there exists a neighborhood, where the matrices $B_{T,c,1}$ corresponding to k_T in the neighborhood have rank s. Giving a similar argument for the other blocks of $B_{0,c}$ and the skewsymmetric part of $A_{0,c}$ it is straightforward to show, that U_{α,β^*} is open in U_α. □

Summing up these results, we can formulate a result similar to Theorem 2.2.1. Denote with ψ_γ the parametrization presented above, where $\gamma = (\alpha, \beta, \gamma^*)$ denotes the multiindex of the balanced canonical form and γ^* corresponds to the additional indices introduced by parametrizing the orthonormal matrices U_i. Let T_γ denote the corresponding parameter space and let $U_\gamma \subset U_{\alpha,\beta}$ denote the set of all transfer functions $k \in U_{\alpha,\beta}$ corresponding to the index γ^*. Also define $|\gamma| = n$ in analogy to the echelon case. The index γ thus can be decomposed into three parts: $\gamma = (\alpha, \beta, \gamma^*)$, where α corresponds to the multiplicities of the second order modes, β defines the structure of the positive upper triangular forms of the submatrices B_i and the r-balanced forms in the skewsymmetric part of A, and finally γ^* refers to the parametrization of the orthonormal matrices and therefore describes the charts corresponding to an atlas of unit spheres. Then the following theorem has been proven above, the analogon to Theorem 2.2.1 for the echelon case:

Theorem 2.2.8 *The state space parametrization $\psi_\gamma : U_\gamma \to T_\gamma$ has the following properties:*

(i) T_γ is an open subset of $\mathbb{R}^{d(\gamma)}$, where the dimension $d(\gamma)$ depends on the multiindex γ.

(ii) $\psi_\gamma : U_\gamma \to T_\gamma$ is a T_{pt} homeomorphism.

(iii) $\{U_\gamma, |\gamma| = n\}$ is a disjoint partition of M_n

(iv) V_γ is open in $\overline{V_\gamma}$.

PROOF: The first three points have been proved above. (iv) follows from the continuity of the second order modes (see Theorem 2.2.1) and from the fact, that for given α the systems in

$(\overline{V_\gamma} - V_\gamma) \cap U_\alpha$ the entries in A_c and B_c have to be restricted to be 0 according to the restrictions indexed with β, whereas the positivity constraint imposed by β are violated for some entries or there are problems with the parametrization of the orthonormal matrices U_i. All these restrictions leave an open set, which proves the claim.

correspond to points on the boundaries of $\overline{T_\gamma}$, at which the mapping attaching transfer functions to parameter vectors can be extended continuously. □

REMARK 2.2.3 Also for the balanced parametrizations it is possible to derive a semi-ordering as for the case of the Kronecker indices. Consider e.g. the case of $\alpha = (1, \cdots, 1)$. *Then the set of integer parameters* β *consists of the* n *indices of the first nonzero entry in the Ober canonical form for each row of* $B_{c,0}$ *i.e.* $\beta = (t_1, \cdots, t_n)$. *Then we define the semi-ordering as* $\beta = (t_1, \cdots, t_n) \leq \beta' = (t'_1, \cdots, t'_n) \Leftrightarrow t_i \geq t'_i$. *Corresponding to the second order modes, we define the following notation: Define the multi-index* $\alpha = (\alpha_1, \cdots, \alpha_n)$ *as* $\alpha_i = 0$, *if* $\sigma_i = \sigma_{i-1}$ *and* 1 *else, for* $i = 1, \cdots, n-1$ *and* $\alpha_n = 1$ *if* $\sigma_n > 0$ *and* $\alpha_n = 0$ *else, where* $\sigma_1 \geq \cdots \geq \sigma_n$ *denote the singular values represented according to their multiplicity. Then define* $\alpha = (\alpha_1, \cdots, \alpha_n) \leq \alpha' = (\alpha'_1, \cdots, \alpha'_n) \Leftrightarrow \alpha_i \leq \alpha'_i$. *Similar, one can define a semi-ordering for the index* γ^*. *In this semi-ordering the case* $\alpha = (1, \cdots, 1), \beta = (1, \cdots, 1), \gamma^* = ++$ *is the greatest index, which corresponds to the fact, that* U_n^{++} *is a generic set in this case. However, since this semi-ordering is very cumbersome to be written down carefully in full detail, we will refrain from discussing this any further. Also the results obtained for this ordering are not complete: Although it is immediate, that* $\overline{V_\gamma} \subset \bigcup_{\gamma' \leq \gamma} V_{\gamma'}$, *equality seems to be hard to establish and is left for future research.*

2.2.3 Minimum-phase Balanced Canonical Forms

The results in this section correspond mainly to the case of no observed inputs. Thus only the parametrization of k will be investigated. Recall, that in the model set treated in this thesis, we restrict the transfer function $k(z) = E + C(zI - A)^{-1}K$ to be strictly minimum-phase. Using the parametrizations introduced so far, this would lead to nonlinear restrictions in the parameter space, which is inconvenient for the actual implementation. A solution to this problem is to use parametrizations, which have the strict minimum-phase property 'built in'.

In this section, we will distinguish two sets of rational transfer functions. As above let M_n denote the set of all stable rational transfer functions of McMillan degree equal to n. $U_n^{(m)} \subset M_n$ will denote the set of all transfer functions $k \in M_n$, which additionally are strictly minimum-phase having a constant part, which is lower diagonal with strictly positive diagonal entries. The parametrization of the minimum-phase systems is based on the SVD of the matrix $(\Gamma^+)^{-1/2}\bar{\mathcal{H}}$, where $\bar{\mathcal{H}}$ denotes the Hankel matrix of the Markoff parameters $K(i) = CA^{i-1}K$ and $\Gamma^+ = \mathbb{E}Y(t)^+(Y(t)^+)^T$, where $Y(t)^+ = [y(t)^T, y(t+1)^T, \cdots]^T$. Equivalently the decomposition of the block Hankel matrix built of the covariances can be considered: Let $\gamma(j) = \mathbb{E}y(t)y(t-j)^T$ and define the block Hankel matrix $\mathcal{H} = (\gamma(j+i-1))_{i,j \in \mathbb{N}}$. Finally $\Gamma^- = \mathbb{E}Y(t)^-(Y(t)^-)^T$, where $Y(t)^- = [y(t-1)^T, y(t-2)^T, \cdots]^T$. Using this notation, it can be shown, that (for suitably chosen square roots) $(\Gamma^+)^{-1/2}\bar{\mathcal{H}} = (\Gamma^+)^{-1/2}\mathcal{H}(\Gamma^-)^{-T/2}$, where the particular choice of the square roots of the covariance matrices are of no importance for the following analysis. Thus the SVD's of these two matrices also coincide. From the results of realization theory (see e.g. Faurre, 1976) it follows, that $\gamma(j) = CA^{j-1}\bar{C}, j > 0$, where $\bar{C} = KE^T + APC^T = \mathbb{E}x(t+1)y(t)^T$ with $P = \mathbb{E}x(t)x(t)^T$. Thus it is obvious, that the rank of $\mathcal{H}$ will be equal to the rank of $\bar{\mathcal{H}}$ and that $\mathcal{H} = \mathcal{O}\bar{\mathcal{C}}$, where $\bar{\mathcal{C}} = [\bar{C}, A\bar{C}, \cdots]$. Now minimum-phase balancing can be defined in analogy to balancing in the Moore sense as follows:

DEFINITION (MINIMUM-PHASE BALANCING) A minimal realization (A, K, C, E) of a strictly minimum-phase transfer function k is called *minimum-phase balanced*, if $\mathcal{O}^T(\Gamma^+)^{-1}\mathcal{O} = \bar{\mathcal{C}}(\Gamma^-)^{-1}\bar{\mathcal{C}}^T = \Sigma = diag(\sigma_1, \cdots, \sigma_n)$, where $1 > \sigma_1 \geq \cdots \geq \sigma_n > 0$ are called *canonical correlation coefficients*. Here $\Gamma^+ = (\gamma(i-j))_{i,j\in\mathbb{N}}, \Gamma^- = (\gamma(j-i))_{i,j\in\mathbb{N}}, \gamma(j) = CA^{j-1}\bar{C}, j > 0, \gamma(j) = \gamma(-j)^T, j < 0$ and $\gamma(0) = C\Sigma C^T + EE^T$. $\bar{C} = KE^T + A\Sigma C^T$. Here Σ denotes the solution of the Lyapunov equation $\Sigma = A\Sigma A^T + KK^T$.

Note, that Σ exists due to the stability of A and that the strict minimum-phase condition for k ensures the invertibility of the covariance matrices Γ^+ and Γ^-. The motivation for the notion 'canonical correlation coefficients' is quite obvious in the present framework: Let $H_y^+(t)$ denote the Hilbert space spanned by the future of the process (see Appendix A for details) and let $H_y^-(t)$ be the Hilbert space spanned by the past. Then the σ_i are the canonical correlations between $H_y^+(t)$ and $H_y^-(t)$. This also gives a heuristic explanation of the restriction $\sigma_1 < 1$.

Inserting $K = (\bar{C} - APC^T)E^{-T}$ into the Lyapunov equation cited above leads to the algebraic Riccati equation (ARE)

$$P = APA^T + (\bar{C} - APC^T)E^{-T}E^{-1}(\bar{C} - APC^T)^T \tag{2.13}$$

This type of equations has been studied extensively in the literature. For the following discussion, we have to introduce another class of transfer functions:

DEFINITION (POSITIVE REALNESS) A stable transfer function $\Phi(z)$ is said to be positive real, if $\Phi(z) + \Phi(z^{-1})^T \geq 0, \forall |z| = 1$ and *strictly positive real*, if the equation is strict.

Now the following lemma has been proven in (Kalman *et al.*, 1966):

Lemma 2.2.2 *Let $\Phi(z)$ be a stable transfer function with realization $(A, \bar{C}, C, E)$. Then Φ is positive real, if and only if there exists a symmetric solution $P > 0$ to the ARE (2.13).*

(Faurre, 1976) investigated the solution set of the ARE, and found, that the set of all solutions to the ARE (2.13) for a strictly positive real rational transfer function $\Phi(z)$ is bounded in the sense, that for every solution the bounds $P^- \leq P \leq P^+$ hold, where the inequality is in the usual semi-ordering of symmetric matrices and P^- and P^+ denote the lower respectively the upper bound of the solution set. These bounds are of particular interest.

Let $2\pi f_y = \sum_{j=-\infty}^{\infty}\gamma(j)z^j = \sum_{j=-\infty}^{-1}\gamma(j)z^j + \gamma(0) + \sum_{j=1}^{\infty}\gamma(j)z^j = \Phi(z) + \Phi^T(z^{-1})$, where f_y denote the spectrum (extended to the complex plane) of the stationary process $(y(t))_{t\in\mathbb{Z}}$. Then it follows from the properties of the covariance sequence (see e.g. Lindquist and Picci, 1985), that $\Phi(z)$ is positive real. Thus, in this case, there exists at least one solution to the ARE (2.13). It has been shown by (Kalman *et al.*, 1966), that the minimal solution, i.e. the lower bound P^-, corresponds to the minimum-phase transfer function k, where $k(z)k^T(z^{-1}) = 2\pi f_y(z)$ i.e. the representation $(A, \bar{C}, C, \gamma(0)/2)$ of the spectral summand Φ and the representation (A, K, C, E) of the minimum-phase spectral factor k are related by: $\bar{C} = KE^T + AP^-C^T, \gamma(0) = CP^-C^T + EE^T$. Thus we are able to state the following lemma:

Lemma 2.2.3 (Spectral Factorization) *Let $\Phi(z)$ be a strictly positive real, stable, rational transfer function of order n with representation $(A, \bar{C}, C, \gamma(0)/2)$. Then there exists a minimal positive definite solution P^- to the algebraic Riccati equation (2.13). The transfer function $\pi(A, K, C, E)$ is strictly minimum-phase, where $K = (\bar{C} - AP^-C^T)E^{-T}$ and $EE^T = \gamma(0) - CP^-C^T$, E being the lower triangular Cholesky factor.*

Note, that together with $\Phi(z)$ also $\Phi(z)^T$ is strictly positive real and thus there also exists a factorization of $\Phi(z)^T + \Phi(z^{-1}) = k^-(z)k^-(z^{-1})^T$, which uses the inverse of the upper bound

P^+. This factorization of the spectrum corresponds to the backward representation (see Desai *et al.*, 1985). Lemma 2.2.3 states the existence of a mapping from the set $U_n^{(p)}$ of all stable, strictly positive real transfer functions of order n, whose constant term is symmetric, to the set $U_n^{(m)}$ of all stable and strict minimum-phase transfer functions of order n, whose constant term is lower triangular with positive entries on the diagonal. It will be used in the derivation of the parametrization of $U_n^{(m)}$. This will be incorporated by using a transformation of M_n onto $U_n^{(m)}$: From now on, all arguments will be stated for the continuous time representations, however we will omit the subscript $_c$ for notational convenience. For continuous time transfer functions, the strict minimum-phase assumption restricts the zeros to lie in the open left halfplane of the complex plane. The strict positive real assumption corresponds to $\Phi(i\omega) + \Phi^T(-i\omega) > 0, \omega \in \mathbb{R}$. The main sources for the results below are (Ober, 1996; Ober, 1991).

We need to define another concept of balancing, which however will not be commented any further:

DEFINITION (POSITIVE REAL BALANCING) A (continuous time) representation $(A, \bar{C}, C, D), D = D^T$ of a stable, strictly positive real transfer function $\Phi(s)$ is called *positive real balanced*, if the following two (continuous time) Riccati equations hold:

$$\begin{aligned} A_P\Sigma + \Sigma A_P^T &= -\Sigma C^T R_P^{-1} C\Sigma - \bar{C} R_P^{-1} \bar{C}^T \\ A_P^T\Sigma + \Sigma A_P &= -\Sigma \bar{C} R_P^{-1} \bar{C}^T \Sigma - C^T R_P^{-1} C \end{aligned} \tag{2.14}$$

where $A_P = A - \bar{C}R_P^{-1}C$ and $R_P = D + D^T$.

Let $M_{n,1}^c$ denote the set of all (continuous time) transfer functions $k \in M_n^c$, whose greatest second order mode σ_1 is smaller than 1. Here M_n^c denotes the set of all continuous time transfer function, which is related to M_n by the bijection i (see (2.6)). For this set (Ober, 1996) proves, that there exists a bijection I_P, that maps $M_{1,n}^c$ onto $U_n^{(p),c}$, the continuous time analogon to $U_n^{(p)}$. This mapping can be represented in terms of system representations as follows: Let (A, B, C, D) be a realization of the stable transfer function $k \in M_{1,n}^c$ with $AP + PA^T = -BB^T$, $A^TQ + QA = -C^TC$. Then:

$$I_P(A, B, C, D) = (A - B(BQ - C)(I - PQ)^{-1}, BR_P^{1/2}, R_P^{1/2}C(I - PQ)^{-1}, \frac{1}{2}R_P) \tag{2.15}$$

where $R_P = D + D^T$, is positive real. The solution P to the Lyapunov equation $AP + PA^T = -BB^T$ depends continuously on (A, B) and the same is true for Q and (A, C) respectively. Thus the mapping I_P is easily seen to be continous due to $P^{1/2}QP^{1/2} < I$, where $P^{1/2}$ denotes the uniquely defined positive definite square root of P. Using the tools provided in (Chou, 1994) it is possible to prove, that also the inverse mapping is continous. Thus the mapping in fact is a homeomorphism from $M_{n,1}^c$ onto $U_n^{(p),c}$. (Ober, 1996) further showed, that for a Lyapunov balanced system (A, B, C, D), the image $(A_P, B_P, C_P, D_P) = I_P(A, B, C, D)$ can be transformed easily to a positive real balanced representation: If we use $T = (I - \Sigma^2)^{1/2}$ as a basis transformation, where Σ denotes the Lyapunov balanced Gramian of the stable system, the corresponding system $(TA_PT^{-1}, TB_P, C_PT^{-1}, D_P)$ is positive real balanced with Gramian $\Sigma < 1$.

(Ober, 1987) states the following lemma, which can be seen as the continuous time analogon to Lemma 2.2.3:

Lemma 2.2.4 *The mapping S_P, defined by the following equations*

$$S_P(A_P, B_P, C_P, D_P) = (A_P, B_P, (2D_P)^{-1/2}(C_P - B_P^T P^-), \sqrt{2}D_P^{1/2}) \tag{2.16}$$

where $D_P = D_P^T > 0$ and $D_P^{1/2}$ denotes the lower triangular Cholesky factor of D_P, is bijective and preserves system equivalence. P^- denotes the minimal solution to the ARE (2.14).

Again using the results in (Chou, 1994) it is possible to show, that S_P is a homeomorphism preserving balancing i.e. (A_P, B_P, C_P, D_P) is positive real balanced if and only if $S_P(A_P, B_P, C_P, D_P)$ is minimum-phase balanced.

Thus a parametrization of the set $U_n^{(m)}$ of all rational, stable and strict minimum-phase transfer functions of order n can be obtained as follows: The homeomorphism i (ref. 2.6), attaches continous time system representations to discrete time system representations, the homeomorphism S_P^{-1} attaches positive real system representations to minimum-phase system representations, the homeomorphism I_P^{-1} attaches stable transfer functions to positive real transfer functions. Thus composing these homeomorphisms we obtain a homeomorphism attaching stable transfer functions with $\sigma_1 < 1$ to strictly minimum-phase transfer functions in $U_n^{(m)}$. The inverse, $\mathcal{T}$ say, of this homeomorphism can be written in terms of (continuous time) system representations as follows, (A, B, C, D) here denotes a Lyapunov balanced system:

$$(A^{(m)}, B^{(m)}, C^{(m)}, D^{(m)}) = (W^{-1/2}AW^{1/2} - B_n(B_n^T\Sigma - C_n), B_nR^{1/2}, C_n - B_n^T\Sigma, R^{1/2})$$

where $W = (I - \Sigma^2), B_n = W^{-1/2}B, C_n = CW^{-1/2}$ and $R = D + D^T$. $R^{1/2}$ here denotes the lower triangular Cholesky factor of R. The representation $(A^{(m)}, B^{(m)}, C^{(m)}, D^{(m)})$ is strictly minimum-phase balanced with Gramian Σ.

Now we may use the parametrization based on the Ober canonical forms to obtain a parametrization of $U_n^{(m)}$: The only remaining problems lies in the restriction $\sigma_1 < 1$, which has to be imposed for the mapping $\mathcal{T}$ to be defined properly. Define the mapping $\mathcal{S}$ attaching to every $k \in M_n$ a transfer function $k_1 \in M_{n,1}$ by transforming the second order modes as $p_i = \frac{\sigma_i}{\sqrt{1+\sigma_i^2}}$, where p_i denotes the second order modes of k_1 and σ_i denotes the second order modes of k. This mapping is a homeomorphism, as can be seen by the continuity of the mapping attaching singular values to transfer functions and the properties of the balanced parametrizations. This can be seen as follows: Note, that the results in Appendix B imply, that for $k_T \to k_0, k_0, k_T \in M_n$ the corresponding Hankel matrices converge in operator norm, and thus the singular values converge and the singular spaces converge in the gap metric. Thus it is possible to prove, that there exist balanced realizations (A_T, B_T, C_T, D_T) of k_T and balanced realizations $(A_{0,T}, B_{0,T}, C_{0,T}, D_{0,T})$ of k_0, such that the distance of these realizations converges to zero (in the embedding $\mathbb{R}^{(n+s+m)^2}$). This implies the same result for the continuous time analoga. Now consider the following transformation defined on M_n^c: Let $\sigma_i' = \sigma_i/\sqrt{1+\sigma_i^2}$, $B_i' = B_i, C_i' = C_i, D', D, A_{ii}^{sk,'} = (1+\sigma_i^2)^{3/2}A_{ii}^{sk}$. Clearly this defines a new continuous time realization (A', B', C', D'), and clearly this mapping attaches balanced equivalence classes to balanced equivalence classes. It is easily verified, that the convergence of the distance between (A_T, B_T, C_T, D_T) and $(A_{0,T}, B_{0,T}, C_{0,T}, D_{0,T})$ is equivalent to the convergence of the distance between the corresponding transformed realizations. Also the inverse can easily be constructed. This shows the claim.

Thus concatenating $\mathcal{S}$ and $\mathcal{T}$ we obtain a mapping $\mathcal{T}'$ attaching transfer functions $k_m \in U_n^{(m)}$ to transfer functions $k \in M_n$.

Since $\mathcal{T}'$ is bijective, the decomposition of M_n into the pieces $U_{\alpha,\beta}$ induces a decomposition of $U_n^{(m)}$ into the pieces $U_{\alpha,\beta}^{(m)}$. Since $\mathcal{T}'$ is a homeomorphism, all results concerning the topology in M_n also hold for $U_n^{(m)}$: Again the systems for $\alpha = (1, \cdots, 1), \beta = (1, \cdots, 1)$ are generic i.e. open and dense. For each piece characterized by α and β, there exists a subset $U_{\alpha,\beta}^{(m),++}$, which can be parametrized continuously (see Theorem 2.2.7). For each α there exists a β^*, such that $U_{\alpha,\beta^*}^{(m)}$ is open and dense in $U_\alpha^{(m)}$. The parametrization for $k \in U_n^{(m)}$ takes the following form (Ober, 1991):

Corollary 2.2.1 (Ober) *Given the integer parameters $n_1, \cdots, n_k, \sum_{j=1}^k n_j = n$ and the multiindex β as defined in section 2.2.2, the matrices $(A_c^{(m)}, K_c^{(m)}, C_c^{(m)}, E_c^{(m)})$ are determined from:*

a) $\sigma_1 > \cdots > \sigma_k > 0$. Define $p_i = \frac{\sigma_i}{\sqrt{1+\sigma_i^2}}$.

b) $K_c^{(m)} = \begin{bmatrix} K_{c,1} \\ \vdots \\ K_{c,k} \end{bmatrix} E_c^{(m)}$, *where the* $K_{c,i} \in \mathbb{R}^{n_i \times s}$ *are in positive upper triangular form, specified by* β. *Denote the rank of* $K_{c,i}$ *by* r_i.

c) $C_{c,i}^{(m)} = [U_i(K_{r,i}K_{r,i}^T)^{1/2} - p_i K_{r,i}^T, 0]$, *where* $K_{r,i}$ *denotes the matrix containing only the nonzero rows in* $K_{c,i}$ *and* $(.)^{1/2}$ *denotes the symmetric positive definite square root of a matrix.* $U_i^T U_i = I_{r_i}, U_i \in \mathbb{R}^{s \times r_i}$.

d) $A_{c,ii}^{(m)} = A_{c,ii}^{skew} - \frac{1+p_i^2}{2p_i} K_{c,i}K_{c,i}^T + K_{c,i}C_{c,i}$, $A_{c,ij}^{(m)} = \frac{1}{p_i^2 - p_j^2}(p_j(1-p_i^2)K_{c,i}K_{c,j}^T - p_i(1-p_j^2)C_{c,i}^T C_{c,j}) + K_{c,i}C_{c,j}$, $C_{c,i} = [U_i(K_{r,i}K_{r,i}^T)^{1/2}, \mathbf{0}]$. *Here* $A_{c,ii}^{skew} \in \mathbb{R}^{n_i \times n_i}$ *denotes a skewsymmetric matrix in r-balanced form.*

e) $E_c^{(m)}$ *lower diagonal with positive entries on the diagonal.*

For arbitrary parameter values, the resulting matrices $(A_c^{(m)}, K_c^{(m)}, C_c^{(m)}, E_c^{(m)})$ *correspond to a transfer function of McMillan degree n, which is strictly minimum-phase and stable.*

Note additionally, that the parameter space for the case $\alpha = (1, \cdots, 1), \beta = (1, \cdots, 1)$ and no U_i is equal to a 'southpole' is equal to T_n^{++}. Using this parametrization, the stability and the strict minimum-phase restriction thus result in simple nonnegativity restrictions (Using e.g. logarithmic transformations even these restrictions can be avoided).

2.2.4 Frequency weighted balancing

Frequency weighted balancing has been introduced by (Enns, 1985). In contrast to the previous sections, we will not try to establish a parametrization in this section, since the practical relevance of such a parametrization must be doubted. However, we will introduce a canonical form, since the structure theory provided in passing will turn out to be useful in the analysis of the subspace algorithms.

In this section, we will again adopt the notation of denoting the pair of transfer functions $(k, l), k(z) \in \mathbb{R}^{s \times s}, l(z) \in \mathbb{R}^{s \times m}$ by a single transfer function $k \in \mathbb{R}^{s \times (m+s)}$ (with slight abuse of notation). Consider again the Hankel matrix $\bar{\mathcal{H}}$ of the Markoff coefficients. Analogously to the last section, output frequency weighted balancing (see Enns, 1985) can be obtained by using the SVD of the matrix $W^+\bar{\mathcal{H}}$, where $W^+ = (W_{ij})_{i,j \in \mathbb{N}}$ has a special structure: $W_{ij} = K_W(i-j), K_W(l) = 0, l < 0$, $K_W(0)$ nonsingular. $K_W(l)$ are the Markoff parameters of some stable and rational transfer function $k_W(\omega) = \sum_{j=0}^{\infty} K_W(j) z^{-j}$, specifying the output weighting.

DEFINITION (FREQUENCY WEIGHTED BALANCING) A minimal system representation (A, B, C, D) is called *(output) frequency weighted balanced*, if $\mathcal{O}^T(W^+)^T W^+ \mathcal{O} = \mathcal{C}\mathcal{C}^T = \Sigma = diag(\sigma_1, \cdots, \sigma_n)$. $\sigma_1 \geq \cdots \geq \sigma_n \geq 0$ are called *frequency weighted singular values*.

Note, that in this setup also the Lyapunov balancing is included for $W^+ = I$ corresponding to $k_W(z) = I$. In the case, that $l \equiv 0$, the minimum-phase balancing turned out to provide simple parameter spaces. It is not straightforward, how to represent this type of balancing in the present context. However we are able to prove the following statement (see also McGinnie, 1993):

Lemma 2.2.5 *A minimal system representation* (A, K, C, E) *of a strictly minimum-phase transfer function* $k \in U_n^{(m)}$ *is minimum-phase balanced with Gramian* Σ, *if* $(TAT^{-1}, TK, CT^{-1}, E)$, *where*

$T = (I-\Sigma^2)^{-1/4}$, is frequency weighted balanced with Gramian $\bar{\Sigma} = \Sigma(I-\Sigma^2)^{-1/2}$ for the frequency weighting transfer function $k^{-1}(z)$.

PROOF: Consider a minimum-phase balanced representation (A, K, C, E). The corresponding observability matrix $\mathcal{O}$ then fulfills $\mathcal{O}^T(\Gamma^+)^{-1}\mathcal{O} = \Sigma$. Γ^+ is the covariance matrix of $Y(t)^+ = \mathcal{O}x(t) + \mathcal{E}E(t)^+$, where $\mathcal{E} = (\mathcal{E}_{ij})_{i,j\in\mathbb{N}}, \mathcal{E}_{ij} = K(i-j), K(j) = 0, j < 0$. $E(t)^+ = [\varepsilon(t)^T, \varepsilon(t+1)^T, \cdots]^T$. $\Sigma = A\Sigma A^T + KK^T$ denotes the covariance of $x(t)$. For both balancing concepts thus the same controllability part of the Lyapunov equations is used. Consider $\Gamma^+ = \mathcal{O}\Sigma\mathcal{O}^T + \mathcal{E}\mathcal{E}^T = \mathcal{O}\Sigma\mathcal{O}^T + R$, where R is easily seen to be nonsingular. Using the matrix inversion lemma we obtain with $x_\Gamma = \mathcal{O}^T(\Gamma^+)^{-1}\mathcal{O}$ and $x_R = \mathcal{O}^T R^{-1}\mathcal{O}$ the following equation:

$$x_\Gamma = x_R - x_R(\Sigma^{-1} + x_R)^{-1}x_R$$

This shows, that $x_\Gamma = \Sigma$ is equivalent to $x_R = \Sigma(I-\Sigma^2)^{-1}$ for $\sigma_1 < 1$. Now examine $R = \mathcal{E}\mathcal{E}^T$ more closely. The matrix $\mathcal{E}$ is nonsingular, since its (block)diagonal entries are equal to E, which was assumed to be nonsingular. The inverse of this matrix can be computed quite easily as $\mathcal{E}^{-1} = (\mathcal{E}^-_{ij})_{i,j\in\mathbb{N}}, \mathcal{E}^-_{ij} = K^-(i-j)$, where $K^-(j)$ denotes the power series coefficients of the transfer function $k^{-1}(z)$, which is stable due to the strict minimum-phase condition. Using $W^+ = \mathcal{E}^{-1}$ proves the statement. □

Thus minimum-phase balancing and frequency weighted balancing with the inverse transfer function coincide for the true transfer function up to a basis transformation in the state space with a diagonal scaling matrix.

The theory developed for stable systems in the previous sections can also be used for the case of frequency weighted balancing, as can be seen from the following arguments: The theorems in section 2.2.2 corresponding to the topological structure basically rely on two properties. One property we often use, is the continuity of the SVD. This continuity was derived on the basis of the results in (Chatelin, 1983, see also Appendix B), where the continuity of the SVD for a sequence of transfer functions $k_T \to k_0$ was shown using the convergence in norm of the Hankel operators $\bar{\mathcal{H}}_T$ corresponding to k_T to the Hankel operator $\bar{\mathcal{H}}_0$ corresponding to k_0. Here the sequence k_T as well as the limit k_0 were assumed to lie in various sets (e.g. M_n or $U_{\alpha,\beta}$) to obtain the continuity results of different quality (e.g. for the singular values or the equivalence classes). All these results also hold for arbitrary W^+ provided, the operator norm of W^+ is bounded, which is ensured by stability of the weighting transfer function $k_W(z)$. The second property often used is the analyticity of the Gramians as functions of the entries in the system matrices. This will be shown next: Let (A, B, C, D) denote a realization of $k(z) \in \mathbb{C}^{s\times(m+s)}$. The controllability Gramian W_c is defined as $W_c = \sum_{j=0}^{\infty} A^j BB^T(A^j)^T$ and thus obviously is analytic in the entries of A and B.

For the observability Gramian consider any realization (A_w, B_w, C_w, D_w) of the transfer function k_W. Then we may build the augmented system matrices:

$$\tilde{A} = \begin{bmatrix} A & 0 \\ B_wC & A_w \end{bmatrix}, \tilde{B} = \begin{bmatrix} B \\ B_wD \end{bmatrix}, \tilde{C} = [D_wC, C_w], \tilde{D} = D_wD$$

which is a (not necessarily minimal) realization of the stable and rational transfer function $k_W(z)k(z)$. It follows from simple but tedious calculations, that $\mathcal{O}^T(W^+)^TW^+\mathcal{O}$ is the heading $n \times n$ submatrix of the observability Gramian $\tilde{W}_o$ formed using the matrices $(\tilde{A}, \tilde{C})$. The stability of both, A and A_w, implies the stability of $\tilde{A}$, and thus the analyticity of the Gramian $\tilde{W}_o$ as a function of the entries in $(\tilde{A}, \tilde{C})$ follows and in particular the result for the heading $n \times n$ submatrix holds. This shows, that again we are able to provide a structure theory for the case of arbitrary stable and rational (output) frequency weighting. Again the set $U_n^+(k_W) \subset M_n$ defined by $k \in U_n^+(k_W)$ iff $k \in M_n$ and the frequency weighted singular values are distinct, is open and dense in M_n. We also may establish a canonical form $\varphi(k_w)$ attaching frequency weighted balanced system representations (A, B, C, D) to transfer functions $k \in M_n$ e.g. by choosing the submatrices of V^T corresponding to equal singular values to be positive upper triangular i.e. in

$W^+\bar{\mathcal{H}} = U\Sigma V^T$ the submatrices $V^T_{1:n_1}, V^T_{(n_1+1):(n_1+n_2)}, \cdots V^T_{(n-n_k+1):n}$ are positive upper triangular, where the indices denote the rows of V^T in **MATLAB** notation, e.g. $1:n_1$ denotes the rows with index $1, 2, \cdots, n_1$. For $k \in U_n^+(k_W)$ this implies, that in every row of V^T the first nonzero entry is positive. This leads in a natural way to a canonical form, indexed by two multiindices (α, β), where α relates to the multiplicities of the singular values and β determines the zero constraints in the positive upper triangular form. Note, however, that now the decomposition of M_n into pieces $U_{\alpha,\beta}$ of transfer functions corresponding to the multiindex (α, β) also depends on the weighting k_W.

Analogously to the Lyapunov balanced case, it is possible to prove, that the mappings $\varphi_{\alpha,\beta} : U_{\alpha,\beta} \to D_{\alpha,\beta}$ attaching the frequency weighted balanced canonical form $(A, B, C, D) \in D_{\alpha,\beta} \subset S_n$ to the transfer functions $k \in U_{\alpha,\beta}$ are continuous. Also note, that for $k_0 \in U_n^+(k_W)$ there always exists a mapping φ_{k_0} defined in some neighborhood $\mathcal{V}(k_0)$ of k_0, which attaches frequency weighted balanced realizations to transfer functions $k \in \mathcal{V}(k_0)$ and additionally is continuous on $\mathcal{V}(k_0)$. Such a mapping can be constructed by fixing one arbitrary nonzero entries in each column of V corresponding to k_0 to be positive. If the same entries are chosen positive for the matrices V corresponding to $k \in \mathcal{V}(k_0)$, then the mapping attaching the realizations corresponding to this choice of the orientation of the singular vectors contained in V to transfer functions $k \in \mathcal{V}(k_0)$ will be continuous (see Appendix B). The construction of this mappings will be heavily used in Chapter 4.

2.2.5 Model Reduction using Balanced Realizations

The primary motivation for including frequency weighting methods into this thesis lies in their model reduction capabilities and connections to subspace algorithms (see section 5.3). (Al-Saggaf and Franklin, 1987) show, that for the Lyapunov balanced case (which corresponds to $k_W = I$), a reasonable approximation to a high order transfer function can be found quite simple: Consider a partitioning of a Lyapunov balanced discrete time realization of a n-th order transfer function k according to $n_1 \leq n$, where the system order n is large:

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}, C = [C_1, C_2]$$

where $A_{11} \in \mathbb{R}^{n_1 \times n_1}, B_1 \in \mathbb{R}^{n_1 \times s}, C_1 \in \mathbb{R}^{s \times n_1}$ and the other matrices are of conformable dimensions. Then the infinity norm $\|k(e^{-i\omega}) - k_1(e^{-i\omega})\|_\infty = \max_{\omega \in [-\pi,\pi]} \bar{\sigma}(k(e^{-i\omega}) - k_1(e^{-i\omega}))$ (where $\bar{\sigma}(.)$ denotes the largest singular value) of the transfer function $k_1 = \pi(A_{11}, B_1, C_1, D)$ is smaller or equal to $2\sum_{n_1+1}^{n} \sigma_i$, where the σ_i denote the Lyapunov balanced singular values (comp. Al-Saggaf and Franklin, 1987).

Similar results also hold for the general frequency weighted case: (Kim *et al.*, 1995) give a proof, that in the continuous time case, truncation of the frequency weighted balanced system matrices (A_c, B_c, C_c, I) corresponding to the (continuous time) transfer function $G(s)$ of order n, where the frequency weightings are equal to $W(s)$ and $V(s)$ (Note that this is a more general setup, since they also consider input weightings $V(s)$, whereas in our context, we restrict ourselves to $V(s) = I$), leads to a reduced order model (A_r, B_r, C_r, I) corresponding to the transfer function $G_r(s)$, which is of order $r < n$, such that

$$\|W(s)(G(s) - G_r(s))V(s)\|_\infty \leq 2 \sum_{k=r+1}^{n} \sqrt{\sigma_k^2 + (\alpha_k + \beta_k)\sigma_k^{3/2} + \alpha_k\beta_k\sigma_k}$$

where α_k and β_k depend on the true system and on the weightings and where σ_k denotes the frequency weighted balanced singular values. The proof given in (Kim *et al.*, 1995) can be transferred also to the discrete time case in parts, however this is very cumbersome and thus omitted. We only remark, that it is possible to obtain bounds on the relative error, which are of the following form:

$$\|k_W(e^{i\lambda})(k(e^{i\lambda}) - k_r(e^{i\lambda}))\|_\infty \leq \sqrt{C_1\sigma_{r+1}^2 + C_2\sigma_{r+1}^{3/2}}$$

for some constants C_1 and C_2, which are independent of σ_{r+1}, but depend on the true (n-th order) system $k(z)$ as well as on the frequency weighting $k_W(z)$.

Another tool in the analysis of the model reduction properties using frequency weighted balanced realizations are the results of (Lu and Lee, 1985), which state that the norm of the matrices A_{21}, A_{12}, B_2 and C_2 in the above equations can be bounded by a constant times $\sigma_{r+1}^{1/2}$, i.e. times the square root of the first neglected singular value. Again the constant depends on the true system. This result has been extended to the frequency weighted case in (Diab *et al.*, 1997).

Note however, that the bounds, which contain a not further specified constant, which depends on the true high order system, may not be very useful for assessing the actual approximation error. There seems to be need for a better evaluation of the constants. Nevertheless we will make use of these concepts in section 5.3.2, where the effects of a misspecification of the order are investigated.

2.3 Comparison of the two Parametrizations

Having presented two alternative parametrizations for the set M_n of transfer functions of order n, a comparison between the two alternatives is in order. In this section we will try to provide a comparison corresponding to the following topics:

- Topological structure of the pieces and the parameter spaces corresponding to the pieces.
- Topological structure of the closures of the pieces.
- Properties of boundary points in the parameter spaces.

For the optimization of the likelihood, the topological structure of the parameter spaces has to be known, since we have to take into account possible restrictions. Furthermore, the structure of the parameter spaces determines the simplicity of the optimization algorithms. Knowledge of the topological structure of the pieces of M_n is important for the analysis of statistical properties of the ML estimates, as will be apparent from Theorem 2.4.1. In the optimization of the likelihood, certain boundary points cannot be avoided (comp. Hannan and Deistler, 1988). Thus it is important to know, which systems lie at the boundaries of the various pieces, i.e. the closures of the various pieces have to be investigated. Also hierarchical structures within the pieces, as demonstrated e.g. for the echelon case by Theorem 2.2.1, are a useful tool for designing efficient optimization algorithms (see e.g. Bauer and Deistler, 1997). And finally also the boundaries of the parameter spaces are important, due to the same reasons. Thus the analysis of the topological structure of the parametrizations is important for the analysis of the statistical properties of the ML estimates obtained by using the parametrizations, and for the design of optimization algorithms.

2.3.1 Topological Structure of the Pieces

For both parametrizations, there exist integer parameters defining a partitioning of M_n (and thus of $U_n^{(m)}$) into pieces, where we can find continuous parametrizations. In the case of the echelon parametrizations presented in section 2.2.1 this multiindex consists of the Kronecker indices. For the case of the balanced parametrizations, this multiindex consists of $\gamma = (\alpha, \beta, \gamma^*)$ (The motivation for the partitioning of this multiindex will become clear in the next sections).

All parameter spaces are open subsets of $\mathbb{R}^d$ for some integer d depending on the multiindex specifying the actual piece. For the case of the echelon forms, the parameters are entries in the matrices (A, B, C, D) and thus admit for a direct interpretation. Also there is a close relation to ARMAX parametrizations (see Hannan and Deistler, 1988), which are widely used in econometrics. However, nonlinear constraints have to be imposed due to the stability condition. From

the standpoint of the asymptotic theory presented in section 2.4 there is no reason to actually implement this nonlinear constraint: It can be shown, that for systems having a pole on the unit circle, asymptotically the value of the likelihood will tend to $-\infty$. Since the likelihood depends on the spectrum rather than on the transfer function k itself, unstable systems (i.e. systems having poles for $|z| > 1$) correspond to an equivalent stable spectral factor, giving the same value of the likelihood. Thus we may assume without restriction of generality, that for T_0 large enough the likelihood estimate will be stable and we do not have to impose the stability restriction explicitely. In finite samples however the restriction might well cause problems especially for gradient based algorithms. A more severe problem are the systems of lower order, which are represented by affine subsets in the parameter spaces. These subsets have to be taken into account during the optimization in order to avoid problems due to the nonuniqueness of the realization of the system. Note, that at each point on the affine linear subspace representing the equivalence class of a system of lower degree, the likelihood has the same value, thus e.g. gradient based methods will have problems in the neighborhood of these points. In this situation the estimate of the parameter vector will converge to the whole equivalence in the sense, that the distance of the estimated parameter vector to the equivalence class converges to zero. This does not imply convergence for the parameter vector itself (comp. the discussion in Hannan and Deistler, 1988).

For the case of the balanced parametrizations, the situation is quite different. The parameter spaces are open subsets of $\mathbb{R}^{d(\gamma)}$, which are bounded by simple nonnegativity restrictions. These nonnegativity restrictions can be easily implemented and monitored during the optimization. However, the price to be paid for that is that there is no direct interpretation of the parameters, since they are not just entries of the system matrices. This point might be judged to be minor, since we are dealing with black box modeling. The system matrices are obtained from the parameters in a rather complicated way, which also means that for the implementation of e.g. Gauss-Newton or gradient based techniques for the optimization of the likelihood, derivatives of the function attaching system matrices to parameter vectors have to be provided (for a discussion, see Chou and Maciejowski, 1997). On the other hand the parameters contain easily accessible information of the appropriateness of the integer parameters (see Bauer and Deistler, 1997).

2.3.2 Properties of the closures of the pieces

Corresponding to the boundaries of the pieces, the situation is clarified in the case of the echelon parametrization by Theorem 2.2.1: The closure of the set of all transfer functions corresponding to the Kronecker index α is equal to the union $\bigcup_{\beta \leq \alpha} V_\beta$ i.e. the semi ordering of the Kronecker indices corresponds to the nested structure of the pieces. For the single-output case, this is also true for the parameter spaces, i.e. every transfer function in $\overline{V_\alpha}$ can be represented by a point on the boundary of the corresponding parameter space T_α. For the multi-output case this result no longer holds. For every transfer function $k \in V_\beta \cap \overline{V_\alpha}$, $\beta \not\leq \alpha$ there is no point in $\overline{T_\alpha}$ representing k in the sense, that for $k_T \to k$, $k_T \in V_\alpha$, the corresponding parameter vectors will diverge to infinity in norm. This has to be taken into account for the identification procedure.

For the balanced canonical forms, the theory corresponding to the topology of the boundaries of the pieces is not so complete. Introduce a somewhat different notation for α, using a binary representation, which has also been used in the section following Theorem 2.2.8: $\alpha = (\alpha_1, \cdots, \alpha_n)$, $\alpha_i = 1$ if $\sigma_i \neq \sigma_{i+1}$ and $\alpha_i = 0$ else for $i = 1, \cdots, n-1$ and $\alpha_n = 0$, if $\sigma_n = 0$ and $\alpha_n = 1$ else. Define $\alpha' \leq \alpha$ iff $\alpha'_i \leq \alpha_i$ for all i. Then it is easy to see from the continuity of the mapping attaching the second order modes to the transfer functions $k \in M(n)$, that $\overline{U_\alpha} \subset \bigcup_{\alpha' \leq \alpha} U_{\alpha'}$ and that U_α is open in $\overline{U_\alpha}$. Denseness however is not easy to deal with. Theorem 2.2.5 states the denseness for the case $\alpha = (1, \cdots, 1)$. Corresponding to the other indices β and γ^* the situation is much clearer. Theorem 2.2.3 states, that the mapping attaching balanced equivalence classes to transfer functions $k \in U_\alpha$ is a homeomorphism on U_α. Thus once α is fixed, the topology in U_α can be analyzed using the analysis in S_n. There it is easy to see, that there exists a semi-ordering of indices β, which is consistent with the hierarchical structure of the sets $U_{\alpha,\beta}$, i.e. $\bigcup_{\beta \leq \beta'} U_{\alpha,\beta} \cap U_\alpha = \overline{U_{\alpha,\beta'}} \cap U_\alpha$. Note, that the structure indexed with β is due to the special choice taken by the Ober canonical form. Different choices leading to different balanced canonical forms will also lead to a different

partitioning of U_α. On the other hand, the partitioning due to α is due to using Lyapunov balancing, and thus the problems due to these indices cannot be avoided by any parametrization based on Lyapunov balancing.

Finally the situation for the index γ^* corresponds to the manifold structure imposed by parametrising sets of orthonormal vectors, which in our case is via stereographic projections. Again the structure induced by these indices depends heavily on the parametrization chosen and is not inherent to balanced canonical forms. This motivates the decomposition of the indices into α (imposed by the Lyapunov balancing), β (due to the Ober canonical form) and γ^* (relating to the parametrization of the orthonormal matrices U_i).

2.3.3 Properties of the Boundaries

For the echelon parametrizations the boundary points of T_α correspond in a natural sense to transfer functions with Kronecker indices $\alpha' < \alpha$, since π can be extended to the subset of S_n corresponding to $\overline{T_\alpha}$: Note, that the elements of the vectors $\tau \in \overline{T_\alpha}$ are entries in the matrices (A, B, C, D), and thus the convergence in the parameter space is directly related to convergence in S_n.

For the balanced parametrizations, however, the situation is more complicated. A simple, but nevertheless typical, example demonstrates the difficulties: Consider the class of systems given by the following continuous time system representations:

$$A_c = \begin{bmatrix} -\frac{1}{2} & -\frac{g}{2-\sigma} + \frac{2g}{(2-\sigma)\sigma} \\ -\frac{g}{2-\sigma} - \frac{1-\sigma}{2-\sigma}\frac{2g}{\sigma} & -\frac{g^2}{2(1-\sigma)} \end{bmatrix}, B_c = \begin{bmatrix} 1 \\ g \end{bmatrix}, C_c = [1, -g]$$

These systems have balanced Gramian $\Sigma = \mathrm{diag}(1, 1-\sigma)$. For $0 < \sigma < 1, g > 0$ the matrices correspond to $k \in U_2^{++}$. For $g = 0$ and $\sigma > 0$ the systems are not minimal and A_c is singular and thus not stable, even though the Lyapunov equations (2.10) are fulfilled with nonsingular Gramian. For $\sigma \to 1$, the smaller second order mode tends to zero, therefore the systems also converge (in Hankel norm) to a nonminimal system. For $\sigma \to 0$, the limit of A_c depends on the limiting behavior of $x_\sigma = \frac{g}{\sigma}$, however, the point on the boundary of T_n^{++} will be the same irrespective of x_σ. Since the offdiagonal entries in A_c are free parameters for $\alpha = (0, 1)$, different limits of x_σ lead to different limiting transfer functions. Note, that for $x_\sigma \to 0$ the limiting system is not minimal. Thus the points at the boundary of T_n^{++} may correspond to not just a single transfer function, but to a set of transfer functions including transfer functions of lower order. This observation has consequences for ML estimation: If one optimizes the likelihood over the parameter space $T_{\alpha,\beta}^{++}$, one cannot guarantee the existence of an optimum, since the parameter spaces are open sets. By maximizing the likelihood over the closure, we would have to extend the mapping $\pi \circ \psi^{-1}$ to the closure of $T_{\alpha,\beta}^{++}$. However, as is shown by the example, this is not possible in a continuous way. Thus the identification algorithm has to detect the case of misspecification of α. A statistically correct way of doing this is still a topic of research. Note, however, that information about the choice of α is contained in the estimates of the singular values σ_i. Misspecification of α can thus be detected by assessing the 'significance' of the estimate of $\sigma_i - \sigma_{i+1}$. The correct meaning of 'significance in this statement is still an open problem.

In U_α, the situation is much clearer: As has been stated already, there exists a semi ordering of indices β, which corresponds to the nested structure of U_α. Note, that β determines the elements, which are strictly positive. Points on the boundary of $T_{\alpha,\beta,\gamma^*}$, which do not fulfill these restrictions (but do fulfill the restrictions on the singular values) correspond to transfer functions in $\overline{U_{\alpha,\beta,\gamma^*}} \cap U_\alpha$: The mapping $\psi_{\alpha,\beta,\gamma^*}^{-1}$ can be extended continuously to be defined also at these points. Note, however, that the extended mapping will no longer be injective. The information contained in the restricted elements i.e. their 'significance' could be used to design efficient algorithms, where again the meaning of 'significance' is an unsolved question. Similar arguments apply to the case of southpoles. Note that here similar problems as in the case of the echelon forms for nonminimal systems indicated above might occur, since for multiplicities larger than one, the equivalence classes in $\overline{T_{\alpha,\beta,\gamma^*}}$ corresponding to systems with $\beta' < \beta$ might have a manifold structure. The difference

however lies in the fact, that the information about the appropriateness of the integer parameters is provided more easily accessible in the case of the balanced parametrizations.

2.4 Asymptotic Results

This section provides some results on the statistical properties of the ML approach. First we will investigate the consistency of the estimates, then we will turn to the asymptotic distribution of the estimates, and finally we will deal with the order estimation problem. The source for these results is (Hannan and Deistler, 1988). For the proofs see the original. The standard assumptions of this section will be the following: The input is required to be a pseudo stationary sequence (see Appendix A), persistently exciting in the sense that the transfer functions (k,l) can be obtained from the knowledge of the joint spectrum, and uncorrelated with the noise. The ergodic noise and the true transfer functions are assumed to fulfill the standard assumptions given in Chapter 1. Though some theorems also hold for weaker assumptions, we will always use these assumptions in order to provide a unified framework.

In order to investigate the consistency of the ML approach, we have to introduce certain assumptions on the set, over which the optimization is performed. The statements will be made coordinate free i.e. in terms of transfer functions rather than parameter vectors. Thus consider the set Θ of pairs of transfer functions (k,l) corresponding to the model class under consideration i.e. $k \in \bigcup_{j\leq n} U_j^{(m)}, l \in \bigcup_{j\leq n} M_j$. Then in the optimization of the likelihood certain boundary points cannot be avoided and we indeed optimize the likelihood over the bigger set $\hat{\Theta} = \{(k,l) \in \bar{\Theta} : (k(z),l(z))$ has no pole for $|z| = 1\}$. In the set Θ there could be pairs of transfer functions of different orders, however we have to assume, that $\Theta \subset \overline{M_n^j}$ for some n. For most of the results it is necessary to assume, that the true process lies in the model set i.e. $(k_0,l_0) \in \Theta$. This implies, that in these cases we assume, that an upper bound for the order is known a priori. Later in this section we will investigate procedures to estimate the order and in section 2.5 we will deal with the effects of underestimating the order.

Now we are ready to state a consistency result (see Hannan and Deistler, 1988)

Theorem 2.4.1 (Consistency of ML approach) *Let the assumptions $(k_0,l_0) \in \Theta$ and the standard assumptions hold and let the pseudo stationary input be persistently exciting of order sn in the sense of Appendix A. Then if $(\hat{k}_T, \hat{l}_T)$ denote the optimizers of the likelihood over $\hat{\Theta}$, the ML estimators are strongly consistent, that is:*

$$lim_{T\to\infty}(\hat{k}_T, \hat{l}_T) \to (k_0,l_0), a.s. \tag{2.17}$$

Note that the statement of the theorem is coordinate free and thus can be used for any parametrization at the continuity points: Consider the parametrization $\psi : k \to \tau$. Then, if the parametrization is continuous at (k_0,l_0), also the parameter vector will be a.s. consistent. This is the justification for the investigation of the continuity properties of the parametrizations provided in the last section.

In order to obtain results on the asymptotic distribution, we introduce a parametrization $\psi_\gamma : (k,l) \in U_\gamma \to \tau \in T_\gamma$, for some parameter space $T_\gamma \subset \mathbb{R}^d$, where $d \in \mathbb{N}$ is determined by the multiindex γ. Furthermore, for notational simplicity, we impose the restriction $\Theta = U_\gamma$ i.e. that the integer parameters have been specified correctly. As has been stated already, this means, that the likelihood optimization is performed over the corresponding set $\hat{\Theta}$. For the proof we will need the requirement, that U_γ is open in $\overline{U_\gamma}$, which has been shown for all the parametrizations presented in the previous sections. Then due to the strong consistency result given in Theorem 2.4.1, $(\hat{k}_T, \hat{l}_T) \to (k_0,l_0)$ will imply that $(\hat{k}_T, \hat{l}_T) \in \Theta$ for T large enough, where $(\hat{k}_T, \hat{l}_T)$ denotes the optimizing argument of the likelihood function in $\hat{\Theta}$. Thus we obtain $\hat{\tau}_T = \psi_\gamma(\hat{k}_T, \hat{l}_T)$ for T large enough.

In the sequel we also assume, that the transfer function is twice continuously differentiable with respect to the entries of the parameters. However, this seems to be a mild requirement, since it is fulfilled for all the parametrizations presented before: Note that $(k,l) = \pi(A,B,C,D,E,K)$ and π is easily seen to be differentiable in the entries of the matrices (A,B,C,D,E,K) as it only includes multiplications and a matrix inversion, which is differentiable, since the eigenvalues of A do not lie on the unit circle by the stability assumption. Now in the echelon case, the parameters are just entries in the matrices (A,B,C,D,E,K), whereas in the case of balanced parametrizations the situation is a bit more complicated. However, simple but tedious calculations show, that also in this case, the transfer function depends differentiably on the parameter vector. Moreover we will assume, that on T_γ, the parameter space corresponding to U_γ, the mapping attaching system matrices to parameter vectors is analytic. Again this is trivial for the echelon case and straightforward to show in the case of the balanced parametrizations. Then the following theorem holds:

Theorem 2.4.2 (Asymptotic Distribution) *Let (k_0,l_0) be an inner point of $\Theta = U_\gamma$ and let $\tau_0 \in T_\gamma$ be the corresponding parameter vector. Denote with $\hat{\tau}_T$ the optimizing argument of the likelihood function over the set $\overline{T_\gamma}$ given the observations $y_t, u_t, 1 \le t \le T$. Then under the standard assumptions a central limit theorem (CLT) holds for the ML estimator $\hat{\tau}_T$ i.e. $\sqrt{T}(\hat{\tau}_T - \tau_0)$ converges in distribution to a Gaussian random variable. Assume, that the parametrization is such, that $(\bar{k}_0, l_0)$, where $\bar{k}_0 = k_0 E_0^{-1}$, and E_0 are parametrized independently, i.e. that the vector τ can be partitioned as $\tau = (\theta,\sigma)$, where $(\bar{k}_0, l_0) = (\bar{k}_0(\theta), l_0(\theta))$ does not depend on σ and $E_0(\sigma)$ does not depend on θ. Then $\sqrt{T}(\hat{\theta}_T - \theta_0)$ and $\sqrt{T}(\hat{\sigma}_T - \sigma_0)$ are asymptotically uncorrelated, where the variance of $\sqrt{T}(\hat{\theta}_T - \theta_0)$ is equal to I_0^{-1}. Here I_0 denotes the Fisher information matrix, which is defined as:*

$$I_0 = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\{\frac{\partial \log L(Y^+_{1,T}, U^+_{1,T}; k(\tau_0), l(\tau_0))}{\partial \tau}(\frac{\partial \log L(Y^+_{1,T}, U^+_{1,T}; k(\tau_0), l(\tau_0))}{\partial \tau})^T\}$$

Here $L(Y^+_{1,T}, U^+_{1,T}; k(\tau_0), l(\tau_0))$ denotes the likelihood function as a function of the parameter vector τ_0 and the notation denotes the derivative evaluated at the parameter vector τ_0.

PROOF: We will only give a sketch of the proof. For full details see (Hannan and Deistler, 1988). Consider the likelihood function as given in equation (2.2). The transfer functions (k,l) enter the likelihood in two matrices (assuming that (k,l) generated the process): $\Gamma^+_T(k)$ and $\overline{Y^+_{1,T}}(l)$. Here $\Gamma^+_T(k) = \mathbb{E}(Y^+_{1,T} - \overline{Y^+_{1,T}}(l))(Y^+_{1,T} - \overline{Y^+_{1,T}}(l))^T = \mathbb{E}(\mathcal{O}_T x(0) + \mathcal{E}_T E^+_{1,T})(\mathcal{O}_T x(0) + \mathcal{E}_T E^+_{1,T})^T = \mathcal{O}_T \Sigma \mathcal{O}_T^T + \mathcal{E}_T \mathcal{E}_T$, where $\mathcal{O}_T = [C^T, A^T C^T, \cdots, (A^T)^{T-1} C^T]^T$, $E^+_{1,T}$ denotes the vector built of the noise $\varepsilon(t)$ analogous to $Y^+_{1,T}$, $\mathcal{E}_T = (K(i-j))_{i,j=1,\cdots,T}$, $K(j) = 0, j < 0$ and Σ denotes the state covariance. Here we used the orthogonality of the state and the future innovations as well as the white noise property of $\varepsilon(t)$. Note, that $y(t) = CA^{t-1}x(0) + \sum_{j=0}^{t}(K(j)\varepsilon(t) + L(j)u(t-j))$ and thus $v(t,l) = y(t) - \overline{y(t)}(l) = CA^{t-1}(x(0) - \overline{x(0)}(l)) + \sum_{j=0}^{t} K(j)\varepsilon(t-j)$, where $\overline{x(0)}(l)$ denotes the part of the initial state $x(0)$, which is due to the input process $(u(t))_{t\in\mathbb{Z}}$. Of course $K(j)$ and $L(j)$ denote the power series coefficients of the transfer functions $k(z)$ and $l(z)$ respectively. The two equations (see Hannan and Deistler, 1988)

$$\begin{aligned} \frac{1}{T}\log\det\Gamma^+_T(k) &= \log\det EE^T + o(\frac{1}{\sqrt{T}}) \\ \frac{1}{T}(V^+_{1,T}(l))^T(\Gamma^+_T(k))^{-1}(V^+_{1,T}(l)) &= \frac{1}{T}(V^+_{1,T}(l))^T(\mathcal{E}_T)^{-T}(\mathcal{E}_T)^{-1}(V^+_{1,T}(l)) + o_P(\frac{1}{\sqrt{T}}) \end{aligned}$$

imply

$$-\frac{2}{T}\log L(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; (k,l)) = \log\det EE^T + (V^+_{1,T}(l))^T(\mathcal{E}_T)^{-T}(\mathcal{E}_T)^{-1}(V^+_{1,T}(l)) + o_P(\frac{1}{\sqrt{T}})$$

where $V^+_{1,T}(l) = Y^+_{1,T} - \overline{Y^+_{1,T}}(l)$. Note that $(\mathcal{E}_T)^{-1}$ is a lower triangular block Toeplitz matrix, whose blocks are the coefficients of the power series of k^{-1} i.e. $(\mathcal{E}_T)^{-1} = (K^-(i-j))_{i,j=1,\cdots,T}, K^-(j) = 0, j < 0, k^{-1}(z) = E^{-1} + \sum_{j=1}^{\infty} K^-(j)z^{-j}$. Thus $\nu_t(k,l) = E\varepsilon_t(k,l)$, where $\varepsilon_t(k,l)$ denotes the t-th block row of $(\mathcal{E}_T)^{-1}(V^+_{1,T}(l))$ can be computed as $\nu_t(k,l) = y_t - Cx_t(k,l) - Du_t, x_{t+1}(k,l) = (A - KE^{-1}C)x_t(k,l) + (B - KE^{-1}D)u_t + KE^{-1}y_t$, where the dependence on the transfer functions k,l is made explicit in the notation. (A,B,C,D,E,K) is a realization of $(k,l) \in U_\gamma$. Now let $\psi_\gamma : U_\gamma \to \mathbb{R}^d$ be the parametrization attaching to every pair $(k,l) \in U_\gamma$ the corresponding parameter vector $\tau \in \mathbb{R}^d$. Then the log likelihood can be written as a function of the parameter vector τ rather than as a function of (k,l):

$$\begin{aligned}\mathcal{L}(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \tau) &= -\frac{2}{T}\log L(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \psi_\gamma^{-1}(\tau)) \\ &= \log\det EE^T + \frac{1}{T}tr[(EE^T)^{-1}(\sum_{t=1}^{T}\nu_t(\psi_\gamma^{-1}(\tau))\nu_t(\psi_\gamma^{-1}(\tau))^T)] + o_P(T^{-1/2})\end{aligned}$$

Here $tr[.]$ denotes the trace of a matrix. In the sequel we will simplify the notation by using $\nu_t(\tau) = \nu_t(\psi_\gamma^{-1}(\tau))$. This approximation to the likelihood builds the basis for the proof: Recall, that the ML estimate, $\hat{\tau}_T$ say, was defined as the maximizing argument of $L(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \tau)$ over $\overline{T_\gamma}$. Due to the assumption, that (k_0, l_0) is an inner point of Θ, the ML estimate $\hat{\tau}_T = \psi_\gamma(\hat{k}_T, \hat{l}_T)$ a.s. from some T_0 onwards, where $(\hat{k}_T, \hat{l}_T)$ denotes the ML estimate, where optimization is performed over $\hat{\Theta}$. Maximization of L is thus (for T large enough) equivalent to minimization of $\mathcal{L}$ with respect to the parameter vector τ. Since $\mathcal{L}$ is differentiable with respect to the parameter vector (if the system matrices are differentiable with respect to the parameters) and since τ_0 and therefore also $\hat{\tau}_T$ for T large enough are interior points of $\overline{T_\gamma}$, the derivative at $\hat{\tau}$ is equal to zero:

$$\frac{\partial\mathcal{L}(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \hat{\tau}_T)}{\partial\tau} = 0$$

where the notation indicates the derivative of $\mathcal{L}$ evaluated at $\hat{\tau}_T$. Now a mean value expansion at $\tau_0 = \psi_\gamma(k_0, l_0)$ gives

$$0 = \frac{\partial\mathcal{L}(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \tau_0)}{\partial\tau} + \frac{\partial^2\mathcal{L}(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \bar{\tau})}{\partial\tau\partial\tau^T}(\hat{\tau} - \tau_0) \tag{2.18}$$

Here $\bar{\tau}$ denotes some vectors of intermediate values with respect to $\hat{\tau}$ and τ_0 (not necessarily the same in each row). Thus (for invertible $\frac{\partial^2\mathcal{L}(Y^+_{1,T}, (u(t))_{t\in\mathbb{Z}}; \bar{\tau})}{\partial\tau\partial\tau^T}$, which will be shown afterwards) we obtain

$$\sqrt{T}(\hat{\tau} - \tau_0) = -(\frac{\partial^2\mathcal{L}(Y^+_{1,T}; (u(t))_{t\in\mathbb{Z}}, \bar{\tau})}{\partial\tau\partial\tau^T})^{-1}\left(\sqrt{T}\frac{\partial\mathcal{L}(Y^;_{1,T}(u(t))_{t\in\mathbb{Z}}, \tau_0)}{\partial\tau}\right)$$

and the central limit theorem is proved, if we show the asymptotic normality of $\sqrt{T}$ times the first derivative and the convergence of the second derivative to a nonsingular matrix. The convergence of the second order derivative can be shown, using the strong consistency result given in Theorem 2.4.1, which implies, that $\bar{\tau} \to \tau_0$ a.s. Ergodicity of $\varepsilon(t)$ and the assumed stability and the strict minimum-phase condition then give the result. Corresponding to the central limit theorem for the first order derivative we will use the approximation given above: $\mathcal{L}(Y^+_{1,T}; (u(t))_{t\in\mathbb{Z}}, \tau) = \log\det EE^T + \frac{1}{T}tr[(EE^T)^{-1}(\sum_{j=1}^{T}\nu_t(\tau)\nu_t(\tau))^T)] + o_P(T^{-1/2})$. The same approximation can be shown to hold for the derivative with respect to τ, implying, that the asymptotic distribution of $\sqrt{T}\frac{\partial\mathcal{L}(Y^;_{1,T}(u(t))_{t\in\mathbb{Z}}, \tau_0)}{\partial\tau}$ is equal to the asymptotic distribution of

$$2\sqrt{T}\{tr[(EE^T - (\sum_{j=1}^{T}\nu_t(\tau_0)\nu_t(\tau_0))^T))\partial(EE^T)^{-1}] + tr[(EE^T)^{-1}(\sum(\partial\nu_t(\tau_0))\nu_t(\tau_0)^T]\}$$

where ∂X is a short notation for $\frac{\partial X}{\partial \tau_i}$. Note, that $\nu_t(\tau_0)$ differs from the innovations sequence only in the initial conditions, which can be shown to be of no importance for the asymptotic distribution, thus we may replace $\nu_t(\tau_0)$ with ν_t without changing the limit. Now the central limit theorem for $\sqrt{T}(\sum_{j=1}^{T} \nu_t \nu_t^T - EE^T)$ holds under the standard assumption (see e.g. Hannan and Deistler, 1988, or Chapter 4), which proves asymptotic normality for the first term. Next note, that in $\sum_{j=1}^{T} \nu_t (\partial \nu_t(\tau_0))^T$, where $\partial \nu_t(\tau_0) = (\partial k^{-1}(z))(y_t - l(z)u_t)$ (with slight abuse of notation), terms of the form $\sum_{j=1}^{T} \nu_t y_{t-j}^T$ and $\sum_{j=1}^{T} \nu_t u_{t-j}^T$, where $y_j = u_j = 0, j < 1$ occur. The exponential decrease of the power series coefficients of k^{-1} ensures the exponential decrease of the power series coefficients of ∂k^{-1}, which in term implies (using the forthcoming Lemma 4.1.4, see Chapter 4) that the essential terms for the asymptotic normality are sums of the form $\sum \varepsilon(t)\varepsilon(t-j), \sum \varepsilon(t)u(t+j)$. The standard assumptions on the white noise $\varepsilon(t)$ and the pseudo stationarity of $u(t)$ ensure the (joint) asymptotic normality of these terms.

It remains to show the nonsingularity of the matrix of the second derivatives as well as the verification, that I_0^{-1} is in fact the asymptotic covariance matrix. Here we will use the decomposition of the parameter vector into a part (θ) parametrizing the transfer functions ($\bar{k}_0, l_0$) and the part parametrizing the innovation variance (σ). From the definition of the sequence $\nu_t(k,l)$ it is obvious, that $\nu_t(k,l)$ is independent of σ. Thus it is easy to see, that the asymptotic distribution of $\sqrt{T}(\hat{\sigma}_T - \sigma_0)$ depends on the asymptotic distribution of $\frac{1}{\sqrt{T}} \sum_{t=1}^{T} [\nu_t(\bar{k}_0, l_0)\nu_t(\bar{k}_0, l_0))^T - EE^T]$, whereas it can be shown for the estimates $\hat{\theta}_T$, that the contribution due to these 'squared terms' is zero (see Hannan and Deistler, 1988, Lemma 4.3.2). Thus in order to conclude, that the estimates of θ and σ are asymptotically uncorrelated, we have to show, that the limit of the matrix of the second derivatives is blockdiagonal, since the terms $\sum[\varepsilon(t)\varepsilon(t)^T - I]$ and $\sum \varepsilon(t)\varepsilon(t-j)^T, j > 0, \sum \varepsilon(t)u(t+j)^T$ are uncorrelated asymptotically. Recall, that the log-likelihood can be approximated by $\log\det EE^T + tr[E^{-T}\Omega(\tau)E^{-1}]$, where the difference is of order $o_P(T^{-1/2})$. Here $\Omega(\tau) = (1/T)\sum \nu_t(\tau)\nu_t(\tau)^T$. The same can be shown for the first and the second derivative (see Hannan and Deistler, 1988). Thus we have to investigate the second order derivatives of this function. Let σ denote the vector of the nonzero entries of E. Then we obtain for the partial derivatives with respect to the entries of σ corresponding to the $(i,j), i \geq j$ entry of E:

$$\partial\{\log\det EE^T + tr[(EE^T)^{-1}\Omega(\tau)]\} = 2\delta_{i,j}/E_{ii} - 2[E^{-T}\Omega(\tau)E^{-1}E^{-T}]_{ij}$$

where $\delta_{i,j} = 1, i = j$ and 0 else. From this expression and the fact, that the derivative of $\Omega(\tau)$ with respect to entries of θ converges to zero, it follows, that indeed the estimates for θ_0 and σ_0 are uncorrelated asymptotically. Thus it remains to show the nonsingularity for the second order derivatives.

Corresponding to the entries in σ we already derived the first order derivative above. The second order derivative with respect to the entries in σ, which correspond to the entries $(i,j), i \geq j$ and $(r,s), r \geq s$ of E can be shown to converge to

$$-\frac{2\delta_{i,j,r,s}}{E_{ii}^2} + 4[E^{-1}]_{si}[E^{-1}]_{jr} + 2\delta_{ri}[(EE^T)^{-1}]_{js}$$

where $\delta_{i,j,r,s} = 1$ iff $i = j = r = s$ and zero else. Additional subscripts denote entries of the various matrices. Note, that E^{-1} is lower triangular and thus $[E^{-1}]_{si}$ is only nonzero, if $s \geq i$. Therefore we obtain, that the contribution of the second term is only nonzero, if $i \geq j \geq r \geq s \geq i$ holds. This holds exactly, if $i = j = r = s$. The contribution of the third term is equal to zero unless $i = r$ holds. Thus the matrix of the second order derivatives with respect to entries of σ is block diagonal, where the blocks refer to the rows of E. These blocks are easily seen to be nonsingular, since they are the sum of two terms, where the first term is equal to a diagonal block of $(EE^T)^{-1}$, which is positive definite due to the assumption on E, and the matrix containing only one nonzero entry in the position corresponding to $i = j = r = s$, which is positive. This shows the nonsingularity for the part involving the parameters σ.

For the parameters contained in θ note, that for the corresponding asymptotic distribution the squared terms can be eliminated, as has been stated already. From this it follows, that the

distribution of the noise does not influence the asymptotic distribution of the estimates and thus we may assume without restriction of generality that the noise is Gaussian. Then using the Gaussanity assumption, it follows, that we only have to deal with the evaluation of I_0, since the second order derivative converges to I_0 (see Hannan and Deistler, 1988). To this end, we will again consider the approximation of the likelihood. The partial derivative with respect to an entry in θ can be shown to be equal to

$$2tr[\Omega(\tau)^{-1}[\partial\nu_t(\tau)]\nu(\tau)^T] = 2\varepsilon_t(\tau)^T(\partial\varepsilon_t(\tau))$$

where $\varepsilon(\tau) = E^{-1}\nu_t(\tau)$ and $\partial\varepsilon_t(\tau)$ denotes the partial derivative with respect to the entry in θ. Thus we need to evaluate the expectation of the product of two such terms. Due to the minimum-phase assumption, $\varepsilon_t(\tau) = E^{-1}(y_t - Du_t - C\mathcal{K}Z_t^-)$ and thus $\partial\varepsilon_t(\tau) = E^{-1}\partial Du_t + E^{-1}\partial[C\mathcal{K}]Z_t^-$. Here $\mathcal{K} = [[B - \bar{K}D, \bar{K}], (A - \bar{K}C)[B - \bar{K}D, \bar{K}], \cdots]$, where $\bar{K} = KE^{-1}$. Note, that $\varepsilon_t(\tau_0)$ is uncorrelated with the past of $y(t)$ and thus $\mathbb{E}\partial_i\varepsilon(t;\tau_0)^T\varepsilon(t;\tau_0)\varepsilon(t;\tau_0)^T\partial_j\varepsilon(t;\tau_0) = \partial_i\varepsilon(t;\tau_0)^T\partial_j\varepsilon(t;\tau_0)$ can be shown by conditioning on $\mathcal{F}_{t-1}$. Here the index of ∂_i indicates the variable, with respect to which the partial derivative is calculated. Additionally the realizations y_t, u_t have been replaced by the random variables $y(t), u(t)$ in the expressions. Thus the remaining part of the proof will be devoted to showing, that the matrix $[\mathbb{E}(\partial_i\varepsilon(t;\tau_0))^T(\partial_j\varepsilon(t;\tau_0))]_{i,j=1,\cdots,d(\theta)}$ is nonsingular. This will be done by constructing a contradiction to the fact, that θ_0 is an inner point of the parametrization of the corresponding manifold of dimension $d(\theta)$.

Assume, that the matrix were singular. Then there existed scalars $a_i, 1 \leq i \leq d(\theta)$ such that

$$\sum_{i,j=1}^{d(\theta)} a_i a_j \mathbb{E}(\partial_i\varepsilon(t;\tau_0))^T(\partial_j\varepsilon(t;\tau_0)) = \mathbb{E}(\sum_i a_i(\partial_i\varepsilon(t;\tau_0)))(\sum_i a_i\partial_i\varepsilon_t(\tau_0))^T = 0$$

implying that $\sum_j \partial_j\varepsilon(t;\tau_0) = 0$ and thus due to the assumptions on $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ we obtain $\sum_i a_i[\partial_i D, \partial_i[C\mathcal{K}]] = 0$. This implies, that $\sum_i a_i\partial_i D + \sum_{j=1}^{\infty}\sum_i a_i\partial(C(A - \bar{K}C)^{j-1}[B - \bar{K}D, \bar{K}])z^{-j}$ is a nonminimal realization of zero. Thus in particular $\sum_i a_i\partial_i D = 0$. From this $\sum_i a_i \mathrm{vec}[\partial_i A, \partial_i B, \partial_i C, \partial_i D, \partial_i\bar{K}]$ follows using a realization of this transfer function, which can be derived using the derivative of the corresponding transfer function $\sum_i a_i\partial D_i + \sum_i a_i\partial_i[C(zI - A + \bar{K}C)^{-1}[B - \bar{K}D, \bar{K}]]$. Thus we obtain that the tangent space of the parameter space in S_n is of dimension smaller than $d(\theta)$, which is a contradiction to the assumption, that the particular piece of M_n^j, which is under consideration is a manifold of dimension $d(\theta)$. Here we use the analyticity property of the mapping attaching system matrices to parameter vectors.

Since the asymptotic distribution of $\hat{\theta}$ does not depend on the squared terms $1/\sqrt{T}\sum(\varepsilon_t\varepsilon_t^T - I)$ it is straightforward to show, that the asymptotic distribution does not depend on the distribution of the noise. Thus the asymptotic distribution is the same, as if the noise would have been Gaussian. This implies, that the asymptotic variance of $\sqrt{T}(\hat{\theta} - \theta_0)$ is equal to I_0^{-1} and also, that the estimation is asymptotically efficient. □

The theorem does not only state asymptotic normality of the estimator, but also implies, that the asymptotic variance for the estimation of $(\bar{k}_0, l_0)$ reaches the Cramer-Rao bound, i.e. with the ML approach one obtains generically almost sure converging, asymptotically optimal estimates of the free parameters contained in θ. Therefore the ML approach is a benchmark for all other procedures for estimating linear systems of the form (1.1), which motivates the inclusion of this part into the thesis.

In all the evaluations above we did assume, that the process in fact has been generated by a system of the form (1.1) and that we are given a suitable upper bound (cf. Theorem 2.4.1) or even the correct specification of the multiindex γ (cf. Theorem 2.4.2). However in practice, we have to estimate the model order. We will confine the discussion to a particular class of estimation criteria, since they have attained the most attention in the literature and are also quite standard for the estimation of linear systems. In the literature, they are termed *information criteria* and have been pioneered by (Akaike, 1976). Famous members of this class of estimation principles

are the Akaike Information Criterion (`AIC`) and `BIC`, sometimes also called Rissanens information criterion or Schwarz criterion.

These criteria are based on a comparison of the estimated variance of the innovations $\nu(t)$: Assume that we chose a model order n and obtained a ML estimate $\hat{\tau}_{T,n}$) by optimizing the likelihood over $\overline{M_n^j}$ using the observations $y_1, u_1, \cdots, y_T, u_T$. Let the parameter vector $\hat{\tau}_{T,n}$ correspond to system matrices $(\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n, \hat{E}_n, \hat{K}_n)$. Then assuming that the initial state $x(0)$ is zero, we obtain estimates $\hat{\nu}_t(\hat{\tau}_{T,n}) = y_t - \hat{C}_n\hat{x}_t - \hat{D}_n u_t, \hat{x}_{t+1} = \hat{A}_n\hat{x}_t + \hat{B}_n u_t + \hat{K}_n(\hat{E}_n)^{-1}\hat{\nu}_t(\hat{\theta}_{T,n})$. From the estimates $\hat{\nu}_t(\hat{\tau}_{T,n})$ we obtain an estimate $\hat{E}_n\hat{E}_n^T$ of the innovation variance $\Omega = EE^T$. Denote this estimate with $\Omega(\hat{\tau}_{T,n})$. Then the information criteria use as an estimate of the order the minimizing argument of:

$$\chi(n) = \log\det\Omega(\hat{\tau}_{T,n}) + d(n)C_T/T, 0 \leq n \leq H_T \tag{2.19}$$

where $d(n) = n(2s+m) + sm + s(s+1)/2$ is the dimension of the parameter vector. Note, that this is not the usual definition of `AIC`, since we also accounted for the parameters contained in D and E. However this does not make any difference, since one is only interested in the location of the minimum, which is not affected by the higher number of parameters. Here C_T has to be specified such that $C_T > 0$ and $C_T/T \to 0$. Typical choices of C_T are $C_T = 2$ (`AIC`) and $C_T = \log T$ (`BIC`). H_T is a suitable upper bound. A suggestion for a choice (from a theoretical point of view) will be given in the following and motivated later on. Then $\hat{n} = \arg\min_{0\leq n\leq H_T}\chi(n)$ defines the estimate of the order. Using this notation, a heuristic interpretation of the information criteria is rather obvious: The choice of the model order is performed by comparing the goodness of fit of the various models, measured by the log of the determinant of the estimated innovations (the one step ahead prediction errors, see section 2.5), to the number of included parameters weighted by C_T/T. According to the restriction $C_T > 0$, additional parameters have to improve the fit to be included, whereas the condition $C_T/T \to 0$ implies, that the penalty on the number of parameters in the criterion vanishes with growing sample size.

In order to obtain results on the asymptotic properties of these estimates, it will be convenient to impose some assumptions on the input sequence. It is noted however, that these assumptions are far from being minimal, and are only chosen, since they have also been considered for the subspace case:

> DEFINITION (ASSUMPTIONS ON INPUT) The input $(u(t))_{t\in\mathbb{Z}}$ is a stationary process, such that the joint process $z(t) = [y(t)^T, u(t)^T]^T$ fulfills the standard assumptions.

This implies, that $u(t)$ is generated by a rational, stable and strictly minimum-phase transfer function $k_u(z) = \sum_{j=0}^{\infty} K_u(j)z^{-j}$ and a noise sequence $\varepsilon_u(t)$, which fulfills the standard assumptions. For weaker conditions on $u(t)$ see (Hannan and Deistler, 1988). Then the asymptotic properties of the estimates of the order can be stated as follows:

Theorem 2.4.3 *If $(y(t))_{t\in\mathbb{Z}}$ is generated by a system of the form (1.1) and thus of finite order n, the standard assumptions hold and also the assumption on the inputs stated above, then for $H_T = o((\log T)^a)$ for some $a < \infty$,*

- *if $\liminf_{T\to\infty} C_T/(2\log\log T) > 1$, $\hat{n} \to n$ a.s.*
- *if $\limsup_{T\to\infty} C_T < \infty$ then convergence in probability fails.*

This implies, that `BIC` is strongly consistent in our setting, whereas `AIC` does not converge in probability. Indeed there is a positive probability of overestimating the order (see e.g. Shibata, 1980). The reader should however keep in mind, that the consistency is not as its own a reason to favour `BIC`. It can be shown, that despite being not consistent, `AIC` selects the optimal order of an autoregression corresponding to the one step ahead prediction error (Shibata, 1980) and the squared relative error in the estimate of the spectrum (Shibata, 1981). Note, however, that the

evaluation of the information criterion requires the optimization of the likelihood for all possible orders $0 \leq n \leq H_T$, which is a computationally demanding task. Therefore procedures have been developed to overcome this problem by performing a multi-step procedure (Hannan and Kavalieris, 1984).

Also another result will prove very useful in the rest of the thesis. This result is concerned with the behaviour of `BIC` and `AIC` for the following situation: Assume, that $(y(t))_{t\in\mathbb{Z}}$ is generated by a system of the form (1.1) and suppose, that the matrix $A - KE^{-1}C$ is not nilpotent i.e. $(A - KE^{-1}C)^n \neq 0, \forall n \in \mathbb{N}$. This is equivalent to the statement, that $(y(t))_{t\in\mathbb{Z}}$ is not an ARX process i.e. $y(t)$ cannot be written as $y(t) = -A_1 y(t-1) - \cdots - A_p y(t-p) + B_0 u(t) + \cdots + B_p u(t-p) + \nu(t)$, where the $\nu(t)$ are the innovations. Assume further, that we want to model the observations y_t by an ARX process. Of course under our assumptions, this will be a misspecification. Even though, one reason to do so, lies in the fact, that the matrices A_i, B_i can be obtained by simple least squares once the order p is known. Furthermore the least-squares fit can be calculated recursively in the order p using the Durbin-Levinson recursion. This is the reason, why models of this form are often used to provide initial estimates for the optimization of the likelihood function in many procedures, for example the Hannan & Rissanen procedure (see Hannan and Kavalieris, 1984) or (Wahlberg, 1989).

Note, that the assumptions on the inputs imply, that the sample covariance estimates of $u(t)$ fulfill the uniform convergence as stated in Theorem A.3.1:

$$sup_{0\leq t\leq H_T} ||\hat{\gamma}_{u,u}(t) - \gamma_{u,u}(t)|| = O((\frac{\log\log T}{T})^{1/2}) \tag{2.20}$$

for $H_T = O((\log T^a), a < \infty$. It is possible to prove this bound also for weaker assumptions on the inputs (see Hannan and Deistler, 1988). Bounds of the form (2.20) are the basis for the proof of Theorem 2.4.4. Note, that for the case of no observed inputs, we do not add any additional conditions. In this framework, the following theorem holds:

Theorem 2.4.4 *Let the observations $y_t, 1 \leq t \leq T$, be generated by a process of the form (1.1) with $n > 0$ and $A - KE^{-1}C$ not nilpotent, let the standard assumptions hold and assume, that the input fulfills the assumptions mentioned above. If the order p is estimated using* `BIC`*, then $\lim_{T\to\infty} -2\hat{p}\log|\rho_0|/\log T = 1$ a.s. Here $\hat{p}$ denotes the minimizing argument of the* `BIC` *criterion for $p \leq H_T$, $H_T = o((\log T)^a)$ and $\rho_0 = \lambda_{max}(A - KE^{-1}C) \neq 0$.*

This also gives a hint on the behaviour of the `AIC` estimates, since $\hat{p}_{BIC} \leq \hat{p}_{AIC}$. This follows from the fact, that `BIC` penalizes the inclusion of additional parameters more severe. Thus we obtain

$$\lim_{T\to\infty} \hat{p}_{AIC}(-2\log\rho_0)/\log T \geq 1, \ a.s.$$

Also note, that the assumption $n > 0$ and $A - KE^{-1}C$ not nilpotent are not of particular importance, since in the set $U_n^{(m)}$ the subset fulfilling the above two conditions is open and dense.

2.5 Approximations to the likelihood function

Since the optimization of the likelihood is a computationally very demanding task (note, that the matrix $\Gamma_T^+(k)$ contains $(Ts)^2$ entries) for big sample sizes, one usually uses approximations of the likelihood function, in order to implement the ML approach. The most common approximation to the likelihood function are known as the Prediction Error Methods (PEM): According to the evaluations in Appendix A under our assumptions on the noise sequence, the best linear mean square prediction of $y(t)$ given the whole past $y(s), s = t-1, t-2, \cdots$ and the whole input process $(u(t))_{t\in\mathbb{Z}}$ is denoted by $y(t|t-1)$ and can be calculated as $y(t|t-1) = Cx(t) + Du(t)$, such that the prediction error is equal to the innovation $\nu(t)$. The minimumphase assumption makes it possible to write the innovations $\nu(t)$ as a function of the outputs $y(t)$ and inputs $u(t)$ as follows: $\nu(t) = y(t) - Cx(t) - Du(t), x(t+1) = (A - KE^{-1}C)x(t) + (B - KE^{-1}D)u(t) + KE^{-1}y(t)$.

Analogously the sequence ν_t is calculated by replacing stochastic variables with observations in the above mentioned recursions. As has been stated in the proof of the central limit theorem for maximum likelihood estimator, the log likelihood function $\mathcal{L}(Y_{1,T}^+, U_{1,T}^+, \tau)$ can be approximated by $\log\det EE^T + tr[(EE^T)^{-1}(\sum_{t=1}^T \nu_t(\tau)\nu_t(\tau)^T)]$ up to terms, which are of order $o_P(T^{-1/2})$. Minimizing this function for fixed τ with respect to EE^Twe obtained the optimum for $EE^T = \sum_{t=1}^T \nu_t(\tau)\nu_t(\tau)$ with a minimal value of $\log\det\Omega(\tau) + s$, where $\Omega(\tau) = \sum_{t=1}^T \nu_t(\tau)\nu_t(\tau)^T$. Introducing the matrix $\bar{K} = KE^{-1}$ we see, that $\nu_t(\tau)$ only depends on $(A, B, C, D, \bar{K})$ but not on E. Thus in the situation, where E and $(A, B, C, D, \bar{K})$ are parametrized independently, as is the case for the parametrizations presented in this thesis, the log likelihood function can be optimized for E and $(A, B, C, D, \bar{K})$ seperately, which leads to the problem of minimizing $\log\det\Omega(\tau)$. Following the proof of the central limit theorem it is clear, that the estimates obtained with this criterion function will be asymptotically equivalent to the ML estimates in the sense, that the asymptotic distribution of $\sqrt{T}(\hat{\tau}_T - \tau_0)$ is identical.

For many applications however it may seem reasonable to minimize a different scalar function of $\Omega(\tau)$ rather than $\log\det(\Omega(\tau))$. Since in general there is no global optimum to this matrix function, the choice of a suitable function $\ell : \mathbb{R}^{s\times s} \to \mathbb{R}$ attaching the $s \times s$ matrices $\hat{\Omega}(\tau)$ to a scalar criterion value, will influence the properties of the estimate. The function ℓ is chosen such that the semi-ordering of symmetric matrices is reflected by the criterion i.e. $X \leq Z \Rightarrow \ell(X) \leq \ell(Z)$. Note, that in the SISO case, all criterion functions, that differ in the choice of ℓ are in fact equivalent, as is obvious from the fact, that in this case the semi-ordering of matrices reduces to the ordering of scalars. In the *Prediction Error Methods* we try to minimize this function rather than an approximation to the likelihood function, i.e. the PEM estimate over the set U_γ, say $\tilde{\tau}$, is defined as follows:

$$\tilde{\tau} = \arg\min_{\tau\in\overline{T_\gamma}} \ell(\hat{\Omega}(\tau))$$

If the function ℓ is chosen to be $\det(.)$ or $tr[(EE^T)^{-1}.]$, the resulting estimates are asymptotically equivalent to the ML estimates. However, clearly also the same parametrization problems occur, as for the ML approach. For a discussion on the properties of the estimators for general ℓ (see e.g. Söderström and Stoica, 1989).

(Ljung, 1985; Ljung and Yuan, 1985; Ljung, 1987; Wahlberg and Ljung, 1986) also investigated a generalization of the PEM approach for the SISO case, using an even more general criterion : Note, that $\nu_t(\tau)$ in our framework denotes the one step ahead prediction errors i.e. $y(t+1) - y(t+1|t)$ (see Appendix A). However in some applications, it might also be reasonable to seek a system, which provides optimal r step ahead predictions. Then a natural idea would be to incorporate this optimality criterion directly in the estimation by replacing $\nu_t(\tau)$ in the criterion by the r step ahead prediction errors as a function of the model used i.e. let $\nu_t(r,\tau) = y_{t+r} - y_{t+r|t}(\tau)$ denote the r step ahead prediction error according to the parameter τ using the observations, then optimize $\ell(\Omega(r,\tau)) = \ell(\frac{1}{T}\sum_{t=1}^T \nu_t(r,\tau)\nu_t(r,\tau)^T))$, where again ℓ is a mapping attaching scalars to matrices in $\mathbb{R}^{s\times s}$ maintaining the semi-ordering of symmetric matrices. The ideas presented above suggest a more general formulation, including also a weighted version of prediction errors to different prediction horizons i.e. to minimize $\ell(\sum_{r=1}^R W_r\Omega(r,\tau)W_r^T)$ for some finite $R \in \mathbb{N}$ and weightings $W_r \in \mathbb{R}^{s\times s}$. This general framework has been investigated in (Wahlberg and Ljung, 1986). Their main conclusion can be stated as follows: If the true process lies in the considered model class, then the one step ahead prediction error approach is always asymptotically optimal, since it is asymptotically equivalent to the ML approach and it therefore does not make sense to use a generalization of the criterion function. The additional degrees of freedom introduced by the choice of the weightings W_r and of $R > 1$ can only deteriorate the performance. However, if the true process is not in the considered model class, the bias can be shaped over the frequency range by these design variables. Simulation results show a similar behaviour for the subspace algorithms (compare section 5.3).

Chapter 3

Subspace Algorithms

3.1 General Structure

The term 'subspace algorithms' is used for a big variety of algorithms, which are all based on the same fact. In this chapter, we will try to give a brief survey on the variety of the different available methods and then at the end concentrate on one type of algorithm, which will be further investigated in this thesis.
All subspace algorithms hinge on the fact, that the state contains, in a certain sense, all the information from the past, that is relevant for the future (see also Appendix A). Assume for the moment, that $u(t)$ is a stationary stochastic process, jointly with $y(t)$. This assumption is only introduced for a unified framework during the presentation of the algorithms. Later on weaker conditions for $u(t)$ will be used. Let $Y^+(t) = [y(t)^T, y(t+1)^T, \cdots]^T$, $Y^-(t) = [y(t-1)^T, y(t-2)^T, \cdots]^T$ and let $U^-(t), U^+(t)$ and $E^+(t)$ be the analogously defined vectors of the past of the input, the future of the input and the noise respectively. Additionally let $Z^-(t) = [y(t-1)^T, u(t-1)^T, y(t-2)^T, u(t-2)^T]^T$ denote the vector of the stacked past of observed inputs and outputs. Then it follows from the system equations (1.1) and the strict minimumphase-condition that $Y^+(t) = \mathcal{O}x(t) + \mathcal{U}U^+(t) + \mathcal{E}E^+(t)$ and $x(t) = \mathcal{K}Z^-(t)$. Here $\mathcal{O} = [C^T, A^TC^T, (A^2)^TC^T, \cdots]^T$ denotes the observability matrix, $\mathcal{K} = [[B - \bar{K}D, K], (A - \bar{K}C)[B - \bar{K}D, \bar{K}], (A - \bar{K}C)^2[B - \bar{K}D, \bar{K}], \cdots]$, where $\bar{K} = KE^{-1}$.

$$\mathcal{E} = \begin{bmatrix} E & 0 & \cdots & 0 \\ CK & E & \ddots & \vdots \\ CAK & \ddots & \ddots & 0 \\ \vdots & & CK & E \end{bmatrix}, \mathcal{U} = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ CAB & \ddots & \ddots & 0 \\ \vdots & \cdots & CB & D \end{bmatrix}$$

Combining both equations leads to:

$$Y^+(t) = \mathcal{O}\mathcal{K}Z^-(t) + \mathcal{U}U^+(t) + \mathcal{E}E^+(t) \tag{3.1}$$

Note, that only for notational convenience we use a realization (A, B, C, D, E, K) of the true system (k, l). However the equations are easily seen to be independent of the particular realization in the sense, that the additive decomposition is the same for each choice of the basis in the state space. Also note, that we included a direct feedthrough term D, which also enters the definition of $\mathcal{K}$. The presentation of the algorithm holds also for the case, where D is restricted to be zero. Since the past of the output series $(y(t))$ and the future of the noise are uncorrelated under our standard assumptions, the best linear least squares predictor of $Y^+(t)$ given the past $(Y^-(t))$ and the whole series $(u(t))_{t\in\mathbb{Z}}$, equals $\mathcal{O}\mathcal{K}Z^-(t) + \mathcal{U}U^+(t)$. This can be expressed as a projection of the future of $y(t)$ onto the space spanned by the past of $y(t)$ and the whole series $u(t)$ i.e. $Y^+(t)|_{(Z^-(t),U^+(t))} = \mathcal{O}\mathcal{K}Z^-(t) + \mathcal{U}U^+(t) = \mathcal{O}x(t) + \mathcal{U}U^+(t)$, where $X|_Y$ denotes the orthogonal

projection in the Hilbert space H_z (for an exact definition of 'projection' see Appendix A). Thus one way to obtain the state space is to extract the term due to the state in this projection:

$$(Y^+(t)|_{(Z^-(t),U^+(t))})|_{(U^+(t))^\perp} = \mathcal{O}[\mathcal{K}Z^-(t)|_{(U^+(t))^\perp}] \tag{3.2}$$

where $(U^+(t))^\perp$ denotes the orthogonal complement in H_z of the space $H_u^+(t)$ (see Appendix A). Now $\mathcal{K}Z^-(t)|_{(U(t)+)^\perp} = x(t)|_{(U^+(t))^\perp}$. The components of the vector on the right hand side of equation (3.2) span the state space $X(t) = \{x_i(t), 1 \leq i \leq n\}$ projected onto the orthogonal complement of the future of the inputs $u(t)$. Therefore we have to impose additional conditions to ensure, that the orthogonal projection does not change the dimension of the space. The formulation of such conditions depends on the assumptions on the inputs, i.e. whether the inputs are assumed to be the realization of a stochastic process or a pseudo-stationary series (see Appendix A), and will be discussed in more detail, whenever it is of importance.

The relevance of equation (3.2) is the following: Commencing from the knowledge of $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ rather than from the system matrices, the projections on the left hand side of the equation can be calculated. Then every choice of the basis in the state space leads to a decomposition of $(Y^+(t)|_{(Z^-(t),U^+(t))})|_{(U^+(t))^\perp}$ into $\mathcal{O}x(t)|_{(U^+(t))^\perp}$, where $x(t)|_{(U^+(t))^\perp} = \mathcal{K}Z^-(t)|_{(U^+(t))^\perp}$. Thus in this way under suitable assumptions we obtain the state as $x(t) = \mathcal{K}Z^-(t)$. From the knowledge of the state, the system matrices can be recovered using projections in the system equations (1.1). Another possibility consists in exploiting the structure in $\mathcal{O} = [C^T, A^TC^T, \cdots]$. Knowing A and C, the matrices $\mathcal{U}$ and $\mathcal{E}$ can be used to obtain B, D, K and E.

Note, that equation (3.2) also holds in H_∞^∞ (for a definition see Appendix A), if the symbols $y(t), u(t)$ and $\varepsilon(t)$ are replaced by y_t, u_t and ε_t, realizations of the stationary processes, and the projections are interpreted in the appropriate sense as the ergodic limits of sample covariances (see Appendix A). Also in the case, where we are given one realization $[y_t^T, u_t^T]^T, t \in \mathbb{Z}$ of the ergodic process, it is possible to obtain the true transfer functions by calculating the projections discussed above (comp. Peternell *et al.*, 1996). Now in practice of course we have to deal with a finite amount of data in a time interval $1 \leq t \leq T$ of the realization of the processes $(y_t)_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$. Thus we cannot use the (infinite dimensional) equation (3.2), but only a finite section of this equation. Let f denote the integer parameter indicating the number of block rows included in the finite dimensional equation. Analogously we cannot use the infinite past to build the state and thus we also have to decide on another integer parameter, p say, indicating the number of lags to be included in the state estimation. For obvious reasons f will be called *row truncation index* and p will be called *column truncation index*. Let $Y_{t,f}^+ = [y_t^T, y_{t+1}^T, \cdots, y_{t+f-1}^T]^T \in \mathbb{R}^{fs}, U_{t,f}^+ = [u_t^T, u_{t+1}^T, \cdots, u_{t+f-1}^T]^T \in \mathbb{R}^{fm}, E_{t,f}^+ = [\varepsilon_t^T, \varepsilon_{t+1}^T, \cdots, \varepsilon_{t+f-1}^T] \in \mathbb{R}^{fs}$ and $Z_{t,p}^- = [y_{t-1}^T, u_{t-1}^T, \cdots, y_{t-p}^T, u_{t-p}^T]^T$. Further let $\mathcal{O}_f$ denote the matrix built of the first f block rows of $\mathcal{O}$, $\mathcal{K}_p$ the matrix built of the first p block columns of $\mathcal{K}$. Then the following equation holds:

$$Y_{t,f}^+ \quad = \quad \mathcal{O}_f\mathcal{K}_pZ_{t,p}^- + \mathcal{O}_f(A - KE^{-1}C)^p[\mathcal{K}Z_{t-p}^-] + \mathcal{U}_fU_{t,f}^+ + \mathcal{E}_fE_{t,f}^+ \tag{3.3}$$

Here $\mathcal{E}_f$ denotes the heading $fs \times fs$ submatrix of $\mathcal{E}$ (i.e. the northwest corner), and finally $\mathcal{U}_f$ the heading $fs \times fm$ submatrix of $\mathcal{U}$. Note that this equality holds for any $t \in \mathbb{Z}$.

REMARK 3.1.1 There are two main frameworks to present the details of the algorithms: (Moonen et al., *1989) introduced a notation, which has become quite popular in the literature, using what they call 'data Hankel matrices': With a slight reordering of some matrices, equation (3.3) can be written as* $Y_{t,f}^+ = \mathcal{O}_f\tilde{\mathcal{K}}_pZ_{t-p,p}^+ + \mathcal{O}_f(A - KE^{-1}C)^px_{t-p} + \mathcal{U}_fU_{t,f}^+ + \mathcal{E}_fE_{t,f}^+$. *Note, that this is a vector equality. Writing this equation simultaneously for* $t = p+1, \cdots, T-f$ *we obtain a matrix equality, where the columns are the vector equalities given above:*

$$Y_{p+1}^f = \mathcal{O}_f\tilde{\mathcal{K}}_pZ_1^p + \mathcal{U}_fU_{p+1}^f + \mathcal{E}_fE_{p+1}^f$$

Here $Y_{p+1}^f = [Y_{p+1,f}^+, Y_{p+2,f}^+, \cdots, Y_{T-f,f}^+]$ *and* Z_1^p, U_{p+1}^f *and* E_{p+1}^f *are defined analogously, these matrices are called 'data Hankel matrices' for obvious reasons. We will use the term 'future' for*

all matrices indexed with f and all matrices indexed with p will be given the attribute 'past'. The discussion of the algorithms as presented e.g. by (Van Overschee and DeMoor, 1994) uses these matrices and explains the procedures using projections in the rowspaces of these matrices. This is also related to the finite sample Hilbert spaces as discussed in Appendix A.

The other possibility to present the algorithm uses a regression framework. In the following we will adopt the regression point of view, since in our opinion it fits better into the framework of projections. However, we emphasize, that the arguments are completely the same, only the interpretation changes (also refer to Appendix A).

Note that the two truncation indices f and p have to be chosen by the user. The consequences of these choices will be discussed in Chapter 5.1. In this section we will also give an indication on how to choose these indices for a particular class of algorithms.

Equation (3.3) is the starting point for all subspace methods. The basic structure of subspace algorithms can be described as follows (Note, that this scheme is intended to be general in order to cover all popular subspace algorithms. This leads to rather imprecise statements, which will be clarified in the following subsections):

1. Calculate an estimate $\hat{\beta}_z$ of $\mathcal{O}_f\mathcal{K}_p$.
2. Decide on the model order n and approximate $\hat{\beta}_z$ by the product $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ where $\hat{\mathcal{O}}_f \in \mathbb{R}^{fs\times n}$ and $\hat{\mathcal{K}}_p \in \mathbb{R}^{n\times p(m+s)}$.
3. Calculate estimates of the system matrices using the results of the previous steps.

The different possibilities of performing these three steps of the general scheme will be specified in the following. Although this thesis will deal mainly with a special class of algorithms, this section is intended to give an overview of the various proposed approaches, to contribute to a clarification of some confusion arising from the fact, that quite different procedures are called 'subspace methods'. Since the number of proposed algorithms is permanently increasing, the discussion cannot not deal with all algorithms, but mainly with the most popular methods.

3.2 Step 1: Estimation of β_z

The most popular way to perform the calculations in this step is by first using $U_{t,f}^+$ as a regressor in the equation (3.3) to eliminate the contribution of $U_{t,f}^+$. Let the residuals from these regressions be denoted by an additional superscript Π. Thus for example $Y_{t,f}^{+,\Pi}$ denotes the residuals from the regression of $Y_{t,f}^+$ onto $U_{t,f}^+$ (Of course $Y_{t,f}^{+,\Pi} = Y_{t,f}^+$ and $Z_{t,p}^{-,\Pi} = Y_{t,p}^-$ in the case $m = 0$). Then in a second step, $Y_{t,f}^{+,\Pi}$ is regressed onto $Z_{t,p}^{-,\Pi}$ in order to obtain the estimate $\hat{\beta}_z$. Most of the popular algorithms use this method, like e.g. the `MOESP` algorithm of (Verhaegen, 1994), `N4SID` (Van Overschee and DeMoor, 1994) and the canonical variables method suggested by Larimore (This algorithm will be denoted with `CCA` in the following, (Larimore, 1983)).

However, this approach neglects the structure in the matrix $\mathcal{U}_f$. (Peternell, 1995; Peternell *et al.*, 1996) proposed two different procedures to use the information about the block Toeplitz-structure of the matrix $\mathcal{U}_f$: Recall the definition of $\mathcal{U}_f$:

$$\mathcal{U}_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ CA^{f-1}B & \cdots & CB & D \end{bmatrix} \tag{3.4}$$

Now this structure can be used in the estimation of β_z in several ways: The first approach proposed in (Peternell *et al.*, 1996) restricts the estimate of $\mathcal{U}_f$ to have the lower triangular block

Toeplitz structure i.e. the least squares solution to $Y_{t,f}^+ = \hat{\beta}_z Z_{t,p}^- + \hat{L}_f U_{t,f}^+ + \hat{N}_{t,f}^+$ is calculated, where $\hat{N}_{t,f}^+$ denotes the residuals of the regression, and where $\hat{L}_f$ is restricted to be of the following form:

$$\hat{L}_f = \begin{bmatrix} \hat{L}(0) & 0 & \cdots & 0 \\ \hat{L}(1) & \hat{L}(0) & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ \hat{L}(f-1) & \cdots & \hat{L}(1) & \hat{L}(0) \end{bmatrix} \tag{3.5}$$

(Peternell *et al.*, 1996) give explicit expressions for the least squares estimate. The second method proceeds in two steps: In a first step, we may use any consistent estimate of the system matrices (A, B, C, D), and in the second step, this estimate is used to estimate $\mathcal{U}_f$ by simply replacing the true values by their estimates in equation (3.4). These algorithms will be dealt with at the end of Chapter 4. For the following discussion assume, that β_z is estimated without using any structure of $\mathcal{U}_f$.

Let f and p be fixed for a moment. Then due to the uniform convergence of the sample covariance estimates (see Theorem A.3.1), the estimated covariances will converge a.s. to the true covariances i.e. to $\beta_z = \Gamma_{Y_f^+, Z_p^-, \Pi}(\Gamma_{Z_p^-, \Pi, Z_p^-, \Pi})^{-1}$, where $\Gamma_{a,b}$ denotes the covariance between a and b e.g. $\Gamma_{Y_f^+, Z_p^-, \Pi} = \mathbb{E} Y_f^+(t)(Z_p^{-,\Pi}(t))^T$. From equation (3.3) it can be seen, that under the condition $f, p \geq n$ the rank of β_z is smaller or equal to n, the true system order. Now it is clear that, for the identification to produce useful results it will be required that this limit has full rank n. β_z is of full rank, if and only if $\mathcal{R}_{f,p}$ is of full rank, where $\mathcal{R}_{f,p}$ is defined as:

$$\mathcal{R}_{f,p} = \mathbb{E} \begin{bmatrix} x(t) \\ U_f^+(t) \end{bmatrix} [Y_p^-(t)^T, U_p^-(t)^T, U_f^+(t)^T] \tag{3.6}$$

and $\mathcal{R}$ is obtained by using the limiting values of f and p (It will be part of the next section to derive possible scenarios for the indices f and p.). Here $\mathbb{E}$ denotes expectation for random variables and limits of sequences of the form $\frac{1}{T}\sum_{t=1}^{T} u_t \varepsilon(t)$ for pseudostationary sequences u_t, which have to be assumed to exist. Here of course e.g. $Y_p^-(t)$ denotes the vector of the first ps coordinates of $Y^-(t)$. Somewhat surprisingly it is not easy to develop explicit conditions to ensure that $\mathcal{R}_{f,p}$ is of full rank, when f and p are fixed and finite. (Jansson and Wahlberg, 1997) show, that it is not sufficient to impose conditions of persistence of excitations of the type given in equation (A.3) on the process $(u(t))_{t\in\mathbb{Z}}$, when f and p are small.

However, a sufficient condition can be found, if we allow p to tend to infinity at a certain rate. A sufficient condition on the input process in this case ensuring consistency of a particular algorithm has been given in (Peternell *et al.*, 1996): For $p \to \infty$ at a suitable rate with $T \to \infty$ it can be shown, that the estimates of the covariance sequence converge uniformely to the true covariances (comp. Theorem A.3.1). The limiting expression for β_z now clearly also involves the matrix Γ_u^-, the covariance of the infinite past of the observed inputs (or the equivalent for pseudostationary sequences). This matrix is required to be of bounded norm and nonsingular with bounded inverse for β_z to be defined for $p = \infty$. A sufficient condition for this to hold is, that $(u(t))_{t\in\mathbb{Z}}$ is a stationary process, jointly with $(y(t))_{t\in\mathbb{Z}}$, having a spectrum, which is bounded uniformely from above and from below on the unit circle, where the input process $(u(t))_{t\in\mathbb{Z}}$ is generated by a white noise sequence $\varepsilon_u(t), t \in \mathbb{Z}$ fulfilling the standard assumptions of Chapter 1, which is filtered using a transfer function $k_u(z) = \sum_{j=0}^{\infty} K_u(j) z^{-j}$. Finally we also have to assume $\sum_{j=0}^{\infty} j^{1/2} \|K_u(j)\| < \infty$ in order for results like Theorem A.3.1 to hold (see Hannan and Deistler, 1988, section 6.6). In fact we will in the analysis use the stronger assumption, that the joint process $(y(t)^T, u(t)^T)^T, t \in \mathbb{Z}$ is generated by a system of the form (1.1) fulfilling the standard assumptions.

It has been argued (see e.g. Wahlberg and Jansson, 1994) that these conditions on the input process are rather restricting due to the following fact: Especially in technical applications like control of industrial processes, it might be useful to apply special input signals such as sums of sinusoid to the plant, which is to be identified. In these cases, the input will not be persistently

Method	$\hat{W}_f^+$	$\hat{W}_p^-$
CCA	$(\hat{\Pi}_f^+)^{-1/2}$	$(\hat{\Pi}_p^-)^{1/2}$
N4SID	I	$(\hat{\Gamma}_p^-)^{1/2}$
MOESP	I	$(\hat{\Pi}_p^-)^{1/2}$

Table 3.1: Choice of the weighting matrices. $\hat{\Pi}_f^+$ denotes the sample covariance of $Y_{t,f}^{+,\Pi}$, $\hat{\Pi}_p^-$ the sample covariance of $Z_{t,p}^{-,\Pi}$, $\hat{\Gamma}_p^-$ denotes the sample covariance of $Y_{t,p}^-$, and $Q^{1/2}$ denotes any square root of a positive semidefinite square matrix Γ such that $\Gamma = Q^{1/2}Q^{T/2}$.

exciting of infinite order. In particular, if we are not really interested in the noise characteristics, but rather in the transfer function l describing the relation between observed inputs and outputs, we may use variants of subspace algorithms, which produce consistent estimates for (A, B, C, D) also for finite truncation index p (see Chapter 6). (Jansson, 1997) gives explicit conditions on the input process, which ensure consistency in several special situations (i.e. for special choices of inputs in connection with certain fixed values of f and p), however the general case is unsolved, to the best of the authors knowledge. As has been stated already, it follows from the results in (Jansson and Wahlberg, 1997), that only imposing conditions on the input process of the type discussed above, will not be sufficient to ensure consistency of the methods, if we fix f and p.

3.3 Step 2: Approximation of rank n

In this step we use as input the estimate $\hat{\beta}_z$ from step 1. This matrix is an estimate for the matrix β_z, which is an approximation to $\mathcal{O}_f\mathcal{K}_p$. From equation (3.3) it is clear, that the matrix β_z has rank less or equal to the system order n. The estimate however will typically have full rank due to noise effects. In a first step let us assume, that the true order n of the system is known. Then in order to identify a system of order n it is thus necessary to find a rank n approximation of the (typically) full rank matrix $\hat{\beta}_z$. This is done using a singular value decomposition of a weighted version of this matrix: Let $\hat{U}\hat{\Sigma}\hat{V}^T = \hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$, where $\hat{W}_f^+ \in \mathbb{R}^{fs\times fs}$ and $\hat{W}_p^- \in \mathbb{R}^{p(m+s)\times p(m+s)}$ are weighting matrices. These weighting matrices may be chosen by the user. It has been noted by (Van Overschee, 1995) and (Peternell, 1995), that some of the proposed algorithms differ in the choice of the weighting matrices. The choices taken in the three most popular subspace methods (`CCA`, `N4SID` and `MOESP`) can be found in table 3.1. The rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ is obtained as $\hat{\mathcal{O}}_f = [(\hat{W}_f^+)^{-1}\hat{U}_n(\hat{\Sigma}_n)^{1/2}]$, $\hat{\mathcal{K}}_p = [(\hat{\Sigma}_n)^{1/2}\hat{V}_n^T\hat{W}_p^{-1}]$, where $\hat{U}_n$ denotes the first n columns of $\hat{U}$, $\hat{V}_n$ the first n columns of $\hat{V}$ and $\hat{\Sigma}_n$ denotes the heading $n\times n$ submatrix of $\hat{\Sigma}$, the diagonal matrix containing the singular values ordered decreasing in size.

In later chapters we will impose further restrictions on $\hat{W}_f^+$ and $\hat{W}_p^-$. For the moment it is sufficient to assume, that both matrices are nonsingular (at least from a certain T_0 onwards, almost everywhere). The relevance of the weighting matrices lies in the fact, that they determine the importance of certain directions in $\hat{\beta}_z$ for the estimation. This vague statement will be clarified in section 5.3, where it will be demonstrated, that the choice of the weighting matrices influences the (asymptotic) bias, if the estimation order is smaller than the true one, and that they also influence the asymptotic variance, if the true system is in the model class used for estimation.

In this step usually the order is estimated. It is common practice (see e.g. Van Overschee, 1995; Verhaegen and DeWilde, 1993) to specify the order by inspection of the singular values and include these singular values in the following steps, that are assessed to be significantly nonzero by visual inspection. Another method to estimate the order is based on the information contained in the singular values. Following (Fuchs, 1990) it is possible to define a criterion function similar to the usual information criteria like e.g. `AIC` or `BIC` (see e.g. Hannan and Deistler, 1988). This approach was investigated in (Peternell, 1995) and shown to be consistent

under the standard assumptions in this thesis (For this result the truncation indices f and p have to be fixed and finite). We will investigate the problem of order estimation in section 5.2.1, where also some simulations will be provided comparing different methods.

3.4 Step 3: Estimation of the System Matrices

In this section, we use as input from step 2 the rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ of $\hat{\beta}_z$. Starting from this estimate there are many different procedures to estimate the matrices (A, B, C, D, E, K). Some of the methods only apply to the case of no observed inputs (i.e. $B = 0, D = 0$). In the following we will give a review of the most important proposed methods, indicating the advantages and disadvantages without aiming at a complete discussion. Consider the following facts and definitions:

- $\mathcal{O}_f = [C^T, A^TC^T, (A^T)^2C^T, \cdots, (A^T)^{f-1}C^T]^T$ and thus $\mathcal{O}_{f-1}A = \mathcal{O}_f^{\uparrow}$, where $\mathcal{O}_f^{\uparrow}$ denotes the matrix obtained from $\mathcal{O}_f$ by omitting the first block row. This fact is known as the *shift invariance property*.
- The matrix $\mathcal{U}_f$ is linear in $\mathrm{vec}[B, D]$ i.e. $\mathrm{vec}[\mathcal{U}_f] = L\mathrm{vec}[B, D]$ (see equation 3.4).
- $x(t) = \mathcal{K}Z^-(t)$.

Note, that $\mathcal{O}_f$ and $\mathcal{K}$ depend on the particular realization of the true system.

Realization Based Methods

Realization based methods have their origin in the work of (Faurre, 1976; Akaike, 1975). They only apply to the case of no observed inputs ($B = 0, D = 0$). These methods are based on the fact, that (block) Hankel matrices calculated from the covariances $\gamma(j)$ of a stationary process $(y(t))_{t\in\mathbb{Z}}$ of the form (1.1) can be decomposed as:

$$\mathcal{H}_{f,p} = \begin{bmatrix} \gamma(1) & \gamma(2) & \cdots & \gamma(p) \\ \gamma(2) & \gamma(3) & \ddots & \\ \vdots & \ddots & & \\ \gamma(f) & & & \gamma(f+p-1) \end{bmatrix} = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{f-1} \end{bmatrix} [\bar{C}, A\bar{C}, \cdots, A^{p-1}\bar{C}] = \mathcal{O}_f\bar{\mathcal{C}}_p \quad (3.7)$$

Let P^- denote the minimal solution of the algebraic Riccati equation (ARE):

$$P = APA^T + (\bar{C} - APC^T)\Omega^{-1}(\bar{C} - APC^T)^T \quad (3.8)$$

where $\Omega = \gamma(0) - CPC^T = EE^T$. Here E denotes the lower triangular Cholesky factor of Ω (see also equation (2.13)). This equation can be solved, if $(A, \bar{C}, C, \gamma(0)/2)$ corresponds to a (strictly) positive real system (comp. also Lemma 2.2.3). In this case, $K = (\bar{C} - AP^-C^T)E^{-1}$ and the transfer function $\pi(A, K, C, E)$ is (strictly) minimumphase.

An obvious way to use this fact in a setting, where $B = 0, D = 0$ is to use the *shift invariance property*, i.e. $\mathcal{O}_{f-1}A = \mathcal{O}_f^{\uparrow}$, to obtain estimates $(\hat{A}, \hat{C})$ of (A, C): Note, that the matrix $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$, which is decomposed in step 2 of the algorithm is just a weighted version of the estimate of the covariance Hankel matrix, since there are no observed inputs present and therefore $\hat{\beta}_z = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}$, where $\hat{\mathcal{H}}_{f,p}$ denotes the sample covariance of $Y_{t,f}^+$ and $Y_{t,p}^-$, $\hat{\Gamma}_p^-$ denotes the sample variance of $Y_{t,p}^-$. Here the initial conditions for $y_j, j \leq 0$ are chosen, such that $\hat{\mathcal{H}}_{f,p}$ is a block Hankel matrix and $\hat{\Gamma}_p^-$ is a block Toeplitz matrix. However it is easy to see, that the choice of the initial conditions has no influence on the asymptotic behaviour (see also the proof of Lemma 4.1.4). Thus

the approximation in step 2 in fact calculates a rank n approximation to the covariance Hankel matrix as $\hat{\mathcal{O}}_f\hat{\bar{\mathcal{C}}}_p$, where $\hat{\mathcal{O}}_f$ has been calculated in step 2 and $\hat{\bar{\mathcal{C}}}_p = \hat{\mathcal{K}}_p\hat{\Gamma}_p^-$. Inserting the estimate $\hat{\mathcal{O}}_f$ instead of $\mathcal{O}_f$ in $\mathcal{O}_{f-1}A = \mathcal{O}_f^{\uparrow}$, we may obtain an estimate $\hat{A}$ for $f > n$ by ordinary least squares as $\hat{A} = \hat{\mathcal{O}}_{f-1}^{\dagger}\hat{\mathcal{O}}_f^{\uparrow}$, where $X^{\dagger}$ denotes the Moore-Penrose pseudoinverse $X^{\dagger} = (X^TX)^{-1}X^T$. (Chui and Maciejowski, 1996) show (see also section 5.4.1), that a slightly different algorithm leads to guaranteed stability of the estimated matrix $\hat{A}$: Let $\overline{\hat{\mathcal{O}}_f} = [(\hat{\mathcal{O}}_f^{\uparrow})^T, 0]^T$, where the matrix $0 \in \mathbb{R}^{n\times s}$. Then $\hat{A} = \hat{\mathcal{O}}_f^{\dagger}\overline{\hat{\mathcal{O}}_f}$ (using the same symbol with a slight abuse of notation) is guaranteed to be asymptotically stable i.e. $|\lambda_{max}(\hat{A})| \leq 1$. The stability of $\hat{A}$ is vital for the algorithm, since the Riccati equation (3.8) for $\hat{A}, \hat{C}, \hat{\bar{C}}$ and $\hat{\gamma}(0)$ can only be solved for P, if $\hat{A}$ is stable and the matrices $(\hat{A}, \hat{\bar{C}}, \hat{C}, \hat{\gamma}(0)/2)$ define a positive real transfer function (see section 2.2.2). Note, however, that for f fixed the introduction of the zero block may also lead to an asymptotic bias in the estimation. This can be seen as follows: Assume we obtain from step 2 a consistent estimate of $\mathcal{O}_f$ (which of course will be shown in the next chapter), then due to the minimality also the pseudoinverse of the estimate will converge to the true pseudoinverse $\mathcal{O}_f^{\dagger}$. If the matrix A is not nilpotent, we obtain $\mathcal{O}_f A - \overline{\mathcal{O}_f} = [0, \cdots, 0, (A^T)^f C^T]^T$ and thus A may not be found by using the Pseudoinverse $\mathcal{O}_f^{\dagger}$. Here $\overline{\mathcal{O}_f}$ denotes the limit of $\overline{\hat{\mathcal{O}}_f}$. This asymptotic bias obviously converges to zero for $f \to \infty$.

An estimate $\hat{C}$ of C is obtained from the first s rows of $\hat{\mathcal{O}}_f$. Finally $\hat{\bar{C}}$ may be obtained from the first s columns of $\hat{\bar{\mathcal{C}}}_p$. Together with the usual estimate $\hat{\gamma}(0)$, we can try to solve the ARE (3.8). However, the algorithm does not ensure, that the estimated function $\hat{\Phi}(z) = \hat{\gamma}(0)/2 + \hat{C}(zI - \hat{A})^{-1}\hat{\bar{C}}$ is positive real. As has been pointed out by (Lindquist and Picci, 1996), the situation, where the positive realness of the estimated transfer function $\hat{\Phi}(z)$ is violated leads to a breakdown in the algorithm in the sense, that the algorithm in this case fails to provide estimates at all (as opposed to bad estimates). For a statistical analysis of this method see section 6.1.

The Method used in `MOESP`

In this section we will present the algorithm of (Verhaegen, 1994). Although originally presented for a certain choice of W_f^+, W_p^-, the algorithm also applies for general weighting matrices, fulfilling the standard assumptions of nonsingularity in the limit. We will discuss the algorithm denoted with '`PO` scheme' in (Verhaegen, 1994), since this is the most general one of the `MOESP` class of algorithms and does not use other a-priori information (e.g. assumptions on the input process to be white noise, as does the basic `MOESP` algorithm). `PO-MOESP` was designed to estimate the transfer function $l(z)$, however there are obvious ways to modify the procedure to the estimation of l and k.

Given the rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ of $\hat{\beta}_z$ calculated in step 2, we may obtain estimates $(\hat{A}, \hat{C})$ of (A, C) using the shift invariance approach as described above. Assume, that the input u_t is a pseudostationary sequence, which is persistently exciting of order $(f + p)$ (see Appendix A). Note, that this definition includes the statement, that sums of the form $\frac{1}{T}\sum_{t=1}^{T-j} u_t u_{t+j}^T$ converge for $0 \leq j \leq f + p$. Also assume, that u_t and $\varepsilon(t)$ are wholly independent in the sense, that sums of the form $\frac{1}{T}\sum_{t=\max(1,j)}^{\min(T-j,T)} u_t\varepsilon_{t+j}^T, -(f + p) \leq j \leq (f + p)$, converge to zero a.s. uniformely in $|j| \leq H_T, H_T = O((\log T)^a)$.

Recall the definition of $\mathcal{U}_f$:

$$\mathcal{U}_f = \begin{bmatrix} D & 0 & \cdots & 0 \\ CB & D & \ddots & \vdots \\ \vdots & CB & \ddots & \\ CA^{f-2}B & & \ddots & D \end{bmatrix}$$

Note, that $\mathcal{U}_f$ depends linearily on B and D. Now from equation (3.3) we obtain $Y^+_{t,f} = \mathcal{O}_f\langle x_t, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1}U^+_{t,f} + \mathcal{U}_f U^+_{t,f} + \mathcal{E}_f\langle E^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1}U^+_{t,f} + Y^{+,\Pi}_{t,f}$, using the shorthand notation $\langle x_t, y_t\rangle = \frac{1}{T}\sum_{t=1}^{T} x_t y_t^T$, where for observations outside the range $1 \le t \le T$ the conditions are set such, that $\langle a_t, b_t\rangle = \langle a_{t+j}, b_{t+j}\rangle$, $|j| \le f + p$. Here the assumptions on the input process $(u(t))_{t\in\mathbb{Z}}$ ensure, that the inverse $\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1}$ exists for T large enough. Thus $\langle Y^+_{t,f}, U^+_{t,f}\rangle = \mathcal{O}_f\langle x_t, U^+_{t,f}\rangle + \mathcal{U}_f\langle U^+_{t,f}, U^+_{t,f}\rangle + \mathcal{E}_f\langle E^+_{t,f}, U^+_{t,f}\rangle$. Now $\langle E^+_{t,f}, U^+_{t,f}\rangle$ converges to zero under our assumptions. Next note, that $\mathcal{O}_f \in \mathbb{R}^{fs\times n}$ is of rank n and thus there exists a matrix $\mathcal{O}_f^\perp \in \mathbb{R}^{(fs-n)\times fs}$ such that $\mathcal{O}_f^\perp$ is of full rank and $\mathcal{O}_f^\perp\mathcal{O}_f = 0$. Thus

$$\mathcal{O}_f^\perp\langle Y^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1} = \mathcal{O}_f^\perp\mathcal{U}_f + \mathcal{O}_f^\perp\langle E^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1} \tag{3.9}$$

Finally replace $\mathcal{O}_f^\perp$ with the full rank matrix $\hat{\mathcal{O}}_f^\perp \in \mathbb{R}^{(fs-n)\times fs}$ satisfying $\hat{\mathcal{O}}_f^\perp\hat{\mathcal{O}}_f = 0$. Then the convergence of $\langle E^+_{t,f}, U^+_{t,f}\rangle$ to zero and the consistency of $\hat{\mathcal{O}}_f$ (which will be proved in section 6.2) show the convergence of $\hat{\mathcal{O}}_f^\perp\langle Y^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1}$ to $\mathcal{O}_f^\perp\mathcal{U}_f$. Thus it seems natural to obtain $\hat{B}$ and $\hat{D}$ from

$$(\hat{B}, \hat{D}) = \arg\min_{B,D} \|\hat{\mathcal{O}}_f^\perp\langle Y^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1} - \hat{\mathcal{O}}_f^\perp\mathcal{U}_f(\hat{A}, B, \hat{C}, D)\|_{Fr} \tag{3.10}$$

where the dependence of $\mathcal{U}_f$ on the system matrices has been made explicit. Since this problem is linear in (B, D), vectorization of the matrices reduces the optimization to a simple regression of $\Xi_1 = \mathrm{vec}[\hat{\mathcal{O}}_f^\perp\langle Y^+_{t,f}, U^+_{t,f}\rangle\langle U^+_{t,f}, U^+_{t,f}\rangle^{-1}]$ onto $L_1\mathrm{vec}[B, D]$, where the derivation of L_1 is straightforward. Similar arguments as given above lead to the equation

$$\langle Y^+_{t-p,p}, U^+_{t-p,p}\rangle - \mathcal{O}_p\langle x_{t-p}, U^+_{t-p,p}\rangle - \mathcal{E}_p\langle E^+_{t-p,p}, U^+_{t-p,p}\rangle = \mathcal{U}_p\langle U^+_{t-p,p}, U^+_{t-p,p}\rangle$$

This equation can be used analogously to equation (3.9) to add rows in the least squares problem of estimating B and D: $\Xi_2 = L_2\mathrm{vec}[B, D]$.

(Verhaegen, 1994) suggests to estimate (B, D) such that $\|\Xi - L\mathrm{vec}[B, D]\|_2$ is minimal, where $\Xi^T = [\Xi_1^T, \Xi_2^T], L^T = [L_1^T, L_2^T]$. In (Verhaegen, 1994) also an efficient implementation of the procedure is given for $f = p$ using the QR decomposition of a matrix composed of the data Hankel matrix (comp. REMARK 3.1.1):

$$\begin{bmatrix} U^f_{p+1} \\ U^p_1 \\ Y^p_1 \\ Y^f_{p+1} \end{bmatrix} = \begin{bmatrix} R_{11} & 0 & \cdots & 0 \\ R_{21} & R_{22} & \ddots & \vdots \\ R_{31} & R_{32} & R_{33} & 0 \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{bmatrix}$$

In this framework, all the regressions and inner products, which have to be estimated during the procedure, can be calculated easily in terms of the matrices R_{ij} i.e. the matrices Q_i do not have to be stored during the identification. For example $\langle Y^+_{t,f}, U^+_{t,f}\rangle = \frac{1}{T}R_{41}R_{11}^T, \langle U^+_{t,f}, U^+_{t,f}\rangle = \frac{1}{T}R_{11}R_{11}^T$. However from a computational point of view, the matrix on the left hand side of the above equation will be of dimension $(s + m)(f + p) \times T - f - p$. For big sample sizes T and high f and p, this matrix becomes quite big and thus the QR decomposition might be quite slow and it might be favourable to use the sample covariance estimates to achieve the information reduction in the data. From an asymptotic point of view, there is no difference between using the QR decomposition or the sample covariances (since this basically amounts in using different initial conditions). However in finite samples this does make a difference. The techniques used in Chapter 4 can be used to analyze this type of algorithms. For a discussion of some statistical properties of these estimates see section 6.2.

A slight adaptation of this method can be easily derived from equation (3.9): Consider the minimization problem

$$(\hat{B},\hat{D}) = \arg\min_{B,D} \sum_{t=1}^{T-f} \|\hat{\mathcal{O}}_f^{\perp} Y_{t,f}^{+} - \hat{\mathcal{O}}_f^{\perp} \mathcal{U}_f(\hat{A},B,\hat{C},D) U_{t,f}^{+}\|_{Fr} \tag{3.11}$$

This basically amounts in using a generalized weighting matrix in the LS fit in equation (3.10). This idea has an interpretation as minimizing the prediction error and has been given by (Ljung and Kelvey, 1996).

The `MOESP - PO` scheme originally was intended to be used for the identification of the deterministic part of the system i.e. the transfer function $l(z)$. However, several extensions to the estimation of $k(z)$ are quite obvious. The most popular of these seem to be the following two:

For $A, B, C, D, \mathbb{E}x(t)u(t)^T, \mathbb{E}u(t)u(t)^T$ (where the expectation has to be interpreted in the appropriate sense) and the state covariance P the following equations hold:

$$\mathbb{E}y(t)y(t)^T = [C,D] \begin{bmatrix} P & \mathbb{E}x(t)u(t)^T \\ (\mathbb{E}x(t)u(t)^T)^T & \mathbb{E}u(t)u(t)^T \end{bmatrix} \begin{bmatrix} C^T \\ D^T \end{bmatrix} + EE^T$$

$$P = [A,B] \begin{bmatrix} P & \mathbb{E}x(t)u(t)^T \\ (\mathbb{E}x(t)u(t)^T)^T & \mathbb{E}u(t)u(t)^T \end{bmatrix} \begin{bmatrix} A^T \\ B^T \end{bmatrix} + KK^T$$

Thus a natural idea would be to replace the true quantities with the estimated quantities, which come out of the estimation of the deterministic part of the system as well as the estimates of the state $x(t)$. In this way we obtain from the equations estimates of EE^T and KK^T respectively, which however are not guaranteed to be positive semidefinite. Thus here again problems with the positivity of the estimates might occur, which have to be taken into account in the algorithm. Also the algorithm has to deal with the fact, that the matrix $KK^T \in \mathbb{Z}^{n\times n}$ has rank equal to s, which will typically be smaller than n. Further note, that in general the estimate of the state and of the state covariance, will be asymptotically biased due to the finiteness of p and thus also the estimates of the noise systems may be biased, if we let p fixed, in contrast to the estimate of the transfer function $l(z)$ (which can be estimated consistently also for fixed f and p, see section 6.2).

A second approach could be to use the estimates of (A,B,C,D) in the system equations (1.1) as follows: Since $E\varepsilon_t = y_t - Cx_t - Du_t$ it seems to be natural to replace true quantities with their estimates, in order to achieve estimates of the unknown quantities i.e. $\hat{E}\hat{\varepsilon}_t = y_t - \hat{C}\hat{x}_t - \hat{D}u_t$. In this way, we obtain the estimate $\hat{E}$ as the lower triangular Cholesky factor of $\hat{E}\hat{E}^T = \langle y_t - \hat{C}\hat{x}_t - \hat{D}u_t, y_t - \hat{C}\hat{x}_t - \hat{D}u_t\rangle$, which also gives an estimate for $\hat{\varepsilon}_t$. If $\hat{E}$ is not invertible, we use the Pseudoinverse to calculate $\hat{\varepsilon}_t$. Note however, that in this setup the nonnegativity of $\hat{E}\hat{E}^T$ is 'built in'. $\hat{K}$ is obtained from least squares in $\hat{x}_{t+1} - \hat{A}\hat{x}_t - \hat{B}u_t = \hat{K}\hat{\varepsilon}_t + \hat{\rho}_t$. In this way, again the low rank property of $\hat{K}$ is ensured. Also note, that the estimates might again be asymptotically biased, due to the bias in the state estimation.

We will not attempt to provide a discussion of the statistical properties of the estimates of (E,K) obtained by either of the methods, since it is believed, that this method is only chosen, if one really only aims at obtaining an estimate of the transfer function l describing the input-output relations. However it is noted, that using the tools presented in the next chapter and using the results on the asymptotic behaviour of the estimates of (A,B,C,D), the asymptotic distributions of $\hat{E}$ and $\hat{K}$ could be derived quite easily (at least) for the second approach.

Using the State as a Regressor

In this section we will present the algorithm, which will be central in the analysis provided in this thesis. The reason for this lies in the simplicity of the approach, which makes it easy to analyze, and in the fact, that it was proposed in several of the most popular algorithms, e.g. in the presentation of the canonical variate approach (in this thesis we will always refer to this approach as `CCA`) by (Larimore, 1983) and in a version of `N4SID` (Van Overschee and DeMoor, 1994, algorithm 2) . Maybe the strongest argument for choosing this class of algorithms to be the central in this thesis lies in the good performance, as can be seen e.g. in (Peternell, 1995), where it is compared to

realization based approaches, and in the fact, that the asymptotic variance in the case where no inputs have been observed proved to be close to the Cramer-Rao bound in a couple of examples (comp. Bauer *et al.*, 1997b).

The main idea in this algorithm is similar to the one used in the last paragraph of the last section: From the rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ of $\hat{\beta}_z$ calculated in step 2, we obtain an estimate of the state sequence $\hat{x}_t = \hat{\mathcal{K}}_p Z^-_{t,p}$. This estimate is used as a regressor together with u_t in the equation $y_t = Cx_t + Du_t + E\varepsilon_t$, resulting in

$$y_t = \hat{C}\hat{x}_t + \hat{D}u_t + \hat{\nu}_t \tag{3.12}$$

where $\hat{\nu}_t$ denotes the residuals. From this equation we obtain estimates $\hat{C}$ of C, $\hat{D}$ of D and an estimate $\hat{\nu}_t$ of the innovation sequence. This estimate also provides an estimate $\hat{\Omega} = \hat{E}\hat{E}^T$ of the innovation variance. Note that the Cholesky factor $\hat{E}$ always exists due to the construction of $\hat{\Omega}$. The uniqueness of the Cholesky factor follows from the nonsingularity of the matrix $\hat{\Omega}$ for large T, which will be proved in the next chapter. From this an estimate of the noise sequence is obtained as $\hat{\varepsilon}_t = \hat{E}^{-1}\hat{\nu}_t$. In order to use the state equation also an estimate of x_{t+1} is needed, which can be obtained for example from simply reusing the state estimate $\hat{x}_t$ shifted one time instant (comp. Larimore, 1983), or from a new SVD (comp. Van Overschee and DeMoor, 1994). Both approaches can be subsummed in the general statement $\tilde{x}_{t+1} = \tilde{\mathcal{K}}_{p+1} Z^-_{t+1,p+1}$. With this estimate the matrices A, B and K may be estimated using the following regression:

$$\tilde{x}_{t+1} = \hat{A}\hat{x}_t + \hat{B}u_t + \hat{K}\hat{\varepsilon}_t + \hat{\rho}_t \tag{3.13}$$

Again note, that even for $T \to \infty$, $\hat{x}_t$ will (in general) not converge to x_t, if the column truncation index p is fixed, since we neglect the information from the far past. Thus in this framework the column truncation index p has to tend to infinity for the estimates to be consistent (comp. Chapter 4).

Using this algorithm, there is no problem with the positivity condition as occured for the realization based procedures, since here we estimate directly the system matrices, without the need for spectral factorization. On the other hand, it is also not guaranteed, that the estimated system fulfills the stability or the miniphase assumption. A discussion on the properties of the estimates corresponding to the location of the poles and the zeros of the estimated transfer functions may be found in section 5.4.

3.5 The Main Algorithm

From the last sections it should be clear, that there are many different algorithms, that are termed 'subspace algorithms'. They all use as the basic fact the property of the state to be the interface between the past and the future of the process y_t in the sense, that the best linear prediction of the future (using the knowledge of the whole input sequence) is a linear function of the state. This property is also documented in equation (3.2). There have been many suggestions, how to use this information for the estimation of the system matrices. It is beyond the scope of this thesis to give a complete analysis of all the different algorithms. As has already been stated, we will restrict ourselves mainly to the analysis of one particular algorithm. As a byproduct, the essential facts for some of the other algorithms will be provided. This will be used, in order to provide the most obvious results for the other algorithms. However, we do not aim at a complete description of all algorithms presented above corresponding to their statistical properties, let alone a comparison among them.

The class of algorithms considered in this thesis can be described as follows:

1. Estimate β_z from equation (3.3) by regression of $Y^{+,\Pi}_{t,f}$ onto $Z^{-,\Pi}_{t,p}$ (see section 3.2 for a definition of the notation) resulting in an estimate $\hat{\beta}_z$

2. Find a rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p$ of $\hat{\beta}_z$ using the SVD of $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$ as described in section 3.3.

3. Estimate the system matrices using the state as a regressor as described in section 3.4, using $\tilde{x}_{t+1} = \hat{x}_{t+1}$

The regressions in step 1 are performed using initial conditions such that all sample covariances corresponding to the same lag are identical. Thus e.g. the sample covariance of $Y_{t,f}^+$, denoted with $\hat{\Gamma}_f^+$, will be a block Toeplitz matrix and $\hat{\mathcal{H}}_{f,p} = \langle Y_{t,f}^+, Y_{t,p}^- \rangle$ will be a block Hankel matrix. Corresponding to the asymptotic properties of the estimates we are free to choose initial conditions, since the effects of these choices are negligible. Again we want to emphasize, that also for the 'data Hankel matrices' setup the same statement holds true.

Chapter 4

Statistical Analysis of the Main Algorithm

In this chapter we will present the main results of this thesis: For the class of subspace algorithms presented in the last chapter, we will analyze the asymptotic properties. The discussion will center on the asymptotic distribution of the estimates and the asymptotic variance. Since the analysis is already complex, we will split the results up into several parts: First we will deal with the case of no observed inputs ($B = 0, D = 0$) and fixed finite $f \geq n$. This case is included, since it will be shown in one example (comp. section 5.3), that for some choices of the weighting matrices it might be reasonable to keep the row truncation index f finite, since the performance seems to decrease with increasing f. Then we will extend the results to the case of $f \to \infty$ (see section 4.2). Again the reason for doing so are calculations showing that for `CCA` the performance of the algorithms increases with increasing f in some examples. Finally we will also include observed inputs (see section 4.3). Throughout this chapter we will assume the system order n to be known and used for estimation. However, in section 5.2.1 we will discuss procedures for estimating the order and in section 5.3 we discuss the effects of choosing the order too low.

4.1 The Case of fixed $f, m = 0$

In this section it is assumed throughout, that there are no observed inputs present. We will also assume, that in the algorithms $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$. The motivation for this restriction lies in the fact, that this choice -to the best of the authors knowledge - is the only one, proposed in the literature for the main class of algorithms considered here. A heuristic motivation of the restriction could be the following argument: Consider the singular value decomposition $\hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-T/2} = \hat{U}\hat{\Sigma}\hat{V}^T$. Now computing the vector $\hat{x}_t^{aug} = (\hat{\Sigma})^{1/2}\hat{V}^T(\hat{\Gamma}_p^-)^{-1/2}Y_{t,p}^-$ we obtain, that $\langle \hat{x}_t^{aug}, \hat{x}_t^{aug} \rangle = \hat{\Sigma}$ holds and thus the components of $\hat{x}_t^{aug}$ are uncorrelated in finite samples. The main class of algorithms considered here uses the first n components of $\hat{x}_t^{aug}$ as an estimate of the state x_t. With the particular choice of the weighting $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$ thus the estimate of the state in finite sample has the property of decomposing the future of the outputs $Y_{t,f}^+$ into a part due to the past of the inputs (i.e. $\hat{\mathcal{O}}_f \hat{x}_t$) and a part, which is orthogonal to the state. Note, that this holds for any specification of the order. Thus for any specification of the order, the estimated state shows the property of splitting the information in the past of the output into an (in some sense) essential part, contributing to the prediction of the output, and into an orthogonal part, which is due to the noise. The same property for the true state is documented in equation (3.1). However, it is not clear, that this particular choice of the weighting $\hat{W}_p^-$ leads to better (in terms of the asymptotic variance) estimates. The simulation results given in (Peternell, 1995) provide some empirical evidence for this conjecture. Thus in the sequel we will only discuss the case of $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$. We would like to emphasize, that this restriction is not a necessary condition for our results to hold, but that the

result will also hold with more general weighting matrices, where possibly the restriction on the truncation indices have to be adapted.

In order to be able to state the main result, we need a few more definitions and preliminary facts, which will be given in the next section.

4.1.1 The Set U_n^+

In order to state a central limit theorem for the estimates $(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T)$ of the system matrices, first the particular realization of the true transfer function k_0, which is the limit of the estimates for $T \to \infty$, will be determined. As is well known, for minimal state space representations, the class of all (minimal) observationally equivalent system matrices corresponding to the transfer function k is given by different choices of the basis in the state space (comp. Chapter 2). In the algorithms considered here, the choice of the basis is done implicitely by decomposing the rank n approximation to $\hat{\beta}_{f,p}$ into $\hat{\mathcal{O}}_f \hat{\mathcal{K}}_p$, where in the notation the dependence of $\hat{\beta}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}$ on the indices f and p has been stressed. This decomposition is performed using the SVD of $\hat{W}_f^+ \hat{\beta}_{f,p}(\hat{\Gamma}_p^-)^{1/2}$ and is unique, if the first n singular values of this matrix have multiplicity one and if the orientation of the singular vectors is fixed. Let Γ^- denote the population variance of $Y^-(t)$ and let $\mathcal{H}_f$ denote the population covariance between $Y_f^+(t)$ and $Y^-(t)$. As will be shown in the proof in section 4.1.3, the matrix $\hat{X}_{f,p} = \hat{W}_f^+ \hat{\beta}_{f,p}(\hat{\Gamma}_p^-)\hat{\beta}_{f,p}^T(\hat{W}_f^+)^T = \hat{W}_f^+ \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T(\hat{W}_f^+)^T$ converges to $\bar{X} = W_f^+ \mathcal{H}_f(\Gamma^-)^{-1}\mathcal{H}_f^T(W_f^+)^T = W_f^+ \mathcal{O}_f \mathcal{K}\Gamma^- \mathcal{K}^T \mathcal{O}_f^T(W_f^+)^T$ a.s., if the index p is a function of the sample size T, which tends to infinity at a certain rate (see Theorem 4.1.1). For the moment we only note, that this follows from the uniform convergence of the sample covariance estimates (see Theorem A.3.1).

In this section we will restrict the set of all possible weighting matrices to two possible choices: Either $\hat{W}_f^+ \in \mathbb{R}^{fs \times fs}$ is assumed to be equal to $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ (i.e. `CCA` is used), where Γ_f^+ denotes the population variance of $Y_f^+(t)$, or the weighting is in close correspondence to a rational transfer function $k_W = \sum_{j=0} K_W(j) z^{-j}$, where $K_W(0)$ is assumed to be nonsingular, in the following way (see also McKelvey, 1995): Let $W_f^+(k_W)$ denote the weighting matrix connected to $k_W(z)$ as:

$$W_f^+(k_W) = \begin{bmatrix} K_W(0) & 0 & \cdots & 0 \\ K_W(1) & K_W(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ K_W(f-1) & \cdots & K_W(1) & K_W(0) \end{bmatrix} \tag{4.1}$$

Although it is not difficult to be more general corresponding to the choice of the weighting matrices (some possible generalizations include the choice of the weighting matrices depending on the sample size), we refrain from doing so for two reasons: First it seems complicated to find minimal conditions on the weighting matrices to ensure our result to hold, and second the merits of admitting more general weightings must be doubted in the context of asymptotic theory. For example letting the W_f^+ depend on the sample size T, where $\|W_f^+(T) - W_f^+\| = o(T^{-1/2})$, where W_f^+ is nonsingular, in some matrix norm is sufficient to ensure the asymptotic distribution result to hold, with the same asymptotic distribution, which is obtained by using W_f^+ for all T.

As will be shown, from the a.s. convergence of the sample covariance estimates (see Theorem A.3.1) it follows, that if the n nonzero eigenvalues of $\bar{X}$ are distinct, then for T large enough, the n largest singular values of $(\hat{W}_f^+)\hat{\beta}_{f,p}(\hat{\Gamma}_p^-)^{1/2}$ will be distinct too. Let $U_n^{(m)}$ denote the set of all rational, stable, strictly minimum-phase transfer functions k of McMillan degree n with a constant term, which is lower diagonal and has strictly positive diagonal entries. Furthermore let $U_n^+ \subset U_n^{(m)}$ denote the subset of all transfer functions $k \in U_n^{(m)}$, for which $\bar{X}$ has n distinct (non-zero) eigenvalues. Then we may state the following lemma:

Lemma 4.1.1 *U_n^+ is generic in $U_n^{(m)}$ in the sense that it is an open and dense subset, where $U_n^{(m)}$ is endowed with the so called pointwise topology (comp. section 2.2.2).*

PROOF: The proof for this statement is similar to the proof of Theorem 2.2.5 and thus we will only give an outline here.

Let $S \subseteq \mathbb{R}^{(n+s)^2-s(s-1)/2}$ denote the set of quadruples (A, K, C, E), where A and $A - KE^{-1}C$ are stable and E is lower triangular with strictly positive diagonal elements. Furthermore let $S_n \subset S$ denote the set of all minimal realizations. It is easy to see, that the set S is an open and nonvoid subset of $\mathbb{R}^{(n+s)^2-s(s-1)/2}$ and that the set of all minimal realizations S_n is an open and dense subset of S. Consider the set S_n^+, where the corresponding matrix $\bar{X}$ has n distinct (non-zero) eigenvalues. Clearly S_n^+ is a subset of S_n and the next step is to prove that S_n^+ is open and dense in S.

First note that $\mathcal{H}_f(\Gamma^-)^{-1}\mathcal{H}_f^T = \mathcal{O}_f\mathcal{C}\mathcal{C}^T\mathcal{O}_f^T$, where $\mathcal{C} = [K, AK, A^2K, \ldots]$ is the controllability matrix. Therefore the n nonzero eigenvalues of $\bar{X}$ are equal to the eigenvalues of $\bar{Z} = [\mathcal{O}_f^T(W_f^+)^T W_f^+ \mathcal{O}_f][\mathcal{C}\mathcal{C}^T]$. Since A is stable $P = \mathcal{C}\mathcal{C}^T = \sum A^j K K^T (A^j)^T$ is an analytic function of (A, K), The analyticity of P also implies that the autocovariances $\gamma(0) = CPC^T + EE^T$, $\gamma(j) = CA^{j-1}(APC^T + KE^T)$, $j > 0$ are analytic functions of (A, K, C, E). Finally since E has full rank and by the minimum-phase assumption it follows that the entries of $\bar{Z}$ are analytic functions of (A, K, C, E) on S for the `CCA` case. For general weighting, the argument given above holds analogously. As in the proof of Theorem 2.2.5 the analyticity of the determinant of the Sylvester matrix R corresponding to the characteristical polynomial of $\bar{Z}$ and its first derivative follows for all $(A, K, C, E) \in S$. This shows the openness.

Now suppose that S_n^+ were not dense in S_n. Then $\det R$ were zero on an open subset $\mathcal{V} \subset S_n$. Now from the analyticity of $\det R$ is follows that $\det R$ is zero on the largest pathwise connected subset of S, which contains $\mathcal{V}$. Since S is pathwise connected, we conclude that $\det R$ is zero on S. To find the desired contradiction it is sufficient to construct an element in S_n^+, i.e. to prove that S_n^+ is not empty. This is done in a recursive manner. Clearly for $n = 1$, $S_1^+ = S_1$ is non void. Now suppose that the conjecture is true for $n - 1$. Therefore there exists a transferfunction $k_0 \in U_{n-1}^{(m)}$ for which $\bar{Z}$ has $n - 1$ non zero distinct eigenvalues. There exists a nonminimal realization $(A_0, K_0, C_0, E_0) \in S_n$ of k_0 and since S_n^+ is dense in S_n we obtain a sequence $(A_T, K_T, C_T, E_T) \to (A_0, K_0, C_0, E_0)$. The continuity of π together with the continuity of the eigenvalues of the finite dimensional matrix Z then shows the nonemptyness of S_n^+. □

As was mentioned already, the SVD of $W_f^+\mathcal{H}_f(\Gamma^-)^{-T/2}$ is used to fix a particular realization of $k_0 \in U_n^{(m)}$. However, even for distinct singular values, the SVD is not unique due to the possibility to choose the orientation of the singular vectors. One way to obtain uniqueness is to fix the sign of certain prescribed entries e.g. in U_n to be strictly positive, where $\bar{X} = U_n \Sigma^2 U_n^T$. This clearly can be done only, if these entries are nonzero. Choosing one entry in each column to be positive, we obtain a unique SVD of $W_f^+\mathcal{H}_f(\Gamma^-)^{-T/2}$. Choosing the SVD such that the same entries in the n singular vectors corresponding to the largest n singular values are positive also in a neighborhood of $\bar{X}$ then results in uniqueness of the essential part of the SVD for T large enough, since (as will be shown in the proof) $\hat{X}_{f,p} \to \bar{X}$ a.s. We will assume throughout that the SVD is performed according to this normalization and thus the actual estimates might differ from the ones of the 'ideal' subspace algorithm described above by a basis change of the form $T = diag(\pm 1, \cdots, \pm 1)$. The results stated in Theorem 4.1.1 will hold also for this different normalization, as long as the continuity of the essential part of the SVD at the true system (i.e. for $T = \infty$) holds. If the choice of the orientation is chosen such that the essential part of the SVD is not continuous at the true system, then a different normalization has to be used, in order to achieve consistency and asymptotic normality of the system matrix estimates. Note, however, that there are only 2^n different realizations corresponding to the different choices of the orientations of the singular vectors, which have to be taken into account.

Since we are only interested in the part of the SVD, which corresponds to the nonzero singular values, we will adopt the following notion, in order to simplify the arguments:

DEFINITION (ESSENTIAL CONTINUITY OF THE SVD) The SVD of a matrix $X_0 = U_n \Sigma_n V_n^T$ is said to be *essentially continuous*, if $X_T \to X_0$, where the corresponding SVD can be written as $X_T = U_{T,n}\Sigma_{T,n}V_{T,n}^T + R_{T,n}$ and $\Sigma_{T,n}$ contains the largest n singular values ordered decreasing in size, implies $U_{T,n} \to U_n, \Sigma_{T,n} \to \Sigma_n$ and $V_{T,n} \to V_n$.

It follows from the results in Appendix B, that the SVD is essentially continuous at X_0, if the nonzero singular values of X_0 are distinct, and if the orientation of the singular vectors is chosen suitably, e.g. by fixing the sign of a nonzero entry, as has been done above. Thus we will always assume, that the normalization of the singular vectors has been chosen, such that the SVD is essentially continuous.

4.1.2 A Central Limit Theorem

In this section the main result of this part of the thesis i.e. asymptotic normality of the subspace estimates will be stated. As has already been stated, in our framework, the truncation index p has to tend to infinity in order for the algorithm to produce consistent estimates. This is essentially due to the fact, that in the first step a regression is performed, neglecting information from the far past (contained in $x(t-p)$). For a central limit theorem convergence of the estimates of order $\sqrt{T}$ is needed, thus a lower bound on the increase of the index p has to be imposed in order to ensure that the effect of neglecting the far past does not show up in the limiting distribution. On the other hand, the limited amount of data imposes upper bounds for the increase of p in order to ensure a uniform convergence of the estimates of the covariance sequence (comp. Theorem A.3.1). The following theorem states a result on the asymptotic distribution of the estimates obtained by the main class of algorithms:

Theorem 4.1.1 *Let $(y_t)_{t\in\mathbb{Z}}$ be generated by the true transfer function $k_0 \in U_n^+$, where the ergodic white noise $(\varepsilon(t))_{t\in\mathbb{Z}}$ fulfills the following conditions (the standard assumptions):*

$$\begin{aligned} \mathbb{E}\{\varepsilon(t)|\mathcal{F}_{t-1}\} &= 0 \\ \mathbb{E}\{\varepsilon(t)\varepsilon(t)^T|\mathcal{F}_{t-1}\} = \mathbb{E}\{\varepsilon(t)\varepsilon(t)^T\} &= I \\ \mathbb{E}\{\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)|\mathcal{F}_{t-1}\} &= \omega_{a,b,c} \\ \mathbb{E}\{\varepsilon_i(t)^4\} &< \infty \end{aligned}$$

where $\mathbb{E}$ denotes expectation, $\mathcal{F}_t$ the σ-algebra spanned by the past of the noise and additional subscripts here indicate components of the vector $\varepsilon(t)$. $\omega_{a,b,c}$ is a constant and $f \geq n$ is a fixed integer. Let $\hat{W}_f^+$ be chosen either to be $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ (i.e. `CCA`) or $\hat{W}_f^+ = W_f^+(k_W)$ for some rational transfer function k_W with nonsingular constant term.

If p fulfills the following conditions:

1. *$p \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ for some $d > 1$, where $\rho_0 = \lambda_{max}(A_0 - K_0 E_0^{-1} C_0)$*
2. *$p/(\log T)^a \to 0$ for some $a < \infty$.*

then

$$\sqrt{T}vec[\hat{A}_T - A_0, \hat{K}_T - K_0, \hat{C}_T - C_0, \hat{E}_T - E_0] \xrightarrow{d} Z$$

where Z is a multivariate normal random variable with zero mean and (singular) variance V_f.

The special realization (A_0, K_0, C_0, E_0) of the true transfer function k_0, which is the limit for $T \to \infty$ corresponds to the normalization of the singular vectors described in the previous section. This realization depends on the weighting matrix $\hat{W}_f^+$. The asymptotic variance V_f and the set U_n^+ also depend on the choice of the weighting matrices. However this is not emphasized in the expressions for notational convenience. Note, that the assumptions on the noise process are exactly the same as those given in section 2.4, where the asymptotic normality of maximum

likelihood estimates is stated. It is important to note, that the asymptotic distribution will also depend on the fourth moment of the noise $\varepsilon(t)$ and thus the result is not independent of the distribution of the noise. Also for ML estimation the distribution of the estimates of the system matrices will depend on the distribution of the noise, however the dependence on the distribution can be removed, if we estimate the transfer function in a slightly different setup: Consider $k_0(z) = (I + C_0(zI - A_0)^{-1}\bar{K}_0)E_0 = \bar{k}_0(z)E_0$. Then the asymptotic distribution of the ML estimate of $\bar{k}_0$ in the set $\overline{M_n}$ is independent of the distribution of the noise, whereas the asymptotic distribution of the ML estimate of E_0 is not (see e.g. Hannan and Deistler, 1988). For the subspace estimates an analogous result will be established in section 5.5.

Finally note that the lower bound of the increase of the truncation index p depends on the true system. However there are possibilities to estimate this bound consistently in a certain sense (see section 5.1).

4.1.3 Proof of the Theorem

To simplify the notation, we will only consider the case of Larimore's procedure `CCA`, i.e. the case where $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$. The case of general constant weighting can be treated completely analogously. In fact for this case some steps simplify considerably.

Preliminaries

For the class of algorithms considered in this paper it is straightforward to prove, that the estimates $(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T)$ are a nonlinear function of the sample autocovariances. For given indices f and p the estimates depend only on the sample covariances up to lag $f+p-1$ i.e. on $\hat{\gamma}(0), \hat{\gamma}(1), \cdots, \hat{\gamma}(f+p-1)$, where $\hat{\gamma}(j)$ denotes the estimate $\frac{1}{T}\sum_{t=1}^{T} y_t y_{t-j}^T$ of $\gamma(j) = \mathbb{E}y(t)y(t-j)^T$. As has been stated already, for consistency of the main algorithm the column truncation index p has to tend to infinity at a certain rate (comp. Deistler *et al.*, 1995). Thus also the number of included covariance estimates tends to infinity, which is the main technical complication in the proof.

Using the shorthand notation $\langle a_t, b_t \rangle = \frac{1}{T}\sum_{t=1}^{T} a_t b_t^T$ again, the regressions in step 3 can be written as follows:

$$\hat{C}_T = \langle y_t, \hat{x}_t \rangle \langle \hat{x}_t, \hat{x}_t \rangle^{-1} \tag{4.2}$$

$$\hat{E}_T = (\hat{\gamma}(0) - \hat{C}_T \langle \hat{x}_t, \hat{x}_t \rangle \hat{C}_T^T)^{1/2} \tag{4.3}$$

$$\hat{A}_T = \langle \hat{x}_{t+1}, \hat{x}_t \rangle \langle \hat{x}_t, \hat{x}_t \rangle^{-1} \tag{4.4}$$

$$\hat{K}_T = \langle \hat{x}_{t+1}, \hat{\varepsilon}_t \rangle = (\langle \hat{x}_{t+1}, y_t \rangle - \langle \hat{x}_{t+1}, \hat{x}_t \rangle \hat{C}_T^T) \hat{E}_T^{-T} \tag{4.5}$$

where the noise sequence ε_t is estimated by $\hat{\varepsilon}_t = \hat{E}_T^{-1}(y_t - \hat{C}_T \hat{x}_t)$. Note that the estimates of the noise are orthogonal to the estimates of the state, i.e. $\langle \hat{x}_t, \hat{\varepsilon}_t \rangle = 0$, and thus $\hat{A}_T$ and $\hat{K}_T$ may be obtained by the two seperate regressions (4.4) and (4.5).

In the following the above expressions will be further analyzed. The arguments will be given for the `CCA` case $(\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2})$ only, however the generalizations to the case $\hat{W}_f^+ = W_f^+(k_W)$ are obvious. The estimates of the states are defined by $\hat{x}_t = \hat{\mathcal{K}}_p Y_{t,p}^-$ and $\hat{x}_{t+1} = \hat{\mathcal{K}}_p Y_{t+1,p}^-$. As noted in the previous chapter, this choice of the estimate of the state x_{t+1} is just one possibility, however for the proof it is the most convenient one. The tools provided in this chapter can be used to obtain similar results for different choices of $\hat{x}_{t+1}$. The matrix $\hat{\mathcal{K}}_p$ is computed from the matrix $\hat{X}_{f,p} = \hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-1} \hat{\mathcal{H}}_{f,p}^T (\hat{W}_f^+)^T$ in the following way: Recall from section 4.1.1, that $\hat{\Sigma}_n$ contains the square roots of the largest n eigenvalues of the matrix $\hat{X}_{f,p}$ and $\hat{U}_n$ contains the corresponding eigenvectors, i.e.

$$\hat{X}_{f,p} \hat{U}_n = \hat{U}_n \hat{\Sigma}_n^2 \text{ and } \hat{U}_n^T \hat{U}_n = I_n$$

The matrices $\hat{\mathcal{O}}_f$ and $\hat{\mathcal{K}}_p$ are defined as

$$\begin{aligned}\hat{\mathcal{O}}_f &= (\hat{\Gamma}_f^+)^{1/2}\hat{U}_n\hat{\Sigma}_n^{1/2} \\ \hat{\mathcal{K}}_p &= \hat{\Sigma}_n^{1/2}\hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2} = \hat{\Sigma}_n^{-1}\hat{\mathcal{O}}_f^T(\hat{\Gamma}_f^+)^{-1}\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\end{aligned}$$

Furthermore let

$$T_k = \begin{bmatrix} I_s \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{ks\times s} \text{ and } S_p = \begin{bmatrix} 0 & \cdots & & \\ I_s & 0 & & \\ 0 & I_s & 0 & \\ & \ddots & \ddots & \ddots \\ & & & I_s \quad 0 \end{bmatrix} \in \mathbb{R}^{ps\times ps}.$$

where $I_s \in \mathbb{R}^{s\times s}$ denotes the identity matrix. Using the identities $y_t = T_f^T Y_{t,f}^+ = T_p^T Y_{t+1,p}^-$ and $Y_{t+1,p}^- = T_p T_f^T Y_{t,f}^+ + S_p Y_{t,p}^-$ it is straightforward to derive the following expressions:

$$\langle \hat{x}_t, \hat{x}_t\rangle = \hat{\Sigma}_n \tag{4.6}$$

$$\langle y_t, \hat{x}_t\rangle = T_f^T\hat{\mathcal{O}}_f\hat{\Sigma}_n \tag{4.7}$$

$$\langle y_t, \hat{x}_{t+1}\rangle = T_p^T\hat{\mathcal{H}}_{f,p}^T(\hat{\Gamma}_f^+)^{-1}\hat{\mathcal{O}}_f\hat{\Sigma}_n^{-1} \tag{4.8}$$

$$\begin{aligned}\langle \hat{x}_{t+1}, \hat{x}_t\rangle = {} & \hat{\Sigma}_n^{-1}\hat{\mathcal{O}}_f^T(\hat{\Gamma}_f^+)^{-1}(\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}T_p)T_f^T\hat{\mathcal{O}}_f\hat{\Sigma}_n + \\ & \hat{\Sigma}_n^{-1}\hat{\mathcal{O}}_f^T(\hat{\Gamma}_f^+)^{-1}(\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}S_p\hat{\mathcal{H}}_{f,p}^T)(\hat{\Gamma}_f^+)^{-1}\hat{\mathcal{O}}_f\hat{\Sigma}_n^{-1}\end{aligned} \tag{4.9}$$

Inspection of the above expressions shows that the estimates of the parameter matrices are obtained using a nonlinear mapping from $\hat{\gamma}(0), \ldots, \hat{\gamma}(f)$ and the finite dimensional matrices $\hat{X}_{f,p} = \hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T(\hat{W}_f^+)^T$, $\hat{Y}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}S_p\hat{\mathcal{H}}_{f,p}^T$ and $\hat{Z}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}T_p$.

In order to outline the proof of the CLT some further notation is introduced. Here vec(.) denotes the vector of stacked vectorizations of the corresponding matrices.

$$\begin{aligned}&\hat{\theta}_T = \text{vec}(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T) \\ &\theta_0 = \text{vec}(A_0, K_0, C_0, E_0) \\ &\hat{m}_p = \text{vec}(\hat{\gamma}(0), \ldots, \hat{\gamma}(f), \hat{X}_{f,p}, \hat{Y}_{f,p}, \hat{Z}_{f,p}) \\ &m_p = \text{vec}(\gamma(0), \ldots, \gamma(f), X_{f,p}, Y_{f,p}, Z_{f,p}) \\ &m = \lim_{p\to\infty} m_p = \text{vec}(\gamma(0), \ldots, \gamma(f), X_0, Y_0, Z_0) \\ &\hat{g}_{T,h} = \text{vec}(\hat{\gamma}(0), \ldots, \hat{\gamma}(h-1)) \\ &g_h = \text{vec}(\gamma(0), \ldots, \gamma(h-1))\end{aligned}$$

Here for example $X_{f,p}$ is defined as $X_{f,p} = W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}\mathcal{H}_{f,p}^T(W_f^+)^T$ and $X_0 = \lim_{p\to\infty} X_{f,p} = W_f^+\mathcal{H}_f(\Gamma^-)^{-1}\mathcal{H}_f^T(W_f^+)^T$. $Y_{f,p}, Y_0$ and $Z_{f,p}, Z_0$ are defined analogously. It will be part of the proof to show, that the limit $\lim_{p\to\infty} m_p$ exists. As has been stated already $\hat{m}_p$ is a function of the sample autocovariances $\hat{g}_{T,f+p}$, $\hat{m}_p = \phi(\hat{g}_{T,f+p})$ say. The population counterparts are connected via $m_p = \phi(g_{f+p})$. The estimates $\hat{\theta}_T$ of the system matrices are obtained as a function $\hat{\theta}_T = \psi(\hat{m}_p)$. Therefore the proof of the CLT may be decomposed into the following four steps:

1. A central limit theorem for the covariances $\hat{g}_{T,h}$, with $h = (f+p) \to \infty$ suitably (see section 4.1.3).

2. The proof, that $\sqrt{T}(m_p - m_0) \to 0$, and that $\psi(m_0) = \theta_0$ where θ_0 corresponds to the realization of the true system as described in section 4.1.1.

3. A central limit theorem for $\hat{m}_p$: $\sqrt{T}(\hat{m}_p - m_0) \xrightarrow{d} \mathcal{N}(0, V^m)$, for $p \to \infty$ suitably.

4. The proof of the differentiability of the mapping ψ at the point m_0. In other words it is proved that $\psi(m_0 + \delta m) - \psi(m_0) = J_\psi(\delta m) + o(||\delta m||)$ and thus by 3. one obtains

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = \sqrt{T}(\psi(\hat{m}_p) - \psi(m_0)) \xrightarrow{d} Z$$

where Z is a multivariate Gaussian variable with mean zero and variance $V_f = (J_\psi)V^m(J_\psi)^T$.

This structure will also be used for all the other cases discussed below (i.e. the case, where f is allowed to tend to infinity, and the case of observed inputs): We first define a finite dimensional vector, from which the estimates of the system matrices can be calculated. Then the proof essentially consists of two parts: The derivation of the asymptotic distribution of this finite dimensional vector and secondly the verification of the properties of the mapping attaching the system matrix estimates to vectors $\hat{m}_p$. However, the definition of the vectors $\hat{m}_p$ will differ in the various proofs.

Some Useful Lemmas

In the various proofs in this chapter, the following technical results will be used quite often. First we deal with the vectorization of products of structured matrices.

Lemma 4.1.2 *Let $x(a) \in \mathbb{R}^a$ be a sequence of vectors, such that $a||x(a) - x_{0,a}||_1 \to 0$, where $x_{0,a}$ denotes the vector of the first a components of the sequence x_0. Assume, that x_0 has the property, that $|(x_0)_i| \leq C\rho_x^i$ for some constant C and some scalar $\rho_x < 1$. Here $(x_0)_i$ denotes the i-th component of x_0. Let $y(a)$ denote a vector with the same properties as $x(a)$, i.e. $a||y(a) - y_{0,a}||_1 \to 0, |(y_0)_i| \leq C\rho_y^i, \rho_y < 1$. Let $H = [h_{i+j-1}]_{i,j \in \mathbb{N}}$ denote a Hankelmatrix and let $T = [t_{i-j}]_{i,j \in \mathbb{N}}$ denote a Toeplitzmatrix. Let $H_{a,b} \in \mathbb{R}^{a \times b}$ denote the northeast corner of the Hankel matrix H and let $T_a \in \mathbb{R}^{a \times a}$ denote the northeast corner of the Toeplitzmatrix T.*

Then $x^T(a)H_{a,b}y(b) = L_{a,b}^H(h_1, h_2, \cdots, h_{a+b-1})'$, where $||L_{a,b}^H - L_{0,a,b}^H||_1 \to 0$ for $a = b \to \infty$. Here $L_{0,a,b}^H$ denotes the vector of the first $a + b - 1$ components of L_0^H, where $|(L_0^H)_i| \leq C(\max\{\rho_x, \rho_y\} + \varepsilon)^i$, where $\varepsilon > 0$ is arbitrary.

Furthermore $x^T(a)T_a y(a) = L_{a,+}^T(t_0, t_1, \cdots, t_{a-1})' + L_{a,-}^T(t_{-1}, \cdots, t_{-a+1})'$, where $||L_{a,+}^T - L_{0,a,+}^T||_1 \to 0$ and for $||L_{a,-}^T - L_{0,a,-}^T||_1 \to 0$ $a \to \infty$. Here $L_{0,a,\pm}^T$ denotes vectors of the first $a - 1$ components of $L_{0,\pm}^T$, where $|(L_{0,\pm}^T)_i| \leq C(\max\{\rho_x, \rho_y\} + \varepsilon)^i$, where $\varepsilon > 0$ is arbitrary.

PROOF: First consider the case of the Hankel matrix: It is easily seen, that the vectorization of $x_0^T H y_0 = L_0(h_1, h_2, \cdots)'$ exists, since the entries corresponding to the i-th element are equal to $\sum_{j=1}^i (x_0)_{i-j+1}(y_0)_j$ and thus the modulus of this entry can be bounded by $\sum_{j=1}^i |(x_0)_{i-j+1}(y_0)_j| \leq C\sum_{j=1}^i \rho_x^{i-j+1}\rho_y^j \leq C(\max\{\rho_x, \rho_y\} + \varepsilon)^i$. Here ε can be set equal to zero, if $\rho_x \neq \rho_y$, whereas $\varepsilon > 0$ is needed, if $\rho_x = \rho_y$. However, it is obvious, that ε can be chosen, such that $\max\{\rho_x, \rho_y\} + \varepsilon < 1$ holds. This proves the exponential decrease of the entries and the existence of the vector L_0^H as an element of ℓ^1, the space of all absolutely summable sequences. Finally the result for the Hankel matrix case follows from

$$\begin{aligned} |\sum_{j=1}^i (x_0)_j(y_0)_j - (x(a))_j(y(b)_j| &\leq \sum_{j=1}^i |(x_0)_{i+1-j} - (x(a))_{i+1-j}|||y(b)||_\infty + |(y_0)_j - (y(b))_j|||x_0||_\infty \\ &\leq ||y(b)||_\infty ||x(a) - x_{0,a}||_1 + ||x_0||_\infty ||y(b) - y_{0,b}||_1 \end{aligned}$$

Here $||x||_\infty = \max_{1 \leq j \leq i} |x_i|$ for $x \in \mathbb{R}^i$.

Similar arguments show the same result for the case of the Toeplitz matrix, where the i-th entry of the limiting vector $L_{0,+}^T$ for example is equal to $\sum_{j=0}^\infty (x_0)_{i+j}(y_0)_j$, and thus the difference to the corresponding entry in $L_{a,+}^T$ can be bounded by

$$
\begin{aligned}
|\sum_{j=0}^{\infty}(x_0)_{i+j}(y_0)_j - \sum_{j=0}^{a-i}(x(a))_{i+j}(y(a))_j| &\leq |\sum_{j=a-i+1}^{\infty}(x_0)_{i+j}(y_0)_j| + \\
&\quad \sum_{j=1}^{a-i}|(x_0)_{i+j}(y_0)_j - (x(a))_{i+j}(y(a))_j| \\
&\leq C(\rho_x^a + \|x(a) - x_{0,a}\|_1 + \|y(a) - y_{0,a}\|_1
\end{aligned}
$$

This shows the result. □

Although this result has been formulated for the case of Hankel (and Toeplitz) matrices, it is obvious, that the lemma generalizes to the case of block-Hankel (and block Toeplitz respectively) matrices. It can also be seen, that the condition $a = b \to \infty$ has only been introduced for convenience and is not vital for the result. In order to ensure the convergence of the $L_{a,b}^H$ to L_0^H in ℓ^1 norm it is sufficient, that $(a+b)\|x(a) - x_{0,a}\|_1 \to 0$ and $(a+b)\|y(a) - y_{0,a}\|_1 \to 0$ respectively. Also the case, where only b tends to infinity, convergence results for $L_{a,b}^H$ can be derived easily.

In this lemma, the exponential decrease of the entries in x_0 and y_0 has been vital. The next lemma states conditions, under which the exponential decrease of $x_0 = Ty_0$ follows from the exponential decrease in y_0:

Lemma 4.1.3 *Let $y(a)$ be as in Lemma 4.1.2 and let $T = [T_{i,j}]_{i,j\in\mathbb{N}}$ denote a matrix with the property, that $|T_{i,j}| \leq C\rho_T^{i-j}$. Then $x(a) = T_a y(a)$ fulfills $a\|x(a) - x_{0,a}\|_1 \to 0$ where x_0 denotes the vector Ty_0, which has entries decreasing exponentially, i.e. $|(x_0)_i| \leq C(\max\{\rho_y, \rho_T\} + \varepsilon)^i$ for all $\varepsilon > 0$. Here T_a denotes the heading $a \times a$ submatrix of T.*

PROOF: The proof follows from simple algebraic manipulations: Consider $a\|x(a) - x_{0,a}\|_1 = a\|T_a y(a) - T_a y_{0,a} - T_{a,a+1:\infty}(y_0)_{a+1:\infty}\|_1 \leq \|T_a\|_1 a\|y(a) - y_{0,a}\|_1 + \|T_{a,a+1:\infty}\|_1\|y_{a+1:\infty}$. Here $T_{a,a+1:\infty}$ denotes the matrix obtained from the matrix containing the first a rows of T by omitting the first a columns. $y_{a+1:\infty}$ denotes the vector obtained from y_0 by omitting the first a components. Thus the result on the norm $\|x(a) - x_{0,a}\|_1$ follows from the boundedness of the various matrix norms, which follow from the exponential decrease in the elements of T and from the fact, that $\|(y_0)_{a+1:\infty}\|_1 \leq C\rho_y^a$. It remains to prove the exponential decrease of the entries in x_0 and thus the existence of the limit. Consider e.g. the i-th entry of x_0:

$$
\begin{aligned}
|\sum_{j=1}^{\infty}(y_0)_i T_{i,j}| &\leq \sum_{j=1}^{i}|(y_0)_j T_{i,j}| + \sum_{j=i+1}^{\infty}|(y_0)_i T_{i,j}| \\
&\leq C\sum_{j=1}^{i}\rho_y^j \rho_T^{i-j} + C\rho_y^i
\end{aligned}
$$

In the case $\rho_T > \rho_y$ the first sum can be bounded by $C\rho_T^i \sum_{j=1}^{i}(\rho_y/\rho_T)^j \leq C\rho_T^i$, if the converse equality holds, the sum is bounded by $C\rho_y^i$. Finally for $\rho_T = \rho_y$ the sum is equal to $Ci\rho_y^i$ and thus we need to introduce the $\varepsilon > 0$ in order for the result to hold. This shows the exponential decrease of the entries of x_0, where the rate is given in the lemma. □

Also this result can be easily adapted to block Toeplitz matrices. These two results will be often used in the sequel.

A CLT for the Covariance Estimates

As has already been stated, the main technical complication lies in the fact, that the index p and thus also the dimension of the stacked vector of autocovariances $\hat{g}_{T,f+p}$ has to tend to infinity.

Tools to handle the growth of dimensions are provided in a paper by (Lewis and Reinsel, 1985). Their techniques are used to prove the following lemma:

Lemma 4.1.4 *For fixed h let $V_h^g \in \mathbb{R}^{hs^2 \times hs^2}$ denote the covariance matrix of $\sqrt{T}(\hat{g}_{T,h} - g_h)$ defined above. Let h depend on T such that $h = o((\log T)^\alpha)$ for some $\alpha < \infty$. Then under the standard assumptions on the process (y_t) and on the noise sequence (ε_t) given in Theorem 4.1.1, for every sequence of vectors $l(T) \in \mathbb{R}^{hs^2}$ satisfying $0 < c_1 \leq l(T)^T V_h^g l(T)$ and $\|l(T)\|_1 \leq c_2 < \infty$ it follows that:*

$$\sqrt{T} \frac{l(T)^T (\hat{g}_{T,h} - g_h)}{(l(T)^T V_h^g l(T))^{1/2}} \xrightarrow{d} Z \tag{4.10}$$

where here Z denotes a normally distributed scalar random variable with zero mean and unit variance and $\xrightarrow{d}$ denotes convergence in distribution.

PROOF: In a first step the lemma is proved for the case $y_t = \varepsilon_t$. In this case $\sqrt{T}(\hat{g}_{T,h} - g_h) = \frac{1}{\sqrt{T}} \sum_{t=1}^T \text{vec}[\varepsilon(t)\varepsilon(t)^T - I, \varepsilon(t)\varepsilon(t-1)^T, \cdots, \varepsilon(t)\varepsilon(t-h+1)^T] = \text{vec}[e_0, e_1, \cdots, e_{h-1}]$, where $e_i = \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon(t)\varepsilon(t-i)^T, i = 1, \cdots, h-1$ and $e_0 = \frac{1}{\sqrt{T}} \sum_{t=1}^T \varepsilon(t)\varepsilon(t)^T - I$. Now let a sequence $l(T) \in \mathbb{R}^{hs^2}$ satisfying $0 < c_1 \leq l(T)^T V_h^\varepsilon l(T), \|l(T)\|_1 \leq c_2$ be given, where the notation indicates, that V_h^ε corresponds to $y(t) = \varepsilon(t)$. Furthermore let $v_{T,h}^2 = l(T)^T V_h^\varepsilon l(T)$ denote the variance of $\sqrt{T} l(T)^T (\hat{g}_{T,h} - g_h)$, which is bounded due to the bound on the norm of $l(T)$ and the (block diagonal) structure of V_h^ε (see Remark 4.1.1 below). Then following the proof of Theorem 3 in (Lewis and Reinsel, 1985) $\sqrt{T} l(T)^T (\hat{g}_{T,h} - g_h)/v_{T,h} = \sum_{t=1}^T X_t(T) = \sum_{t=1}^T \frac{1}{\sqrt{T}} l(T)^T \text{vec}[\varepsilon(t)\varepsilon(t)^T - I, \cdots, \varepsilon(t)\varepsilon(t-h+1)^T]/v_{T,h}$. Then the assumptions on $\varepsilon(t)$ imply that $X_t(T)$ and $X_s(T)$ are uncorrelated for $t \neq s$ with expectation equal to zero and that the variance of $X_t(T)$ is equal to $\frac{1}{T}$. $\sum_{t=1}^n X_t(T), n = 1, \cdots, T$ is a martingale sequence for each T. In order to prove the convergence of $\sqrt{T} l(T)^T (\hat{g}_{T,h} - g_h)/v_{T,h}$ to a normal distribution, it is thus sufficient to show, that

(a) $\sup_{t \leq T} X_t^2(T) \xrightarrow{p} 0$ for $T \to \infty$

(b) $\sum_{t=1}^{\lfloor \tau T \rfloor} X_t^2(T) \xrightarrow{p} \tau, 0 < \tau \leq 1$ for $T \to \infty$

where $\xrightarrow{p}$ denotes convergence in probability and where $\lfloor x \rfloor$ denotes the integer part of x. Write $X_t(T) = X_{t,0}(T) + X_{t,1}(T)$, where $X_{t,0}(T) = \frac{1}{\sqrt{T v_{T,h}^2}} l_0(T)^T \text{vec}(\varepsilon(t)\varepsilon(t)^T - I)$, where $l_0(T)$ denotes the vector of the first s^2 elements of $l(T)$. Then $\sup_{t \leq T} X_{t,0}(T)^2 \xrightarrow{p} 0$, since $\sup_{t \leq T} X_{t,0}(T)^2 \leq \sup_{t \leq T} X_{t,0}(t)^2$ and $\mathbb{E} X_{t,0}(t)^2 \to 0$ for $t \to \infty$ (compare Hannan and Deistler, 1988, p.149). $\sup_{t \leq T} X_{t,1}(T)^2 \xrightarrow{p} 0$ follows from the arguments in (Lewis and Reinsel, 1985). Thus the convergence of $\sup_{t \leq T} X_t(T)^2$ to zero follows from $X_t(T)^2 \leq 2(X_{t,0}(T)^2 + X_{t,1}(T)^2)$.

For condition (b) note, that $\sum X_{t,0}(T)^2$ converges due to ergodicity of $\varepsilon(t)$ and thus of $(\varepsilon(t)\varepsilon(t)^T - I)^2$ and the assumption of finite fourth moments. The convergence of $\sum X_{t,1}(T)^2$ can be seen, using the arguments of (Lewis and Reinsel, 1985).

Finally the contribution of the mixed terms $\sum X_{t,0}(T) X_{t,1}(T) \xrightarrow{p} 0$, which can be seen as follows: $|\sum X_{t,0}(T) X_{t,1}(T)| \leq \sup_{t \leq T} |X_{t,0}(T)| \sum |X_{t,1}(T)|$. Now $\sup_{t \leq T} |X_{t,0}(T)|$ converges in probability to zero, since $\sup_{t \leq T} X_{t,0}(T)^2$ does, and $\mathbb{P}\{\sum |X_{t,1}(T)| > \delta\} \leq \sum \mathbb{P}\{|X_{t,1}(T)| > \delta\} = \sum \mathbb{P}\{X_{t,1}(T)^2 > \delta^2\} \to 0$ (see the proof in Lewis and Reinsel, 1985). This shows the convergence in distribution to a random variable, which is normally distributed with mean zero and variance unity. Note, that the normalization by $v_{T,h}$ is necessary here, if no conditions on the limiting behaviour of $l(T)^T V_h^\varepsilon l(T)$ except for its boundedness are imposed.

In order to extend this result to the autocovariances of a process $(y(t))_{t \in \mathbb{Z}}$, generated by the model (1.1), note that $y(t) = \sum_{j=0}^\infty K(j)\varepsilon(t-j)$ holds, and that the Markoff parameters $K(i)$ converge exponentially to zero i.e. $\|K(j)\| \leq c(\rho_p')^j$ for some constant $c > 0$ and for arbitrary

$\rho_p' > |\rho_p|$ and $\rho_p = |\lambda_{\max}(A)| < 1$ denotes the modulus of a largest eigenvalue of A. The following assessement will be often used:

$$\mathbb{E}\| \sum_{t\in\mathcal{I}} (\varepsilon(t)\varepsilon(t-i)^T - \delta_{0,i}I)\|^2 \le \sum_{t\in\mathcal{I}} \mathbb{E}\|(\varepsilon(t)\varepsilon(t-i)^T - \delta_{0,i}I)\|^2 \le c|\mathcal{I}|. \tag{4.11}$$

Here $|\mathcal{I}|$ denotes the number of elements of the indexset $\mathcal{I}$. The above inequalities follow from the fact that the terms $\mathrm{vec}(\varepsilon(t)\varepsilon(t-i)^T - \delta_{0,i}I)$ and $\mathrm{vec}(\varepsilon(s)\varepsilon(s-i)^T - \delta_{0,i}I)$ are uncorrelated for $t \neq s$.

Now substituting $y(t) = \sum K(j)\varepsilon(t-j)$ in $\hat{\gamma}(j)$ and in $\gamma(j)$ one obtains:

$$\sqrt{T}(\hat{\gamma}(j) - \gamma(j)) = \sum_{i,l=0}^{a} K(i)\bar{e}_{j+l-i}K(l)^T + r(j) \tag{4.12}$$

where $\bar{e}_j = e_j$ for $j \ge 0$ and $\bar{e}_j = e_{-j}^T$ for $j < 0$. The term $r(j)$ may be decomposed into four components:

The first one is due to the replacement of $\mathbb{E}\hat{\gamma}(j) = ((T-j)/T)\gamma(j)$ by $\gamma(j)$. This term may be bounded by ch/T, and thus clearly is $o(T^{-1/2})$. The second contribution is due to the approximation of the process $y(t)$ by the finite sums $\sum_{i=0}^{a} K(i)\varepsilon(t-i)$. Because of (4.11) and the exponential decrease of the Markoff parameters the expectation of the Frobenius norm of this term is bounded by $c\rho_p^a$. Therefore this term converges to zero if a converges to infinity quicker than $-\log T/(2\log|\rho_p|)$, implying that $\rho_p^a = o(T^{-1/2})$.

The third contribution is due to the fact, that sums like $\sum_{t=j+1-i}^{T-i}(\varepsilon(t)\varepsilon(t-j-l+i)^T - \delta_{0,j+l-i}I)$ are replaced by e_{j+l-i}, i.e. by sums where the summation index runs from $1,\ldots,T$. The difference is the sum of at most $(2a+1)$ summands. Therefore by (4.11) and by the exponential decrease of the Markoff parameters the expectation of the Frobenius norm of the contribution due to this term may be bounded by $ca/\sqrt{T}$.

The last contribution stems from the replacement of e_i with e_{-i}^T for $i < 0$. Now e_i and e_{-i}^T differ only in the first i and last i summands. Therefore by the same reasonings as above this term may be bounded by $c(a+h)/\sqrt{T}$.

Putting together these considerations imply that

$$\mathbb{E}\|r(j)\| \le C_1(a+h)\frac{1}{\sqrt{T}} + C_2\rho_p^a \tag{4.13}$$

where the constants C_1, C_2 do not depend on j.

Now the first term in equation (4.12) is analyzed in more detail. Let $M_l^a = \sum_{i=\max(0,-l)}^{\min(a,a-l)} K(l+i)\otimes K(i)$ and $M_l = \lim_{a\to\infty} M_l^a = \sum_{i=\max(0,-l)}^{\infty} K(l+i)\otimes K(i)$, then

$$\mathrm{vec}\left(\sum_{i,l=0}^{a} K(i)\bar{e}_{j+l-i}K(l)^T\right) = \sum_{l=-a}^{a} M_l^a \mathrm{vec}(\bar{e}_{j+l})$$

Note that $M = (\ldots, M_{-1}, M_0, M_1, \ldots)$ has rows which are elements of ℓ_1 and that $M^a = (\ldots, M_{-1}^a, M_0^a, M_1^a, \ldots)$ converges to M by Lemma 4.1.2. The proof of Lemma 4.1.2 also shows, that $\|M^a - M\| \le c\rho_p^a$. Here the sequence $M_l^a = 0 \in \mathbb{R}^{s^2\times s^2}$ for $|l| > a$. Now let the permutation matrix $\Pi \in \mathbb{R}^{s^2\times s^2}$ be defined such that $\Pi\mathrm{vec}(H) = \mathrm{vec}(H^T)$ for every matrix $H \in \mathbb{R}^{s\times s}$. Then it follows that the first term in (4.12) can be written as

$$\mathrm{vec}\left(\sum_{i,l=0}^{a} K(i)\bar{e}_{j+l-i}K(l)^T\right) = \sum_{l=0}^{a+h} L_{j,l}^a \mathrm{vec}(e_l) = L_j^{a,h}\mathrm{vec}(e_0, e_1, \ldots, e_{a+h})$$

where $L_{j,l}^a = M_{-j}^a$ for $l = 0$ and $L_{j,l}^a = M_{l-j}^a + M_{-l-j}^a\Pi$ for $l > 0$. Clearly the rows of the matrix $L_j^{a,h}$, when extended with zeros to lie in $\mathbb{R}^{s^2\times\infty}$ converge in the ℓ_1 sense to the rows of the $(s^2\times\infty)$

matrix $L_j = (L_{j,0}, L_{j,1}, \ldots)$, where $L_{j,l} = \lim_{a\to\infty} L^a_{j,l}$. In fact by the convergence of $M^a \to M$, it follows that $\|L^a_j - L_j\|_1 \leq c\rho^a_p$ holds, where the constant does not depend on j.

Assembling the expressions for the covariances $\sqrt{T}(\hat{\gamma}(j) - \gamma(j))$ in the vector $\sqrt{T}(\hat{g}_{T,h} - g_h)$ then leads to

$$\sqrt{T}(\hat{g}_{T,h} - g_h) = L^{a,h}\mathrm{vec}(e_0, e_1, \ldots, e_{a+h}) + \mathrm{vec}(r(0), r(1), \ldots, r(h))$$

where the matrix $L^{a,h} \in \mathbb{R}^{hs^2\times(a+h)s^2}$ has as its j-th block row the matrix $L^{a,h}_j$. By the convergence of the rows of $L^{a,h}$ and by the block diagonal structure of V^ε_{a+h} it follows that, for fixed h, the variance of $\sqrt{T}(\hat{g}_{T,h} - g_h)$ exists and is given by $V^g_h = \lim_{a\to\infty} L^{a,h}V^\varepsilon_{a+h}(L^{a,h})^T$, see also REMARK 4.1.1 below.

Finally let $l(T) \in \mathbb{R}^{hs^2}$ be a sequence of vectors fulfilling $0 < c_1 \leq l(T)^T V^g_h l(T)^T$ and $\|l(T)\|_1 \leq c_2 < \infty$. Then

$$\sqrt{T}l(T)^T(\hat{g}_{T,h} - g_h) = l(T)^T L^{a,h}\mathrm{vec}(e_0, e_1, \ldots, e_{a+h}) + l(T)^T\mathrm{vec}(r(0), r(1), \ldots, r(h))$$

Now from (4.13) it follows that the second term converges to zero in probability. In addition the convergence of L^a_j to L_j and the bounded ℓ_1-norm of $l(T)$ implies that

$$\lim\left(l(T)^T L^{a,h}V^\varepsilon_{a+h}(L^{a,h})^T l(T) - l(T)^T V^g_h l(T)\right) = 0$$

It follows that $l(T)^T L^{a,h}V^\varepsilon_{a+h}(L^{a,h})^T l(T)$ is a sequence bounded from below and above and thus the first part of the proof gives the desired result.

Finally note, that the result is not changed, if in the definition of the sample covariances the summation is changed, such that the difference is a sum of up to $(f + p)$ terms. This can be seen using similar reasoning as in equation (4.11), the only difference being the fact, that $y(t)y(t-j)^T$ and $y(s)y(s-j)^T$ are not uncorrelated, and thus $|\mathcal{I}|$ has to be replaced by $|\mathcal{I}|^2$ in equation (4.11). □

REMARK 4.1.1 Recall, that the variance matrix V^g_h can be calculated as follows:

$$V^g_h = \lim_{a\to\infty} L^{a,h}V^\varepsilon_{h+a}(L^{a,h})^T$$

This limit exists due to the structure of V^ε_x and $L^{a,h}$ and due to the convergence of $L^{a,h}$ for fixed h. Now the covariance matrix of $vec[e_0, \cdots, e_{x-1}]$, which has been denoted as $V^\varepsilon_x \in \mathbb{R}^{xs^2\times xs^2}$, where x is an arbitrary integer, has as its $[js^2+(b-1)s+a, is^2+(d-1)s+c]$ entry the following expression: $\frac{1}{T}\sum_{t,s=1}^T[\mathbb{E}(\varepsilon_a(t)\varepsilon_b(t-j) - \delta_{j0}\delta_{ab})(\varepsilon_c(s)\varepsilon_d(s-i) - \delta_{i0}\delta_{cd})]$, where δ denotes the Kronecker delta function. For $s < t$ conditional expectation on $\mathcal{F}_{t-1}$ shows, that the contribution is zero and vice versa. Thus we have to examine only the contribution for $t = s$. If $j \neq i$, then again taking the expectation conditional on $\mathcal{F}_{t-\min(i,j)}$ shows, that the expectation is zero (here we need the assumption on the third order moments to simplify the expressions). Thus V^ε_x is blockdiagonal. In order to calculate the blockdiagonal entries, we distinguish the two cases $i = j > 0$ and $i = j = 0$. As can easily be seen, for $j > 0$, the expectation $\mathbb{E}(\varepsilon_a(t)\varepsilon_b(t-j))(\varepsilon_c(t)\varepsilon_d(t-j)) = \delta_{ac}\delta_{bd}$, which is equal to 1, if $a = c$ and $b = d$ and zero else, thus the variance matrix in this case is equal to the identity. For $j = 0$, the expectation $\mathbb{E}(\varepsilon_a(t)\varepsilon_b(t) - \delta_{ab})(\varepsilon_c(t)\varepsilon_d(t) - \delta_{cd}) = \mathbb{E}\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)\varepsilon_d(t) - \delta_{ab}\delta_{cd}$. Now for Gaussian $\varepsilon(t)$, the fourth moment $\mathbb{E}\varepsilon_a(t)\varepsilon_b(t)\varepsilon_c(t)\varepsilon_d(t)$ is equal to $\delta_{ab}\delta_{cd}+\delta_{ac}\delta_{bd}+\delta_{ad}\delta_{bc}$. Thus the expectation $\delta_{ac}\delta_{bd}+\delta_{ad}\delta_{bc}$ is equal to 2, if $a = b = c = d$, equal to 1, if $a \neq b$ and $a = c \wedge b = d$ or $a \neq b$ and $a = d \wedge b = c$, and equal to zero else. This is exactly, where the dependence on the distribution of the noise enters the asymptotic distribution. For non-Gaussian noise $\varepsilon(t)$ the asymptotic variance of $e_0 = T^{-1/2}\sum_{t=1}^T(\varepsilon(t)\varepsilon(t)^T - I)$ will not equal the equations derived above. In the proof of the central limit theorem for ML estimation given in (Hannan and Deistler, 1988) it is thus shown, that for the estimation of the transfer function $\bar{k}_0 = k_0E_0^{-1}$, the term due to e_0 can be eliminated without changing the asymptotic distribution. This should be compared with section 5.5.

REMARK 4.1.2 The lemma also shows, that the condition $L_h V_h^g L_h^T \to V^m, L_h \in \mathbb{R}^{m \times hs^2}$ *(m fixed and finite), where the rows of* L_h *are of bounded* ℓ_1 *norm uniformely in h, is a sufficient condition for* $\sqrt{T} L_h(\hat{g}_{T,h} - g_h) \xrightarrow{d} Z$, *where Z is multivariate normally distributed with mean zero and variance equal to* V^m *(see e.g. Anderson, 1971, Theorem 7.7.7). Since* V_h^g *is a matrix with elements of bounded infinity norm for a stable system, where the bound is independent of h, it is straightforward to see, that a sufficient condition for* $L_h V_h^g L_h^T \to V$ *to hold for some V is that the rows of* L_h *embedded in the* ℓ_1 *converge in the* ℓ_1 *norm to an infinite dimensional vector, having elements decreasing exponentially. As can be seen from* $V_h^g = \lim_{a\to\infty} L^{a,h} V_{h+a}^{\varepsilon} (L^{a,h})^T$ *and the convergence of the rows of* L_h *also the limit* V^m *is independent from the rate of increase of h within the bounds given in the theorem. This will be the condition used in the sequel. Note, that the requirement* $l(T)^T V_h^g l(T) > 0$ *is only needed for the normalization of the variance and thus can be dropped for our purposes.*

Convergence of $m_p \to m_0$

It will be proved that $\|m_p - m_0\| = O(|\rho_0^{'}|^p), \forall |\rho_0| < |\rho_0^{'}| < 1$, and ρ_0 denotes an eigenvalue of $(A - KE^{-1}C)$ of maximum modulus. For this purpose the following lemma is used:

Lemma 4.1.5 $\|\mathcal{H}_{f,p}(\Gamma_p^-)^{-1} - \mathcal{O}_f \mathcal{K}_p\|_{Fr} = O(|\rho_0^{'}|^p)$

PROOF: The relation $\mathcal{H}_f(\Gamma^-)^{-1} = \mathcal{O}_f \mathcal{K}$ implies

$$[\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}, 0](\Gamma^-)_p - \mathcal{O}_f \mathcal{K}(\Gamma^-)_p = 0$$

where $(\Gamma^-)_p$ denotes the first p block columns of the infinite dimensional matrix Γ^-. From this it follows, that $\mathcal{H}_{f,p}(\Gamma_p^-)^{-1} - \mathcal{O}_f \mathcal{K}_p = \mathcal{O}_f (A - KE^{-1}C)^p \mathcal{K} \tilde{\mathcal{H}}_p (\Gamma_p^-)^{-1}$. Here $\tilde{\mathcal{H}}_p$ denotes the matrix obtained by omitting the first p block rows in $(\Gamma^-)_p$, which is a (reordered) part of the covariance Hankelmatrix $\mathcal{H} = \mathbb{E} Y^+(t)(Y^-(t))^T$ and thus has finite Frobenius norm, independently of p. The Frobenius norm of $\mathcal{O}_f$ can also be bounded (independently of f), Γ_p^- has bounded eigenvalues independently of p and finally $\mathcal{K}$ is of finite Frobenius norm. Thus the Frobenius norm of the error can be bounded by the Frobenius norm of $(A - KE^{-1}C)^p$ times a constant, which depends only on the underlying system and not on the choice of the truncation indices. Now $\|(A - KE^{-1}C)^p\|$ can be bounded by $C|\rho_0^{'}|^p$, where ρ_0 denotes an eigenvalue of $(A - KE^{-1}C)$ of maximal modulus. Here $\rho_0^{'}$ is introduced due to the possibility of multipel eigenvalues of modulus $|\rho_0|$. If there exists only one eigenvalue of modulus $|\rho_0|$, then $\rho_0^{'} = \rho_0$ can be used. □

This lemma immediately implies that

$$\|Z_{f,p} - Z_0\| = \|\mathcal{H}_{f,p}(\Gamma_p^-)^{-1} T_p - \mathcal{H}_f(\Gamma^-)^{-1} T_\infty\| = O(|\rho_0'|^p)$$

Furthermore it is easy to see that

$$\begin{aligned} \|X_{f,p} - X_0\| &= \|\hat{W}_f^+ \mathcal{H}_{f,p}(\Gamma_p^-)^{-1} \mathcal{H}_{f,p}^T (\hat{W}_f^+)^T - W_f^+ \mathcal{H}_f(\Gamma^-)^{-1} \mathcal{H}_f^T (W_f^+)^T\| \\ &\le \|\hat{W}_f^+ (\mathcal{H}_{f,p}(\Gamma_p^-)^{-1} - \mathcal{O}_f \mathcal{K}_p) \mathcal{H}_{f,p}^T (W_f^+)^T\| \\ &\quad + \|W_f^+ \mathcal{O}_f (A - KE^{-1}C)^p \mathcal{K}\| \|\mathcal{H}_f^T (W_f^+)^T\| \\ &\quad + 2\|(\hat{W}_f^+ - W_f^+)\| \|\mathcal{O}_f \Sigma_n \mathcal{O}_f^T\| \end{aligned}$$

is of the desired order $O(|\rho_0'|^p)$. Also

$$\|Y_{f,p} - Y_0\| = \|\mathcal{H}_{f,p}(\Gamma_p^-)^{-1} S_p \mathcal{H}_{f,p}^T - \mathcal{H}_f(\Gamma^-)^{-1} S_\infty \mathcal{H}_f^T\| = O(|\rho_0'|^p)$$

as follows with the techniques used in Lemma 4.1.5.

In order to obtain $\sqrt{T}\|m_p - m_0\| \to 0$ it is sufficient that $\sqrt{T}|\rho_0'|^p \to 0$ holds, which is guaranteed by the condition $p \ge -(d \log T)/(2 \log |\rho_0|)$ for some $d > 1$. These considerations show, that one has to impose a lower bound on the convergence rate of $p \to \infty$ to ensure, that the convergence of

the matrices (A, K, C, E) obtained by performing the subspace algorithm using the true covariances instead of sample estimates, is of order $o(T^{-1/2})$.

Next it is proved that $\psi(m_0) = \theta_0$. In other words the subspace algorithms considered in this section are an implementation of a realization algorithm, if the true covariances are used rather than sample estimates and if the column truncation index $p = \infty$. Note that by construction the eigenvalue decomposition $X_0 = U_n \Sigma_n^2 U_n^T$ determines a factorization of the rank n matrices $\mathcal{H}_f = \mathcal{O}_f \bar{\mathcal{C}}$ and $\mathcal{H}_f(\Gamma^-)^{-1} = \mathcal{O}_f \mathcal{K}$, where

$$\begin{aligned} \mathcal{O}_f &= (C_0^T, A_0^T C_0^T, (A_0^2)^T C_0^T, \ldots, (A_0^{f-1})^T C_0^T)^T \\ \bar{\mathcal{C}} &= (\bar{C}_0, A_0 \bar{C}_0, A_0^2 \bar{C}_0, \ldots) \\ \mathcal{K} &= (K_0 E_0^{-1}, (A_0 - K_0 E_0^{-1} C_0) K_0 E_0^{-1}, \ldots) \\ E_0 E_0^T &= \gamma(0) - C_0 \Sigma_n C_0^T \\ \bar{C}_0 &= A_0 \Sigma_n C_0^T + K_0 E_0^T \end{aligned}$$

and the matrices (A_0, K_0, C_0, E_0) are the realization of the transfer function k_0, described in section 4.1.1.

Now the mapping ψ evaluated at the point m_0 is investigated more closely. Let $(A, K, C, E) = \psi(m_0)$. Then from (4.2) and (4.7) and the factorization of $\mathcal{H}_f(\Gamma^-)^{-1}$ as given above one obtains that $C = T_f^T \mathcal{O}_f = C_0$ and thus from (4.3) and $\mathcal{K}\Gamma^-\mathcal{K}^T = \Sigma_n = \mathbb{E}x(t)x(t)^T$ that $E = E_0$. The term (4.9) simplifies to

$$\begin{aligned} &\Sigma_n^{-1} \mathcal{O}_f^T (\Gamma_f^+)^{-1} [\mathcal{H}_f(\Gamma^-)^{-1} T_\infty T_f^T \mathcal{O}_f \Sigma_n + (\mathcal{H}_f(\Gamma^-)^{-1} S_\infty \mathcal{H}_f^T)(\Gamma_f^+)^{-1} \mathcal{O}_f \Sigma_n^{-1}] = \\ &\Sigma_n^{-1} \mathcal{O}_f^T (\Gamma_f^+)^{-1} [\mathcal{O}_f K_0 E_0^{-1} C_0 \Sigma_n + \mathcal{O}_f (A_0 - K_0 E_0^{-1} C_0) \Sigma_n] = A_0 \Sigma_n \end{aligned}$$

since $\mathcal{H}_f(\Gamma^-) T_\infty = \mathcal{O}_f \mathcal{K}_1 = \mathcal{O}_f K_0 E_0^{-1}$, $T_f^T \mathcal{O}_f = C_0$, $\mathcal{H}_f(\Gamma^-)^{-1} S_\infty = \mathcal{O}_f (A_0 - K_0 E_0^{-1} C_0)\mathcal{K}$ and finally $\mathcal{K}\mathcal{H}_f^T(\Gamma_f^+)^{-1}\mathcal{O}_f = \Sigma_n^2$ and $\mathcal{O}_f^T(\Gamma_f^+)^{-1}\mathcal{O}_f = \Sigma_n$. Thus it follows that $A = A_0$.

Now (4.8) together with equation (4.5) imply $K = K_0$.

A CLT for $\hat{m}_p$

In this subsection we will use the results of the previous subsections to prove a central limit theorem for the vector $\hat{m}_p = \mathrm{vec}(\hat{\gamma}(0), \ldots, \hat{\gamma}(f-1), \hat{X}_{f,p}, \hat{Y}_{f,p}, \hat{Z}_{f,p})$. This will be done by linearizing the map ϕ attaching $\hat{m}_p$ to the sample covariances $\hat{g}_{T,f+p}$. It will be shown that $\hat{m}_p = m_p + L_p(\hat{g}_{T,p+f} - g_{p+f}) + o_P(T^{-1/2})$. In order to apply Lemma 4.1.3 it then remains to show, that the rows of L_p converge in the ℓ_1 norm to vectors with elements decreasing exponentially (comp. *REMARK* 4.1.2).

First the term $(\hat{W}_f^+)^{-1}\hat{X}_{f,p}(\hat{W}_f^+)^{-T} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T$ is considered. By linearizing this expression one obtains that

$$\begin{aligned} \sqrt{T}(\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T - \mathcal{H}_{f,p}(\Gamma_p^-)^{-1}\mathcal{H}_{f,p}^T) &= (\sqrt{T}(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p}))\beta_{f,p}^T - \\ &\quad \beta_{f,p}(\sqrt{T}(\hat{\Gamma}_p^- - \Gamma_p^-))\beta_{f,p}^T + \\ &\quad \beta_{f,p}(\sqrt{T}(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})^T) + \\ &\quad \text{high order terms} \end{aligned}$$

where $\beta_{f,p} = \mathcal{H}_{f,p}(\Gamma_p^-)^{-1}$. To prove that the high order terms are of order $o_P(1)$, the uniform convergence of the sample autocovariances will be used (comp. Theorem A.3.1): Under the upper bound on the increase of p, it follows that $\max_{|j| \le f+p-1} \|\hat{\gamma}(j) - \gamma(j)\| = O(Q_T)$ a.s., where $Q_T = \sqrt{\log\log T/T}$. This also implies that $\|\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p}\|_{Fr} = O(\sqrt{p}Q_T)$ a.s., $\|\hat{\Gamma}_p^- - \Gamma_p^-\|_{Fr} = O(pQ_T)$ a.s. In addition it can be shown that $\|(\Gamma_p^-)^{-1}\|_{Fr}$ and $\|(\hat{\Gamma}_p^-)^{-1}\|_{Fr}$ are of order $O(p)$ a.s. (see e.g. Peternell *et al.*, 1996). Using these bounds some simple but tedious calculations show that the high order terms are of order $\sqrt{T}p^k Q_T^j$ for some integer $k \le 6$, $j \ge 2$ and thus convergence to zero a.s. follows.

In order to apply Lemma 4.1.4 it remains to show that the rows of L_p corresponding to the term $\hat{X}_{f,p}$ converge in the ℓ_1 norm to vectors with elements decreasing exponentially. This follows immediately from the proof of Lemma 4.1.2 and the convergence of $\beta_{f,p}$.

The terms $\hat{Y}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1} S_p \hat{\mathcal{H}}_{f,p}^T$ and $\hat{Z}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1} T_p$ can be analyzed in a completely analogous manner, by showing that

$$\|\mathcal{H}_{f,p} S_p^T (\Gamma_p^-)^{-1} - (\mathcal{H}_f S_\infty^T (\Gamma^-)^{-1})_p\| = o(T^{-1/2})$$

and that $\mathcal{H}_f S_\infty^T (\Gamma^-)^{-1}$ has rows with exponentially decreasing entries, which follows from the same arguments as have been used in proving Lemma 4.1.3. Thus application of Lemma 4.1.4 shows the asymptotic normality of $\sqrt{T}(\hat{m}_p - m_0)$.

Differentiability of ψ

The last part of the proof consists in the proof of the differentiability of ψ at the value $m_0 = \text{vec}[\gamma(0), \ldots, \gamma(f), X_0, Y_0, Z_0]$. Thus it remains to show, that in the neighborhood of m_0, the approximation $\psi(m_0 + \delta m) = \psi(m_0) + J_\psi \delta m + o(\|\delta m\|)$ holds. For this purpose the essential steps of the computation of ψ are repeated.

For `CCA` a Cholesky factorization $(\Gamma_f^+)^{1/2}$ of Γ_f^+ and the inverse of Γ_f^+ and of $(\Gamma_f^+)^{1/2}$ have to be computed. Since Γ_f^+ is positive definite, these computations are differentiable in the entries of the corresponding matrices. For the Cholesky factorization this follows from the recursions defining the Cholesky factor (see e.g. Golub and VanLoan, 1989). Also the largest n eigenvalues Σ_n^2 and the corresponding eigenvectors U_n of the matrix X_0 have to be computed. The differentiability of the mapping attaching the eigenvectors and the eigenvalues to the matrix X_0 holds for $k_0 \in U_n^+$, see Lemma B.2.1 in Appendix B. The remaining steps are easily analyzed. One only has to keep in mind that Σ_n and $(\gamma(0) - C_0 \Sigma_n C_0^T)$ are positive definite and thus the Cholesky factorizations and the inverses, which have to be computed, are differentiable. This concludes the proof.

4.1.4 Implications of the result

In the last section, we derived the asymptotic behaviour of the estimates of the system matrices, when estimation is performed by a subspace procedure out of the class described in section 3.5. The result clarifies the asymptotic properties of the subspace estimates on the generic subset $U_n^+ \subset U_n^{(m)}$ (see section 4.1.1). The main result states under our standard assumptions on the noise sequence, that the subspace estimates are consistent (in probability) and that they fulfill a central limit theorem. From the proof it is clear, how to calculate the asymptotic variance of these estimates, compare *REMARK* 4.1.1 and *REMARK* 4.1.2. However unfortunately, the expressions are too complicated to serve as a tool for comparing the asymptotic variances for different choices of weighting matrices or truncation indices on an analytical basis.

The asymptotic variance can be written formally as $V_f^\theta = \lim_{p\to\infty} L_{f,p} V_{f,p}^g L_{f,p}^T$ as follows from section 4.1.3, where $L_{f,p} = J_\psi L_p$ (L_p is defined in the first paragraph of section 4.1.3 and J_ψ denotes the matrix of partial derivatives of the function ψ). Actual computations provide approximations to $V_f^{\theta_0}$, where the approximation error can be made arbitrarily small by choosing a and p suitably large. This approximation can be calculated along the following lines:

Recall, that the asymptotic variance V_h^g of $\sqrt{T}[\hat{g}_{T,h} - g_h]$ for fixed h (see Lemma 4.1.4) is the limit $\lim_{a\to\infty} L^{a,h} V_{a+h}^\varepsilon (L^{a,h})^T$, where the evaluation of V_{h+a}^ε is documented in Remark 4.1.1. Expressions for $L^{a,h}$ can be found in the proof of Lemma 4.1.4. The convergence of this expression is related to the magnitude of ρ_p^a (see the proof of Lemma 4.1.4). Corresponding to the asymptotic variance, V^m say, of $\sqrt{T}[\hat{m}_p - m_0]$ it has been shown, that $\sqrt{T}[\hat{m}_p - m_0] = \sqrt{T} L_p [\hat{g}_{T,p+f} - g_{p+f}] + o_P(1)$. Thus $V^m = \lim_{p\to\infty} L_p V_{f+p}^g (L_p)^T$, where the existence of the limit has been shown in section 4.1.3. L_p can be found from section 4.1.3. Lemma 4.1.5 shows, that the convergence of L_p and thus of V^m depends heavily on $|\rho_0^p|$. Finally ψ is a mapping between two finite dimensional vector spaces and thus the derivative of ψ can be calculated without any approximation using the results of Lemma B.2.1 and equations (4.2-4.9). Thus the approximation of V_f^θ can be found as $J_\psi L_p L^{a,f+p} V_{a+h}^\varepsilon (L^{a,f+p})^T L_p^T J_\psi^T$ by taking a and p large, where the meaning of large depends on the location of the systems zeros (for p) and poles

(for a).

On the set $U_n^{(m)} - U_n^+$ these results are not valid, due to the properties of the SVD: On this set, at least one singular value of $W_f^+ \mathcal{H}_f (\Gamma^-)^{-1/2}$ is of higher multiplicity and thus we cannot expect a consistency result for the system matrices (A, K, C, E). However, it is quite easy to see from an investigation of the proof, that the only problem consists in the choice of the singular vectors. Thus we are able to state a consistency result for the transfer function for the whole set $U_n^{(m)}$ as a corollary. This result has the same form as the consistency result obtained in (Peternell *et al.*, 1996) and thus is not new. Nevertheless it is included for the sake of a rather complete description of the asymptotic properties of the main algorithm.

Corollary 4.1.1 *Let $(y_t)_{t\in\mathbb{Z}}$ be generated by a true transfer function $k_0 \in U_n^{(m)}$, where the ergodic white noise $(\varepsilon_t)_{t\in\mathbb{Z}}$ fulfills the standard assumptions. Let $f \geq n$ be a fixed integer. If p fulfills the condition $p/(\log T)^a \to 0$ for some $a < \infty$, then there exists a realization (A_0, K_0, C_0, E_0) of the true transfer function k_0 and a sequence of orthonormal basis transformations S_T, such that*

$$\|vec(S_T \hat{A}_T S_T^{-1} - A_0, S_T \hat{K}_T - K_0, \hat{C}_T S_T^{-1} - C_0, \hat{E}_T - E_0)\| \to 0$$

PROOF: As can be seen from the proof of Theorem 4.1.1, $\hat{m}_p - m_0 = L_p(\hat{g}_{T,f+p} - g_{f+p}) + o_P(T^{-1/2})$, where the term $o_P(T^{-1/2})$ indeed is $o(T^{-1/2})$ a.s. Then $\|\hat{m}_p - m_0\| \to 0$ a.s. follows from convergence of the sample covariances (see Theorem A.3.1) and the bounds on the ℓ_1 norm of L_p. The result then is easy to see from Theorem B.2.1 and Theorem B.2.2. □

As another corollary, the central limit theorem for the matrix estimates also implies a central limit theorem for other quantities, which are derived from the system matrix estimates by differentiable mappings:

Corollary 4.1.2 *Let $g : \mathbb{R}^{n^2+2ns+(s^2+s)/2} \to \mathbb{R}^m$ be a mapping attaching the vector $x \in \mathbb{R}^m$ to the matrices (A, K, C, E) which is differentiable at (A_0, K_0, C_0, E_0). Let J_g denote the matrix of the partial derivatives with respect to the entries in (A, K, C, E). Then the following holds:*

$$\sqrt{T}(g(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T) - g(A_0, K_0, C_0, E_0)) \xrightarrow{d} Z \tag{4.14}$$

where Z is multivariate normally distributed with mean zero and variance $V = J_g V_f J_g^T \in \mathbb{R}^{m\times m}$.

In particular there are three applications of this corollary, we bear in mind:

- The poles of the system depend differentiable on the entries in the matrix A, if the eigenvalues are distinct (see Lemma B.2.1). Thus we obtain a CLT for the estimates of the system poles on the set $U_n^+ \cap U_n^d$, where U_n^d denotes the set of all transfer functions $k \in U_n^{(m)}$, which have distinct poles. It is easy to see, that U_n^d is generic in $U_n^{(m)}$ and thus also $U_n^+ \cap U_n^d$ is generic in $U_n^{(m)}$. The same statement is true for the estimates of the system zeros.
- For fixed frequency ω the transfer function evaluated at ω is equal to $k_0(e^{i\omega}) = E_0 + C_0(e^{i\omega} I - A_0)^{-1} K_0$. Thus we also obtain a central limit theorem for the estimates of the transfer function at fixed frequency points. This can be used, to compare different choices of procedures (i.e. different choices of f and of $\hat{W}_f^+$) for given system, as will be done in section 5.3
- The transformation of system matrices to echelon coordinates is differentiable for system matrices corresponding to a transfer function in the generic neighborhood of the echelon parametrization. Thus we get a CLT for the echelon parameter estimates (see section 2.2.1) on a generic neighborhood (since the intersection of two generic sets is still a generic set). This can also be used to compare different procedures (different weightings, different choices for f) corresponding to their asymptotic behaviour (for calculations related to this point comp. section 5.3).

4.2 The Case $f \to \infty, m = 0$

In this section, we will investigate the changes from f being fixed and finite to the case, where f is allowed to tend to infinity at a certain rate. From a theoretical point of view the approach, where f is allowed to increase has the advantage, that we do not have to invest a priori knowledge about the system order. On the other hand, there exist examples, where the asymptotic accuracy seems to decrease as f increases for a certain choice of weightings as well as examples, where the asymptotic accuracy seems to increase as f increases (comp. section 5.1). This motivates the inclusion of this section.

It will prove necessary to further constrain the set of feasible weighting matrices. Recall, the we already restricted the weighting matrix $\hat{W}_f^+$ to be either of the following two weightings: Either $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ (which corresponds to `CCA`) or $W_f^+ = W_f^+(k_W) = (K_W(i-j))_{i,j=1,\cdots,f}$, $K_W(j) = 0, j < 0$ is a lower triangular block Toeplitz matrix, where the blocks correspond to the rational transfer function $k_W(z) = K_W(0) + \sum_{j=1}^{\infty} K(j) z^{-j}$ with nonsingular constant term $K_W(0)$. In this case therefore

$$W_f^+(k_W) = \begin{bmatrix} K_W(0) & 0 & \cdots & 0 \\ K_W(1) & K_W(0) & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ K_W(f-1) & \cdots & K_W(1) & K_W(0) \end{bmatrix}$$

In addition we will now restrict k_W to be stable and strictly minimum-phase. The stability of the weighting transfer function will be of importance in the proof. The strict minimum-phase requirement is needed to ensure the invertibility of the matrix $W_f^+(k_W)$ for $f = \infty$. The rationality is merely a technical requirement, which does not seem to be a minimal requirement. However, the merit of seeking for more general results must be doubted. Typically the weighting transfer function will be chosen as some filter (see also chapter 5.3) with desired frequency characteristics and thus will be rational. There are also some results on the approximation properties of subspace algorithms using the rationality of k_W (see section 5.3).

In order to state the main result of this section, it is again necessary to define some subsets of $U_n^{(m)}$, the set of all stable, strictly minimum-phase rational transfer functions of order n. In the case $f \to \infty$ the a.s. limit of $\hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-T/2}$ will be shown to be equal to $W^+ \mathcal{H} (\Gamma^-)^{-T/2}$, where convergence is in operator norm. Thus we have to ensure, that the SVD of this matrix is essentially continuous, that is to ensure that the nonzero singular values of $W^+ \mathcal{H} (\Gamma^-)^{-T/2}$ are all distinct. Here W^+ denotes the infinite dimensional matrix obtained by letting f (and T) tend to infinity, Γ^- again denotes the covariance of $Y^-(t)$, $\mathcal{H}$ denotes the block Hankel-matrix of the covariance of $Y^+(t)$ and $Y^-(t)$ and $\Gamma^{1/2}$ denotes any square root of a positive definite matrix Γ, such that $(\Gamma)^{1/2}(\Gamma)^{T/2} = \Gamma$. Let U_n^+, using the same symbol in the `CCA` case, denote the set of all transfer functions $k \in U_n^{(m)}$, such that the nonzero singular values of $W^+ \mathcal{H} (\Gamma^-)^{-T/2}$are distinct. Consider the choice $\mathcal{O} = (W^+)^{-1} U_n \Sigma_n^{1/2}, \mathcal{K} = \Sigma_n^{1/2} V_n^T (\Gamma^-)^{-1/2}$. With this choice $\mathcal{O}^T (W^+)^T W^+ \mathcal{O} = \Sigma_n = \mathcal{K} \Gamma \mathcal{K}^T$ and thus for $W^+ = W^+(k_W)$, Σ_n contains the frequency weighted balanced singular values. Thus the singular values corresponding to k_0 and the weighting transfer function k_W are distinct, if $k_0 \in U_n^+$ as defined in section 2.2.4. For the `CCA` case, the relevant balancing is the minimum-phase balancing described in section 2.2.3. The realizations of the transfer function k_0 corresponding to this SVD will thus be either frequency weighted balanced or minimum-phase balanced (in the `CCA` case). In both cases, the evaluations in Chapter 2 show, that the set U_n^+ is open and dense in $U_n^{(m)}$. Also in both cases, there exists a choice of the orientation of the singular vectors, such that the SVD is essentially continuous in the neighborhood of $W^+ \mathcal{H} (\Gamma)^{-T/2}$. Let the corresponding realization be denoted with (A_0, K_0, C_0, E_0). The same remarks corresponding to the actual implimentation of the SVD as in the case f fixed apply also in this case: If the normalization is done, such that the resulting SVD is essentially continuous at the limit $W^+ \mathcal{H} (\Gamma)^{-T/2}$, then the results will hold, otherwise a new normalization has to be used in order to ensure the following results to hold.

Note, that U_n^+ and the realization (A_0, K_0, C_0, E_0) depend on the weighting scheme. Now we may formulate the main result of this section:

Theorem 4.2.1 *Let $(y_t)_{t\in\mathbb{Z}}$ be generated by a true transfer function $k_0 \in U_n^+$, where the ergodic white noise $(\varepsilon(t))_{t\in\mathbb{Z}}$ fulfills the standard assumptions.*

If f and p fulfill the following conditions:

1. $p \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ *for some* $d > 1$, *where* $\rho_0 = \lambda_{max}(A_0 - K_0 E_0^{-1} C_0)$,
2. $\max(f,p)/(\log T)^a \to 0$ *for some* $a < \infty$.
3. *If a weighting matrix* $W_f^+ = W_f^+(k_W)$ *is used,* $f \geq -\frac{d}{2}\frac{\log T}{\log|\rho_p|}, \forall T > T_0$ *for some* $d > 1$, *where* ρ_p *is a pole of* $k_0 + k_W$ *of maximum modulus*
4. *if* `CCA` *is used,* $f \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ *for some* $d > 1$

then

$$\sqrt{T}vec[\hat{A}_T - A_0, \hat{K}_T - K_0, \hat{C}_T - C_0, \hat{E}_T - E_0] \xrightarrow{d} Z$$

holds, where Z is a multivariate normal random variable with zero mean and variance V.

Again the asymptotic variance V depends on the noise distribution, in particular on the fourth moments of the noise $\varepsilon(t)$. However, as will be shown in section 5.5, the asymptotic distribution of the estimates of the transfer function $\bar{k}_0 = k_0 E_0^{-1}$ is independent of the noise distribution. Note, that there are two different requirements on the integer f, corresponding to which weighting is used. It should be noted, that the lower bound on the increase of f might not be sharp for all cases. However, since the increase is quite slow and there is still the problem of choosing the constant, this does not seem to be much of a problem for application. The choice of f might be critical for some choices of the weighting transfer function k_W (see the example given in section 5.1.1).

Also note, that in the `CCA` case, the choice $f = p$ leads to the validity of the argument. Since this lower bound can be estimated consistently in a certain sense (see section 5.1), we thus obtain a completely automatic procedure based on data only, the only choice of the user being to use `CCA` rather than any other subspace algorithm, if we also incorporate an estimation of the order in the algorithm. In section 5.2 we will discuss several approaches to estimation of the order.

4.2.1 The Structure of the Proof

The structure of the proof is completely analogous to the proof for fixed f, only the details change. As has been stated in section 4.1, the estimates of the system matrices are a nonlinear transformation of the covariance estimates $\hat{\gamma}(0), \cdots, \hat{\gamma}(f+p-1)$. Again we may consider equations (4.7,4.8,4.9) as the central part. Analogously to the proof for fixed f we may define the following five matrices:

$$\begin{aligned}
\hat{x}_1 &= T_f^T \hat{\mathcal{O}}_f \\
\hat{x}_2 &= T_p^T \hat{\mathcal{H}}_{f,p}^T (\hat{W}_f^+)^T \hat{W}_f^+ \hat{\mathcal{O}}_f \\
\hat{x}_3 &= \hat{\Sigma}_n \\
\hat{x}_4 &= \hat{\mathcal{O}}_f^T (\hat{W}_f^+)^T \hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-1} T_p \\
\hat{x}_5 &= \hat{\mathcal{O}}_f^T (\hat{W}_f^+)^T \hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-1} S_p \hat{\mathcal{H}}_{f,p}^T (\hat{W}_f^+)^T \hat{W}_f^+ \hat{\mathcal{O}}_f
\end{aligned} \tag{4.15}$$

Here we neglected the dependence of the quantities on the indices f and p for notational convenience.

Denoting $\hat{x}_6 = \hat{\gamma}(0)$, we can define $\hat{m}(f,p) = \text{vec}[\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_6]$. Note that for fixed f and p each matrix $\hat{x}_i$ is a function of finitely many covariance estimates. Thus there exist mappings ϕ_i attaching the estimates $\hat{x}_i$ to the vectorization of the estimated covariance sequence $\hat{g}_{T,f+p} = \text{vec}[\hat{\gamma}(0), \cdots, \hat{\gamma}(f+p-1)]$. Using these mappings we define $x_i(f,p) = \phi_i(g_{f+p})$. Let

$m(f,p) = \mathrm{vec}[x_1(f,p), \cdots, x_6(f,p)]$. It will be shown, that $x_{i,0} = \lim_{f,p\to\infty} x_i(f,p)$ exists under the assumptions of the theorem. Let $m_0 = \lim_{f,p\to\infty} m(f,p)$. Note, that the system matrix estimates are a nonlinear function ψ of $\hat{m}_{f,p}$, also note, that the dimension of $\hat{m}_{f,p}$ is fixed and finite. In analogy to the proof with fixed finite f we may thus decompose the proof into several parts. The outline of the rest of this section can be described as follows:

1. Prove, that $\sqrt{T}(\hat{g}_{T,f+p} - g_{f+p})$ is asymptotically Gaussian distributed in the sense of Lemma 4.1.4 for f and p suitable functions of the sample size T.
2. Show that, $\sqrt{T}(m(f,p) - m_0) \to 0$ for $f,p \to \infty$ as a function of the sample size at the rate given in the theorem.
3. Show, that $\sqrt{T}(\hat{m}(f,p) - m(f,p))$ is asymptotically Gaussian distributed.
4. Verify, that ψ is differentiable at m_0.

The first item is identical with Lemma 4.1.4. we will thus start with proving point number two.

4.2.2 The details of the proof

The proof will be given seperately for the two different choices of the weighting matrix. First we deal with the case $W_f^+ = W_f^+(k_W)$. The proof for `CCA` is postponed to the end of the section.

The first part of this section will be devoted to the proof, that $\sqrt{T}(m(f,p) - m_0) \to 0$ for $f,p \to \infty$ at the rate given in the theorem. The proof will be given in two steps, which form a natural decomposition of the problem: In the algorithm we use the singular value decomposition of $W_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2} \in \mathbb{R}^{fs\times ps}$, which can be calculated from the eigenvaluedecomposition of $W_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T(W_f^+)^T$. Thus we will in a first step investigate the convergence properties of

$$X_{f,p} = W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}\mathcal{H}_{f,p}^T(W_f^+)^T$$

and then use these results to deal with the convergence of $\sqrt{T}(m(f,p) - m_0)$.

Since the dimension of $X_{f,p}$ increases with increasing f, we first have to clarify the setting, in which convergence may take place: Consider $X_{f,p}$ as a mapping acting on ℓ^2, the vector space of square summable sequences, acting on the subspace $\mathcal{X}_f = \{(x_1, x_2, \cdots), x_i = 0, i > fs\}$. Then the matrix description of the mapping on ℓ^2 defined by the matrix $X_{f,p}$ contains the matrix $X_{f,p}$ as its $fs \times fs$ heading submatrix, the remaining elements being zero. By this construction we also obtain a natural framework for the analysis, in particular the notion of convergence is well defined in this setting. Let $X_0 = W^+\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(W^+)^T$, using the same notation as in the previous section. Due to the stability of k_0 and k_W and the strict minimum-phase condition all occuring matrices have bounded norm, and thus the resulting operator is a self adjoint, bounded operator. Due to the finite rank of $\mathcal{H}$ the operator moreover is compact. Clearly the operator $X_{f,p}$ also is compact. Thus convergence of the sequence of operators $X_{f,p}$ for $f,p \to \infty$ suitably can be proved by showing, that the spectral radius of the difference converges to zero. As a first step, recall that $\mathcal{H} = \mathcal{O}\bar{\mathcal{C}}$ and $\mathcal{H}_{f,p} = \mathcal{O}_f\bar{\mathcal{C}}_p$. Using the same argument as in Lemma 4.1.5 we obtain $\|\bar{\mathcal{C}}_p(\Gamma_p^-)^{-1}\bar{\mathcal{C}}_p^T - \bar{\mathcal{C}}(\Gamma^-)^{-1}\bar{\mathcal{C}}^T\|_{Fr} = O(|\rho_0^{'}|^p), \forall |\rho_0| < |\rho_0^{'}| < 1$. Here as usual ρ_0 denotes a zero of k_0 of maximum modulus. Clearly

$$\begin{aligned} X_{f,p} - X_0 \quad = \quad & (W_f^+\mathcal{O}_f - W^+\mathcal{O})\bar{\mathcal{C}}_p(\Gamma_p^-)^{-1}\bar{\mathcal{C}}_p^T\mathcal{O}_f^T(W_f^+)^T + \\ & W^+\mathcal{O}[\bar{\mathcal{C}}_p(\Gamma_p^-)^{-1}\bar{\mathcal{C}}_p^T - \bar{\mathcal{C}}(\Gamma^-)^{-1}\bar{\mathcal{C}}^T]\mathcal{O}_f^T(W_f^+)^T + \\ & W^+\mathcal{O}\bar{\mathcal{C}}(\Gamma^-)^{-1}\bar{\mathcal{C}}^T(W_f^+\mathcal{O}_f - W^+\mathcal{O})^T \end{aligned} \tag{4.16}$$

where finite dimensional matrices are always interpreted as operators defined on ℓ^2 by adding zeros as shown above for $X_{f,p}$. The arguments given above show the convergence of the second term independent of f. The desired result now follows from the convergence of $(W_f^+\mathcal{O}_f - W^+\mathcal{O})$, since this term governs the norm of the first and the third term. It is easy to see, that the lower triangularity of W_f^+ as well as W^+ implies that the first f blocks of $(W_f^+\mathcal{O}_f - W^+\mathcal{O})$ are indeed equal

to zero. Thus the norm of the difference is determined by the modulus of the elements with index greater than fs in $W^+\mathcal{O}$. These elements are the entries of $\sum_{j=0}^{i} K_W(j)CA^{i-j}, i > f$. Due to the assumed stability of both, k_0 and k_W, we obtain $\|CA^i\| \leq C|\rho_{p,0}^{'}|^i, \forall|\rho_{p,0}| < |\rho_{p,0}^{'}| < 1$, where $\rho_{p,0} = \lambda_{max}(A)$ and the same is true for $K_W(i)$ using $\rho_{p,k_W}^{'}, |\rho_{0,k_W}| < |\rho_{0,k_W}^{'}| < 1$, where ρ_{0,k_W} denotes a pole ρ_{p,k_W} of maximum modulus. Thus $\|\sum_{j=0}^{i} K_W(j)CA^{i-j}\| \leq \sum_{j=0}^{i} \|K_W(j)\|\|CA^{i-j}\| \leq Ci(\max\{|\rho_{p,0}^{'}|, |\rho_{p,k_W}^{'}|\})^i$. The norm of the convolution is thus governed by the maximum of these two moduli and $\sqrt{T}Ci(\max\{|\rho_{p,0}^{'}|, |\rho_{p,k_W}^{'}|\})^i \to 0$ under the condition of the Theorem. Then $\rho_p = \max\{|\rho_{p,0}^{'}|, |\rho_{p,k_W}^{'}|\}$. Note, however, that this is no sharp bound. It seems possible, that in some cases the bound will be very unrealistic. However the importance of obtaining tight bound for practical situations might be doubted.

Thus we obtain the desired rate of convergence of $X_{f,p}$ to X_0 in operator norm, i.e. $\|X_{f,p} - X_0\| = o(T^{-1/2})$. Using the assumption $k_0 \in U_n^+$ we obtain the differentiability of the singular values and of the singular vectors contained in U_n (see Lemma B.2.1 in Appendix B), since in this case the singular values of X_0 will be distinct.

Now we are ready to prove the order of convergence of $\|x_i(f,p) - x_{i,0}\|$: First consider $x_3(f,p) = \Sigma(f,p)$ (where the notation makes the dependence of $\Sigma(f,p)$ on the truncation indices explicit). The error $x_3(f,p) - x_{3,0}$ is of order $o(T^{-1/2})$ under the conditions of Theorem 4.2.1, as follows from the properties of the singular values derived in Lemma B.2.1. Next $x_1(f,p) = T_f^T \mathcal{O}_f(f,p)$, where again the dependence of $\mathcal{O}_f(f,p) = (W_f^+)^{-1}U_n(f,p)(x_3(f,p))^{1/2}$ on f and p has been stressed. Now the convergence of $x_3(f,p)$ and of $U_n(f,p)$ (apply again Lemma B.2.1) and the lower triangularity of $(W_f^+)^{-1}$ show the desired result. The strict minimum-phase assumption on k_W ensures the boundedness of $(W_f^+)^{-1}$ independent of f. The exponential decrease of $\mathcal{O}$ and the results shown above imply

$$\|\mathcal{O}_f^T(W_f^+)^T W_f^+ \mathcal{O}_f - \mathcal{O}^T(W^+)^T W^+ \mathcal{O}\| = o(T^{-1/2})$$

since

$$\begin{bmatrix} W_f^+ \mathcal{O}_f \\ 0 \end{bmatrix} - W^+\mathcal{O} = \begin{bmatrix} U_n(f,p) \\ 0 \end{bmatrix} \Sigma_n(f,p)^{1/2} - U_n \Sigma_n^{1/2}$$

Together with $\|[\bar{\mathcal{C}}_p(\Gamma_p^-)^{-1}, 0] - \bar{\mathcal{C}}(\Gamma^-)^{-1}\| = o(|\rho_o^{'}|^p)$ this proves the desired result for the remaining terms $x_i(f,p)$ and thus $\sqrt{T}(m(f,p) - m_0) \to 0$.

In the next step we deal with the term $\sqrt{T}(\hat{m}(f,p) - m(f,p))$. Again the central matrix will be $X_{f,p}$, which is estimated by $\hat{X}_{f,p}$. The norm of the difference can be bounded by

$$\begin{aligned} \|\sqrt{T}(\hat{X}_{f,p} - X_{f,p})\|_{Fr} \quad \leq \quad & 2\sqrt{T}\|W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})\beta_{f,p}^T(W_f^+)^T\|_{Fr} + \\ & \sqrt{T}\|W_f^+\beta_{f,p}(\hat{\Gamma}_p^- - \Gamma_p^-)\beta_{f,p}^T(W_f^+)^T\|_{Fr} + o_P(1) \end{aligned} \tag{4.17}$$

where again $\beta_{f,p} = \mathcal{H}_{f,p}(\Gamma_p^-)^{-1}$. Here the $o_P(1)$ term is due to neglecting the higher order terms, which again can be shown to be of the required order using the uniform convergence of the sample covariances stated in Theorem A.3.1. Thus $\sqrt{T}(\hat{X}_{f,p} - X_{f,p})$ essentially is the sum of the three terms: $W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})\beta_{f,p}^T(W_f^+)^T$, the transpose of this term and $W_f^+\beta_{f,p}(\Gamma_p^- - \hat{\Gamma}_p^-)\beta_{f,p}^T(W_f^+)^T$. Again we will first center on the properties of the estimates for the singular values and the singular vectors and afterwards deal with the asymptotic distribution of $\sqrt{T}(\hat{m}(f,p) - m(f,p))$.

Corresponding to the singular values, note that $\hat{\sigma}_i(f,p)$ is the square root of $\hat{\lambda}_i(f,p)$, the i-th eigenvalues of $\hat{X}_{f,p}$ and the same is true for the singular value $\sigma_i(f,p)$ and the eigenvalue $\lambda_i(f,p) = \sigma_i(f,p)^2$ of $X_{f,p}$. Equation (B.3) contains the eigenvector $u_i(f,p)$ of $X_{f,p}$, which is equal to $W_f^+(\mathcal{O}_f)_i/\sigma_i^{1/2}$, where $(\mathcal{O}_f)_i$ denotes the i-th column of $\mathcal{O}_f$. Thus

$$\begin{aligned} \sqrt{T}(\hat{\lambda}_i(f,p) - \lambda_i(f,p)) \quad \doteq \quad & 2\sqrt{T}u_i(f,p)^T W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})\mathcal{K}_p^T\mathcal{O}_f^T(W_f^+)^T u_i(f,p) \\ + \quad & \sqrt{T}u_i(f,p)^T W_f^+ \mathcal{O}_f \mathcal{K}_p(\Gamma_p^- - \hat{\Gamma}_p^-)\mathcal{K}_p^T\mathcal{O}_f^T(W_f^+)^T u_i(f,p) \end{aligned}$$

where again $\doteq$ denotes equality up to terms of order $o_P(1)$. Note, that in both matrices $\sqrt{T}(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})$ and $\sqrt{T}(\hat{\Gamma}_p^- -\Gamma_p^-)$ the entries are elements of $\sqrt{T}(\hat{\gamma}(j)-\gamma(j)), j=0,\cdots,f+p-1$. Thus the asymptotic normality of $\sqrt{T}(\hat{\lambda}_i(f,p)-\lambda_i(f,p))$ is proved, if we find a vector $l(f,p)$, such that $\sqrt{T}(\hat{\lambda}_i(f,p)-\lambda_i(f,p)) = l(f,p)^T\sqrt{T}(\hat{g}_{T,f+p}-g_{f+p})$ and $l(f,p)$ converges to l in the ℓ_1 norm, where l has elements decreasing sufficiently fast. Now note, that $(W_f^+)^T u_i(f,p)$ converges to $(W^+)^T u_i$, where u_i denotes the i-th eigenvector of X_0, which has elements decreasing exponentially. The convergence follows from the convergence of $X_{f,p}$ to X_0 and from the upper triangularity of $(W^+)^T$. It is also easy to see, that the entries of $(W^+)^T u_i = (W^+)^T W^+(\mathcal{O})_i \sigma_i^{-1/2}$, where $(\mathcal{O})_i$ denotes the i-th column of $\mathcal{O}$, decrease exponentially and that $\mathcal{O}_f^T(W_f^+)^T u_i(f,p)$ converges to $\mathcal{O}^T(W^+)^T u_i$. In all the evaluations above, convergence means convergence in ℓ_1. Finally the first term can be written as

$$2\mathrm{vec}[\sqrt{T}\overline{u_i}^T(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})\mathcal{K}_p^T\mathcal{O}_f^T\overline{u_i}] = 2(\overline{u_i}^T\mathcal{O}_f\mathcal{K}_p\otimes\overline{u_i}^T)\sqrt{T}\mathrm{vec}[\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p}]$$

where $\otimes$ denotes the Kroneckerproduct and where $\overline{u_i} = (W_f^+)^T u_i(f,p)$. Both vectors in this Kroneckerproduct converge in the ℓ_1 norm to vectors with elements decreasing exponentially. The block Hankel structure of $\sqrt{T}(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})$ then implies the convergence of the vector $l(f,p)$ due to the first term (see Lemma 4.1.2). The second term has as the essential part $\mathcal{K}_p\sqrt{T}(\hat{\Gamma}_p^- - \Gamma_p^-)\mathcal{K}_p^T$. This term has the same form as the expressions analyzed in Lemma 4.1.2, and therefore the techniques used in the proof of Lemma 4.1.4 imply the asymptotic normality of the eigenvalues and thus the asymptotic normality of $\sqrt{T}(\hat{x}_3(f,p)-x_3)$.

Recall the second equation in (B.3):

$$\hat{u}_i(f,p) = u_i + \sum_{j=1,j\neq i}^{n}\frac{u_j^T(\hat{X}_{f,p}-X_0)u_i}{\lambda_i-\lambda_j} + \frac{1}{\lambda_i}P_0(\hat{X}_{f,p}-X_0)u_i + o_P(T^{-1/2})$$

Here P_0 denotes the projection onto the orthogonal complement of $span\{u_i : i=1,\cdots,n\}$. The contribution due to the sum on the right hand side of this equation can be examined using the same arguments as given above. Thus it remains to investigate the term including P_0. Since

$$\begin{aligned}\hat{X}_{f,p}-X_{f,p} &= W_f^+(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})\mathcal{K}_p^T\mathcal{O}_f^T(W_f^+)^T + W_f^+\mathcal{O}_f\mathcal{K}_p(\Gamma_p^- - \hat{\Gamma}_p^-)\mathcal{K}_p^T\mathcal{O}_f^T(W_f^+)^T + \\ &\quad W_f^+\mathcal{O}_f\mathcal{K}_p(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})^T(W_f^+)^T + \text{higher order terms}\end{aligned}$$

the term involving P_0 only matters for $W_f^+(\hat{\mathcal{H}}_{f,p}-\mathcal{H}_{f,p})\mathcal{K}_p^T\mathcal{O}_f^T(W_f^+)^T$ and is zero else, since $W_f^+\mathcal{O}_f \to W^+\mathcal{O} \in span\{u_1,\cdots,u_n\}$. Now for $\hat{x}_1$, this expression is premultiplied by T_f^T, which is the matrix containing the transpose of the first n vectors of the canonical basis. Consider e.g. e_1, the first vector of the canonical basis. Then the contribution due to $e_1^T P_0(\hat{X}_{f,p}-X_0)u_i$ contains the vector P_0e_1, which is of finite ℓ_1 norm, since $\|e_1\|_1 = 1$ and $\|(I-P_0)e_1\|_1 < c < \infty$, which follows from the exponential decrease of the elements of U_n. Then convergence in the ℓ_1 norm to a vector of bounded ℓ_1 norm is not hard to obtain for the vectorization of this term and the same is true for the exponential decrease. The convergence of the remaining terms follows from the same arguments as given above. This finally shows the asymptotic normality of $\sqrt{T}(\hat{x}_1(f,p)-x_{1,0})$.

For the remaining 3 expressions x_2, x_4 and x_5 a reuse of the arguments involved in the proof above shows, that

$$\sqrt{T}(\hat{m}(f,p)-m_0) \doteq L(f,p)\mathrm{vec}[\hat{\gamma}(0)-\gamma(0),\cdots,\hat{\gamma}(f+p-1)-\gamma(f+p-1)]$$

and additionally the fact, that the rows of $L(f,p)$ converge in the ℓ_1 norm to the rows of the operator L, say, which have bounded ℓ_1 norm. Also the exponential decrease of these rows is straightforward to derive. Therefore Lemma 4.1.4 together with the Cramer-Wold device (see e.g. Anderson, 1971, Theorem 7.7.7, comp. also *REMARK* 4.1.2) show the asymptotic normality of $\sqrt{T}(\hat{m}(f,p)-m_0)$. The verification of the last point of the proof (the differentiability of

ψ) is straightforward and hence omitted. Thus we have shown the result for the case, that $W_f^+ = W_f^+(k_W)$.

In the case of `CCA` the weighting matrix $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ is used. This leads to two changes in the argumentation: First the weighting matrix is no longer fixed, but depends on the data, and secondly it is no longer true, that W_f^+ is the heading submatrix of W^+ and lower triangular. However, the arguments will be very similar to the ones given in the case $W_f^+ = W_f^+(k_W)$ and thus we will only comment on the changes in the proof, when dealing with `CCA` in the following.

Also in this case, we will augment the finite dimensional matrices to operators on ℓ^2 by adding zeros. Then equation (4.16) still holds. The proof, that the second term in (4.16) is of order $o(|\rho_0^{'}|^p)$ is unchanged. However, the proof for the two other terms is quite involved: We want to prove, that $\begin{bmatrix} (\Gamma_f^+)^{-1/2}\mathcal{O}_f \\ 0 \end{bmatrix} - (\Gamma^+)^{-1/2}\mathcal{O}$ is of order $o(|\rho_0^{'}|^f), \forall |\rho_0| < |\rho_0^{'}| < 1$. Therefore $p \geq -d \log T/(2 \log |\rho_0|), d > 1$ implies $\sqrt{T}|\rho_0'|^p \to 0$. Here we will need the results of Lemma B.2.2. Note, that the eigenvalues of Γ_f^+ can be bounded uniformely in f by two constants $c_1 > 0$ and $c_2 < \infty$ (see e.g. Hannan and Deistler, 1988) such that the spectrum of $(\Gamma_f^+)^{-1}$ is contained in the interval (c_1, c_2), for all $f \geq f_0$. Also c_1 and c_2 can be chosen, such that the spectrum of the operator $(\Gamma^+)^{-1}$ is contained in the same interval under our assumptions (see e.g. Hannan and Deistler, 1988). Now

$$\begin{bmatrix} (\Gamma_f^+)^{-1/2}\mathcal{O}_f \\ 0 \end{bmatrix} = \begin{bmatrix} (\Gamma_f^+)^{-1/2} & 0 \\ 0 & (\Gamma^+)^{-1/2} \end{bmatrix} \begin{bmatrix} \mathcal{O}_f \\ 0 \end{bmatrix}$$

Thus the spectrum of all matrices, of which we need the square root lie in an interval (c_1, c_2). Thus we may use the Dunford integral (see equation (B.5)) to define the inverse square root. For example $(\Gamma^+)^{1/2} = ((\Gamma^+)^{-1})^{-1/2}$ can be defined as

$$(\Gamma^+)^{1/2} = \sum_{i=0}^{\infty} a_i [(\Gamma^+)^{-1} - cI]^i$$

where $c = (c_1 + c_2)/2$ and the power series $\sum_{j=0}^{\infty} a_i (z-c)^i = z^{-1/2}$ has a convergence region containing the closed disc (in the complex plane) with center c and radius $c - c_1$. The fact, that the spectrum of $(\Gamma^+)^{-1} - cI$ is contained in the interval $(c_1 - c, c - c_1)$ thus shows the convergence of the power series.

Let us introduce some more notation:

$$\begin{aligned} G_f(c) &= \begin{bmatrix} (\Gamma_f^+)^{-1} - cI & 0 \\ 0 & (\Gamma^+)^{-1} - cI \end{bmatrix} \\ G(c) &= [(\Gamma^+)^{-1} - cI] \\ \bar{O}_f &= \begin{bmatrix} (\Gamma_f^+)^{-1}\mathcal{O}_f \\ 0 \end{bmatrix} \\ \bar{O} &= (\Gamma^+)^{-1}\mathcal{O} \\ \Gamma^+ &= \begin{bmatrix} \Gamma_f^+ & \tilde{\mathcal{H}}_f^T \\ \tilde{\mathcal{H}}_f & \Gamma^+ \end{bmatrix} \end{aligned}$$

where $\tilde{\mathcal{H}}_f = \mathcal{O}\tilde{\mathcal{C}}_f$. In the sequel we will need the following equations:

$$\begin{aligned} (\Gamma^+)^{-1} &= \begin{bmatrix} (\Gamma_f^+)^{-1} & 0 \\ 0 & 0 \end{bmatrix} \\ &\quad + \begin{bmatrix} -(\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T\mathcal{O}^T \\ I \end{bmatrix} (\Gamma^+ - \mathcal{O}\tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T\mathcal{O}^T)^{-1}[-\mathcal{O}\tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1}, I] \qquad (4.18) \\ &= R^{-1} - R^{-1}\mathcal{O}(\Sigma^{-1} + \mathcal{O}^T R^{-1}\mathcal{O})^{-1}\mathcal{O}^T R^{-1} \qquad (4.19) \end{aligned}$$

Here $R = \Gamma^+ - \mathcal{O}\Sigma\mathcal{O}^T$. Note, that $\tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T = \bar{\mathcal{C}}_f(\Gamma_f^-)^{-1}\bar{\mathcal{C}}_f^T = \Sigma + O(|\rho_0'|^f)$, which can be shown, using the techniques of the proof of Lemma 4.1.5. Thus $\Gamma^+ - \mathcal{O}\tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T\mathcal{O}^T = R + O(|\rho_0'|^f)$. Using these facts, we obtain:

$$G_f(c) - G(c) \doteq \begin{bmatrix} -(\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T x_R \tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1} & (\Gamma_f^+)^{-1}\tilde{\mathcal{C}}_f^T\mathcal{O}^T R^{-1} \\ R^{-1}\mathcal{O}\tilde{\mathcal{C}}_f(\Gamma_f^+)^{-1} & -R^{-1}\mathcal{O}(\Sigma^{-1} + x_R)^{-1}\mathcal{O}^T R^{-1} \end{bmatrix} \tag{4.20}$$

where $\doteq$ denotes equality up to terms of order $O(|\rho_0'|^f)$ and where $x_R = \mathcal{O}^T R^{-1}\mathcal{O}$. Using

$$(\Gamma_f^+)^{-1/2}\mathcal{O}_f = [(\Gamma_f^+)^{-1}]^{-1/2}(\Gamma_f^+)^{-1}\mathcal{O}_f, (\Gamma^+)^{-1/2}\mathcal{O} = [(\Gamma^+)^{-1}]^{-1/2}(\Gamma^+)^{-1}\mathcal{O}$$

we obtain

$$\begin{bmatrix} (\Gamma_f^+)^{-1/2}\mathcal{O}_f \\ 0 \end{bmatrix} - (\Gamma^+)^{-1/2}\mathcal{O} \;=\; \sum_{i=0}^{\infty} a_i \left(G_f(c)^i\bar{O}_f - G(c)^i\bar{O}\right)$$

Thus the essential terms are $G_f(c)^i\bar{O}_f - G(c)^i\bar{O} = G_f(c)^i(\bar{O}_f - \bar{O}) + (G_f(c)^i - G(c)^i)\bar{O}$. Note, that the norm of $G_f(c)^i$ is of order $O[(c-c_1)^i]$. Using the techniques of the proof for Lemma 4.1.5 it is straightforward to show $\|\bar{O}_f - \bar{O}\| = O(|\rho_0'|^f)$. Therefore the contribution due to the term $G_f(c)^i(\bar{O}_f - \bar{O})$ can be bounded by

$$(\sum_{i=0}^{\infty} a_i(c-c_1)^i)O(|\rho_0^{'}|^f)$$

since the sum is bounded due to the absolute convergence of the power series.

For the remaining terms it will prove to be convenient to use the following bound:

Lemma 4.2.1 *Let X be a vector with the property, that for its j-th (block) entry, X_j say, $\|X_j\|_1 \leq c_X|\rho|^j$ holds for some $|\rho_0'| < |\rho| < 1$. Then:*

$$\|[G_f(c)^j - G(c)^j]X\|_1 \leq C^0|\rho|^f j(c-c_1)^{j-1} \tag{4.21}$$

where C^0 is independent of i, but depends on c_X and $|\rho|$.

PROOF: The proof of this statement is inductive: First let $j = 1$. Thus we consider $(G_f(c) - G(c))X$. It is easy to see from equation (4.20), that the vector X may be split into a first component, $X_{1:f}$ say, containing the first fs entries and the remaining part, which has 1-norm bounded by $c_X'|\rho|^f$, where c_X' depends on c_X and on $|\rho|$. Thus we only have to deal with the first part, whose contribution can be easily seen to be bounded by $C_1\|\tilde{\mathcal{C}}_f^T(\Gamma_f^+)^{-1}X_{1:f}\|_1$, where C_1 depends only on the true system and not on X or f. Due to

$$\|\tilde{C}_f(\Gamma_f^+)^{-1} - [(A - KE^{-1}C)^{f-1}K, (A - KE^{-1}C)^{f-2}K, \cdots, K]\|_1 = O(|\rho_0'|^f)$$

and $\|(A - KE^{-1}C)^jK\|_1 \leq C_k|\rho_0^{'}|^j$ for some constant C_k we obtain that $\|\tilde{\mathcal{C}}_f^T(\Gamma_f^+)^{-1}X_{1:f}\|_1 \leq C_2|\rho|^f$, where C_2 depends on X, but is independent of f. This shows $\|[G_f(c) - G(c)]X\|_1 \leq C^0|\rho|^f$ for $j = 1$, where C^0 does not depend on f.

Thus assume, that equation (4.21) holds up to j. Then

$$[G_f(c)^{j+1} - G(c)^{j+1}]X = G_f(c)^j[G_f(c) - G(c)]X + [G_f(c)^j - G(c)^j]G(c)X$$

The ℓ^1-norm of the first term can be shown to be bounded by $(c-c_1)^jC^0|\rho|^f$ using the arguments above and the bound on the spectrum of $G_f(c)$. It thus remains to show, that $Y = G(c)X$ has the property, that the j-th (block) entry, Y_j say, can be bounded by $\|Y_j\|_1 \leq c_Y|\rho|^j$. This will be shown next. Note, that

$$G(c)X = [(\Gamma^+)^{-1} - cI]X = [R^{-1} - R^{-1}\mathcal{O}(\Sigma^{-1} - x_R)^{-1}\mathcal{O}^T R^{-1} - cI]X$$

Thus

$$\|Y_j\|_1 \leq \|(R^{-1}X)_j\|_1 + c\|X_j\|_1 + \|(R^{-1}\mathcal{O})_j\|_1 \|(\Sigma^{-1} - x_R)^{-1}\mathcal{O}^T R^{-1}X\|_1$$

where $(.)_j$ denotes the j-th block row of the corresponding expression. The second term in this expression is bounded by $cc_X|\rho|^j$ by the assumption on X. Corresponding to the remaining terms, we note, that $R^{-1} = \mathcal{E}^{-T}\mathcal{E}^{-1}$, where $\mathcal{E}^{-1}$ is the block Toeplitz lower triangular matrix, containing the power series coefficients of the transfer function k_0^{-1} as its blocks. Thus the norm of these blocks decreases exponentially with the distance to the main diagonal. Therefore Lemma 4.1.3 shows, that $\|(\mathcal{E}^{-1}X)_j\|_1 \leq c_k|\rho|^j$ for some constant c_k depending on X and k_0. Due to the structure of $\mathcal{E}^{-T}$ this implies, again using Lemma 4.1.3, that $\|(R^{-1}X)_j\|_1 \leq C_3|\rho|^j$. For the third term finally note, that $\|(\Sigma^{-1} - x_R)^{-1}\mathcal{O}^T R^{-1}X\|_1$ is bounded and independent of j and f. $R^{-1}\mathcal{O} = \mathcal{E}^{-T}[\mathcal{E}^{-1}\mathcal{O}] = \mathcal{E}^{-T}[C^T E^{-T}, (A - KE^{-1}C)^T C^T E^{-T}, \cdots]^T$ and thus $\|(R^{-1}\mathcal{O})_j\|_1 \leq C_4|\rho|^j$ follows. This shows, that $Y = G(c)X$ can be used in equation (4.21) with the same ρ as X and a constant c_Y, i.e. $\|[G_f(c)^j - G(c)^j]Y\|_1 \leq C^1(c - c_1)^{j-1}j|\rho|^f$. Taking the maximum of the finitely many constants, which all are independent of j and f, it follows, that equation (4.21) holds for $j \in \mathbb{N}$. □

It is easily verified using Lemma 4.1.3, that $X = \bar{O}$ fulfills the assumption $\|X_j\|_2 \leq c_X|\rho_0'|^j$ and thus $\|[G_f(c)^i - G(c)^i]\bar{O}\|_1 \leq C^0(c - c_1)^{i-1}i|\rho_0'|^f$ holds. Concluding this results in the following bound:

$$\|\sum_{j=0}^{\infty} a_i(G_f(c)^i\bar{O}_f - G(c)^i\bar{O})\| \leq C(\sum_{j=0}^{\infty} a_i(c - c_1)^i + a_i i(c - c_1)^{i-1})|\rho_0'|^f \leq C'|\rho_0'|^f$$

where the last equality follows from the fact, that the sum is the power series expansion of the inverse square root and therefore absolutely convergent on the disc with center 0 and radius $(c - c_1)$, which implies, that $\sum ia_i z^{i-1}$ is also absolutely convergent with convergence region including the disc with radius $(c - c_1)$. Thus we have shown, that $\|X_{f,p} - X_0\| = O(T^{-1/2})$.

Following the proof for general weighting matrices, the next step consists in using the result derived above for proving $\sqrt{T}(m(f,p) - m_0) \to 0$. Again for $x_3(f,p)$ this follows immediatly from the differentiability of the singular values (see Lemma B.2.1). From Lemma B.2.1 also the convergence of the singular vectors follows: $\sqrt{T}\|U_{n,f} - U_n\| = o(1)$. From the definitions in equation (4.15), it is straightforward, that the convergence of $U_{n,f}$ and $(\Gamma_f^+)^{-1/2}\mathcal{O}_f$ implies the convergence of the terms $x_2(f,p), x_4(f,p)$ and $x_5(f,p)$. Finally for $x_1(f,p) = T_f^T\mathcal{O}_f = T_f^T(\Gamma_f^+)^{1/2}U_{n,f}$ note, that we can use completely the same arguments as given above, by replacing $\bar{O}_f$ with T_f and $\bar{O}$ with T_∞ to show that $(\Gamma_f^+)^{1/2}T_f \to (\Gamma^+)^{1/2}T_\infty$. Finally the power series expansion of $(\Gamma^+)^{1/2}$ also shows the exponential decrease of $(\Gamma^+)^{1/2}T_\infty$. Thus we have shown $\sqrt{T}(m(f,p) - m_0) \to 0$ under the assumptions on the increase of f and p given in Theorem 4.2.1.

The next step is to examine $\sqrt{T}(\hat{X}_{f,p} - X_{f,p})$. In the case of `CCA` in equation (4.17) two more terms have to be considered, which are due to $\hat{W}_f^+ - W_f^+$. It follows from Theorem A.3.1 that $\|\hat{W}_f^+ - W_f^+\|_{Fr} = O(Q_T f)$ a.s. (see also below), where $Q_T = (\log\log T/T)^{1/2}$, and therefore in the case of `CCA` we obtain:

$$\begin{aligned}\|(\hat{X}_{f,p} - X_{f,p})\| \leq\ & 2\|W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})(\Gamma_p^-)^{-1}(\mathcal{H}_{f,p})^T(W_f^+)^T\| + \\ & \|W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}(\hat{\Gamma}_p^- - \Gamma_p^-)(\Gamma_p^-)^{-1}\mathcal{H}_{f,p}^T(W_f^+)^T\| + \\ & 2\|(\hat{W}_f^+ - W_f^+)\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}\mathcal{H}_{f,p}^T(W_f^+)^T\| + o_P(T^{-1/2})\end{aligned} \tag{4.22}$$

Thus in this case, $\sqrt{T}(\hat{X}_{f,p} - X_{f,p})$ is equal to (up to order $o_P(1)$) the sum of five terms: $(\hat{W}_f^+ - W_f^+)\mathcal{H}_{f,p}\beta_{f,p}^T(W_f^+)^T$, its transpose, $W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})\beta_{f,p}^T(W_f^+)^T$, its transpose and finally

$-W_f^+\beta_{f,p}(\hat{\Gamma}_p^- - \Gamma_p^-)\beta_{f,p}^T(W_f^+)^T$, where again $\beta_{f,p} = \mathcal{H}_{f,p}(\Gamma_p^-)^{-1}$. Let $u_i(f,p)$ denote the i-th eigenvector of $X_{f,p}$ and let $\lambda_i(f,p)$ denote the corresponding eigenvalue. Then it follows from the rate of convergence of $X_{f,p} - X_0$, that the $u_i(f,p)$ converge in ℓ_1 norm to $u_{i,0}$ and $\lambda_i(f,p) \to \lambda_{i,0}$, where $u_{i,0}$ and $\lambda_{i,0}$ denote the i-th eigenvector and -value respectively of X_0. Using the argument given above it is also straightforward, that $(W_f^+)^T u_i(f,p) \to (W^+)^T u_{i,0}$ in ℓ_1. As for the general weighting case, the elements of $(W^+)^T u_{i,0}$ decrease exponentially and $\mathcal{O}_f^T(W_f^+)^T u_i(f,p) \to \mathcal{O}^T(W^+)^T u_{i,0}$. Thus the arguments for the terms $W_f^+(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})\beta_{f,p}^T(W_f^+)^T$, $W_f^+\beta_{f,p}(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p})^T(W_f^+)^T$ and $-W_f^+\beta_{f,p}(\hat{\Gamma}_p^- - \Gamma_p^-)\beta_{f,p}^T(W_f^+)^T$ are analogous to the corresponding arguments for the case of general weighting matrix $\hat{W}_f^+$. For the remaining two terms, we use the power series expansion of $(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2}$ having a convergence region including the closed disc with center $c > 0$ and radius $(c - c_1) > 0$:

$$(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2} = \sum_{j=1}^{\infty} a_i([\hat{\Gamma}_f^+ - cI]^i - [\Gamma_f^+ - cI]^i)$$

Note, that the coefficients a_i are independent of f, since they are the power series coefficients of the complex function $z^{-1/2}$, where the power series expansion is performed at c. Let $\hat{G}_f(c) = [\hat{\Gamma}_f^+ - cI]$ and let $G_f(c) = [\Gamma_f^+ - cI]$, with slight abuse of notation. Then

$$\begin{aligned}\hat{G}_f(c)^i - G_f(c)^i &= \hat{G}_f(c)^{i-1}(\hat{\Gamma}_f^+ - \Gamma_f^+) + \hat{G}_f(c)^{i-2}(\hat{\Gamma}_f^+ - \Gamma_f^+)G_f(c) + .. + (\hat{\Gamma}_f^+ - \Gamma_f^+)G_f(c)^{i-1} \\ &\doteq G_f(c)^{i-1}(\hat{\Gamma}_f^+ - \Gamma_f^+) + .. + (\hat{\Gamma}_f^+ - \Gamma_f^+)G_f(c)^{i-1}\end{aligned}$$

where $\doteq$ denotes equality up to order $o(T^{-1/2})$ a.s., which holds due to the uniform convergence of the sample covariances (see Lemma A.3.1). Now consider $x^T(f,p)(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2})y(f,p)$, where $x(f,p), y(f,p) \in \mathbb{R}^{fs\times 1}$ are vectors, such that $f\|x(f,p) - x_0\|_1 \to 0$, $f\|y(f,p) \to y_0\|_1 \to 0$, where $x_0, y_0 \in \ell^2$ have elements decreasing exponentially. Then

$$x^T[(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2}]y \doteq \sum_{i=1}^{\infty} a_i \sum_{j=0}^{i-1} x^T G_f(c)^j(\hat{\Gamma}_f^+ - \Gamma_f^+)G_f(c)^{i-j-1}y$$

where we omitted the argument (f,p) for $x(f,p)$ and $y(f,p)$ for notational convenience. In each summand we multiply $\hat{\Gamma}_f^+ - \Gamma_f^+$ with $x^T(f,p)G_f(c)^j$ from the left and with $G_f(c)^{i-j-1}y(f,p)$ from the right. Now e.g. $G_f(c)^{i-j-1}y(f,p)$ converges to $G(c)^{i-j-1}y_0$, where the 1-norm of the difference can be bounded by $C(\|y(f,p) - y_0\|_1(c-c_1)^{i-j-1} + (c-c_1)^{i-j-2}(i-j-1)|\rho_y'|^f)$, where $\rho_y' = (\max\{|\rho_p|, |\rho_y|\}) + \varepsilon, \varepsilon > 0$ arbitrarily small. Here ρ_p denotes a pole of $k(z)$ of maximum modulus and ρ_y denotes the rate of decrease of the elements of y_0. This bound follows from

$$\|G_f(c)^{i-j-1}y(f,p) - G(c)^{i-j-1}y_0\|_1 \le \|G_f(c)\|_1^{i-j-1}\|y(f,p) - y_0\|_1 + \|[G_f(c)^{i-j-1} - G(c)^{i-j-1}]y_0\|_1$$

using analogous arguments as those, which have been used to prove equation (4.21). The same arguments show $\|G_f(c)^j x(f,p) - G(c)^j x_0\|_1 \le C(\|x(f,p) - x_0\|_1(c-c_1)^j + |\rho_x'|^f(c-c_1)^{j-1}j$. Here $\rho_x' = (\max\{|\rho_p|, |\rho_x|\}) + \varepsilon, \varepsilon > 0$ arbitrarily small.

These arguments show, that for each summand we obtain a situation similar to the situation involving the term $\hat{\Gamma}_p^- - \Gamma_p^-$ in the proof of Lemma 4.1.4: Consider

$$\mathrm{vec}[x^T(f,p)G_f(c)^j(\hat{\Gamma}_f^+ - \Gamma_f^+)G_f(c)^{i-j-1}y(f,p)] = L_f(x(f,p), y(f,p), i, j)[\hat{g}_{T,f} - g_f]$$

Since $\|(G_f(c)^i - G(c)^i)x_0\|_1 = O((c-c_1)^{i-1}i|\rho_x'|^i)$, $\|(G_f(c)^i - G(c)^i)y_0\|_1 = O((c-c_1)^i i|\rho_y'|^i)$ show the exponential decrease of $G(c)^i y_0$ and $G(c)^j x_0$ respectively, it follows from Lemma 4.1.2, that there exists a vector $L(x_0, y_0, i, j)$ such that:

$$\|L_f(x,y,i,j) - L(x_0,y_0,i,j)\|_1 \le C\max\{\|x - x_0\|_1, \|y - y_0\|_1, |\rho|^f\})f(c-c_1)^{i-1}(i+1)$$

where $1 > |\rho| = \max\{|\rho'_x|, |\rho'_y|\}$ and where we omitted the argument (f,p) for $x(f,p)$ and $y(f,p)$ for notational convenience. Here the constant $C < \infty$ is independent of i and j. Thus finally, we obtain

$$\begin{aligned} & x^T(f,p)[(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2}]y(f,p) && \doteq \\ \sum_{i=1}^{\infty} & \sum_{j=0}^{i-1} a_i x^T(f,p) G_f(c)^j (\hat{\Gamma}_f^+ - \Gamma_f^+) G_f(c)^{i-j-1} y(f,p) && = \\ & \left[\sum_{i=1}^{\infty} a_i \sum_{j=0}^{i-1} L_f(x(f,p), y(f,p), i, j)\right] (\hat{g}_{T,f} - g_f) && = \\ & L_f(x(f,p), y(f,p))(\hat{g}_{T,f} - g_f) \end{aligned}$$

Now using the bounds on the norm of $L_f(x(f,p), y(f,p), i, j) - L(x_0, y_0, i, j)$ derived above (which are independent of j) and the fact, that

$$\sum_{i=1}^{\infty} \sum_{j=0}^{i-1} a_i (c - c_1)^{i-1} (i+1) = \sum_{i=1}^{\infty} a_i (c - c_1)^{i-1} (i+1) i < \infty$$

we obtain the convergence in the ℓ_1 norm of $L_f(x(f,p), y(f,p))$ to a vector $L(x_0, y_0)$ since $|\rho|^f \to 0$ and $f\|x(f,p) - x_0\|_1 \to 0, f\|y(f,p) - y_o\|_1 \to 0$ due to the assumptions on $x(f,p)$ and $y(f,p)$. It can also be verified, that $L(x_0, y_0)$ has elements decreasing exponentially, which follows from the representation of the elements of $L(x_0, y_0, i, j)$ as the sum of entries in $G_f(c)^j x(f,p)$ and $G_f(c)^{i-j-1} y(f,p)$, and from the exponential decrease of $G(c)^j x_0$ and $G(c)^{i-j-1} y_0$ respectively.

Thus the theorem is proved, if we can prove, that whenever $(\hat{\Gamma}_f^+)^{-1/2} - (\Gamma_f^+)^{-1/2}$ occurs, the pre- and postmultiplying vectors converge in ℓ_1 norm to a vector with elements decreasing exponentially. The following vectors may occur as $x(f,p)$ or $y(f,p)$ in the expressions above (see equations (4.15,4.22)): $u_i(f,p), (\mathcal{O}_f)_i, (\Gamma^+)^{1/2} e_i, P_0(\Gamma^+)^{1/2} e_i$, where P_0 denotes the projection onto the orthogonal complement of the column space of $(\Gamma^+)^{-1/2}\mathcal{O}$. Here $e_i, 1 \leq i \leq n$ denote the first n vectors of the canonical basis in ℓ_1. Clearly all these vectors converge in ℓ_1 norm to vectors with elements decreasing exponentially. Thus the result (finally) follows also in the case of `CCA`.

4.3 The Case of Observed Inputs

In this section we will generalize the result presented above to the case of additional observed inputs $u(t)$. The reader is reminded, that the thesis is mainly concerned with the case of no observed inputs. For this reason, some of the results are quite incomplete corresponding to the case of observed inputs. Especially the imposed conditions on the inputs do not seem to be minimal. However there are two strong reasons for the inclusion of this section: First of all it seems to be desirable to extend the result to the case of observed inputs, since this case will occur quite often in practice. And secondly conceptually (in the sense that the structure of the proofs is quite similar) there are only minor changes as compared to the case of no observed inputs, under some suitable assumptions on the inputs. Let $(u(t))_{t \in \mathbb{Z}}$ denote the process generating the observed input series and let $u_1, \cdots, u_T$ denote the corresponding observations.

DEFINITION (ASSUMPTIONS OF THIS SECTION ON THE INPUTS) Throughout this section it will be assumed, that the process $(z(t))_{t\in\mathbb{Z}} = ([y(t)^T, u(t)^T]^T)_{t\in\mathbb{Z}}$ is a (linearily regular, see Appendix A) stationary process i.e. that $(z(t))_{t\in\mathbb{Z}} \in (\mathbb{R}^{s+m})^{\mathbb{Z}}$, $z(t) = [y(t)^T, u(t)^T]^T$ is generated by a white noise sequence $\eta(t) = [\varepsilon(t)^T, \varepsilon_u(t)^T]^T$ fulfilling the standard assumptions as

$$\begin{aligned} y(t) &= \sum_{j=0}^{\infty} K(j)\varepsilon(t-j) + L(j)u(t-j) \\ u(t) &= \sum_{j=0}^{\infty} K_u(j)\varepsilon_u(t-j) \end{aligned}$$

where the noise $\eta(t)$ has variance equal to identity (and thus the input $u(t)$ is totally uncorrelated with the noise $\varepsilon(t)$) and where $k_u(z) = \sum_{j=0}^{\infty} K_u(j)z^{-j} \in U_r^{(m)}$ for some integer r.

This assumption implies, that for some constants $0 < C_1 < C_2 < \infty$ the following equation holds:

$$0 < C_1 I \leq f_u(\lambda) \leq C_2 I < \infty \tag{4.23}$$

where $f_u(\lambda) = \frac{1}{2\pi} k_u(e^{i\lambda}) k_u(e^{i\lambda})^*$ denotes the spectrum of $(u(t))_{t\in\mathbb{Z}}$.

REMARK 4.3.1 This set of assumptions seems to be rather restrictive and it seems to be likely, that the result holds under weaker conditions on the input, e.g. only assuming that the estimates of the first and the second moments converge suitably fast (comp. the results in Hannan and Deistler, 1988). However, no attempt was made to generalize the result to this setup. In section 6.2 we will derive the asymptotic distribution of the estimates obtained by the `MOESP` *class of algorithms, where weaker assumptions on the input process will suffice to consistently estimate the transfer function l, describing the relation from the observed inputs to the outputs. The motivation for this type of strong assumptions lies in the fact, that using these assumptions we can immediately show, that for the joint process $z(t)$ the estimates of the covariance sequence fulfill the uniform convergence (see Theorem A.3.1). Also the results of Lemma 4.1.4, which were central to the proof of the asymptotic normality, hold unchanged in this case. These assumptions ensure the existence and the (uniform in $f+p$) boundedness of the covariance matrices $\Gamma_{u,u} = \mathbb{E}U_{f+p}^+(t)(U_{f+p}^+(t))^T$ and their inverses, which will be needed throughout this section. Also the exponential decrease of the entries of $\Gamma_{u,u}^{-1}$ in the case $f+p = \infty$ with the distance to the main diagonal will be used often, i.e. that for the block Toeplitz matrix $\Gamma_{u,u}^{-1} = [\gamma_{u,u}^-(i-j)]_{i,j\in\mathbb{N}}$, say, the bound $\|\gamma_{u,u}^-(j)\| \leq C|\rho|^j, 0 < C < \infty, |\rho| < 1$ holds.*

One simple extension would be to allow for a nonzero mean in $(u(t))_{t\in\mathbb{Z}}$ i.e. to define $\tilde{u}(t) = u(t) + \mu$. Then it follows from the ergodicity of the noise and thus of the process $(u(t))_{t\in\mathbb{Z}}$, that the sample mean converges a.s. to the true mean i.e. $\bar{u} = \frac{1}{T}\sum_{t=1}^{T} u_t = \mu$ a.s. Thus the uniform convergence of the sample covariances still holds for the variables $\hat{u}(t) = \tilde{u}(t) - \bar{u} = u(t) + \mu - \bar{u}$. This shows, that all results, which are derived from the uniform convergence of the sample covariances stated in Theorem A.3.1, will still hold in this situation. In particular the consistency results will hold unchanged.

Recall from section 3.5 the class of algorithms considered in this part of the thesis: Let $Y_f^+(t) = [y(t)^T, \cdots, y(t+f-1)^T]^T$, $Z_p^-(t) = [y(t-1)^T, u(t-1)^T, \cdots, y(t-p)^T, u(t-p)^T]^T$ and let $U_f^+(t) = [u(t)^T, \cdots, u(t+f-1)^T]^T$. In the sequel we will adopt the following notation for the various covariance matrices: $\Gamma_{a,b}$ will denote the covariance of a and b, where the indices a and b are used to denote the quantities, which are involved in the expression: u is used for $U_f^+(t)$, z for $Z_p^-(t)$, x for $x(t)$ and y for $Y_f^+(t)$. At some places, we will also use the same symbols for denoting

the corresponding infinite matrices, e.g. $\Gamma_{u,u}$ will also be used to denote $\Gamma_{u,u} = \mathbb{E}U^+(t)(U^+(t))^T$. Whenever there might be confusions about the correct meaning, we note this. Unless the contrary is stated explicitely, we will however always use the notation for the finite matrices. Thus e.g. $\Gamma_{x,u} = \mathbb{E}x(t)(U_f^+(t))^T$. Estimates of these covariance matrices will be denoted with $\hat{\Gamma}_{a,b}$. Using this notation it is straightforward, that $Y_f^{+,\Pi}(t) = Y_f^+(t) - \Gamma_{y,u}(\Gamma_{u,u})^{-1}U_f^+(t)$, where $Y_f^{+,\Pi}(t)$ denotes the residuals from the projection of $Y_f^+(t)$ onto the space spanned by the components of $U_f^+(t)$. Now the regression calculated in the first step of the considered subspace procedure can be written as $\hat{\beta}_z = (\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{-1}$. A weighted version of this matrix is decomposed in the algorithm, i.e. $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^- = \hat{U}\hat{\Sigma}\hat{V}^T = \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T + \hat{R}$, where the largest n singular values are contained in $\hat{\Sigma}_n$ and appear in decreasing order. Here we will in a first step restrict the choice of $\hat{W}_p^-$ to be equal to $(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$ similar to the case of no observed inputs. Note, however, that with this restriction we exclude the `N4SID` procedure, as described in section 3.3. The case of the weighting used in `N4SID`, i.e. $\hat{W}_p^- = (\hat{\Gamma}_{z,z})^{1/2}$ will be dealt with afterwards (see REMARK 4.3.3).

Corresponding to the weighting matrix $\hat{W}_f^+$ we will for the moment restrict the choice to $\hat{W}_f^+ = (\hat{\Gamma}_{y,y} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,y})^{-1/2}$. This is done, since this case is the most complicated one. The proof for $\hat{W}_f^+ = W_f^+(k_W)$ proceeds completely analogously. As for the case of no observed inputs we will first discuss the case, that f is fixed and then extend the proof to the case of $f \to \infty$.

4.3.1 The Case of f fixed

The proof of a central limit theorem in this case proceeds analogous to the proof given in section 4.1.3: Again we will define a number of finite dimensional matrices $\hat{x}_i, i = 1, \cdots, 7$, whose vectorization $\hat{m}_p = [\mathrm{vec}(x_1)^T, \cdots, \mathrm{vec}(x_7)^T]^T$ is a function of the estimates of the sample covariances $\hat{\gamma}_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f+p-1)$ of the joint process $z(t)$. Let m_p denote the vector obtained by using the true covariances $\gamma_{z,z}(0), \cdots, \gamma_{z,z}(f+p-1)$ instead of the sample estimates. It will be part of the proof to show, that m_p converges sufficiently fast as p increases as a function of the sample size T within some bounds given in the theorem i.e. $\|m_p - m_0\| = O(|\rho_0^{'}|^p), \forall |\rho_0| < |\rho_0^{'}| < 1$, where $m_0 = \lim_{p\to\infty} m_p$. Then we will show a central limit theorem for $\hat{m}_p$ i.e. that $\sqrt{T}(\hat{m}_p - m_0)$ converges in distribution to a random variable, which is distributed Gaussian with zero mean and variance V^m. Finally, the estimates $(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{E}_T, \hat{K}_T)$ are a nonlinear function ψ of $\hat{m}_p$ by the construction of $\hat{m}_p$ and thus linearization of ψ at the true system shows the result. Thus the structure of the proof is completely the same as the structure of the proof for the case of no observed inputs. Some of the results of section 4.1.3 will be used directly. Note, that the central limit theorem for the sample covariance estimates in the sense of Lewis & Reinsel also holds under our assumptions, since the process $(z(t))_{t\in\mathbb{Z}}$ fulfills the conditions of Lemma 4.1.4.

Consider the two regressions, which are used to estimate the matrices, once the state has been estimated as $\hat{x}_t = \hat{\mathcal{K}}_p Z_{t,p}^-$, where $\hat{\mathcal{K}}_p = \hat{\Sigma}_n^{1/2}\hat{V}_n^T(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{-1/2} = \hat{\Sigma}_n^{-1/2}\hat{U}_n^T(\hat{\Gamma}_{y,y} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,y})^{-1/2}\hat{\beta}_z$ (Note, that here we do not invest any a priori knowledge on a possible delay from inputs to outputs and thus D contains free parameters. The case, where a delay is postulated and therefore $D = 0$, can be analyzed completely analogously.):

$$
\begin{aligned}
[\hat{C}_T, \hat{D}_T] &= \langle y_t, \begin{bmatrix} \hat{x}_t \\ u_t \end{bmatrix} \rangle \begin{bmatrix} \langle \hat{x}_t, \hat{x}_t \rangle & \langle \hat{x}_t, u_t \rangle \\ \langle u_t, \hat{x}_t \rangle & \langle u_t, u_t \rangle \end{bmatrix}^{-1} \\
\left[\hat{A}_T, \hat{B}_T, \hat{K}_T\right] &= \langle \hat{x}_{t+1}, \begin{bmatrix} \hat{x}_t \\ u_t \\ \hat{\varepsilon}_t \end{bmatrix} \rangle \begin{bmatrix} \langle \hat{x}_t, \hat{x}_t \rangle & \langle \hat{x}_t, u_t \rangle & 0 \\ \langle u_t, \hat{x}_t \rangle & \langle u_t, u_t \rangle & 0 \\ 0 & 0 & \langle \hat{\varepsilon}_t, \hat{\varepsilon}_t \rangle \end{bmatrix}^{-1} \\
\hat{E}_T &= (\langle y_t - \hat{C}_T\hat{x}_t - \hat{D}_T u_t, y_t - \hat{C}_T\hat{x}_t - \hat{D}_T u_t \rangle)^{1/2}
\end{aligned}
$$

where $(.)^{1/2}$ here denotes the lower triangular Cholesky factor and $\hat{\varepsilon}_t = \hat{E}_T^{-1}(y_t - \hat{C}_T\hat{x}_t - \hat{D}_T u_t)$. From these equations it can be seen, that the estimates are a nonlinear function of the following inner products: $\langle y_t, \hat{x}_t\rangle, \langle y_t, u_t\rangle, \langle u_t, \hat{x}_t\rangle, \langle u_t, u_t\rangle, \langle \hat{x}_t, \hat{x}_t\rangle, \langle \hat{x}_{t+1}, \hat{x}_t\rangle, \langle \hat{x}_{t+1}, y_t\rangle, \langle y_t, y_t\rangle$ and $\langle \hat{x}_{t+1}, u_t\rangle$. Note that with the particular choice of the weighting matrix $\hat{W}_p^- = (\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$ the sample state covariance will no longer be diagonal. For later reference we note, that the diagonality of the state covariance matrix is obtained with $\hat{W}_p^- = (\hat{\Gamma}_{z,z})^{1/2}$.

Now $\hat{x}_t = \hat{\mathcal{K}}_p Z_{t,p}^-, \hat{x}_{t+1} = \hat{\mathcal{K}}_p Z_{t+1,p}^-$, where $\hat{\mathcal{K}}_p = \hat{\Sigma}_n^{-1/2}\hat{U}_n^T\hat{W}_f^+\hat{\beta}_z$ depends on the first f covariance estimates $\hat{\gamma}_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f-1)$ of the joint process $z(t)$, the matrices $\hat{\Sigma}_n, \hat{U}_n$, which are all of fixed finite dimension for f fixed and finite, as well as $\hat{\beta}_z$, whose dimension depends on p and thus increases with $p \to \infty$ as a function of T. Note, that $\hat{\Sigma}_n$ and $\hat{U}_n$ are calculated as singular values and singular vectors respectively of the matrix $\hat{W}_f^+\hat{\beta}_z(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$ or equivalently as the square roots of the eigenvalues and the eigenvectors respectively of the matrix $\hat{W}_f^+\hat{\beta}_z(\hat{\Gamma}_{z,y} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,y})\hat{\beta}_z^T(\hat{W}_f^+)^T \in \mathbb{R}^{fs\times fs}$. In

$$\langle \hat{x}_t, \hat{x}_t\rangle = \hat{\mathcal{K}}_p\langle Z_{t,p}^-, Z_{t,p}^-\rangle\hat{\mathcal{K}}_p^T = \hat{\Sigma}_n^{-1/2}\hat{U}_n^T\hat{W}_f^+\hat{\beta}_z\hat{\Gamma}_{z,z}\hat{\beta}_z^T(\hat{W}_f^+)^T\hat{U}_n\hat{\Sigma}_n^{-1/2}$$

the central term is equal to $\hat{\beta}_z\hat{\Gamma}_{z,z}\hat{\beta}_z^T$. Similar calculations for the other inner products lead to the following essential terms, which are used to construct $\hat{m}_p$ in this case:

$$\begin{aligned}
\hat{x}_1 &= \hat{\beta}_z(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})\hat{\beta}_z^T \\
\hat{x}_2 &= \hat{\beta}_z\hat{\Gamma}_{z,z}\hat{\beta}_z^T \\
\hat{x}_3 &= \langle z_t, Z_{t,p}^-\rangle\hat{\beta}_z^T \\
\hat{x}_4 &= \hat{\beta}_z\bar{T}_p \\
\hat{x}_5 &= \hat{\beta}_z\bar{S}_p\langle Z_{t,p}^-, z_t\rangle \\
\hat{x}_6 &= \hat{\beta}_z\bar{S}_p\hat{\Gamma}_{z,z}\hat{\beta}_z^T \\
\hat{x}_7 &= [\mathrm{vec}[\hat{\gamma}_{z,z}(0)]^T, \cdots, \mathrm{vec}[\hat{\gamma}_{z,z}(f-1)]^T]^T
\end{aligned} \tag{4.24}$$

where

$$\bar{T}_k = \begin{bmatrix} I_{s+m} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \in \mathbb{R}^{k(s+m)\times(s+m)} \text{ and } \bar{S}_p = \begin{bmatrix} 0 & \cdots & & & \\ I_{s+m} & 0 & & & \\ 0 & I_{s+m} & 0 & & \\ & \ddots & \ddots & \ddots & \\ & & & I_{s+m} & 0 \end{bmatrix} \in \mathbb{R}^{p(s+m)\times p(s+m)}.$$

where $I_{s+m} \in \mathbb{R}^{(s+m)\times(s+m)}$ denotes the identity matrix. The term $\hat{x}_7$ can be shown to fulfill a central limit theorem easily using the result in Lemma 4.1.4. Thus we will only deal with the remaining 6 terms in the following.

Note, that all terms are functions of the estimates of the first $f+p$ covariances of the joint process $(z(t))_{t\in\mathbb{Z}}$, $\hat{\gamma}_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f+p-1)$ say. First consider m_p, i.e. the vector, where the sample covariance estimates are replaced by the true values: $\beta_z = (\Gamma_{y,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,y})(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^{-1}$. Recall equation (3.3): $Y_f^+(t) = \mathcal{O}_f\mathcal{K}_p Z_p^-(t) + \mathcal{O}_f(A - KE^{-1}C)^p x(t-p) + \mathcal{U}_f U_f^+(t) + \mathcal{E}_f E_f^+(t)$. Thus (here the subscript x_- is used for $x(t-p)$)

$$\Gamma_{y,u} = \mathcal{O}_f\mathcal{K}_p\Gamma_{z,u} + \mathcal{O}_f(A - KE^{-1}C)^p\Gamma_{x_-,u} + \mathcal{U}_f\Gamma_{u,u}$$

and

$$\Gamma_{y,z} = \mathcal{O}_f\mathcal{K}_p\Gamma_{z,z} + \mathcal{O}_f(A - KE^{-1}C)^p\Gamma_{x_-,z} + \mathcal{U}_f\Gamma_{u,z}$$

resulting in

$$\begin{aligned}
\Gamma_{y,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z} &= \mathcal{O}_f\mathcal{K}_p(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z}) \\
&\quad + \mathcal{O}_f(A - KE^{-1}C)^p(\Gamma_{x_-,z} - \Gamma_{x_-,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})
\end{aligned}$$

Therefore we have

$$\beta_z - \mathcal{O}_f\mathcal{K}_p = \mathcal{O}_f(A-KE^{-1}C)^p(\Gamma_{x_-,z} - \Gamma_{x_-,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^{-1}$$

where the inverse exists under our assumptions on the process $z(t)$ (comp. Peternell *et al.*, 1996). Since the norm of all matrices in this expression is bounded uniformely in p, the norm of the difference is bounded by $c\|(A-KE^{-1}C)^p\|$ and thus is $O(|\rho_0^{'}|), \forall |\rho_o| < |\rho_0^{'}| < 1$. Here $\rho_0 = \lambda_{max}(A_0 - K_0E_0^{-1}C_0)$. This ensures the convergence of β_z to $\mathcal{O}_f\mathcal{K}_p$ at a suitable rate determined by $p = p(T) \to \infty$. Since this is the central term in all the six variables $x_1(p), \cdots, x_6(p)$, making the dependence on $p = p(T)$ more explicit, it is straightforward, that $\|m_p - m_0\| = O(|\rho_0^{'}|^p)$ holds. Thus the uniform convergence of the sample covariance estimates (comp. Theorem A.3.1) for $f + p = O((\log T)^a)$ ensures the consistency for $x_i(p)$ (where $x_i(p)$ denotes the expressions obtained by replacing in the definition of $\hat{x}_i$ the estimated covariances with the true covariances), which is straightforward to see using the same arguments as given in the proof in section 4.1.3 i.e. we obtain $\|\hat{m}_p - m_0\| = o(1)$ a.s. Note, that this result holds for $(k,l) \in M_n^j$ and not only in the generic subset $U_n^+(u(t))$.

In the next step the asymptotic distribution of $\sqrt{T}(\hat{m}_p - m_p)$ will be examined. First consider

$$\begin{aligned}
\hat{x}_1 - x_1 &= \hat{\beta}_z(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})\hat{\beta}_z^T - \beta_z(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\beta_z \\
&= (\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{-1}(\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^T - \\
&\quad (\Gamma_{y,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^{-1}(\Gamma_{y,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^T \\
&= (\hat{\Gamma}_{y,z} - \Gamma_{y,z})\beta_z^T - (\hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\beta_z^T + \\
&\quad \beta_z(\Gamma_{z,z} - \hat{\Gamma}_{z,z})\beta_z^T - \beta_z(\Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})\beta_z^T + \\
&\quad \beta_z(\hat{\Gamma}_{y,z} - \Gamma_{y,z})^T - \beta_z(\hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^T + o_P(T^{-1/2})
\end{aligned}$$

where the $o_P(T^{-1/2})$ term is due to neglecting higher order terms, which does not affect the asymptotic distribution. This again follows from the uniform convergence of the sample covariances. The results in Lemma 4.1.4 show the asymptotic normality: Consider the first term for example: $\mathrm{vec}[(\hat{\Gamma}_{y,z} - \Gamma_{y,z})\beta_z^T] = L_{f,p}\mathrm{vec}[\hat{\gamma}_{z,z}(0) - \gamma_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f+p-1) - \gamma_{z,z}(f+p-1)]$, where $L_{f,p}$ converges in ℓ_1 to a matrix, whose rows have elements decreasing exponentially, since $\beta_z \to \mathcal{O}_f\mathcal{K}$ as $p \to \infty$ (see the remarks below Lemma 4.1.2). Analogous arguments for the remaining terms show the asymptotic normality of $\sqrt{T}(\hat{x}_1 - x_1)$.

The asymptotic normality of $\sqrt{T}(\hat{x}_2 - x_2)$ can be shown in a similar way, since $\beta_z\Gamma_{z,z}(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}\Gamma_{u,z})^{-1}$ has elements decreasing exponentially (where here the subscript $_z$ is used, although the infinite matrices are investigated, e.g. $\Gamma_{z,z} = \mathbb{E}Z(t)^-(Z(t)^-)^T$), which follows from the exponential decrease of $\mathcal{K}$ and the Toeplitz structure of $\Gamma_{z,z}$, where the entries decrease exponentially with the distance to the main diagonal and the same property of $\Gamma_{z,z}^{-1}$. Using the same arguments for the other terms finally results in the conclusion, that $\sqrt{T}(\hat{m}_p - m_p)$ is asymptotically normal.

The last step consists in showing the differentiability of ψ at $m_0 = \lim_{p\to\infty} m_p$. Note, that the essential step in ψ consists in the eigenvaluedecomposition of $\hat{X}_{f,p} = \hat{W}_f^+\hat{\beta}_z(\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^T(\hat{W}_f^+)^T$. As has been noted already, this estimate converges almost surely to $X_f = W_f^+\mathcal{O}_f\mathcal{K}(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\mathcal{K}^T\mathcal{O}_f^T(W_f^+)^T$. In order to ensure the differentiability of the eigenvaluedecomposition of this matrix, we have to impose suitable restrictions on the true pair of transfer functions $(k_0,l_0) \in M_n^j$: From Lemma B.2.1 we know, that the eigenvaluedecomposition is differentiable for all eigenvalues of multiplicity 1. Thus we have to ensure, that the nonzero eigenvalues of X_f are distinct. For given input process $(u(t))_{t\in\mathbb{Z}}$ let $U_n^+(u(t)) \subset M_n^j$ denote the set of all pairs $(k,l) \in M_n^j$, such that the nonzero eigenvalues of X_f are all distinct (again using the same symbol for the case $W_f^+ = W_f^+(k_W)$ and the `CCA` case). Analogously to the case of no observed inputs it can be shown, that $U_n^+(u(t))$ is in fact an open and dense subset of M_n^j: The essential property in this respect was the analyticity of the Gramians as functions of the entries in the system matrices. In the new setting the Gramians are equal to $\mathcal{O}_f^T(W_f^+)W_f^+\mathcal{O}_f$ and $\mathcal{K}(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\mathcal{K}^T$, which are in fact analytic in the entries of

(A, B, C, D, E, K), as tedious calculations show. Here again we use e.g. $\Gamma_{z,z}$ for the infinite-dimensional matrix $\mathbb{E}Z^-(t)(Z^-(t))^T$. For `CCA` the matrix $(W_f^+)^T W_f^+ = (\Gamma_{y,y} - \Gamma_{y,u}\Gamma_{u,u}\Gamma_{u,y}) = \mathbb{E}(\mathcal{O}_f x(t)^\Pi + \mathcal{E}_f E_f(t))(\mathcal{O}_f x(t)^\Pi + \mathcal{E}_f E_f(t))^T = \mathcal{O}_f \mathbb{E}x(t)^\Pi (x(t)^\Pi)^T \mathcal{O}_f^T + \mathcal{E}_f \mathcal{E}_f^T$ and thus we obtain the same structure as in the case of no observed inputs, with the change, that the state covariance must be replaced with the covariance of $x(t)^\Pi$, which is analytic in the entries of the system matrices for fixed input spectrum $f_u(\lambda)$ fulfilling the assumptions of this section (see equation (4.23)). Here we use the exponential decrease of the Markoff coefficients of $k_u(z)$. Similar arguments also show the analyticity of the controllability Gramian, since each element of $\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z}$ is an analytic function of the entries and the entries in $\mathcal{K}$ converge exponentially to zero.

Once the analyticity of the Gramians is proved, the genericity follows easily from the evaluations in section 4.1.1: Openess follows immediately and denseness follows from the fact, that for $k \in U_n^+$ the pair $(k, 0) \in U_n^+(u(t))$. Again we have to deal with the problem of the choice of the orientation of the singular vectors. We again impose nonnegativity restrictions on one particular entry in each column of the matrix U_n containing the eigenvectors of X_0. This will lead to an essentially continuous SVD in the neighborhood of X_0, if the elements, which are restricted to be nonnegative, are in fact positive, which will be assumed in the sequel. Note, that such a choice always exists. Again the same remark as in section 4.1.1 is in order: The actual implementation of the SVD might choose another normalization of the orientation of the singular vectors. If these normalizations are such, that the corresponding SVD in the neighborhood of X_0 is essentially continuous, then the asymptotic behaviour is described by the following result. Thus assume, that the singular vectors contained in U_n are chosen, such that the SVD in the neighborhood of X_0 is essentially continuous. Let the corresponding realization of $(k_0, l_0) \in U_n^+(u(t))$ be denoted with $(A_0, B_0, C_0, D_0, E_0, K_0)$. Thus we have shown the following extension of Theorem 4.1.1:

Theorem 4.3.1 *Let $(y(t))_{t\in\mathbb{Z}}$ be generated by a true pair $(k_0, l_0) \in U_n^+(u(t))$ and let the process $\begin{bmatrix} y(t) \\ u(t) \end{bmatrix}$ be a stationary process, which fulfills the assumptions of this section. Let the row truncation index $f \geq n$ be fixed and the estimation be performed using the true order n. Let the weighting matrix $\hat{W}_f^+$ be chosen as in `CCA` or as $\hat{W}_f^+ = W_f^+(k_W)$, where k_W denotes a rational weighting transfer function.*

If p fulfills the following conditions:

- $p \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ *and for some* $d > 1$, *where* $\rho_0 = \lambda_{max}(A_0 - K_0 E_0^{-1} C_0)$
- $p/(\log T^a) \to 0$ *for some* $a < \infty$

then for the particular realization $(A_0, B_0, C_0, D_0, E_0, K_0)$ presented above, we obtain:

$$\sqrt{T}vec(\hat{A}_T - A_0, \hat{B}_T - B_0, \hat{C}_T - C_0, \hat{D}_T - D_0, \hat{E}_T - E_0, \hat{K}_T - K_0) \xrightarrow{d} Z$$

where Z is multivariate normally distributed with zero mean and variance V.

Here again the same remarks apply as below Theorem 4.1.1: The asymptotic variance V and the set $U_n^+(u(t))$ depend on the row truncation index f, however this has not been emphasized in the notation. Note, that the set $U_n^+(u(t))$ depends on the true input spectrum $f_u(\lambda)$. Also the asymptotic variance V depends on the fourth moments of the noise $\eta(t)$.

As in the case of no observed inputs we can now state as a corollary some direct consequences of the central limit theorem:

Corollary 4.3.1 *Let $(y_t)_{t\in\mathbb{Z}}$ be generated by a pair of transfer functions $(k_0, l_0) \in M_n^j$ using an input process $(u(t))_{t\in\mathbb{Z}}$, such that the joint process $([y(t)^T, u(t)^T]^T)_{t\in\mathbb{Z}}$ fulfills the assumptions of this section. Let $f \geq n$ be a fixed integer. Let the estimation be performed with the true system order and let the weighting matrix $\hat{W}_f^+$ be chosen as in Theorem 4.3.1. If p fulfills the condition $p/(\log T)^a \to 0$ for some $a < \infty$, then there exists a realization $(A_0, B_0, C_0, D_0, E_0, K_0)$ of the true pair of transfer functions (k_0, l_0) and a sequence of orthonormal basis transformations S_T, such that*

$$||vec(S_T \hat{A}_T S_T^{-1} - A_0, S_T \hat{B}_T - B_0, \hat{C}_T S_T^{-1} - C_0, \hat{D}_T - D_0, \hat{E}_T - E_0, \hat{K}_T - K_0)|| \to 0, \ a.s.$$

Furthermore, let $(k,l) \in U_n^+(u(t))$ *and let* $g : \mathbb{R}^{(n+s)(n+m+s)} \to \mathbb{R}^r$ *be a mapping attaching the vector* $x \in \mathbb{R}^r$ *to the matrices* (A, B, C, D, E, K). *Furthermore let* g *be differentiable at* $(A_0, B_0, C_0, D_0, E_0, K_0)$ *with* J_g *denoting the matrix of the partial derivatives with respect to the entries in* (A, B, C, D, E, K). *Then the following holds:*

$$\sqrt{T}(g(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}_T, \hat{E}_T, \hat{K}_T) - g(A_0, B_0, C_0, D_0, E_0, K_0)) \xrightarrow{d} Z$$

where Z *is multivariate normally distributed with zero mean and variance* $V_g = J_g V J_g^T \in \mathbb{R}^{m \times m}$.

PROOF: Recall, that we obtained $||\hat{m}_p - m_0|| = o(1)$ a.s. for $(k,l) \in M_n^j$. The result then follows from Theorem B.2.1 and Theorem B.2.2. The second part is immediate from Theorem 4.3.1. □

REMARK 4.3.2 Throughout this section, it was assumed, that the matrix D is estimated without any restrictions. However, it is straightforward to see, that the same results hold for $D = 0$, *i.e. if a delay from the inputs to the outputs is present. Note, that in this case* $\langle u_t, \hat{\varepsilon}_t \rangle = 0$ *is not satisfied and thus the matrices* $(\hat{A}_T, \hat{B}_T, \hat{K}_T)$ *cannot be obtained from two seperate regressions in this setup.*

REMARK 4.3.3 The discussion above relates only to the case $\hat{W}_p^- = (\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$. *The arguments for the case of* $\hat{W}_p^- = (\hat{\Gamma}_{z,z})^{1/2}$ *are simpler, since in this case* x_1 *is not needed, as the eigenvaluedecomposition is performed for* $\hat{X}_{f,p} = \hat{W}_f^+ \hat{\beta}_z \hat{\Gamma}_{z,z} \hat{\beta}_z^T (\hat{W}_f^+)^T$. *Also the definition of the set* $U_n^+(u(t))$ *has to be adopted to this eigenvaluedecomposition.*

4.3.2 The Case $f \to \infty$

To show the same result for the case $f \to \infty$ we will in a first step restrict the weighting matrix to be equal to $W_f^+ = W_f^+(k_W)$ as in the case of no observed inputs, and again we will discuss the **CCA** case seperately afterwards. Thus examine the definition of the matrices $x_1, \cdots, x_7$ in equation (4.24): These quantities were introduced for two reasons: The system matrix estimates are a nonlinear function of these quantities and they all are of fixed dimension for f fixed. Note, that x_7 was only needed, since we discussed the **CCA** case of $\hat{W}_f^+ = (\hat{\Gamma}_{y,y} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,y})^{-1/2}$. If we use a general weighting matrix, we will only need $\hat{\gamma}_{z,z}(0) - \gamma_{z,z}(0)$. Now in the case $f \to \infty$, the matrices $x_1, \cdots, x_7$ will no longer be of fixed dimension, but instead of increasing dimension as f tends to infinity as a function of the sample size T. Thus we have to replace these variables with another set of related variables, which is obtained by replacing $\hat{\beta}_z$ in (4.24) with $\hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z = \hat{\Sigma}_n \hat{\mathcal{K}}_p$:

$$\begin{aligned}
\hat{x}_1 &= \hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z (\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z}) \hat{\beta}_z^T (W_f^+)^T W_f^+ \hat{\mathcal{O}}_f \\
\hat{x}_2 &= \hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z \hat{\Gamma}_{z,z} \hat{\beta}_z^T (W_f^+)^T W_f^+ \hat{\mathcal{O}}_f \\
\hat{x}_3 &= \langle z_t, Z_{t,p}^- \rangle \hat{\beta}_z^T (W_f^+)^T W_f^+ \hat{\mathcal{O}}_f \\
\hat{x}_4 &= \hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z \bar{T}_p \\
\hat{x}_5 &= \hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z \bar{S}_p \langle Z_{t,p}^-, z_t \rangle \\
\hat{x}_6 &= \hat{\mathcal{O}}_f^T (W_f^+)^T W_f^+ \hat{\beta}_z \bar{S}_p \hat{\Gamma}_{z,z} \hat{\beta}_z^T (W_f^+)^T W_f^+ \hat{\mathcal{O}}_f \\
\hat{x}_7 &= \langle z_t, z_t \rangle
\end{aligned} \tag{4.25}$$

using the notation $z_t^T = [y_t^T, u_t^T]$. Most of the remaining details of the proof combine the arguments of the proof for the case of no observed inputs and the case of additional observed inputs for f fixed given above, and hence are omitted. We only cite, that the exponential decrease of $K_u(j)$ is used to show the exponential decrease of $\mathcal{K}\Gamma_{z,u}\Gamma_{u,u}^{-1}$, where here $\Gamma_{z,u} = \mathbb{E}Z^-(t)(U^+(t))^T$ and $\Gamma_{u,u} = \mathbb{E}U^+(t)(U^+(t))^T$ are used to denote the infinite dimensional matrices. The definition of

$U_n^+(u(t)) \subset M_n^j$ also has to be adopted to the new situation: $(k,l) \in U_n^+(u(t))$ if and only if the n nonzero eigenvalues of $W^+\mathcal{O}\mathcal{K}(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\mathcal{K}^T\mathcal{O}^T(W^+)^T$ are distinct (or equivalently the n eigenvalues of $\mathcal{K}(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\mathcal{K}^T\mathcal{O}^T(W^+)^TW^+\mathcal{O})$. The proof, that $U_n^+(u(t))$ is open and dense in M_n^j follows again from the analyticity of the Gramians as a function of the entries in the system matrices and from the nonemptyness of $U_n^+(u(t))$, which follows from the fact, that $k_0 \in U_n^+$ implies $(k_0, 0) \in U_n^+(u(t))$. Analogously to the case of no observed inputs we obtain a particular realization $(A_0, B_0, C_0, D_0, E_0, K_0)$ of $(k_0, l_0) \in U_n^+(u(t))$ by fixing the sign of certain entries of U_n, the matrix containing the singular vectors of X_0. Using this choice also in the neighborhood of X_0 we obtain an essentially unique SVD. Again the same remark as in the case f fixed is in order: The choice of the entries, which are fixed, is done, such these particular entries in U_n are nonzero. This ensures, that the SVD is essentially continuous at X_0. The actual implementation of the SVD may take another choice of the signs of the singular vectors. As long as the corresponding SVD is essentially continuous at X_0, the results stated below will hold true.

For the case of `CCA` the situation is a bit more complicated. In this case we have to examine the matrix $\Gamma_{y,y} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,y}$ more closely: Recall, that $Y_{t,f}^+ = \mathcal{O}_f x_t + \mathcal{U}_f U_{t,f}^+ + \mathcal{E}_f E_{t,f}^+$ holds. Thus it is easy to see, that

$$\Gamma_{y,y} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,y} = \mathcal{E}_f\mathcal{E}_f^T + \mathcal{O}_f\Sigma\mathcal{O}_f^T$$

and thus the weighting is of the same structure as in the case of no observed inputs. Note also, that $\sqrt{T}\|\beta_z - \mathcal{O}_f\mathcal{K}_p\| \to 0$, so that we have to deal with the difference of $(\mathcal{E}_f\mathcal{E}_f^T + \mathcal{O}_f\Sigma\mathcal{O}_f^T)^{-1/2}\mathcal{O}_f$ and $(\mathcal{E}\mathcal{E}^T + \mathcal{O}\Sigma\mathcal{O}^T)^{-1/2}\mathcal{O}$, where the matrices without the subscript $_f$ denote the infinite dimensional matrices, obtained by letting $f \to \infty$. This difference has been shown to be of order $o(T^{-1/2})$ under the conditions of Theorem 4.2.1. Again here we have to define a set $U_n^+(u(t)) \in M_n^j$, such that for $(k_0, l_0) \in U_n^+(u(t))$ the matrix $W^+\beta_z W^-$ has distinct singular values. The proof, that $U_n^+(u(t))$ for given input process $(u(t))_{t\in\mathbb{Z}}$, and thus given spectrum $f_u(\lambda)$ fulfilling the restriction (4.23), is open and dense in M_n^j is completely analogous to the proof for no observed inputs: The Gramians are shown to be analytic functions of the entries of the system matrices, where we use the structure of the observability Gramian as described above: The Gramian is equal to $\mathcal{O}^T(\mathcal{E}\mathcal{E}^T + \mathcal{O}\Sigma\mathcal{O}^T)^{-1}\mathcal{O}$. Since Σ is an analytic function of the entries in (A, B, C, D, E, K) for stable A and the same is true for the entries of $\mathcal{E}^{-1}$ for stable $A - KE^{-1}C$, the result follows from the matrix inversion lemma and arguments analogous to the case of no observed inputs. This proves openess of $U_n^+(u(t))$ in this case. Denseness follows from the nonemptyness of $U_n^{++} \in U_n^{(m)}$ corresponding to `CCA` for $m = 0$ (see section 4.1.1) and the fact, that $k \in U_n^+$ implies $(k, 0) \in U_n^+(u(t))$.

Thus the results presented in section 4.2 also apply in this situation and we obtain the central limit theorem also in this case. Concluding we may state the last result of this section:

Theorem 4.3.2 *Let $(y(t))_{t\in\mathbb{Z}}$ be generated by a true pair $(k_0, l_0) \in U_n^+(u(t))$ and let the process $z(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}$ be a stationary process, which fulfills the assumptions of this section. Let the estimation be performed using the true order n.*

If the following conditions hold:

1. $p \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ *for some* $d > 1$, *where* $\rho_0 = \lambda_{max}(A_0 - K_0E_0^{-1}C_0)$,
2. $\max(f,p)/(\log T)^a \to 0$ *for some* $a < \infty$.
3. *If a general weighting matrix $W_f^+ = W_f^+(k_W)$ is used, where k_W is a stable, strictly minimum-phase rational transfer function, $f \geq -\frac{d}{2}\frac{\log T}{\log|\rho_p|}, \forall T > T_0$ for some $d > 1$, where ρ_p is a pole of $k_0 + k_W$ of maximum modulus*
4. `CCA` *is used, and $f \geq -\frac{d}{2}\frac{\log T}{\log|\rho_0|}, \forall T > T_0$ for some $d > 1$*

then for the particular realization $(A_0, B_0, C_0, D_0, E_0, K_0)$ of (k_0, l_0) described above, the following holds:

$$\sqrt{T}vec(\hat{A}_T - A_0, \hat{B}_T - B_0, \hat{C}_T - C_0, \hat{D}_T - D_0, \hat{E}_T - E_0, \hat{K}_T - K_0) \xrightarrow{d} Z$$

where Z is multivariate normally distributed with zero mean and variance V.
Moreover, the conclusions in Corollary 4.3.1 hold true also under the assumption on the truncation indices and the weighting matrices in this theorem.

The statement of Corollary 4.3.1 corresponding to the asymptotic distribution of a function of the system matrices is a simple consequence of the CLT for the system matrix estimates, and the a.s. consistency of the transfer function estimates has been proven already in (Peternell *et al.*, 1996). Again we note, that the set $U_n^+(u(t))$ and the asymptotic variance V will depend on the weighting scheme, however we neglected this fact in the notation. Again the variance depends on the fourth moments of $\eta(t)$ generating the joint process $z(t)$. Also note, that REMARK 4.3.3 is still valid in this setup, again the definition of the set $U_n^+(u(t))$ has to be adopted.

4.3.3 Using the Structure in $\mathcal{U}_f$

In chapter 3 we cited also two methods of using the Toeplitz structure in $\mathcal{U}_f$ to obtain an estimate of $\hat{\beta}_z$, which have been proposed in (Peternell *et al.*, 1996). In this section we will extend the results of the previous section also to the case, where these more complicated versions of the main algorithm are used.

In this section, only the case where f is fixed will be discussed. The evaluations in this case already are quite complex, thus it is not attempted to provide a result for $f \to \infty$ in this case. From the description of the algorithm in chapter 3 we recall the three steps: In the first step, we obtain an estimate $\hat{\beta}_z$ of β_z, which is used in the second step as follows: A weighted version of $\hat{\beta}_z$ is decomposed as $\hat{W}_f^+ \hat{\beta}_z \hat{W}_p^- = \hat{U}_n \hat{\Sigma}_n \hat{V}_n^T + \hat{R} = \hat{W}_f^+ \hat{\mathcal{O}}_f \hat{\mathcal{K}}_p \hat{W}_p^- + \hat{R}$, where the largest n singular values are retained and the remaining singular values (contained in $\hat{R}$) are neglected. Then in the last step, the estimate $\hat{x}_t = \hat{\mathcal{K}}_p Z_{t,p}^-$ is used to obtain estimates of the system matrices, where $\hat{\mathcal{K}}_p = \hat{\Sigma}^{1/2} \hat{V}_n^T (\hat{W}_p^-)^{-1}$.

Up to now, we did not take into account the structure of $\mathcal{U}_f$. The basic version of the main algorithm gets rid of the effects of $U_{t,f}^+$ by simply projecting onto the orthogonal complement:

$$\hat{\beta}_z = (\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u} \hat{\Gamma}_{u,u}^{-1} \Gamma_{u,z})(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u} \hat{\Gamma}_{u,u}^{-1} \hat{\Gamma}_{u,z})^{-1}$$

In the two other methods proposed in (Peternell *et al.*, 1996), the effects of the future of the inputs are estimated and subtracted: In the first method, we obtain an estimate of β_z as

$$\hat{\beta}_z^c = \arg\min_{b_z, L_f} \sum_{j=p+1}^{T-f} \|Y_{t,f}^+ - b_z Z_{t,p}^- - L_f U_{t,f}^+\|_2^2$$

where L_f is subject to the constraint stated in equation (3.5) and the additional superscript ^{c} stands for *constrained* regression.

In the second method, the estimate is obtained as

$$\hat{\beta}_z^i = (\hat{\Gamma}_{y,z} - \mathcal{U}_f(\hat{A}, \hat{B}, \hat{C}, \hat{D}) \hat{\Gamma}_{u,z}) \hat{\Gamma}_{z,z}^{-1}$$

where in the notation it is emphasized, that the estimate $\hat{\mathcal{U}}_f$ is built using some preliminary estimates $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$. These estimates could be obtained e.g. by the basic version of the main subspace procedure or by `MOESP`. Here the additional superscript ^{i} refers to *iterative* method, which will also be called *two stage method* in the sequel.

For the analysis of these two methods, all we have to do, is to replace the old expressions for $\hat{\beta}_z$ by the new expression ($\hat{\beta}_z^c$ and $\hat{\beta}_z^i$ respectively) in equation (4.24). The two essential properties of the estimate $\hat{\beta}_z$, which were used in the proof for the basic version are:

$$\begin{aligned} \|\beta_z - \mathcal{O}_f\mathcal{K}\| &= o(T^{-1/2}) \\ \sqrt{T}\mathrm{vec}[\hat{\beta}_z - \beta_z] &\xrightarrow{d} Z \end{aligned}$$

where β_z denotes the matrix obtained by using the true covariances instead of sample covariances, $\xrightarrow{d}$ denotes convergence in distribution and Z is a (infinite dimensional) multivariate normally distributed random variable with zero mean and finite variance (in the sense of Lemma 4.1.4): The second equation has been shown by linearizing the mapping attaching sample covariances to matrices $\hat{\beta}_z$:

$$\mathrm{vec}[\hat{\beta}_z - \beta_z] = L_{f+p}(\hat{g}_{T,f+p} - g_{f+p}) + o_P(T^{-1/2})$$

where L_{f+p} denotes a matrix, whose rows are of bounded ℓ_1 norm (uniformely in $f+p$) and converge (in ℓ_1 norm) to vectors with elements decreasing exponentially. Thus it is sufficient to show the same for $\hat{\beta}_z^c$ ($\hat{\beta}_z^i$ respectively).

For the constrained least squares method we use the notation introduced in (Peternell *et al.*, 1996): Define the matrix N_{22} as follows:

$$\mathrm{vec}[\hat{L}_f] = N_{22}\mathrm{vec}[\hat{L}(0), \hat{L}(1), \cdots, \hat{L}(f-1)]$$

For a definition of these matrices see equation (3.5). Then it follows from straightforward calculations, that

$$\mathrm{vec}[\hat{\beta}_z^c] = \mathrm{vec}[\hat{\Gamma}_{y,z}\hat{\Gamma}_{z,z}^{-1}] - [(\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u}) \otimes I_{fs}]N_{22}\hat{N}_{u,z}N_{22}^T\mathrm{vec}[\hat{\Gamma}_{y,u} - \hat{\Gamma}_{y,z}\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u}] \tag{4.26}$$

where $\hat{N}_{u,z} = [N_{22}^T(\hat{\Gamma}_{u,u} - \hat{\Gamma}_{u,z}\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u})N_{22}]^{-1}$ has been introduced to shorten the notation. The arguments provided in the last section directly imply the asymptotic normality of the contribution of the term $\mathrm{vec}[\hat{\Gamma}_{y,z}\hat{\Gamma}_{z,z}^{-1}]$ to $\hat{m}_p$ (see equation (4.25)). For the second term note, that $\mathrm{vec}[\hat{x}_1] = \mathrm{vec}[\hat{\beta}_z^c(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})(\hat{\beta}_z^c)^T] = [\hat{\beta}_z^c(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z}) \otimes I_{fs}]\mathrm{vec}[\hat{\beta}_z^c]$ and thus the contribution of the second term is equal to

$$[\hat{\beta}_z^c(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u} \otimes I_{fs}]N_{22}\hat{N}_{u,z}N_{22}^T\mathrm{vec}[\hat{\Gamma}_{y,u} - \hat{\Gamma}_{y,z}\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u}]$$

Note, that all matrices in square bracketts are of fixed dimension. The asymptotic nonsingularity of $\hat{N}_{u,z}$ has been shown in (Peternell *et al.*, 1996). Then application of the previously obtained results implies the asymptotic normality of the contribution due to this term: Note, that only the matrices, which have the subscript $_z$ attached, contain matrices of dimensions growing in size with $p \to \infty$. However all these expressions are analogous to expressions, which already occured several times in previous sections and hence the derivation is omitted. Thus we have shown, that $\sqrt{T}(\hat{m}_p - m_p)$ is asymptotically normal also in this case.

For the two stage method, we assume, that the estimated system matrices $(\hat{A}, \hat{B}, \hat{C}, \hat{D})$ are such, that asymptotically $(\hat{A}, \hat{B}, \hat{C}, \hat{D}) = L_{f,p}\hat{g}_{T,f+p} + o_p(T^{-1/2})$, where $L_{f,p} \in \mathbb{R}^{f^2s^2 \times (f+p)s^2}$ has rows converging in ℓ_1 norm to vectors having elements decreasing exponentially. This has been shown for the case, where (A, B, C, D) are estimated using the main class of subspace algorithms (see the proof in the last section). Since in this case $\hat{\beta}_z^i = (\hat{\Gamma}_{y,z} - \mathcal{U}_f(\hat{A}, \hat{B}, \hat{C}, \hat{D})\hat{\Gamma}_{u,z})\hat{\Gamma}_{z,z}^{-1}$ holds, the asymptotic normality for $\sqrt{T}(\hat{m}_p - m_p)$ is straightforward to prove.

Thus in order to complete the proof, we still have to show, that $\sqrt{T}\|m_p - m_0\| \to 0$ and that the mapping ψ is differentiable at the true system. In order to prove the first point, it is sufficient to prove, that $\sqrt{T}\|\beta_z^c - \mathcal{O}_f\mathcal{K}\|_2 \to 0$ ($\sqrt{T}\|\beta_z^i - \mathcal{O}_f\mathcal{K}\|_2 \to 0$ respectively), where β_z^c (β_z^i) denotes the result of the regressions, if the true covariances are used instead of the estimated covariances. For the constrained regression approach, this can be seen as follows: In equation (4.26) consider

the second term, where we replace estimated covariances with true covariances. The essential part of this term is $\Gamma_{y,u} - \Gamma_{y,z}\Gamma_{z,z}^{-1}\Gamma_{z,u} = \mathcal{U}_f(\Gamma_{u,u} - \Gamma_{u,z}\Gamma_{z,z}^{-1}\Gamma_{z,u}) + \mathcal{O}_f(A_0 - K_0E_0^{-1}C_0)^p(\Gamma_{x^-,u} - \Gamma_{x^-,z}\Gamma_{z,z}^{-1}\Gamma_{z,u})$, where $\Gamma_{x^-,u} = \mathbb{E}x(t-p)(U_f^+(t))^T$ and $\Gamma_{x^-,z} = \mathbb{E}x(t-p)(Z_p^-(t))^T$. The second term in this expression can be neglected, since its Frobenius norm is of order $o(T^{-1/2})$ under our assumptions on the processes $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ and on the true system (k_0, l_0). Thus we obtain:

$$\begin{aligned}
\mathrm{vec}[\beta_z^c] &= \mathrm{vec}[\Gamma_{y,z}\Gamma_{z,z}^{-1}] - (\Gamma_{z,z}^{-1}\Gamma_{z,u} \otimes I)N_{22}N_{u,z}N_{22}^T\mathrm{vec}[\mathcal{U}_f(\Gamma_{u,u} - \Gamma_{u,z}\Gamma_{z,z}^{-1}\Gamma_{z,u})] + o(T^{-1/2}) \\
&= \mathrm{vec}[\Gamma_{y,z}\Gamma_{z,z}^{-1}] - (\Gamma_{z,z}^{-1}\Gamma_{z,u} \otimes I)\mathrm{vec}[\mathcal{U}_f] + o(T^{-1/2}) \\
&= \mathrm{vec}[\Gamma_{y,z}\Gamma_{z,z}^{-1}] - \mathrm{vec}[\mathcal{U}_f\Gamma_{u,z}\Gamma_{z,z}^{-1}] + o(T^{-1/2}) \\
&= \mathrm{vec}[\mathcal{O}_f\mathcal{K}_p + \mathcal{O}_f(A_0 - K_0E_0^{-1}C_0)^p\Gamma_{x^-,z}\Gamma_{z,z}^{-1}] + o(T^{-1/2})
\end{aligned}$$

and thus $\sqrt{T}\|\beta_z^c - \mathcal{O}_f\mathcal{K}_p\| \to 0$. Here equality up to terms of order $o(T^{-1/2})$ is understood in the sense, that the ℓ^2 norm of the difference is $o(T^{-1/2})$.

For the two stage method, the evaluations are simpler, if we use the main class of algorithms in the first step. From the proof of Theorem 4.3.1 we obtain $\|\mathrm{vec}(A_p, B_p, C_p, D_p) - \mathrm{vec}(A_0, B_0, C_0, D_0)\| = o(T^{-1/2})$. Here (A_p, B_p, C_p, D_p) denotes the system $\psi(m_p)$ obtained in the first stage of the two stage procedure. Thus we obtain $\|\mathcal{U}_f(A_p, B_p, C_p, D_p) - \mathcal{U}_f(A_0, B_0, C_0, D_0)\| = o(T^{-1/2})$, if we use the main class of algorithms in the first step. Conditions for other methods, which still ensure the result to hold, are stated in REMARK 4.3.4. Continuing, we have shown, that

$$\begin{aligned}
\beta_z^i &= (\Gamma_{y,z} - \mathcal{U}_f(A_p, B_p, C_p, D_p)\Gamma_{u,z})\Gamma_{z,z}^{-1} \\
&= \mathcal{O}_f\mathcal{K}_p + (\mathcal{U}_f(A_0, B_0, C_0, D_0) - \mathcal{U}_f(A_p, B_p, C_p, D_p)\Gamma_{u,z}\Gamma_{z,z}^{-1} + o(T^{-1/2}) \\
&= \mathcal{O}_f\mathcal{K}_p + o(T^{-1/2})
\end{aligned}$$

holds for the two stage procedure, where in the first step one of the algorithms out of the main class is chosen. Again equality up to terms of order $o(T^{-1/2})$ means, that the ℓ^2 norm of the difference is of order $o(T^{-1/2})$. Thus also in this case we obtain the result $\sqrt{T}\|m_p - m_0\| \to 0$.

Finally the differentiability depends only on the limit of the estimate $\hat{\beta}_z$ and not on the method of estimating $\mathcal{U}_f$. Thus for the constrained regression method, the CLT has been shown to hold on the set $U_n^+(u(t))$, defined just before Theorem 4.3.1. For the two stage method there is another subtlety, which has to be noted: It is possible to use a different weighting matrix in the both stages and even the choice of f could be changed, i.e. we could use e.g. $f \to \infty$ in the first stage and then e.g. using the knowledge on the state dimension estimated in the first step, we could use $f \geq n$ and fixed in the second stage. The effect will be the following: In the two stage method, we perform two SVD's. Thus in order to ensure the asymptotic normality for the estimates in the first step, we require $(k_0, l_0) \in U_n^+(u(t), f_1, (W^+)_1, (W^-)_1)$ for the choices of f_1 and $(W^+)_1$ as well as $(W^-)_1$ taken in the first stage, where now the dependence of the set $U_n^+(u(t), f_1, (W^+)_1, (W^-)_1)$ on the truncation index f and the weighting matrices has been made explicit in the notation. In order for the second stage to provide estimates, which fulfill a CLT, we have to assume, that $(k_0, l_0) \in U_n^+(u(t), f_2, (W^+)_2, (W^-)_2)$. Thus we obtain the assumption $(k_0, l_0) \in U_n^+(u(t), f_1, (W^+)_1, (W^-)_1) \cap U_n^+(u(t), f_2, (W^+)_2, (W^-)_2) = U_n^+(u(t))^2$, which however is still a generic subset of the set M_n^j, since the intersection of two generic subsets is also generic. Thus we have proved the following result:

Theorem 4.3.3 *Let $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ fulfill the assumptions of Theorem 4.3.1. If the estimation is performed using the constrained regression method, the results of Theorem 4.3.1 and Corollary 4.3.1 still hold under the conditions that:*

- $(k_0, l_0) \in U_n^+(u(t), f, W^+, W^-)$,
- $f \geq n$ *is fixed*
- p *fulfills the restrictions in Theorem 4.3.1.*

If the two stage method is used, the results still hold under the conditions that:

- *In the first step a subspace method is used, which fulfills the conditions of Theorem 4.3.1 or Theorem 4.3.2.*
- $(k_0, l_0) \in U_n^+(u(t))^2$.
- $f \geq n$ *fixed*
- p *fulfills the restrictions of Theorem 4.3.1.*

Again the statements of Corollary 4.3.1 can be seen easily. The consistency part has been proved in (Peternell *et al.*, 1996).

The same remark as below every result in this chapter is in order: The asymptotic variance of the estimates, the sets $U_n^+(u(t))$ and $U_n^+(u(t))^2$ as well as the limiting realization of the true pair depend on the particular choices of f and the weighting matrices, however this is not emphasized in the notation. Also the asymptotic variance V depends on the fourth moments of the noise $\eta(t)$. Again the asymptotic variance can be approximated on a computer and thus compared for given system. And again the expressions seem to be too complicated to be evaluated analytically. There are no results on an analytical basis, on whether the more complicated methods (constrained regression, two stage procedure) indeed result in a decrease of the asymptotic variance in some situations. However, simulation results and calculations suggests, that this could be the case for coloured inputs $(u(t))_{t\in\mathbb{Z}}$ (comp. Peternell *et al.*, 1996, and also the calculations in section 5.3.3). Since the difference between the more complicated methods and the basic algorithm only lies in the extraction of the part due to the future of the inputs (i.e. $\mathcal{U}_f U_{t,f}^+$), it is not surprising, that for white noise inputs, the asymptotic distribution of the three methods coincides, which has already been noted in (Peternell *et al.*, 1996):

Theorem 4.3.4 *Let $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ fulfill the assumptions of Theorem 4.3.3, where $(u(t))_{t\in\mathbb{Z}}$ is white noise. Also let the truncation indices f and p fulfill the assumptions of Theorem 4.3.3. Then the following holds:*

- *The asymptotic distribution of the estimates of the system matrices obtained by the constrained regression approach coincide with the estimates obtained by the basic approach.*
- *If in the two stage method, the estimates of the first stage fulfill the condition $\|vec[\hat{A}_T - S_T A_0 S_T^{-1}, \hat{B}_T - S_T B_0, \hat{C}_T - C_0 S_T^{-1}, \hat{D}_T - D_0]\| = o(Q_T p^k)$ a.s. for some $k < \infty$, where $0 < c_1 < |S_T| < c_2 < \infty$ for some constants c_1 and c_2, then also the estimates obtained with the two stage methods are asymptotically equivalent.*

PROOF: As has been noted already, the various estimates differ only in the treatment of the future of the input process, i.e. in the estimation of $\hat{\beta}_z$. Thus it is sufficient to show, that $\sqrt{T}\|\hat{\beta}_z - \hat{\beta}_z^c\|_{Fr} \to 0$ or $\sqrt{T}\|\hat{\beta}_z - \beta_z^i\|_{Fr} \to 0$ respectively.

Consider the case of the constrained regression first. It is possible to find a (full rank) matrix R, such that the restrictions on the matrix $\hat{L}_f$ are equivalent to $R\text{vec}[\hat{L}_f] = 0$. Then it follows from the theory of optimization under equality constraints, that the following equation holds:

$$\text{vec}[\hat{\beta}_z^c, \hat{\beta}_u^c] = \text{vec}[\hat{\beta}_z, \hat{\beta}_u] - (\hat{\Gamma} \otimes I)^{-1} R^T (R(\hat{\Gamma} \otimes I)^{-1} R^T)^{-1} R\text{vec}[\hat{\beta}_z, \hat{\beta}_u]$$

Here $\hat{\beta}_u^c$ denotes the estimate of $\mathcal{U}_f$ resulting from the constrained regression approach and $\hat{\beta}_u$ denotes the corresponding estimate of the unconstrained regression. $\hat{\Gamma} = \langle \begin{pmatrix} Zt, p^- \\ U_{t,f}^+ \end{pmatrix}, \begin{pmatrix} Zt, p^- \\ U_{t,f}^+ \end{pmatrix} \rangle$.

Since $\hat{\beta}_z^c$ is unrestricted, R has the form $R = [0, R_2]$. Also the inverse of $R(\hat{\Gamma} \otimes I)^{-1} R^T$ exists under our assumptions on the processes $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$, since we may choose R to be of full rank. $R\text{vec}[\hat{\beta}_z, \hat{\beta}_u] = R_2\hat{\beta}_u = R_2(\hat{\beta}_u - \mathcal{U}_f)$, since $R_2\mathcal{U}_f = 0$. Now it follows from the uniform convergence of the sample covariances, that $\|R_2\hat{\beta}_u\|_{Fr} = O(p^2 Q_T)$. Finally consider the term $(\hat{\Gamma} \otimes I)^{-1} R^T$. $\hat{\Gamma}$ is composed of 4 terms: $\hat{\Gamma}_{z,z}, \hat{\Gamma}_{z,u}, \hat{\Gamma}_{u,z}$ and $\hat{\Gamma}_{u,u}$. Due to the assumed white noise property of $(u(t))_{t\in\mathbb{Z}}$, we obtain that $\|\hat{\Gamma}_{z,u}\|_{Fr} = O(Q_T p)$. Since we are only interested in the estimation of β_z we may only consider the first part of the term, and since $R = [0, R_2]$ we are only interested in the north-east corner of the inverse $(\hat{\Gamma} \otimes I)^{-1}$. This can be seen to be equal to $-(\hat{\Gamma}_{z,z} \otimes I)^{-1}(\hat{\Gamma}_{z,u} \otimes I)[(\hat{\Gamma}_{u,u} - \hat{\Gamma}_{u,z}\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{u,z}) \otimes I]^{-1}$. Since the norm of all other terms is bounded uniformely in p we conclude, that the norm of the difference $\text{vec}[\hat{\beta}_z^c - \hat{\beta}_z]$ is of order $o(p^3 Q_T^2) = o(T^{-1/2})$ a.s. and thus the both terms are asymptotically equivalent.

For the two stage method, the assumptions imply, that the Markoff parameters CA^iB are estimated with accuracy $O(p^k Q_T)$ a.s. Recall that

$$\hat{\beta}_z^i = (\hat{\Gamma}_{y,z} - \mathcal{U}_f(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}_T)\hat{\Gamma}_{u,z}\hat{\Gamma}_{z,z}^{-1}$$

and that $\hat{\beta}_z = (\hat{\Gamma}_{y,z} - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\Gamma_{u,z})(\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{-1}$. Then the result follows from straightforward evaluations noting that $\|\mathcal{U}_f(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}_T) - \hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1}\|_{Fr} = O(Q_T p^k)$ and $\|\hat{\Gamma}_{u,z}\|_{Fr} = O(Q_T p)$. □

REMARK 4.3.4 The formulation of the Theorems suggest, that the main algorithm should be used to perform the estimation in the first step. This restriction has only been used for notational reasons. From the proof given above, it is relatively clear, that also other estimates may be used as the initial guess needed: The key property concerning the CLT is the existence of a matrix $L_{f,p}$ such that $\sqrt{T}\text{vec}[\hat{A}_T - A_0^i, \hat{B}_T - B_0^i, \hat{C}_T - C_0^i, \hat{D}_T - D_0^i] = \sqrt{T}L_{f,p}\,\text{vec}[\hat{g}_{T,f+p} - g_{f+p}] + o_P(1)$ holds, where $L_{f,p}$ has rows of bounded ℓ_1 norm and converging to rows with elements decreasing exponentially. Here $(A_0^i, B_0^i, C_0^i, D_0^i)$ denotes a particular representation of the true transfer function, which need not be the same as (A_0, B_0, C_0, D_0). The asymptotic normality of the system matrix estimates obtained from the two stage procedure then holds, if this property holds for an algorithm for obtaining the initial estimates $(\hat{A}_T^i, \hat{B}_T^i, \hat{C}_T^i, \hat{D}_T^i)$. For a.s. consistency of the estimates obtained by the two step procedure the initial estimates only have to be a.s. consistent.

4.4 Summary of the Results

In this chapter we presented many results on the asymptotic properties of subspace algorithms in various situations. The bottom line of all these results can be stated as follows: Using the main class of algorithms will for most transfer functions (k_0, l_0) (in the sense, that the results hold on generic subsets) produce estimates, which are consistent (a.s.) and fulfill a central limit theorem, whenever the weighting matrices are chosen out of a special class of block Toeplitz lower triangular matrices or according to `CCA`, and if the truncation indices fulfill a number of restrictions. The significance of these results lies in the fact, that for given system the asymptotic variance can be calculated and thus we could use this knowledge to obtain e.g. confidence regions for certain parameters. Another application is the comparison of the various procedures, corresponding to the asymptotic accuracy of the estimates. This will be done in the next chapter.

The main points of criticism about the results presented in this chapter are probably the following two:

- The results assume that a true system of finite order exists and that moreover this true order is known. In response to the second point, we refer the reader to section 5.2, where various procedures for the estimation of the order are discussed. There it will be shown, that the order can be estimated consistently a.s. in the case, where the true system indeed is finite dimensional. Section 5.3 presents some extensions of the results obtained so far, also to the

case, where estimation is performed using a lower order than the true system order. Thus the assumption on the knowledge of the system order is thought to be not too restrictive, since similar results also hold for other situations, such that the results in this section may be seen as preliminary.

The assumption of rationality of the true transfer functions however is more severe. In this respect it would be desirable to clarify the asymptotic properties of the subspace estimates, when the true system is infinite dimensional. In this situation, some of the tools presented in this chapter could be of importance. However, we are not able to present any results for the case of infinite dimensional systems.

- The second point of criticism relates to the comparison of the various methods: The expressions obtained so far for the asymptotic variance are too complicated to be analyzed on an analytical basis. Thus we may only compare the various approaches for given system, which of course restricts the possibility to give guidelines for the user. The next chapter will try to provide some evidence on how to set the user choices in order to use the full strength of subspace methods. However the lack of explicit results on a comparison of the asymptotic properties of the various algorithms and user choices in fact is a disadvantage of the methods presented in this chapter.

Finally we also want to mention, that the restrictions on the input process $(u(t))_{t\in\mathbb{Z}}$ are thought to be too restrictive. It seems to be desirable to generalize the results by allowing for other types of input processes, including e.g. inputs, which have been pretreated for the extraction of a nonzero expectation or deterministic trend components, to give just one example.

Chapter 5

Some Further Considerations

This chapter will deal with problems, which have not been addressed up to now in the thesis: The main part of this thesis describes the asymptotic properties of subspace estimates under some conditions on the design variables: The design variables are the truncation indices, the order of the system and the choice of the weighting matrices. Up to now, most of the results were shown to hold in many different situations. In this chapter we will try to develop and motivate some algorithms to choose these design variables in accordance to data and modeling objectives. This will be done on a theoretical basis relying on asymptotic theory and also from a more practical point of view, demonstrating the finite sample properties in simulations studies. Therefore the next three sections deal with the truncation indices, the estimation of the order, and the choice of the weighting matrices. Note, that the effects of the design variables are interconnected, so that we will obtain different recommendations for the various design variables for different choices of the other variables. This chapter will be closed with two somewhat technical issues: Section 5.4 will investigate stability and minimum-phase properties of the estimates, and section 5.5 shows, that the asymptotic distribution of the estimates of the transfer function $\bar{k}_0(e^{i\lambda} = k_0(e^{i\lambda})E_0^{-1}$ does not depend on the distribution of the white noise sequence $\varepsilon(t)$ fulfilling the standard assumptions of this thesis. This gives a justification for neglecting the distribution in the simulations and calculations provided in this chapter.

5.1 The Choice of the Truncation Indices

From the description of the algorithm it is clear, that the truncation indices f and p are design variables, which have to be set by the user for all subspace algorithms. Having a freedom to choose is only useful, if we have some knowledge of the effect of the choices. Up to now, there have not been too many suggestions in the thesis on how to make the choice of the indices based on the data. From a point of view of asymptotical theory, as presented in the theorems in chapter 4, there are two restrictions for the index p: The truncation index p has to tend to infinity at a certain rate with $T \to \infty$, where the rate is bounded from below by $-\frac{d \log T}{2 \log |\rho_0|}$ for some $d > 1$ and from above by $C(\log T^a), a < \infty$. In practical circumstances, both restrictions are not much of a help for the user, since the upper bound contains a not further specified constant, which can be arbitrarily big, and the lower bound depends on the true system, which is not known. Concerning the row truncation index f we have either $f \geq n$ fixed and finite, or $f \to \infty$ with prescribed rates for $T \to \infty$. Again we see the same problems. Additionally the choice of f as fixed and finite invokes the question, whether it is advisable to choose f relatively large or relatively small, once an upper bound of the system order is known (or postulated). Thus we will in this section try to find some hints on how to choose the indices in correspondence to the asymptotic theory as well as to finite data length.

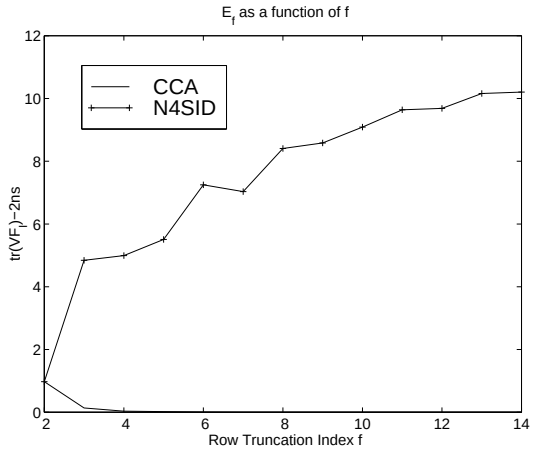

Figure 5.1: Here the measure $E_f = tr(V_{ech} I_0) - 2ns$ is plotted for $f = 2, \cdots, 14$ and two different weighting schemes: `CCA` (—) and `N4SID` (-x-).

5.1.1 An illustrative Example

To start with, we will use an example to demonstrate some of the possible effects: Consider the following SISO system, with no observed inputs.

$$A = \begin{bmatrix} 0 & 1 \\ -0.7 & 0.5 \end{bmatrix}, K = \begin{bmatrix} 1.3 \\ 0.3 \end{bmatrix}, C = [1, 0], E = 1$$

This system has also been investigated in (Bauer *et al.*, 1997b). Figure 5.1 shows a measure of the asymptotic efficiency for the estimation of $\bar{k}_0 = k_0 E_0^{-1}$: Let I_0 denote the Fisher information matrix (defined in Theorem 2.4.2). Theorem 2.4.2 states, that the inverse of I_0 is equal to the Cramer-Rao bound and also equal to the asymptotic variance of the estimates of the parameters corresponding to $\bar{k}_0$ obtained by using the maximum likelihood approach. Of course the Fisher information matrix as well as the asymptotic covariance of the ML estimates depends on the actual parametrization chosen. Corollary 4.1.2 shows the effect of transformations of the state coordinates corresponding to the asymptotic distribution. This makes it possible to compare the asymptotic variance of the estimates of $\bar{k}_0$ obtained by using subspace algorithms, which afterwards have been transformed into echelon coordinates, to the Fisher information matrix. Let V_{ech} denote the asymptotic variance of the transformed subspace estimates. Then clearly $V_{ech} \geq I_0^{-1}$ holds. Thus $tr(V_{ech} I_0) = tr((V_{ech} - I_0^{-1}) I_0) + 2ns \geq 2ns$, since $(V_{ech} - I_0^{-1})$ is positive semidefinite and I_0 is positive definite. Therefore we use as a measure of asymptotic optimality the quantity $E_f = tr(V_{ech} I_0) - 2ns \geq 0$. This is the quantity, which is plotted in Figure 5.1 for various values of f and two different weighting schemes. It is clearly visible, that for `CCA` for increasing f the measure E_f decreases, whereas for `N4SID` the opposite is true. Also observe, that for `CCA` and large f the quantity E_f is close to zero (We obtain $E_{15} = 0.0019$). This shows, that for certain choices of the weighting matrix $\hat{W}_f^+$ it seems to be better to choose f as close to the true order as possible, whereas for `CCA` (at least in the example) increasing f also leads to better accuracy of the estimates. The same behaviour is observed in a different example in (Bauer *et al.*, 1998).

This is the justification for including both cases (f fixed and $f \to \infty$) in the thesis. The discussion provided above only gives hints about choosing the truncation index f, but we did not deal with the choice of p up to now. For the result on the asymptotic distribution to hold, p has to fulfill certain bounds, depending both on the sample size as well as the true system. Note, that in the case of `CCA` the restrictions on f (for $f \to \infty$) coincide with the restrictions on p. Thus in the next section we will investigate a possible choice in the `CCA` framework (see also Peternell, 1995; Deistler *et al.*, 1995), where $f = p$ is estimated from the data.

5.1.2 Proposition for CCA

Recall, that the requirement on the truncation index p is that it increases with the sample size T within two bounds:

$$-\frac{d}{2}\frac{\log T}{\log|\rho_0|} \leq p \leq (\log T^a)$$

for some constants $d > 1$ and $a < \infty$. Here again $\rho_0 = \lambda_{max}(A_0 - K_0 E_0^{-1} C_0)$. Let the conditions of Theorem 2.4.4 be fulfilled. Then we may obtain an estimate, $\hat{p}$ say, using the BIC estimate of a long autoregression as described in Theorem 2.4.4. The upper bound on the increase of $\hat{p}$ then holds due to the imposed upper bound in the estimation of the order of the autoregression. The lower bound holds a.s. due the result in Theorem 2.4.4 for $d > 1$ (In fact this is the reason for the strict inequality). Thus in the case of CCA a simple procedure would be to use $f = p = d\hat{p}$ for some given constant d. Alternatives could be to use rather AIC than BIC to estimate the truncation index, since the estimates of AIC are always larger than the ones obtained with BIC and we want to ensure, that the lower bound is not violated.

Although this procedure is mainly motivated for CCA, it is obvious, that the choice of p in any procedure can be made according to the suggestion given above. Corresponding to the choice of f it is not so clear, that the same procedure will lead to good results (compare section 5.1.1). However, if there is no a priori knowledge on the system order (such as a known upper bound), it seems to be a useful procedure. Note, that the choice of f restricts the possible order of the estimates to be smaller than fs. Thus choosing f too small may lead to bad estimates (If $fs < n$ the estimates will not be consistent for obvious reasons). Thus our suggestion for a fully automatic procedure (without investing any a-priori knowledge on the system order) would be to use $f = p = d\hat{p}$ in any case, where $\hat{p}$ denotes the estimated order in an autoregression using either AIC or BIC, where $d > 1$ should be 'small', even though this might not lead to the validity of the asymptotic results.

5.1.3 Simulations results

Corresponding to the vague statement at the end of the last section, we will in this section investigate the finite sample properties of the CCA procedure, in a situation, where the true system order is known comparing the procedure presented above using either AIC or BIC for various values of d (In section 5.2 we will deal with the question of estimating the order.).

For this purpose, we generated 1000 time series of sample length $T = 100$ by filtering Gaussian white noise (mean zero and variance unity) through the system used in section 5.1.1. The truncation indices were calculated using AIC and BIC order selection of the long autoregression, resulting in an estimate $\hat{p}_{AIC}$ and $\hat{p}_{BIC}$ respectively. Then the CCA subspace algorithm was applied with $f = p = d\hat{p}_{AIC}$ and $f = p = d\hat{p}_{BIC}$ respectively, where d was chosen to be either $d = 1$, $d = 2$ or $d = 8$. Typical estimates of $\hat{p}_{AIC}$ were $\hat{p}_{AIC} = 3$ or $\hat{p}_{AIC} = 4$ and the same is true for $\hat{p}_{BIC}$. Thus for $d = 1$ or $d = 2$ respectively typically f and p will not be large compared to the sample length, whereas for $d = 8$ the typical values of $f = p$ are $f = p = 32$ and thus the first 63 covariances enter the estimation algorithm, which is very large compared to the sample size $T = 100$. The results can be seen in Figure 5.2. The left picture shows the modulus of the deviation of the sample mean from the true transfer function. The picture on the right shows the mean of the squared modulus of the deviation of the estimated transfer function from the true transfer function.

Corresponding to the mean values, we observe, that for BIC the estimates using $d = 1$ and $d = 8$ show a considerably bigger bias, than in the case $d = 2$. For AIC this situation is less pronounced (undocumented). Corresponding to the mean squared deviation to the true transfer function only the case $d = 8$ gives lower accuracy, while the two other cases are very similar with small advantages for $d = 2$. This time the plot shows the results for AIC, however there is no difference to be observed in the case of BIC. Thus (at least) in this example $d = 2$ seems to be a reasonable choice, whereas $d = 1$ especially in the BIC case leads to a risk of using too low truncation indices. Up to now, however, we only dealt with the idealized situation, that the true

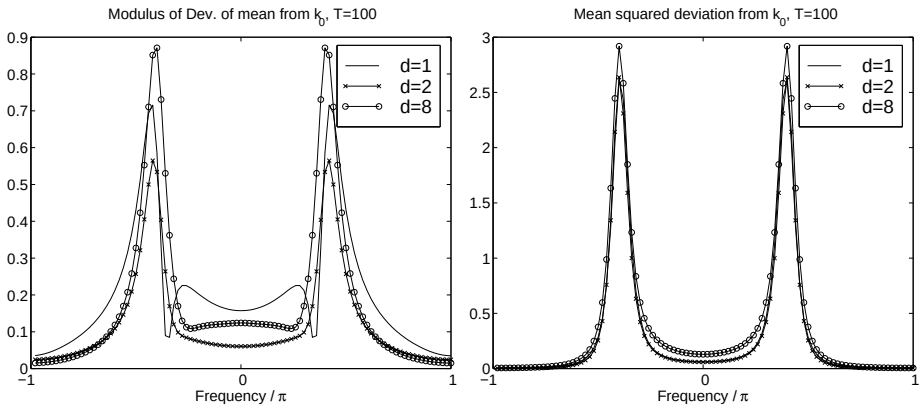

Figure 5.2: Here the absolute value of the deviation of the sample mean to the true transfer function is shown in the left picture. $f = p = d\hat{p}_{BIC}$ is used for the estimation, where $d = 1$ (—), $d = 2$ (-x-) and $d = 8$ (-o-). The right image shows the mean squared error of the estimates of the transfer function with respect to the true transfer function. $f = p = d\hat{p}_{AIC}$ is used, where $d = 1$ (—), $d = 2$ (-x-) and $d = 8$ (-o-). For both plots 1000 time series of sample length $T = 100$ have been generated and estimation is performed using `CCA`.

order of the system is known. The choice of the parameter d will also be important for the accuracy of the order estimation procedures presented in the next section.

5.2 Estimating the Order

This section is devoted to the estimation of the order of the model in the context of subspace identification. The discussion will deal with both, the case of no observed inputs and afterwards with the case of additional exogenous inputs. We will review two existing methods and then propose and motivate two procedures, which have not been suggested in the literature so far, to the best of the authors knowledge, despite the fact, that they are quite obvious.

5.2.1 Estimation Procedures

In the literature the topic of estimation of the order in the context of subspace identification methods is not addressed to very often: Most of the authors suggest to find the order of the model by visual inspection of a plot of the logarithms of the singular values (see e.g. Van Overschee and DeMoor, 1994; Verhaegen, 1994): As follows from the discussion in Chapter 4 the singular values will converge to the true singular values, if there exists a true system of finite order, and if the truncation indices fulfill the assumptions given in the various theorems. Thus for large T it will be possible to group the singular values into two groups, one which corresponds to the nonzero singular values, and the other group, which is only nonzero due to the noise effects. In between these two groups there will be a gap, and this gap will become more and more pronounced as the sample size increases. This heuristic argument can also be formalized to give an automatic procedure of order estimation.

This seems to be the idea behind the order estimation procedure implemented in the `N4SID` version of the `MATLAB` system identification toolbox of (Ljung, 1991): Let $\hat{\sigma}_i$ denote the estimate of the i-th singular value, then Ljung specifies the order as the greatest integer $\hat{n}$, such that $\log(\hat{\sigma}_{\hat{n}}) > \frac{1}{2}(\log(\hat{\sigma}_1) + \log(\hat{\sigma}_k)$, where $\hat{\sigma}_1$ denotes the greatest singular value of $\hat{W}_f^+ \hat{\beta}_z \hat{W}_p^-$ and $\hat{\sigma}_k$, where $k = \min\{fs, p(m+s)\}$, the smallest singular value. (Ljung, 1991) uses the weightings according to the `N4SID` procedure.

Another procedure has been proposed and analyzed by (Peternell, 1995): Using the results in Chapter 4 it is possible to show, that $\sum_{j=n+1}^{\min(fs,p(s+m))} \hat{\sigma}_j^2 = O(Q_T^2)$ a.s. for f and p fixed (see also the discussion below). Here $Q_T = (\log\log T/T)^{1/2}$. (Peternell, 1995) suggests to estimate the order as the minimizing argument of the function $\chi(n)$:

$$\chi(n) = \sum_{j=n+1}^{\min(fs,p(s+m))} \hat{\sigma}_j^2 - d(n)C_T/T, 0 \leq n \leq \min\{fs, p(s+m)\} \tag{5.1}$$

where $d(n) = (ps-n)s + (fs-n)s - s^2$ for the case of no observed inputs and $d(n) = (fs-n)(p(m+s)-n)$ for the case of additional observed inputs. From the a.s. convergence result above it follows, that the estimate of n will be consistent a.s. if $C_T/T \to 0, \liminf C_T/\log\log T \to \infty$. Thus $C_T = \log T$ is a possible choice ensuring the consistency of the approach. However, it will be shown in the simulations section, that the accuracy of the estimation of the order depends heavily on the values of f and p. Note, that the (a.s.) consistency result only holds for fixed f and p. In the main class of algorithms of this thesis it is required, that the index p tends to infinity for consistency of the estimates. Thus from an asymptotic point of view, in order to obtain consistent estimates of the system matrices, it would be necessary to perform a two stage procedure, where in the first step some indices f and p are specified and the order is estimated as described above, and then in a second step the algorithm is used with the specified order and a different set of indices f and p, where p in the second step increases with the sample size T. An alternative would be to use a greater penalty term C_T.

Another argument for the insufficiency of these two procedures are the simulations presented in section 5.2.2. There it will be obvious, that both methods do not perform satisfactory, if the indices f and p are estimated from data as described in section 5.1.

The Case of no exogenous inputs

The discussion given above motivates the seek for other possibilities for the estimation of the order. In the following we will deal with the two most obvious ideas and provide also an analysis of some of the properties, without aiming for a complete discussion. This will be done for the case of no additional observed inputs in this section, the case of exogenous inputs will be treated in the next section.

The first procedure is very much related to Peternell's method. The principle is exactly the same: First define the following criterion function:

$$SVC(n) = \hat{\sigma}_{n+1}^2 + \frac{d(n)C_T}{T}, 0 \leq n \leq \min(fs, ps) \tag{5.2}$$

where SVC stands for *singular value criterion*. Here $d(n) = 2ns$ and $C_T > 0$ denotes a penalty function to be specified later on. Although the criterion is formulated analogously to the information criteria specified in equation (2.19) we do not intend to conjecture a connection between the two definitions. The essential part of the criterion is the penalty term $d(n)C_T/T$, which prevents the model builder from choosing a large order. The higher we specify the penalty term, the lower the risk of overestimating the order and the higher the risk of underestimating the order. Using this notation we obtain the following result for the case of no additional observed inputs:

Theorem 5.2.1 *Let the process $(y(t))_{t\in\mathbb{Z}}$ be generated by a true system $k_0 \in U_{n_0}^{(m)}$ for some integer n_0, where the ergodic white noise fulfills the standard assumptions given in the introduction. If the truncation indices fulfill either of the conditions of the Theorems in Chapter 4 and if the order is estimated as the minimizing integer of the criterion function $SVC(n), 0 \leq n \leq \min(fs, ps)$, i.e.*

$$\hat{n} = \arg \min_{0 \leq n \leq \min(fs,ps)} SVC(n)$$

then $\hat{n} \to n$ a.s. if $\lim C_T/T \to 0, \liminf C_T/(fp\log\log T) \to \infty$.

PROOF: The proof only uses the uniform convergence of the sample covariances stated in Theorem A.3.1. Note, that

$$\begin{aligned}\hat{\sigma}_{n+1} &= \|\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2} - \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T\|_2 \\ &\leq \|\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2} - W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-T/2}\|_2 + \|W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-T/2} - \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T\|_2\end{aligned}$$

For the first term, note that $\|\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p}\|_2 \leq \|\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p}\|_{Fr} = \sqrt{fp}O(Q_T)$. Furthermore it has already been stated, that all matrices $\hat{W}_f^+, W_f^+, \hat{W}_p^-$ and W_p^- have finite 2-norm independent of f and p. Then the bound for the first term follows from the properties of the square root of positive definite matrices as used in section 4.2: This shows e.g., that each element of $(\hat{\Gamma}_p^-)^{1/2} - (\Gamma_p^-)^{1/2}$ is of order $O(Q_T)$. Then the exponential decrease in the elements of $\mathcal{H}_{f,p}$ imply the desired bound on the 2-norm. Similar arguments show the result also for $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$. Corresponding to the second term note, that $W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-T/2} = U_{n_0}\Sigma_{n_0}V_{n_0}^T$ for $f, p \geq n_0$. Thus it follows, that for $n < n_0$ the norm of the second term can be bounded by a positive constant for T large enough i.e. $\exists T_0 : \|W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-T/2} - \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T\|_2 > c > 0$ a.s. for $T > T_0$. Since $\lim C_T/T \to 0$, we conclude $\hat{n} \geq n_0$ a.s. for T large enough.

Thus consider $n \geq n_0$. Then

$$\begin{aligned}\hat{\sigma}_{n+1} &= \|\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2} - \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T\|_2 \\ &= \|\hat{U}^T(\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2}) - \begin{bmatrix}\hat{\Sigma}_n\hat{V}_n^T \\ 0\end{bmatrix}\|_2 \\ &\leq \|\hat{U}_2\hat{U}_2^T(\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2})\|_2\end{aligned}$$

where $\hat{U}_2$ denotes the second block column of $\hat{U} = [\hat{U}_{n_0}, \hat{U}_2]$, which corresponds to the omitted singular values. Following (Chatelin, 1983) it can be shown, that $\|\hat{U}_2\hat{U}_2^T - U_2U_2^T\|_2 = \|\hat{U}_{n_0}\hat{U}_{n_0}^T - U_{n_0}U_{n_0}^T\|_2 = O(Q_T\sqrt{f})$, due to the analyticity result given in Theorem B.2.1 (see also Chatelin, 1983, p. 141): Here we use, that $I = \hat{U}_{n_0}\hat{U}_{n_0}^T + \hat{U}_2\hat{U}_2^T$. First consider the case $k_0 \in U_{n_0}^+$: Since the elements of the matrix $\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}\hat{\mathcal{H}}_{f,p}^T(\hat{W}_f^+)^T$, on which the eigenvaluedecomposition is performed, have an error of order $O(Q_T)$ and since the vectors u_i have elements decreasing exponentially (of course in the case of f fixed this is meaningless), the result in this case follows from the equation (B.4), which is shown in the sequel: In this formula terms of the form $u_j^T(\Delta X)u_i$ appear in the first order approximation, where $\Delta X = \hat{X}_{f,p} - X_0$. In our case, it is straightforward to see, that ΔX is a matrix, whose entries are of order $O(Q_T)$. The first order approximation contains two terms: The first term is a finite sum, where terms similar to the ones dealt with above appear. Thus we obtain, that the contribution of the sum to the norm of the error $\hat{u}_i - u_i$ is of order $O(Q_T)$. Consider the contribution of the second term:

$$\frac{1}{\lambda_i}P_0\Delta X u_i = \frac{1}{\lambda_i}(I - U_{n_0}U_{n_0}^T)\Delta X u_i$$

where P_0 denotes the projection onto the nullspace of X_0, i.e. $span\{u_1, \cdots, u_n\}^\perp$. Using the arguments given above it is straightforward to show, that $\|\Delta X u_i\|_2 = O(\sqrt{f}Q_T)$ and thus $\|\hat{u}_i - u_i\|_2 = O(\sqrt{f}Q_T)$. This shows $\|\hat{U}_{n_0}\hat{U}_{n_0}^T - U_{n_0}U_{n_0}^T\|_2 = O(\sqrt{f}Q_T)$ in the case $k_0 \in U_{n_0}^+$. For $k_0 \in U_{n_0}^{(m)} - U_{n_0}^+$ the claim follows from (Chatelin, 1983, Proposition 3.25).

Using $\|\hat{W}_f^+\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1/2} - W_f^+\mathcal{H}_{f,p}(\Gamma_p^-)^{-1/2}\|_2 = O(Q_T\sqrt{fp})$ we obtain $\hat{\sigma}_{n+1} = O(\sqrt{fp}Q_T)$. The result then follows from the fact, that $T(SVC(n) - SVC(n_0))/C_T = (d(n) - d(n_0)) + T(\hat{\sigma}_{n+1}^2 - \hat{\sigma}_{n_0+1}^2)/C_T = 2s(n - n_0) + O(fp\log\log T/C_T)$, since under the assumption on C_T stated in the theorem the second term will tend to zero, while the first term stays positive independently of T with minimum at $n = n_0$. □

REMARK 5.2.1 The condition on the penalty is stronger than the one given in (Hannan and Deistler, 1988) for the information criteria. Also the penalty used in `BIC`*, i.e.* $C_T = \log T$ *is not included in the result, if we allow* f *and* p *tend to infinity within the bounds stated e.g. in Theorem 4.2.1. However, the proof of the Theorem given above is rather brute force, bounding the 2-norm by the Frobenius norm and the result may hold also for weaker requirements on* C_T*. We also want to emphasize, that the consistency of the estimation of the order may not be the only useful criterion to assess the quality of the estimation method, compare the discussion below Theorem 2.4.3.*

REMARK 5.2.2 Also note, that for f *and* p *fixed we obtain the same condition for* C_T *as stated in (Peternell, 1995) for* $\chi(n)$*. However simulation studies (see section 5.2.2) show, that in the case, where* f *and* p *are allowed to tend to infinity and* $C_T = \log T$ *is used both for* $\chi(n)$ *and* $SVC(n)$*, the performance of* $\chi(n)$ *in comparison to* `SVC` *deteriorates for increasing sample size (note that this* C_T *does not fulfill the condition for consistency). The reason for this lies in the fact, that since* $\chi(n)$ *compares the sum of squared singular values to the penalty term, the dimension of the matrix* $\hat{\Sigma}$ *and thus the number of singular values is crucial to the performance, whereas for the* `SVC` *criterion only the largest neglected singular value enters, which seems to be more robust to an increase in the dimensions. It is not difficult to see that for penalty terms, which fulfill* $C_T/(fp\log\log T) \to \infty$*, e.g.* $fp\log T$ *both criteria lead to consistent estimates, however it is believed, that such a high penalty term leads to quite a high risk of underestimating the order in finite samples (compare section 5.2.2).*

REMARK 5.2.3 As a last remark we would like to emphasize, that the theorem holds for any weighting scheme, i.e. independently of the particular choice of $\hat{W}_f^+$*. However, clearly the accuracy of the estimates for different weighting will be different (see also the simulations in section 5.2.2).*

Finally we will also deal with another criterion, which can be used for estimating the order of the system in the context of subspace identification in a computationally appealing way. Note, that we still only consider the case, where no additional observed inputs are present. Recall the definition of the information criteria, presented in section 2.4 (comp. equation (2.19)):

$$IC(n) = \log\det\Omega(\hat{\tau}_n) + \frac{C_T\, d(n)}{T}, 0 \le n \le H_T$$

Here $\Omega(\hat{\tau}_n)$ denotes the innovation variance according to the model corresponding to $\hat{\tau}_n$, where $\hat{\tau}_n$ denotes the maximum likelihood estimate of the true transfer function in the set $\overline{U_n^{(m)}}$. Again $C_T > 0$ denotes a penalty term, $d(n) = 2ns$. The estimate of the order, $\hat{n}$ say, minimizes the above criterion. Some of the properties of this estimator are stated in Theorem 2.4.3 and depend heavily on the penalty term.

In the subspace case, the innovation variance according to the estimated model $(\hat{A}_n, \hat{K}_n, \hat{C}_n, \hat{E}_n) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times s} \times \mathbb{R}^{s\times n} \times \mathbb{R}^{s\times s}$ is estimated as $\hat{E}_n\hat{E}_n^T$, where in the notation we emphasized the order used in the estimation. Thus it seems tempting to replace the information criterion by the following criterion:

$$IVC(n) = \log\det(\hat{E}_n\hat{E}_n^T) + \frac{C_T\, d(n)}{T}, 0 \le n \le \min(fs, ps) \tag{5.3}$$

Here IVC stands for *innovation variance criterion*. Note, however, that we do not want to conjecture, that there are close connections between these two criteria. One attractive feature of this approach lies in the fact, that computationally the calculation of the criterion for different values of n is very cheap: It is straightforward to see, that for the particular choice $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$ the estimate $\hat{x}_t(n) = (\hat{\Sigma}_n)^{1/2}\hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2}Y_{t,p}^-$ leads to $\langle\hat{x}_t(n), \hat{x}_t(n)\rangle = \hat{\Sigma}_n$. Also

$$\langle y_t, \hat{x}_t(n)\rangle = [I,0]\hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-T/2}\hat{V}_n(\hat{\Sigma}_n)^{1/2} = [I,0](\hat{W}_f^+)^{-1}\hat{U}_n(\hat{\Sigma}_n)^{3/2} = [I,0]\hat{\mathcal{O}}_f\hat{\Sigma}_n$$

Thus the estimate of C for any specified order n is equal to the first n columns of the matrix consisting of the first s rows of $(\hat{W}_f^+)^{-1}\hat{U}_n(\hat{\Sigma}_n)^{1/2}$. Concluding we obtain $\hat{E}_n\hat{E}_n^T = \hat{\gamma}(0) - \hat{C}_n\hat{\Sigma}_n\hat{C}_n^T$.

Due to the diagonality of $\hat{\Sigma}$ the calculation of all possible values of $\hat{E}_n\hat{E}_n^T$ can be done recursively in n with performing just one regression for the maximal order.

In order to analyze the asymptotic properties of the estimates of the order obtained by minimizing $IVC(n)$ we will again resort to arguments relying on the a.s. convergence of the sample covariances (see Theorem A.3.1). It has been stated in Corollary 4.1.1, that for each $k \in U_{n_0}^{(m)}$ there exists a sequence of (orthonormal) matrices S_T, such that $\|\hat{C}_{n_0}S_T^{-1} - C_{n_0}\| \to 0$. It follows from the construction of the proof, that also $\|S_T\hat{\Sigma}_{n_0}S_T^T - \Sigma_{n_0}\| \to 0$ holds. Thus we obtain $\|\hat{C}_{n_0}\hat{\Sigma}_{n_0}\hat{C}_{n_0}^T - C_{n_0}\Sigma_{n_0}C_{n_0}^T\| \to 0$. Therefore in the case, that the last column of C_{n_0} is nonzero, we obtain for $n < n_0$:

$$\begin{aligned} \hat{E}_n\hat{E}_n^T - \hat{E}_{n_0}\hat{E}_{n_0}^T &= \hat{C}_{n+1:n_0}\hat{\Sigma}_{n+1:n_0}\hat{C}_{n+1:n_0}^T \\ &\to C_{n+1:n_0}\Sigma_{n+1:n_0}C_{n+1:n_0}^T \\ &\geq c > 0 \end{aligned} \tag{5.4}$$

a.s. for some constant $c > 0$. Here $\hat{C}_{n+1:n_0}$ denotes the matrix of the last $n_0 - n$ columns of $\hat{C}_{n_0}$ and $C_{n+1:n_0}$ is defined analogously. $\hat{\Sigma}_{n+1:n_0}$ denotes the trailing $(n_0-n)\times(n_0-n)$ submatrix of $\hat{\Sigma}_n$. If the last column of C_{n_0} (which depends on the actual weighting scheme and the choice of the truncation index f) is zero, we cannot conclude, that asymptotically the correct order will be estimated. Thus assume for the moment, that the last column of C_{n_0} is nonzero. Below Theorem 5.2.2 we will discuss this condition in some special cases.

Under this assumption equation (5.4) and $C_T/T \to 0$ show, that asymptotically $\hat{n} \geq n_0$ holds. Thus consider the difference $IVC(n) - IVC(n_0)$ for $n \geq n_0$: For convenience of presentation we will slightly change the state basis, which is used in estimation, by defining $\hat{\mathcal{K}}_p = \hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2}$ resulting in $\langle\hat{x}_t(n), \hat{x}_t(n)\rangle = I$ (where we emphasized the dimension of $\hat{x}_t(n)$ notationally). This corresponds to a state space transformation with the matrix $T = (\hat{\Sigma}_n)^{-1/2}$ i.e. is just a scaling of the coordinates. The transfer function estimate is not changed by this transformation, irrespective of the specified order n. For $n > n_0$ partition the state vector as $\hat{x}_{t,1} = \hat{x}_{t,1:n_0}, \hat{x}_{t,2} = \hat{x}_{t,n_0+1:n}$, i.e. $\hat{x}_{t,1}$ denotes the first n_0 coordinates of $\hat{x}_t(n)$ and $\hat{x}_{t,2}$ the remaining $n - n_0$. Then if $\hat{C}_{n_0+1:n}$ denotes the columns with index $n_0+1, \cdots, n$ of the matrix $\hat{C}_n$, we obtain $\hat{C}_{n_0+1:n} = \langle y_t, \hat{x}_{t,2}\rangle = \langle y_t - C_{n_0}\hat{x}_{t,1}, \hat{x}_{t,2}\rangle$ due to the orthogonality of $\hat{x}_{t,1}$ and $\hat{x}_{t,2}$. Now

$$\langle y_t - C_{n_0}\hat{x}_{t,1}, \hat{x}_{t,2}\rangle = E_{n_0}\langle\varepsilon_t, \hat{x}_{t,2}\rangle - C_{n_0}\langle\hat{x}_{t,1} - x_{t,1}, \hat{x}_{t,2}\rangle$$

The ℓ_2 norm of the first term may be bounded by $c\|\langle\varepsilon_t, Y_{t,p}^-\rangle\|_2$ for some constant $c < \infty$, since $\hat{x}_{t,2} = \hat{V}_2^T(\hat{\Gamma}_p^-)^{-1/2}Y_{t,p}^-$, where $\hat{V}_2$ denotes the columns with indices $n_0+1, \cdots, n$ of the matrix $\hat{V}$ and thus has ℓ_2 norm smaller than one. The norm of $(\hat{\Gamma}_p^-)^{-1/2}$ is bounded since the norm of $\hat{\Gamma}_p^-$ is bounded and the spectrum of the square root of a bounded self adjoint operator is the square root of the spectrum and thus again bounded. $\|\langle\varepsilon_t, Y_{t,p}^-\rangle\|_2 = O(\sqrt{p}Q_T)$ holds due to the uniform convergence of the sample covariances (see Hannan and Deistler, 1988, Theorem 5.3.1). For the second term note, that using dual arguments to the ones used in the proof of Theorem 5.2.1 it is possible to prove $\|\hat{\mathcal{K}}_p - \mathcal{K}_p\|_2 = O(\sqrt{p}Q_T)$. Therefore $\|\hat{C}_{n_0+1:n}\|_2 = O(\sqrt{p}Q_T)$. Thus $\|\hat{C}_{n_0+1:n}I_{n-n_0}\hat{C}_{n_0+1:n}^T\|_2 = O(pQ_T^2)$ and finally

$$\frac{T}{C_T}(IVC(n) - IVC(n_0)) = 2s(n-n_0) + O(pQ_T^2T/C_T)$$

will be positive a.s. for $\liminf C_T/(p\log\log T) \to \infty$. Thus we have shown the following result:

Theorem 5.2.2 *Let the process $(y(t))_{t\in\mathbb{Z}}$ be generated by a true system $k_0 \in U_{n_0}^{(m)}$ for some integer n_0, where the ergodic white noise fulfills the standard assumptions given in the introduction. If the truncation indices fulfill either of the conditions of the theorems in Chapter 4, if the last column of C_{n_0} is nonzero, where $(A_{n_0}, K_{n_0}, C_{n_0}, E_{n_0})$ denotes a realization of k_0 corresponding to the SVD of $W_f^+\mathcal{H}_f(\hat{\Gamma}^-)^{-1/2} = U_{n_0}\Sigma_{n_0}V_{n_0}^T$, and if the order is estimated as the minimizing integer of the criterion function $IVC(n), 0 \leq n \leq \min(fs, ps)$, i.e.*

$$\hat{n} = \arg \min_{0 \leq n \leq \min(fs,ps)} IVC(n)$$

then $\hat{n} \to n$ *a.s. if* $\lim C_T/T \to 0, \liminf C_T/(p \log\log T) \to \infty$.

REMARK 5.2.4 The same remarks as below Theorem 5.2.1 are in order: The result holds for arbitrary weighting schemes, and all the various scenarios corresponding to the truncation indices used in Chapter 4. Again the proof relies on very brute force arguments and it seems likely, that the theorem also holds for weaker conditions on the penalty term.

The result stated above has one big drawback, which is the requirement, that the last column of C_{n_0} in the limiting realization is nonzero. For the case of $k_W = I$ (i.e. `N4SID`) it can be shown recursively in the order n, that the set of transfer functions fulfilling this condition is in fact generic in U_n^+. Note, that in this case the limiting system (A_0, K_0, C_0, E_0) is Lyapunov balanced. Clearly the norm of the last row of C_n is an analytic function of the balanced parameters on the parameter set T_n^{++} (see section 2.2.2). Since T_n^{++} is connected for $s > 1$, the norm is either generically nonzero or zero on the whole parameter set and thus on U_n^+. Thus it is sufficient to show, that there exists for each $n \in \mathbb{N}$ a system having a last column of C_n, which is nonzero. This will be shown by induction in the system order n: For $n = 1$ the result is obvious. Next assume, that the result has been shown for $n-1$. The induction step will use the parametrization of Ljapunov balanced systems derived in Chapter 2. We obtain a system of order n having the property, that the last column of C_n is not zero by augmenting the $(n-1)$-th order system with an additional large singular value $\sigma \gg \sigma_1$, where σ_1 denotes the largest singular value of the $(n-1)$-th order system, and arbitrary entries for the other new parameters. Then the matrix inversion lemma applied to the formulas defined in equation (2.6) together with continuity arguments show, that the last column of C_n is nonzero for σ large enough. For $s = 1$ the parameter space is the union of finitely many pathwise connected sets. However, it is straightforward to choose the parameter values in order to show, that in each of these subsets there exists a system with the desired properties: This can be achieved by selecting the parameters for K_c and C_c to have the same or the opposite sign respectively.

Using the minimum-phase balancing of Chapter 2 together with choosing p_1 close to one shows the same result for `CCA`. This shows, that in the cases $k_W = I$ or `CCA` and $f \to \infty$ the condition $C_{n_0-1:n_0} \neq 0$ is not too restrictive. Letting the index f tend to infinity seems to be a natural setting in the case, where the order of the system is not known, since the choice of f as fixed would restrict the possible orders. However the restriction to $k_W = I$ is a severe restriction. It is desirable to obtain (at least) a result, that the case of $C_{n_0-1:n_0} \neq 0$ holds generically in $U_n^{(m)}$ for all weightings. No results for other weighting matrices have been obtained, this question is still an open problem. Also the results of the estimation will be poor, even if the last column of C_{n_0} corresponding to the true system is small albeit nonzero (see section 5.2.2, Example I). Thus the quality of the estimates in finite samples will heavily rely on the magnitude of the last column in comparison to the sample size.

The Case of exogenous inputs

So far we have established two procedures, which allow for a consistent estimation of the order of the system in the case of unobserved white noise inputs (under suitable conditions). The extensions of these two methods to the case of additional observed inputs are obvious. The definition of the two criteria $SVC(n)$ and $IVC(n)$ is unchanged. For $SVC(n)$ note, that the analysis of its properties relies only on two arguments: The central result, which is used very often is the uniform convergence of the sample covariances. Thus we have to ensure, that this result holds also in the case of additional observed inputs, which can be done using the restrictions on the input process used in section 4.3. In the case of no exogenous inputs the uniform convergence was used to conclude that $\|\hat{W}_f^+ \hat{\beta}_{f,p} \hat{W}_p^- - W_f^+ \beta_{f,p} W_p^-\|_2 = O(\sqrt{fp} Q_T)$ a.s.

In the case of the unrestricted algorithms (i.e. the algorithms, where the structure in the matrix $\mathcal{U}_f$ is not used), the arguments are nearly unchanged and thus omitted. For the constrained

regression procedure the arguments in (Peternell *et al.*, 1996) show the same result. Finally we deal with the two stage method. It is easy to see, that a sufficient condition for the result to hold also in this case, is that the convergence of $(\hat{A}_T, \hat{B}_T, \hat{C}_T, \hat{D}_T)$ is of order $O(Q_T)$. This will be shown in the following. For convenience we will impose the restriction $(k_0, l_0) \in U_n^+$, implying that all the nonzero singular values of the limiting matrix $W^+\mathcal{O}\mathcal{K}W^-$ are distinct and we may use the differentiability properties of the singular value decomposition. In Chapter 4 it has been shown, that under the assumption, that the true system order has been used for the estimation,

$$\sqrt{T}\mathrm{vec}(\hat{A}_T - A, \hat{B}_T - B, \hat{C}_T - C, \hat{D}_T - D, \hat{E}_T - E, \hat{K}_T - K) \doteq L_{f,p}\sqrt{T}(\hat{g}_{T,f+p} - g_{f+p}) \quad (5.5)$$

where $\doteq$ stands for equality up to terms of order $o_P(1)$. Here (A, B, C, D, E, K) denotes the realization of the true transfer function, which is obtained by using the true covariances instead of the sample estimates in the subspace procedure. It has also been shown, that

$$\|\mathrm{vec}(A, B, C, D, E, K) - \mathrm{vec}(A_0, B_0, C_0, D_0, E_0, K_0)\| = o(T^{-1/2})$$

and thus we may replace (A, B, C, D, E, K) by $(A_0, B_0, C_0, D_0, E_0, K_0)$ in equation (5.5) without changing the error term. Using the uniform convergence of the sample covariances, it can be shown, that the term $o_P(1)$ indeed is $o(1)$ a.s. (In fact, this reasoning has been used in most places in Chapter 4). Each element of $\sqrt{T}(\hat{g}_{T,f+p} - g_{f+p})$ is of order $O(\sqrt{T}Q_T)$ a.s. and the rows of $L_{f,p}$ are of bounded ℓ_1 norm and have been shown to converge in ℓ_1 norm to vectors with elements decreasing exponentially. This shows the desired order of convergence of the system matrix estimates, if the order has been chosen correctly.

REMARK 5.2.5 Note that this result in fact has the form of a law of the iterated logarithm, except for the fact, that we did not evaluate the constant exactly.

For the two stage procedure however, estimation of the order has to be performed in the first step as well. If the order in the first step is specified smaller than the true one, then equation (5.5) still holds, where $(A_0, B_0, C_0, D_0, E_0, K_0)$ have to be replaced by the realizations of the limiting lower order system (see the forthcoming section 5.3.2). Thus the estimates of $\mathcal{U}_f$ will not be consistent and the estimation of the order in the second step will typically lead to an extremely high order, since in this case $\tilde{\beta}_z = (\Gamma_{y,z} - \tilde{\mathcal{U}}_f\Gamma_{u,z})(\Gamma_{z,z}^-)^{-1} = (\mathcal{O}_f\Gamma_{x,z} - (\tilde{\mathcal{U}}_f - \mathcal{U}_f)\Gamma_{u,z})\Gamma_{z,z}^{-1}$, where $\tilde{\mathcal{U}}_f$ denotes the lower triangular block Toeplitz matrix of the limiting Markoff coefficients of the lower order approximation to (k_0, l_0). This matrix might well be of high rank compared to n.

Finally we examine the case, where in the first step an order is specified, which is larger than the true order. Essential for the second step proved to be the accuracy of the Markoff coefficients estimates obtained in the first step. Thus consider the situation, where in the first step the order $n > n_0$ is used. Recall, that $\hat{\mathcal{K}}_p = (\hat{\Sigma}_n)^{1/2}\hat{V}_n^T(\hat{W}_p^-)^{-1}$. For the following discussion it will prove convenient to use a slightly different definition: In the sequel let $\hat{\mathcal{K}}_p = \hat{V}_n^T(\hat{W}_p^-)^{-1}$, which is equivalent to the state transformation $\hat{x}_{t,new} = (\hat{\Sigma}_n)^{-1/2}\hat{x}_{t,old}$. Since we are only interested in the accuracy of the Markoff coefficients, the choice of the basis is of no importance, and with the new normalization some of the calculations become easier. Also we will reorder the matrices (C, D) and (A, B, K) such that in the new ordering the coordinates $n_0 + 1, \cdots, n$ of the estimated state $\hat{x}_t$ always take the last $n - n_0$ coordinates of the regressors. This is done only for notational convenience. Thus e.g. (C, D) are estimated by regressing y_t onto $(\hat{x}_{t,1:n_0}^T, u_t, \hat{x}_{t,n_0+1:n}^T)^T$, where the additional indices refer to the coordinates of the vectors in **MATLAB** notation (i.e. $i : j$ stands for $i, i + 1, \cdots, j$). Here we assume, that the matrix D is estimated and not set to zero a priori. However, also the case of a time delay (i.e. $D = 0$) can be treated with the techniques used below. In order to simplify the notation, let $\hat{x}_{t,1} = \hat{x}_{t,1:n_0}$ and $\hat{x}_{t,2} = \hat{x}_{t,(n_0+1):n}$. Then we may use the block matrix inversion to obtain

$$\langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \\ \hat{x}_{t,2} \end{pmatrix}, \begin{pmatrix} \hat{x}_{t,1} \\ u_t \\ \hat{x}_{t,2} \end{pmatrix} \rangle^{-1} = \begin{bmatrix} \hat{M}^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -\hat{M}^{-1}\hat{X}_{1,2} \\ I \end{bmatrix} (\hat{X}_{2,2}^{\perp})^{-1}[-\hat{X}_{1,2}^T\hat{M}^{-1}, I]$$

and thus

$$[\hat{C}_{1:n_0}, \hat{D}_n, \hat{C}_{n_0+1:n}] = [\hat{C}_{n_0}, \hat{D}_{n_0}, 0] + [\langle \hat{E}_{n_0}\hat{\varepsilon}_t, \hat{x}_{t,2}\rangle](\hat{X}^{\perp}_{2,2})^{-1}[-\hat{X}^T_{1,2}\hat{M}^{-1}, I] \tag{5.6}$$

where

$$\begin{aligned}
\hat{M} &= \langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix}, \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix} \rangle \\
\left[\hat{C}_{n_0}, \hat{D}_{n_0}\right] &= \langle y_t, \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix} \rangle \hat{M}^{-1} \\
\hat{E}_{n_0}\hat{\varepsilon}_t &= y_t - [\hat{C}_{n_0}, \hat{D}_{n_0}] \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix} \\
\hat{X}^{\perp}_{2,2} &= \langle \hat{x}_{t,2}, \hat{x}_{t,2}\rangle - \langle \hat{x}_{t,2}, \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix} \rangle \hat{M}^{-1} \langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix}, \hat{x}_{t,2}\rangle \\
\hat{X}_{1,2} &= \langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix}, \hat{x}_{t,2}\rangle
\end{aligned}$$

Note, that $\hat{x}_{t,1} = \hat{\mathcal{K}}_p Z^-_{t,p}$. For each element, $(\hat{\mathcal{K}}_p)_{i,j}$ say, of $\hat{\mathcal{K}}_p$ we obtain $\|(\hat{\mathcal{K}}_p)_{i,j} - (\mathcal{K}_p)_{i,j}\| = O(Q_T)$ using similar arguments as in the proof of Theorem 5.2.1. Therefore it follows from straightforward calculations that $\|\hat{M} - M\|_2 = O(Q_T)$ and that $\|\hat{M}^{-1}\|_2 \leq c > 0$. Thus the central term in equation 5.6 is equal to $\langle y_t - \hat{C}_{n_0}\hat{x}_{t,1} - \hat{D}_{n_0}u_t, \hat{x}_{t,2}\rangle = \langle E_0\varepsilon_t, \hat{x}_{t,2}\rangle + \langle \hat{E}_{n_0}\hat{\varepsilon}_t - E_0\varepsilon_t, \hat{x}_{t,2}\rangle$. In this section we obtained the result $\|\text{vec}[\hat{C}_{n_0} - C_0, \hat{D}_{n_0} - D_0, \hat{E}_{n_0} - E_0]\|_2 = O(Q_T)$. As in the case of no observed inputs the uniform convergence of the sample covariances implies $\|\langle E_0\varepsilon_t, \hat{x}_{t,2}\rangle\|_2 = O(\sqrt{p}Q_T)$. Also $\hat{E}_{n_0}\hat{\varepsilon}_t - E_0\varepsilon_t = (\hat{C}_{n_0}\hat{x}_{t,1} - C_0x_{t,1}) + (\hat{D}_{n_0} - D_0)u_t$ and thus $\|\langle \hat{E}_{n_0}\hat{\varepsilon}_t, \hat{x}_{t,2}\rangle\|_2 = O(\sqrt{p}Q_T)$. Thus it remains to prove, that $(\hat{X}^{\perp}_{2,2})^{-1}$ and $\hat{X}_{1,2}$ have bounded norm (a.s. for T large enough). For $\hat{X}_{1,2}$ this follows from the finite norm of $\hat{\mathcal{K}}_p$ and the boundedness of the matrix $\hat{\Gamma}_{z,z}$. $(\hat{X}^{\perp}_{2,2})^{-1}$ has bounded norm, if the smallest eigenvalue of

$$\langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \\ \hat{x}_{t,2} \end{pmatrix}, \begin{pmatrix} \hat{x}_{t,1} \\ u_t \\ \hat{x}_{t,2} \end{pmatrix} \rangle$$

is bounded away from zero a.s. and uniformely in T. However

$$\begin{pmatrix} \hat{x}_{t,1} \\ u_t \\ \hat{x}_{t,2} \end{pmatrix} = \begin{bmatrix} 0 & 0 & \hat{V}^T_{1:n_0}(\hat{W}^-_p)^{-1} \\ 0 & I & 0 \\ 0 & 0 & \hat{V}^T_{n_0+1:n}(\hat{W}^-_p)^{-1} \end{bmatrix} Z^-_{t+1,p+1}$$

and thus the result follows from the fact, that the matrix on the left side is of full rank, and from the uniform bound on the smallest eigenvalue of $\langle Z^-_{t+1,p+1}, Z^-_{t+1,p+1}\rangle$ (see e.g. Peternell *et al.*, 1996). Thus we obtain for the estimates using the order n, that

$$\|\text{vec}[\hat{C}_n - (C_0, 0), \hat{D}_n - D_0]\|_2 = O(\sqrt{p}Q_T)$$

Corresponding to the estimation of (A, B) note, that the regression in the first n_0 coordinates is of completely the same form as the regression of y_t onto $\hat{x}_t, u_t$. Thus the same arguments as given above show, that $\|\text{vec}[\hat{A}_{1:n_0,1:n_0} - A_0, \hat{B}_{1:n_0,1:m} - B_0]\|_2 = O(\sqrt{p}Q_T)$ and that $\|\text{vec}[\hat{A}_{1:n_0,n_0+1:n}]\|_2 = O(\sqrt{p}Q_T)$, where again the subscripts indicate the various submatrices of the estimates in **MATLAB** notation. Corresponding to the remaining matrices $\hat{A}_{(n_0+1):n,1:n}$ and $\hat{B}_{(n_0+1):n,1:m}$, we obtain a bound on the 2-norm.

Concluding we obtain for the estimates of the Markoff coefficients, that $\|\hat{C}_n\hat{B}_n - C_0B_0\|_2 = O(\sqrt{p}Q_T)$ and the same is true for $\|\hat{C}_n(\hat{A}_n)^j\hat{B}_n - C_0A_0^jB_0\|_2$ for $0 \leq j \leq K$, where K is some

fixed integer. In order to extend these results also for $0 \leq j \leq f$ in the case of $f \to \infty$, we need also some results on the stability of the estimate $\hat{A}_n$. Such a result does not exist, to the best of the authors knowledge. Thus, also according to the results in section 4.3.3 we will restrict the choice of f to be fixed in the second stage. Then we may use the result above for $K = f$. The arguments presented above show, that each element of $\mathcal{U}_f(\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n) - \mathcal{U}_f(A_0, B_0, C_0, D_0)$ is of order $O(\sqrt{p}Q_T)$, where we stressed the dependence of $\mathcal{U}_f$ on the system matrices in the notation. Thus the elements of $\hat{\Gamma}_{y,z} - \Gamma_{y,z} - (\mathcal{U}_f(\hat{A}_n, \hat{B}_n, \hat{C}_n, \hat{D}_n)\hat{\Gamma}_{u,z} - \mathcal{U}_f(A_0, B_0, C_0, D_0)\Gamma_{u,z})$ are of order $O(\sqrt{p}Q_T)$. This shows, that the elements of $\hat{\beta}_z - \beta_z$ are of the same order, in contrast to the other two procedures, where we obtained the result, that the elements of $\hat{\beta}_z - \beta_z$ are of the order $O(Q_T)$.

Using this result, it is straightforward to show, that

$$||\hat{W}_f^+ \hat{\beta}_z \hat{W}_p^- - W_f^+ \beta_z W_p^-||_2 = O(\sqrt{fp^j}Q_T)$$

a.s., where $j = 1$ for the unconstrained and the constrained method and $j = 2$ for the two stage method, and thus the proof of the consistency of the $SVC(n)$ criterion holds unchanged in all three cases, where only the penalty term C_T has to be larger for the two stage procedure.

In the arguments given above, we also derived many details to examine the properties of the estimates of the order obtained by minimizing the $IVC(n)$ criterion: Consider first the case of underestimating the order: In this case again we will need a condition on the limiting system matrices to ensure, that asymptotically the criterion does not lead to underestimation of the order. In this case however, we do not obtain such a simple expression as in the case of no observed inputs. Let $\Omega(n) = E_n E_n^T$ denote the (a.s.) limit of the sample covariance matrix of the residuals of the regression of y_t onto $\hat{x}_t(n)$, where $\hat{x}_t(n) \in \mathbb{R}^n$ denotes the estimate of the state using the order n (and u_t, if D is also estimated). Then we have to impose the condition $\Omega(n_0) - \Omega(n_0 - 1) > 0$ in order to ensure, that there is no risk of asymptotically underestimating the order, when using $IVC(n)$. The analysis of this condition is even more complex than in the case of no observed inputs, since in the present case, the covariance matrix of $x(t)$ may not be diagonal, depending on the weighting scheme used. However, note that the condition does not depend on the decision to use the unstructured approach, the constrained regression approach or the two stage procedure. However it does depend on the choice for $\hat{W}_f^+$ and $\hat{W}_p^-$. As an interpretation of this condition, we note, that $\Omega(n_0) - \Omega(n_0 - 1) = 0$ means, that the last coordinate of the state does not contribute to the presence of the output, but only to the future. We did not obtain any results on the relevance of this restriction.

In the next step, we will investigate the risk of overestimating the order. Therefore consider $T/C_T(IVC(n) - IVC(n_0)) = (d(n) - d(n_0)) + T/C_T(\hat{E}_n \hat{E}_n^T - \hat{E}_{n_0} \hat{E}_{n_0}^T)$. Recall, that $\hat{E}_n \hat{E}_n^T = \hat{\gamma}_{y,y}(0) - [\hat{C}_n, \hat{D}_n]\langle \begin{pmatrix} \hat{x}_t(n) \\ u_t \end{pmatrix}, \begin{pmatrix} \hat{x}_t(n) \\ u_t \end{pmatrix} \rangle [\hat{C}_n, \hat{D}_n]^T$ and thus we have to examine the norm of the difference of the various expressions. All the expressions involving $\hat{x}_t(n)$ will again split into the two parts $\hat{x}_{t,1} \in \mathbb{R}^{n_0}$ and $\hat{x}_{t,2} \in \mathbb{R}^{n-n_0}$. Now for example $\langle \hat{x}_{t,1}, \hat{x}_{t,1} \rangle = \langle x_{t,1}, x_{t,1} \rangle + O(\sqrt{p^{j-1}}Q_T) = \gamma_{x,x}(0) + O(\sqrt{p^{j-1}}Q_T)$, where $j = 1$ for the unconstrained regression method and the constrained method and $j = 2$ for the two stage method, which follows from the fact, that the elements of $\hat{\beta}_z - \beta_z$ are of order $O(\sqrt{p^{j-1}}Q_T)$ and that the elements of $\hat{\mathcal{K}}_p - \mathcal{K}_p$ are of order $O(\sqrt{p^{j-1}}Q_T)$.

Then the result follows from the properties of the matrix $\mathcal{K}_p(\Gamma_{z,z}^-)$, in particular from the bounded ℓ_1 norm and the exponential decrease of the limiting matrix $\mathcal{K}(\Gamma_{z,z}^-)$. The same is true for $\langle \hat{x}_{t,1}, u_t \rangle$. $\langle u_t, u_t \rangle$ of course is of order $O(Q_T)$ a.s. This shows, that $||\hat{M} - M||_2 = O(\sqrt{p^{j-1}}Q_T)$ (see also equation (5.6)). For all terms involving $\hat{x}_{t,2}$ we only obtain a bound on the norm, uniformely in f, p and T. Thus it remains to investigate $(\hat{C}_n - [C_{n_0}, 0], \hat{D}_n - D_0)$. Recall equation (5.6). Since the bound on $\hat{X}_{1,2}$ and $(\hat{X}_{2,2}^{\perp})^{-1}$ holds independent of the estimation method, the crucial term is equal to $\langle \hat{E}_{n_0} \hat{\varepsilon}_t, \hat{x}_{t,2} \rangle = \langle E_0 \varepsilon_t, \hat{x}_{t,2} \rangle + \langle C_0 x_{t,1} - \hat{C}_{n_0} \hat{x}_{t,1}, \hat{x}_{t,2} \rangle - \langle D_0 u_t + \hat{D}_{n_0} u_t, \hat{x}_{t,2} \rangle$. Since $||\hat{C}_{n_0} - C_0||_2 = O(Q_T \sqrt{p^{j-1}})$ and $||\hat{D}_{n_0} - D_0||_2 = O(Q_T \sqrt{p^{j-1}})$ (see equation (5.5)) the remaining term is $\langle x_{t,1} - \hat{x}_{t,1}, \hat{x}_{t,2} \rangle = O(\sqrt{p^j}Q_T)$, where again $j = 1$ or $j = 2$ respectively, depending

on the method. Finally we obtain also $\|\hat{E}_n\hat{E}_n^T - \hat{E}_{n_0}\hat{E}_{n_0}^T\|_2 = O(\sqrt{p^j}Q_T)$ uniformely in f and $n_0 \leq n \leq \min(fs, p(s+m))$, a.s. Thus $T/C_T(IVC(n)-IVC(n_0)) = d(n)-d(n_0)+O(T\sqrt{p^j}Q_T/C_T)$ and this will be positive a.s. for large T, if $C_T/(\sqrt{p^j}T^{1/2}\sqrt{\log\log T}) \to \infty$. Note, that this lower bound on the increase of the penalty is much larger, than the previously obtained, since it contains the term $T^{1/2}$. In some special situations we obtain much weaker bounds on the increase of C_T, e.g. in the situation, where $\hat{W}_p^- = (\hat{\Gamma}_{z,z})^{1/2}$ and where $D = 0$ is assumed: In this special case only $C_T/(p^jQ_T^2) \to \infty$ is needed, which follows from the diagonality of $\langle \hat{x}_t, \hat{x}_t \rangle$ and arguments similar to the case of no exogenous inputs. Concluding we obtain the following theorem:

Theorem 5.2.3 *Let $(y(t))_{t\in\mathbb{Z}}$ and $(u(t))_{t\in\mathbb{Z}}$ fulfill the assumptions of Theorem 4.3.1 or Theorem 4.3.2 respectively, depending on the choice of f. Then if the estimate, $\hat{n}$ say, is obtained by minimizing $IVC(n)$ over $0 \leq n \leq \max(fs, p(s+m))$, then $\hat{n} = n_0$ a.s. under the condition that*

- $\Omega(n_0) - \Omega(n_0 - 1) > 0$
- $C_T > 0, C_T/T \to 0, C_T/(\sqrt{p^jT\log\log T}) \to \infty$

where $j = 2$ if the two stage method is used and $j = 1$ else.

If the estimate $\hat{n}$ is obtained by minimizing $SVC(n)$ over $0 \leq n \leq \max(fs, p(s+m))$, then $\hat{n} = n_0$ a.s. under the conditions $C_T > 0, C_T/T \to 0, C_T/(fp^j \log\log T) \to \infty$.

Note, that the first part of this theorem is not very powerful, since both conditions are not satisfactory: There is no knowledge about the structure of the first condition. The penalty term is restricted to be very big, which in finite sample will lead to a high risk of underestimating the order. Thus it is believed that in the case of additional observed inputs the SVC criterion leads to a better theoretical foundation of the procedure of order estimation. Also the implimentation of the IVC criterion is not as simple as in the case of no observed inputs, since in the present case, for each order a regression has to be calculated, which has no simple structure, since $\langle \hat{x}_t(n), \hat{x}_t(n) \rangle$ may not be diagonal for general weighting matrix $\hat{W}_p^-$. Thus it is believed, that in the case of additional observed inputs the SVC criterion is preferable from a theoretical point of view. Of course, this has to be complemented with experience from practice and simulation studies.

5.2.2 Comparison

This section presents some simulation studies, which compare the various order estimation procedures presented in the last section in finite sample situations. We will examine the effect of the choices of f and p as well as the effects of the weighting matrices on the estimation of the order. This will be done in four examples: A second order SISO system (Example I), a (nearly nonminimal) eighth order system (Example II), a MIMO system (Example III) and finally one example with observed inputs (Example IV). We will mainly use systems, which have also been used for simulations in (Peternell, 1995). This is done, since these models are also used for the calculations in later sections, where also the asymptotic distribution of the estimates will be discussed. There it will be convenient to consider the simulation studies in (Peternell, 1995) as some kind of benchmark to which our results can be compared.

Example I

As a first example consider the system defined by the following matrices:

$$A = \begin{bmatrix} 0 & 1 \\ -0.7 & 0.5 \end{bmatrix}, K = \begin{bmatrix} 1.3 \\ 0.3 \end{bmatrix}, C = [1, 0], E = 1$$

This system has Lyapunov balanced Gramian of approximately $\Sigma = diag(2.55, 1.7792)$. The system poles are at $0.25 \pm 0.7984i$ and the zeros at $-0.4 \pm 0.4359i$.

	T	100			1000		
	Order	< 2	2	> 2	< 2	2	> 2
`CCA`	$IVC1$	0.00	0.83	0.17	0.00	0.77	0.23
	$IVC2$	1.00	0.00	0.00	0.82	0.18	0.00
	$SVC1$	0.00	0.96	0.04	0.00	0.93	0.07
	$SVC2$	1.00	0.00	0.00	0.86	0.14	0.00
`N4SID`	$IVC1$	0.82	0.03	0.15	0.68	0.06	0.26
	$IVC2$	1.00	0.00	0.00	1.00	0.00	0.00
	$SVC1$	0.00	0.66	0.34	0.00	0.40	0.60
	$SVC2$	0.45	0.55	0.00	0.04	0.96	0.00
`butter(12,0.5)`	$IVC1$	0.25	0.52	0.23	0.00	0.59	0.41
	$IVC2$	1.00	0.00	0.00	0.51	0.49	0.00
	$SVC1$	0.00	0.64	0.36	0.00	0.12	0.88
	$SVC2$	0.01	0.71	0.28	0.00	0.16	0.84

Table 5.1: Here the probability of estimating the indicated order for 1000 time series of sample size T is shown for three different weighting schemes (`CCA`, `N4SID` and a low pass filter) and 4 different estimation methods: $IVC(n)$ with $C_T = \log T$ ($IVC1$) and $C_T = fp \log T$ ($IVC2$) and SVC with $C_T = \log T$ ($SVC1$) and $C_T = fp \log T$ ($SVC2$). $f = p = \hat{p}_{AIC}$ has been used.

In a first simulation we demonstrate the effect of different penalty terms on the estimation of the order. In the estimation we use three different weighting schemes: `CCA`, `N4SID` and a low pass weighting generated with the `MATLAB` command `butter(12,.5)`. Both truncation indices are equal and estimated from the data according to the suggestion made in section 5.1. Here the constant $d = 1$ and `AIC` is used (see section 5.1). For each of the weighting schemes, the order is estimated using four different methods: IVC and SVC with $C_T = \log T$ (denoted with $IVC1$ and $SVC1$ respectively in the sequel), IVC and SVC with $C_T = \log T fp$ (denoted with $IVC2$ and $SVC2$ respectively). Note, that only for the last two procedures the consistency results have been derived. 1000 time series of length 100 and 1000 respectively have been generated and used for estimation. Table 5.1 shows the results for $T = 100$ and $T = 1000$ respectively.

The results in Table 5.1 show, that the performance of the order estimation procedure depends heavily on the weighting scheme: For `CCA` the $IVC1$ method works well, whereas it shows problems to estimate the true order, when used with `N4SID`. This is due to the fact, that in the Lyapunov balanced realization of the true system, the entry $C_{1,2}$ is equal to -0.0146 and thus close to zero (compare the discussion below Theorem 5.2.2). This leads to a high risk of underestimating the order using IVC together with `N4SID` in this example. For `CCA` it is observed, that as has been expected, the higher penalty term results in a high risk of underestimation, while reducing the risk of overestimation. For `N4SID` the SVC method outperforms IVC and we also observe, that for $C_T = fp \log T$ the accuracy increases with the sample size, whereas the lower penalty term does not seem to lead to consistent order estimates. For the frequency weighted method using the low pass filter, none of the methods performs good, however $IVC1$ gives better results for $T = 1000$ than the other three criteria. Thus we observe, that the performance of the order estimation procedures depends crucially on the weighting scheme (at least in this example). Based on these results, we will restrict ourselves to $C_T = \log T$ for IVC since $IVC1$ outperforms $IVC2$ in all cases. Corresponding to SVC we also make this restriction for the remaining simulations, although it is not so obvious in all situations, that this choice is superior.

In the following we will compare the four different estimation methods for different choice of the truncation indices. In the following, IVC using $C_T = \log T$ will be denoted with IVC, SVC with $C_T = \log T$ with SVC, the criterion derived by Peternell (see also equation (5.1)) by NIC and finally the method implemented in the system identification toolbox of `MATLAB` will be denoted

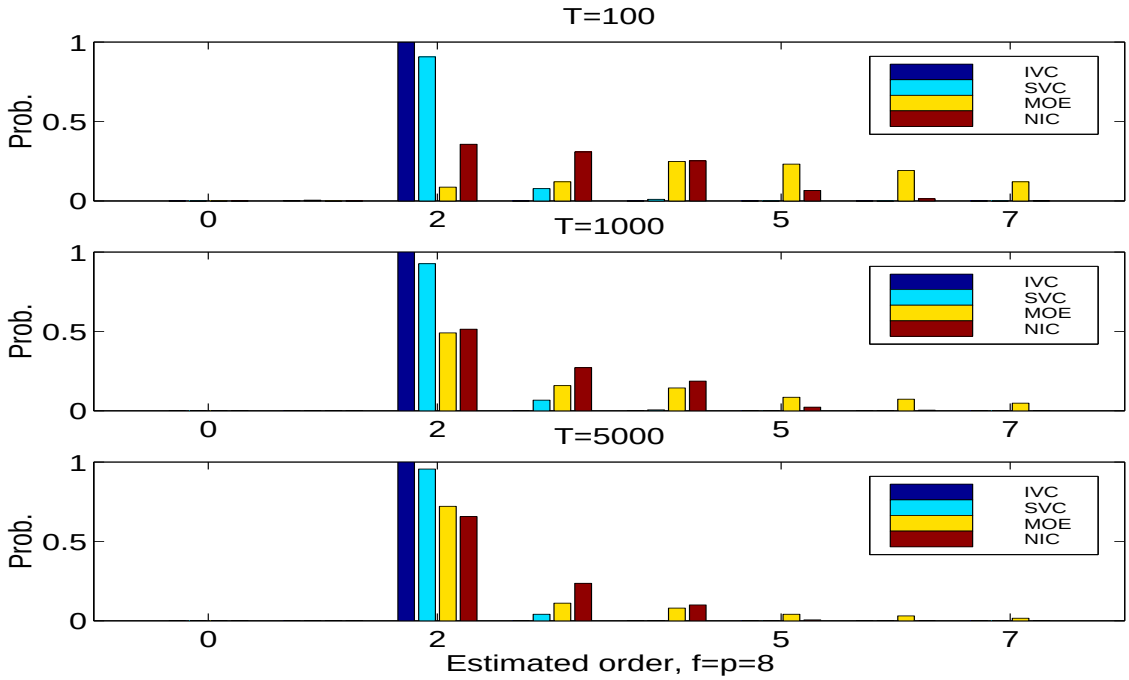

Figure 5.3: Example I: Here the probability of estimating the order n is plotted as a function of n. The probability has been estimated using 1000 time series of sample size T, `CCA` and $f = p = 8$. The three rows correspond to $T = 100$, $T = 1000$ and $T = 5000$. The bars correspond to IVC, SVC, MOE, NIC (see legend).

with MOE (see Ljung, 1991). We will use three different setups: To start with, f and p are fixed and equal to 8. In the second setup p is estimated from the data using the procedure suggested in section 5.1, where $d = 2$ was used and $f = 6$ was fixed. Finally the third setup uses estimated f and p according to the procedure of section 5.1.

In the case, that $f = p = 8$ is fixed, the results of 1000 simulation runs for the sample sizes $T = 100, T = 1000$ and $T = 5000$ can be seen in Figure 5.3. In all procedures the `CCA` weighting scheme is used.

For all four methods, the percentage of correct decisions increases with increasing sample size. In this case, IVC and SVC outperform the remaining procedures (especially for small sample sizes). Note, however, that fixing p results in estimates of the system matrices, which are not consistent in general (see the discussion in Chapter 4). Thus when estimating the order using fixed f and p, in order to obtain consistent estimates of the transfer function, it will be necessary to perform the subspace algorithm twice, where in the first step the order is estimated, and in the second step the transfer function estimate for the specified order is calculated. This higher computational load can be prevented by letting p tend to infinity as a function of the sample size (and the data), as will be done next.

In the second setup, $f = 6$ is fixed and p is estimated from the data (for details, see section 5.1). The results are documented in Figure 5.4. All procedures except for the NIC method show convergent behaviour. IVC outperforms the other procedures for $T = 100$, where only in one case the order $\hat{n} = 1$ was estimated and in the remaining cases the correct order has been specified. NIC on the other hand shows a tendency to overestimate the order even at sample size $T = 5000$.

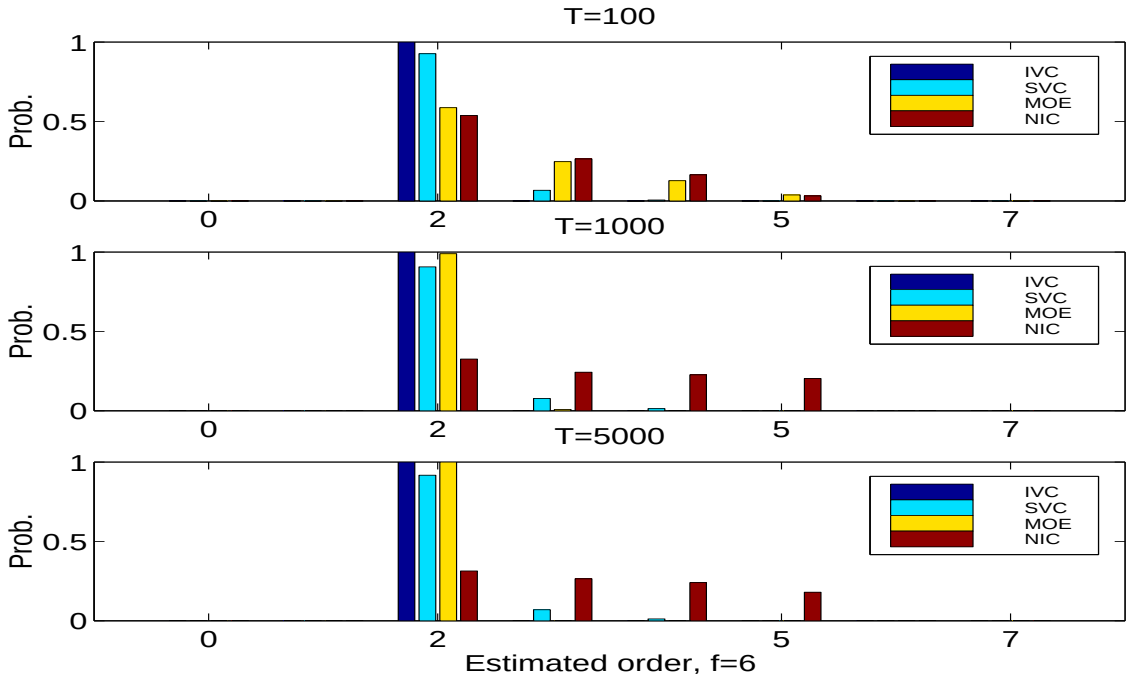

Figure 5.4: Example I: Here the probability of estimating the order n is plotted as a function of n. The probability has been estimated using 1000 time series of sample size T, using `CCA`, $f = 6$ and $p = 2\hat{p}_{AIC}$ is estimated from the data. The three rows correspond to $T = 100$, $T = 1000$ and $T = 5000$. The bars correspond to IVC, SVC, MOE, NIC (see legend).

Finally the third setup deals with the situation, where both truncation indices are allowed to tend to infinity and are estimated from the data. Figure 5.5 documents the performance of the four different methods for increasing sample size and $d = 2$ (see section 5.1 for details). In this case, it is clearly visible, that NIC and MOE do not perform well for all sample sizes compared to IVC and SVC, which detect the correct order in more than 85% of the cases. Moreover for $T = 5000$ especially for NIC the performance is worse than for sample size $T = 100$. In this case, there is a high risk of choosing a very high order (> 12). In fact this behaviour was the reason for the development of the new procedures IVC and SVC.

Figure 5.6 gives a comparison for various values of d. The results are quite different for the various procedures: While IVC performs better with increasing d, for NIC the opposite is true, for $d = 4$ only in one case the correct order has been estimated. Also the accuracy of SVC decreases with increasing d however less dramatically. MOE shows about the same behaviour as NIC.

The most striking fact about the simulations given above probably is the dependence of the performance of NIC and MOE on the choice of the truncation indices: In the case, where both truncation indices are fixed all procedures work satisfactory, however, if we estimate both truncation indices from data, NIC and MOE perform significantly worse than SVC and IVC. Thus these simulations suggest to use SVC or IVC rather than NIC or MOE if truncation indices are estimated. The simulations also showed, that the estimates provided by IVC and SVC improve with increasing sample size in all setups. Thus these methods seem to be relatively robust with respect to the choice of the truncation indices in this example.

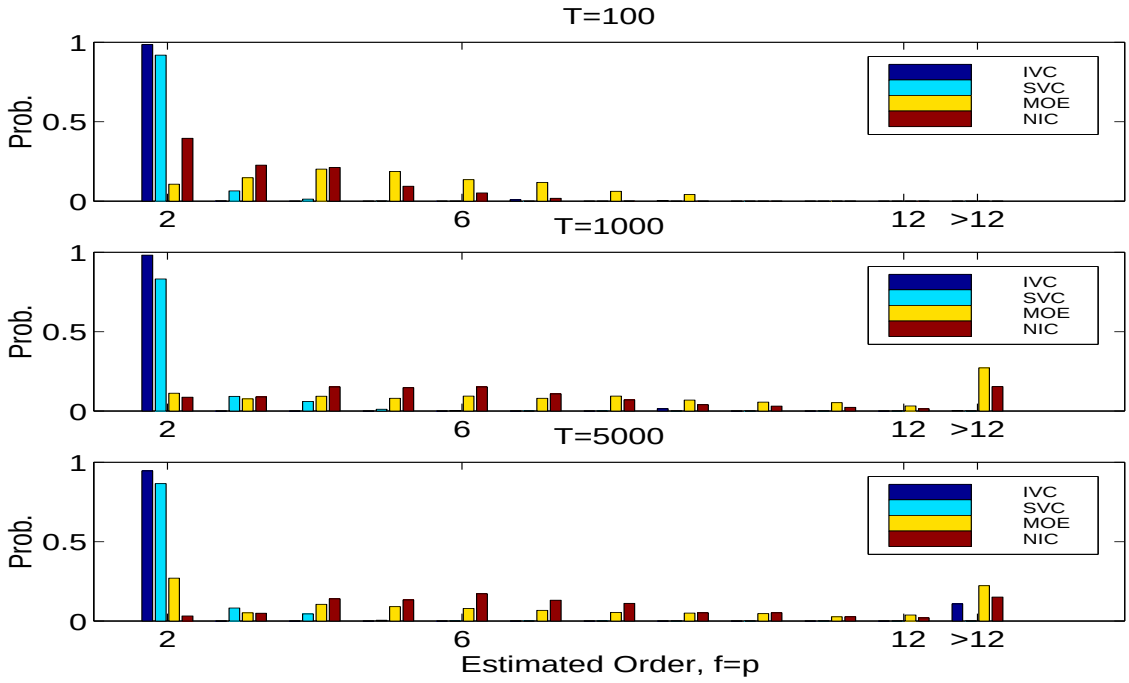

Figure 5.5: Example I: Here the probability of estimating the order n is plotted as a function of n. The probability has been estimated using 1000 time series of sample size T, using `CCA`, $f = p = 2\hat{p}_{AIC}$ are estimated from the data. The three rows correspond to $T = 100$, $T = 1000$ and $T = 5000$. The bars correspond to IVC, SVC, MOE, NIC (see legend).

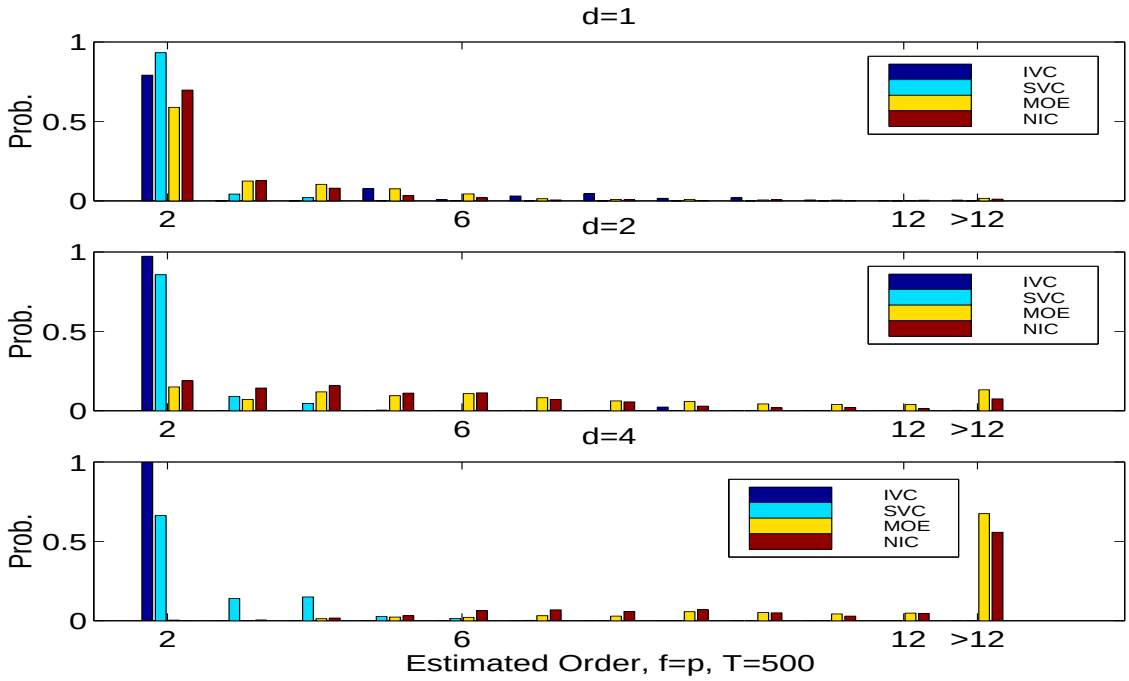

Figure 5.6: Example I: Here the probability of estimating the order n is plotted as a function of n. The probability has been estimated using 1000 time series of sample size $T = 500$, using `CCA`, $f = p = d\hat{p}_{AIC}$ are estimated from the data with several values of d. The three rows correspond to $d = 1$, $d = 2$ and $d = 4$. The bars correspond to IVC, SVC, MOE, NIC (see legend).

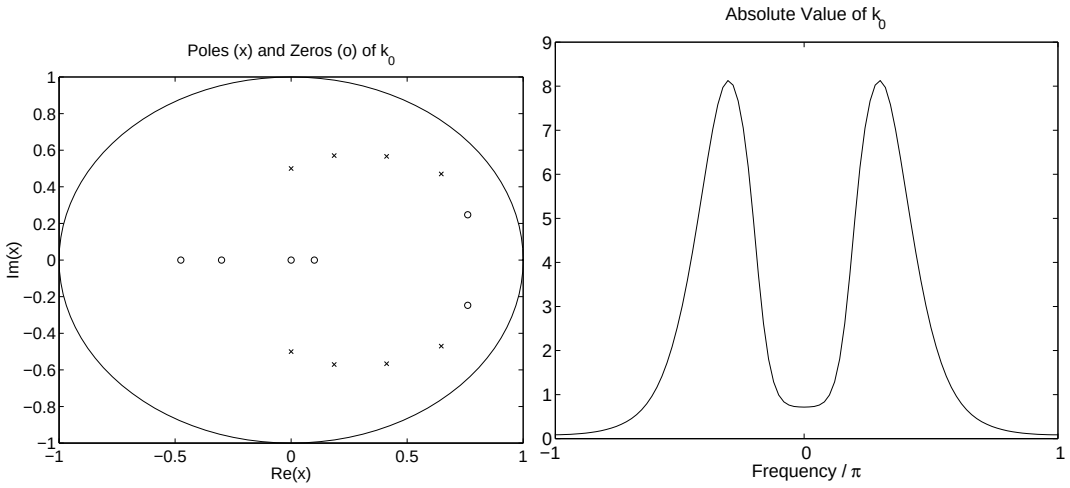

Figure 5.7: Example II: The left figure shows the poles (x) and the zeros (o) of the transfer function k_0, the right figure shows a plot of the absolute value of k_0 for the angular frequency range $(-\pi, \pi)$.

Example II

In this example we compare the estimation of the order for the four algorithms for an eighth order system, whose pole/zero configuration and absolute value of the transfer function k_0 is shown in Figure 5.7. The system has poles at $z = 0.8e^{\pm 0.2i\pi}$, $z = 0.7e^{\pm 0.3i\pi}$, $z = 0.5e^{\pm 0.5i\pi}$, $z = e^{\pm 0.4i\pi}$ and zeros at $z = 0.8e^{\pm 0.1i\pi}$, $z = -0.4755$, $z = 0.1$, $z = 0.3$, $z = 0$.

From the pole-zero plot it may be suspected, that the system will not be easy to estimate and this indeed is the case, as can be seen from the distribution of the system order estimates, which is shown in Figure 5.9 for sample size $T = 10000$ and for three different weighting schemes: `CCA`, a low pass filter generated by the `MATLAB` command `butter(12,.5)` and a high pass filter generated by `butter(12,.5,'high')`. The absolute values of the transfer functions describing the filters clearly show their properties (see Figure 5.8).

Even for this relatively large sample size the percentage of correct specifications of the order is rather low. It can also be seen, that the IVC criterion gives the highest rates of correct specification for the two weighting schemes corresponding to the filters, however, IVC is also the only criterion, which underestimates the order. MOE and NIC in contrast overestimate the order in most of the cases: The risk of obtaining an order higher than 15 is greater than 50% in all cases for these two procedures. Also SVC leads to overestimation of the order, except in the `CCA` case, where it shows the best accuracy. Again we see, that the performance of the order estimation procedure depends heavily on the choice of the weighting scheme.

However, correct specification of the order is not the only criterion, which is relevant for the identified system. Since the aim of the identification is the estimation of the transfer function k, one aspect of order estimation also lies in the results corresponding to estimation errors in terms of the transfer function (see also the discussion below Theorem 2.4.3). In Figure 5.10 the deviation of the sample mean over 1000 time series of sample length $T = 10000$ using `CCA` and $f = p = 2\hat{p}_{AIC}$, where $\hat{p}_{AIC}$ denotes the estimate of the order in the long autoregression (see section 5.1), is plotted for the four different order estimation procedures. The second plot shows the mean of the squared estimation error of the transfer function. It is observed, that MOE and NIC show less deviations to the true transfer function, which had been expected, since the estimated order is higher in average (see Figure 5.9). Despite this fact, SVC does not perform much worse with respect to the square of the modulus of the deviations.

Figure 5.11 shows the same plots for $T = 1000$. In these plots it can be seen again, that despite the fact, that the bias is much larger for SVC and IVC, the square of the modulus of the deviation

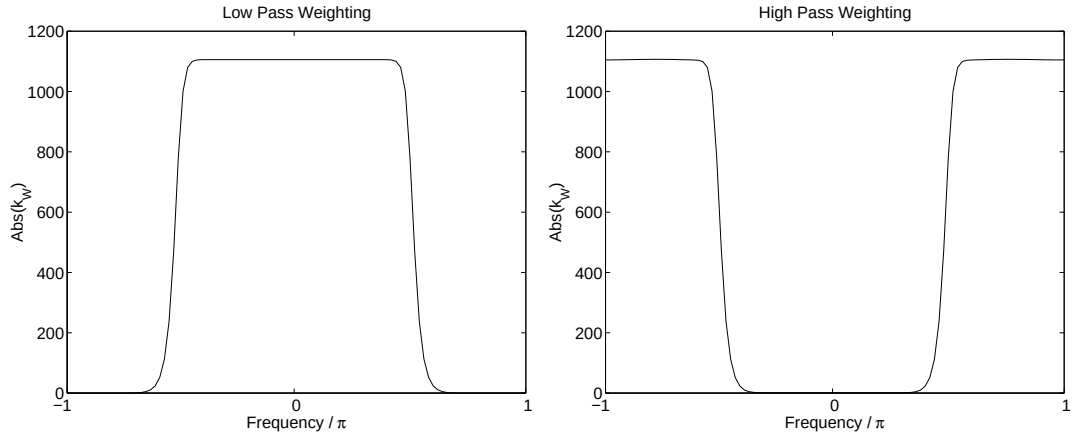

Figure 5.8: In the left figure, the absolute value of the low pass filter `butter(12,.5)`, in the right figure the absolute value of the high pass filter `butter(12,.5,'high')` is plotted for the angular frequency range $(-\pi, \pi)$.

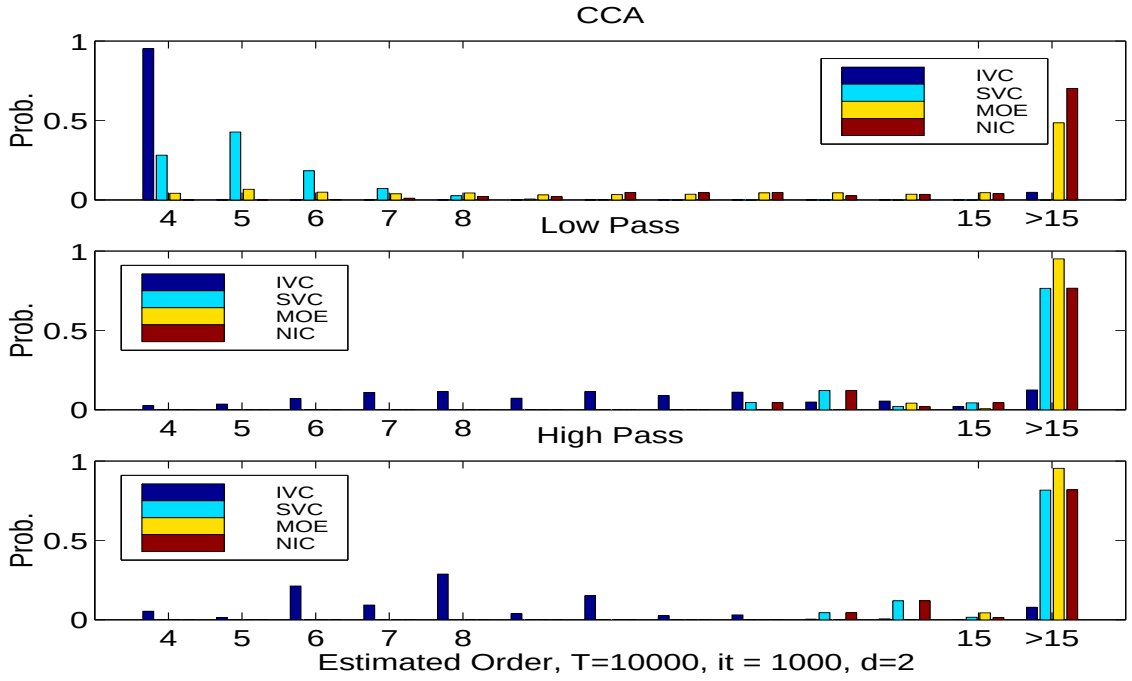

Figure 5.9: Example II: In this figure the estimates of the order of 1000 times series of sample size $T = 10000$ are shown for three different weighting schemes: `CCA` (picture on the top), the low pass filter (picture in the middle) and the high pass filter (picture on the bottom). The figures show the histogram of the estimated orders for IVC, SVC, MOE and NIC (see legend). $f = p = 2\hat{p}_{AIC}$.

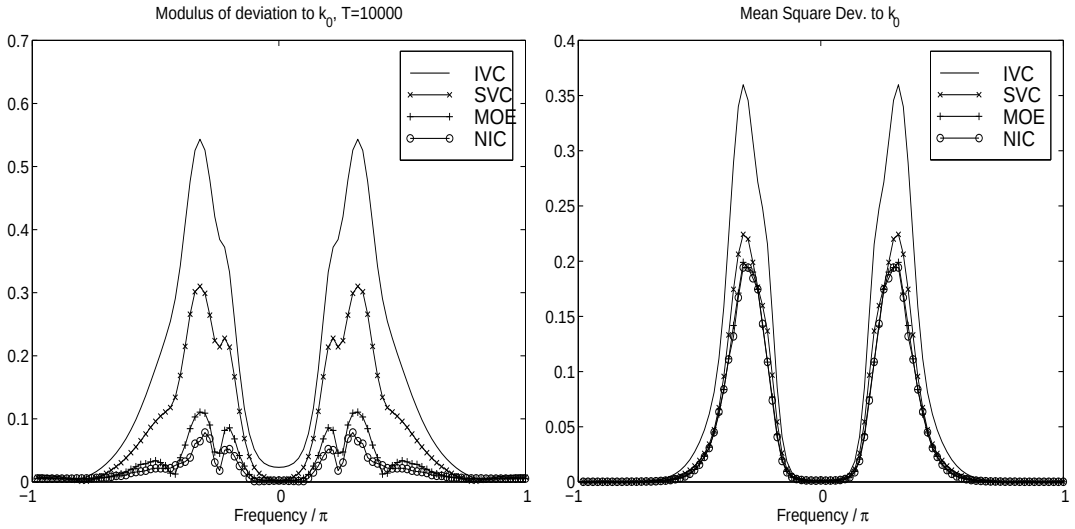

Figure 5.10: Example II: The left picture shows the absolute value of the deviation of the sample mean from k_0 for the four order estimation procedures (see legend). Here the sample mean is calculated from 1000 time series of sample length $T = 10000$, for which systems are estimated using `CCA` and $f = p = 2\hat{p}_{AIC}$. The right picture shows the mean of the square of the modulus of the deviation from k_0.

shows no large difference between the procedures.

Corresponding to the performance of the order estimation procedure it is also interesting to compare the results to the standard methodology of using `AIC` or `BIC` in combination with ML estimation. The next picture shows a comparison of the estimated orders using four different setups: As before, IVC and SVC are used together two order estimation procedures, which are labelled with AIC and BIC respectively. In the latter two cases, the transfer functions estimates for the same range of possible orders, as has been used in the subspace setup, are obtained by using the `MATLAB` function `PEM`(see Ljung, 1991) for each of the orders $1 \leq n \leq 2\hat{p}_{AIC}$, where $\hat{p}_{AIC}$ has been obtained as suggested in section 5.1. For these estimates the criterion function (2.19) is calculated for the `AIC` penalty term $C_T = 2$ and the `BIC` penalty $C_T = \log T$ respectively. The results can be seen in Figure 5.12. From this example we cannot see any striking difference between the accuracy of the order estimation in the context of subspace algorithms and in the context of ML estimation. As has been observed already, IVC and SVC tend to underestimate the order, and this is also true for `BIC` and `AIC`. `AIC` does slightly better than the other procedures.

Finally we also deal with the effects of different weighting schemes on the estimation of the transfer function. Figure 5.13 plots the absolute value of the deviation of the sample mean from the true transfer function (left picture) and the mean of the squared modulus of the differences of the transfer function estimates to k_0 (right picture) for sample size $T = 1000$ and the three different weighting schemes. Here the order is estimated using SVC. Both figures show the superiority of `CCA` in this case. This also true for the other estimation procedures (not documented). A reason for this could be the illconditioning of the weighting matrix $\hat{W}_f^+$ for relatively large values of f and p. For example for $f = 30$ the ratio of largest singular value to smallest singular value for the low pass weighting is of order 10^{13}. For later reference we note, that also in the high frequency region, the estimates using the high pass filter are not better than for the other weighting schemes. Concluding this example, we want to note, that the procedure used in the `N4SID` implimentation in `MATLAB` (see Ljung, 1991) performs surprisingly good in this example, despite of its simple structure. This statement especially holds for the weightings schemes using the pass band filters.

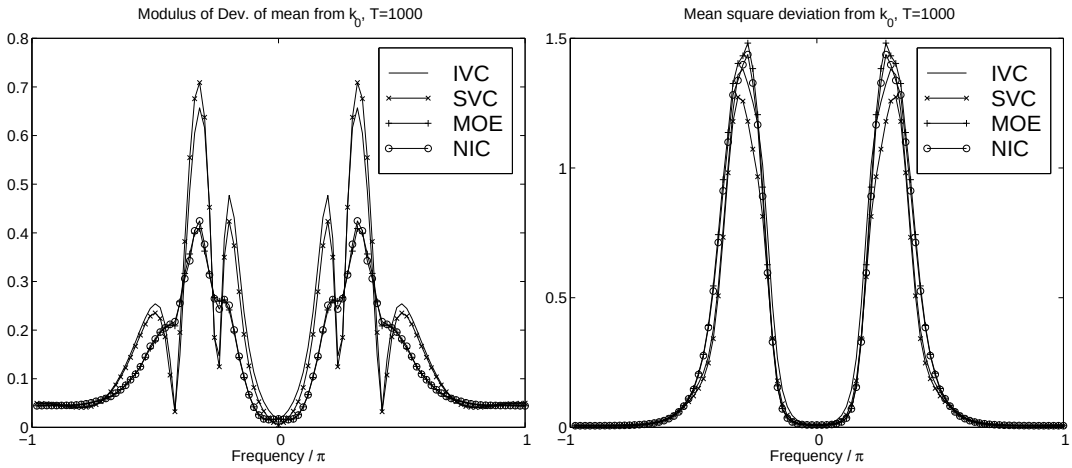

Figure 5.11: Example II: The left figure shows the modulus of the deviation of the mean of the transfer function estimates over the angular frequency range $(-\pi, \pi)$ for 1000 time series of sample size $T = 1000$, where the order is estimated using the four different approaches (see legend). The right figure shows the mean square of the modulus of the deviation of the transfer function estimate to k_0. The transfer functions are estimated using `CCA` and $f = p = 2\hat{p}_{AIC}$.

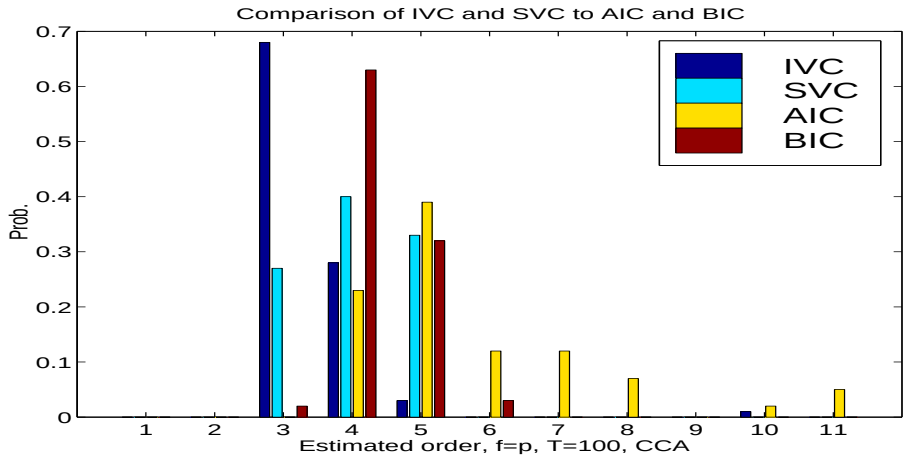

Figure 5.12: Example II: In this figure, the order estimates obtained with IVC and SVC are compared to the estimates obtained in the ML framework using the information criteria `AIC` and `BIC` (see legend). The figure has been obtained using 100 time series of sample length $T = 100$, $f = p = 2\hat{p}_{AIC}$ and `CCA` has been used.

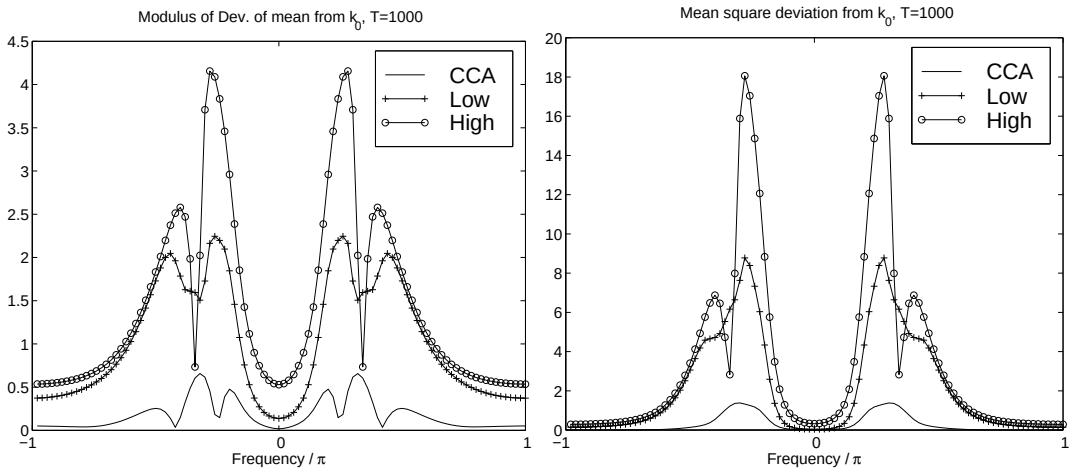

Figure 5.13: Example II: The left figure shows the absolute value of the deviation of the sample mean of the transfer function estimates in the angular frequency range $(-\pi, \pi)$ for 1000 time series of sample size $T = 1000$, where the order is estimated using SVC and three different weighting schemes (see legend). The right figure shows the plot of the mean of the squared modulus of the estimation error.

Example III

The last example, which deals with the estimation of the order in the case of no observed inputs, will be the multi-input multi-output system given by the system matrices (see also Peternell, 1995, Model 4):

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0.25 & 0 & 0.25 \\ 1 & 0 & 0 \end{bmatrix}, K = \begin{bmatrix} 0.00547 & 0.063 \\ 0.119 & 0.157 \\ 0.674 & 0.000666 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, E = I_2$$

This system has a pole-zero configuration as presented in Figure 5.14.

Only one experiment has been performed: The estimation has been performed on 1000 time series of sample length $T = 2000$ with all four estimation procedures for three weighting schemes (`CCA`, `N4SID` and the high pass weight `butter(6,0.5,'high')`). The truncation indices have been estimated as $f = p = 2\hat{p}_{AIC}$ as presented in section 5.1. Figure 5.15 shows the results of the simulation: Again we see the dependence of the performance of all procedures on the weighting scheme used. As in Example I we observe, that in the situation, where f and p are estimated, MOE and NIC lead to bad estimates and that IVC and SVC are superior. IVC leads to good results for the high pass filter, whereas SVC performs best for `CCA` and `N4SID`. For these two weighting schemes, IVC shows a high risk of underestimating the order.

As the bottom line of the three examples we may argue, that the two proposed estimation procedures outperform the existing ones in most cases, especially, when the truncation indices are allowed to depend on the data.

Example IV

Finally we also present an example, where exogenous inputs are present. Consider the following system (see also Peternell, 1995, Model 7):

$$A = \begin{bmatrix} 0.8 & 0.2 \\ -0.4 & -0.5 \end{bmatrix}, B = \begin{bmatrix} 0 & -1 \\ 1 & 0.5 \end{bmatrix}, K = \begin{bmatrix} 1.5 & 0 \\ -0.2 & -0.8 \end{bmatrix}, D = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

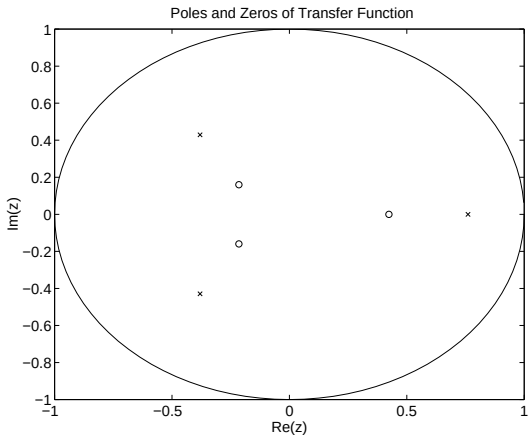

Figure 5.14: Example III: Poles (x) and zeros (o) of k_0.

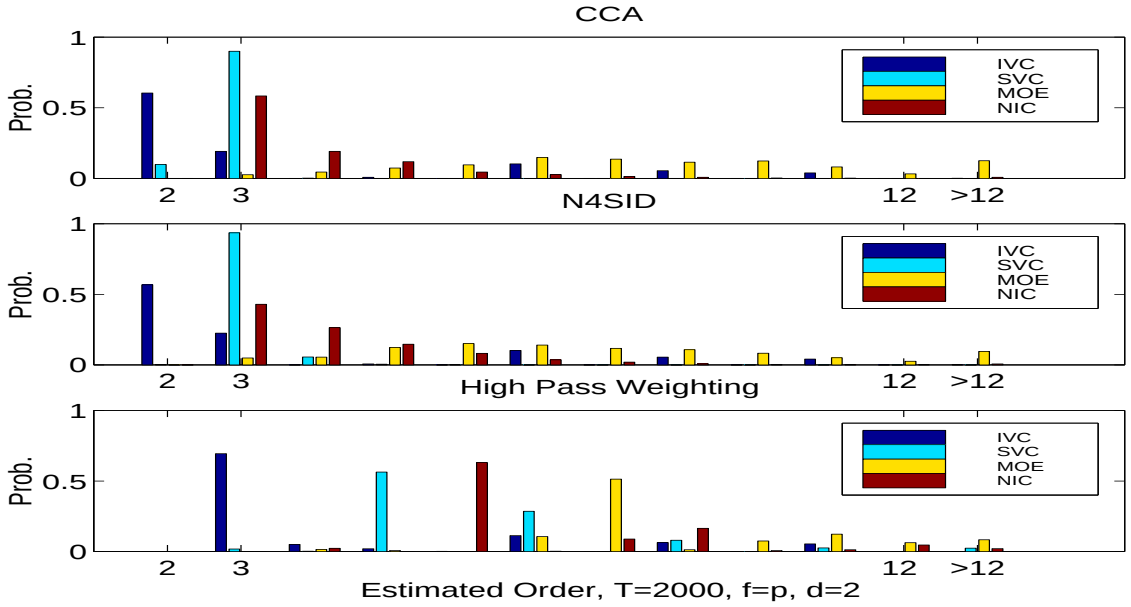

Figure 5.15: Example III: This figure shows the probability distribution of the four different order estimation procedures. The distributions were calculated from 1000 time series of sample size $T = 2000$ using three different weighting schemes: `CCA` (top row), `N4SID` (middle row) and the high pass weighting `butter(6,.5,'high')` (bottom row). $f = p = 2\hat{p}_{AIC}$ was chosen.

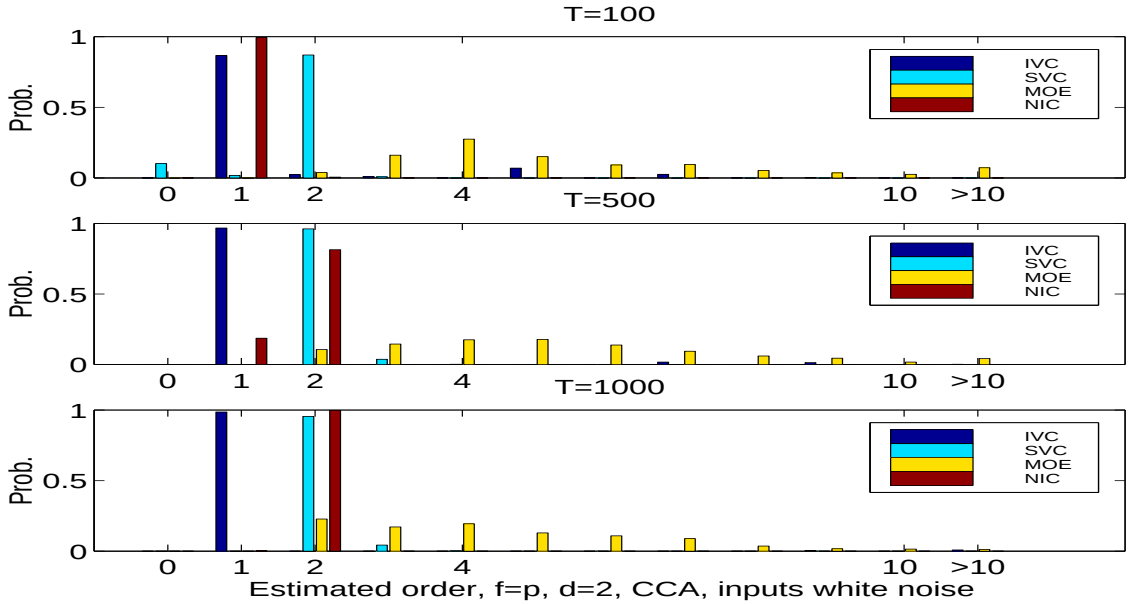

Figure 5.16: Example IV: This figure shows the result of 1000 simulation runs using sample sizes $T = 100$ (top row), $T = 500$ (middle row), $T = 1000$ (bottom row). Each picture shows the probabilities of estimating the order using the four estimation algorithms: The weightings have been chosen according to `CCA`. The truncation indices have been chosen as $f = p = 2\hat{p}_{AIC}$. The inputs are i.i.d. uniformely distributed white noise normalized to zero mean and unit variance.

and $C = E = I_2$, the two-by-two identity matrix. The poles of the system are 0.7352 and -0.4352, the zeros of k_0 are at -0.6583 and 0.2583, whereas the zeros of l_0 are at $-0.1000 \pm 0.9327i$. Figure 5.16 shows the probabilities of the order estimates for different sample sizes, where the input is i.i.d uniformely distributed white noise with zero mean and unit variance. It can be seen, that IVC has a tendency to underestimate the order in all cases, NIC only for $T = 100$. In this example SVC and NIC show the best performance. Note, that for the case of additional exogenous inputs (Peternell, 1995) suggests to use $d(n) = (p(s+m) - n)(fs - n)$ (comp. equation (5.1)), which essentially leads to a bigger penalty term of the form $n(sf + p(m+s))C_T$.

Figure 5.17 shows the results using sample size $T = 500$ for three different weighting schemes: `CCA`, `N4SID` and the high pass weighting `butter(6,0.5,'high')`. In this example, only the `CCA` weighting leads to acceptable performances of the estimation algorithms.

Finally we deal with the situation, where the input process $(u(t))_{t\in\mathbb{Z}}$ is a coloured noise process. The coloured noise has been generated using the following system:

$$A = \begin{bmatrix} 0 & 1 \\ -0.7 & 0.5 \end{bmatrix}, K = \begin{bmatrix} -1.3 & 0 \\ 0 & 1 \end{bmatrix}, C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

using uniformely distributed white noise, which is normalized to mean zero and unit variance. The results are shown in Figure 5.18. It can be seen, that generally the performance now is worse than in the case of white noise inputs, as has been expected. NIC performs best and SVC reasonably. The results obtained by IVC are very poor.

The results in the same setup, when using Gaussian white noise to generate u_t are nearly unchanged (also quantitatively), and thus are not documented.

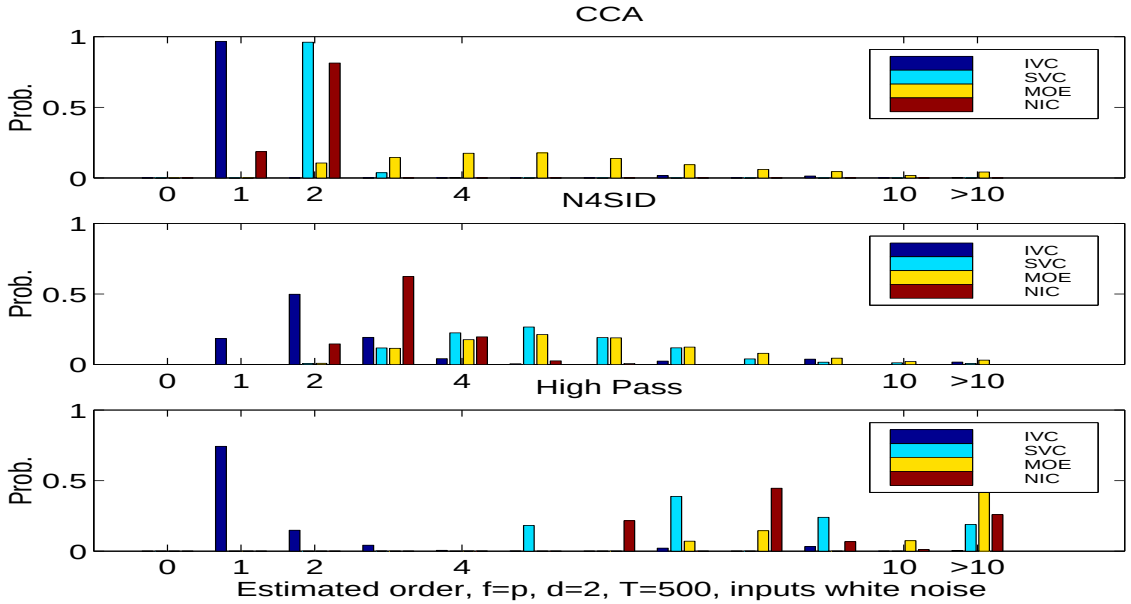

Figure 5.17: Example IV: This figure shows the result of 1000 simulation runs using sample size $T = 500$. Each picture shows the probabilities of estimating the order using the four estimation algorithms: IVC, SVC, MOE and NIC (see legend). The three rows correspond to the different weighting schemes: `CCA` (top row), `N4SID` (middle row) and the high pass weighting `butter(6,.5,'high')` (bottom row). The truncation indices have been chosen as $f = p = 2\hat{p}_{AIC}$. The inputs are i.i.d. uniformely distributed white noise normalized to zero mean and unit variance.

Conclusions

It has been shown in the simulations, that there are situations, where the existing order estimation procedures do not perform very well (see Example I). This led to the development of two new methods: IVC and SVC. Both criteria have been shown to provide consistent estimates in all cases treated in this thesis using suitably chosen penalty terms C_T, however for IVC an additional condition had to be imposed, which is neither conceptually satisfactory, nor can be checked in practice. It has also been shown, that in finite samples also being close to this restriction will lead to bad estimates of the order (see Examples I and IV). Summarizing, the simulation studies in this thesis did not give us a clear picture in favour of one of the estimation algorithms, however they show, that SVC work satisfactory in most situations. We want to emphasize, that especially the choice of the penalty term in the criteria is rather ad hoc and not really motivated from a theoretical point of view. Moreover, we do not claim any optimality for one of the procedures. It is the opinion of the author, that there seems to be need for a further investigation of the properties of the various procedures. As to now, our suggestion would be to use SVC with the penalty term $C_T = \log T$, since this on average led to acceptable performance.

5.3 The Choice of the Weighting Matrices

This section is mainly devoted to the case, that $u(t) = 0, \forall t$ i.e. the case of no observed inputs. The reason for this is the fact, that the theory in this case is very closely related to the frequency weighted balancing, which has been discussed in section 2.2.4. These connections will lead to some very nice results, giving some hints, on how to choose the weighting matrices in order to achieve a desired result. The case of additional observed inputs will only be treated in the simulations part of this section (see section 5.3.3).

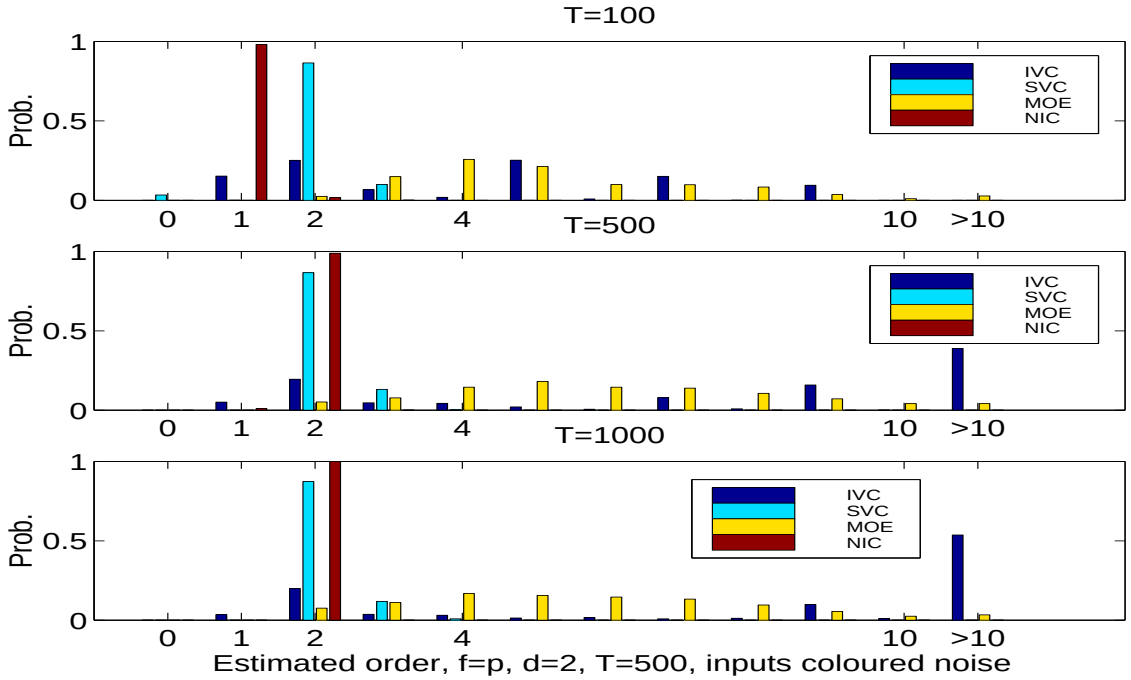

Figure 5.18: Example IV: In this figure the results of the order estimation procedures IVC, SVC, MOE and NIC (see legend) are documented, where the estimation was performed for 1000 data sets of sample size $T = 100$ (top row), $T = 500$ (middle row) and $T = 1000$ (bottom row) using `CCA` and $f = p = 2\hat{p}_{AIC}$ with coloured inputs.

From the description of the algorithm it is clear, that the two weighting matrices, $\hat{W}_f^+$ and $\hat{W}_p^-$ respectively, are user choices in any subspace algorithm. Having the freedom to choose design parameters, it is desirable to know the effects of certain choices on the (asymptotic) properties of the estimates. The discussion will be divided into two different situations: In both situations we will assume that there exists a true system of the form (1.1) generating the process, being finite dimensional with minimal state dimension equal to n. In the first case we will further assume, that we know the true order and perform the estimation with the true order i.e. that the specified order n_{est} is equal to n. Afterwards we deal with the case, that the order specified for estimation is lower than the true order, i.e. $n_{est} < n$. Note, that even in this situation it might be necessary to assume, that an upper bound on the system order is a priori known, namely in the situation, that the row truncation index f is fixed.

5.3.1 Using the True Order in the Estimation

The main result of this thesis describes the asymptotic distribution of the subspace estimates, in several situations corresponding to the truncation indices and the weighting matrices, i.e. asymptotic normality of the estimated system matrices is stated (which also includes a consistency result) on a generic subset U_n^+ of the set $U_n^{(m)}$. For the remaining transfer functions in $U_n^{(m)} - U_n^+$ we still have consistency for the transfer function estimates. The result has been proven in all the situations for two different scenarios corresponding to the weighting matrices. The weighting matrix $\hat{W}_p^-$ has been fixed throughout as $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$, which can be motivated heuristically as follows: This choice is the only choice, that makes the state sequence and the residuals $\hat{\eta}_{t,f}^+$ in $Y_{t,f}^+ = \hat{\mathcal{O}}_f \hat{x}_t(n) + \hat{\eta}_{t,f}^+$ orthogonal for all possible orders, where $\hat{x}_t(n) \in \mathbb{R}^n$, which corresponds to the decomposition into prediction and prediction error. Another motivation for doing so lies in the fact, that this particular choice is apparently the only one supposed in the literature for the main class of algorithms. Also simulation evidence shows the suitability of this choice (comp. Peternell, 1995). However, we want to emphasize, that the main result also holds for quite general choices of the weighting matrix $\hat{W}_p^-$, only the rate of increase of the truncation index p as a function of the sample size has to be adopted. We will not go into details any more on this topic.

Corresponding to the weighting matrix $\hat{W}_f^+$ we discussed two different scenarios in Chapter 4: Either $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ is chosen, which has been denoted as `CCA` throughout the thesis, or $W_f^+ = W_f^+(k_W)$ is used (comp. equation (4.1)), i.e. W_f^+ is chosen to be block Toeplitz, lower triangular, where the blocks are closely related to a weighting transfer function k_W. The weighting transfer function k_W does not need to be stable or strictly minimum-phase for f fixed, however in the case, where f also tends to infinity, we have to impose the stability condition and the strict minimum-phase condition on the weighting k_W in order to guarantee the existence of all the objects occuring in the proof. The rationality of the frequency weighting k_W is only needed for the interpretation of the particular realization, which is the limit for the sample size tending to infinity, in the case, where both truncation indices increase as a function of the sample size, and where no observed inputs are present: In this case the limiting realization will be (output) frequency weighted balanced in the sense of section 2.2.4 (see also Enns, 1985) with frequency weighting k_W. This fact has been noted by (McKelvey, 1995) for realization based algorithms and by (Van Overschee, 1995) for subspace algorithms. Again we want to emphasize, that it is not really difficult to allow for more general weighting matrices, the various proofs give hints on the actually needed assumptions. As mentioned already, we will refrain from being more general, for two reasons: First of all it seems to be rather complicated to give a simple characterization of minimal requirements for the weighting matrix $\hat{W}_f^+$, if we include the possibility of estimating the matrix from data, as well as for the case, where additional observed inputs are present. And secondly, it is necessary to be able to obtain some intuition on how to choose the weighting matrix in order to be able to use the freedom to choose. Thus from now on, let W_f^+ be lower triangular and block Toeplitz, generated by a transfer function $k_W(z) = \sum_{j=0}^{\infty} K_W(j) z^{-j}$ i.e. $W_f^+ = (K_W(i-j))_{i,j=1,\cdots,s}$, $K_W(j) = 0, j < 0, K_W(0)$ nonsingular.

REMARK 5.3.1 Note, that the assumptions on the frequency weighting k_W ensure, that the matrix $W_f^+(k_W)$ is always invertible. As will be clear from section 5.3.2, useful choices of the frequency weighting include band pass filters, which are designed to have good 'cut-off' properties, i.e. to have a high modulus inside the desired frequency range and a low modulus in the remaining frequencies, with a very clear cut between the frequency bands. This can be achieved by placing poles and zeros close to the unit circle. For the weighting matrix W_f^+ this means, that the condition number can be rather high, even for moderate indices f. Thus there are two conflicting aims involved in the choice: On the one hand, one would like to have a very good 'cut-off' property of the frequency weighting and thus poles and zeros as close to the unit circle as possible, on the other hand they should not be too close in order for the calculations to be not too bad conditioned. We want to emphasize, that this point is a merely numerical one, and is not related to asymptotic theory. In practice however, the user will have to take a compromise between these two aims.

Note, that the derivation of the asymptotic distribution of the system matrix estimates also gives us a tool to compare the asymptotic variance for various choices of weighting transfer functions at different frequency points, since the transfer function evaluated at fixed frequency can easily be shown to be a differentiable function of the entries in the matrices (A, B, C, D, E). Comparisons of this type will be given in section 5.3.3.

The analysis so far only concentrated on the situation, where the true system order is known and used for identification. In identification, however, this situation may seem to be rather rare. Even if the system order would be known, there might be reasons to use a different order for the identification, among them the aim to use the system for controller design, where low order systems are needed in order to obtain low order controllers. In such a situation, the analysis above only shows part of the picture, since then the true system will not be in the model class and will have to be approximated by the system in the model class. The asymptotic properties in this situation are discussed in the next section.

5.3.2 Misspecification of the Order

In this section we will use the results derived in the last chapter to investigate the properties of the estimates obtained by the main class of subspace algorithms considered in the thesis, if the order used in identification, n_{est} say, is chosen smaller than the true system order n. In this section we will restrict the discussion to the case, where there are no observed inputs. At the end of the section we will deal also with the case, where observed inputs are present.

From the derivation of the asymptotic properties given in the last chapter we know, that for the correctly specified estimation order the limit of the estimates will be a particular realization of the true system i.e. $\lim_{T\to\infty}(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T) = (A_0, K_0, C_0, E_0)$ a.s. Note, that this realization also depends on the choice of f. In the case $f \to \infty$, the discussion in section 2.2.4 shows, that (A_0, K_0, C_0, E_0) will be frequency weighted balanced with frequency weighting k_W. Given n_{est}, partition these matrices as follows:

$$A_0 = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, K_0 = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}, C_0 = [C_1, C_2] \tag{5.7}$$

and analogously partition $(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T)$. Also let $\Sigma_n = diag(\Sigma_1, \Sigma_2), \Sigma_1 \in \mathbb{R}^{n_{est}\times n_{est}}, \Sigma_2 \in \mathbb{R}^{(n-n_{est})\times(n-n_{est})}$. Here $A_{11} \in \mathbb{R}^{n_{est}\times n_{est}}, A_{12} \in \mathbb{R}^{n_{est}\times(n-n_{est})}, A_{21} \in \mathbb{R}^{(n-n_{est})\times n_{est}}, A_{22} \in \mathbb{R}^{(n-n_{est})\times(n-n_{est})}, K_1 \in \mathbb{R}^{n_{est}\times s}, K_2 \in \mathbb{R}^{(n-n_{est})\times s}, C_1 \in \mathbb{R}^{s\times n_{est}}$ and $C_2 \in \mathbb{R}^{s\times(n-n_{est})}$. The matrices $\hat{\Sigma}_j, \hat{A}_{T,ij}, \hat{K}_{T,i}$ and $\hat{C}_{T,i}$ are of the corresponding dimensions. Let $(\tilde{A}_T, \tilde{K}_T, \tilde{C}_T, \tilde{E}_T)$ denote the estimates of the subspace algorithm using the estimation order $n_{est} < n$. The distribution of these estimates can be seen from an investigation of the algorithm.

Note, that the first time, when the estimation order is used in the algorithm is the formation of the estimate of the state as $\hat{\mathcal{K}}_p Y_{t,p}^-$. Using a lower state dimension corresponds to simply dropping

coordinates of the estimated state vector i.e. $\hat{x}_t = [\hat{x}_{t,1}^T, \hat{x}_{t,2}^T]^T$, where $\hat{x}_{t,1} \in \mathbb{R}^{n_{est}}, \hat{x}_{t,2} \in \mathbb{R}^{n-n_{est}}$. Note, that the two components of the state are orthogonal also in finite samples: $\langle \hat{x}_{t,1}, \hat{x}_{t,2} \rangle = 0$, thus the omission of $\hat{x}_{t,2}$ does not influence the results of the regressions using the first part of the estimated state. Thus $\tilde{C}_T = \hat{C}_{T,1}, \tilde{A}_T = \hat{A}_{T,11}$ follows. Also since $\langle \hat{x}_t, \hat{\varepsilon}_t \rangle = 0$ it follows, that $\tilde{\Omega} = \hat{C}_{T,2}\hat{\Sigma}_2\hat{C}_{T,2}^T + \hat{\Omega}$. Since we derived a central limit theorem for all the involved variables, the (joint) central limit theorems for $\tilde{A}_T, \tilde{C}_T, \tilde{\Omega}$ and the lower triangular Cholesky factor $\tilde{E}_T$ of $\tilde{\Omega}$ are obvious. It remains to calculate the estimate for K_1: Recall, that K_1 is estimated in the regression of $\tilde{x}_{t+1}$ onto the normalized residuals from the first regression $\tilde{\varepsilon}_t = (\tilde{E})^{-1}(y_t - \tilde{C}\tilde{x}_t) = (\tilde{E})^{-1}(y_t - \hat{C}_T\hat{x}_t + \hat{C}_{T,2}\hat{x}_{t,2})$. For the true order n with $\hat{E}^{-1}(y_t - \hat{C}_T\hat{x}_t) = \hat{\varepsilon}_t$ we obtain $\hat{K}_T = \langle \hat{x}_{t+1}, \hat{\varepsilon}_t \rangle$. Thus

$$\begin{aligned} \langle \hat{x}_{t+1,1}, \tilde{\varepsilon}_t \rangle &= \langle \hat{x}_{t+1,1}, \tilde{E}^{-1}(\hat{\varepsilon}_t + \hat{C}_{T,2}\hat{x}_{t,2}) \rangle \\ &= \langle \hat{x}_{t+1,1}, \hat{E}\hat{\varepsilon}_t + \hat{C}_{T,2}\hat{x}_{t,2} \rangle \tilde{E}^{-T} \\ &= (\hat{K}_{T,1}\hat{E}^T + \hat{A}_{T,12}\hat{\Sigma}_2\hat{C}_{T,2}^T)\tilde{E}^{-T} \end{aligned}$$

Again, since for all variables occuring in the above expressions we have obtained central limit theorems as well as strong consistency, the central limit theorem and the strong consistency for $\tilde{K}$ is evident. Thus we have proved the following theorem:

Theorem 5.3.1 *Let the process $(y(t))_{t\in\mathbb{Z}}$ fulfill the conditions of Theorem 4.1.1 (or Theorem 4.2.1 respectively). Also let f and p be functions of T such that the corresponding restrictions hold. Then for the subspace estimates $(\tilde{A}_T, \tilde{K}_T, \tilde{C}_T, \tilde{E}_T)$ using the frequency weighting transfer function k_W and order $n_{est} < n$, where n is the true order of the process, the following holds:*

$$\sqrt{T}\mathrm{vec}[\tilde{A}_T - A_0^r, \tilde{K}_T - K_0^r, \tilde{C}_T - C_0^r, \tilde{E}_T - E_0^r] \xrightarrow{d} Z$$

where Z is a multivariate normal random variable with zero mean and variance $V_f(k_W)$. Here $\Sigma_n = diag(\Sigma_1, \Sigma_2), A_0^r = A_{0,11}, K_0^r = (K_{0,1}E_0^T + A_{0,12}\Sigma_2 C_{0,2}^T)(E_0^r)^{-T}, C_0^r = C_{0,1}, E_0^r(E_0^r)^T = E_0E_0^T + C_{0,2}\Sigma_2 C_{0,2}^T$ and $A_0 = (A_{0,ij}), K_{0,i}, C_{0,i}$ refers to the partitioning of the particular realization (A_0, K_0, C_0, E_0) of k_0 used in the respective theorems, as given in equation (5.7). The analogous result holds, if the estimation is performed using `CCA`.

Furthermore $\|\mathrm{vec}[\tilde{A}_T - A_0^r, \tilde{K}_T - K_0^r, \tilde{C}_T - C_0^r, \tilde{E}_T - E_0^r]\|_2 \to 0$ *holds a.s.*

Note, that the theorem is formulated for both, the case of f fixed and finite as well as the case, that $f \to \infty$ as a function of T. However, the interpretation of these two cases will be different. Also note, that for the case of fixed finite f we still need the assumption $f \geq n$, which means that although the estimation is performed using a lower state dimension, at least an upper bound for the true system order n has to be known. Thus the really interesting case will be the case of $f \to \infty$. In this case the result also provides a nice interpretation: As has been noted several times, the restriction on the weighting matrices of $W_f^+ = W_f^+(k_W)$ and $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{1/2}$ leads to limiting system matrices, which are output frequency weighted balanced (comp. Enns, 1985, and the discussion in section 2.2.4). Thus we may use the results cited there corresponding to the model reduction properties of principal subsystem truncation: As has been cited in section 2.2.5 it is possible to derive bounds on the relative error of the transfer function approximation. Let $k_r(z) = E_0 + C_0^r(zI - A_0^r)^{-1}K_{0,1}$ denote an approximating lower order system, where the matrices refer to the partitioning given in equation (5.7). Then we provided bounds of the form $\sqrt{C_1\sigma_{n_{est}+1}^2 + C_2\sigma_{n_{est}+1}^{3/2}}$ for the relative error $\|k_W(k_0 - k_r)\|_\infty$, where we omitted the argument $e^{i\lambda}$ for notational simplicity. Now corresponding to the formulas given in the theorem we note, that the approximation produced by the subspace algorithm is more complicated than the simple truncation method: The constant is changed to the Cholesky factor of $E_0E_0^T + C_2\Sigma_2C_2$, which in view of section 2.2.5 introduces an additional error $O(\|k_W\|\sigma_{n_{est}+1}^2)$ and a similar statement is true for K_0^r. This leads to an error bound for the approximation error, which shows the dependence of the error on the largest neglected singular value, which thus provides a basis for the choice of the model order in a certain sense.

Of course the bound does not seem to be tight and may not be useful in practice, since the constants depend on the true system, which will not be known. The main reason for the inclusion of this discussion into the thesis lies in the fact, that the connex to frequency weighted model reduction might give hints on how to choose the weighting transfer function, in order to control the distribution of the asymptotic bias over the frequency range.

For the case of additional observed inputs, a similar discussion holds true. Again all the necessary results for a central limit theorem have been given in the last chapter. The actual limits of the reduced order system can be calculated from the knowledge of the true second moments of $x(t), u(t), y(t)$ and $\varepsilon(t)$. However the results do not allow for such a nice interpretation as in the case of no observed inputs and hence are omitted. Simulation results also in this case demonstrate the usefulness of using different weighting matrices in order to enhance different properties of the true system.

5.3.3 Numerical Examples

In this section the theory presented in the last section is applied to a couple of examples, demonstrating the main results given above. As has already been mentioned several times, the asymptotic variance of subspace estimates can be approximated on a computer using the constructive proof of the asymptotic distribution. The variance has been given as the limit of some expressions, which involve mainly infinite sums, where the summands decrease exponentially in norm. Thus we may calculate the sum of the first K, say, terms. This serves as an approximation of the limit. It is also possible to give a bound on the error for each truncated finite sum. The accuracy of the calculated variance is only limited by the computational ressources. This should always be kept in mind, when interpreting the numbers and pictures given in the following. Clearly the accuracy depends heavily on the 'closeness' of the poles and zeros of the systems to the unit circle.

In the following we will provide some calculations, where the output typically will be the plot of some function over the angular frequency range $[-\pi, \pi]$. This will be done, even though the actual expressions have only be calculated on a grid over this frequency range. Also the transfer function is a complex valued function, whereas all the plots we will provide will consider the modulus of the transfer function. For the asymptotic variance note, that the distribution of a scalar transfer function at one fixed angular frequency point can be embedded into $\mathbb{R}^2$ and thus the asymptotic variance at each point in fact is a two-by-two matrix. In order to be able to plot the function however, we chose to plot the trace of this matrix, which is a commonly used possibility. We will always refer to this quantity as the 'absolute value of the asymptotic variance' or even 'the asymptotic variance' in the sequel. Also in the case of no observed inputs we will always plot the asymptotic variance of the estimates of the transfer function $\bar{k}_0 = k_0 E_0^{-1}$. This is done for the simple fact, that these variances will be shown to be independent from the noise distribution in section 5.5. Thus the results are robust with respect to the distribution of the noise.

The Case of no observed inputs

In a first example, we will deal with a system, which has poles at $0.8e^{\pm 0.2i\pi}, 0.79e^{\pm 0.8i\pi}$ and zeros at $0.4, -0.4, \pm 0.6i$. The asymptotic variance of the estimates of the transfer function $\bar{k}_0 E_0^{-1}$ at 100 equally spaced points in the angular frequency range $(-\pi, \pi]$ is shown in Figure 5.19. Here we used `CCA`, a 6-th order Butterworth low pass filter generated with `butter(6,0.5)` (using again the `MATLAB`-code notation) and a 6-th order high pass filter generated with `butter(6,0.5,'high')`. In this example, the results are as expected: The high pass filter results in good accuracy at the high frequency range and the low pass filter achieves a better performance at the low frequency region. Also observe, that `CCA` shows an accuracy close to the minimum of the variance achieved with the two weighting filters.

As a second example, consider the system given in Example I of section 5.2.2. For $f = 15$ the asymptotic variance of the transfer function estimates at 100 equally spaced frequency points in the range $-(\pi, \pi]$ was calculated using three different weighting schemes: `CCA`, the 6-th order low

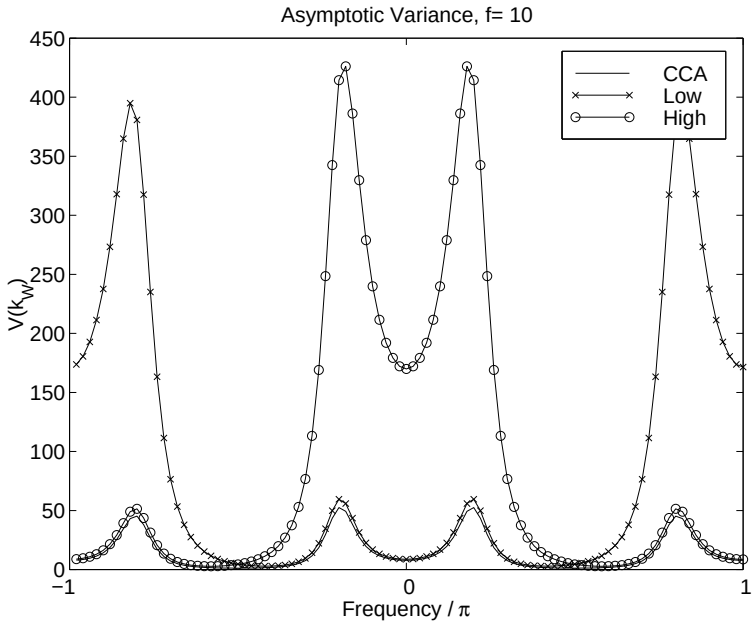

Figure 5.19: Here the asymptotic variance of the transfer function estimates of $\bar{k}_0$ is plotted as a function of the angular frequency for three different weighting schemes: `CCA` (—), a 6-th order Butterworth low pass filter (-x-) and high pass filter (-o-). $f = 10$.

pass Butterworth filter, and the 6-th order high pass butterworth filter. Figure 5.20 shows the result.

In this example, we observe, that the accuracy of the transfer function estimates does not correspond, to what we might have expected: Also for low frequencies, the high pass filter results in a better accuracy. The results also clearly show, that the asymptotic variance of the `CCA` approach is lower at all frequencies.

This fact also relates to the examination of the asymptotic variance of the estimates of the echelon coordinates: Let $F = (I_0)^{-1}$ denote the inverse of the Fisher information matrix, which is the Cramer-Rao bound for the estimates. This may be compared with the asymptotic variance V of the subspace estimates (using `CCA`, $f = 10$), which have been transformed to echelon coordinates:

$$F = \begin{bmatrix} 0.6533 & -0.3691 & 0.0989 & 0.4641 \\ -0.3691 & 1.0117 & 0.1394 & 0.2170 \\ 0.0989 & 0.1394 & 0.9399 & 1.1455 \\ 0.4641 & 0.2170 & 1.1455 & 2.2324 \end{bmatrix}, V = \begin{bmatrix} 0.6544 & -0.3697 & 0.0991 & 0.4649 \\ -0.3697 & 1.0122 & 0.1393 & 0.2167 \\ 0.0991 & 0.1393 & 0.9400 & 1.1457 \\ 0.4649 & 0.2167 & 1.1457 & 2.2331 \end{bmatrix},$$

It can be seen, that the difference is of order 10^{-3}. The difference can be made even smaller by choosing f larger (see section 5.1.1).

Corresponding to the finite sample properties in comparison to ML estimation we refer to the simulations given in (Peternell, 1995) and will only present one example here, using the system occuring in Example II. For this system the mean of the squares of the modulus of the deviations with respect to the true transfer function k_0 are plotted for two different sample sizes, $T = 100$ and $T = 4000$, in Figure 5.21.

It can be observed, that `CCA` outperforms `PEM` (and thus the ML approach) especially for $T = 100$, whereas the differences for $T = 4000$ are less pronounced. Note, that -asymptotically- ML leads to optimal estimates of the transfer function. However, we want to emphasize, that the example has been chosen primarily, because it has occured in the thesis already at several places. It is not conjectured, that `CCA` achieves better results in finite samples, than the ML approach. On

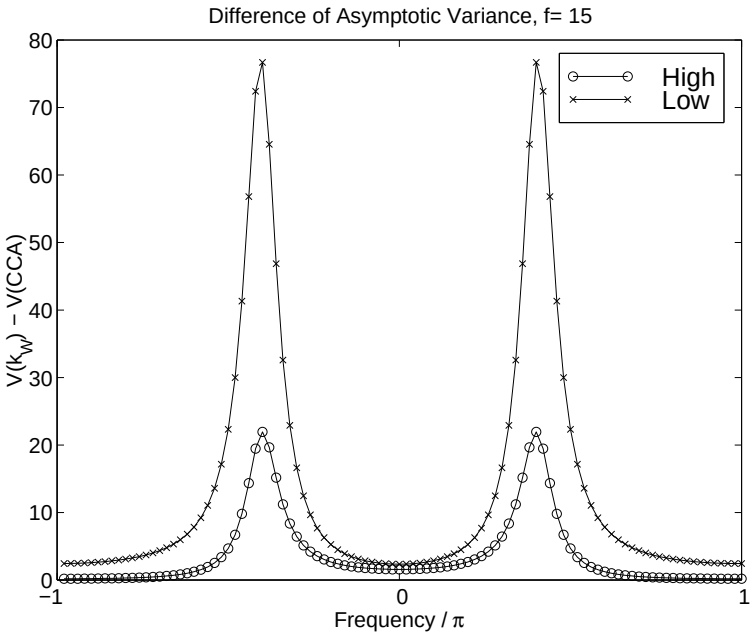

Figure 5.20: Here the difference of the asymptotic variance of the transfer function estimates of $\bar{k}_0$ using a weighting filter k_W minus the asymptotic variance of the transfer function estimates using `CCA` is plotted as a function of the angular frequency for two different weighting schemes: A 6-th order butterworth low pass filter (-x-) and high pass filter (-o-).

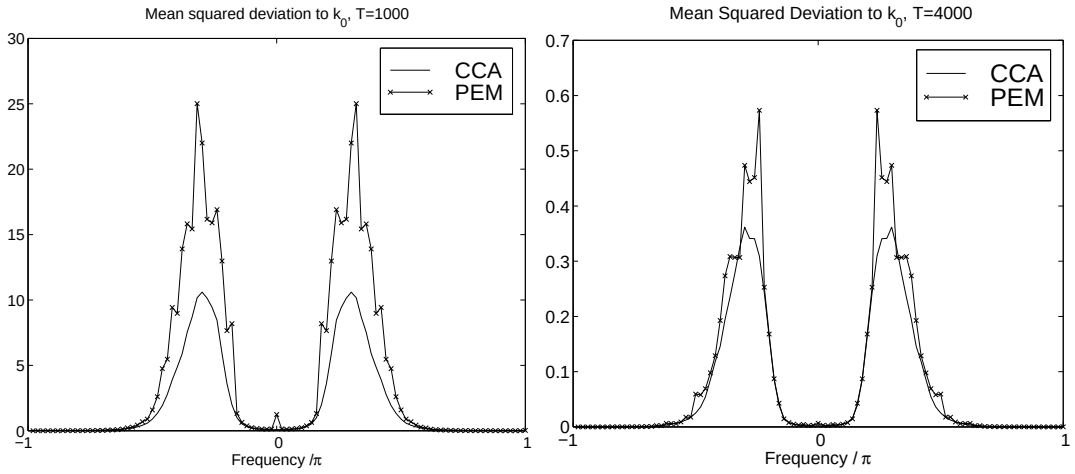

Figure 5.21: Here the mean of the squares of the modulus of the differences of the transfer function estimates of k_0 are plotted. The left picture shows the results for sample size $T = 100$ for `CCA` (—) and `PEM` (-x-), which have been produced using 500 time series. $f = p = \max\{2\hat{p}_{AIC}, 8\}$ has been used, $n = 8$ in all cases. The right picture shows the same plot for $T = 4000$, which has been generated by using 100 times series.

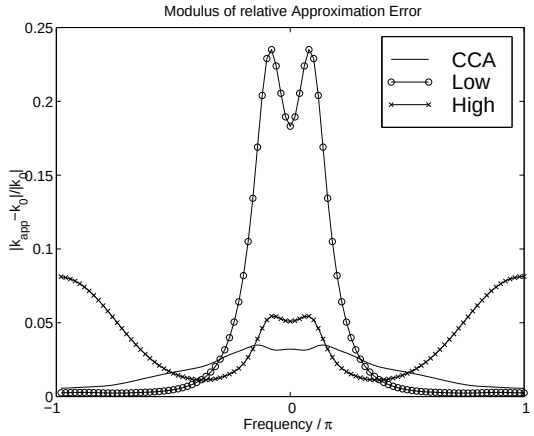

Figure 5.22: In this figure the modulus of the relative error of the 5^{th} order approximation to the true 8^{th} order system used in Example II of section 5.2.2 using three different weighting schemes is plotted as a function of the frequency: `CCA` (—), `butter(12,0.5)` (-x-) and `butter(12,0.5,'high')` (-o-).

the other hand, the example demonstrates, that also the opposite is not true. The reader should also note, that we are dealing here with the idealized situation, that we are given the correct system order.

The model approximation properties as shown in the last section will again be demonstrated for the system used in Example II. In this example an eighth order system has been examined. The estimated orders have been lower than order eight in a relatively large number of cases. Therefore we will approximate the eight order system using fifth order systems and the three weighting schemes also used in section 5.2.2: `CCA`, the low pass filter `butter(12,0.5)` and the high pass filter `butter(12,0.5,'high')`. The resulting relative approximation errors of the transfer function estimates for $f = 10$ are plotted in Figure 5.22. It can be seen clearly, that the low pass weighting provides a better approximation of the true transfer function in the low frequency regions, and the high pass filter results in a good fit at high frequencies. The approximation produced by `CCA` provides a compromise between these two extremes and shows almost constant relative error. This shows, that the frequency weighting transfer function can be used to shape the bias error in a rather intuitive way (as is also supported by the results obtained in the last section). In a simulation study we calculate the estimated transfer function for the three weighting schemes using 500 time series of sample length $T = 1000$ in order to compare the finite sample properties of the estimates. The transfer functions have been estimated using $f = 10, p = 30$ and $n_{est} = 5$. The results can be seen in Figure 5.23. It is interesting to note, that `CCA` obtains a good compromise between the asymptotic variance and the asymptotic bias, which is comparable to the better of the two frequency weighting filters.

The last two calculations in this section relate to a confirmation of the results conjectured in (Peternell, 1995): All the simulation results cited therein show, that `CCA` gives (at least) near to optimal system estimates. Using the tools presented in this thesis, we may calculate the asymptotic variances for given system and thus may confirm e.g. the outcomes of **Model 4** in (Peternell, 1995). We calculated the asymptotic variance of the parameter estimates in echelon coordinates for the `CCA` subspace procedure. Table 5.2 shows the results of the calculations. The simulations in (Peternell, 1995) have been performed using $f = p = \hat{p}_{AIC}$, where $\hat{p}_{AIC}$ is estimated using the `AIC` criterion in an autoregression context, as discussed in section 5.1.

Note, that the fact, that the values for the simulation are sometimes lower than the Cramer Rao bound is no typing error, but maybe caused by the fact, that due to the finite p for the estimation there is still a bias present, whereas the Cramer Rao bound is a lower bound for unbiased estimates.

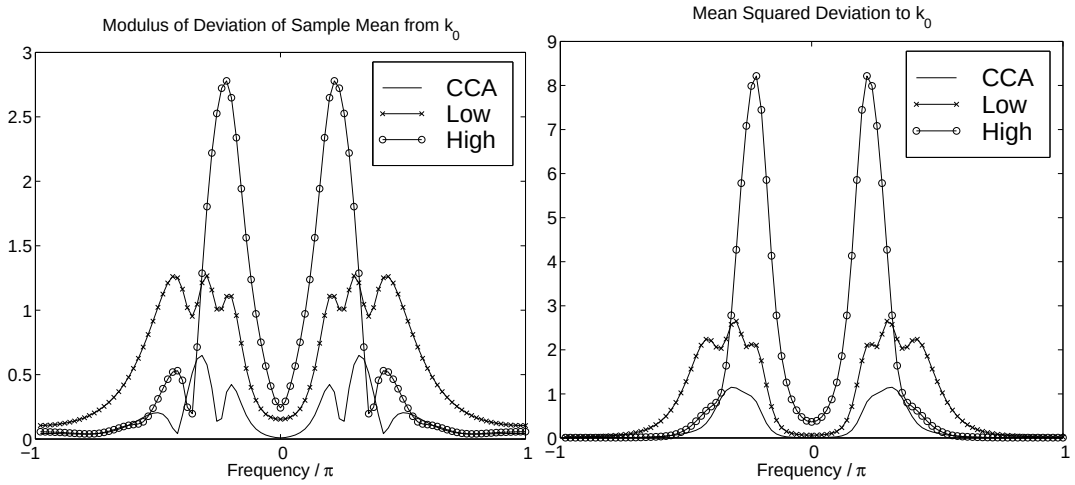

Figure 5.23: The left picture shows the modulus of the deviation of the sample mean from k_0 for Example II of section 5.2.2 using 500 time series of sample length $T = 1000, f = p = 2\hat{p}_{AIC}, n_{est} = 5$ and three different weighting schemes are used: `CCA` (—), `butter(12,0.5)` (-x-) and `butter(12,0.5,'high')` (-o-). The right picture shows the mean of the squares of the modulus of the deviation to k_0 in the same setup.

Elem.	Sim	Var	CR	Elem.	Sim	Var	CR	Elem.	Sim	Var	CR
A_{21}	16.65	16.25	16.24	A_{22}	24.37	22.59	22.58	A_{23}	2.46	2.51	2.51
A_{31}	20.44	19.63	19.62	A_{32}	28.94	27.29	27.28	A_{33}	2.81	2.90	2.90
K_{11}	0.92	0.99	0.99	K_{21}	0.90	0.96	0.95	K_{31}	1.02	1.00	1.00
K_{12}	0.95	0.95	0.95	K_{22}	0.88	0.94	0.94	K_{32}	0.92	1.00	1.00

Table 5.2: In this table the simulation results (Sim) are compared to the asymptotic variance (Var) and to the corresponding Cramer Rao bound (CR) for estimating Model 4 in (Peternell,1995) using `CCA`.

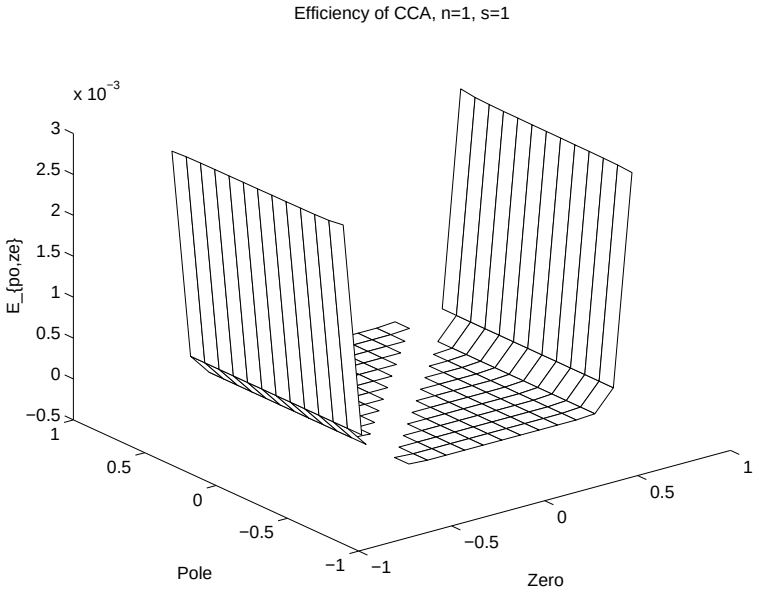

Figure 5.24: Here the measure of the efficiency $E_{po,ze} = tr(V_f I_0) - 2$ of the `CCA` estimates for a scalar system with $n = 1$ is plotted as a function of the zero and the pole.

To conclude the case of no observed inputs we perform some calculations for the simple case of SISO systems of order 1, having $E = 1$. This class of systems can be parametrized using their pole and their zero (see also Bauer *et al.*, 1997b). Figure 5.24 shows the same measure of optimality as has been used in Example 5.1.1 for `CCA` on a grid of the parameter space $|\rho_p| < 0.7, |\rho_z| < 0.7$. Here 0.7 has been chosen, since for greater values of the pole or the zero the truncation of the infinite sums is not very accurate within the bounds imposed by the ressources of the workstation. It can be seen, that the values are nearly equal to zero for pole/zeros near to the origin. For zeros of high modulus we observe, that `CCA` loses efficiency compared to the ML approach. This seems to be due to f finite, as the value of $tr(V_f I_0) - 2$ decreases for increasing f. Thus the conjecture from this plot would be: `CCA` using $f = p$ (and thus $f \to \infty$) is asymptotically equivalent to the ML approach. Unfortunately we do not have an analytic proof of this conjecture. However, all simulation evidence, which has been gathered so far is in no contradiction to the conjecture.

Summing up, the results for the case, where no observed inputs are present, show, that in the case, where the order used for estimation is lower than the true system order, the weighting matrices may be used to shape the bias over the frequency range. In the examples, also the notion of 'frequency weighting transfer function' for k_W is motivated: The results for the low pass filters show good approximation properties in the low frequency range and the high pass filter show good properties for the high frequency range. For the asymptotic variance in the case, where the system order has been specified correctly, the conclusions are not so clearcut. In the examples it has been observed, that the estimates for `CCA` were comparable in all frequency regions to the best results obtained with the filters. This is in accordance with the simulations in (Peternell, 1995), which also show a dominance of `CCA` as well as the simple SISO example, which shows, that `CCA` is (at least) close to optimal in these cases.

The Case of additional observed inputs

Concerning the case of additional inputs, we will deal with the effect of the methods, using the structure in $\mathcal{U}_f$ (see section 3.2) on the asymptotic variance and in a final example show the effect of the weighting matrix in the distribution of the asymptotic bias over the frequency range. We

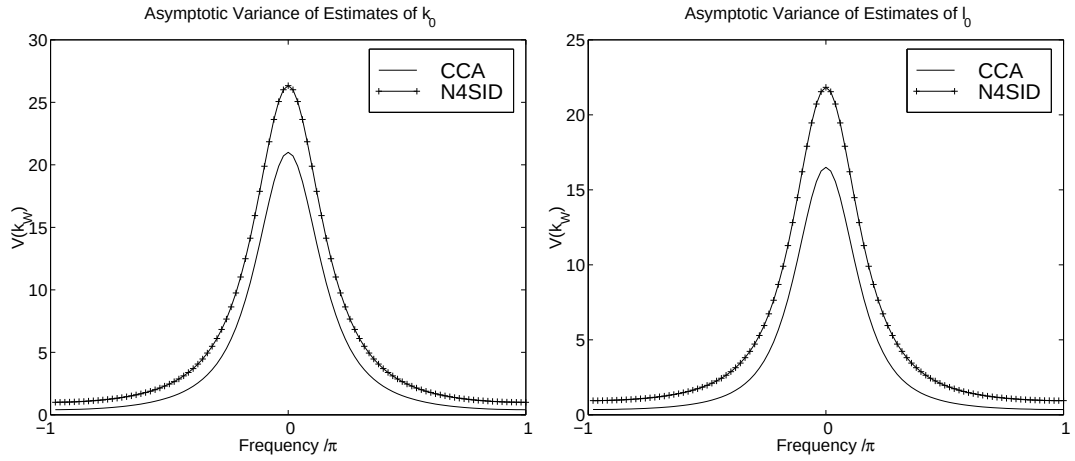

Figure 5.25: Here the modulus of the asymptotic variance of the transfer function estimates at 100 equally spaced angular frequency points is plotted as a function of the angular frequency. $f = 10$ has been chosen and $D = 0.0$ fixed a priori. Two weighting schemes have been used: `CCA` (—) and the weighting corresponding to $k_W = 1$ (-+-). The left plot corresponds to estimation of k_0, the right plot to l_0.

want to emphasize again, that the main concern of this thesis is the case, where no additional inputs are present. Therefore the following calculations are far from covering all aspects, that have been raised in the previous sections. The reason for the inclusion is mainly to demonstrate, that an analysis of these issues for given system is possible using the tools provided in this thesis.

The first example will be the simple SISO example also used in (Peternell, 1995, Model 5):

$$A = 0.5, B = 1.0, C = 1.0, D = 0.0, E = 1.0, K = 1.0$$

For all forthcoming experiments we assume a delay to be present i.e. D is not estimated but set to $D = 0.0$ a priori. Also we will only deal with the choice $\hat{W}_p^- = (\hat{\Pi}_p^-)^{1/2} = (\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$ using the notation of section 4.3. In a first experiment the asymptotic variance of the transfer function estimates at 100 equally spaced frequencies is calculated using Gaussian white noise as input ($u(t)$). The result can be seen in Figure 5.25.

It can be seen, that `CCA` is superior at all frequency points. Also the calculations showed, that there was no difference between the basic approach, the constrained regression and the two stage approach, which uses the basic algorithm in the first step. This is in accordance with the results of (Peternell *et al.*, 1996).

In the next example, the inputs are changed to be coloured noise, which is achieved by filtering white noise through the system used in the Example 5.1.1. The asymptotic variance can be seen from Figure 5.26. In this case the basic approach performs worse than the two other methods corresponding to asymptotic variance. Especially for the transfer function k_0 the estimates of the two stage procedure seem to be more accurate, than using the other two procedures in this case. In contrast to the results obtained in the case of no observed inputs, in this example `CCA` performs worse than `N4SID` for the basic procedure, especially corresponding to the estimation of l. Thus we conclude from this example, that using the methods, which incorporate the structure of $\mathcal{U}_f$, might decrease the asymptotic variance of the transfer function estimates in the case, where the inputs are coloured noise.

For this system we also performed a simulation study. 100 time series of sample size $T = 100$ were generated and estimated using the unstructured `CCA` algorithm (CCA), the constrained regression `CCA` algorithm (CCAC) and the `PEM` implimentation of the system identification toolbox

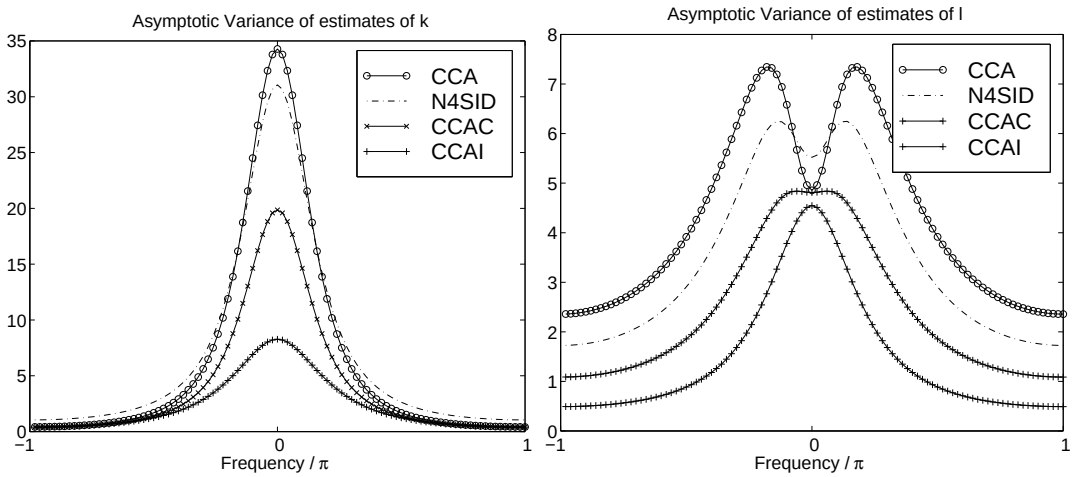

Figure 5.26: Here the absolute value of the variance of the transfer function estimates at 100 equally spaced frequency points is plotted as a function of the angular frequency. $f = 10$ has been chosen and $D = 0.0$ fixed a priori. The basic procedure with the weighting corresponding to `CCA` and with the `N4SID` weights is compared to the constrained regression method (CCAC) and the two stage method using the basic algorithm `CCA` as the first step (CCAI). The left plot corresponds to estimation of k_0, the right plot to l_0. The input is coloured noise.

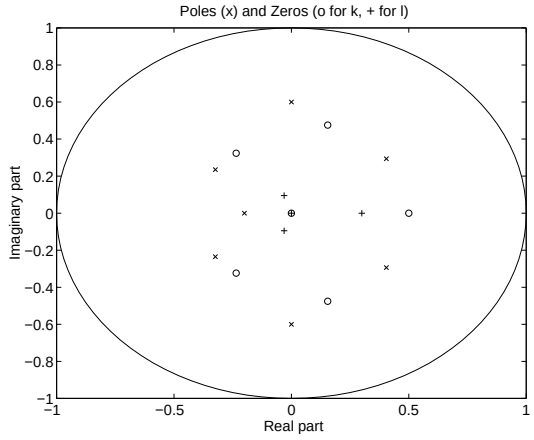

Figure 5.27: Here the poles (x) and the zeros of k_0 (o) as well as the zeros of l_0 (+) for the 7-th order system are shown.

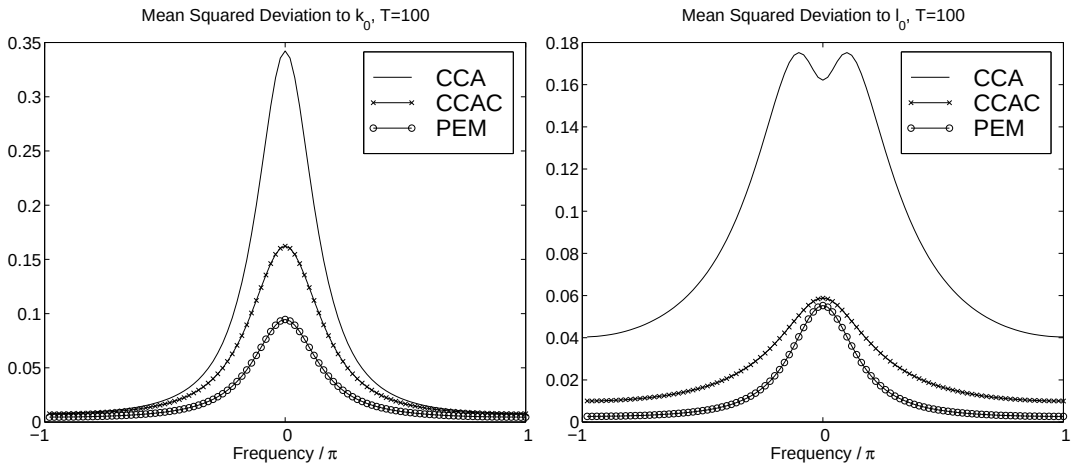

Figure 5.28: The pictures show the mean of the squares of the modulus of the deviation from k_0 (left picture) and l_0 (right picture) respectively. The figures have been produced using 500 time series of sample size $T = 100$ and three different setups: `CCA`, without taking into account the structure of $\mathcal{U}_f$ (—), the constrained regression approach (-x-) and `PEM` (-o-). The subspace estimates have been obtained using $f = p = 2\hat{p}_{AIC}$. The inputs are coloured noise.

of `MATLAB` (Ljung, 1991). The simulation results confirmed the asymptotic results, as can be seen from Figure 5.28.

As a final example consider the 7-th order system, having a pole/zero configuration as depicted in Figure 5.27. This system has been approximated by a 4-th order system using the subspace algorithms of the main class, where both f and p are allowed to tend to infinity. The inputs are obtained by filtering white noise with unit variance through the system considered in Example I in section 5.2.2. As weighting matrices we use the `CCA` approach and two filter-based weightings $W_f^+(k_W)$, where k_W corresponds either to the low pass weighting `butter(6,0.5)` or to the high pass weighting `butter(6,0.5,'high')` again using `MATLAB` code notation. Note, that the approximation depends on the spectral density of the observed inputs. The resulting approximation errors can be seen from Figure 5.29.

The results can be interpreted as follows: The various weighting schemes result in the expected behaviour for the transfer function k_0. The high pass filter results in a good approximation in the high frequency range and a bad approximation for the low frequencies, whereas the behaviour corresponding to the low pass filter is the opposite. `CCA` achieves a compromise between the two other weighting schemes. For the transfer function l_0 the results are similar, however the performance degradation in the low frequency region for the high pass weighting is much lower.

As a resumee of the calculations we may state the following: For the estimation of the transfer function k_0 it seems to be the case, that for white noise inputs the more complex methods (constrained regression and two stage procedure) do not result in better performance, whereas the examples show a decrease in the asymptotic variance in the case of coloured input sequences. Also `CCA` outperforms `N4SID` in the first example (for white noise inputs). However, this is not the case for coloured inputs as is obvious from the example. Corresponding to the estimation of l_0 also the comparison with other estimation algorithms like `MOESP` should be considered in order to judge the properties of the main class of algorithms. Concluding we have to admit, that little is known about a comparison of the various approaches corresponding to their second order asymptotic properties. In this field there remain many open questions.

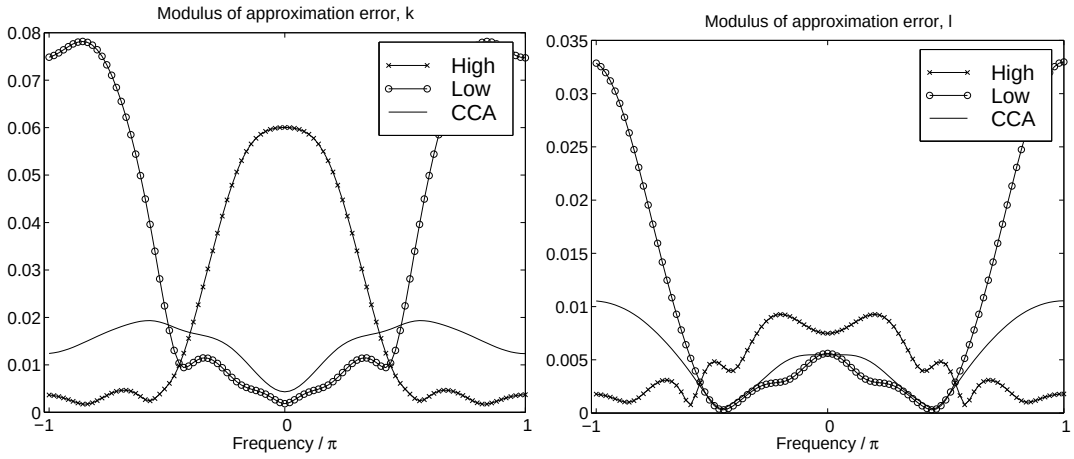

Figure 5.29: Here the modulus of the approximation error of a 4-th order approximation to the 7-th order true system is plotted for the transfer function k (left plot) and for l (right plot). The approximations have been calculated using the main class of subspace algorithms, using $f = p = \infty$ and three different weightings: `CCA` (—), the low pass filter `butter(6,0.5)` (-o-) and the high pass filter `butter(6,0.5,'high')` (-x-).

5.4 Stability and Minimum-phase Property

Throughout the discussion we assumed, that the true transfer function k_0 is stable and strictly minimum-phase. In the following we will derive some sufficient condition for the estimates to be stable or strictly minimum-phase in the case of unobserved white noise inputs only, i.e. we will not deal with the case, where observed inputs are present. This is followed by some simulation experiments.

5.4.1 Theoretical Considerations

This section is based entirely on the results stated in (Chui and Maciejowski, 1996). Let $\mathcal{O}_f = [C^T, A^T C^T, \cdots, C^T (A^T)^{f-1}]^T$ denote the observability matrix and let $\hat{\mathcal{O}}_f$ denote an estimate of $\mathcal{O}_f$. Define $\bar{\mathcal{O}}_f = [A^T C^T, \cdots, (A^T)^{f-2} C^T, 0]^T$ and analogously define $\hat{\bar{\mathcal{O}}}_f$. Using this notation, (Chui and Maciejowski, 1996) state the following result:

Theorem 5.4.1 (Chui & Maciejowski) *For $\hat{A} = \hat{\mathcal{O}}_f^{\dagger} \hat{\bar{\mathcal{O}}}_f$, where † denotes the Moore-Penrose inverse, $|\lambda_{max}(\hat{A})| \leq 1$.*

For a proof see (Chui and Maciejowski, 1996). Thus when using the shift invariance approach to estimate A, we achieve a (marginally) stable estimate by augmenting the shifted matrix with zeros. This might be of interest in the case of the realization approaches, where the stability of $\hat{A}$ is required for the spectral factorization step. However, note, that in general for fixed f the estimate $\hat{A}$ might be biased due to appending the zeros. Clearly the bias decreases with increasing f, such that $f \to \infty$ leads to consistent estimates of A even using this approach.

In the main class of algorithms treated in this thesis, the shift invariance approach is not used, but the state is estimated in a first step, and estimates of the system matrices are obtained from the state estimate using regression in the system equations (1.1). In section 3.4 we already noticed, that the estimate of the shifted state $\tilde{x}_{t+1}$ need not necessarily be equal to $\hat{x}_{t+1}$, the shifted state estimate. Thus consider the estimates of $\hat{A}, \hat{K}, \hat{C}$ for $\hat{x}_t = \hat{\mathcal{K}}_p Y_{t,p}^-$ and $\tilde{x}_{t+1} = \tilde{\mathcal{K}}_{p+1} Y_{t+1,p+1}^- = \tilde{\mathcal{K}}_{p+1}(1) y_t + \tilde{\mathcal{K}}_{p+1}(2) Y_{t,p}^-$, where $\tilde{\mathcal{K}}_{p+1}(1) \in \mathbb{R}^{n \times s}$ denotes the first block column of $\tilde{\mathcal{K}}_{p+1}$ and

$\tilde{\mathcal{K}}_{p+1}(2) \in \mathbb{R}^{n \times sp}$ the remaining block columns: $\hat{C} = \langle y_t, \hat{x}_t \rangle \langle \hat{x}_t, \hat{x}_t \rangle^{-1}$ is not affected by the choice of $\tilde{\mathcal{K}}$. For $\hat{A}$ we obtain from straightforward calculations:

$$\begin{aligned} \hat{A} &= \langle \tilde{x}_{t+1}, \hat{x}_t \rangle \langle \hat{x}_t, \hat{x}_t \rangle^{-1} = [\tilde{\mathcal{K}}_{p+1}(1), \tilde{\mathcal{K}}_{p+1}(2)] \begin{bmatrix} \hat{\mathcal{H}}_{1,p} \\ \hat{\Gamma}_p^- \end{bmatrix} \hat{\mathcal{K}}_p^T \langle \hat{x}_t, \hat{x}_t \rangle^{-1} \\ &= \tilde{\mathcal{K}}_{p+1}(1)\hat{C} + \tilde{\mathcal{K}}_{p+1}(2)\hat{\Gamma}_p^- \hat{\mathcal{K}}_p^T \hat{\Sigma}_n^{-1} \end{aligned} \tag{5.8}$$

Here we use $\langle \hat{x}_t, \hat{x}_t \rangle = \hat{\mathcal{K}}_p \hat{\Gamma}_p^- \hat{\mathcal{K}}_p^T = \hat{\Sigma}_n$. Corresponding to $\hat{K}$ we obtain:

$$\begin{aligned} \hat{K}\hat{E}^{-1} &= \langle \tilde{x}_{t+1}, \hat{\nu}_t \rangle \hat{\Omega}^{-1} = \tilde{\mathcal{K}}_{p+1} \begin{bmatrix} \hat{\gamma}(0) & \hat{\mathcal{H}}_{1,p} \\ \hat{\mathcal{H}}_{1,p}^T & \hat{\Gamma}_p^- \end{bmatrix} \begin{bmatrix} I_s \\ -\hat{\mathcal{K}}_p^T \hat{C}^T \end{bmatrix} \hat{\Omega}^{-1} \\ &= \tilde{\mathcal{K}}_{p+1}(1) + \tilde{\mathcal{K}}_{p+1}(2)(\hat{\Gamma}_p^-)^{1/2}[(\hat{\Gamma}_p^-)^{-1/2}\hat{\mathcal{H}}_{f,p}^T(\hat{W}_f^+)^T - \hat{V}_n \hat{\Sigma}_n \hat{U}_n^T](\hat{W}_f^+)^{-T} \begin{bmatrix} I_s \\ 0 \end{bmatrix} \hat{\Omega}^{-1} \end{aligned} \tag{5.9}$$

where $\hat{\nu}_t = y_t - \hat{C}\hat{x}_t$ and $\hat{\Omega} = \langle \hat{\nu}_t, \hat{\nu}_t \rangle$. These two equations are the basis for sufficient conditions for the identified system to be minimum-phase: For the true matrices $\mathcal{K}_p$ and $\mathcal{K}_{p+1}(2)$ we have a shift invariance property i.e. $\mathcal{K}_{p+1}(2) = (A - KE^{-1}C)\mathcal{K}_p$. This equation could be used to obtain the estimate $\tilde{\mathcal{K}}_{p+1}(2)$ e.g. in the following way: $\tilde{\mathcal{K}}_{p+1}(2) = [\vec{\mathcal{K}}_p, 0]\hat{\mathcal{K}}_p^\dagger \hat{\mathcal{K}}_p$. Here again † denotes the Moore-Penrose inverse and $\vec{\mathcal{K}}_p$ denotes the matrix obtained from $\hat{\mathcal{K}}_p$ by omitting the first block column. Setting $\tilde{\mathcal{K}}_{p+1}(1) = \hat{\mathcal{K}}_p(1) \in \mathbb{R}^{n \times s}$, where $\hat{\mathcal{K}}_p(1)$ denotes the first s columns of $\hat{\mathcal{K}}_p$, we obtain from equations (5.8,5.9): $\hat{K}\hat{E}^{-1} = \hat{\mathcal{K}}_p(1)$, $\hat{A} = \hat{K}\hat{E}^{-1}\hat{C} + [\vec{\mathcal{K}}_p, 0]\hat{\mathcal{K}}_p^\dagger$. Then it follows from the Theorem given above, that the estimated matrix $\hat{A} - \hat{K}\hat{E}^{-1}\hat{C}$ is marginally stable i.e. has all the roots inside or on the unit circle. Thus we have shown the following corollary:

Corollary 5.4.1 *If $\tilde{\mathcal{K}}_{p+1} = [\tilde{\mathcal{K}}_{p+1}(1), \tilde{\mathcal{K}}_{p+1}(2)]$ is chosen as $\tilde{\mathcal{K}}_{p+1}(1) = \hat{\mathcal{K}}_p(1), \tilde{\mathcal{K}}_{p+1}(2) = [\vec{\mathcal{K}}_p, 0]\hat{\mathcal{K}}_p^\dagger \hat{\mathcal{K}}_p$, then the resulting estimated system $(\hat{A}, \hat{K}, \hat{C}, \hat{E})$ is minimum-phase, i.e. $|\lambda_{max}(\hat{A} - \hat{K}\hat{E}^{-1}\hat{C})| \leq 1$.*

If $\tilde{x}_{t+1} = \hat{x}_{t+1}$ is used, then the resulting matrix $\hat{A}$ will be marginally stable i.e. $|\lambda_{max}(\hat{A})| \leq 1$.

PROOF: The proof of the first point has been given above the Corollary. The second statement follows immediately from the marginal stability of the Yule-Walker estimates (see e.g. Hannan and Deistler, 1988). □

In the case of the main class of algorithms the augmentation of $\vec{\mathcal{K}}_p$ with a zero block does not seem to be restrictive, since $p \to \infty$ in the algorithm. Also in the case $\tilde{x}_{t+1} = \hat{x}_t$ we use $\tilde{\mathcal{K}}_{p+1} = [\hat{\mathcal{K}}_p, 0]$. Note, that the procedure ensuring the minimum-phase condition to hold for the estimates can be interpreted as a realization algorithm applied to $\hat{\mathcal{O}}_f \hat{\mathcal{K}}_p$. Of course the simplest method to ensure, that the minimum-phase condition is satisfied for the estimated system, under the assumption, that the estimates $\hat{A}$ is stable, i.e. $|\lambda_{max}(\hat{A})| < 1$ holds, would be to calculate the covariance function associated with the estimated system $(\hat{A}, \hat{K}, \hat{C}, \hat{E})$ and to perform a spectral factorization using the obtained covariance sequence. This idea can also be applied in the case, where additional exogenous inputs are present, since the matrix pair $(\hat{A}, \hat{C})$ is not changed by this procedure.

Again the fundamental problem with the various proposed methods of choosing $\tilde{\mathcal{K}}_{p+1}$ is the lack of a comparison of the relative efficiency. To the best of the authors knowledge, there is no theory available corresponding to the effects of the various choices for $\tilde{\mathcal{K}}_{p+1}$ on the asymptotic variance. Thus we have no recommendation, which of the proposed methods leads to the best estimates (in the sense of smallest asymptotic variance).

5.4.2 Simulation Evidence

As an example consider the fourth order SISO system having poles at $z = \pm 0.5i, z = 0.3e^{\pm 0.2i\pi}$ and zeros at $z = 0.97e^{\pm 0.6i\pi}, z = 0.99e^{\pm 0.8i\pi}$. Thus the system is nearly non-minimum-phase. The

$f=p$	$n=2$		$n=3$		$n=4$		$n=5$	
	T=500	T=2000	T=500	T=2000	T=500	T=2000	T=500	T=2000
4	7	7	0	0	0	0	0	0
5	83	99	32	18	0	0	0	0
6	73	100	53	42	1	0	0	0
7	79	100	57	50	20	4	0	0
8	70	100	61	55	50	20	3	0

Table 5.3: In this table, the number of unstable zeros of the estimated transfer function is documented for `N4SID` applied to 100 time series of sample size T with n and $f = p$ as specified above.

finite sample properties of the subspace estimates of the system poles and zeros are demonstrated in the following setup: For each value of $1 \leq n \leq 8$ and $1 \leq f = p \leq 8, f \geq n$ the estimated poles and zeros are calculated for 100 time series generated by filtering Gaussian white noise through the system using `CCA` and `N4SID` for two different sample sizes, $T = 500$ and $T = 2000$. The system poles were all smaller than one in modulus, as is clear from the last section due to the choice $\tilde{x}_{t+1} = \hat{x}_{t+1}$. For `CCA` there was not a single occurrence of an unstable zero. On the other hand, for `N4SID` we obtained the results as documented in Table 5.3. It can be observed, that especially for $n = 2$ the probability of estimating a non-minimum-phase system is very high and even increases with sample size. Note that in this case the true transfer function is not contained in the model class and therefore the system has to be approximated using the lower order. As has been stated already, this approximation depends on the weighting scheme chosen (see section 5.3). Obviously the approximation corresponding to `N4SID` contains a unstable zeros. For correctly specified n it is striking, that the probability of unstable zeros increases with increasing $f = p$: Whereas the risk for $f = p = 6$ is negligible, this is not true for $f = p = 7$, where the risk is as high as 20% for $T = 500$. For $f = p = 8$ the same is true also for the larger sample size. Thus again we observe a big influence of the choice of the weighting scheme and the truncation indices on the properties of the estimates.

5.5 The Influence of the Distribution of the Noise

As has already been stated, for the ML approach the asymptotic distribution of the estimates of the transfer function $\bar{k}_0 = k_0 E_0^{-1}$ is independent of the distribution of the noise, and in particular of the fourth moments of the noise. From the proof of the asymptotic normality for the estimates obtained by subspace algorithms it is clear, that the asymptotic distribution of the estimates of $\bar{k}_0$ is independent of the fourth moment of the noise if and only if e_0 (see the proof of Lemma 4.1.4) can be eliminated in the expressions, without changing the asymptotics. Therefore we will investigate the effects of e_0 on the asymptotic distribution of $(\hat{A}, \hat{\bar{K}}, \hat{C})$, where here $\hat{\bar{K}} = \hat{K}\hat{E}^{-1}$. Since we are only interested in the transfer function estimates, we are free to choose as basis of the state space, in which the analysis will be performed. Thus we will consider a slightly different version of the main class of subspace procedure in this section, which can be outlined as follows (compare also Chapter 3):

- Estimate $\hat{\beta}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}$.
- Find a rank n approximation $\hat{\mathcal{O}}_f\hat{\mathcal{K}}_p = (\hat{W}_f^+)^{-1}\hat{U}_n\hat{\Sigma}_n\hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2}$, where $\hat{U}_n, \hat{\Sigma}_n$ and $\hat{V}_n$ correspond to the SVD of $\hat{W}_f^+\hat{\beta}(\hat{\Gamma}_p^-)^{1/2} = \hat{U}\hat{\Sigma}\hat{V}^T$. Here $\hat{\mathcal{K}}_p = (\hat{\Sigma}_n)^{1/2}\hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2}$.
- Use the estimate of the state $\hat{x}_t = \hat{\mathcal{K}}_p Y_{t,p}^-$ in the system equations (1.1) to obtain estimates $(\hat{A}_T, \hat{K}_T, \hat{C}_T, \hat{E}_T)$.

- Transform the system matrix estimates to the coordinates corresponding to the overlapping parametrization discussed in (Hannan and Deistler, 1988, Chapter 2).

Thus the difference to the main class of algorithms is that the estimates will be transformed to coordinates in some overlapping parametrization. These paramterizations are based on 'nice' selections of the basis in the Hankel matrix, and are closely related to the echelon parametrization. The only difference being that not the uniquely defined Kronecker indices are used to define the partitioning of M_n, but any 'nice' selection of basis rows of the Hankel matrix $\bar{\mathcal{H}}$, where 'nice' means, that the basis can be described by the multiindex $\alpha = (n_1, \cdots, n_s)$. Therefore there exists for every transfer function $k \in M_n$ several possible multiindices and thus several realizations. The reason for using overlapping parametrizations will be apparent in the following. It has been shown in Chapter 4, that asymptotically the system matrix estimates have the same distribution as a linear function of the sample covariance estimates $\hat{\gamma}(i), 0 \leq i \leq f+p-1$. It follows from the proof of Lemma 4.1.4, that it is possible to reformulate the result in terms of the sample covariances of the noise process $\varepsilon(t)$ i.e. that

$$\sqrt{T}\mathrm{vec}[\hat{A}_T - A_0, \hat{K}_T - K_0, \hat{C}_T - C_0, \hat{E}_T - E_0] = L_h \mathrm{vec}[e_0, e_1, \cdots, e_h] + o_P(1)$$

where $L_h \to L$ and where $e_j = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \varepsilon(t)\varepsilon(t-j)^T, j > 0, e_0 = \frac{1}{\sqrt{T}} \sum_{t=1}^{T} (\varepsilon(t)\varepsilon(t)^T - I)$. Thus in order to show, that the fourth moments of the noise and thus also the distribution of the noise has no impact on the asymptotic variance, it is sufficient to show, that the columns of L_h corresponding to e_0 converge to zero sufficiently fast, and thus, since it has already been shown, that the columns of L_h converge sufficiently fast, that the limit is zero, i.e. that the corresponding columns of L are zero. This will be shown next. Here sufficiently fast means, that $\|L_h - L\|_1 Q_T \to 0$, which follows from the arguments given in Chapter 4. In this respect we will adopt the following notation: Since L_h is defined by the mapping attaching the vector $\mathrm{vec}[e_0, \cdots, e_h]$ to a vector $x \in \mathbb{R}^{n^2+2ns+s(s+1)/2}$, such that the distribution of this function is asymptotically the same as the distribution of the estimation errors, the columns of L_h corresponding to elements of e_0 can be interpreted as the derivatives of this mapping with respect to this entry in e_0. Thus we will in the following use notations such as ΔA, meaning the first n^2 entries in the particular column of L, which correspond to the element of e_0 under consideration. These entries correspond in turn to the estimation of A.

Consider the (i,j) element of e_0, $e_{0,i,j}$ say, for example. Thus e.g. to give an explicit example, $\Delta\gamma(s) = \sum_{l=0}^{\infty} K(l+s)_i K(l)_j^T, s > 0$, where $K(j)_i$ denotes the i-th row of $K(j)$ (see equation (4.12)). Additionally we will define the following matrices:

$$\bar{\mathcal{H}} = [K(i+j-1)]_{i,j\in\mathbb{N}}, \mathcal{T}_- = [K(j-i)]_{i,j\in\mathbb{N}}, E = [e_{i-j}]_{i,j\in\mathbb{N}}$$

where $K(j) = 0, j < 0$. Thus $\bar{\mathcal{H}}$ denotes the block Hankel matrix of the Markoff parameters, $\mathcal{T}_-$ denotes the upper triangular block Toeplitz matrix containing the Markoff parameters and E denotes a block Toeplitz matrix containing the e_i as blocks. Thus e.g. $\sqrt{T}(\hat{\mathcal{H}}_{f,p} - \mathcal{H}_{f,p}) \doteq \bar{\mathcal{H}}_f E \mathcal{T}_{-,p}^T$, where $\bar{\mathcal{H}}_f$ denotes the first f block rows of $\bar{\mathcal{H}}$, where $\mathcal{T}_{-,p}$ denotes the first p block rows of $\mathcal{T}_-$ and $\doteq$ denotes equality up to terms of order $o_P(1)$ (see equation (4.12)). Thus $\Delta\mathcal{H} = \bar{\mathcal{H}}\Delta E \mathcal{T}_-^T$ follows from straightforward calculations in the sense, that the equation holds for each element of $\Delta\mathcal{H}$. Also the equations $\Delta\Gamma^- = \mathcal{T}_-\Delta E \mathcal{T}_-^T$ and $\mathcal{H} = \bar{\mathcal{H}}\mathcal{T}_-^T, \Gamma^- = \mathcal{T}_-\mathcal{T}_-^T$ and $\beta\mathcal{T}_- = \bar{\mathcal{H}}$, where $\beta = \mathcal{H}(\Gamma^-)^{-1}$, can be shown. Thus we obtain

$$\begin{aligned} \Delta(\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T) &= \bar{\mathcal{H}}\Delta E\mathcal{T}_-^T\beta^T - \beta\mathcal{T}_-\Delta E\mathcal{T}_-^T\beta^T + \beta\mathcal{T}_-\Delta E\bar{\mathcal{H}}^T \\ &= \bar{\mathcal{H}}\Delta E\bar{\mathcal{H}}^T \end{aligned}$$

In the algorithm the eigenvalue decomposition of $W^+\mathcal{H}(\Gamma-)^{-1}\mathcal{H}^T(W^+)^T$ is used. For the moment assume, that $W_f^+ = W_f^+(k_W)$ is used. We will deal with the `CCA` case afterwards. For this choice of the weighting matrix it is obvious, that $\Delta W^+ = 0$ holds. Thus the limit of the derivative of the matrix, which is decomposed in the algorithm is equal to $W^+\bar{\mathcal{H}}\Delta E\bar{\mathcal{H}}^T(W^+)^T$ in this case. This results in the following two expressions for the derivative of the i-th eigenvalue, λ_i say, and the i-th eigenvector, u_i say, respectively $(1 \leq i \leq n)$ again using equations (B.3,B.4):

$$\begin{aligned}\Delta\lambda_i &= u_i^T W^+ \bar{\mathcal{H}} \Delta E \bar{\mathcal{H}}^T (W^+)^T u_i \\ \Delta u_i &= \sum_{j\neq i} \tfrac{1}{\lambda_i - \lambda_j} u_j u_j^T W^+ \bar{\mathcal{H}} \Delta E \bar{\mathcal{H}}^T (W^+)^T u_i\end{aligned}$$

We obtain for the i-th column $\mathcal{O}_i$ of $\mathcal{O} = (W^+)^{-1}U_n$ (respectively $\mathcal{O}_f$, if f is fixed) $\Delta\mathcal{O}_i = (W^+)^{-1}\Delta u_i$. Here we -with slight abuse of notation- changed the state basis by omitting the factor $\Sigma^{1/2}$. However, since we are only interested in the distribution of the transfer function estimates, a change in the state basis does not matter. Consider the observability matrix corresponding to the coordinates in the overlapping parametrization: $\mathcal{O}_e = \mathcal{O}T^{-1}$, where T denotes the state space transformation, which corresponds to a change of the coordinates to the overlapping coordinates. In the sequel, matrices, which correspond to overlapping coordinates are indicated with a subscript $_e$. For every transfer function $k_0 \in U_n^{(m)}$ there exists a piece, U_α say, in the overlapping parametrization (for an exact definition see Hannan and Deistler, 1988) such that:

(i) The mapping attaching the overlapping parameters to the transfer functions $k \in U_\alpha$ is continuous, and

(ii) k_0 is an inner point of U_α, which is open in $U_n^{(m)}$ (see Hannan and Deistler, 1988)

This is in fact the reason for using overlapping descriptions rather than echelon parameters in this case: We already know, that $\hat{k}_T \to k_0$, where $\hat{k}_T$ denotes the estimate of the transfer function obtained by one of the subspace algorithm under consideration. Thus for T large enough $\hat{k}_T \in U_\alpha$ will hold, and thus we find (from a certain T onwards) parameters for $\hat{k}_T$, i.e. we are able to transform the obtained estimates of the system matrices into estimates of the parameters in the overlapping parametrization. Since the new basis is constructed by simply choosing rows of $\bar{\mathcal{H}}$ as the rows of $\mathcal{C}$, the transformation T is formed of rows of $CA^i, i \geq 0$. Thus the transformation is differentiable in the entries of (A, C) in a neighborhood of the true system, which follows from the openess of the set U_α. Therefore $\Delta\mathcal{O}_e = (\Delta\mathcal{O})T^{-1} - \mathcal{O}T^{-1}(\Delta T)T^{-1} = [(\Delta\mathcal{O}) - \mathcal{O}_e(\Delta T)]T^{-1}$, where ΔT is defined for T large enough. Note, that $\mathcal{O}_e^\dagger \mathcal{H}(\Gamma^-)^{-1} = \mathcal{K}_e$, where $X^\dagger = (X^TX)^{-1}X^T$ denotes the Moore-Penrose pseudo-inverse of a matrix X. Thus

$$\Delta\mathcal{K}_e = (\Delta\mathcal{O}_e^\dagger)\mathcal{H}(\Gamma^-)^{-1} + \mathcal{O}_e^\dagger \Delta(\mathcal{H}(\Gamma^-)^{-1})$$

where

$$\begin{aligned}\Delta(\mathcal{H}(\Gamma^-)^{-1}) &= (\Delta\mathcal{H})(\Gamma^-)^{-1} - \mathcal{H}(\Gamma^-)^{-1}(\Delta\Gamma^-)(\Gamma^-)^{-1} \\ &= \bar{\mathcal{H}}\Delta E\mathcal{T}_-^T(\Gamma^-)^{-1} - \bar{\mathcal{H}}\mathcal{T}_-^T\mathcal{T}_-^{-T}\mathcal{T}_-^{-1}\mathcal{T}_-\Delta E\mathcal{T}_-^T(\Gamma^-)^{-1} = 0\end{aligned}$$

Also $\Delta\mathcal{O}_e^\dagger = -(\mathcal{O}_e^T\mathcal{O}_e)^{-1}(\Delta\mathcal{O}_e^T\mathcal{O}_e + \mathcal{O}_e^T\Delta\mathcal{O}_e)\mathcal{O}_e^\dagger + (\mathcal{O}_e^T\mathcal{O}_e)^{-1}\Delta\mathcal{O}_e^T$ holds. Putting these equations together we obtain

$$\Delta\mathcal{K}_e = -\mathcal{O}_e^\dagger(\Delta\mathcal{O}_e)\mathcal{K}_e \tag{5.10}$$

Thus $\Delta\mathcal{K}_e$ is essentially a function of $\Delta\mathcal{O}_e$. Let ΔS denote $-\mathcal{O}_e^\dagger(\Delta\mathcal{O}_e)$. Then we obtain:

$$\begin{aligned}\Delta C_e &= \Delta[\mathcal{H}_1\mathcal{K}_e^T\Sigma_e^{-1}] \\ &= \bar{\mathcal{H}}_1(\Delta E)\mathcal{T}_-^T\mathcal{K}_e^T\Sigma_e^{-1} + \bar{\mathcal{H}}_1\mathcal{T}_-^T(\Delta\mathcal{K}_e^T)\Sigma_e^{-1} - C_e(\Delta\Sigma_e)\Sigma_e^{-1} \\ &= \bar{\mathcal{H}}_1[\Delta E\mathcal{T}_-^T\mathcal{K}_e^T - \mathcal{T}_-^T\mathcal{K}_e^T\Sigma_e^{-1}(\Sigma_e(\Delta S)^T + \Delta S\Sigma_e + \mathcal{K}_e\mathcal{T}_-\Delta E\mathcal{T}_-^T\mathcal{K}_e^T - \Sigma_e(\Delta S)^T)]\Sigma_e^{-1} \\ &= \bar{\mathcal{H}}_1[\Delta E\mathcal{T}_-^T\mathcal{K}_e^T - \mathcal{T}_-\mathcal{K}_e^T\Sigma_e^{-1}\Delta S\Sigma_e - \mathcal{T}_-^T\mathcal{K}_e^T\Sigma_e^{-1}\mathcal{K}_e\mathcal{T}_-(\Delta E)\mathcal{T}_-^T\mathcal{K}_e^T]\Sigma_e^{-1} \\ &= -\bar{\mathcal{H}}_1\mathcal{T}_-\mathcal{K}_e^T\Sigma_e^{-1}\Delta S \\ &= -C_e\Delta S\end{aligned}$$

Next, let $\mathcal{K}_e(2)$ denote the matrix obtained from $\mathcal{K}_e$ by omitting the first block column and let $\langle a_t, b_t\rangle$ here denote $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T a_t b_t^T$ (with abuse of notation), which has been shown to exist for all relevant variables in Chapter 4. Then we obtain for the estimation of A:

$$
\begin{aligned}
\Delta A_e &= \Delta[\langle \hat{x}_{t+1,e}, \hat{x}_{t,e}\rangle \Sigma_e^{-1}] = \Delta[\mathcal{K}_e \begin{bmatrix} \mathcal{H}_1 \\ \Gamma^- \end{bmatrix} \mathcal{K}_e^T \Sigma_e^{-1}] \\
&= (\Delta\mathcal{K}_e(1))C_e + \bar{K}_e \Delta(\mathcal{H}_1 \mathcal{K}_e^T \Sigma_e^{-1}) + \\
&\quad (\Delta\mathcal{K}_e(2))\Gamma^- \mathcal{K}_e^T \Sigma_e^{-1} + \mathcal{K}_e(2)[(\Delta\Gamma^-)\mathcal{K}_e^T \Sigma_e^{-1} + \Gamma^-(\Delta\mathcal{K}_e^T)\Sigma_e^{-1} - \Gamma^- \mathcal{K}_e \Sigma_e^{-1}(\Delta\Sigma_e)\Sigma_e^{-1}] \\
&= \Delta S \bar{K}_e C_e + \bar{K}_e(\bar{\mathcal{H}}_1 \Delta E \mathcal{T}_-^T \mathcal{K}_e^T \Sigma_e^{-1} + C_e \Sigma_e \Delta S^T \Sigma_e^{-1}) \\
&\quad -\bar{K}_e C_e \Delta\Sigma_e \Sigma_e^{-1} + \Delta S(A_e - \bar{K}_e C_e) + (A_e - \bar{K}_e C_e)\mathcal{K}_e[(\Delta\Gamma^-)\mathcal{K}_e^T \Sigma_e^{-1} + \Gamma^-(\Delta\mathcal{K}_e^T)\Sigma_e^{-1} \\
&\quad -\Gamma^- \mathcal{K}_e \Sigma_e^{-1}((\Delta\mathcal{K}_e)\Gamma^- \mathcal{K}_e^T + \mathcal{K}_e(\Delta\Gamma^-)\mathcal{K}_e^T + \mathcal{K}_e\Gamma^-(\Delta\mathcal{K}_e^T))\Sigma_e^{-1}] \\
&= \bar{K}_e(\bar{\mathcal{H}}_1 \Delta E \mathcal{T}_-^T \mathcal{K}_e^T \Sigma_e^{-1} + C_e \Sigma_e \Delta S^T \Sigma_e^{-1} - C_e(\Delta S \Sigma_e + \mathcal{K}_e \mathcal{T}_- \Delta E \mathcal{T}_-^T \mathcal{K}_e^T + \Sigma_e \Delta S^T)\Sigma_e^{-1} \\
&\quad +\Delta S \bar{K}_e C_e + \Delta S(A_e - \bar{K}_e C_e) + (A_e - \bar{K}_e C_e)\Delta S \\
&= \Delta S A_e - A_e \Delta S
\end{aligned}
$$

Here we also use $\Sigma_e = \mathcal{K}_e \Gamma^- \mathcal{K}_e^T$. Since $\bar{K}_e$ denotes the first s columns of $\mathcal{K}_e$ we immediately obtain from equation (5.10) that $\Delta\bar{K}_e = \Delta S \bar{K}_e$ holds. Thus we observe, that the estimate of the transfer function $\bar{k}_0 = k_0 E_0^{-1}$ does not depend on the fourth moments of the noise $\varepsilon(t)$ if $\Delta S = 0$. Thus we are left to examine $\Delta S = -\mathcal{O}_e^\dagger \Delta\mathcal{O}_e$. Clearly $\Delta S = 0$ if $\mathcal{O}_e^\dagger \Delta\mathcal{O} = \Delta T$. From the definition of the overlapping coordinates we obtain that $\Delta T = [P, 0]\Delta\mathcal{O}$, where $P \in \mathbb{R}^{n \times ns}$ is a selector matrix, having exactly one entry 1 in each row, all other entries being zero. Thus $\Delta S = 0$ if $[\mathcal{O}_e^\dagger - [P, 0]]\Delta\mathcal{O} = 0$. Note, that $[\mathcal{O}_e^\dagger - [P, 0]]\mathcal{O} = T - T = 0$ and thus it is sufficient to show, that the columns of $\Delta\mathcal{O}$ are in the range space of $\mathcal{O}$. The i-th column of $\Delta\mathcal{O}$, denoted by $\Delta\mathcal{O}_i$, is equal to

$$
\Delta\mathcal{O}_i = (W^+)^{-1}\Delta u_i = (W^+)^{-1} \sum_{j \neq i} \frac{u_j}{\lambda_i - \lambda_j}(u_j^T W^+ \bar{\mathcal{H}} \Delta E \bar{\mathcal{H}}^T (W^+)^T u_i)
$$

Since $\bar{\mathcal{H}} = \mathcal{O}\mathcal{C}$ the result follows from $u_j^T W^+ \mathcal{O} = 0$ for all $u_j \perp span\{U_n\}$. Thus we have shown, that the squared terms (i.e. e_0) do not influence the asymptotic distribution of $\bar{k}_0$ in the case, where the weighting matrix $W^+(k_W)$ is chosen fixed. Note, however, that corresponding to the estimation of E_0 the distribution of $\varepsilon(t)$ does matter. Thus also the distribution of the estimates of the transfer function k_0 depend on the distribution of the noise $\varepsilon(t)$ and in particular on the fourth moments.

Corresponding to the `CCA` case, we will need slightly different arguments: Since we already know, that we may neglect every term in $\Delta\mathcal{O}_i$, which is in the column span of $\mathcal{O}$, some arguments are in fact simpler: The discussion will center on $\Delta X = \Delta[W^+ \mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(W^+)^T]$, where here for notational simplicity W^+ denotes either $(\Gamma_f^+)^{-1/2}$ or $(\Gamma^+)^{-1/2}$ respectively, depending on the choice of f fixed or $f \to \infty$ respectively. Also $\bar{\mathcal{H}}$ and $\mathcal{H}_f$ respectively will be denoted by $\mathcal{H}$ in the sequel. The following equation holds:

$$
\begin{aligned}
\Delta X &= (\Delta W^+)\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(W^+)^T + W^+[\Delta\mathcal{H}(\Gamma^-)^{-1}]\mathcal{H}^T(W^+)^T \\
&\quad +W^+\mathcal{H}(\Gamma^-)^{-1}(\Delta\mathcal{H})^T(W^+) + W^+\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(\Delta W^+)^T \\
&= (\Delta W^+)\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(W^+)^T + W^+\bar{\mathcal{H}}(\Delta E)\bar{\mathcal{H}}^T(W^+) + W^+\bar{\mathcal{H}}\bar{\mathcal{H}}^T(\Delta W^+)^T
\end{aligned}
$$

The contribution of the second and the third term to Δu_i lies in the row space of $\mathcal{O}$ and thus can be neglected. Thus we have to consider only the first term. In the `CCA` case the weighting also depends on the data, thus $\Delta\mathcal{O}_i = \Delta[(W^+)^{-1}u_i]$, again omitting the term involving $\Sigma^{1/2}$. Therefore $\Delta\mathcal{O}_i = -(W^+)^{-1}(\Delta W^+)(W^+)^{-1}u_i + (W^+)^{-1}\Delta u_i$, where the essential part of Δu_i according to the last equation is equal to $\sum_{j \neq i} u_j u_j^T \frac{1}{\lambda_i - \lambda_j}(\Delta W^+)(W^+)^{-1}[W^+\mathcal{H}(\Gamma^-)^{-1}\mathcal{H}^T(W^+)^T]u_i = \sum_{j \neq i} u_j u_j^T \frac{\lambda_i}{\lambda_i - \lambda_j}(W^+)^{-1}u_i$. Thus

$$
\Delta\mathcal{O}_i = (W^+)^{-1}[-(\Delta W^+)(W^+)^{-1}u_i + \sum_{j \neq i} u_j u_j^T \frac{\lambda_i}{\lambda_i - \lambda_j}(W^+)^{-1}u_i]
$$

which shows, that also this contribution is in the row span of $\mathcal{O}$, since $\lambda_i/(\lambda_i - \lambda_j) = 1, j > n$. Thus also in this case $\Delta S = 0$.

Finally consider the situation in the case, where additional inputs are used in the estimation: Note, that the asymptotic distribution of the crosscovariances such as $\frac{1}{T}\sum \varepsilon(t)u(t-j)^T$ does not depend on the distribution of the noise $\varepsilon(t)$. Therefore e.g. $\Delta\Gamma_{y,u} = 0$ and $\Delta\beta_z = 0$ follows from straightforward application of the results obtained above. It is easily seen, that also $\Delta\beta_z^c = 0$ and $\Delta\beta_z^i = 0$ holds, if the same is true for the estimates obtained in the first stage of the two stage method (For the constrained regression approach this can be seen from the expression given in the proof of Theorem 4.3.4, the arguments for the two stage procedure this follows from straightforward arguments). Writing down the equations for $\Delta X, \Delta A, \Delta B, \Delta C, \Delta D, \Delta\bar{K}$ in a similar way as for the case m=0, we easily obtain the following result:

Theorem 5.5.1 *Let $(y(t))_{t\in\mathbb{Z}}$ be generated by a system of the form (1.1), where no additional inputs are present. Then the distribution of the estimates of the transfer function $\bar{k}_0 = k_0 E_0^{-1}$, where $k_0 \in U_n^+$, at fixed frequency points does not depend on the distribution of the noise $\varepsilon(t)$ under the conditions of Theorem 4.1.1 or Theorem 4.2.1.*

Let $(y(t))_{t\in\mathbb{Z}}$ be generated by a system of the form (1.1), where the input process $(u(t))_{t\in\mathbb{Z}}$ fulfills the conditions of section 4.3, and let $\hat{W}_p^- = (\hat{\Gamma}_{z,z} - \hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})^{1/2}$. Then the distribution of the estimates of the transfer functions $(\bar{k}_0, l_0) = (k_0 E_0^{-1}, l_0)$, where $(k_0, l_0) \in U_n^+(u(t))$, at fixed frequency points under the conditions of the Theorem 4.3.1 or Theorem 4.3.2 do not depend on the distribution of the noise $\varepsilon(t)$. If the constrained regression approach is used, or if the asymptotic distribution of the estimates of the first step of the two stage method does not depend on the noise distribution, the same is true for the results of Theorem 4.3.3.

Thus the asymptotic variance of the estimates of the transfer function $\bar{k}_0$ does not depend on the distribution of the noise in all the cases treated in Chapter 4. Note, however, that the estimates of k_0 will depend on the fourth moments of the noise, since the estimates of E_0 do. Thus in this respect the subspace estimates show the same behaviour as the ML estimates (comp. Hannan and Deistler, 1988, chapter 4). This concludes this chapter.

Chapter 6

Results for other Subspace Procedures

In this chapter we will briefly be concerned with some statistical properties of the different algorithms presented in chapter 3. This chapter will apply the (somewhat technical) tools, provided in Chapter 4 to the various approaches. We will not aim at a complete discussion in this chapter, but rather restrict ourselves to the most obvious asymptotic results.

6.1 The Realization Approach

The statistical analysis of this method is straightforward using the knowledge presented in this thesis: First note, that the method is only used for the identification of data in the case, where no observed exogenous inputs are present. Thus we will throughout this section assume $(y(t))_{t\in\mathbb{Z}}$ to be a stationary process generated by a rational, causal, stable and strictly minimum-phase transfer function k_0 and a noise process $(\varepsilon_t)_{t\in\mathbb{Z}}$ fulfilling the standard assumptions. Recall, that the realization based identification algorithms can be outlined as followed:

- Obtain an estimate $\hat{\mathcal{H}}_{f,p}$ of a covariance Hankel matrix $\mathcal{H}_{f,p} = \mathcal{O}_f\bar{\mathcal{C}}_p$ using estimates of the covariance sequence, $\hat{\gamma}(j)$ say, $j = 1, \cdots, f+p-1$, where $\min(f,p) > n$, where n denotes the order of k_0.
- Decompose $\hat{\mathcal{H}}_{f,p} = \hat{\mathcal{O}}_f\hat{\bar{\mathcal{C}}}_p + \tilde{R}$ using a SVD of $\hat{W}_f^+\hat{\mathcal{H}}_{f,p}\hat{W}_p^- = \hat{U}_n\hat{\Sigma}_n\hat{V}_n^T + \hat{R}$, where $\hat{\Sigma}_n$ contains the greatest n singular values ordered in size and $\hat{R}$ contains the contribution due to the remaining ones.
- Define $\hat{C}$ as the first s rows of $\hat{\mathcal{O}}_f$, $\hat{\bar{C}}$ as the first s columns of $\hat{\bar{\mathcal{C}}}_p$ and calculate $\hat{A}$ using the shift invariance approach.
- If the transfer function corresponding to $(\hat{A}, \hat{\bar{C}}, \hat{C}, \hat{\gamma}(0)/2)$ is strictly positive real, then $\hat{K}$ and $\hat{E}$ are obtained by performing a spectral factorization of the spectrum obtained from $2\pi\hat{f}(\lambda) = \hat{\gamma}(0) + \hat{\bar{C}}(e^{-i\lambda}I - \hat{A})^{-1}\hat{C} + [\hat{\bar{C}}(e^{-i\lambda}I - \hat{A})^{-1}\hat{C}]^*$.

Here * denotes conjugate transpose. Note, that in this identification scheme there are several pitfalls. These problems have been investigated extensively by (Lindquist and Picci, 1996; Dahlen *et al.*, 1997). Maybe the most severe are the following: First the algorithm, as it stands, does not guarantee, that the estimated matrix $\hat{A}$ is asymptotically stable. However, $\hat{A}$ being asymptotically stable is a necessary condition for $\hat{f}(\lambda)$ to be the spectrum corresponding to a stationary process. In every case, where $\hat{A}$ is not stable, the algorithm fails to produce an estimate (comp. the simulation studies given in Peternell, 1995). A second reason for the algorithm to break down lies in the requirement, that $(\hat{A}, \hat{\bar{C}}, \hat{C}, \hat{\gamma}(0)/2)$ is positive real, which is not ensured by the algorithm. (Lindquist

and Picci, 1996) emphasized, that this problem is more severe, than it might appear: For given integers f, p there exists a nonzero, open subset of the set of all covariances $\gamma(0), \gamma(1), \cdots, \gamma(f+p-1)$, such that a straightforward application of the algorithm will not lead to a positive real transfer function estimate. Therefore, every realization based algorithm has a risk of breaking down This failure is not only due to finite sample problems, as the arguments given in (Lindquist and Picci, 1996) show: Given a finite number of covariances (i.e. $\gamma(0), \gamma(1), \cdots, \gamma(f+p-1)$), two different orders, δ and τ say, may be defined as follows: Let δ be the smallest integer, such that there exists matrices $(A, \bar{C}, C, \gamma(0)/2)$, such that $\gamma(i) = CA^{i-1}\bar{C}, i = 1, \cdots, f+p-1$. Further let $\tau \geq \delta$ denote the smallest integer, such that the corresponding $(A, \bar{C}, C, \gamma(0)/2)$, realizing $\gamma(0), \cdots, \gamma(f+p-1)$, in fact corresponds to a positive real system. For fixed f and p, let the set of all matrix sequences $\gamma(i), 0 \leq i \leq f+p-1$ embedded into $\mathbb{R}^{s^2(f+p)}$ be denoted with S_{f+p}. In (Lindquist and Picci, 1996) it is shown, that in S_{f+p} the subset of all sequences of matrices $\gamma(i)$, such that $\tau > \delta$, contains an open set, i.e. is not a 'thin' set. Consider a transfer function k with the property, that for given $f+p$ the corresponding finite sequence of covariances has the property $\tau > \delta$ and the same is true in the neighborhood of the true covariances. Then the uniform convergence of the sample covariances implies, that the estimated covariances will enter this neighborhood for T large enough a.s. In this case, the estimates of $(A, \bar{C}, C)$ will realize the δ dimensional system, which is not positive real, and thus the algorithm will break down a.s. in this situation. Since this problem relates to f and p finite, a possible way out would be to let f and p tend to infinity as a function of the sample size T (which will also depend on the actual weighting scheme, for central limit theorems to hold).

Another point, which also relates to the problem of positive realness, corresponds to the choice of the order: Consider the case, where the specified order is smaller than the true order. In general there is no guarantee, that the lower order approximation of the system corresponding to $(A_0, \bar{C}_0, C_0, \gamma(0)/2)$ obtained by the subspace procedure is positive real. Thus underestimation of the order may cause a break down of the algorithm.

The last point finally is related to finite sample properties. In this respect simulation examples in (Peternell, 1995) show, that the failure rate depends heavily on the choice of the weighting matrices $\hat{W}_f^+$ and $\hat{W}_p^-$.

For the analysis given below, we will thus impose a couple of assumptions: First we will assume, that f and p are chosen large enough, such that δ and τ as defined above coincide for k_0. As has been noted already, this could be achieved by estimating f and p from data as indicated in section 5.1. Secondly we assume, that the correct order is specified for the estimation. Since we are dealing with asymptotic theory, this seems to be a mild assumption. Methods for consistently estimating the order are easily derived in the spirit of SVC (see section 5.2.1) and thus the derivation of such an estimation is omitted.

In order to obtain results on the asymptotic behaviour of the estimates, we will use the evaluations of Chapter 4, section 2.2.3 and Theorem A.3.1 as well as Lemma 4.1.4. Note, that for fixed f and p, where $\min(f,p) > n$ holds, the estimates of the system matrices are a nonlinear function of the estimates of the sample covariances up to lag $f+p-1$. Thus corresponding to Theorem A.3.1 and Lemma 4.1.4, we only have to prove, that the mapping ψ attaching estimates of the system matrices $(\hat{A}, \hat{K}, \hat{C}, \hat{E})$ to the vector $\text{vec}[\hat{\gamma}(0), \cdots, \hat{\gamma}(f+p-1)]$ is well defined for $T > T_0$, continuous and differentiable at the true covariances $\gamma(0), \cdots, \gamma(f+p-1)$ and that $\psi(\gamma(0), \cdots, \gamma(f+p-1)) = (A_0, K_0, C_0, E_0)$ for some realization (A_0, K_0, C_0, E_0) of the true transfer function k_0, which of course depends on the choice of the weighting matrices.

The last point follows from equation (3.7) and Lemma 2.2.3. The fact, that the mapping is well defined a.s. for $T > T_0$ follows from the a.s. uniform convergence of the sample covariances, the assumptions on f and p and the fact, that the set of all system matrices $(A, \bar{C}, C, \gamma(0)/2)$, which correspond to positive real transfer functions, is open in $\mathbb{R}^{n \times n+n \times s+s \times n+s \times s}$, and thus the estimates of these matrices will correspond to a positive real transfer function from a certain T_0 onwards a.s. Here we also need the essential continuity of the SVD, showing the convergence of the balanced equivalence classes $[\hat{A}, \hat{\bar{C}}, \hat{C}]$ similar to the proof of Theorem 2.2.3. The same

argument together with the continuity properties of the spectral factorization shows the continuity of the mapping attaching transfer function estimates to the covariance estimates. Thus we obtain the consistency of the approach. In order to show a central limit theorem for the case f and p fixed, the last problem consists in the differentiability of the SVD, since the spectral factorization is a differentiable mapping with respect to the entries in $(A, \bar{C}, C, \gamma(0)/2)$ (see section 2.2.3 and (Chou, 1994) for a continuous time analogon, from which the discrete time result can be obtained using the bijection i, see equation (2.6)).

The differentiability of the essential part of the SVD is guaranteed, if the matrix $W_f^+ \mathcal{H}_{f,p} W_p^-$ has distinct nonzero singular values (see Lemma B.2.1). For fixed $f > n$ and $p > n$ let U_n^+ denote the set of all transfer functions $k \in U_n^{(m)}$, such that $W_f^+ \mathcal{H}_{f,p} W_p^-$ has n distinct nonzero singular values, then familiar arguments show the analyticity of this matrix as a function of the system matrices (A, K, C, E) for the common choices of the weighting matrices $\hat{W}_f^+$ and $\hat{W}_p^-$ discussed in section 3.3 and thus the set $U_n^+ \subset U_n^{(m)}$ is open and dense in $U_n^{(m)}$. This finally shows the asymptotic normality of the estimates, using Lemma 4.1.4. Thus we may state the following Theorem:

Theorem 6.1.1 *Let $(y(t))_{t \in \mathbb{Z}}$ be generated by a system of the form (1.1), where the true transfer function $k_0 \in U_n^{(m)}$ and where the ergodic noise $\varepsilon(t)$ fulfills the standard assumptions. Let the identification be performed by the realization approach using $\min(f, p) > n$ fixed such that for the sequence $\gamma(0), \cdots, \gamma(f + p - 1)$ of corresponding covariances the two integers δ and τ coincide, where the weighting matrices $\hat{W}_f^+$ and $\hat{W}_p^-$ are chosen to be either fixed nonsingular and independent of the data as well as the sample size or equal to $\hat{W}_f^+ = (\hat{\Gamma}_f^+)^{-1/2}$ and $\hat{W}_p^- = (\hat{\Gamma}_p^-)^{-T/2}$ respectively. Finally let the specified order be equal to the true order. Then there exists a realization (A_0, K_0, C_0, E_0) of k_0 and transformations $S_T, \|S_T\| < C, \|S_T^{-1}\| < C$ for some constant $C < \infty$, such that*

$$\|vec[\hat{A}_T - S_T A_0 S_T^{-1}, \hat{K}_T - S_T K_0, \hat{C}_T - C_0 S_T^{-1}, \hat{E}_T - E_0]\| \to 0, \; a.s.$$

Furthermore, if $k_0 \in U_n^+$ holds, then

$$\sqrt{T} vec(\hat{A}_T - A_0, \hat{K}_T - K_0, \hat{C}_T - C_0, \hat{E}_T - E_0) \xrightarrow{d} Z$$

where Z is a multivariate normal random variable.

REMARK 6.1.1 The same remark as below (nearly) every Theorem in Chapter 4 is in order. Note, that the realization (A_0, K_0, C_0, E_0), the variance of Z and the set U_n^+ depend on the truncation indices and the weighting matrices, which are used in the algorithm. Note, that the theorem is only formulated for the case f, p fixed. However at this stage of the thesis, it is no miracle, that the proof can be extended to the case, where f or p (or both) are allowed to tend to infinity as a function of the sample size. We will refrain from presenting a proof, since this is not the major part of the thesis.

6.2 The MOESP type of algorithms

In this section we will deal with the type of algorithms, where the system matrices are estimated using the method used in the MOESP class of algorithms (see section 3.4). The discussion in this section will be limited to the estimation of (A, B, C, D) i.e. the transfer function l describing the effects of the inputs on the outputs. The estimation of the remaining matrices is not considered here, since we believe, that in situations, where also the noise dynamics are of interest, another type of algorithm (e.g. the main class of algorithms) will be used instead.

Note, that in this situation, where the emphasis is on the estimation of l, also the assumptions on the inputs will be very decisive. Throughout this section we will assume, that the truncation indices f and p are fixed, $f > n, p \geq n$. We will distinguish between two sets of assumptions.

DEFINITION (WEAK ASSUMPTIONS) The input sequence u_t is a pseudostationary sequence, which is persistently exciting of order $f+p$ (see Appendix A).

Note, that this includes the statement, that limits of the form

$$\tilde{\gamma}_{u,u}(i) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T-i} u_t u_{t+i}^T$$

exist for all i (comp. equation (3.6)). The assumptions are formulated for the case, where the inputs are pseudostationary sequences. However, it will be clear from the analysis, that all results also hold in the case, where the inputs are stationary, jointly with $(y(t))_{t\in\mathbb{Z}}$, if the corresponding persistence of excitation conditions hold.

Recall the structure of the algorithm, which can be outlined as follows:

- Use any of the estimation algorithms of section 3.2 to get an estimate $\hat{\beta}_z$ of β_z, whose column space is identical to the column space of $\mathcal{O}_f$, and an estimate $\hat{\mathcal{U}}_f$ of $\mathcal{U}_f$
- Estimate $\hat{\mathcal{O}}_f$ using the SVD of $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$.
- Obtain the estimates $\hat{A}$ and $\hat{C}$ using the shift invariance approach and obtain the estimates $\hat{B}$ and $\hat{D}$ using the structure of the matrix $\mathcal{U}_f$ and the estimate $\hat{\mathcal{U}}_f$

Again the analysis of the estimates is centered on the clarification of the following two points: Which conditions have to be imposed on the inputs and the noise sequence in order to ensure, that the estimates of the covariances converge suitably to the true values, and which conditions on the process and the inputs ensure the differentiability of the mapping attaching system matrix estimates to covariance estimates at the true covariances.

It is rather simple to give conditions on the input sequence corresponding to the first point: Corresponding to consistency, only the existence of the limits of the covariance estimates is needed, which has been assumed already for the input process, and indeed is true for the stacked process $z(t) = [y(t)^T, u_t^T]^T$ (comp. Hannan and Deistler, 1988, section 4.1.) under our assumptions. For the asymptotic distribution however, we will impose the stronger conditions:

DEFINITION (STRONG ASSUMPTIONS) u_t is a pseudostationary process, which is persistently exciting of order $f+p$. Furthermore

$$\sqrt{T}\|\tilde{\gamma}_{u,u}(i) - \frac{1}{T}\sum_{t=1}^{T-i} u_t u_{t+i}^T\| \to 0$$

holds $\forall |i| \leq f+p$.

It is possible, that these stronger conditions might not be necessary (comp. Hannan and Deistler, 1988, where only the weaker condition of existence of the limits is used), however we will use the stronger conditions, since they fit better into our framework. Also it is easily seen, that for sums of sinusoids (which besides other inputs have been the reason for weakening the assumptions on the persistence of excitation of infinite order) the strong assumptions hold. Note, however, that the strong assumptions do not hold for stationary input processes. However, these assumptions will turn out to be unnecessarily strong in the case of stationary inputs and the results will hold also for (jointly with $(y(t))_{t\in\mathbb{Z}}$) stationary inputs $(u(t))_{t\in\mathbb{Z}}$.

Thus stacking outputs and inputs as $z(t) = [y(t)^T, u(t)^T]^T$ we obtain for the weaker conditions:

$$\hat{\gamma}_{z,z}(j) \to \gamma_{z,z}(j), j = 0, \cdots, f+p-1$$

and for the stronger conditions, that

$$\sqrt{T}\text{vec}[\hat{\gamma}_{z,z}(0) - \gamma_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f+p-1) - \gamma_{z,z}(f+p-1)] \xrightarrow{d} Z$$

where Z denotes a multivariate (degenerate) normal distribution with zero mean. It will be shown, that these are exactly the conditions, which are needed in order to ensure consistency and asymptotic normality respectively of the system matrix estimates in generic situations. In the case, where the input process $(u(t))_{t\in\mathbb{Z}}$ is a stationary process, jointly with $(y(t))_{t\in\mathbb{Z}}$ the conditions on the input $(u(t))_{t\in\mathbb{Z}}$ have to be such, that the central limit theorem given above is still valid. Thus e.g. the assumptions of Theorem 4.3.1 are sufficient.

The second point can be dealt with using similar techniques as have been used in section 4.1.1. First assume, that in step 1 the approach is used, where the effect of the future of the inputs is extracted by projecting onto the orthogonal complement of the future of the inputs (for a discussion see section 3.2). Using the above mentioned assumptions, it follows, that the estimate of $\hat{\beta}_z$ converges to

$$\beta_z = (\Gamma_{y,z} - \Gamma_{y,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^{-1}$$

using the notation of section 4.3. Here the inverse exists due to the assumption of the persistence of excitation on the input sequence and the nonsingularity of E_0. Now the properties of the system matrix estimates depend on the properties of the matrix β_z. Thus consider the various covariances more closely:

$$\Gamma_{y,z} = \mathcal{O}_f\Gamma_{x,z} + \mathcal{U}_f\Gamma_{u,z}, \Gamma_{y,u} = \mathcal{O}_f\Gamma_{x,u} + \mathcal{U}_f\Gamma_{u,u}$$

Thus $\beta_z = \mathcal{O}_f(\Gamma_{x,z} - \Gamma_{x,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})^{-1}$. Since $\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z}$ is nonsingular under our assumptions, for β_z to be of full rank n it is required, that

$$\mathcal{R}_{f,p} = \mathbb{E}\begin{bmatrix} x(t) \\ U_f^+(t) \end{bmatrix}\begin{bmatrix} (Z_p^-(t))^T & (U_f^+(t))^T \end{bmatrix}$$

is of full rank $n + fs$, where $\mathbb{E}$ denotes expectation for random variables and the limits of the estimates of the (pseudo)covariances for pseudostationary series. It is straightforward to show, that the covariances $\gamma_{z,z}(j)$ for given $\tilde{\gamma}_{u,u}(i), i \in \mathbb{Z}$ are analytic functions of the entries in the system matrices (A, B, C, D, E, K). This follows from the stability of A. Thus using the same tools as in section 4.1.1 it is possible to show, that generically the matrix β_z will be of rank n, if we are able to show, that for each choice of f, p and of the input (pseudo)spectrum there exists one pair of transfer functions $(k, l) \in M_n^j$, such that the rank of the corresponding matrix β_z is equal to n. However this follows from the results of the case, where no observed inputs are present by setting $B = 0, D = 0$: In this case, $\Gamma_{y,u} = 0$ and β_z essentially contains the matrix $\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}$, using the notation of section 4.1.1. The remaining entries of β_z are zero. Then the result is obvious from the fact, that $U_n^{(m)}$ is nonempty.
Moreover, if we want to obtain the differentiability of the SVD, which is used in the algorithm, we have to further restrict the set of admissible pairs of transfer functions: For given f and p and covariance matrices $\tilde{\gamma}_{u,u}(j)$ let $U_n^+(u_t) \subset M_n^j$ denote the set of all pairs $(k, l) \in M_n^j$ such that the matrix $W_f^+\beta_z(\Gamma_{z,z} - \Gamma_{z,u}\Gamma_{u,u}^{-1}\Gamma_{u,z})\beta_z^T(W_f^+)^T$ has exactly n distinct nonzero eigenvalues of multiplicities equal to one. Again it follows, that $U_n^+(u_t)$ is open and dense in M_n^j, if we can show, that there exists one pair of transfer functions $(k, l) \in U_n^+(u_t)$, which again is easy to show by setting $B = 0, D = 0$ and using the results in section 4.1.1, since in this case as has already been stated β_z is essentially equal to $\mathcal{H}_{f,p}(\Gamma_p^-)^{-1}$ using the notation of section 4.1.3. Thus a system with distinct singular values can be constructed using induction in the system order and the continuity of the singular values of finite dimensional matrices.

In this context it is interesting to note, that it is not sufficient to impose conditions only on the input sequence, in order to ensure consistency. This is somewhat surprising, because

the persistence of excitation conditions in fact have been developed in the context of ensuring consistency for estimation algorithms (e.g. instrumental variables methods or ML methods). In the context of the subspace algorithms presented in this section, (Jansson, 1997) provided some examples, where consistency does not hold despite the fact, that the inputs are generated by an ARMA system and thus persistently exciting of infinite order.

The a.s. convergence of the sample covariance estimates implies consistency of $\hat{\beta}_z$ for β_z and the asymptotic normality of $\sqrt{T}\text{vec}[\hat{\gamma}_{z,z}(i)-\gamma_{z,z}(i)]$, $i=0,\cdots,f+p-1$ (jointly) under our assumptions implies the asymptotic normality of $\sqrt{T}\text{vec}[\hat{\beta}_z-\beta_z]$. The same result for the estimates $\hat{A}_T$ and $\hat{C}_T$ follows from the continuity and differentiability of the essential part of the singular value decomposition under the assumption of $(k_0,l_0)\in U_n^+(u_t)$ and the fact, that the row space of β_z coincides with the row space of $\mathcal{O}_f$, since $\text{vec}[\hat{A}_T,\hat{C}_T]$ is a differentiable function of $\hat{\beta}_z$, $(\hat{\Gamma}_{z,z}-\hat{\Gamma}_{z,u}\hat{\Gamma}_{u,u}^{-1}\hat{\Gamma}_{u,z})$ and $\hat{W}_f^+$ for $k\in U_n^+(u(t))$ and for all these quantities we obtained a CLT previously. Note, that the row space of $\mathcal{O}_f$ is independent of the particular realization. Choosing a particular singular value decomposition (i.e. choosing the weighting matrices) then fixes a basis in the state space and thus also a particular realization of the true pair of transfer functions. Note, that here again the same problems corresponding to the choice of the orientation of the singular vectors occur. In the same way as for the main class of subspace algorithms we assume, that the normalization of the orientation of the singular vectors is done by fixing one particular entry in each singular vector to be positive, where this entries are nonzero for the true system. The actual implimentation of the SVD might well use a different singular value decomposition. Again the results will also hold for the other normalization as long as this normalization results in an essentially continuous SVD for the true system. Thus using this assumption we obtain the consistency and the asymptotic normality of the estimates $(\hat{A}_T,\hat{C}_T)$. It remains to show the same results for the estimates of B and D, which will be done in the sequel.

Recall the algorithm for estimating $\hat{B}_T$ and $\hat{D}_T$: The basic argument was the fact, that $\Gamma_{y,u}=\mathcal{O}_f\Gamma_{x,u}+\mathcal{U}_f\Gamma_{u,u}$, where $\mathcal{U}_f$ is a linear function of $\text{vec}[B,D]$, i.e. $\text{vec}[\mathcal{U}_f]=\overline{L}\text{vec}[B,D]$. The idea is to use this relation to obtain $\hat{B}_T,\hat{D}_T$ as the least squares solution to $\hat{\Xi}=\hat{L}\begin{bmatrix}B\\D\end{bmatrix}$, where

$$
\begin{aligned}
\hat{\Gamma}_{y,u}\hat{\Gamma}_{u,u}^{-1} &= [\hat{G}_1,\hat{G}_2,\cdots,\hat{G}_f]\\
\hat{\Xi} &= \begin{bmatrix}\hat{\mathcal{O}}_f^\perp\hat{G}_1\\ \hat{\mathcal{O}}_f^\perp\hat{G}_2\\ \vdots\\ \hat{\mathcal{O}}_f^\perp\hat{G}_f\end{bmatrix}\\
\hat{L} &= \begin{bmatrix}\hat{\mathcal{O}}_f^\perp\begin{bmatrix}0^{s\times n}\\ \hat{\mathcal{O}}_{f-1}\end{bmatrix} & \hat{\mathcal{O}}_f^\perp\begin{bmatrix}I\\ 0^{(f-1)s\times m}\end{bmatrix}\\ \hat{\mathcal{O}}_f^\perp\begin{bmatrix}0^{2s\times n}\\ \hat{\mathcal{O}}_{f-2}\end{bmatrix} & \hat{\mathcal{O}}_f^\perp\begin{bmatrix}0^{s\times m}\\ I\\ 0^{(f-2)s\times m}\end{bmatrix}\\ \vdots & \vdots\\ 0 & \hat{\mathcal{O}}_f^\perp\begin{bmatrix}0^{(f-1)s\times m}\\ I\end{bmatrix}\end{bmatrix}
\end{aligned}
$$

where the superscripts of the zero matrices indicate the respective dimensions. Here $\hat{\mathcal{O}}_f^\perp\in\mathbb{R}^{(fs-n)\times fs}$ fulfills the property, that $\hat{\mathcal{O}}_f^\perp\hat{\mathcal{O}}_f=0$ and is of full rank. We will choose $\hat{\mathcal{O}}_f^\perp=\hat{U}_2^T\hat{W}_f^+$, where $\hat{U}=[\hat{U}_n,\hat{U}_2]$ denotes the orthonormal matrix in the SVD $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-=\hat{U}\hat{\Sigma}\hat{V}^T$. Thus we obtain $(\hat{\mathcal{O}}_f^\perp)^T\hat{\mathcal{O}}_f^\perp=(\hat{W}_f^+)^T\hat{U}_2\hat{U}_2^T\hat{W}_f^+=(\hat{W}_f^+)^T(I-\hat{U}_n\hat{U}_n^T)\hat{W}_f^+$ and thus we immediately obtain consistency and a central limit theorem for this matrix from the respective results on the matrix $\hat{W}_f^+\hat{\beta}_z\hat{W}_p^-$ using the continuity and the differentiability of the singular value decomposition. (Note,

that in this way we also avoid problems with the nonuniqueness of the orthogonal complement of $\mathcal{O}_f$). Corresponding to the matrices $\hat{\mathcal{O}}_{f-i}$, which occur in the definition of $\hat{L}$ we note, that we may either take submatrices of $\hat{\mathcal{O}}_f$ or matrices built of the estimates $\hat{A}_T$ and $\hat{C}_T$. Due to the consistency results (and the asymptotic normality) obtained so far, the consistency of these matrices to $\mathcal{O}_{f-i}$ follows in both situations (also the results for the asymptotic distribution holds in both cases, only the asymptotic variance may change, however we will not try to compare the asymptotic variances since this is not a central part of the thesis). Using additionally the consistency of the sample covariances it is straightforward to prove, that $\hat{L}^T\hat{\Xi} \to L^T\Xi$ holds, where Ξ and L denote the matrices obtained by replacing in the definition of $\hat{\Xi}$ and $\hat{L}$ estimates with true quantities. Also $\hat{L}^T\hat{L} \to L^TL$ holds, as can be seen completely analogously. A central limit theorem for these two matrices can be derived from the respective results for $\sqrt{T}(\hat{\gamma}_{z,z}(j) - \gamma_{z,z}(j)), 0 \leq j \leq f+p-1, \sqrt{T}(\hat{A}_T - A_0), \sqrt{T}(\hat{C}_T - C_0), \sqrt{T}(\hat{\mathcal{O}}_f - \mathcal{O}_f)$ and $\sqrt{T}[(\hat{\mathcal{O}}_f^\perp)^T\hat{\mathcal{O}}_f^\perp - (\mathcal{O}_f^\perp)^T\mathcal{O}_f^\perp]$. Note, that these results do not hold for $\hat{\Xi}$ and $\hat{L}$ due to the nonuniqueness of $\mathcal{O}_f^\perp$, the matrix describing the orthogonal complement of $\mathcal{O}_f$. Thus the consistency of the estimates $\hat{B}_T$ and $\hat{D}_T$ (the asymptotic normality respectively) is proved, if we are able to show, that L is of full column rank and thus L^TL nonsingular: In order to prove this, assume, that L would be rank defficient. Then there existed a vector $z = [x^T, y^T]^T$ such that $Lz = 0^{fs\times 1}$. Considering the last block in the second block column of L, this is equivalent to $\mathcal{O}_f^\perp \begin{bmatrix} 0 \\ I \end{bmatrix} y = 0$, i.e. $[0, y^T]$ is in the column space of $\mathcal{O}_f$, which is to say there exists a vector g_f such that $\mathcal{O}_f g_f = [0, y^T]^T$. Considering $f > n$ and the full rank property of $\mathcal{O}_n$ we obtain $g_f = 0$ and thus $y = 0$. From the equation corresponding to the second last block row, we obtain from similar arguments that $[0, x^TC^T] = \mathcal{O}_f g_{f-1}$ holds, resulting in $Cx = 0$. Inserting this result into the third last block row we obtain $CAx = 0$. Iterating this argument we obtain $\mathcal{O}_{f-1}x = 0$ and thus $x = 0$. Therefore we have shown, that L is of full rank and we may state the following result:

Theorem 6.2.1 *Let $(y(t))_{t\in\mathbb{Z}}$ be generated by a system of the form (1.1), where the ergodic noise $\varepsilon(t)$ fulfills the standard assumptions. If the input u_t is a pseudostationary sequence fulfilling the weak assumptions of this section, if $(k_0, l_0) \in U_n^+(u(t))$ and if the* MOESP *algorithm is used to estimate the system matrices for fixed $f > n, p \geq n$, then there exists a (possibly nonminimal) realization (A_0, B_0, C_0, D_0) of the true transfer function l_0, such that*

$$\|vec[\hat{A}_T - A_0, \hat{B}_T - B_0, \hat{C}_T - C_0, \hat{D}_T - D_0]\| \to 0$$

Moreover, if the pseudostationary sequence u_t fulfills the strong assumptions, then

$$\sqrt{T}vec[\hat{A}_T - A_0, \hat{B}_T - B_0, \hat{C}_T - C_0, \hat{D}_T - D_0] \xrightarrow{d} Z$$

where Z denotes a multivariate normal distribution with zero mean.

If $(u(t))_{t\in\mathbb{Z}}$ is a stationary process, jointly with $(y(t))_{t\in\mathbb{Z}}$ fulfilling the assumptions of Theorem 4.3.1, then the results stated above still hold.

REMARK 6.2.1 It is easy to guess, that here the same remark as below most of the Theorems in this thesis applies: The set $U_n^+(u_t)$, the particular realization (A_0, B_0, C_0, D_0) and the asymptotic variance V of Z all depend on the weighting scheme used, and the choice of the truncation indices f and p. It is also interesting to note, that the asymptotic variance also depends on the interpretation of the inputs under our assumptions: When u_t denotes a pseudostationary sequence, the entries corresponding to sample covariances of u_t in $\sqrt{T}vec[\hat{\gamma}_{z,z}(0) - \gamma_{z,z}(0), \cdots, \hat{\gamma}_{z,z}(f+p-1) - \gamma_{z,z}(f+p-1)]$ will be zero, whereas there will be nonzero entries in the case of a stationary input process $(u(t))_{t\in\mathbb{Z}}$. However the particular limiting realization (A_0, B_0, C_0, D_0) does not depend on the interpretation of the inputs, but only on the limiting covariance sequence $\gamma_{z,z}(j)$.

REMARK 6.2.2 The realization of the transfer function l_0, which is obtained as the limit of the estimates, may be nonminimal. This is due to the fact, that the matrix A is used to model both, the noise dynamics as well as the dynamics of the transfer function l. Thus the dimension of the

state will be equal to the joint order n, whereas the order of l_0 may well be smaller. This topic has been addressed already in REMARK 2.2.1. It is not at all clear, how this affects the asymptotic performance of the estimates.

A fundamental difference of this Theorem (as well as Theorem 6.1.1) compared to the results presented in Chapter 4 lies in the fact, that using `MOESP` (or the realization approach) it is possible to obtain consistent transfer function estimates with finite truncation index p. The price to be paid for this in the case of the `MOESP` algorithm is the risk of using too big system order in the estimation of l (In the case of the realization algorithms it is the risk of obtaining no estimates at all due to the positive real condition).

A second issue in this respect is of course the analysis of the asymptotic variance and thus the question of relative asymptotic efficiency, which to the best of the authors knowledge is unsolved for any of the subspace algorithms presented in this thesis.

Note, that we do not get any hint on how to choose the truncation indices out of the theorem. Also there seems to be not much evidence on how to choose these two indices with one exception: It is common in the literature to choose $f = p$. This seems to be supported by some computational advantages (compare also section 3.4) and notational convenience. (Jansson, 1997) provides some calculations, that indicate, that it might be optimal corresponding to the asymptotic variance of the estimates of the poles of the system (i.e. the eigenvalues of A) to choose p close to the system order. However the calculations presented therein give no clear picture on how to choose the truncation indices in general. Also it is a major question, which still is to a large extent unanswered, how the restrictions $f > n, p \geq n$ can be implemented prior to estimation, since the order n is estimated only after the indices f and p have been chosen, and the possible orders of the estimated system are limited to $\max\{fs, p(s+m)\}$ and thus there is an interdependence between choosing the truncation indices and the system order, unless the system order is known a priori.

Finally we want to note, that following the arguments given above it is straightforward to examine the asymptotic properties of mixed procedures. For example it would be possible to examine procedures, where in a first stage `MOESP` is used to obtain estimates of (A, B, C, D), which are then used to estimate also the noise dynamics by inserting these estimates into the two stage method (see section 4.3.3). Also it is possible to examine an algorithm, where in the first step a constrained regression is performed and in the third step of the general algorithm the method used in `MOESP` is used to estimate the system matrices. However we refrain from producing results of this kind, since the relevance of providing such results, without being able to give a performance analysis must be doubted.

Chapter 7

Conclusions and Future Research

The first part of this thesis dealt with maximum likelihood estimation. Consistency and asymptotic normality of the transfer function estimates has been stated, without reference to a particular parametrization. This general result can be used to obtain consistency and asymptotic normality for any parametrization by clarifying the topological properties of the parametrization. In particular the continuity of the parametrization is of importance. Since for MIMO systems there exists no globally continuous parametrization, the set M_n has to be split up into several pieces, which can be parametrized continuously. Thus for assessing the usefulness of a parametrization it is of importance to have some knowledge about the partitioning of M_n, the closures of the pieces and the closures of the parameter sets. An analysis of the topological properties of balanced parametrizations proposed by Ober has been given in the first part of this thesis, where continuity of the parametrization is proved and also the structure of the pieces and the relations in between these pieces is analyzed in some depth. However the closures of the pieces have not been described in full generality and there remain some unsolved questions. The results are also transferred to a canonical form for minimum-phase balanced systems, also developed by Ober. The properties of these parametrizations are compared to echelon parametrizations. The structure theory presented for balanced parametrization shows clear hierarchical structures, which could be used in identification algorithms to reduce the computational cost of optimization of the likelihood, by incorporating the information on the appropriateness of the integer parameters contained in the free parameters. However, a statistically sound way of using this information is still an open question.

The main part of this thesis contains a rather technical discussion of the asymptotic properties of the estimates of linear systems obtained by a class of subspace algorithms. These results have been extended to a more general class of algorithms and the tools presented in passing can be used to investigate a vast variety of different identification procedures, which all hinge on the same basic fact. This thesis shows consistency and asymptotic normality for the system matrix estimates of a large class of subspace algorithms, however the discussion has one fundamental drawback: The nature of the results is more of an affirmative kind, proving that the asymptotic normality indeed holds in a number of different frameworks, whereas it does not lead to direct results about the comparison of different procedures. The proof of the central limit theorem is constructive, and thus expressions for the asymptotic variance can be given. These expressions are too complicated, in order to be examined analytically, but approximations of the asymptotic variances can be calculated, where the approximation error can be made arbitrarily small and is just limited by the system ressources on the computer used for the calculations. Thus we are able to compare the asymptotic performance for different procedures, given a particular system, and thus we may perform the comparison for various situations, which has been done in Chapter 5 for some examples.

One class of algorithms has been investigated in more depth. For this class of algorithms we developed a completely automatic identification procedure (which does not contain data pretreat-

ment such as removing the mean or deterministic trends, dealing with missing observations and outliers etc., which certainly will be needed in order to perform the identification fully automatic, however this is not the subject of this thesis). Furthermore the asymptotic properties of this procedure were analyzed in some depth. For the main class of algorithms also a quite complete discussion on the effects of the user choices has been given in Chapter 5: There it is shown, how the choice of the weighting matrix W_f^+ influences the asymptotic bias, if the specified model order is smaller than the true order. It is also discussed, how the weighting matrix may be chosen in order to shape this asymptotic bias, giving the user the choice to directly influence the quality of the model in certain frequency ranges. This subject has been shown to be closely related to the theory of frequency weighted balancing. Concerning the choice of the truncation indices calculations in Chapter 5 show, that the performance of the algorithm depends on the choice of f, and the results indicate, that also the choice of the weighting matrix $\hat{W}_f^+$ influences the choice of f in the sense, that for different $\hat{W}_f^+$ different choices of f seem to be better. Finally in Chapter 5 also possibilities to estimate the system order are discussed: Two new procedures are analyzed and shown to provide (a.s.) consistent estimates of the order. These two new procedures are compared to existing algorithms for order estimation by means of simulation studies. The analysis of the main class of algorithm thus gives a quite complete picture, with one exception: It has been the declared aim of this thesis project to clarify the question of asymptotic efficiency of the algorithm termed `CCA` throughout the thesis. First steps indicate, that `CCA` for $f, p \to \infty$ as a function of the sample size T is (at least) close to the optimum and thus asymptotically efficient, however we did not succeed in proving this conjecture, nor did anybody else to the best of the authors knowledge.

This leads directly to the question of future research in the area of subspace identification: The interested reader might have noticed, that whilst being confronted with a big number of results on the asymptotic properties of the estimates, there are very few remarks in the thesis, giving recommendations on the many choices in using subspace algorithms. This has been done for a simple reason: We are not able to give these recommendations, since we simply do not know enough. The results lead to expressions for the asymptotic variance of the estimates, which can be interpreted as a measure of the asymptotic performance in some sense, however, as has been stated several times, the expressions are too complicated to be investigated analytically in order to compare different algorithms. It is desirable to obtain more knowledge on the comparison between different algorithms, either on an analytical basis, or by trying many systems and comparing the various algorithms for these systems (using the theory presented in this thesis) and then building an intuition, in which situations to use what kind of algorithms. This in particular relates to the topic of asymptotic efficiency mentioned above.

Finally the last topic we want to emphasize is the following: Throughout the thesis we dealt with asymptotic properties, however asymptotic results are just one side of the coin, so to say. On the other hand also comparisons for finite sample, real world data have to be provided, in order to compare the relevance for application. The key for this comparison probably is the acceptance of the methods for application by scientists, which can only be increased by providing software packages, which offer the user all the flexibility of subspace methods and good guidance for the choices, the user has to take. A basis for this guidance might be some of the results included in this thesis.

Appendix A

Stochastic Setting

A.1 Stationary processes

In this section we will review some basic facts about stationary processes. We will always denote random variables by writing the time index in brackets, whereas for realizations of random variables the time index will be denoted using subscripts. Since the basic facts of stationarity are quite standard and can be found in many books about stationary processes and system identification (see e.g. Hannan and Deistler, 1988; Ljung, 1987; Söderström and Stoica, 1989), we will only cite the results and refer mostly to (Hannan and Deistler, 1988) and the references therein for proofs.

DEFINITION (STRICT STATIONARITY) A random process $(y(t))_{t\in\mathbb{Z}}$ is called *strictly stationary*, if all distributions of finite dimensional vectors $[y(t_0), y(t_0+i_1), \cdots, y(t_0+i_n)], i_j \in \mathbb{Z}, 1 \leq j \leq n$ are independent of the time t_0.

DEFINITION (WIDE SENSE STATIONARITY) A random process $(y(t))_{t\in\mathbb{Z}}$ is called *wide sense stationary*, if $\mathbb{E}y(t) = \mu_0$ and $Cov(y(t), y(s)) = \mathbb{E}(y(t)-\mu_0)(y(s)-\mu_0)^T$ depends only on $t-s$, where $\mathbb{E}$ denotes the expectation of a random variable.

Throughout the thesis *stationarity* will be used in the strict sense. Furthermore, we will assume, that the second moments exist. As a justification for doing so, note that in the Gaussian case both concepts are equivalent. We will assume the expectation μ_0 to be zero. For a stationary process $(y(t))_{t\in\mathbb{Z}}$ of dimension s and of finite variance we may define the covariance function $\gamma_{y,y} : \mathbb{Z} \to \mathbb{R}^{s\times s}, \gamma_{y,y}(j) = Cov(y(j), y(0))$. In the case that $\gamma_{y,y}(j) = \delta_{j,0}\Omega$, where $\delta_{0,0} = 1, \delta_{j,0} = 0, j \neq 0$ denotes the Kronecker delta function, $(y(t))_{t\in\mathbb{Z}}$ is called *white noise*.
Every stationary process of finite variance can be seen as a sequence of vectors, whose elements are random variables in some probability space $L_2(\Omega, \mathcal{F}, \mathbb{P})$ of random variables with finite variances (for more exact definitions and properties of the σ-algebra $\mathcal{F}$ see Hannan and Deistler, 1988). The subspace $span\{y(t)_i | t \in \mathbb{Z}, 1 \leq i \leq s\}$, where the additional subscript $_i$ denotes the i-th component of the vector $y(t)$, can be closed in this L_2. Let $H_y = \overline{span\{y(t)_i | t \in \mathbb{Z}, 1 \leq i \leq s\}}$ denote the closure of $span\{y(t)_i | t \in \mathbb{Z}, 1 \leq i \leq s\}$ in L^2. H_y with the scalar product in L_2 i.e. $\langle a, b\rangle = \mathbb{E}ab$, is a Hilbert space (in the following the subscript will always denote the stationary process under consideration, i.e. H_y will denote the Hilbert space associated with $y(t)$, H_u the Hilbert space corresponding to $u(t)$ and so on). We will also need the subspaces, which contain the past of the process and the future respectively: $H_y^-(t) = \overline{span\{y(r)_i | r \leq t, 1 \leq i \leq s\}}, H_y^+(t) = \overline{span\{y(r)_i | r > t, 1 \leq i \leq s\}}$. $H_y^-(t)$ will be called the *past* of $y(t)$ and $H_y^+(t)$ will be called the *future* of $y(t)$. Denote the best

linear mean square prediction (componentwise) of $y(t+\tau)$ in $H_y^-(t)$ by $y(t+\tau|t)^l$ i.e. $\mathbb{E}(y(t+\tau)_i - y(t+\tau|t)_i^l)^2 = \min_{x \in H_y^-(t)} \mathbb{E}(y(t+\tau)_i - x)^2, i = 1, \cdots, s$. Alternatively this could be stated as $\mathbb{E}(y(t+\tau) - y(t+\tau|t)^l)^T (y(t+\tau) - y(t+\tau|t)^l) = \min_{x, x_i \in H_y^-(t)} \mathbb{E}(y(t+\tau)_i - x)^T (y(t+\tau)_i - x)$. Then a stationary process with zero mean is called *linearily regular*, if $\lim_{\tau \to \infty} y(t|t-\tau) = 0$, where the limit is in mean square. This is equivalent to $\bigcap_{t \in \mathbb{Z}} H_y^-(t) = \{0\}$. A process is called *linearily singular*, if $y(t+\tau|t) = y(t+\tau)$ a.s. for one t, τ and thus for all $t, \tau > 0$ (or $\bigcap_{t \in \mathbb{Z}} H_y^-(t) = H_y^-(t)$). Thus linear singularity means perfect predictability using only linear predictors: From the knowlegde of the whole past, the future of the process is determined a.s. and can be predicted exactly. It can be shown, that every stationary process can be decomposed additively into a regular and a singular component:

Theorem A.1.1 (Wold Decomposition) *Every stationary process $y(t)$ can be represented in a unique way as*

$$y(t) = r(t) + s(t)$$

where $r(t)$ and $s(t)$ are obtainable as linear transformations of $y(t)$, $H_r^-(t) \subset H_y^-(t), H_s^-(t) \subset H_y^-(t)$, $\mathbb{E}r(a)s(b)^T = 0, a, b \in \mathbb{Z}$ and where $r(t)$ is linearily regular and $s(t)$ is linearily singular. Furthermore, every linear regular process $r(t)$ can be represented as (Wold representation)

$$r(t) = \sum_{j=0}^{\infty} K(j)\varepsilon(t-j), \sum_{j=0}^{\infty} \|K(j)\|^2 < \infty$$

where $H_r^-(t) = H_\varepsilon^-(t)$ and where $\varepsilon(t)$ is white noise.

Finite dimensional systems in the form (1.1) without exogenous inputs $u(t)$, fulfilling the stability assumption $|\lambda_{max}(A)| < 1$ describe regular processes for $\varepsilon(t)$ white noise, since then $\|A^n\| \to 0, n \to \infty$ and $y(t+\tau|t)^l = CA^{\tau-1}x_t$ The coefficients $K(j), L(j)$ are equal to $K(0) = E, K(j) = CA^{j-1}K, L(0) = D, L(j) = CA^{j-1}B, j \in \mathbb{N}$ and are sometimes called *Markoff parameters*. The $L(j)$ are called *impulse response coefficients*, since they describe the response in $y(t)$ to a unit impulse in $u(t)$. Assume additionally, that E is nonsingular and that the matrices (A, K, C, E) fulfill the strict miniphase assumption $|\lambda_{max}(A - KE^{-1}C)| < 1$. In this case we are able to define the *spectrum* f of the stationary process by:

$$f(e^{i\lambda}) = \frac{1}{2\pi} k(e^{i\lambda}) k(e^{i\lambda})^* \tag{A.1}$$

where * denotes conjugate transpose, and where $k = \sum_{j=0}^{\infty} K(j)z^{-j} = E + C(zI - A)^{-1}K$ has no poles and no zeros outside the open unit disc of the complex plane (see e.g. Hannan and Deistler, 1988). Under our assumptions this decomposition of the spectrum is unique. The spectrum is connected to the covariance function as follows:

$$\gamma_y(j) = \int_{-\pi}^{\pi} f(\lambda) e^{ij\lambda} d\lambda, 2\pi f(\lambda) = \sum_{j=-\infty}^{\infty} \gamma_y(j) e^{-ij\lambda} \tag{A.2}$$

i.e. the covariance function is the Fourier transform of the spectrum. Again we will indicate the process, to which the spectrum corresponds, with a subscript e.g. f_y denotes the spectrum corresponding to $(y(t))_{t \in \mathbb{Z}}$. Therefore there exists a bijection between the spectrum and the covariance function and both contain the same information about the stationary process.
Now consider the case, where the input process $(u(t))_{t \in \mathbb{Z}}$ is a stationary process, jointly with $y(t)$ i.e. the stacked process $([y(t)^T, u(t)^T]^T)_{t \in \mathbb{Z}}$ is a stationary process with spectral density $f(\lambda) = \begin{bmatrix} f_y(\lambda) & f_{yu}(\lambda) \\ f_{uy}(\lambda) & f_u(\lambda) \end{bmatrix}$. Assume, that both inputs and outputs are regular stationary processes, and assume further, that $f_u(\lambda) > 0, \forall \lambda \in [-\pi, \pi]$. Then the transfer function $l(z) = D + \sum_{j=1}^{\infty} L(j)z^{-j} = D + C(zI - A)^{-1}B$ can be obtained from the joint spectrum f as $l(e^{i\lambda}) = f_{yu}(\lambda) f_u(\lambda)^{-1}$. In this case it is possible to identify l from the joint spectrum of $y(t)$

and $u(t)$, and k from the decomposition $f_y(\lambda) = \frac{1}{2\pi}k(e^{i\lambda})k(e^{i\lambda})^* + l(e^{i\lambda})f_u(\lambda)l(e^{i\lambda})^*$. Note that this corresponds to an orthogonal (in the sense of the Hilbert space $H_y + H_u$) decomposition of $y(t)$ into a component due to $u(t)$ and a component, which is uncorrelated with $u(t)$. Thus we may always assume, that the inputs $u(t)$ and the noise $\varepsilon(t)$ are uncorrelated over all times i.e. $\mathbb{E}u(t)\varepsilon(s)^T = 0, \forall t, s$. Due to the connection of the spectrum with the covariance function (A.2), there exists in this case a mapping attaching (k, l) to the covariance function of the joint process $z(t) = [y(t)^T, u(t)^T]^T$. Of course this mapping depends on the input spectrum as well.

DEFINITION (PERSISTENCE OF EXCITATION) An input process $(u(t))_{t\in\mathbb{Z}}$ is called *persistently exciting* for the system (k, l), if the transfer functions (k, l) can be obtained from the joint spectrum f.

It has been noted by several authors (e.g. Wahlberg and Jansson, 1994; Verhaegen and DeWilde, 1993), that the requirement of $u(t)$ to be a regular process, jointly stationary with $y(t)$ is quite restrictive, since normally in technical applications the inputs are regarded as part of the experimental design, provided by the user to identify the system. As such, they will typically be equal to a weighted sum of sinusoids or similar signals, e.g. binary signals i.e. sequences of 0's and 1's. There are two main reasons for objecting to the assumption that $(u(t))_{t\in\mathbb{Z}}$ is a (jointly with $(y(t))_{t\in\mathbb{Z}}$) regular stationary process:

First, sums of sinusoids are singular processes, and the spectrum will no longer exist, since the spectral distribution function in this case contains of a finite number of jumps at the included frequencies and will be constant everywhere else. This implies, that the transfer function $l(z)$ cannot be obtained from the decomposition of the spectrum of $y(t)$ into the parts stemming from the observed inputs $u(t)$ and the parts due to the unobserved noise $\varepsilon(t)$ in general. However for some algorithms the conditions for persistence of excitation can be weakened. In the discussion of the algorithms given in Chapter 3 we will discuss the particular conditions on the inputs, which the various algorithms require. However for the main results we will use the framework of (Peternell, 1995), where we assume that $(u(t))_{t\in\mathbb{Z}}$ is a regular stationary process, jointly with $(y(t))_{t\in\mathbb{Z}}$, whose spectrum can be using by two positive constants $\rho_1 > 0$ and $\rho_2 < \infty$ i.e. $\rho_1 I < f_u(\lambda) < \rho_2 I, \lambda \in [-\pi, \pi]$.
The weaker requirements on the input process $(u(t))_{t\in\mathbb{Z}}$ will make use of a concept extensively discussed in (Ljung, 1987). A stationary process $u(t) \in \mathbb{R}^m$ is called *persistently exciting of order* α, if the following Toeplitz matrix is of full rank $m\alpha$:

$$\Gamma_{u,\alpha} = \begin{bmatrix} \gamma_{u,u}(0) & \gamma_{u,u}(1) & \cdots & \gamma_{u,u}(\alpha-1) \\ \gamma_{u,u}(-1) & \gamma_{u,u}(0) & \cdots & \\ \vdots & \ddots & \ddots & \\ \gamma_{u,u}(-\alpha+1) & \cdots & & \gamma_{u,u}(0) \end{bmatrix} \tag{A.3}$$

Note, that we do not impose the condition, that the mean of $u(t)$ is zero. Thus $\gamma_{u,u}(j) = \mathbb{E}u(t)u(t-j)^T$ may not be equal to the covariances of the process $(u(t))_{t\in\mathbb{Z}}$. Thus the sum of α sinusoids is persistently exciting of order α (at most). Note, that thus we have two different concepts of persistently excitation. Whenever we use the latter meaning, we will indicate this with adding the term 'of order α'.

Secondly one could argue, that the input is provided by the user, and is therefore a deterministic sequence rather than a stochastic process. In this case, it is still possible to derive statistical results (see e.g. Ljung, 1985; Ljung and Yuan, 1985, for results in the ARMAX context), also compare the discussion in (Hannan and Deistler, 1988), assuming that the observations u_t behave like the realization of a stationary process. Since the realization $u_t, 1 \le t \le T$ is all we know about the inputs u_t, we may assume, that the following limits exist and are independent of t:

$$\begin{array}{rcl} \lim_{T\to\infty} \frac{1}{T}\sum_{j=1}^{T} u_{t+j} & = & \mu \\ \lim_{T\to\infty} \frac{1}{T}\sum_{j=1}^{T} u_{t+j}u_{t+j-s}^{T} & = & \tilde{\gamma}_{u,u}(s), |s| \leq \alpha \text{ for some } \alpha < \infty \end{array} \tag{A.4}$$

However in the thesis we will also need stronger assumptions on the observations. This will be indicated, whenever needed. Note that the above equations have the form of strong laws of large numbers for stochastic sequences, also note the connections to ergodicity. Input sequences, that fulfill the above conditions (A.4) will be called *pseudostationary* sequences. Again we are able to define the concepts of persistence of excitation of order α for these sequences based on the function $\tilde{\gamma}_{u,u}(s)$: A pseudostationary sequence $u_t \in \mathbb{R}^m$ is called *persistently exciting of order* α, if the matrix $\tilde{\Gamma}_{u,\alpha}$ is of full rank $m\alpha$. Note, that this introduces a new notion of persistence of excitation, however as this concept differs only in the interpretation of the inputs from the concept of persistence of excitation of order α, we will not distinguish these two notions in the notation. The restriction, that the inputs u_t are assumed to be uncorrelated of the noise $\varepsilon(t)$ of course translates in this framework to $\lim \frac{1}{T}\sum_{j=1}^{T} u_t\varepsilon(t+s)^T = 0$ in a suitable sense (which will be made more exact, when needed). In this thesis we will always distinguish between pseudostationary sequences u_t and stationary processes $u(t)$, whenever the probabilistic structure of $u(t)$ makes a difference.

A.2 Hilbert Spaces and Projections of Random Variables

Let $(y(t))_{t\in\mathbb{Z}}$ be a s-dimensional stationary random process. Consider the Hilbert space H_y equipped with the scalar product as defined in the last section:

$$\langle x, y\rangle = \mathbb{E}xy \tag{A.5}$$

The predictions $y(t+\tau|t)^l$ have a nice interpretation in this framework: Recall that $y(t+\tau|t) = \arg\min_{x, x_i\in H_y^-(t)} \mathbb{E}(y(t+\tau) - x)^T(y(t+\tau) - x)$. Thus the components of $y(t+\tau|t)$ are just the orthogonal projections of $y(t+\tau)$ onto $H_y^-(t)$, since they are the points in $H_y^-(t)$, which have the smallest distance to $y(t+\tau)_i$. Up to now, we have only treated linear prediction. It would be of no use, to treat only linear predictors, if we can achieve better predictions, using a nonlinear framework. Thus -in order to justify our framework- we have to impose restrictions on the probabilistic structure, to ensure, that the linear prediction is always the best prediction (see Hannan and Deistler, 1988, Chapter 1, section 1.4 for a discussion): Denote with $\mathcal{F}$ the σ-algebra induced by the components of $(y(t))_{t\in\mathbb{Z}}$ and with $\mathcal{F}_t$ the σ-algebra spanned by the components of $y(s), s \leq t$. Then of course $\mathcal{F}_t \subset \mathcal{F}_{t+1} \subset \cdots \subset \mathcal{F}$. $\mathcal{F}_{-\infty} = \bigcap_{j=-\infty}^{\infty} \mathcal{F}_j$. There exists a subset $\mathcal{I} \subset \mathcal{F}$ of invariant sets i.e. the sets $S \subset \Omega$ that differ only in a set of zero measure (with respect to $\mathbb{P}$) from $\mathcal{T}S$, where $\mathcal{T}T$ denotes the time translation operator (for an exact definition and discussion see Hannan and Deistler, 1988). Note, that the time translation operator exists due to the assumed strict stationarity. In this framework, the best predictor (not necessarily linear) in the least squares sense is given by the conditional expectation:

$$y^b(t+\tau|t) = \mathbb{E}\{y(t+\tau)|\mathcal{F}_t\}$$

Now a sufficient condition for $y(t+\tau|t)^b = y(t+\tau|t)^l$ is the following (see Hannan and Deistler, 1988):

$$\mathbb{E}\{\nu(t)|\mathcal{F}_{t-1}\} = 0 \tag{A.6}$$

where $\nu(t) = y(t) - y(t|t-1)^l$. Thus in this case we may drop the superscript in the prediction and denote the best predictor of $y(t+\tau)$ given $\mathcal{F}_t$ with $y(t+\tau|t)$. If condition (A.6) holds, the best prediction is thus described by the orthogonal projections in the Hilbert space H_y. Note, that the $\nu(t)$ in (A.6) are equal to the innovations in the Wold representation.
In this framework, also the role of the state space $X_{t+1} = span\{x_{t,i}|1 \leq i \leq n\}$ can be given a nice interpretation (see Lindquist and Picci, 1985))

$$X_t = H_y^+(t)|_{H_y^-(t)} \tag{A.7}$$

where $A|_B$ denotes the orthogonal projection in H_y, i.e. the state space at any time t is the projection of the future onto the past. Note that this statement is of course independent of the choice of the basis in the state space. Therefore the best mean square prediction of any variable in the future is a function of the state. This property is the crucial fact used in 'subspace algorithms'. However in finite sample we are not able to calculate the projections of random variables. This leads to another concept dealing with finite realizations of stationary processes, which is described in the next section.

A.3 'Finite Sample Hilbert Spaces'

In this section we will approximate the scalar product in H_y by a scalar product in a finite dimensional space, which can be calculated from observations of the process $(y(t))_{t\in\mathbb{Z}}$ in a time interval $1 \leq t \leq T$ i.e. from one realization $y_t, 1 \leq t \leq T$. The key property of the stationary process in this context will be ergodicity:

> DEFINITION (ERGODICITY) A (strictly) stationary process is called *ergodic*, if the set $\mathcal{I}$ of invariant subsets of Ω as defined in the last section contains only sets of measure one or zero.

This property ensures, that the sample moments converge almost sure to the true moments. This is also the reason for assuming strict stationarity instead of only wide sense stationarity. In particular the sample mean converges to zero ($\frac{1}{T}\sum_{t=1}^{T} y_t \to 0$) almost everywhere and the sample covariances converge to the true covariances i.e.

$$\lim_{N\to\infty} \frac{1}{T} \sum_{t=j+1}^{T} y_t y_{t-j}^T = \gamma_{y,y}(j) a.e. \tag{A.8}$$

Thus define the sample analogon $H_N^T \subset \mathbb{R}^T$ to the Hilbert space H_y as follows: Let $Y_{t,T} = [y_t, y_{t+1}, y_{t+2}, \cdots, y_{t+T}] \in \mathbb{R}^{s\times T}$, where $y_{T+j} = 0, j \in \mathbb{N}$. Now $H_N^T = span\{Y_{t,T,i}, 1 \leq t \leq N, 1 \leq i \leq s\}$, where the additional subscript i denotes the i-th row of $Y_{t,T}$. Together with the scalar product in $\mathbb{R}^T$ given by $\langle a, b\rangle = \frac{1}{T} ab^T$, the space H_N^T becomes a finite dimensional (closed) subspace of the Hilbert space $\mathbb{R}^T$. This setup is extensively used in (Van Overschee and DeMoor, 1994). Note that with the notation

$$Y_{1,T}^i = \begin{bmatrix} Y_{1,T} \\ Y_{2,T} \\ \vdots \\ Y_{i,T} \end{bmatrix}$$

the rowspace of the matrix $Y_{1,T}^N$ is equal to H_N^T. Also compare the discussion above with the work of (Lindquist and Picci, 1996).
By letting $T, N \to \infty, N = N(T)$, the scalar product in H_N^T approximates the scalar product in H_y, due to stationarity and ergodicity. However, the quality of this convergence is influenced by the rate, at which $N = N(T)$ converges to ∞. An important property of this approximation, which will be used extensively in this thesis, can be stated as follows (for a proof see e.g. Hannan & Deistler):

Theorem A.3.1 (Uniform Convergence of Sample Covariances) *Let the process $(y(t))_{t\in\mathbb{Z}}$ be stationary generated by a system of the form (1.1) with ergodic noise $\varepsilon(t)$ fulfilling the standard assumptions (1.3). Then:*

$$max_{0\leq j\leq H_T}\|\hat{\gamma}(j)-\gamma(j)\| = O((\log\log(T)/T)^{1/2}), H_T \leq (\log T)^a, a<\infty$$

where $\hat{\gamma}(j) = \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-j}^T = (\langle Y_{j+1,T,a}, Y_{1,T,b}\rangle)_{a,b=1,\cdots,s}$. *and* $f(T) = O(h(T))$ *means that* $f(T)/h(T) \leq C$, *from a certain* T_0 *onwards.*

Thus under standard assumptions on $\varepsilon(t)$, we can approximate the first $(\log T)^a$ covariances with uniform accuracy $O(\log\log(T)/T)^{1/2}$. This gives an indication on the choice of N. Note, that this result als holds true for ergodic processes, where the mean is not equal to zero, in the sense, that the result holds for the covariances and the second moments, i.e. the noncentral covariances.

A.4 Projections and Regressions in H_N^T

Projections in finite sample Hilbert spaces are just the usual orthogonal projections in finite dimensional vector spaces. Thus the projection of a row vector $y \in \mathbb{R}^T$ onto the row space of some matrix $X \in \mathbb{R}^{m\times T}$ is equal to $y|_X = (\frac{1}{T}yX^T)(\frac{1}{T}XX^T)^{-1}X = yX^T(XX^T)^{-1}X$. Therefore the projection coincides with the approximation of y resulting from the regression using the rows of X as regressors. In this sense e.g. the projection $Y_{p+1,T}|_{Y_{1,T}^p}$ is equivalent to the regression $y_t = \hat{\beta}_p Y_{t-p,0}^p + \hat{\nu}_t$. Holding p fixed for the moment and using Theorem A.3.1 we observe, that $\hat{\beta}_p$ converges a.s. to $[\gamma_{y,y}(p), \gamma_{y,y}(p-1), \cdots, \gamma_{y,y}(1)](\Gamma_p^+)^{-1}$, where Γ_p^+ denotes the variance of the random variable generating the vector $Y_{1,0}^p$, and thus the finite sample projections converge a.s. to the projections of the random variables. This ability to approximate projections of random variables using finite data lies at the core of subspace algorithms as presented in section 3. This discussion also clarifies the connections between the regression interpretation and the 'row space projections' interpretation of subspace algorithms.

Appendix B

Operator Theory

In this section we will review some of the basic facts about linear operators, which are needed in the main part of this thesis. The presentation is very short, for details the interested reader may refer to (Chatelin, 1983), which provides the basis for the cited results. Consequently, we will use the notation as introduced therein.

B.1 Definition and Existence of Singular Value Decomposition

We are concerned with mappings $\mathcal{T} : H_1 \to H_2$, where H_1 and H_2 are Hilbert spaces. In our context $H_1 = H_2 = H$ will be equal to ℓ^2 (if the contrary is not explicitely stated), the Hilbert space of all sequences of real numbers $(x_i)_{i\in\mathbb{N}}$, which are square summable i.e. $\sum_{i=1}^{\infty} x_i^2 < \infty$. The inner product of the Hilbert space induces a norm $\|x\| = (\sum_{i=1}^{\infty} x_i^2)^{1/2}$. A linear mapping $\mathcal{T}$ is called *bounded*, if the image of the unit sphere B in H_1, $B = \{x \in H_1 : \|x\| = 1\}$, is bounded. For a bounded linear operator $\mathcal{T}$, the operator norm may be defined as $\|\mathcal{T}\|_{op} = \max_{x\in B} \|\mathcal{T}x\|$. It can be shown, that for a bounded linear operator the limit $r_\sigma(\mathcal{T}) = \lim_{k\to\infty} \|\mathcal{T}^k\|_{op}^{1/k}$ exists. $r_\sigma(\mathcal{T})$ is called the *spectral radius*. For a bounded operator $\mathcal{T}$, the adjoint operator $\mathcal{T}^*$ can be defined as follows: Denote the inner product in H with $\langle x, y\rangle$, then for every $y \in H$ there exists a $z \in H$, such that $\langle y, \mathcal{T}x\rangle = \langle z, x\rangle, \forall x \in H$ (Riesz representation theorem). Since it can be shown, that this z is unique for given y, the mapping $\mathcal{T}^* : y \mapsto z$ is well defined. $\mathcal{T}^*$ is a linear mapping with the properties $\|\mathcal{T}^*\|_{op} = \|\mathcal{T}\|_{op}$ and $\|\mathcal{T}^*\mathcal{T}\|_{op} = \|\mathcal{T}\|_{op}^2$. A linear operator is called *self adjoint*, if $\mathcal{T}^* = \mathcal{T}$. A self adjoint operator is called *positive (semi)definite*, if $\langle x, \mathcal{T}x\rangle > 0$ ($\langle x, \mathcal{T}x\rangle \geq 0$) holds. For self adjoint operators the equality $\|\mathcal{T}\|_{op} = r_\sigma$ holds.

A bounded linear operator is called *compact*, if the image of any bounded set is relatively compact i.e. every sequence $y_n \in \mathcal{T}S$, where $\|x\| < C, \forall x \in S$, contains a convergent subsequence. A bounded operator P is called *projection*, if $P^2 = P$. Every projection defines two spaces: Let $ker(\mathcal{T})$ denote the *kernel* of the operator $\mathcal{T}$, $ker(\mathcal{T}) = \{x \in H : \mathcal{T}x = 0\}$. Then $M = ker(I - P)$, where I denotes the identity, and $N = ker(P)$ span the Hilbert space H i.e. $M + N = H$, where the addition of two subspace means 'linear span'. Note that M is the image of P, since for each $x \in M : (I - P)x = 0$. Conversely every decomposition of H into two closed subspaces M and $N, M \cap N = \{0\}, M + N = H$ defines a projection. Let M be a closed subspace of H. Then there exists a subspace $M^\perp$, such that $M + M^\perp = H$ and $\langle x, y\rangle = 0, \forall x \in M, \forall y \in M^\perp$. $M^\perp$ is called *orthogonal complement of* M. A projection induced by a closed subspace and its orthogonal complement is called *orthogonal projection*. Every orthogonal projection P is self adjoint and positive semidefinite with norm equal to unity. Using orthogonal projections, it is possible to define the *gap* between two closed subspaces M and N:

$$\Theta(M,N) = \max(\sup_{x\in M,\|x\|=1} \|(I-Q)x\|, \sup_{x\in N,\|x\|=1} \|(I-P)x\|) \tag{B.1}$$

where Q denotes the orthogonal projection onto N and P is the orthogonal projection onto M. It can be shown, that $\Theta(M,N) = \|P-Q\|_{op}$. For finite dimensional subspaces M and N it follows from $\|P-Q\|_{op} < 1$ that $dim(M) = dim(N)$, where $dim(.)$ denotes the dimension of a subspace.

An equivalent alternative way to define the gap metric is as follows: Let S denote the unit sphere of H: For $x, y \in S$ define $d_a(x,y) = \sqrt{1-\langle x,y\rangle}$. Then the gap metric is the Hausdorff distance between $M_1 \cap S$ and $M_2 \cap S$. Here the Hausdorff distance is defined as follows: Let S_1 and S_2 be two compact subsets of a metric space with metric d. Then the Hausdorff distance of S_1 and S_2 is defined by:

$$d_H(S_1,S_2) = \max(\max_{x\in S_1} d(x,S_2), \max_{x\in S_2} d(S_1,x))$$

In the subspace procedures a singular value decomposition (SVD) plays a major role (see section 3). In the following we will give an exact definition of singular values and singular vectors and investigate the properties of the SVD. In order to be able to investigate the properties of the SVD, we will investigate the properties of the eigenvalue decomposition of a bounded operator, since this is an equivalent problem in a certain sense, which will be clear in the following. This requires some concepts and definitions from spectral theory:

For a bounded operator $\mathcal{T}$ consider the mappings $\mathcal{T}-zI$, where $z \in \mathbb{C}$. The *resolvent set* $\rho(\mathcal{T})$ is defined as $\rho(\mathcal{T}) = \{z \in \mathbb{C}, (\mathcal{T}-zI)^{-1}$ is a bounded operator $\}$. The complement $\mathbb{C} - \rho(\mathcal{T}) = \sigma(\mathcal{T})$ is called the *spectrum* of $\mathcal{T}$. As shown in (Chatelin, 1983) the resolvent set is an open subset of $\mathbb{C}$ and $\sigma(\mathcal{T})$ is a compact subset. Boundedness of $\sigma(\mathcal{T})$ be seen easily from the fact, that $\lim_{k\to\infty} \|\mathcal{T}^k\|_{op}^{1/k} = r_\sigma(\mathcal{T}) < \infty$ in this case. Then the *Neumann series* $\sum_{j=0}^{\infty} z^{-k-1}\mathcal{T}^k = (\mathcal{T} - zI)^{-1}$ converges for every $|z| > r_\sigma$, where convergence is to be understood in the operator norm.

As an example consider the case of finite dimensional Hilbert space H: In this case every bounded linear operator can be represented by a matrix. Consequently, the spectrum is a finite set and consists only of the pointspectrum $P\sigma(\mathcal{T})$, which consists of the eigenvalues: a complex number z is called *eigenvalue of* $\mathcal{T}$, if $(\mathcal{T}-zI)x = 0$ has a nontrivial solution x i.e. $ker(\mathcal{T}-zI) \neq \{0\}$. Conversely every eigenvalue of a linear operator $\mathcal{T}$ is contained in the pointspectrum. Suppose the spectrum consists of two components σ and τ i.e. $\sigma(\mathcal{T}) = \sigma \cup \tau$, which may be seperated by a closed Jordan curve in $\rho(\mathcal{T})$ i.e. there exists an oriented Jordan curve Γ such that τ is contained in the interior of Γ and σ is contained in the complement of the interior. Then there exist subspaces M and N such that $M+N = H, \mathcal{T}M \subset M, \mathcal{T}N \subset N$ and the restricted mappings $\mathcal{T}_M : M \to H$ and $\mathcal{T}_N : N \to H$ have spectrum equal to σ and τ respectively. This means, that if the spectrum of $\mathcal{T}$ consists of seperated parts, we may decompose the mapping into mappings $\mathcal{T}_{M_i}$ acting on subsets $M_i \subset H$, which have the same properties corresponding to their spectrum as $\mathcal{T}$. This property will be used extensively in this thesis. The spectrum of the adjoint operator $\mathcal{T}^*$ is just the complex conjugate of the spectrum of $\mathcal{T}$: $\sigma(\mathcal{T}^*) = \overline{\sigma}(\mathcal{T})$. The spectrum of a self adjoint operator is real i.e. for $\mathcal{T} = \mathcal{T}^*$ the statement $\sigma(\mathcal{T}) \subset \mathbb{R}$ holds.

For compact operators the situation simplifies substantially (see Chatelin, 1983): The spectrum of a compact operator is a countable set with no accumulation point other than zero. Each nonzero $z \in \sigma(\mathcal{T})$ is an isolated eigenvalue with finite algebraic multiplicity, where the algebraic multiplicity of an eigenvalue is defined as the dimension of the kernel $ker(\mathcal{T}-zI)$. If we additionally consider a compact self adjoint operator $\mathcal{T}$, the operator may be decomposed as $\mathcal{T} = \sum_{j=1}^{\infty} \lambda_j P_j$, where $\lambda \in \sigma(\mathcal{T}) \subset \mathbb{R}$ and P_j are orthogonal projections with the following properties: $P_iP_j = \delta_{ij}P_i$, where δ_{ij} denotes the Kronecker delta function and $dim(P_iH) < \infty$, where P_iH denotes the image of H under P_i i.e. $P_iH = \{x \in H : \exists y \in H : x = Py\}$. Thus for a compact self adjoint operator the Hilbert space H may be decomposed into countably many subspaces, such that the subspaces are pairwise orthogonal, the elements in the subspaces are eigenvectors and the multiplicities of the eigenvalues are finite, with the only exception of 0. This decomposition will be called *eigenvalue-decomposition.*

In an analogous way, we can define singular values and singular vectors: Let $\mathcal{T}$ be a compact operator. Then the adjoint operator $\mathcal{T}^*$ is compact, too, and so are the self adjoint operators $\mathcal{T}^*\mathcal{T}$ and $\mathcal{T}\mathcal{T}^*$. Thus there exist two eigenvalue-decompositions $\mathcal{T}^*\mathcal{T} = \sum_{j=1}^{\infty} \lambda_j P_j, \mathcal{T}\mathcal{T}^* = \sum_{j=1}^{\infty} \lambda'_j P'_j$. It is possible to show, that $\sigma(\mathcal{T}\mathcal{T}^*) = \sigma(\mathcal{T}^*\mathcal{T})$ and thus w.r.o.g. $\lambda_i = \lambda'_i, \forall i$, where additionally the multiplicities $n_i < \infty$ of the eigenvalues λ_i are the same for $\mathcal{T}\mathcal{T}^*$ and $\mathcal{T}^*\mathcal{T}$. Now denote $\mathcal{V}_i = P_i H$ and choose an orthonormal basis $v_i^j, j = 1, \cdots, n_i$ in each $\mathcal{V}_i$. This gives a decomposition of H into $ker(\mathcal{T}) + \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} \{v_i^j\}$. Define $u_i^j = \frac{1}{\sqrt{\lambda_i}} \mathcal{T} v_i^j, j = 1, \cdots, n_i$, which defines a basis for the orthogonal complement of the kernel of $\mathcal{T}\mathcal{T}^*$. It is easily verified, that the u_i^j are orthonormal. Thus $\mathcal{T}$ can be written as:

$$\mathcal{T}x = \sum_{i=1}^{\infty} \sum_{j=1}^{n_i} \sqrt{\lambda_i} u_i^j \langle v_i^j, x \rangle \tag{B.2}$$

Thus for every compact operator $\mathcal{T}$ there exist orthonormal bases u_i^j and v_i^j of the orthogonal complement of the kernels $ker(\mathcal{T}\mathcal{T}^*)$ and $ker(\mathcal{T}^*\mathcal{T})$ respectively, such that $\mathcal{T} v_i^j = \sqrt{\lambda_i} u_i^j$. The real numbers $\sigma_i = \sqrt{\lambda_i}$ are called *singular values*, v_i^j (u_i^j respectively) are called *right (left) singular vectors*. From the discussion above it is clear, that the singular values and singular vectors depend differentiable on the eigenvalues and eigenvectors. For Hilbert spaces with a denumerable basis (such as e.g. ℓ^2) we may represent any bounded operator $\mathcal{T}$ by an (infinite dimensional) matrix. Since this is true for all operators occuring in the thesis, we will throughout identify the notions 'matrix' and 'operator'.
Note that the discussion above of course includes finite dimensional Hilbert spaces such as $\mathbb{R}^n$ equipped with the usual scalar product.

B.2 Continuity and Differentiability Properties

In this section we are concerned with the differentiability of the singular value decomposition given in equation (B.2). In the following we will cite the results as stated in (Chatelin, 1983). For a detailed discussion we refer to (Chatelin, 1983).
First we have to define different concepts of convergence of sequences of operators, in order to be able to retain the notation of (Chatelin, 1983): Let $\mathcal{T}_n$ be a sequence of compact operators and let $\mathcal{T}_0$ be a compact operator. Denote with λ an isolated eigenvalue of $\mathcal{T}_0$ of multiplicity m_0 and let Γ be a closed Jordan curve enclosing an area Δ, which contains only λ and no other point of the spectrum of $\mathcal{T}_0$. Then the following types of convergences can be defined:

- Pointwise convergence: $\mathcal{T}_n \xrightarrow{p} \mathcal{T}_0$, if $\|(\mathcal{T}_n - \mathcal{T}_0)x\| \to 0, \forall x \in H$.
- Stable Convergence: $\mathcal{T}_n - zI \xrightarrow{s} \mathcal{T}_0 - zI$, if $\mathcal{T}_n \xrightarrow{p} \mathcal{T}_0$ and $\exists C, N_0$ such that $\|\mathcal{T}_n - zI\|_{op} < C, \forall n > N_0$.
- Strongly stable convergence: $\mathcal{T}_n \xrightarrow{ss} \mathcal{T}_0$ on Δ, if $\mathcal{T}_n - zI \xrightarrow{s} \mathcal{T}_0 - zI, \forall z \in \Delta - \{\lambda\}$ and $dim(P_n H) = m_0$ for n large enough, where P_n denotes the orthogonal projection associated with the span of all eigenvectors associated with eigenvalues of $\mathcal{T}_n$ contained in Δ.

Now we are ready to state the main results in Chatelin:

Theorem B.2.1 ((Chatelin, 1983),Theorem 5.5) *If for all z in $\Delta - \{\lambda\}, \mathcal{T}_n - zI \xrightarrow{s} \mathcal{T}_0 - zI$, then $lim_{n\to\infty}[\sigma(\mathcal{T}_n) \cap \Delta] = \{\lambda\}$, where the limit is to be understood in the Hausdorff metric.*

However stability is not enough to ensure the convergence of the eigenvalues with preservation of multiplicities. Let M_n denote the subspace spanned by all vectors corresponding to eigenvalues of $\mathcal{T}_n$ contained in Δ and analogously let M_0 be the subspace corresponding to the eigenvalue λ of $\mathcal{T}_0$. Using this notation we get:

Theorem B.2.2 ((Chatelin, 1983),Theorem 5.6, Theorem 5.8) *If $\mathcal{T}_n - zI \overset{ss}{\to} \mathcal{T}_0 - zI$ for $z \in \Delta$, then $\sigma(\mathcal{T}_n) \cap \Delta$ consists, for n large enough, of exactly m_0 eigenvalues, counting their multiplicities and $\Theta(M_n, M_0) \to 0$.*

In the thesis typically we will consider only sequences of operators, which converge in the operator norm i.e. $||\mathcal{T}_n - \mathcal{T}_0||_{op}$. Using the results in (Chatelin, 1983, Chapter 5) we conclude, that $||\mathcal{T}_n - \mathcal{T}_0||_{op} \to 0 \Rightarrow \mathcal{T}_n \overset{ss}{\to} \mathcal{T}_0$ for $z \in \rho(\mathcal{T}) \cup Q\sigma(\mathcal{T}_0)$, where $Q\sigma(\mathcal{T}_0)$ denotes the set of all isolated eigenvalues of finite multiplicity of $\mathcal{T}_0$. Thus convergence in operator norm for $\mathcal{T}_n\mathcal{T}_n^*$ implies the convergence of the eigenvalues of $\mathcal{T}_n\mathcal{T}_n^*$ and thus the convergence of the singular values of $\mathcal{T}_n$, and the convergence in the gap metric of the spaces spanned by the eigenvectors associated with the eigenvalue λ.

However we are able to show more than just continuity in the case, where all the eigenvalues are distinct i.e. the multiplicity is equal to one. In this case it can be shown, that for $\mathcal{T}(t) = \mathcal{T}_0 + t\Delta\mathcal{T}$ the functions $\lambda(t)$ and $\phi(t)$ are analytic in t for t contained in δ_Γ, where $\delta_\Gamma = \{t \in \mathbb{C} : |t| < 1/r_\Gamma\}$ with $r_\Gamma = \max_{z\in\Gamma} r_\sigma[\Delta\mathcal{T}(\mathcal{T}_0 - zI)^{-1}]$ and Γ denotes a closed Jordan curve, whose interior contains only the eigenvalue λ of $\mathcal{T}_0$ but no other point of the spectrum $\sigma(\mathcal{T}_0)$. Here $\phi(t)$ denotes the eigenvector of $\mathcal{T}(t)$ corresponding to the eigenvalue $\lambda(t)$ of $\mathcal{T}(t)$ normalized such that $P_0\phi(t) = \phi_0$ with P_0 denoting the projection onto the eigenvector ϕ_0 of the eigenvalue λ_0 of $\mathcal{T}_0$ i.e. the eigenvector is not of length one. It is possible to show, that these functions are analytic in t with expansion

$$\begin{aligned}\lambda(t) &= \lambda_0 + t\langle\Delta\mathcal{T}\phi_0, \phi_0\rangle + o(t)\\ \phi(t) &= \phi_0 + t[S\Delta\mathcal{T}\phi_0 + \langle-\Delta\mathcal{T}\phi_0, \phi_0\rangle\phi_0] + o(t)\end{aligned}$$

where $S = \lim_{z\to\lambda_0}(\mathcal{T}_0 - zI)^{-1}(I - P_0)$. For the corresponding matrix representations the formulas can be stated as follows:

$$\begin{aligned}\lambda(t) &= \lambda_0 + t\phi_0^T\Delta\mathcal{T}\phi_0 + o(t)\\ \phi(t) &= \phi_0 + \sum_{\lambda_j\neq\lambda_0} \frac{\phi_j^T\Delta\mathcal{T}\phi_0}{\lambda_0 - \lambda_j}\phi_j + o(t)\end{aligned}$$

where the sum is over all eigenvalues λ_j of $\mathcal{T}_0$ with associated eigenvectors ϕ_j, which are not equal to λ_0. The proof of this statement essentially only uses the Neumann series expansion of the resolvent operator $R(t,z) = (\mathcal{T}(t) - zI)^{-1} = (\mathcal{T}_0 + t\Delta\mathcal{T} - zI)^{-1} = R(0,z)(I + t\Delta\mathcal{T}R(0,z))^{-1}$ on the closed Jordan curve Γ. Here t is restricted, so that $r_\sigma(t\Delta\mathcal{T}R(0,z)) < 1$. Now the same result holds, if we replace the restriction of t with the restriction $r_\sigma(\Delta\mathcal{T}R(0,z)) < 1, \forall z \in \Gamma$. For the compact linear operator $\Delta\mathcal{T}$. In this case, the Neumann series expansion also holds true, implying, that the analyticity property also holds for general approximations with compact matrices, not only with the directional approximation $t\Delta\mathcal{T}$. Summing up we obtain the following result:

Lemma B.2.1 (Chatelin) *Let $\mathcal{T}_n$ be a sequence of compact operators converging in operator norm to the compact operator $\mathcal{T}_0$. Then the following statements hold:*

i) For all nonzero eigen- (singular-) values of $\mathcal{T}_0$, the eigen-(singular-) values of $\mathcal{T}_n$ converge to the eigen- (singular-) values of $\mathcal{T}_0$, the eigen- (singular-) spaces of $\mathcal{T}_n$ converge in the gap metric to the eigen- (singular-) spaces of $\mathcal{T}_0$.

ii) For an eigenvalue λ_i of the compact operator $\mathcal{T}_0$ of multiplicity one, the eigenvalue λ_i' of the compact operator $\mathcal{T}'$ and the corresponding eigenvectors ϕ_i and ϕ_i' respectively fulfill the following relation (for small $||\mathcal{T}_0 - \mathcal{T}'||_{op}$):

$$\lambda_i' = \lambda_i + \phi_i^T(\mathcal{T}' - \mathcal{T}_0)\phi_0 + o(||\mathcal{T}' - \mathcal{T}_0||_{op}) \tag{B.3}$$

$$\phi_i' = \phi_i + \sum_{\lambda_j\neq\lambda_i} \frac{\phi_j^T(\mathcal{T}' - \mathcal{T}_0)\phi_i}{\lambda_i - \lambda_j}\phi_j + o(||(\mathcal{T}' - \mathcal{T}_0)||_{op}) \tag{B.4}$$

where again the summation is over all eigenvalues unequal to λ_i. The normalization of ϕ_i' is taken such that $\mathcal{T}_0\phi_i' = \phi_i$.

In the thesis we will however use a different normalization of the singular vectors: These will be normalized to be of length one. Thus we need the analogous result for $\phi_i'/\|\phi_i'\|_2$. The square of the 2-norm is equal to $\|\phi_i'\|^2 = (\phi_i')^T\phi_i' = 1 + \sum_{j\neq i}(\frac{\phi_j^T(\mathcal{T}'-\mathcal{T}_0)\phi_i}{\lambda_i-\lambda_j})^2 + o(\|(\mathcal{T}'-\mathcal{T}_0)\|_{op}^2)$. This follows from the orthonormality of the vectors ϕ_i. Computing the Taylor series expansion of $\frac{1}{\sqrt{1+x^2}}$ at $x = 0$ we obtain, that the linear term is zero. Therefore, the formulas for the linearization of the eigenvectors are unchanged, if ϕ_i' denotes the vector normalized to be of length one.

Finally we will need the definition of a function of an operator, in particular we will need the square root. In general for a bounded linear operator $\mathcal{T}$, a function $f(\mathcal{T})$ can be defined using the so called *Dunford integral* (see e.g. Yosida, 1974) under the following conditions:

DEFINITION (DUNFORD INTEGRAL) Let $\mathcal{T}$ denote a bounded linear operator and let $f : \mathbb{C} \to \mathbb{C}$ be a complex mapping, which is holomorphic in some neighborhood U of the spectrum of $\mathcal{T}$. Then the following integral may be defined:

$$f(\mathcal{T}) = \frac{1}{2\pi i}\int_{\partial U} f(z)R(\mathcal{T}, z)dz$$

Here $R(\mathcal{T}, z) = (\mathcal{T} - zI)^{-1}$ for z in the resolvent set denotes the resolvent operator and ∂U denotes the boundary of the neighborhood U.

It can be shown, that this definition is independent of the choice of the neighborhood U. (Yosida, 1974) presents some properties of the Dunford integral, of which we will only need the following:

Lemma B.2.2 (Dunford) *Let $f : \mathbb{C} \to \mathbb{C}$ be a complex mapping and let $\mathcal{T}$ be a bounded operator, where the spectrum $\sigma(\mathcal{T})$ is contained in some neighborhood U, in which f is analytic. If f has a Taylor series expansion $f(z) = \sum_{j=0}^{\infty}\alpha_j z^j$ convergent in $U \supset \sigma(\mathcal{T})$, then*

$$f(\mathcal{T}) = \sum_{j=0}^{\infty}\alpha_j\mathcal{T}^j \tag{B.5}$$

where convergence is in operator norm.

This lemma can be used to define the square root of a symmetric operator $\mathcal{T}$, if the spectrum is contained in the some interval $(c_1, c_2), c_1 > 0, c_2 < \infty$: Define $c = (c_1 + c_2)/2$, then $(\sigma(\mathcal{T}) - c) = (\sigma(\mathcal{T} - cI) \subset (-c, c)$ and the power series expansion $\sum_{j=0}^{\infty}\alpha_j(z-c)^j = \sqrt{z}$ of the square root in c has a convergence radius of c. Thus we may build $(\mathcal{T})^{1/2} = cI + \sum_{j=0}^{\infty}\alpha_j(\mathcal{T} - cI)^j$, which then converges in operator norm and has the same properties as the square root i.e. $\mathcal{T}^{1/2}\mathcal{T}^{1/2} = \mathcal{T}$.

Appendix C

Notation

This section summarizes the notation used in the thesis. The organization is as follows: The symbols are grouped into main categories, which correspond either to a subject oriented partitioning or a partitioning on the place, where they occur. Symbols, which are only used locally, are not listed here.

General

s: dimension of outputs

m: dimension of inputs

n: dimension of state vector

$\mathbb{R}$: real numbers

$\mathbb{N}$: integers

$\mathbb{N}^+$: integers plus 0

$\mathbb{Z}$: positive and negative integers

$\mathbb{C}$: complex numbers

$\mathbb{P}$: probability

$\mathbb{E}$: expectation

$\mathbb{E}\{.|.\}$: conditional expectation

$\mathcal{F}_t$: sigma-algebra of past events

x_i: i-th component of a vector x

$A_{i,j}$: i,j-element of a matrix A

A_{ij}: i,j subblock of a matrix A

A_i: i-th row or i-th column of the matrix A depending on the particular situation

$x_{i:j}$: subvector of the vector x containing the entries $i, i+1, \cdots, j$

$\|.\|$ or $\|.\|_2$: 2-norm of a matrix or a vector,

$\lambda_{max}(A)$: eigenvalue of a matrix A of maximum modulus

$\|x\|_1 = \sum_{j=1}^{n} |x_i|$

$\|A\|_1 = \max_{\|x\|_1=1} \|Ax\|_1$

$\|A\|_{Fr} = (\sum_{i,j=1}^{m,n} A_{i,j}^2)^{1/2}$: Frobenius norm of a matrix $A \in \mathbb{R}^{m \times n}$

$\|f\|_\infty = \max_\lambda \|f(\lambda)\|$ ∞-norm of a function f, where the maximum is over the range of definition of f

$\rightarrow$: convergence

$\xrightarrow{d}$: convergence in distribution

$\xrightarrow{p}$: convergence in probability

a.s.: almost sure

$\overline{U}$: closure of a set U

$span\{x_1, \cdots, x_n\}$: linear span of the vectors $x_1, \cdots, x_n$

$f(T) = o(g(T)) \Leftrightarrow \lim_{T\to\infty} f(T)/g(T) = 0$, for random variables limit is a.s.

$f(T) = O(g(T)) \Leftrightarrow \limsup_{T\to\infty} |f(T)/g(T)| \leq M$ for some some constant M, for random variables limit is a.s.

$f(T) = o_P(g(T)) \Leftrightarrow \mathbb{P}\{\lim_{T\to\infty} f(T)/g(T) = 0\} = 1$,

$f(T) = O_P(g(T)) \Leftrightarrow \exists M < \infty : \mathbb{P}\{\limsup_{T\to\infty} |f(T)/g(T)| \leq M\} = 1, \forall T > T_0$ constant M,

$\rho_0 = \lambda_{max}(A - KE - 1C)$

$A^\dagger$ Moore-Penrose pseudoinverse, $A^\dagger = (A^T A)^{-1} A^T$ for nonsingular $A^T A$.

$\hat{A}$ estimate of A

System representations

(A, B, C, D, E, K): system matrices, see (1.1)

$k(z) = E + C(zI - A)^{-1}K$: transfer function

$\bar{k}(z) = k(z)E^{-1} = I + C(zI - A)^{-1}KE^{-1}$

$\bar{K} = KE^{-1}$

$l(z) = D + C(zI - A)^{-1}B$: transfer function

$f_{u,u}(\lambda)$: spectrum of $(u(t))_{t\in\mathbb{Z}}$

$f_{y,u}(\lambda) = l(e^{i\lambda})f_{u,u}(\lambda)$

$f(\lambda) = \frac{1}{2\pi}k(e^{i\lambda})k(e^{i\lambda})^* + l(e^{i\lambda})f_{u,u}(\lambda)l(e^{i\lambda})^*$: spectrum

$f_{z,z} = \begin{bmatrix} f_{y,y} & f_{y,u} \\ f_{y,u}^* & f_{u,u} \end{bmatrix}$

$\Phi(\lambda)$: $\Phi(\lambda) + \Phi(-\lambda) = 2\pi f(\lambda)$: spectral summand

$K(j) = CA^{j-1}K, j > 0, K(0) = E, K(j) = 0, j < 0$: Markoff parameters

$L(j) = CA^{j-1}B, j > 0, L(0) = D, L(j) = 0, j < 0$

$K_u(j)$: Markoff parameter for $(u(t))_{t\in\mathbb{Z}}$

$y(t), u(t),\ z(t) = \begin{bmatrix} y(t) \\ u(t) \end{bmatrix}^T, x(t), \varepsilon(t), \eta(t), \varepsilon_u(t)$: stochastic variables

$H_z = \overline{span\{z(t)_i, t \in \mathbb{Z}, i = 1, \cdots, s+m\}}$

$H_y^+(t) = \overline{span\{y(s)_i, s > t, 1 \le i \le s\}}$

$H_y^-(t) = \overline{span\{y(s)_i, s \le t, 1 \le i \le s\}}$

$H_u^-(t) = \overline{span\{u(s)_i, s \le t, 1 \le i \le m\}}$

$y_t, u_t, x_t, \varepsilon_t, \eta_t$: realizations

τ: system parameters

Operators

$\mathcal{O} = [C^T, A^TC^T, \cdots]^T$: observability matrix

$\mathcal{O}^\uparrow = \mathcal{O}A$

$\mathcal{C} = [[B, K], A[B, K], \cdots]$: controllability matrix

$\mathcal{K} = [[B - \bar{K}D, \bar{K}], (A - \bar{K}C)[B - \bar{K}D, \bar{K}], \cdots], \bar{K} = KE{-}1$

$\mathcal{K}_p(1) = [B - \bar{K}D, \bar{K}]$

$\mathcal{K}_p(2) = [(A - \bar{K}C)[B - \bar{K}D, \bar{K}], \cdots, (A - \bar{K}C)^{p-1}[B - \bar{K}D, \bar{K}], 0]$

$\mathcal{E} = [K(i-j)]_{i,j\in\mathbb{Z}}$

$\mathcal{U} = [L(i-j)]_{i,j\in\mathbb{Z}}$

$\bar{\mathcal{H}} = \mathcal{O}\mathcal{C} = [K(i+j-1)]_{i,j\in\mathbb{Z}}$

$\mathcal{T}_f$: first f block rows or columns of (infinite dimensional) matrix $\mathcal{T}$

$\mathcal{T}_{f,p}$: leading submatrix of f block rows and p block columns of (infinite dimensional) matrix $\mathcal{T}$

Sets of system representations

M_n: set of all (discrete time) stable rational transfer functions of order n

M_n^c: continuous time analogon to M_n

$U_n^{(m)}$: set of all $k \in M_n$, which are strictly minimum-phase and have constant term, which is nonsingular and lower diagonal with positive entries on the diagonal

M_n^j: set of all pairs (k, l), such that $k \in U_p^{(m)}, l \in M_r$ and the joint order is equal to n

V_α : Echelon coordinate neighborhood

$U_\cdot$: subset of M_n defined by using properties of the balanced realizations; the indices indicate the restrictions

Θ: set, that builds the basis for likelihood optimization

$\hat{\Theta}$: Θ plus non-strict minimum-phase systems in $\overline{\Theta}$

$T_\cdot$: parameter spaces; additional indices indicate the corresponding class of systems

S_n : set of all system matrices of state dimension n corresponding to $(k, l) \in \overline{M_n^j}$

D_n^+: set of all (discrete time) balanced system realizations of systems with distinct singular values

C_n^+: continuous time analogon of D_n^+

D_n^{++}: generic set in the discrete time Ober canonical form

C_n^{++}: continous time analogon of D_n^{++}

$E(k)$: class of all observationally equivalent realizations of k

$E_b(k)$: class of all observationally equivalent balanced realizations of k

$E^c(k)$: continuous time analogon to $E(k)$

$E_b^c(k)$: continuous time analogon to $E_b(k)$

Mappings

π: attaching transfer functions to system matrices

ψ: attaching parameters to transfer functions; additional indices indicate range of definition

φ: attaching system matrices to transfer functions; additional indices indicate the range of definition

ρ: attaching parameters to system matrices; additional indices indicate range of definition

i: attaching continuous time representations to discrete time representations, see (2.6)

Vectors

$Y_{1,T}^+ = [y(1)^T, \cdots, y(T)^T]^T$

$U_{1,T}^+ = [u(1)^T, \cdots, u(T)^T]^T$

$v_t(l) = y_t - \sum_{j=0}^{\infty} L(j)\, u_{t-j}$

$V_{1,T}^+(l) = [v_1(l)^T, \cdots, v_T(l)^T]^T$

$\overline{Y_{1,T}^+}(l) = Y_{1,T}^+ - V_{1,T}^+(l)$

$Y^+(t) = [y(t)^T, y(t+1)^T, \cdots]^T$

$Y_f^+(t) = [y(t)^T, \cdots, y(t+f-1)^T]^T$

$Y_{t,f}^+ = [y_t^T, \cdots, y_{t+f-1}^T]^T$

$Y^-(t) = [y(t-1)^T, y(t-2)^T, \cdots]^T$

$Y_p^-(t) = [y(t-1)^T, \cdots, y(t-p)^T]^T$

$Y_{t,p}^- = [y_{t-1}^T, \cdots, y_{t-p}^T]^T$

$U^+(t) = [u(t)^T, u(t+1)^T, \cdots]^T$

$U_f^+(t) = [u(t)^T, \cdots, u(t+f-1)^T]$

$U_{t,f}^+ = [u_t^T, \cdots, u_{t+f-1}^T]^T$

$z(t) = [y(t)^T, u(t)^T]^T$

$z_t = [y_t^T, u_t^T]^T$

$Z^-(t) = [z(t-1)^T, z(t-2)^T, \cdots]^T$

$Z_p^-(t) = [z(t-1)^T, \cdots, z(t-p)^T]^T$

$Z_{t,p}^- = [z_{t-1}^T, \cdots, z_{t-p}^T]^T$

$E^+(t) = [\varepsilon(t)^T, \varepsilon(t+1)^T, \cdots]^T$

$E_f^+(t) = [\varepsilon(t)^T, \cdots, \varepsilon(t+f-1)^T]^T$

$E_{t,f}^+ = [\varepsilon_t^T, \cdots, \varepsilon_{t+f-1}^T]^T$

$\mathrm{vec}(A)$ vectorization of the matrix A

$\mathrm{vec}(A,B,C) = [\mathrm{vec}[A]^T, \mathrm{vec}[B]^T, \mathrm{vec}[C]^T]^T$

$A \otimes B \in \mathbb{R}^{mn \times rs}$ Kronecker product of two matrices $A \in \mathbb{R}^{n \times r}, B \in \mathbb{R}^{m \times s}$

Covariance matrices

$\gamma(s) = \mathbb{E} y(t) y(t-s)^T$

$\gamma_{a,b}(s) = \mathbb{E} a(t) b(t-s)^T$, where $a, b = y, u, z$

$\tilde{\gamma}_{u,u}(s) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=s+1}^{T} u_t u_{t-s}^T$ for pseudostationary u_t

$\Gamma_T(k) = \mathbb{E}\overline{Y_{1,T}^+}\,\overline{(Y_{1,T}^+)}^T$

$\Gamma_f^+ = \mathbb{E}\overline{Y_f^+(t)}\,\overline{Y_f^+(t)}^T$

$\Gamma^+ = \mathbb{E} Y^+(t) Y^+(t)^T$

$\Gamma_p^- = \mathbb{E} Y_p^-(t) Y_p^-(t)^T$

$\Gamma^- = \mathbb{E} Y^-(t) Y^-(t)^T$

$\mathcal{H} = \mathbb{E} Y^+(t) Y^-(t)^T$

$\bar{\mathcal{C}}$: $\mathcal{H} = \mathcal{O}\bar{\mathcal{C}}$

$\bar{C} = \bar{\mathcal{C}}_1$

$\mathcal{H}_{f,p} = \mathbb{E} Y_f^+(t) Y_p^-(t)^T$

$\Gamma_{a,b} = \mathbb{E} A(t) B(t)^T$, where $a = y \Leftrightarrow A(t) = Y_f^+(t)$, $a = u \Leftrightarrow A(t) = U_f^+(t)$, $a = z \Leftrightarrow A(t) = Z_p^-(t)$, $a = x \Leftrightarrow A(t) = x(t)$

$A_f^{+,\Pi} = A(t)|_{(U_f^+(t))^\perp}$

$\Gamma_{a,b}^{\Pi} = \mathbb{E} A_f^{+,\Pi}(t) B_f^{+,\Pi}(t)^T$ where $a = y \Leftrightarrow A(t) = Y_f^+(t)$, $a = z \Leftrightarrow A(t) = Z_p^-(t)$, $a = x \Leftrightarrow A(t) = x(t)$

$\Pi_p^- = \mathbb{E} Z_p^{-,\Pi}(t) (Z_p^{-,\Pi})^T$

Chapter 2

$u_i(v_i)$: left (right) singular vector of $\bar{\mathcal{H}}$

$\bar{U}_i = span\{u_1, \cdots, u_{n_i}\}$, where $u_1, \cdots, u_{n_i}$ correspond to the i-th singular value

$\bar{V}_i$: analogous for v_i

$\mathcal{U}_n = \{\bar{U}_i, i = 1, \cdots, k\}$

$\mathcal{V}_n = \{\bar{V}_i, i = 1, \cdots, k\}$

$L(Y_{1,T}^+, (u(t))_{t\in\mathbb{Z}}; k, l)$: likelihood function, (2.2)

I_0: Fisher information matrix, see Theorem 2.4.2

$\mathcal{L}(Y_{1,T}^+, (u(t))_{t\in\mathbb{Z}}; \tau)$: log likelihood function

C_T penalty term for information criteria, see (2.19)

$Q_T = \sqrt{\log\log T/T}$

H_T upper bound for the order estimation, usually $O(\log T^a), a < \infty$

χ_n information criterion, (2.19)

$\hat{p}_{AIC}$: order of an autoregression estimated with `AIC`

$\hat{p}_{BIC}$: order of an autoregression estimated with `BIC`

$U\Sigma V^T$ SVD of $\overline{\mathcal{H}}$

$\alpha, \beta, \gamma, \gamma^*$ multiindices, describing the partitioning of M_n

$\ell(\theta)$: prediction error criterion value evaluated at θ

Chapter 3

$\hat{W}_f^+, W_f^+, \hat{W}_p^-, W_p^-$ weighting matrices

$(\beta_z, \beta_u) = [\Gamma_{y,z}, \Gamma_{y,u}] \begin{bmatrix} \Gamma_{z,z} & \Gamma_{z,u} \\ \Gamma_{u,z} & \Gamma_{u,u} \end{bmatrix}^{-1}$

$W_f^+ \beta_z W_p^- = U\Sigma V^T$ singular value decomposition

$\hat{U}\hat{\Sigma}\hat{V}^T = \hat{W}_f^+ \hat{\mathcal{H}}_{f,p} (\hat{\Gamma}_p^-)^{-1/2}$ (no exogenous inputs) or $\hat{U}\hat{\Sigma}\hat{V}^T = \hat{W}_f^+ \hat{\beta}_z \hat{W}_p^-$ (with exogenous inputs)

$\langle a_t, b_t\rangle = \frac{1}{T}\sum_{t=1}^T a_t b_t^T$

$X|_Y$ orthogonal projection of the random vector X onto the space spanned by the components of the vector Y

$X|_{Y^\perp} = X - X|_Y$

$Y_{t,f}^{+,\Pi} = Y_{t,f}^+ - \langle Y_{t,f}^+, U_{t,f}^+\rangle\langle U_{t,f}^+, U_{t,f}^+\rangle^{-1} U_{t,f}^+$

$x_t^\Pi = x_t - \langle x_t, U_{t,f}^+\rangle\langle U_{t,f}^+, U_{t,f}^+\rangle^{-1} U_{t,f}^+$

$Z_{t,p}^{-,\Pi} = Z_{t,p}^- - \langle Z_{t,p}^-, U_{t,f}^+\rangle\langle U_{t,f}^+, U_{t,f}^+\rangle^{-1} U_{t,f}^+$

$\mathcal{R}_{f,p} = \mathbb{E}\begin{bmatrix} x(t) \\ U_f^+(t) \end{bmatrix} [Z_p^-(t)^T, U_f^+(t)^T]$

$\hat{L}_f$ see (3.5)

$\mathcal{U}_f$ see (3.4)

$\mathcal{O}_f^\perp$ a matrix fulfilling $\mathcal{O}_f^\perp \mathcal{O}_f = 0, \mathcal{O}_f^\perp \in \mathbb{R}^{(fs-n)\times fs}$

$\Xi_1 = \mathrm{vec}[\hat{\mathcal{O}}_f^\perp \langle Y_{t,f}^+, U_{t,f}^+\rangle\langle U_{t,f}^+, U_{t,f}^+\rangle^{-1}]$

Chapter 4

$W_f^+(k_W)$ weighting matrix associated with transfer function k_W, (4.1)

$W^+, W^+(k_W)$ limit of weighting matrix W_f^+ ($W_f^+(k_W)$ resp.) for $f \to \infty$

W^- limit of weighting matrix $\hat{W}_p^-$ for p and $T \to \infty$

$\delta_{i,0} = 1, i = 0, \delta_{i,0} = 0$, else

$e_j = \sqrt{T}\langle \varepsilon_t, \varepsilon_{t-j}\rangle - \delta_{j,0} I$

$V_h^\varepsilon = \lim_{T\to\infty} \mathbb{E}\mathrm{vec}[e_0, .., e_{h-1}]\mathrm{vec}[e_0, .., e_{h-1}]^T$

$X_{f,p}, X_0, Y_{f,p}, Y_0, Z_{f,p}, Z_0, m_{f,p}, m_0, \theta_p, \theta_0$ see section 4.1.3, Preliminaries

λ_i: eigenvalue of X_0

u_i: eigenvector of x_0 corresponding to eigenvalue λ_i

$\hat{\beta}_{f,p} = \hat{\mathcal{H}}_{f,p}(\hat{\Gamma}_p^-)^{-1}$

T_k, S_k: see below Lemma 4.1.4

V_f asymptotic variance of subspace estimates

$\hat{g}_{T,h} = \mathrm{vec}[\hat{\gamma}(0), \cdots, \hat{\gamma}(h-1)]$

$g_h = \mathrm{vec}[\gamma(0), \cdots, \gamma(h-1)]$

$V_h^g = \lim_{T\to\infty} \sqrt{T}\mathbb{E}(\hat{g}_{T,h} - g_h)(\hat{g}_{T,h} - g_h)^T$

ρ_p a pole of maximum modulus of $k + k_W$

$L_j^{a,h}, L_j^h, L^{a,h}, L_h$ see Lemma 4.1.4

$L_{f,p}$ matrix describing the linearization of the mapping attaching system matrix estimates to $\hat{\gamma}_{f+p}$

ψ mapping attaching system matrix estimates to $\hat{m}_p$

J_ψ $= \partial\psi/\partial m_p$

V^m asymptotic variance of $\hat{m}_p$

V_f asymptotic variance of subspace estimates

$\hat{x}_i, x_i$ defined in (4.15),(4.24),(4.25) respectively

β_z^c matrix β_z calculated using the constrained regression approach

β_z^i matrix β_z calculated using the two stage procedure

U_n^+ set of all transfer functions $k \in U_n^{(m)}$, such that $W_f^+\overline{\mathcal{H}}_f$ (or $W^+\overline{\mathcal{H}}$ resp.) has distinct singular values

$U_n^+(u(t))$ set of all transfer functions $(k,l) \in M_n^j$, such that $W_f^+\beta_z W^-$ (or $W^+\beta_z W^-$ resp.) has distinct singular values

N_{22}: a matrix containing only 0 and 1, such that $\mathrm{vec}[\hat{L}_f] = N_{22}\mathrm{vec}[\hat{L}(0),\cdots,\hat{L}(f-1)]$, see 3.5

$\hat{N}_{u,z}$ $= [N_{22}^T(\hat{\Gamma}_{u,u} - \hat{\Gamma}_{u,z}\hat{\Gamma}_{z,z}^{-1}\hat{\Gamma}_{z,u})N_{22}]^{-1}$

Chapter 5

V_{ech} asymptotic variance of subspace estimates of transfer function $\bar{k}$, which are transformed to echelon coordinates

E_f $= tr(V_{ech}I_0) - 2ns$

$\chi(n)$ see (5.1)

SVC see (5.2)

IVC see (5.3)

ΔX $= \hat{X}_{f,p} - X_0$, see Chapter 4

$\hat{\tau}_n$ parameter vector corresponding to ML estimate over $\overline{U_n^{(m)}}$

$(\hat{A}_n, \hat{K}_n, \hat{C}_n, \hat{E}_n)$ system matrix estimates using subspace algorithms and order n

$\hat{x}_t(n)$ $= \hat{\Sigma}_n\hat{V}_n^T(\hat{\Gamma}_p^-)^{-1/2}Y_{t,p}^-$

$C_{i:j}$ columns numbered $i, i+1, \cdots, j, j \le n_0$ of C_{n_0}

$\hat{C}_{i:j}$ columns numbered $i, \cdots, j$ of $\hat{C}_n$

$\hat{\Sigma}_{i:j}$: diagonal subblock with row and column indices equal to $i, i+1, \cdots, j$ of $\hat{\Sigma}_n$

$x_t(n)$ $= [x_{t,1}^T, x_{t,2}^T]^T, x_{t,1} \in \mathbb{R}^{n_0}, x_{t,2} \in \mathbb{R}^{n-n_0}$

$\hat{M}$ $= \begin{bmatrix} \langle \hat{x}_{t,1}, \hat{x}_{t,1}\rangle & \hat{x}_{t,1}, u_t\rangle \\ \langle u_t, \hat{x}_{t,1}\rangle & \langle u_t, u_t\rangle \end{bmatrix}$

$\hat{X}_{1,2}$ $= \langle \begin{pmatrix} \hat{x}_{t,1} \\ u_t \end{pmatrix}, \hat{x}_{t,2}\rangle$

$\hat{X}_{2,2}^\perp$ $= \langle \hat{x}_{t,2}, \hat{x}_{t,2}\rangle - \hat{X}_{1,2}^T\hat{M}^{-1}\hat{X}_{1,2}$

$\Omega(n)$ $= E_nE_n^T$, where $E_n = \lim_{T\to\infty}\hat{E}_n$ a.s.

MOE: order estimation procedure implemented in `MATLAB` (Ljung, 1991)

NIC: order estimation procedure suggested by (Peternell, 1995)

$(\tilde{A}_T, \tilde{K}_T, \tilde{C}_T, \tilde{E}_T)$ system matrix estimates using the order $r < n$, see Theorem 5.3.1

$(A_0^r, K_0^r, C_0^r, E_0^r)$ reduced order system, where $r < n$, see Theorem 5.3.1

$\bar{\mathcal{O}}_f$ $= [A^TC^T, \cdots, (A^T)^{f-1}C^T, 0]^T$

$\tilde{\mathcal{K}}_{p+1}$ $= [\tilde{\mathcal{K}}_{p+1}(1), \tilde{\mathcal{K}}_{p+1}(2)]$ used for estimating x_{t+1} as $\tilde{x}_{t+1} = \tilde{\mathcal{K}}_{p+1}Y_{t+1,p+1}^-$

$\mathcal{T}_-$ $= [K(j-i)]_{i,j\in\mathbb{N}}$ (block Toeplitz matrix of Markoff parameters

E $= [e_{i-j}]_{i,j\in\mathbb{N}}$

ΔA $= \partial A/\partial e_{0,r,s}$ (partial derivative of matrix A with respect to the (r,s) entry of e_0)

(A_e, K_e, C_e, E_e) realization of k_0 in overlapping coordinates

$\mathcal{O}_e$: observability matrix in echelon coordinates

$\mathcal{K}_e$ $= \mathcal{O}_e^\dagger\mathcal{H}(\Gamma^-)^{-1}$

$\mathcal{K}_e(1)$ $= \bar{K}_e$

$\mathcal{K}_e(2)$ $= (A_e - K_eE_e^{-1}C_e)\mathcal{K}$

$\mathcal{O}_i$: i-th column of $\mathcal{O}$

Bibliography

Akaike, H. (1975). Markovian representation of stochastic processes by canonical variables. *SIAM Journal of Control* **13**(1), 162–172.

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In: *System Identification: Advances and Case Studies* (R. Mehra and D. Lainiotis, Eds.). pp. 27–96. Academic Press Inc.

Akaike, H. (1997). On the role of statistical reasoning in the process of identification. In: *Proceedings of the 11th IFAC Symposium on System Identification*. IFAC. Fukuoka, Japan. pp. 1–8.

Al-Saggaf, U. and G. F. Franklin (1987). An error bound for a discrete reduced order model of a linear multivariable system. *IEEE transactions on Automatic Control* **32**(9), 815 – 819.

Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. John Wiley & Sons.

Bauer, D. and M. Deistler (1996). Balanced parametrizations: A structure theory for identification. In: *Lecture Notes in Statistics, Athens Conference on Applied Probability and Time Series, Volume II: Time Series Analysis, In Memory of E.J.Hannan* (Robinson P.M. and Rosenblatt M., Eds.). Springer, New York. pp. 27–41.

Bauer, D. and M. Deistler (1997). Balanced canonical forms for system identification. *submitted to IEEE Transactions on Automatic Control*.

Bauer, D., M. Deistler and W. Scherrer (1997b). The analysis of the asymptotic variance of subspace algorithms. *Proceedings of the* 11^{th} *IFAC Symposium on System Identification, Fukuoka, Japan* pp. 1087–1091.

Bauer, D., M. Deistler and W. Scherrer (1998). User choices in subspace algorithms. In: *To Appear in Proceedings of the CDC'98 Conference*. Tampa, Florida.

Box, G. and M. Jenkins (1970). *TIME SERIES ANALYSIS Forecasting and Control*. Holden-Day, San Francisco.

Chatelin, F. (1983). *Spectral Approximation of Linear Operators*. Academic Press.

Chou, C. T. (1994). Geometry of Linear Systems and Identification. PhD thesis. University of Cambridge.

Chou, C. T. and J. M. Maciejowski (1997). System identification using balanced parametrizations. *IEEE Transactions on Automatic Control*. forthcoming.

Chui, N.L.C. and J. M. Maciejowski (1996). Realization of stable models with subspace methods. Technical Report CUED/F-INFENG/TR222. Eng. Department, Cambridge University.

Dahlen, A., Lindquist A. and Mari J. (1997). Experimental evidence showing that stochastic subspace identification methods may fail. Technical Report TRITA/MAT-97-OS10. KTH. Stockholm.

Deistler, M., K. Peternell and W. Scherrer (1995). Consistency and relative efficiency of subspace methods. *Automatica* **31**, 1865–1875.

Desai, U. B., D. Pal and R. D. Kirkpatrick (1985). A realization approach to stochastic model reduction. *International Journal of Control* **42**(4), 821–838.

Diab, M., V. Sreeram and W.Q. Liu (1997). Frequency weighted model reduction via quasi kalman decomposition. In: *Proceedings of the ECC'97 Conference*. Brussels.

Enns, D. H. (1985). Model Reduction for Control System Design. PhD thesis. Stanford University. Stanford, CA.

Faurre, P. L. (1976). Stochastic realization algorithms. In: *System Identification: Advances and Case Studies* (Mehra R. K. and Lainiotis D. G., Eds.). Academic Press. New York. pp. 1–25.

Fuchs, J. J. (1990). Structure and order estimation of multivariable stochastic processes. *IEEE Transactions on Automatic Control* **35**, 1338–1341.

Glover, K. (1984). All optimal hankel norm approximations of linear multivariable systems and their l_∞-error bound. *International Journal of Control* **39**, 1115–1193.

Golub, G. and C. VanLoan (1989). *Matrix Computations*. second ed.. John Hopkins University Press. Maryland.

Hannan, E. J. and L. Kavalieris (1984). Multivariate linear time series models. *Adv. Appl. Prob.* **16**, 492–561.

Hannan, E. J. and M. Deistler (1988). *The Statistical Theory of Linear Systems*. John Wiley. New York.

Hannan, E.J., W. Dunsmuir and M. Deistler (1980). Estimation of vector armax models. *J. Multivariate Analysis* **10**, 275–295.

Hanzon, B. and R. Ober (1997). Overlapping block-balanced forms and parametrizations: The stable siso case. *SIAM J. Control and Optimization* **35**(1), 228–242.

Hazewinkel, M. and R. E. Kalman (1976). Invariants, canonical forms and moduli for linear, constant, finite dimensional, dynamical systems. In: *Lecture Notes in Economics and Mathematical Systems* (Marchesini G. and Mitter S. K., Eds.). pp. 48–60. Springer-Verlag. Berlin.

Jansson, M. (1997). On Subspace Methods in System Identification and Sensor Array Signal Processing. PhD thesis. KTH, Stockholm.

Jansson, M. and Bo Wahlberg (1997). Counterexample to general consistency of subspace system identification methods. In: *Proceedings of SYSID'97*. Fukuoka, Japan. pp. 1677 – 1682.

Kabamba, T. (1985). Balanced forms: Canonicity and parametrization. *IEEE Transactions on Automatic Control* **30**, 1106–1109.

Kalman, R. E., P.L. Falb and M. A. Arbib (1966). *Topics in Mathematical System Theory*. International Series in Pure and Applied Mathematics. McGraw Hill.

Kim, S. W., B.D.O. Anderson and Madievski A. G. (1995). Error bound for transfer function order reduction using frequency weighted balanced truncation. *Systems and Control Letters* **24**, 183–192.

Larimore, W. E. (1983). System identification, reduced order filters and modeling via canonical variate analysis. In: *Proc. 1983 Amer. Control Conference 2.* (H. S. Rao and P. Dorato, Eds.). Piscataway, NJ. pp. 445–451. IEEE Service Center.

Lewis, R. and G. Reinsel (1985). Prediction of multivariate time series by autoregressive model fitting. *Journal of Multivariate Analysis* **16**, 393–411.

Lindquist, A. and G. Picci (1985). Realization theory for multivariable stationary gaussian processes. *SIAM Journal on Control and optimization* **23**, 809–857.

Lindquist, A. and G. Picci (1996). Canonical correlation analysis, approximate covariance extension, and identification of stationary time series. *Automatica* **32**(5), 709–733.

Ljung, L. (1985). Asymptotic variance expressions for identified black-box transfer function models. *IEEE Transactions On Automatic Control* **30**(9), 834–844.

Ljung, L. (1987). *System Identification: Theory for the User*. Prentice Hall, Englewood Cliffs.

Ljung, L. (1991). *System Identification Toolbox, User's Guide*. The MathWorks.

Ljung, L. and T. Mc Kelvey (1996). A least squares interpretation of sub-space methods for system identification. In: *Proccedings of the CDC96 Conference*. Kobe, Japan.

Ljung, L. and Z. D. Yuan (1985). Asymptotic properties of black-box identification of transfer functions. *IEEE Transactions on Automatic Control* **30**(6), 514–530.

Lu, W. S. and E.B. Lee (1985). Model reduction via a quasi-kalman decomposition. *IEEE Transactions on Automatic Control* **30**(8), 786–790.

Maciejowski, J. M. (1985). Balanced realizations in system identification. *Proc. 7th IFAC Symposium on Identification and Parameter Estimation*.

McGinnie, B. P. (1993). A Balanced View of System Identification. PhD thesis. University of Cambridge.

McKelvey, T. (1995). On State Space Models in System Identification. PhD thesis. Dept. of Electr. Eng., Linköping.

Moonen, M., B. De Moor, L. Vandenberghe and J. Vandewalle (1989). On- and off-line identification of linear state-space models. *Int. Journal of Control* **49**(1), 219–232.

Moore, B. C. (1981). Principal component analysis in linear systems: Controllability, observability and model reduction. *IEEE Transactions on Automatic Control* **26**, 17–31.

Ober, R. (1987). Balanced realizations: Canonical forms, parametrization, model reduction. *Int. J. Control* **46**(2), 643–670.

Ober, R. (1991). Balanced parametrization of classes of linear systems. *SIAM J. Control and Optimization* **29**(6), 1251–1287.

Ober, R. (1996). Balanced canonical forms. In: *Identification, Adaptation, Learning* (S. Bittanti and G. Picci, Eds.). Chap. Balanced Canonical Forms, pp. 120–183. Springer.

Pernebo, L. and L. Silverman (1982). Model reduction via balanced state space systems. *IEEE Transactions on Automatic Control* **27**, 382–387.

Peternell, K. (1995). Identification of Linear Dynamic Systems by Subspace and Realization-Based Algorithms. PhD thesis. TU Wien.

Peternell, K., W. Scherrer and M. Deistler (1996). Statistical analysis of novel subspace identification methods. *Signal Processing* **52**, 161–177.

Rissanen, J. (1978). Modeling by shortest data description. *Automatica* **14**, 465–471.

Shibata, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statistics* **8**(1), 147–164.

Shibata, R. (1981). An optimal autoregressive spectral estimate. *The Annals of Statistics* **9**(2), 300–306.

Söderström, T. and P. Stoica (1989). *System Identification*. Prentice Hall.

Van Overschee, P. (1995). Theory, Implementation and Practice of Subspace Identification Algorithms. PhD thesis. Leuven.

Van Overschee, P. and B. DeMoor (1994). N4sid: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica* **30**, 75–93.

Verhaegen, M. (1994). Identification of the deterministic part of mimo state space models given in innovations form from input-output data. *Automatica* **30**(1), 61–74.

Verhaegen, M. and P. DeWilde (1993). Subspace model identification: Part 3. analysis of the ordinary output-error state-space model identification algorithm. *International Journal of Control* **58**(3), 555–586.

Verriest, E. and T. Kailath (1983). On generalized balanced realizations. *IEEE Transactions on Automatic Control* pp. 833–844.

Verriest, E. I. and W. S. Gray (1995). A geometric approach to the minimum sensitivity design problem. *SIAM J. Control and Optimization* **33**(3), 863–881.

Wahlberg, B. (1989). Estimation of autoregressive moving- average models via high-order autoregressive approximations. *Journal of Time Series Analysis* **10**, 283 – 299.

Wahlberg, B. and L. Ljung (1986). Desing variables for bias distribution in transfer function estimation. *IEEE Transactions on Automatic Control* **31**, 134 – 144.

Wahlberg, B. and M. Jansson (1994). 4sid linear regression. In: *Proc. of the 33rd Conference on Decision and Control*. Orlando, USA. pp. 2858–2863.

Yosida, K. (1974). *Functional Analysis*. Springer.

Lebenslauf

Name:	Dietmar W. Bauer
geboren am:	21. Juni 1972
1978 - 1982	Jonas-Volkschule St.Pölten
1982 - 1990	neusprachliches Gymnasium Josefstraße, St.Pölten
1990 - 1995	Studium der technischen Mathematik, TU Wien
1993, 1994	Tutor am Institut f. Numerische Mathematik, TU Wien
1995 -	Doktoratstudium der technischen Mathematik, TU Wien
1995 -	Vertragsassistent, Institut f. Ökonometrie, Operations Research und Systemtheorie, TU Wien
1995 -	Projektmitarbeiter im Rahmen des FWF- Projektes P11213-MAT