ON APPROXIMATE SYSTEM IDENTIFICATION WITH SYSTEM BASED ORTHONORMAL FUNCTIONS

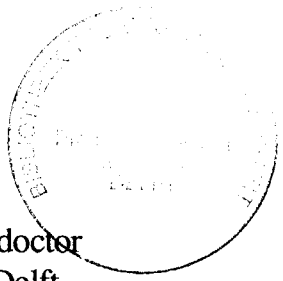
PETER HEUBERGER

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK. DEN HAAG

Heuberger. Petrus Servatius Chrétien

On approximate system identification with system based orthonormal functions /
Petrus Servatius Chrétien Heuberger. - Delft : Delft University of Technology,
Faculty of Mechanical Engineering and Maritime Technology -[1].
Proefschrift Delft. - Met samenvatting in het Nederlands.
ISBN 90-370-0045-2
SISO 517 UDC 519.87 (043.3)
Trefw.: systeemidentificatie / modelreductie / orthonormale functies

ON APPROXIMATE SYSTEM IDENTIFICATION WITH SYSTEM BASED ORTHONORMAL FUNCTIONS



PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus,
Prof. Drs. P.A. Schenck,
in het openbaar te verdedigen
ten overstaan van een commissie
aangewezen door het College van Dekanen
op donderdag 31 januari 1991 te 14.00 uur

door

Petrus Servatius Chrétien Heuberger

geboren te Maastricht

Doctorandus in de Wiskunde

Dit proefschrift is goedgekeurd door de promotor
Prof.ir. O.H. Bosgra

There is no permanent place in
the world for ugly mathematics

G.H. Hardy
A Mathematician's Apology

Dit proefschrift is opgedragen
aan mijn ouders en vrienden

VOORWOORD

De voltooiing van een proefschrift wordt vaak omschreven als "een moeilijke bevalling". Zonder de pretentie te hebben de ervaringen tijdens een zwangerschap te kunnen doorgronden, wil ik hier laten zien dat de vergelijking met het doen van een promotie–onderzoek op meerdere fronten opgaat.

In den beginne is men vaak verguld dat men in de gelegenheid wordt gesteld om tijd te besteden aan het ontwikkelen en creëren van iets nieuws, sommigen zijn er zelfs trots op. De eerste periode is er dan nog niet zoveel aan de hand, vaak zet men de lijn van leven en werken, die tot dan toe gehanteerd werd, gewoon door, ook al worden sommigen af en toe niet goed van al dat nieuwe gedoe.
Na verloop van tijd begint er dan toch iets te dagen, je begint je over bepaalde zaken dik te maken en er wordt een bepaalde druk voelbaar, nog niet echt hinderlijk en sommigen vinden dit zelfs prettig. Je hoort er dan zo'n beetje bij en de omgeving herkent dit verschijnsel en men begint je iets anders te benaderen.
De laatste periode daarentegen verandert er plotseling van alles, de druk neemt zienderogen toe en je bent je vaak op pijnlijke wijze van je situatie bewust. Ook je directe omgeving begint er onder te lijden, alles moet op den duur wijken voor deze ene onderneming, waar je ook nog eens zelf een naam voor moet verzinnen. Je wordt een lastig, soms niet bijzonder sociaal mens en het is maar een geluk dat men je situatie herkent en je vertelt dat dat er nou eenmaal bijhoort. Dat is ook het argument dat men je toewerpt als je je bezorgd begint te maken over de kosten van deze onderneming. Er komt namelijk van alles bij kijken, er moet van alles geregeld worden met officiele instanties, je moet afspraken maken met een drukker om de wereld kond te doen van het resultaat, er hoort speciale kleding bij, je moet regelmatig overleggen met een of meer doktoren, al doen sommigen het af met een kwakzalver. Zelfs bij het moment supreme wordt je niet alleen gelaten, juist daarin zijn de doktoren, voor de gelegenheid nog eens in speciale kledij, speciaal geinteresseerd.
En waar doe je het voor? Die vraag heb je je zelf de afgelopen tijd al vaak gesteld. Het enige afdoende antwoord dat je na verloop van tijd nog over hebt is dat je er nu eenmaal aan begonnen bent en dat je er niet meer onderuit kan. Het idee dat je er zo graag bij wil horen ben je dan al lang vergeten. Het enige dat je kunt hopen is dat het heerlijk zal zijn als je het achter de rug hebt. Dat hoop ik dus maar.

Iets serieuzer nu. Mijn keuze om na de middelbare school wiskunde te gaan studeren, betekende een lichte teleurstelling voor mijn moeder, die stilletjes gehoopt had dat ook haar derde zoon medicijnen zou gaan studeren. Drie dokters in haar gezin leek haar een ideale en geborgen situatie. Ik koester de hoop door middel van het voltooien van dit proefschift, waarmee ik de titel van doctor hoop binnen te halen, deze teleurstelling na al die jaren enigzins weg te kunnen nemen. Mijn vader daarentegen gaf mij van het begin af aan zijn volle steun. Zijn eerste reactie was:

" Es dee jóng dat wèlt, daan laot em toch, heer moot demèt leve ".

Deze scherts is bedoeld als een hommage aan mijn ouders, die mij de gelegenheid boden om te zijn wat ik ben en die mijn daarin meestal volledig gesteund hebben. Zij zijn dan ook de eersten die ik wil bedanken voor hun bijdrage aan de totstandkoming van dit proefschrift.

Als ik me dan richt op de plek waar dit werk werd uitgevoerd, dan zijn er vele collega's en studenten, die mij hebben gesteund en geholpen in de vorm van vriendschap, discussie, motivatie, werk, ondersteuning en stimulans. Verreweg de meeste invloed is uitgegaan van Paul van den Hof; hij is niet alleen van een uistekende komaf, hij heeft dit werk tot op de letter nagelezen, en gelezen, en gelezen... Hij wist zijn eigen ervaring op dit gebied op een stimulerende wijze over te brengen, droeg ideeën aan, was altijd bereid tot discussie, hield pep–talks, kortom hij was een zeer waardevolle collega en vriend en hij werd en wordt dan ook zeer gewaardeerd.

Een speciaal woord is er ook voor Hans van Helmont, Frank Sperling en Henk Aling, samen met mij de eerste lichting van de "Good Old Boys", een groepje aerobics–fanaten die het na de lichamelijke inspanning niet konden nalaten om onder het genot van alcohol en tabak te discussieren over "het vak". Ze hebben me veel geleerd en waren altijd bereid tot luisteren en kwamen met inspirerende ideeen aandragen. Hun kritiek en de eeuwige herhaling van "waarom?" heeft meer invloed gehad dan ze zelf zullen vermoeden.

Met plezier denk ik ook terug aan de middagen met Peter Bongers en Ruud Schrama, onze sessies met krijt en papier en pen mis ik nu al. Van 'mijn' studenten wil ik graag Huib Aernoudse bedanken, ik denk dat wij allebei erg veel van elkaar geleerd hebben. Ook Sjoerd de Wolf mag hier niet ontbreken, hij deed mij beseffen dat het allemaal toch nog niet zo erg kon zijn.

Okko Bosgra wordt hierbij bedankt voor de kans die hij mij geboden heeft om aan dit werk te beginnen en voor de begeleiding erbij. Dit werk is grotendeels gefinancierd door het Koninklijke/Shell Laboratorium Amsterdam (KSLA), en vanuit dit bedrijf ben ik door diverse mensen gesteund en begeleid. Met name Wim Schinkel en Hennie Swaanenburg worden hierbij voor hun belangrijke bijdrage bedankt.

Zeer belangrijk zijn de mensen die de randcondities in de gaten houden, zoals de financiën, de rekenfaciliteiten, de bibliotheek en de secretariaatswerkzaamheden; Sjoerd Dijkstra, Piet Ruinard, Ruud Dezentjé, Peter Valk, Cor Kremers, Els Arkesteijn, Diana van Dijk en Trinette Scholte worden hierbij uitdrukkelijk bedankt.

Keer ik nu weer terug naar het persoonlijk vlak dan is daar allereerst Petra Tim, al sinds jaar en dag mijn baken in de zondvloed, een niet te vermurwen steun en bron van feedback, feedforward en zelfvertrouwen. Haar ongezouten mening over de problemen die ik op mijn weg tegenkwam heeft ervoor gezorgd dat ik het juiste perspectief kon behouden. Verder is er een hele reeks vrienden, die de laatste tijd veel te kort zijn gekomen en mij desondanks bleven steunen.
Last but zeker niet least komt mijn dank aan Liny Berentschot, haar bijdrage is van essentieel belang geweest. Haar niet aflatend geloof, begrip en stimulans waren onmisbaar, waarvan akte.

Delft, 28 November 1990.

Financial support by Shell Research B.V., The Hague, is hereby gratefully acknowledged.

CONTENTS

SUMMARY

The use of mathematical models to describe dynamic processes is common in various sciences. This thesis deals with aspects of the modeling process, i.e the creation of such models. A distinction is made between the case that an existing model has to be approximated with a model of a lower complexity (model reduction) and the case that a model has to be created on the basis of measurement data of the dynamic process (system identification). While model reduction intrinsically has an approximative character, in the past decade the conviction has risen that the same is valid for system identification: the real process is too complicated to be described exactly by a simple, linear, time invariant, finite dimensional model. This observation reveals the close relationship between model reduction and system identification; both fields deal with the creation of (optimal) approximative models of dynamic processes.

In the literature numerous methods are proposed to deal with both problems, but these methods are hardly related or comparable. Often the character of the approximation and the quality of the resulting model are unclear. In our point of view a situation should be created in which a variety of methods is available to deal with both problems in a comparable manner, such that the character and result of the approximation are thoroughly understood.

The results, that are presented in this thesis, contribute to the development of such a situation. For this purpose attention is given, first, to the problem of model reduction and, secondly, to the application of orthonormal functions for model descriptions and system identification. The use of orthogonality in this context is motivated by the close relation of this concept with the principle of approximation, such as is for instance the case with projections. The underlying idea is that, if a model reduction method is based on a certain orthogonality principle, then this may lead to specific orthonormal functions that can be applied for approximate system identification.

For model reduction a method is presented that generalizes existing reduction methods, that are based on balanced realizations. With this method the frequency behavior of the approximations can be improved considerably. The combination of this method with fractional descriptions of linear systems yields another model reduction method.

For the application of orthonormal functions first the use of discrete Laguerre polynomials for the description of signals and systems is analyzed. This class of polynomials is then extended to a broader class of orthonormal functions, that are induced by linear systems. These functions constitute a basis of the function space ℓ_2 and are used to derive alternative descriptions of signals and systems in this new basis. If a system is described in the basis, which is induced by the system itself, then in the alternative description the system has only two non zero Markov parameters. In this case the dynamic behavior of the system is more or less covered by the orthonormal functions. These 'system based' orthonormal functions are induced by balanced state space realizations, which are the basis of several model reduction methods.

The application of classical identification methods to these alternative system descriptions yields new methods to deal with the problem of approximate system identification. From these methods the estimation of ARX and FIR models in orthogonal terms is considered in detail and it is shown that the alternative methods can improve the results of their classical counterparts substantially. The search for an optimal set of orthonormal functions in this context is complementary to the aim of identification, i.e. the determination of the dynamic structure that is present in the measurement data.

We conclude that the application of orthonormal functions, combined with classical identification methods, yields a promising approach to deal with the problem of approximate system identification.

1. INTRODUCTION, MOTIVATION AND PROBLEM SETTING

1.1. INTRODUCTION

In this preliminary chapter we will describe the background, the problem setting and the contents of this thesis. The research that resulted in this manuscript was initiated with the motivation that classical prediction error identification methods failed to meet certain demands, such as a fair estimate of low frequent system behavior. For applications, such as the development of adequate control systems, demands of this nature can be of high importance. In the next section we briefly explore the field of what we will refer to as system approximation, which is the central theme of this thesis. This results in section 1.3 in a problem setting and a number of subjects that are studied in the remaining chapters. In section 1.4 we describe the contents of the thesis and an overview of the main results.

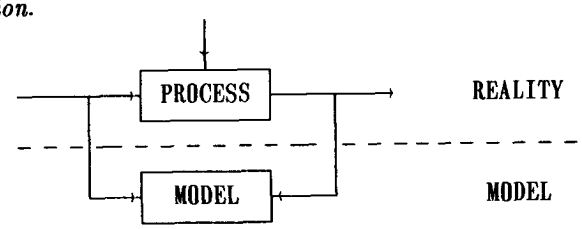
1.2 SYSTEM APPROXIMATION.

The problem of modeling can in general terms be described as the task to develop (mathematical) models of dynamic processes. Such models play a significant role in many sciences, such as economics, physics and control engineering. In the many possible approaches for this *modeling process* one can distinguish between two extremes, the so called black box and white box modeling. The latter method is based on the formalization of fundamental knowledge of a dynamic process, such as physical laws and relations. This method will describe the apple that falls off a tree with the use of Newtons second law. Black box modeling does not use any knowledge of the system, it is solely based on the behavior of the process, which is described by measurable signals, the so called measurement data. A black boxer measures the apple, the tree, the velocity, the time etc, and will try to come up with a model of the phenomenon purely on the basis of these data. For obvious reasons a combination of both methods is referred to as grey box modeling. A greyish person will apply Newtons law and will use the measurement data to determine the influence of for instance the shape of the apply and the friction caused by the air.

There are many reasons for creating mathematical models, some of which are:

- to get a better understanding of the process, which can be used for several purposes as for the development of new processes and for fault detection;
- to predict the behavior of the model in the future;
- to validate other existing models, for instance the result of theoretical modeling;
- to use the model as a training device for operators;
- to design control systems for the process.

In this thesis we will merely consider black box modeling, i.e. the creation of mathematical models on the basis of measurement data, commonly also referred to as *system identification.*



For this approach we can recognize a number of essential steps in any procedure which is used to tackle the problem (Ljung, 1987):

a. Experiment design and data collection.
b. Data manipulation.
c. Choice of the class of models to be considered.
d. Parametrization of this model class, leading to a set of unknown parameters, that have to be *identified.*
e. Choice of the *identification criterion,* i.e how to discriminate between the different elements of the model class.
f. Calculation of the models that are optimal with respect to the criterion.
g. Validation of the resulting *model estimates.*

Often the model is required to be of a restricted complexity, for instance because the computer that will be used to evaluate or apply the model can only deal with models of a certain complexity. In this case we can add an extra step:

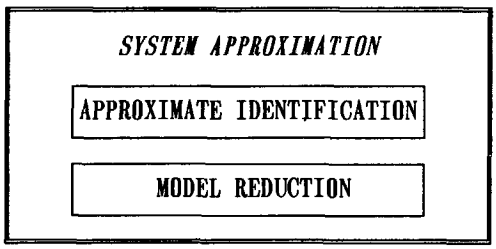
h. Model reduction, the approximation of a model with a lower order model.

These steps cannot be carried out independently, each one will influence the outcome of some of the following steps. Often it will be necessary to repeat previous steps with different settings, due to obtained knowledge. It is for example very well possible that validation of the model (g) implies that the class of models (c) is too restricted and this may cause the need for a new experiment (a).

For the aim of an identification procedure one can in principle choose between two premises; either one has the goal of finding the true model of the process assuming that such a model does exist, or one aims at finding a model that approximates the dynamic process in a best way, defined by a specific criterion. While classical identification merely used the first premise (Eykhoff, 1974), in the past decade the conviction has risen that, especially given the commonly considered restrictive model classes, there is no such thing as the true model (Ljung, 1987).

> The position of the moon and the stars will influence the speed of the apple, but such phenomena will not be taken into account.

Instead one considers the problem of approximate identification, or approximate modeling. Given a model class and a criterion, which model describes the behavior of the process in the best way with respect to that criterion. This conceptual change of philosophy has a great impact on the evaluation of identification procedures. Furthermore it motivates to link the field of identification with the field of model reduction that deals with the approximation of a given model, based on a description of that model. In the philosophy of Willems (1986), where systems are described by all–possible signal trajectories, we can consider model reduction as an approximation based on the knowledge of all signal trajectories, while identification makes an approximation on the basis of one element of all these data. We combine both problems, the approximation on the basis of measurement data and the approximation on the basis of model descriptions under the term *system approximation*. We feel that both fields deal essentially with the same problem, *the optimal approximation of the behavior of a dynamic process.*



In an ideal situation one would like to be able to choose from multiple approximation criteria, with a clear understanding of the character of the approximation and of the pros and cons of the resulting model estimates. Unfortunately there are few results known on the essence of the approximation, both for model reduction and approximate identification.

For identification only the use of least squares criteria for prediction error models is relatively well understood (Swaanenburg and co–workers, 1985; Wahlberg and Ljung, 1986; Van den Hof and Janssen, 1987). For model reduction, though many different methods have been proposed and have been proven to be of practical use, only the optimal Hankel norm approximation method is thoroughly understood in the sense that this method is optimal with respect to a specific criterion. Nor the use of least squares criteria in model reduction nor a Hankel norm criterion for approximate system identification have so far led to comparable results. While identification often uses simple criteria without understanding of the approximation character, there are many model approximation methods known for which it is unclear which criterion is used.
The recent developments and achievements in the area of H∞–control design clearly motivate the approximation of systems on the basis of robustness criteria, but there is yet no adequate theory to solve this problem. These observations lead to the conclusion that first there is still a gap of knowledge to be bridged in order to understand the general problem of system approximation and, secondly, that the theory has to be enriched with approximation on the basis of new criteria.

The most commonly applied identification procedures are methods that aim at the minimization of the one step ahead prediction error. Such PE methods often have the property that they emphasize the high frequency characteristics of a process and this effect is magnified if the sample frequency is increased. For applications with the resulting model, such as long range prediction and control engineering, it is of importance to have a fair approximation of other frequency ranges as well. Many systems include dynamical effects that work on different time scales and these effects are too important to neglect. Consider for instance the fast respectively slow transitions of pressure and temperature changes in steam boilers, both effects have a considerable influence on the behavior of the process, but PE methods cannot estimate both this high and low frequent behavior. The resulting problem can in principle be overcome with the application of prefilters on the measurement data in the data manipulation step of the identification procedure. However in general it is unknown at beforehand which prefilter should be used to obtain satisfactory results. On the contrary, such a prefilter can only be designed with knowledge of the to be identified process and only after the estimation procedure has been performed one can evaluate the quality of the applied prefilter (Ljung, 1987, 1989).
Another problem that occurs in the application of PE methods with the common least squares criteria is that for the general model structures, as proposed by Ljung (1987), no analytical solution exists to the resulting estimation problem. The criterion function is non–linear in the unknown parameters and time consuming computer optimization algorithms have to be used to tackle the problem. No guaranty can be given that such

programs will actually come up with an optimal solution. However, two specific model structures exist, the so called ARX and FIR model classes, which lead to criterion functions that are linear in the parameters and hence for these model classes analytical solutions to the estimation problem do exist and are easily calculated. These model structures have the drawback that they are either not very feasible from a practical point of view (ARX), or that they have an intrinsic over–parametrization (FIR). The latter can cause severe problems when the number of data is relatively small.
This leads to another observation in the context of approximation. Classical identification procedures are (theoretically) based on the use of an infinite number of data. In practice this will never be the case and in many situations the number of available data will be relatively small.

How many times can the same apple fall off the same tree?

This effect of finite data on the model estimates is an important issue in identification and approximation, which is hardly understood.

1.3 PROBLEM STATEMENT AND MOTIVATION

In the previous section we showed that the problem of system approximation raises a lot of questions, which are not answered yet, and that the problem needs and deserves much more research. The aim of our research was the development of approximate identification methods that are either in a natural way related to or intrinsically based on model reduction methods, with well defined optimization criteria. This resulted in the following problem statement:

Analyze the possibilities to approach the problem of approximate system identification through the use of orthogonal functions and prediction error techniques.

This was motivated with the following observations and questions in mind:

- The notion of orthogonality is closely connected with the principle of approximation, with many examples such as in
  - signal analysis and Fourier transformation
  - Kalman filtering
  - orthogonal projections
  - singular value decompositions

  In this light the approximation of models asks for an orthogonal representation of the model, such that the approximation can be considered as a projection on a lower dimensional space.

- In the field of model reduction and model description the concept of orthogonality is intrinsic to for instance
  - Balanced realizations (Moore, 1981)
  - Hankel norm approximation (Adamyan, Arov and Krein, 1968, 1978)
  - Inner/outer and coprime factorizations (Vidyasagar, 1985)

  and though many model reduction methods are not directly based on an orthogonal projection, it is worthwhile to consider the question if specific orthogonal functions exist such that a projection in terms of these functions is equivalent with a model reduction procedure, such as balanced truncation and Hankel norm approximation. Such functions, based on model reduction, might be used for system identification, for instance to approximate signals, and would interrelate both fields.
- The question whether there exists a class of orthogonal functions that can be used to derive alternative descriptions of systems, such that essential properties of a system can in a transparent way be obtained from the new description.
- The latter question raises a new one: does it make sense to apply PE methods to such alternative system descriptions, and can this procedure improve the frequency behavior of the resulting models?

We do not have the pretension that we are able to solve the problem, we merely want to explore alternative ideas and approaches, linking on the one side with the ideas of classical methods and expanding the horizon on the other side.
To this end we will first consider model reduction as a problem of its own and study c.q generalize existing methods in order to get a better understanding of the underlying structures and mechanisms.
In very general terms we aim at alternative descriptions of linear systems, based on decompositions in orthogonal ($\approx$ independent) components. The motivation is that this will enable us to construct approximations using the most important components. In the process we want to establish a connection between the mechanisms of model reduction methods and the orthogonal bases that are used for the system approximation. This yields the so called system based orthogonal functions, i.e orthogonal bases that are based on specific linear systems. These bases shall be used to derive alternative descriptions of signals and systems in the so called orthogonal domain. The concept can be compared with the description of systems and signals in frequency domain terms. We propose to use these new descriptions, that on data level can be considered as prefilter operations, to derive identification algorithms in the orthogonal domain. This will result in a number of identification methods based on the application of pre–specified prefilters and the use of least squares criteria with simple model structures.

We present a theory with new ideas and identification methods, that have appealing properties and can be applied successfully, though the mathematical background and properties have not completely been developed. We left many open ends and questions for future research. The relation between model reduction and system identification remains a problem, but the use of orthonormal functions opens interesting and promising perspectives to approach this problem

1.4 OVERVIEW OF CONTENTS AND RESULTS

This thesis is divided into four major chapters, next to this introduction and the conclusions in chapter six.

Chapter two is devoted to the basic (mathematical and system theoretical) tools that will be used in the remainder. This includes the notions of finite dimensional linear time–invariant systems, transfer functions and state space realizations. We present a short overview of related concepts such as Gramians, Hankel singular values, balanced realizations and all–pass functions. For the latter we derive a new result on the description of these functions in the discrete time domain, which will be used to derive formulas for discrete time coprime factorizations. This will enable us in chapter three to define a new model reduction method. The description of all–pass functions will be the basis for the theory of chapter five.

In chapter three we explore three different types of model reduction procedures. First we consider the methods that are based on balanced realizations. It will be explained that the existing methods can be embedded in a broader class of methods, by introduction of a new design variable, the so called reduction factor. It is shown that the freedom can be used to improve the behavior of the reduced order models, where we put emphasis on the frequency characteristics in terms of the infinity norm of the error transfer function.
The second method we consider is the optimal Hankel norm approximation and we explore the structure of this approximation problem for discrete time systems.
The application of the theory on coprime factorization (chapter two) in combination with the generalized reduction method based on balanced realizations (see above) yields a third approach for model reduction. Contrary to the other methods, that are solely based on model descriptions, this method can also be considered in terms of the signal behavior of the system.

The concept of orthogonal functions is introduced in chapter four, where we consider the discrete Laguerre polynomials. The set of these functions is used as an alternative basis

for the space of ℓ_2 functions. This is applied to define coefficients of a function expansion in this new basis and induces a transformation between a time function and the set of its coefficients. This transformation is explained in some detail, with remarks on the approximation character if such a transformation is applied to finite data series. The application of this concept to the input/output signals of a linear system yields an alternative description of a system in terms of the orthonormal coefficients of the latter signals. We consider three approaches for system identification in this new domain, illustrated with examples. The Laguerre polynomials introduce one degree of freedom, that for the applications to identification can be considered as a new design variable. Further it is shown that the transformation of signals with these polynomials can in a classical sense be interpreted as a prefilter operation.

In chapter five we consider the question if a linear system gives rise to a specific set of orthonormal functions and we will show that indeed for every system we can define an orthonormal basis. The derivation is based on the description of discrete all–pass functions and on input balanced realizations (see chapter two). It is shown that such functions are a generalization of the set of Laguerre polynomials, with the introduction of more degrees of freedom. These system based functions can again (as in chapter four) be used to transform time series and linear systems to an orthonormal domain. The theory is more involved than in the Laguerre case and we go through many steps to proof the property that a transformed system is again a linear system with the same McMillan degree, Hankel norm and infinity norm. If a system is transformed with the set of functions that are based on this system itself, then the transformed system will have the special property that it has only uncoupled poles in z=0 and hence only two non zero Markov parameters. We say that in this case the dynamic behavior is removed by the transformation. Again the transformation of data with these general orthonormal functions can be interpreted as prefiltering in a classical sense.
We present two approaches to use these functions and transformations for approximate identification, which can be seen as the application of equation error identification in the orthonormal domain and a generalization of the estimation of a finite number of Markov parameters. These methods will be illustrated with examples, where they will be compared with their time domain equivalents.
The search for an optimal set of orthonormal functions is complementary to the identification problem, in the sense that both are aimed at an explanation of the structure in the data. Different orthonormal functions will result in different model estimates, they emphasize different parts of the behavior.

2. DYNAMICAL SYSTEMS AND REALIZATIONS

2.1 INTRODUCTION

This chapter is devoted to the basics of this thesis and will treat a number of basic mathematical and system theoretical tools, that are fundamental for the theory which will be developed in the subsequent chapters. In the scope of this thesis we consider those dynamical systems that can be described by proper transfer functions, and next to this representation we will use state space models and Markov parameter descriptions. This more or less classical definition of a dynamical system coincides with the behavioral approach of Willems (1986,1988) if a restriction is made to controllable systems and if an a priori distinction is made between input and output signals. Within this class of systems we only consider finite dimensional linear time invariant non anticipating input output systems. Further we make the distinction between continuous time systems and discrete time systems.

The main notions are presented in the sections 2.3 and 2.4, where we also consider some related concepts, such as controllability, observability and Gramians of state space representations, Hankel operators and Hankel singular values. In section 2.5 we briefly review a transformation between continuous and discrete time systems that will be used extensively in this thesis, the bilinear ω–transformation. Another transformation, but now of state space representations of dynamical systems, known as balancing transformations, is considered in section 2.6. A special class of systems, the all–pass systems are the subject of section 2.7, where also the implications of this property for state space representations are evaluated. This is a new result in case the system is discrete. The consequences of this result for the existence of coprime factorizations are the subject of section 2.8, where an also new result for factorizations of discrete time systems is presented. Though this chapter is meant as a review of system theoretical concepts, we choose to include these new results because they are fundamental for the theory of the remaining chapters. Sections 2.9 and 2.10 are devoted to basic mathematical concepts which are used in this thesis, Kronecker calculus, norms and spaces. We conclude with a short discussion in section 2.11. The proofs of the derived results are given in appendices at the end of the chapter.

2.2. DYNAMICAL SYSTEMS

For our purposes we consider a dynamical system to be a physical system or a mathematical model to which is applied an input signal u and which reproduces output signals y. Thus a system can be considered as a mapping from inputs to outputs, in the sense that the outputs are –besides initial conditions– completely determined by the input. The input and output are real or complex valued functions of time and generally u(t) will be an m–vector and y(t) a p–vector. A multivariable system is one in which max(p,m)>1 and a common way to divide this field is to talk about SISO (p=m=1), SIMO (p>1,m=1), MISO (p=1,m>1) and MIMO (p,m>1) systems, abbreviations for Multi/Single Input/Output systems. We restrict attention to finite dimensional linear time invariant non anticipating systems, which can be described by their transfer function. In the context of the behavioral definition of systems from Willems (1986,1988), where systems are considered as sets of possible signal trajectories that obey the underlying laws and restrictions of a system, our system concept is equivalent with non anticipating controllable input output systems. Next to the transfer function description we will mainly be concerned with two other representations or descriptions of systems, the state space models and the Markov parameter descriptions.
We will make a distinction between two time domains, the continuous time domain, where u(t) and y(t) are functions with time domain T=$\mathbb{R}$ and the discrete time case, where the time domain is T=$\mathbb{Z}$. The field in which these functions take on their values is often denoted with $\mathbb{F}$, which can be either $\mathbb{R}$ or $\mathbb{C}$:

continuous time systems: $u:\mathbb{R} \to \mathbb{F}^m$ $y:\mathbb{R} \to \mathbb{F}^p$

discrete time systems: $u:\mathbb{Z} \to \mathbb{F}^m$ $y:\mathbb{Z} \to \mathbb{F}^p$

In the next two sections we consider systems in both time domains and we use the abbreviation **FDLT system** to denote that we consider Finite Dimensional Linear Time invariant systems.

2.3 REPRESENTATIONS OF CONTINUOUS TIME SYSTEMS

In this section we consider the case that the time set T=$\mathbb{R}$, to which we will refer as the continuous time case. We define the notion of a dynamical input/output system and present two representations of a system, where we restrict systems to be finite dimensional, linear, time invariant and causal. We emphasize that for a given input signal there exists more than one output signal; the outputs are determined by the input, the system and a set of initial conditions. For the state space representation we introduce the notions of controllability, observability and Gramians and we give a definition of the Hankel singular values of a system.

We define a p×m dynamical input/output system as a mapping $G:(\mathbb{F}^m)^{\mathbb{R}} \to 2^{(\mathbb{F}^p)^{\mathbb{R}}}$ [1] and we write $\mathcal{Y}=Gu$, if u and $y\in\mathcal{Y}$ are functions $u:\mathbb{R} \to \mathbb{F}^m$, $y:\mathbb{R} \to \mathbb{F}^p$. Hence, we explicitly state that a signal u corresponds with more than one signal y. We refer to the signal u as an **input** signal and the signals $y\in\mathcal{Y}$ are called **output** signals. A pair of signals (u,y) with $\mathcal{Y}=Gu$ and $y\in\mathcal{Y}$ is called an input/output pair of G and with abuse of notation we will write y=Gu. In the scope of this thesis we only consider **finite dimensional linear time invariant systems.** The restriction to this class of systems implies that we can identify the mapping G with the **transfer function matrix G(s),** which is a p×m matrix with entries $G_{ij}(s)\in\mathbb{F}(s)$. Furthermore, we restrict the systems to be **causal** or non anticipating, which is equivalent with $\lim_{s\to\infty} G_{ij}(s)=k_{ij}<\infty$ [2]. Such transfer functions are called **proper** and if $\forall ij\ k_{ij} = 0$ we call them **strictly proper.** In this thesis we will refer to an element of this class of systems as a **FDLT** system.

A transfer function can be interpreted as a mapping of time functions if we identify the intermediate s with the differential operator d/dt:

$$(su)(t):=\frac{du(t)}{dt} \tag{2.3.1}$$

and if we decompose $G(s)=P(s)^{-1}Q(s)$, with P(s) and Q(s) matrices with entries $P_{ij}(s),Q_{ik}(s)\in\mathbb{F}[s]$, with the restriction that $\det(P(s))\neq 0$ and P(s) and Q(s) are left coprime with respect to $\mathbb{F}[s]$. Let u be a sufficiently smooth m–dimensional time function, then we define $\mathcal{Y}=Gu$ as the set of signals y that obey

$$P(s)y(t)=Q(s)u(t) \tag{2.3.2}$$

Besides a finite number of initial conditions on y, this relation completely determines the p–dimensional signal y. Let $y_0(t)$ with $y_0(-\infty)=0$ [3] be a solution to (2.3.2) and let $y_1(t)$ be a solution to $P(s)y_1(t)=0$ then $y=y_0+y_1$ is also a solution to (2.3.2). In general there will be infinitely many such signals y_1 and they can only be restricted with a set of initial conditions.

It is well known (Kailath, 1980) that if we denote by Y(s) and U(s) the Laplace transforms of y(t) respectively u(t), with the restrictions that $u(-\infty)=0=y(-\infty)$, then these transforms are related by Y(s)=G(s)U(s).

Remark 2.3.1. We present the notion of a system in this formal manner to emphasize that for a given transfer function G(s) and a given input there exist many outputs such

---

[1] Where 2^A stands for the set of all subsets of A: $2^A := \{ X \mid X\subseteq A \}$.

[2] This is a slight abuse of notation, since s is an indeterminate. A more formal definition can be given by defining $G_{ij}(s)=p(s)/q(s)$ and requiring that degree(p)≤degree(q).

[3] In the sense that $\exists M<0$ with $y(t)=0$ for $t<M$.

that (u,y) is an input/output pair of G. Consider for example the transfer function $G(s)=\frac{1}{s+1}$ then $G(s)=P(s)^{-1}Q(s)$ as in (2.3.2) with $P(s)=s+1$ and $Q(s)=1$. Hence all input/output pairs (u,y) of G(s) are characterized by:

$$\dot{y}(t)+y(t)=u(t) \tag{2.3.3a}$$

Choose a specific input $u(t)=0$, then it follows that for every constant $c\in\mathbb{F}$

$$y_c(t)=ce^{-t} \tag{2.3.3b}$$

is a solution to equation (2.3.3a) and a specific solution can only be obtained by imposing an ('initial') condition $y(t^0)=y^0$. To omit this phenomenon some authors restrict the outputs to those signals y with $y(-\infty)=0$, in which case an input u corresponds with only one output y. o

Any p×m FDLT system G(s) can be described with a set of first order ordinary differential equations in the sense that there exists a scalar $n\in\mathbb{N}$ and (constant) matrices $A\in\mathbb{F}^{n\times n}$, $B\in\mathbb{F}^{n\times m}$, $C\in\mathbb{F}^{p\times n}$, $D\in\mathbb{F}^{p\times m}$, such that for every $u\in(\mathbb{F}^m)^{\mathbb{R}}$ and $y\in(\mathbb{F}^p)^{\mathbb{R}}$ with $y=Gu$ there exists an n dimensional differentiable function $x\in(\mathbb{F}^n)^{\mathbb{R}}$ such that:

$$\dot{x}(t)=Ax(t) + Bu(t) \tag{2.3.4a}$$

$$y(t)=Cx(t) + Du(t) \tag{2.3.4b}$$

The representation (2.3.4) is called a **state space representation** of the system G(s) and the signal x is called the **state** vector of this representation.
A signal y that obeys y=Gu, in the sense of (2.3.2), with a set of initial conditions on y can always be described in a state space representation with an initial condition on x. The relation between G(s) and the set of matrices {A,B,C,D} in (2.3.4) is described by:

$$G(s)=D + C[sI-A]^{-1}B \tag{2.3.5}$$

and we call the quadruple [A,B,C,D] a (**state space**) **realization** of G(s).

It is well known that there exist many different realizations of the same transfer function G(s), hence the state space representation (2.3.4) is not unique. If [A,B,C,D] is a realization of G(s) and $T\in\mathbb{F}^{n\times n}$ is non singular, then $[T^{-1}AT,T^{-1}B,CT,D]$ is also a realization of G(s); such transformations are known as **similarity transformations** with the **similarity matrix** T. Furthermore there exist realizations with state vectors of different dimensions.

A second way to represent a FDLT system is the **Markov parameter description**, obtained by applying a Laurent series expansion around $s=\infty$ to the transfer function G(s):

$$G(s) = \sum_{i=0}^{\infty} h_i s^{-i} \tag{2.3.6}$$

where the $\{h_i\}$ are called the Markov parameters. They are related to the state space realization (2.3.4) by $h_o=D$ and $h_i=CA^{i-1}B$.

Remark 2.3.2. It should be noted that if we consider the Markov parameter description in terms of the operator $s=d/dt$ or $s^{-1}=\int_{-\infty}^{t}$, and examine which signals y can be described by $\dot{y}(t)=(\sum_{i=0}^{\infty} h_i s^{-i})u(t)$ that we will find only one such signal y. Consider for example, as in remark 2.3.1 an input $u=0$, then the only possible output is $y=0$. The reason for this phenomenon is that for the expansion (2.3.6) we have to interpret G(s) as a function of the Laplace operator s and only consider signals whose Laplace transforms obey $Y(s)=G(s)U(s)$, which is equivalent with considering the initial conditions $y(-\infty)=0=u(-\infty)$. This fact, that each input u results with this description in a unique output y, is is the reason why the representation (2.3.6) is often used to define systems, instead of the definition in terms of transfer functions. o

Remark 2.3.3. For every output $y\in\mathcal{Y}=Gu$ one can distinguish two signals y_0 and y_1, such that $y=y_0+y_1$, where the signal y_0 is uniquely determined by the input signal u, as in remark 2.3.2, and where the signal y_1 results from the initial conditions. In the behavioral approach of Willems (1986,1988) this is formalized by Van den Hof (1989) who presents a decomposition of systems in an initial zero part (y_0) and an autonomous part (y_1). We further note that the definition of systems by transfer functions is equivalent with the behavioral notion of controllable systems under an a priori selection of input and output signals. o

The notions of **controllability** and **observability** (Kailath, 1980) are intrinsic to state space descriptions and with each realization [A,B,C,D] we identify the controllability matrix M_c and the observability matrix M_o,

$$M_c = \left[B|AB|A^2B|A^3B\ \cdots\right] \tag{2.3.7a}$$

$$M_o = \left[C^*|A^*C^*|A^{*^2}C^*|A^{*^3}C^*\ \cdots\right]^* \tag{2.3.7b}$$

The realization is controllable respectively observable if M_c respectively M_o^* has full row rank n. A realization that is both controllable and observable is called **minimal**. In that case A (for a given G(s)) is of minimal dimension and we call n the **McMillan degree** of G(s). See Kailath (1980) for a thorough treatment of this notion.

A realization [A,B,C,D] is called (**asymptotically**) **stable**, if $Real(\sigma(A)) < 0$, where $\sigma(A)$ is the set of eigenvalues of the matrix A. A transfer function G(s) is stable if any minimal realization of G(s) is stable. A stable FDLT system will be denoted by a **FDLTS** system. If a realization [A,B,C,D] is stable (all eigenvalues of A are in the open left half plane), then we can define the **controllability Gramian** P and the **observability**

Gramian Q,

$$P := \int_0^\infty e^{At}BB^*e^{A^*t}dt \qquad Q := \int_0^\infty e^{A^*t}C^*Ce^{At}dt \qquad (2.3.8a,b)$$

It is straightforward to show, using (2.3.8), that P and Q satisfy the following **Lyapunov equations,**

$$AP + PA^* + BB^* = 0 \qquad A^*Q + QA + C^*C = 0 \qquad (2.3.9a,b)$$

From (2.3.8) it is clear that $P \geq 0$ and $Q \geq 0$. It is standard (Kailath, 1980) that for A stable: $P>0$ iff [A,B,C,D] controllable $Q>0$ iff [A,B,C,D] observable.

An interpretation of these Gramians is given in, among others, (Wahlberg, 1984; Glover, 1984). There it is shown that the minimal amount of input energy, necessary to reach a state $\bar{x}$ at $t=\infty$, starting from $x(0)=0$, is equal to $\bar{x}^*P^{-1}\bar{x}$, hence if P is nearly singular there will be some states that can only be reached with a large amount of energy. Such states are therefore called weakly controllable.
If $x(0)=\bar{x}$ and $u(t)\equiv 0$, $t \geq 0$, then the amount of energy in the output $y(t)$ on $[0,\infty)$ is $\bar{x}^*Q\bar{x}$, showing that if Q is nearly singular some states will have little effect on the output. Such states are thus weakly observable.

The equations (2.3.9) make it possible to calculate P and Q in a numeric efficient way, contrary to calculation of the integral expressions (2.3.8), (Bartels and Stewart, 1972; Stewart, 1973; Laub, 1980). Note that the Gramians depend strongly on the realization which is considered; however, their product has an invariance property.

Proposition 2.3.4. [Glover, 1984] Let G(s) be a FDLTS system and [A,B,C,D] a stable realization of G(s) with Gramians P and Q. The eigenvalues of the product PQ are invariant under similarity transformations. ○

Note that we do not require minimality, but if the realization is not minimal we know that PQ will have at least one zero eigenvalue. The invariant eigenvalues of PQ play a central role in the field of system and control theory and also in this thesis. In defining them we restrict ourselves to the non–zero elements to omit ambiguity.

Definition 2.3.5. Let G(s) be a FDLTS system and [A,B,C,D] a stable realization of G(s) with Gramians P and Q. Let $\{\lambda_k\}_{k=1}^n = \sigma(PQ)$ (the set of eigenvalues of PQ). The **Hankel singular values** of G(s) are defined as

$$\sigma_i(G(s)) = \{\lambda_i^{\frac{1}{2}}, \lambda_i \neq 0\} \qquad (2.3.10)$$

○

If we would restrict the realization to be stable and minimal then both P and Q would be positive and hence their product would be non–singular and there would not be any zero eigenvalues. These invariants play a central role in the field of model reduction, c.f. balanced truncation (section 3.2), Hankel norm approximation (section 3.6), but also in various other areas as H∞–control (Francis, 1987; Doyle and co–workers, 1989), all–pass transfer functions (Glover, 1984), coprime factorizations (Meyer, 1988) and system identification (Maciejowski and Ober, 1988).
The term Hankel singular values stems from the fact that these invariants can also be interpreted (Glover, 1984) as the singular values of the **Hankel operator**

$$\mathscr{H}: L_2(0,\infty) \to L_2(0,\infty), \quad \text{(for } L_2 \text{ see section 2.10)}$$
$$(\mathscr{H}u)(t) = \int_0^\infty Ce^{A(t+\tau)}Bu(\tau)d\tau \qquad (2.3.11)$$

the operator that in a certain sense (Glover, 1984) maps past inputs into future outputs.

2.4 REPRESENTATIONS OF DISCRETE TIME SYSTEMS

In this section we give the discrete analog (time domain $\mathbb{Z}$) of the concepts of continuous time systems, without going into much detail. Again we note that for a given system G and input u there is no unique output signal y, but a set of possible outputs, which are determined by specific initial conditions. This is explained in a similar way as in section 2.2.3. For this class of systems we will define the notion of the (signal) behavior of a system, which will be used in the chapters four and five.

A finite dimensional linear time invariant causal (**FDLT**) system is defined by the transfer function G(z), that maps inputs in a set of outputs, in the same sense as was explained in section 2.3. The indeterminate z should in an operator sense be considered as the forward shift operator and its inverse as the backward shift:

$$(zu)(t)=u(t+1) \qquad (2.4.1a)$$
$$(z^{-1}u)(t)=u(t-1) \qquad (2.4.1b)$$

A **state space representation** of a FDLT discrete time dynamical system is given by a set of difference equations,

$$x_{k+1}= Ax_k + Bu_k \qquad (2.4.2a)$$
$$y_k= Cx_k + Du_k \qquad (2.4.2b)$$

The matrices $A\in\mathbb{F}^{n\times n}$, $B\in\mathbb{F}^{n\times m}$, $C\in\mathbb{F}^{p\times n}$ and $D\in\mathbb{F}^{p\times m}$ are assumed to be constant and the vectors $x_k\in\mathbb{F}^n$, $u_k\in\mathbb{F}^m$, and $y_k\in\mathbb{F}^p$ are respectively the **state**, the **input** and the **output** of the system. An initial condition that determines a specific output $y\in\mathcal{Y}=G(z)u$ can always be represented by an initial condition on x in the state space representation.

The relation between the transfer function G(z) and the matrices of the state space representation is given by:

$$G(z) = D + C[zI-A]^{-1}B \tag{2.4.3}$$

As in section 2.3 we call the quadruple [A,B,C,D] a (**state space**) **realization** of G(z) and note that these realizations are not unique.

A second way to represent a system is the **Markov parameter description**, through application of a Laurent series expansion around $z=\infty$ of G(z):

$$G(z) = \sum_{i=0}^{\infty} M_i z^{-i} \tag{2.4.4a}$$

where the **Markov parameters** $\{M_i\}$ can be expressed in the matrices of a state space representation (2.4.2) by

$$M_o = D \text{ and } M_i = CA^{i-1}B \tag{2.4.4b}$$

Again it follows that in this description, an input u results in a unique output y.

With each realization [A,B,C,D] we identify the controllability matrix M_c and the observability matrix M_o:

$$M_c = \left[B|AB|A^2B|A^3B\ \cdots\right] \tag{2.4.5a}$$

$$M_o = \left[C^*|A^*C^*|A^{*^2}C^*|A^{*^3}C^*\ \cdots\right]^* \tag{2.4.5b}$$

We call the realization **controllable** respectively **observable** iff M_c respectively M_o has full row rank n. This notion of controllability is better known as **reachability** or controllability p.s.f.o (pointwise state from the origin) (Kailath, 1980; Rosenbrock, 1970). Our notion of observability is better known as **constructability**.
A realization that is both controllable and observable is called **minimal**. In that case A is of minimal dimension and we call n the **McMillan degree** of G(z). A realization is called (**asymptotically**) **stable**, if $|\sigma(A)|<1$. A transfer function G(z) is called stable if any minimal realization of F(z) is stable and again we denote stable FDLT systems as **FDLTS** systems.

If a state space realization is stable (all eigenvalues of A lie inside the unit circle), then we can define the **controllability Gramian** P and **observability Gramian** Q,

$$P = \sum_{k=0}^{\infty} A^k BB^* A^{*k} \qquad Q = \sum_{k=0}^{\infty} A^{*k} C^* CA^k \tag{2.4.6a,b}$$

In this time domain the Gramians can directly be related to the controllability and observability matrices M_c and M_o, as defined by (2.4.5),

$$P = M_c M_c^* \qquad Q = M_o^* M_o \tag{2.4.7a,b}$$

It is straightforward to show, using (2.4.6), that P and Q satisfy the following **Lyapunov equations**,

$$APA^* + BB^* = P \qquad A^*QA + C^*C = Q \qquad (2.4.8a,b)$$

From (2.4.7) it is clear that $P \geq 0$ and $Q \geq 0$ and will be off full rank iff M_c respectively M_o^* have full row rank. So we have that – for A stable –

$$P>0 \text{ iff } [A,B,C,D] \text{ controllable} \qquad Q>0 \text{ iff } [A,B,C,D] \text{ observable.}$$

We have again an invariance property for the eigenvalues of the product PQ :

Proposition 2.4.1. [Glover, 1984] Let G(z) be a FDLTS system and [A,B,C,D] a stable realization of G(z) with Gramians P and Q. The eigenvalues of the product PQ are invariant under similarity transformations. o

Definition 2.4.2. Let G(z) be a FDLTS system and [A,B,C,D] be a stable realization of G(z) with Gramians P and Q. Let $\{\lambda_k\}_{k=1}^{n} = \sigma(PQ)$. The **Hankel singular values** of G(z) are defined as:

$$\sigma_i(G(z)) = \{\lambda_i^{\frac{1}{2}}, \lambda_i \neq 0\} \qquad (2.4.9)$$

o

As in the continuous time case we could make a restriction to minimal realizations in which case the product PQ will have no zero eigenvalues. We can again consider the Hankel operator of G(z) and in discrete time it is much easier to see that the singular values of this operator are the Hankel singular values of definition 2.4.2. In discrete time this operator is the so called (block) **Hankel matrix** $\mathscr{H}$,

$$\mathscr{H} = \begin{bmatrix} M_1 & M_2 & M_3 \cdots \\ M_2 & M_3 & M_4 \cdots \\ M_4 & M_5 & M_6 \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \qquad (2.4.10)$$

where the Markov parameters $\{M_i\}$ are defined by (2.4.4). So $\mathscr{H}$ is built up from impulse response parameters of the system G(z). This Hankel matrix can be seen as a map from 'past' inputs into 'future' outputs as follows:

let $u_t=0$ $t>0$ and let $\{u_t, y_t, -\infty \leq t \leq \infty\}$ be an input/output pair of G(z) with $y(-\infty)=0$. Consider the vectors of past inputs $\underline{U}:=[u_0^* | u_{-1}^* | u_{-2}^* | \cdots]^*$ and future outputs $\underline{Y}:=[y_1^* | y_2^* | y_3^* | \cdots]^*$ then it follows that $\underline{Y} = \mathscr{H} \cdot \underline{U}$

It is easy to see that $\mathscr{H} = M_o M_c$ and thus $\text{rank}(\mathscr{H}) \leq n$, with equality iff the realization is minimal. Thus we have that $\text{rank}(\mathscr{H})$ = McMillan degree of G(z). The singular values of $\mathscr{H}$ are the square roots of the eigenvalues of $\mathscr{H}\mathscr{H}^*$:

$$\sigma_i^2(\mathscr{H}) = \lambda_i(\mathscr{H}\mathscr{H}^*) = \lambda_i(M_o M_c M_c^* M_o^*) = \lambda_i(M_c M_c^* M_o^* M_o) = \lambda_i(PQ) \qquad (2.4.11)$$

Hence the Hankel singular values of the system G(z) are the non–zero singular values of the Hankel matrix (2.4.10) that corresponds with G(z).

In the remainder of this thesis the notation G(s) respectively G(z) for a system implies that we are dealing with a continuous respectively discrete time system. We thus identify 's' respectively 'z' with the continuous respectively discrete time domain. Whenever we make a statement that holds for both time domains, we will either use G(p), or just G.

Though we did not adopt the definition of dynamical systems from Willems (1986,1988) we will on a few occasions use the concept of the **behavior** of a discrete time system, which is the set of all possible input/output trajectories. We will also consider the restriction of the behavior to square summable signals, which is denoted by ℓ_2 (see section 2.10 for a formal definition) and the restriction of the latter to $\mathbb{Z}_+$.

Definition 2.4.3. Let G(z) be a (p×m) FDLT system. We define the **behavior** of G(z), denoted with $\mathcal{B}(G)$, as the set of admissible input/output trajectories:

$$\mathcal{B}(G) := \{ (y,u)\in(\mathbb{F}^p\times\mathbb{F}^m)^{\mathbb{Z}} \mid \exists \mathcal{Y}\in 2^{\mathbb{F}^m\mathbb{Z}} \quad y\in\mathcal{Y}=G(z)u \} \qquad \circ$$

Definition 2.4.4. Let G(z) be a (p×m) FDLT system. We define the **ℓ_2–behavior** of G(z), denoted with $\mathcal{B}_2(G)$, as the set of admissible ℓ_2 input/output trajectories:

$$\mathcal{B}_2(G) := \mathcal{B}(G)\cap\ell_{2\mathbb{Z}}^{p+m} \qquad \circ$$

Definition 2.4.5. Let G(z) be a (p×m) FDLT system. We define the **restricted ℓ_2–behavior** of G(z), denoted with $\mathcal{B}_2^+(G)$, as the set of admissible ℓ_2 input/output trajectories restricted to the time domain $\mathbb{Z}_+$:

$$\mathcal{B}_2^+(G) := \{w^+:\mathbb{Z}_+ \to \mathbb{F}^{p+m} \mid \exists w\in\mathcal{B}_2(G)\ w^+(t)=w(t)\ \forall t\in\mathbb{Z}_+\} \qquad \circ$$

2.5 THE ω–TRANSFORMATION

In this thesis we will frequently switch between the continuous and discrete time domains, mostly to prove that results in one time domain are also valid in the other one. The reason for this is that calculations in discrete time are often much more involved then in continuous time. Consider for example the Lyapunov equations (2.3.9) and (2.4.8), where in continuous time only products of two matrices appear, but in discrete time one has to deal with triple matrix products. This implies that the calculations are more involved than for continuous time realizations. For this purpose we will nearly always use a well known bilinear map ω: $\mathbb{C}\to\mathbb{C}$, which is defined by,

$$\omega : s \to z=\frac{(1+s)}{(1-s)} \qquad \omega^{-1} : z \to s=\frac{(z-1)}{(z+1)} \tag{2.5.1}$$

Notice that if z and s are subject to (2.5.1) then:

$$|z|<1 \Leftrightarrow \text{Real}(s)<0 \quad \text{and} \quad |z|=1 \Leftrightarrow \text{Real}(s)=0 \tag{2.5.2}$$

so the imaginary axis is transformed into the unit circle and the open left half plane into the open unit disk. With a slight abuse of notation we will use ω to denote both the forward and backward transformation, it will be apparent from the context which transformation is considered. On state space level we can represent this transformation as follows (Glover, 1984):

Let G(s) be a FDLT system with realization $[A_c,B_c,C_c,D_c]$ where A_c has no eigenvalues in s=1. Define the discrete time system F(z) by:

$$F(z) := G\left[\omega(s)\right] = G\left[\frac{(1+s)}{(1-s)}\right] \tag{2.5.3a}$$

then F(z) is a FDLT system and has a realization $[A_d,B_d,C_d,D_d]$ and both these realizations satisfy the following equations:

$$A_d = [I+A_c][I-A_c]^{-1} \qquad A_c = [A_d-I][A_d+I]^{-1} \tag{2.5.3a}$$
$$B_d = \sqrt{2}\,[I-A_c]^{-1}B_c \qquad B_c = \sqrt{2}\,[A_d+I]^{-1}B_d \tag{2.5.3b}$$
$$C_d = \sqrt{2}\,C_c[I-A_c]^{-1} \qquad C_c = \sqrt{2}\,C_d[A_d+I]^{-1} \tag{2.5.3c}$$
$$D_d = D_c + C_c[I-A_c]^{-1}B_c \qquad D_c = D_d - C_d[A_d+I]^{-1}B_d \tag{2.5.3d}$$

The condition that A_c should not have poles in s=1, assures the validity of these equations. However if this is the case one can define alternative bilinear transformations (Al–Saggaf, 1986), to avoid problems.

Proposition 2.5.1. [Glover, 1984] The ω–transformation preserves stability and –under the condition that they do exist– the Gramians and Hankel singular values of $[A_c,B_c,C_c,D_c]$ and $[A_d,B_d,C_d,D_d]$ are the same. o

Proof: The preservation of stability follows from (2.5.2): $|z|<1 \Leftrightarrow \text{Real}(s)<0$.
If the systems are stable then we can define the controllability Gramian P (2.3.5a) of $[A_c,B_c,C_c,D_c]$, i.e $\quad 0 = A_cP + PA_c^* + B_cB_c^*$
this implies $\quad 0 = \left[I+A_d\right]\left[A_cP+PA_c^*+B_cB_c^*\right]\left[I+A_d^*\right]$
substitute A_c and B_c as given by (2.5.3)

$$0 = \left[A_d-I\right]P\left[I+A_d^*\right] + \left[I+A_d\right]P\left[A_d^*-I\right] + 2B_dB_d^*$$

which can be written as $\quad 0 = 2\left[A_dPA_d^*+B_dB_d^*-P\right]$

Thus $A_dPA_d^* + B_dB_d^* = P$ and from (2.4.8a) we conclude that P is the controllability Gramian (2.4.5a) of $[A_d,B_d,C_d,D_d]$. Equality of the observability Gramians is proven in a similar way. The invariance of the Hankel singular values is a direct consequence of this. □

The consequence of this proposition is thus that the ω–transformation leaves stability, controllability and observability invariant. We will frequently use this property in the remainder of this thesis. We must be aware that the transformation must be well defined, thus that $[I-A_c]$ and $[I+A_d]$ in (2.5.3) must be non–singular. In almost all cases where this transformation will be applied the realizations will be stable, in which case this condition is satisfied.

2.6 BALANCING TRANSFORMATIONS

In this section we will explain the notion of balanced realizations, which were introduced by Moore (1981). This is a well known concept in the literature and we treat it only very briefly, giving the most important definitions and properties. In words one may say that a balanced realization of a system has the property that the amount of controllability of a certain component of the state vector is equal to the amount of observability of this component. As shown by for instance Enns (1984) and Glover (1984) we can more or less consider the Gramians of a system as a tool to measure the controllability and observability of a realization. This is used in the balanced realization approach. This concept will be of central importance in chapter three on model reduction. We begin with a formal definition.

Definition 2.6.1. Let G(p) be a FDLTS system and [A,B,C,D] be a stable realization of G with controllability Gramian P and observability Gramian Q. If

$$P = Q = \Sigma \tag{2.6.1}$$

where Σ is a positive semi–definite diagonal matrix, then we call [A,B,C,D] a realization that is *(internally) balanced* w.r.t Σ. o

This definition seems to leave a lot of freedom, but the next proposition shows that this freedom is in fact very limited. It is a direct consequence of definitions 2.3.4 and 2.4.2.

Proposition 2.6.2. [Moore, 1981; Glover, 1984] Let A,B,C,D,P,Q,Σ be as in definition 2.6.1. Then the non–zero elements of Σ are equal to the Hankel singular values of G. o

Note that we restricted G in definition 2.6.1 to be stable, which assures that the Hankel singular values are well defined. The following proposition states that it is always possible to find a balanced realization of a stable system.

Proposition 2.6.3. [Laub, 1980; Glover, 1984; Enns, 1984] Let G(p) be a FDLTS system and let [A,B,C,D] be a stable realization of G , $A \in \mathbb{F}^{n \times n}$, $n \geq N$, where N is the McMillan

degree of G. Let $\Sigma \in \mathbb{R}^{nxn}$ be a diagonal matrix with on the diagonal the N Hankel singular values of G and (n–N) zeros. Then there always exists a non–singular transformation matrix $T \in \mathbb{F}^{nxn}$, such that $[TAT^{-1},TB,CT^{-1},D]$ is balanced w.r.t. Σ. o

Remark 2.6.4. In general there exists an infinite number of balanced realizations, since there are no restrictions on the dimension n, nor on the ordering of the elements of Σ. However it is the convention to fix n to N , hence $\Sigma>0$ and the realization is minimal. Furthermore we use a decreasing ordering on the diagonal elements of Σ. In the sequel we will use this convention. o

Convention 2.6.5. Unless otherwise stated we will assume that if [A,B,C,D] is a balanced realization w.r.t $\Sigma = \mathrm{diag}(\sigma_1,\sigma_2,\cdots,\sigma_n)$, that

1. [A,B,C,D] is a minimal realization
2. $\sigma_i \geq \sigma_{i+1} > 0 \quad i=1,n-1$ o

Corollary 2.6.6. [Moore, 1981; Glover, 1984]. If G(p) is a FDLTS system and [A,B,C,D] is a balanced realization w.r.t. Σ of G, then the following Lyapunov equations are valid:

	Controllability:	Observability:	
Discrete :	$A\Sigma A^*+BB^*=\Sigma$	$A^*\Sigma A+C^*C=\Sigma$	(2.6.2a)
Continuous:	$A\Sigma+\Sigma A^*+BB^*=0$	$A^*\Sigma+\Sigma A+C^*C=0$	(2.6.2b)

o

An interpretation of balanced realizations can be given as follows:
The controllability Lyapunov equation shows that, given a balanced realization, the amount of energy to reach a state $x_o=x(\infty)$ from $x(0)=0$ is equal to $x_o^*\Sigma^{-1}x_o$. Thus if σ_i is very small it will take a large amount of energy to reach a state x_o, with $x_o=e_i$, so such states are almost unreachable. The observability equation shows that, with $u(t)\equiv 0$ $t\geq 0$, the amount of energy from the output on $[0,\infty)$ is given by $x(0)^*\Sigma x(0)$. Hence initial states $x(0)=e_i$, with small σ_i make a small contribution to the output and therefore they are almost unobservable. The equality of the energies leads to the term 'balancing'. These realizations were introduced by Moore (1981) in the context of model reduction and later have been proved to be of major importance in various applications. Furthermore it has been shown (Gray and Verriest, 1987; Mullis and Roberts, 1976; Prabhakara, 1989) that these realizations are numerically superior to others, minimizing the roundoff errors in simulation and minimizing parameter sensitivity.

Closely connected with the concept of internally balanced realizations are the input balanced and output balanced realizations (Enns, 1984) and in fact they can be derived from a balanced realization through a similarity transformation with a diagonal similarity matrix. We will only give the formal definitions.

DEFINITION 2.6.7. Let G(p) be a FDLTS system and let [A,B,C,D] be a minimal realization of G. Let P and Q be respectively the controllability and observability Gramian of this realization and Σ the diagonal matrix with Hankel singular values in decreasing order. We call [A,B,C,D]

1. an ***input balanced realization*** of G w.r.t. Σ if $P = I$ and $Q = \Sigma^2$ (2.6.3a)
2. an ***output balanced realization*** of G w.r.t Σ if $P = \Sigma^2$ and $Q = I$ (2.6.3b)

o

The interpretation of the Gramians of an input balanced realization shows that the components of the state vector are (equally) perfectly controllable and the energy with which these components contribute to the input/output behavior of the system is shifted to the observation part. If $[A_b,B_b,C_b,D]$ is a balanced realization w.r.t Σ then we can create an input balanced realization if we apply a similarity transformation with similarity matrix $T=\Sigma^{\frac{1}{2}}$, so let $A_i=\Sigma^{-\frac{1}{2}}A_b\Sigma^{\frac{1}{2}}$, $B_i=\Sigma^{-\frac{1}{2}}B_b$, $C_i=C_b\Sigma^{\frac{1}{2}}$ then $[A_i,B_i,C_i,D]$ is an input balanced realization w.r.t Σ. A similar argument applies for output balanced realizations.

2.7 ALL–PASS TRANSFER FUNCTIONS

The notion of all–pass functions, especially of all–pass transfer functions, plays an important role in the field of model reduction and control design. For instance the solution of the Nehari–problem (Adamyan, Arov and Krein, 1968,1978), the optimal Hankel norm approximation (Glover, 1984), inner/outer factorization (Vidyasagar, 1985) and H_∞ control design (Francis, 1987) make extensive use of the properties of these functions. The theory on this subject has been developed mainly in the continuous time time domain, where in (Glover, 1984) a basic theorem is given for state space realizations of continuous time all–pass transfer functions. In this paragraph we will derive the discrete counterpart of this theorem, which will be of central importance for the theory of chapters three and five.

First we define the notion of all–pass transfer functions.

DEFINITION 2.7.1. A continuous time transfer function matrix E(s) with dimensions p×m is called an all–pass transfer function if:

$$\begin{aligned} E(s)\bar{E}^T(-s)&=I \quad p\leq m \\ \bar{E}^T(-s)E(s)&=I \quad p\geq m. \end{aligned} \tag{2.7.1}$$

o

This concept is easily understood for SISO–systems where this condition comes down to the fact that the Bode amplitude of the transfer function is equal to one for all

frequencies. In general all-pass systems have the property that the amount of energy in the output is equal to the amount of energy in the input.

Glover (1984) gives the following theorem for state space realizations of square all-pass functions, showing that the Hankel singular values are equal to unity, without making assumptions on stability. In the unstable case the Gramians of a realization are not defined but the Lyapunov equations can have solutions.

THEOREM 2.7.2. [Glover, 1984] Given a realization [A,B,C,0], (not necessarily stable), $A \in \mathbb{C}^{n\times n}, B \in \mathbb{C}^{n\times m}, C \in \mathbb{C}^{m\times n}$, then

(1) If [A,B,C,0] is controllable and observable the following two statements are equivalent:

a. $\exists\ D \in \mathbb{C}^{m\times m}$ such that $G(s)G^*(-\bar{s}) = \sigma^2 I\ \forall s \in \mathbb{C}$, where $G(s) := D+C[sI-A]^{-1}B$

b. $\exists\ P,Q \in \mathbb{C}^{n\times n}$ such that

(i) $P = P^*,\ Q = Q^*$

(ii) $PQ = \sigma^2 I$

(iii) $AP + PA^* + BB^* = 0$

(iv) $A^*Q + QA + C^*C = 0$

(2) Given that (1)(b) are satisfied then $\exists$ D satisfying

(i) $DD^* = D^*D = \sigma^2 I$

(ii) $D^*C + B^*Q = 0$

(iii) $DB^* + CP = 0$

and any such D will satisfy (1)(a).

(Note that here observability and controllability are not assumed) ○

We refer to Glover (1984) for an interpretation of this result and proceed with the discrete counterpart of this theorem, where we will make the distinction between square and non-square transfer functions.

DEFINITION 2.7.3. A discrete time p×m transfer function matrix E(z) is called

right all-pass if $p \le m$ and $E(z)\bar{E}^T(\frac{1}{z}) = I$ (2.7.2a)

left all-pass if $p \ge m$ and $\bar{E}^T(\frac{1}{z})E(z) = I$ (2.7.2b)

all-pass if $p = m$ and $E(z)\bar{E}^T(\frac{1}{z}) = \bar{E}^T(\frac{1}{z})E(z) = I$ (2.7.2c)

○

Note that in definition 2.7.3 we do not pose any restrictions on the poles of the transfer function E(z). In the literature left (right) all-pass functions, that are restricted to have all poles inside the unit circle are called inner (co-inner) functions. It will follow (remark 2.7.9) that these functions cannot have poles on the unit circle.

We will use the following theorem on these functions. It is the discrete time version of theorem 2.7.2, previously reported in (Heuberger, 1990[a],1990[b]). It shows that for an all–pass function all the Hankel singular values are equal to unity and gives conditions for a state space realization of (square) all–pass functions.

THEOREM 2.7.4. Given a realization [A,B,C,0] of a discrete time system (not necessarily stable), $A \in \mathbb{C}^{n\times n}$,$B \in \mathbb{C}^{n\times m}$,$C \in \mathbb{C}^{m\times n}$, then

(1) If [A,B,C,0] is controllable and observable then the following two statements are equivalent:

a. $\exists\ D \in \mathbb{C}^{m\times m}$ such that $G(z)\bar{G}^T(\frac{1}{z})=\sigma^2 I$, where $G(z):= D + C[zI-A]^{-1}B$.

b. $\exists\ P,Q \in \mathbb{C}^{n\times n}$, such that

(i) $P=P^*$, $Q=Q^*$ (2.7.3a)

(ii) $A^*QA+C^*C=Q$ (2.7.3b)

(iii) $APA^*+BB^*=P$ (2.7.3c)

(iv) $PQ=\sigma^2 I$ (2.7.3d)

(2) Without condition on controllability or observability:

Given that (1)(b) are satisfied then $\exists$ D satisfying

(i) $D^*D+B^*QB=\sigma^2 I$ (2.7.4a)

(ii) $BD^*+APC^*=0$ (2.7.4b)

(iii) $DD^*+CPC^*=\sigma^2 I$ (2.7.4c)

(iv) $C^*D+A^*QB=0$ (2.7.4d)

and any such D satisfies (1)(a) ○

PROOF: See appendix 2A.

REMARK 2.7.5.

1. Note that if the matrix A in theorem 2.7.4 is not stable then P and Q cannot be seen as the Gramians of the realization, since these are only defined for stable realizations. Nevertheless P and Q are unique if A has no eigenvalues on the unit circle. If A has eigenvalues on the unit circle there may be an infinite number of solutions for (1)(b)(ii) and (iii), some of which will satisfy (1)(b)(iv) iff (1)(b)(i) is satisfied.

2. The condition $PQ=\sigma^2 I$ does not imply minimality, if A is not stable. Take for example A=I,B=0,C=0, then $P=Q=I$ but the realization is neither observable nor controllable.

3. An important feature of theorem 2.7.4 is the fact that the system is not required to be stable. The stable version represents in fact the positive real equations. An even more important feature is that the system may have poles in z=0, which is for instance excluded in (Chu, 1988). Many discrete time systems have poles in z=0 (time–delays).

○

Example 2.7.6. A simple example of an all–pass system with poles in z=0 is $G(z) = \frac{1}{z}$. All minimal realizations of G are parametrized by a parameter $x \neq 0$ and are given by $[A,B,C,D]_x = [0,x,x^{-1},0]$ with Gramians $P=|x|^2$ and $Q=|x|^{-2}$. This is clearly an all–pass function: $G(z)G^T(\frac{1}{z}) = \frac{1}{z}\cdot z = 1$. The conditions $C^*D+A^*QB=0$ and $BD^*+APC^* = 0$ are fulfilled since A=D=0 and $DD^*+CPC^*=D^*D+B^*QB=1$ also since D=0 and $CPC^*=x^{-1}xx^*(x^*)^{-1}=1$ and $B^*QB= x^*(xx^*)^{-1}x=1$. o

In general one will not be dealing with square transfer functions and the following theorem shows how the left (right) –all–pass property of a system can be related to the matrices of a state space realization. This results in weaker conditions. The theorem treats only the right all–pass case, but the left all–pass case is derived from this by taking the Hermitian transpose of the system.

Theorem 2.7.7. Let G(z) be a p×m (p<m) FDLT system and let [A,B,C,D] be a minimal realization of G(z). Then G(z) is a right all–pass transfer function, so $G(z)\bar{G}^T(\frac{1}{z})=I_p$ if and only if there exists a matrix $P=P^*$ such that:

$$APA^*+BB^*=P \qquad (2.7.5a)$$

$$CPC^*+DD^*=I \qquad (2.7.5b)$$

$$APC^*+BD^*=0 \qquad (2.7.5c)$$

Proof: See appendix 2B. o

This theorem will be used in the next section in order to derive an algorithm to calculate coprime factorizations for arbitrary transfer functions. Note again that we do not assume stability in theorem 2.7.7; in the stable case theorem 2.7.7 is known as the discrete positive real lemma.

Example 2.7.8. An example of an inner function with a pole in z=0 and an unstable pole is $G(z)= \frac{1}{2}\sqrt{2}\begin{bmatrix} \frac{1}{z} & \frac{2z-1}{z-2} \end{bmatrix}$. A minimal realization [A,B,C,D] of G(z) is given by:

$$A=\begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix} \quad B=\begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix} \quad C=\tfrac{1}{2}\sqrt{2}[1\ 1] \quad D=[0\ \sqrt{2}]$$

and the solution P to the controllability Lyapunov equation (2.7.5a) is $P=\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}$ Now we check the remaining conditions of theorem 2.7.7:

(2.7.5b): $CPC^*+DD^*= \frac{1}{2}[1\ 1]\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} + 2 = -1 + 2 = 1$

(2.7.5c): $APC^*+BD^*= \frac{1}{2}\sqrt{2}\begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix}\begin{bmatrix}0\\ \sqrt{2}\end{bmatrix} = \frac{1}{2}\sqrt{2}\begin{bmatrix}0\\ -6\end{bmatrix}+\begin{bmatrix}0\\ 3\sqrt{2}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$ o

H is a non-singular matrix (this follows from the proof in Appendix 2C). Hence the condition (2.8.5c) chooses a special case of all possible *rcf*'s, which is the normalized one. Note that there is still freedom in the solution to (2.8.5b), because if H is a solution then also HV will be a solution for any unitary matrix V. This is a direct consequence of the fact that if $G=NM^{-1}$, then also $G=(NV)(MV)^{-1}$. In chapter three we will show how the *nrcf* defines a representation of the ℓ_2 behavior of the system G(z) and can be used to obtain a model reduction method based on approximating this ℓ_2 behavior.

It will be clear that this procedure can also be used to obtain a normalized *left* coprime factorization. This is given in the following corollary.

COROLLARY 2.8.4 [Bongers and Heuberger,1990[a],1990[b]] Given a minimal realization [A,B,C,D] of a transfer function matrix G(z). Let $\tilde{M}$ and $\tilde{N}$ be transfer function matrices with realizations

$$\tilde{M}\text{: } [A-KC,K,-LC,L] \qquad \tilde{N}\text{: } [A-KC,-B+KD,-LC,LD] \tag{2.8.6a,b}$$

then $(\tilde{M},\tilde{N})$ is a normalized left coprime factorization of G(z) iff

$$\text{a. } K = \left[APC^* + BD^*\right]\left[I + D^*D + CPC^*\right]^{-1} \tag{2.8.7a}$$

and

$$\text{b. } L^*L = \left[I + D^*D + CPC^*\right]^{-1} \tag{2.8.7b}$$

and

$$\text{c. } 0 = P - APA^* - BB^* + K\left[I + D^*D + CPC^*\right]K^* \tag{2.8.7c}$$

o

ALGORITHM 2.8.5 The procedure to calculate a nrcf (nlcf) is thus:

1. Solve the Riccati equation in Q (P) with (2.8.5c/2.8.7c)
2. Calculate F (K) with (2.8.5a/2.8.7a)
3. Choose H (L) such that (2.8.5b/2.8.7b) holds.

o

REMARK 2.8.8 There is no condition on the eigenvalues of the system matrix in theorem 2.8.3, which is a result of theorem 2.6.4. In Chu (1988) the author uses the condition that A is non singular, to derive coprime factors with an inner numerator; however this condition is superfluous which can be verified by application of theorem 2.7.4. In section 3.6 these factorizations will be used to define a model reduction procedure.

o

EXAMPLE 2.7.6. A simple example of an all-pass system with poles in z=0 is $G(z) = \frac{1}{z}$. All minimal realizations of G are parametrized by a parameter $x \neq 0$ and are given by $[A,B,C,D]_x = [0,x,x^{-1},0]$ with Gramians $P=|x|^2$ and $Q=|x|^{-2}$. This is clearly an all-pass function: $G(z)G^T(\frac{1}{z}) = \frac{1}{z}\cdot z = 1$. The conditions $C^*D+A^*QB=0$ and $BD^*+APC^* = 0$ are fulfilled since A=D=0 and $DD^*+CPC^*=D^*D+B^*QB=1$ also since D=0 and $CPC^*=x^{-1}xx^*(x^*)^{-1}=1$ and $B^*QB= x^*(xx^*)^{-1}x=1$. o

In general one will not be dealing with square transfer functions and the following theorem shows how the left (right) -all-pass property of a system can be related to the matrices of a state space realization. This results in weaker conditions. The theorem treats only the right all-pass case, but the left all-pass case is derived from this by taking the Hermitian transpose of the system.

THEOREM 2.7.7. Let G(z) be a p×m (p<m) FDLT system and let [A,B,C,D] be a minimal realization of G(z). Then G(z) is a right all-pass transfer function, so $G(z)\tilde{G}^T(\frac{1}{z})=I_p$ if and only if there exists a matrix $P=P^*$ such that:

$$APA^*+BB^*=P \qquad (2.7.5a)$$

$$CPC^*+DD^*=I \qquad (2.7.5b)$$

$$APC^*+BD^*=0 \qquad (2.7.5c)$$

PROOF: See appendix 2B. o

This theorem will be used in the next section in order to derive an algorithm to calculate coprime factorizations for arbitrary transfer functions. Note again that we do not assume stability in theorem 2.7.7; in the stable case theorem 2.7.7 is known as the discrete positive real lemma.

EXAMPLE 2.7.8. An example of an inner function with a pole in z=0 and an unstable pole is $G(z)= \frac{1}{2}\sqrt{2}\begin{bmatrix} \frac{1}{z} & \frac{2z-1}{z-2} \end{bmatrix}$. A minimal realization [A,B,C,D] of G(z) is given by:

$$A=\begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix} \quad B=\begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix} \quad C=\tfrac{1}{2}\sqrt{2}[1\ 1] \quad D=[0\ \sqrt{2}]$$

and the solution P to the controllability Lyapunov equation (2.7.5a) is $P=\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}$ Now we check the remaining conditions of theorem 2.7.7:

(2.7.5b): $CPC^*+DD^*= \frac{1}{2}[1\ 1]\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} + 2 = -1 + 2 = 1$

(2.7.5c): $APC^*+BD^*= \frac{1}{2}\sqrt{2}\begin{bmatrix}0 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}1 & 0\\ 0 & -3\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} + \begin{bmatrix}1 & 0\\ 0 & 3\end{bmatrix}\begin{bmatrix}0\\ \sqrt{2}\end{bmatrix} = \frac{1}{2}\sqrt{2}\begin{bmatrix}0\\ -6\end{bmatrix}+\begin{bmatrix}0\\ 3\sqrt{2}\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$ o

Remark 2.7.9. From theorem 2.7.4 and 2.7.7 we can deduct that (left/right) all–pass functions can not have poles on the unit circle. Let [A,B,C,D] be a minimal realization of a (right) all–pass function, then (2.7.3c) and (2.7.5a) show that $\exists\ P=P^*$ with $APA^*+BB^*=P$. Now suppose $\exists\ \lambda,x$ with $|\lambda|=1$ $x\neq 0$ and $xA=\lambda x$, then:

$$xPx^* = xAPA^*x^* + xBB^*x^* = |\lambda|^2xPx^* + xBB^*x^* = xPx^* + xBB^*x^*$$

Thus $xB=0$, but the Popov– Belevitch–Hautus test (Kailath, 1980) shows this is in contradiction with the controllability of (A,B). This is of course not very surprising since a pole on the unit circle, that is not compensated by a zero, would cause the transfer function to explode in that pole and hence the amplitude can never be equal to one in such situations. o

2.8 COPRIME FACTORIZATIONS

Coprime factorizations play an increasingly important role in system theory. In the work of Desoer (1980) and Vidyasagar (1982,1984) the benefits of using coprime factorizations in stability analysis of controlled systems are shown. In Glover and McFarlane (1988,1989) and in McFarlane (1988) the importance of normalized coprime factorization is shown for the H_∞ controller design. In Ober and McFarlane (1988) the normalized coprime factorization is used to derive canonical forms of balanced realizations. Almost all of this work in done in the continuous time domain and in (Nett et.al, 1984; Meyer and Franklin, 1988; Vidyasagar,1988) state space representations for the normalized coprime factors are derived. For the discrete time Chu (1988) gave state space representations for coprime factors with an inner numerator, provided the system has no poles in the origin. This condition, which is severe since in discrete time often time delays are incorporated in the system or model, was removed in (Bongers and Heuberger, 1990[a],1990[b]) In chapter three we will show how coprime factorizations can be used to define model reduction methods with appealing properties. This section deals with the definition and calculation of discrete time coprime factorizations. The results of section 2.7 on left and right all–pass functions will appear to be of central importance.

The following definitions are due to Vidyasagar (1984) and Huang and Liu (1987).

Definition 2.8.1.
Define G as a ring with identity, X a subring of G and F a quotient field of X. o

In our context of linear systems we consider G as the space of matrices with as elements proper rational functions having no poles on the unit circle (discrete time systems) or no poles on the imaginary axis (continuous time systems).

This is often denoted with $G=\mathbb{R}L_\infty$ or $G=\mathbb{C}L_\infty$, where the infinity norm (see 2.10) is used. The subspace X is identified with $\mathbb{R}H_\infty$ or $\mathbb{C}H_\infty$, the functions with all poles in the open left half plane (continuous time) or in the open unit disc (discrete time).
The field F then describes the set of all proper rational transfer function matrices in terms of ratios of stable proper transfer functions.

DEFINITION 2.8.2. A transfer function $G(z)\in F$ has a right (left) fractional representation if there exist N,M $(\tilde{N},\tilde{M}) \in X$ such that:

$$G = NM^{-1} \ (\tilde{M}^{-1}\tilde{N}) \tag{2.8.1}$$

Furthermore we say that the pair N,M $(\tilde{N},\tilde{M})$ is right (left) coprime (***rcf*** or ***lcf***) if there exist U,V $(\tilde{U},\tilde{V}) \in X$ such that:

$$UN + VM = I \ \ (\tilde{N}\tilde{U} + \tilde{M}\tilde{V} = I) \tag{2.8.2}$$

The pair N,M $(\tilde{N},\tilde{M})$ is normalized right (left) coprime (***nrcf*** or ***nlcf***) if in addition to (2.8.2):

$$MM^* + NN^* = I \ \ (\tilde{M}^*\tilde{M} + \tilde{N}^*\tilde{N} = I) \tag{2.8.3}$$

o

Meyer and Franklin (1988) present an explicit method to calculate the normalized coprime factorization in continuous time. Here we will restrict our attention to the discrete time case and in the following theorem we give sufficient conditions for the existence of a *nrcf* of a discrete time plant, which implicitly leads to a constructive algorithm.

THEOREM 2.8.3. [Bongers and Heuberger, 1990[a],1990[b]] Given a minimal realization [A,B,C,D] of a transfer function matrix G(z), let M and N be transfer function matrices with realizations:

$$M: [A-BF,BH,-F,H] \qquad N: [A-BF,BH,C-DF,DH] \tag{2.8.4a,b}$$

then (M,N) is a normalized right coprime factorization of G(z) iff

$$\text{a.} \quad F^* = \left[A^*QB + C^*D\right]\left[I + DD^* + B^*QB\right]^{-1} \tag{2.8.5a}$$

and

$$\text{b.} \quad HH^* = \left[I + DD^* + B^*QB\right]^{-1} \tag{2.8.5b}$$

and

$$\text{c.} \quad 0 = Q - A^*QA - C^*C + F^*\left[I + DD^* + B^*QB\right]F \tag{2.8.5c}$$

PROOF: See appendix 2C. Also in (Bongers and Heuberger, 1990[a],1990[b]). o

This theorem shows that an *nrcf* is in fact obtained by finding a stabilizing state feedback matrix F, that is created with the solution to the Riccati equation (2.8.5c). The matrix Q is then the observability Gramian of the resulting closed loop system, which is the reason why we used the same symbol as before. If we only wish to find a coprime factorization then we may relax the conditions (2.5.8b,c) to the condition that

H is a non-singular matrix (this follows from the proof in Appendix 2C). Hence the condition (2.8.5c) chooses a special case of all possible *rcf*'s, which is the normalized one. Note that there is still freedom in the solution to (2.8.5b), because if H is a solution then also HV will be a solution for any unitary matrix V. This is a direct consequence of the fact that if $G=NM^{-1}$, then also $G=(NV)(MV)^{-1}$. In chapter three we will show how the *nrcf* defines a representation of the ℓ_2 behavior of the system G(z) and can be used to obtain a model reduction method based on approximating this ℓ_2 behavior.

It will be clear that this procedure can also be used to obtain a normalized *left* coprime factorization. This is given in the following corollary.

Corollary 2.8.4 [Bongers and Heuberger,1990a,1990b] Given a minimal realization [A,B,C,D] of a transfer function matrix G(z). Let $\tilde{M}$ and $\tilde{N}$ be transfer function matrices with realizations

$$\tilde{M}\text{: } [A-KC,K,-LC,L] \qquad \tilde{N}\text{: } [A-KC,-B+KD,-LC,LD] \tag{2.8.6a,b}$$

then $(\tilde{M},\tilde{N})$ is a normalized left coprime factorization of G(z) iff

$$\text{a. } K = \left[APC^* + BD^*\right]\left[I + D^*D + CPC^*\right]^{-1} \tag{2.8.7a}$$

and

$$\text{b. } L^*L = \left[I + D^*D + CPC^*\right]^{-1} \tag{2.8.7b}$$

and

$$\text{c. } 0 = P - APA^* - BB^* + K\left[I + D^*D + CPC^*\right]K^* \tag{2.8.7c}$$

o

Algorithm 2.8.5 The procedure to calculate a nrcf (nlcf) is thus:

1. Solve the Riccati equation in Q (P) with (2.8.5c/2.8.7c)
2. Calculate F (K) with (2.8.5a/2.8.7a)
3. Choose H (L) such that (2.8.5b/2.8.7b) holds.

o

Remark 2.8.8 There is no condition on the eigenvalues of the system matrix in theorem 2.8.3, which is a result of theorem 2.6.4. In Chu (1988) the author uses the condition that A is non singular, to derive coprime factors with an inner numerator; however this condition is superfluous which can be verified by application of theorem 2.7.4. In section 3.6 these factorizations will be used to define a model reduction procedure.

o

2.9. KRONECKER CALCULUS

In chapter 5 we will use the concept of Kronecker products and perform calculations with these products. We briefly repeat the main characteristics of this concept, a complete treatment can be found in (Bellman, 1970; Stewart, 1973).

DEFINITION 2.9.1. Let $X \in \mathbb{C}^{n \times m}$, $Z \in \mathbb{C}^{p \times q}$. The Kronecker product $X \otimes Z \in \mathbb{C}^{np \times mq}$ and the Kronecker vector $\mathrm{Vec}(X) \in \mathbb{C}^{nm \times 1}$ are defined by:

$$X \otimes Z = \begin{bmatrix} X_{11}Z & X_{12}Z & \cdots & X_{1m}Z \\ X_{21}Z & \cdots & & \vdots \\ \vdots & & & \vdots \\ X_{n1}Z & \cdots & \cdots & X_{nm}Z \end{bmatrix} \qquad \mathrm{Vec}(X) = \begin{bmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \\ X_{12} \\ \vdots \\ X_{nm} \end{bmatrix}$$

o

We will need this definition to simplify calculations. It is a well known concept in matrix theory and the main properties are summarized in the next proposition.

PROPOSITION 2.9.2. [Bellman, 1970].

1. If $X \in \mathbb{C}^{n \times n}$, $Z \in \mathbb{C}^{m \times m}$ are square matrices and $\{\lambda_i,\ i=1 \cdots n\}$, $\{\mu_j, j=1 \cdots m\}$ are the sets with eigenvalues of X respectively Z, then:
   - the characteristic roots of $X \otimes Z$ are $\{\lambda_i \mu_j,\ i=1 \cdots n,\ j=1 \cdots m\}$, and
   - the characteristic roots of $I_m \otimes X + Z \otimes I_n$ are $\{\lambda_i + \mu_j,\ i=1 \cdots n,\ j=1 \cdots m\}$.
2. $A \otimes (B \otimes C) = (A \otimes B) \otimes C = A \otimes B \otimes C$
   $(A+B) \otimes (C+D) = A \otimes C + B \otimes C + A \otimes D + B \otimes D$
   $(A \otimes B)(C \otimes D) = (AC) \otimes (BD)$
3. $AXB = C \Leftrightarrow (B^T \otimes A)\mathrm{Vec}(X) = \mathrm{Vec}(C)$ o

REMARK 2.9.3. The definition of Vec(X) is here given by stacking the columns of X, which in other textbooks is sometimes done by stacking the rows. In that case proposition 2.8.2(3) is different. We will work with definition 2.8.1. Kronecker products can for instance be used to solve Lyapunov equations, as $AP+PA=C$, where A and C are given and P has to be found. Use proposition 2.8.2(3) to transform this equation into $(I \otimes A + A^T \otimes I)\mathrm{Vec}(P) = \mathrm{Vec}(C)$ which is a linear equation which has a unique solution if the matrix $(I \otimes A + A^T \otimes I)$ is non– singular. Proposition 2.8.2(1) shows that this is the case if A has no eigenvalues λ and μ with $\lambda + \mu = 0$. o

2.10 NORMS AND SPACES

In this paragraph we define the function spaces and norms, which will be used in this thesis. We make a distinction between continuous and discrete time systems and signals. We refrain from going into detail, more complete definitions are given by Hoffman (1962), Rudin (1966) and Francis (1987).

DEFINITION 2.10.1. Let G(s) and F(z) :$\mathbb{C}\to\mathbb{C}^{p\times m}$ be the transfer function of a continuous respectively discrete time FDLT system; let X(t): $\mathbb{R}\to\mathbb{F}^n$, Y(t): $\mathbb{Z}\to\mathbb{F}^n$ be time functions and $M\in\mathbb{F}^{p\times m}$. Let furthermore $a,b\in\mathbb{R}\cup\{-\infty,\infty\}$ and $c\le d\in\mathbb{Z}\cup\{-\infty,\infty\}$ and let $\sigma_{max}(M)$ denote the maximal singular value of the matrix M.

1. $G\in H_\infty$ if G is analytic and bounded in the open right half plane,
   $F\in H_\infty$ if F is analytic and bounded outside the closed unit disk [4]. We denote the H_∞-norm of F and G by:
$$\|G\|_\infty := \sup_{\omega\in\mathbb{R}}\left[\sigma_{max}\left[G(j\omega)\right]\right]$$
$$\|F\|_\infty := \sup_{\omega\in\mathbb{R}}\left[\sigma_{max}\left[F(e^{j\omega})\right]\right]$$

2. Let $G(s) = G_u(s)+G_s(s)$, with $G_s(s)$ the stable part of G(s),
   Let $F(z) = F_u(z)+F_s(z)$, with $F_s(z)$ the stable part of F(z). Let $G_s(s)$ and $F_s(z)$ have Hankel singular values $\{\sigma_i\}$ respectively $\{\rho_i\}$. The **Hankel norm** of G(s) and F is denoted by:
$$\|G(s)\|_H = \|G_s(s)\|_H := \max_i \sigma_i$$
$$\|F(z)\|_H = \|F_s(z)\|_H = \max_i \rho_i$$

3. $X\in L_{2(a,b)}$ if $\int_a^b X^*(t)X(t)\,dt < \infty$ and we denote the $L_{2(a,b)}$ **norm** of X(t) as:
$$\|X(t)\|_2 := \left[\int_a^b X^*(t)X(t)\,dt\right]^{\frac{1}{2}}$$

4. $Y\in \ell_{2(c,d)}$ if $\sum_{t=c}^{d} Y^*(t)Y(t) < \infty$ and we denote the $\ell_{2(c,d)}$ **norm** of Y(t) as:
$$\|Y(t)\|_2 := \left[\sum_{t=c}^{d} Y^*(t)Y(t)\right]^{\frac{1}{2}}$$

5. Let M have singular values $\{\sigma_i\}$ then we denote the **spectral norm** of M as
$$\|M\|_s = \max_i \sigma_i$$
o

[4] For real rational transfer functions the common notation is RH_∞. $G,F\in H_\infty$ is equivalent with stating that G and F are stable and proper.

Note that the Hankel norm of a system is not influenced by any unstable part which is added, which is of course not the case for the infinity–norm. Furthermore we know from 2.3.10 and 2.4.9 that the Hankel singular values of a system depend only on the {A,B,C} matrices of a realization, or similarly only on the strictly proper part of a transfer function. Hence G(p)+D has the same Hankel norm as G(p). This follows also from the fact that the Hankel norm is equal to the largest singular value of the Hankel matrix 2.4.10, on which the D–matrix has no influence.

2.11 DISCUSSION OF CHAPTER 2

In this chapter we defined the notion of dynamical systems, with different representations and related concepts. We used a classical definition of a system and showed how this fits in the behavioral approach of Willems (1986). A distinction was made between two types of systems, depending on the function domain of the signals that define the system, the continuous respectively discrete time systems. We showed that the bilinear ω–transformation can be used to transform a system in one domain to a new system in the other domain, with preservation of the main system properties. This transformations will enable us to expand certain results from one domain to the other.
For a specific class of dynamical systems, the square all–pass systems, explicit conditions were derived for the matrices that constitute a state space realization of the system. A new feature of this result is that no restriction have to made concerning the poles of the system. These representations will be applied in chapter three to approach model reduction of discrete time systems with a Hankel norm criterion and will be a basic tool for the theory of chapter five. We derived an analogous result for non square systems and the application of the latter result yielded explicit conditions and algorithms for the calculation of discrete coprime factorizations. These factorizations will in the next chapter be used to define a model reduction method that can be considered to be both an approximation based on model descriptions and an approximation based on behaviors.

2.12 APPENDICES OF CHAPTER 2

Appendix 2A proof of theorem 2.7.4.

For the proof of theorem 2.7.4 we need 2 standard linear algebra results and a rank condition. They are given in the following 3 lemma's.

LEMMA 2A.1. [Glover, 1984] Given matrices $X \in \mathbb{C}^{n\times m}$, $Y \in \mathbb{C}^{n\times r}$ $r\geq m$ with $XX^*=YY^*$ then $\exists U \in \mathbb{C}^{m\times r}$ such that $Y=XU$ and $UU^*=I$. ○

LEMMA 2A.2. Given matrices $X \in \mathbb{C}^{n\times m}$, $Y^*=Y \in \mathbb{C}^{n\times n}$, $n\geq m$,$\alpha \in \mathbb{R}$, $\alpha>0$. with $X[\alpha I+X^*YX]X^*\geq 0$, then $\alpha I+X^*YX\geq 0$. ○

PROOF: If $X=0$, this is trivial. If X has full column rank m, let $Z=\alpha I+X^*YX$. We get : $XZX^*\geq 0 \Rightarrow (X^*X)^{-1}X^*\cdot XZX^*\cdot X(X^*X)^{-1}\geq 0 \Rightarrow Z\geq 0$.

If $0<\text{rank}(X)<m$, Let $X=USV^*$ be the svd of X with $S=\begin{bmatrix}S_1 & 0\\ 0 & 0\end{bmatrix}$, where $S_1>0$. Denote $\begin{bmatrix}P_1 & P_2\\ P_2^* & P_3\end{bmatrix}:=U^*YU$, then $SU^*YUS=\begin{bmatrix}S_1P_1S_1 & 0\\ 0 & 0\end{bmatrix}$. Substitute $X=USV^*$ in $X[\alpha I+X^*YX]X^*\geq 0$:

$$\Rightarrow USV^*[\alpha I+VSU^*YUSV^*]VSU^*\geq 0 \Rightarrow S[\alpha I+SU^*YUS]S\geq 0 \Rightarrow S_1[\alpha I+S_1P_1S_1]S_1\geq 0$$

$$\Rightarrow \alpha I+S_1P_1S_1\geq 0 \Rightarrow \alpha I+\begin{bmatrix}S_1\\ 0\end{bmatrix}\begin{bmatrix}P_1 & P_2\\ P_2^* & P_3\end{bmatrix}\begin{bmatrix}S_1\\ 0\end{bmatrix}\geq 0 \Rightarrow \alpha I+SU^*YUS\geq 0$$

$$\Rightarrow V^*[\alpha I+SU^*YUS]V\geq 0 \Rightarrow \alpha I+X^*YX\geq 0$$ □

LEMMA 2A.3. If the matrices {A,B,C} obey the conditions of theorem 2.7.4.(1b) then $\text{rank}(BB^*) = \text{rank}(C^*C)$. ○

PROOF: $APA^* + BB^* = P \Rightarrow APA^*Q + BB^*Q = \sigma^2 I$ (since $PQ=\sigma^2 I$)

$A^*QA + C^*C = Q \Rightarrow PA^*QA + PC^*C = \sigma^2 I$ (idem)

If $\lambda\in\sigma(APA^*Q-\sigma^2 I)$, then $(\lambda+\sigma^2)\in\sigma(APA^*Q)$ and thus also $(\lambda+\sigma^2)\in\sigma(PA^*QA)$ which shows that λ is an eigenvalue of $PA^*QA-\sigma^2 I$. Hence $\text{rank}(BB^*Q)=\text{rank}(PC^*C)$ and the regularity of P and Q ($PQ=\sigma^2 I$) shows that $\text{rank}(BB^*) = \text{rank}(C^*C)$ □

PROOF of theorem 2.7.4.

Part 1. Since we have no knowledge of the eigenvalues of A, we will use a variant of the ω-transformation to transform the problem to continuous time.

Let $k\in\mathbb{C}$, with $|k|=1$, such that $(I+kA)$ has full rank. Since A is finite dimensional such a k will always exist. Consider the bilinear transformation:

$$s=k\frac{z-1}{z+1} \quad z=\frac{k+s}{k-s} \quad \text{and let } G_2(s)=G(\tfrac{k+s}{k-s}).$$

define : $A_2=[kA+I]^{-1}[kA-I]$ $\qquad$ $B_2=\sqrt{2k}[kA+I]^{-1}B$;

$C_2=\sqrt{2k}C[kA+I]^{-1}$ $\qquad$ $D_2=D+C[kA+I]^{-1}B$

In All–Sagaff (1986) it is shown that (i) this is a bijective map, (ii) that $\{A_2,B_2,C_2,D_2\}$ is a realization of G_2 and (iii) that minimality of $\{A,B,C\}$ is equivalent with minimality of $\{A_2,B_2,C_2\}$. Further:

$$APA^* + BB^* = P \Leftrightarrow A_2P + PA_2^* + B_2B_2^* = 0$$
$$A^*QA + C^*C = Q \Leftrightarrow A_2^*Q + QA_2 + C_2^*C_2 = 0.$$

It is immediate that G all–pass $\Leftrightarrow$ G_2 all–pass. Now it is easy to see that the proof of (1)(a)$\Leftrightarrow$(1)(b) is a direct result from part 1 of theorem 2.6.2. (part 1)□

Part 2. Postmultiply the Lyapunov equation $APA^*+BB^*=P$ with QA:

$$APA^*QA + BB^*QA = \sigma^2A \Rightarrow AP[Q-C^*C] + BB^*QA = \sigma^2A$$
$$\Rightarrow APC^*C=BB^*QA \quad (2A.1)$$
$$\Rightarrow APC^*CPA^* = BB^*QAPA^* = BB^*Q[P-BB^*] = B[\sigma^2I-B^*QB]B^* \quad (2A.2)$$

Since the left side of (2A.2) is semi positive definite lemma 2A.2 shows that $\sigma^2I-B^*QB\geq 0$. So $\exists\ E \in \mathbb{C}^{m\times m}$, with $EE^*=\sigma^2I-B^*QB$ or $APC^*CPA=BEE^*B$ (2A.3)

Apply lemma 2A.1 on (2A.3) to show $\exists\ U\mathbb{C}^{m\times m}$, $UU^*=I$, with $APC^*=BEU$ (2A.4)

Choose $D=-(EU)^*$. Then we have: $D^*D=\sigma^2I-B^*QB$ (2A.5)

$$BD^*+APC^*=0 \quad (2A.6)$$

Now assume that B and C^* have full column rank. We have shown (2A.1):

$APC^*C=BB^*QA$ or (with 2A.6) $BD^*C=-BB^*QA$ and thus $D^*C=-B^*QA$ (2A.7)

Premultiply (2A.7) with D: $DD^*C= -DB^*QA= CPA^*QA= -CPC^*C+\sigma^2C$

and with the rank condition this implies that $DD^* = -CPC^*+\sigma^2I$ (2A.8)

At this point (2A.5–8) show the validity of (2.7.4) if B and C have full column rank. If B and C do not have full column rank, lemma 2A.3. shows that they still have equal column rank $r<m$. So $\exists W,Z \in \mathbb{C}^{m\times m}$ with

$$\tilde{B}:=BW=[\hat{B}\ 0] \qquad W^*W=ZZ^*=I$$
$$\tilde{C}^*:=C^*Z=[\hat{C}^*\ 0] \qquad \hat{B},\hat{C}\in\mathbb{C}^{n\times r}$$
$$\left[\text{Let } B=U\begin{bmatrix}S\\0\end{bmatrix}V^* \text{ (svd) and take } W=V\right]$$

Note that $BB^*=\hat{B}\hat{B}$ and $C^*C=\hat{C}^*\hat{C}$ and that $\{A,\hat{B},\hat{C},P,Q\}$ fulfill the conditions of theorem 2.7.4 (1b) (with m=r) and therefore we can conclude that $\exists\hat{D}$ with

$$\hat{D}^*\hat{D}=\sigma^2I-\hat{B}^*Q\hat{B} \qquad \hat{D}\hat{D}^*=\sigma^2I-\hat{C}P\hat{C}^*$$
$$\hat{B}\hat{D}^*+AP\hat{C}^*=0. \qquad \hat{C}^*\hat{D}+A^*Q\hat{B}=0$$

Now define $\tilde{D} := \begin{bmatrix}\hat{D} & 0\\ 0 & \sigma I\end{bmatrix}$ then it is immediate that

$$\tilde{D}^*\tilde{D}=\sigma^2I-\tilde{B}^*Q\tilde{B} \qquad \tilde{D}\tilde{D}^*=\sigma^2I-\tilde{C}P\tilde{C}^*$$
$$\tilde{B}\tilde{D}^*+AP\tilde{C}^*=0. \qquad \tilde{C}^*\tilde{D}+A^*Q\tilde{B}=0$$

Finally we take $D := W^*\tilde{D}Z$ which yields:

$$D^*D=\sigma^2I-B^*QB \qquad DD^*=\sigma^2I-CPC^*$$
$$BD^*+APC^*=0. \qquad C^*D+A^*QB=0$$

and these are exactly the relations which we wanted to proof. This shows the validity of the 4 equations (2.7.4) independent of the column rank of B and C.
It remains to proof that such a D satisfies theorem 2.7.4 (1)(a). Therefore we consider the state space realization of:

$$G(z)\bar{G}^T(\tfrac{1}{z}) = \{\ C[zI-A]^{-1}B+D\ \} \times \{B^*[\tfrac{1}{z}I-A^*]^{-1}C^*+D^*\}.$$

Some simple calculus shows that this can be written as:

$$G(z)\bar{G}^T(\tfrac{1}{z}) = -[C \mid DB^*+CPA^*] \begin{bmatrix} zI-A & 0 \\ 0 & \frac{1}{z}I-A^* \end{bmatrix}^{-1} \begin{bmatrix} BD^*+APC^* \\ C^* \end{bmatrix} + DD^* + CPC^*$$

Now we substitute the equations (2.7.4) which we have just proven:

$$G(z)\bar{G}^T(\tfrac{1}{z}) = -[C \mid 0] \begin{bmatrix} zI-A & 0 \\ 0 & \frac{1}{z}I-A^* \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ C^* \end{bmatrix} + \sigma^2 I. \quad \text{So } G(z)\bar{G}^T(\tfrac{1}{z})=\sigma^2 I \qquad \begin{matrix}(\text{part 2})\square \\ (\text{theorem})\square\end{matrix}$$

Appendix 2B proof of theorem 2.7.7.

1. $\Rightarrow$
Since (A,C) is observable there exists a $K \in \mathbb{F}^{n \times p}$ such that (A–KC) is stable, i.e. has all its poles inside the unit circle. Now we define:

$$A_k := A - KC \qquad X(z) := I_p - C[zI-A_k]^{-1}K \qquad (2B.1,2)$$

An easy calculation shows that $X(z)G(z) = D + C[zI-A_k]^{-1}(B-KD)$ (2B.3)
We proceed with an evaluation of $X(z)[I-G(z)G^*(z)]X^*(z)$:

$$0 = X(z)[I-G(z)G^*(z)]X^*(z) = X(z)X^*(z) - X(z)G(z)G^*(z)X^*(z) \qquad (2B.4a)$$

$$= (I-DD^*) - C[zI-A_k]^{-1}\left[K+(B-KD)D^*\right] - \left[K+(B-KD)D^*\right]^*[\tfrac{1}{z}I-A_k^*]C^* \qquad (2B.4b)$$

$$+ C[zI-A_k]^{-1}\left[KK^* - (B-KD)(B-KD)^*\right][\tfrac{1}{z}I-A_k^*]C^* \qquad (2B.4c)$$

The stability of A_k shows there exists a matrix $P=P^*$ that is the unique solution to the following Lyapunov equation: $P = A_kPA_k^* - KK^* + (B-KD)(B-KD)^*$ (2B.5)
We use this equation (2B.5) to evaluate (2B.4c) further:

$$C[zI-A_k]^{-1}\left[KK^* - (B-KD)(B-KD)^*\right][\tfrac{1}{z}I-A_k^*]^{-1}C^*$$

$$= C[zI-A_k]^{-1}\left[A_kPA_k^* - P\right][\tfrac{1}{z}I-A_k^*]^{-1}C^*$$

$$= -CPC^* - C[zI-A_k]^{-1}A_kPC^* - CPA_k^*[\tfrac{1}{z}I-A_k^*]^{-1}C^* \qquad (2B.6)$$

If we substitute (2B.6) in (2B.4) we get:

$$0 = (I-DD^*-CPC^*) - C[zI-A_k]^{-1}\left[K+(B-KD)D^*+A_kPC^*\right]$$
$$- \left[K+(B-KD)D^*+CPA_k^*\right]^*[\tfrac{1}{z}I-A_k^*]C^* \qquad (2B.7)$$

This implies immediately that $(I-DD^*-CPC^*)=0$ (2B.8)
(notice that (2B.7) is of the form $0 = \tilde{D} + \tilde{F}(z) + \tilde{F}^*(z)$ with F a stable strictly proper transfer function).

Since (A,C) is observable we know that (A_k,C) is also observable and hence equation (2B.7) implies also $K+(B-KD)D^*+A_kPC^* = 0$ (2B.9)
Substitution of (2B.1) in (2B.9) yields $K(I-DD^*) + BD^* + APC^* - KCPC^* = 0$
$\Rightarrow K(I-DD^*-CPC^*) + BD^* + APC^* = 0 \overset{2B.8}{\Longrightarrow} APC^* + BD^* = 0$ (2B.10)
Finally we substitute (2B.1) in (2B.5):

$$\begin{aligned} P &= (A-KC)P(A-KC)^* - KK^* + (B-KD)(B-KD)^* \\ &= APA^*-K(CPA^*+DB^*)-(APC^*+BD^*)K^*-K(I-DD^*-CPC^*)K^*+BB^* \end{aligned} \quad (2B.11)$$

and substitution of (2B.10) and (2B.8) in (2B.11) results in

$$P = APA^* + BB^* \quad (2B.12)$$

We conclude that equations (2B.8), (2B.10) and (2B.12) are valid and therefore the ($\Rightarrow$) part of the theorem is proven. ($\Rightarrow$)□

2. $\Leftarrow$
If the equations (2.7.5) are valid then straightforward calculation shows that

$$G(z)G^*(z)= -\begin{bmatrix} C & DB^*+CPA^* \end{bmatrix}\begin{bmatrix} zI-A & 0 \\ 0 & \frac{1}{z}I-A^* \end{bmatrix}^{-1}\begin{bmatrix} BD^*+APC^* \\ C^* \end{bmatrix} + DD^* + CPC^* = I.$$

This completes the proof of theorem 2.7.7 □

Appendix 2C. proof of theorem 2.8.3.

Part 1: $\Leftarrow$
We follow definition 2.8.2 to check that (M,N) is indeed a *nrcf.* This leads to 4 steps we have to proof.

Step 1. $M,N \in X$. Note that (A–BF) is stable since F is the stabilizing solution to the Riccati equation (2.8.5c), and thus we have assured that $M,N \in X$. (step 1)□

Step 2. $G = NM^{-1}$. M^{-1} has a realization $[A,B,FH^{-1},H^{-1}]$ and thus $y=NM^{-1}u$ is given by the state space equations

$$\begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ w_k &= H^{-1}Fx_k + H^{-1}u_k \\ z_{k+1} &= (A-BF)z_k + BHw_k \\ y_k &= (C-DF)z_k + DHw_k \end{aligned} \quad \Rightarrow \quad \begin{aligned} x_{k+1} &= Ax_k + Bu_k \\ z_{k+1} &= (A-BF)z_k + BFx_k + Bu_k \\ y_k &= (C-DF)z_k + DFx_k + Du_k \end{aligned}$$

Hence if we subtract the state space equations we get
$(x_{k+1}-z_{k+1}) = (A-BF)(x_k-z_k)$, which shows that on transfer function level, where $x_o=z_o$ we get that $x_k=z_k$, which results in $x_{k+1} = Ax_k + Bu_k$ and $y_k = Cx_k + Du_k$ and thus NM^{-1} has a realization [A,B,C,D] or $G = NM^{-1}$ (step 2)□

Step 3. $\exists$ U,V $\in X$ with VM + UN = I. Let K be such that (A–KC) is stable, the minimality of [A,B,C,D] assures the existence of such a K. Define (U,V) with the following realizations U: $[A-KC,K,H^{-1}F,0]$ V: $[A-KC,B-KD,H^{-1}F,H^{-1}]$ then we have that U,V $\in X$. Consider [V U] and $\begin{bmatrix}M\\N\end{bmatrix}$ each as one transfer function matrix and evaluate the state space equations of their product $y=[V\ U]\begin{bmatrix}M\\N\end{bmatrix}u$:

$$x_{k+1} = (A-BF)x_k + BHu_k$$

$$w_k = \begin{bmatrix}-F\\C-DF\end{bmatrix}x_k + \begin{bmatrix}H\\DH\end{bmatrix}u_k \qquad x_{k+1} = (A-BF)x_k + BHu_k$$

$$z_{k+1} = (A-KC)z_k + [B-KD\ K]w_k \Rightarrow \qquad z_{k+1} = (A-KC)z_k - (BF-KC)x_k + BHu_k$$

$$y_k = H^{-1}Fz_k + [H^{-1}\ 0]w_k \qquad y_k = H^{-1}Fz_k - H^{-1}Fx_k + u_k$$

Again we set $x_o=z_o$ and get $x_k=z_k$, which leads to $y_k=u_k$, or VM+UN=I. (step 3)□

Step 4. $MM^*+NN^* = I$ ([M N] is inner). Recall from theorem 2.7.7 that [M N] is inner if the observability Gramian $\tilde{Q}$ of it's realization $[\tilde{A},\tilde{B},\tilde{C},\tilde{D}]$ exists and obeys the equations

(2.7.5a): $\tilde{A}^*\tilde{Q}\tilde{A}+\tilde{C}^*\tilde{C}=\tilde{Q}$ (2.7.5b): $\tilde{D}^*\tilde{D}+\tilde{B}^*\tilde{Q}\tilde{B}=I$ (2.7.5c): $\tilde{C}^*\tilde{D}+\tilde{A}^*\tilde{Q}\tilde{B}=0$.

We claim that $\tilde{Q} = Q$, the solution to (2.8.5c).

2.7.5.a: $(A-BF)^*Q(A-BF) + F^*F + (C^*-F^*D^*)(C-DF) = Q \Leftrightarrow$

$A^*QA + C^*C - F^*\left[I + DD^* + B^*QB\right]F = Q$ which is (2.8.5c)

2.7.5b: $[H^*\ H^*D^*]\begin{bmatrix}H\\DH\end{bmatrix} + H^*B^*QBH = I \Leftrightarrow$

$H^*[I + D^*D + B^*QB]H = I$ and this is equivalent with (2.8.5b)

2.7.5c: $[-F^*\ C^*-F^*D^*]\begin{bmatrix}H\\DH\end{bmatrix} + (A-BF)^*QBH = 0 \Leftrightarrow$

$\left[-F^*\left[I + D^*D + B^*QB\right] + C^*D + A^*QB\right]H = 0 \Leftrightarrow$

$F^* = \left[A^*QB + C^*D\right]\left[I + DD^* + B^*QB\right]^{-1}$ which is (2.8.5a) (step 4)□

This concludes the proof of part 1. (part 1)□

Part 2. $\Rightarrow$

We now have to proof that the equations (2.8.5) are valid if (M,N) is a *nrcf* for G(z). Consider therefore [M N] as one system, then this system is inner and has a realization [(A–BF), BH, [–F C–DF], [H DH]]. Step 4 of part 1 now shows that the equations (2.8.5) are valid. (part 2)□

(theorem 2.8.3)□

3. MODEL REDUCTION

3.1 INTRODUCTION

In this chapter we will focus on the problem of model reduction, which deals with the creation of reduced order mathematical models of given large order models. When mathematical models are created for various purposes, as mentioned in chapter one, then it is often important that these models are of a low complexity; this may be necessary to reduce the computing time for predictions or to assure that control design methods lead to low order controllers, in order to to be able to implement the controller. There are also identification methods which lead to models of a high McMillan degree (Wahlberg, 1984), such as the Finite Impulse Response estimation (Ljung, 1987), while the underlying system is of an essentially lower order and model reduction will be an essential part of the identification method. One can also consider a model reduction method as an identification method, by creating from measurement data an exact model explaining these data completely and reducing this high order model to a low order approximation.
These are some of the reasons why model reduction has since long played an important role in system and control theory with a large growth in the past two decades. As indicated in chapter one we consider the problem of model reduction as an intrinsic element of the problem of system approximation and we believe that the latter problem can only be solved with a thorough understanding of approximation on the basis of model descriptions as well as on the basis of behavior or measurement data. However, the criteria that are used in both fields are hardly compatible. For this reason we aim in this chapter at a better understanding of the structure and character of a number of promising model reduction techniques. Especially the work on Hankel norm approximations and reduction techniques based on balanced realizations have proven to be not only model reduction methods with favorable properties but also to be of major importance in various other areas.
For the solution of the Hankel norm approximation problem the work of Adamyan, Arov and Krein (1968,1971,1978) has been the key that led to several elegant and numeric efficient algorithms, like the algorithms of Glover (1984) and Savonov et.al. (1987,1990). These algorithms are almost all developed for continuous time systems, in which time domain the solution of the problem seems to be much easier than in the discrete time domain. The earlier work on Hankel norm approximations was merely done in discrete time, (Kung and Lin, 1981[a], 1981[b]), but this resulted in complicated algorithms, which

could only solve the problem for a selective class of systems. However, the discrete time problem has been solved by Ball and Ran (1987[a],1987[b]) but the algorithms still are very complicated and require much more calculations than the continuous time analog. This is the reason why in all discrete time applications the latter is used in combination with the ω–transformation (2.6.1). The first efficient algorithm was given by Glover (1984) and makes an extensive use of balanced realizations and properties of all–pass functions as described in section 2.7. In section 3.6 we will show that the same line of reasoning for the discrete time case, using the properties of discrete time all–pass functions, does lead to a number of (new) sufficient equations, that are however hard to solve.

Balanced realizations where introduced by Moore (1981) in the context of model reduction, which led to the so called balanced truncation method (section 3.2), and its discrete time analog (section 3.3). Other reduction techniques based on balanced realizations have been introduced by various authors, (Fernando and Nicholson, 1982, 1983; Al–Saggaf, 1986; Al–Saggaf and Franklin, 1986,1987,1988; Liu and Anderson, 1989), some of which will be treated in sections 3.2 and 3.3. In section 3.4 we will show that these methods and algorithms can be unified and are special cases of a new general model reduction method that leads to a one parameter family of reduced order models. The main result of this chapter is theorem 3.4.5, which deals with the conditions under which this generalized reduction method leads to stable and minimal reduced order models. Furthermore this theorem establishes an bound for the error in the frequency domain, i.e. the L_∞ norm of the difference transfer function. Since the development of H_∞ control theory (Francis, 1987, Doyle et.al., 1989) this norm has become an important issue in system and control theory and we will show that the presented general model reduction procedure can improve this error considerably if compared with the standard methods. We will thus consider the frequency behavior of the different reduced order model as a means to compare the reduction methods.

A third approach towards model reduction combines the new general balanced reduction with the concept of coprime factorization, as an extension of the method presented by Meyer (1988). This approach has two major advantages, first it can be applied to both stable and unstable systems, secondly it can be interpreted as an approximation of the behavior of a system in terms of the possible input/output trajectories.

The chapter is organized as follows: in sections 3.2 and 3.3 the standard methods based on balanced realizations are reviewed, respectively for continuous and discrete time systems; section 3.4 deals with the general procedure, that unifies these standard methods and in section 3.5 we will show by means of three examples what this new method can achieve in terms of the frequency response. In section 3.6 we first review the Hankel norm algorithm of Glover (1984) and continue with application of his line of reasoning to the discrete time case. In section 3.7 we combine the general model reduction method, that is presented in section 3.4, with the concept of coprime

factorizations (section 2.8) and derive a so called general fractional model reduction method. We conclude with an evaluation in section 3.8.

We will frequently use a partitioning of the matrices that constitute a state space realization of a transfer function matrix. If [A,B,C,D] is a realization of a transfer function G(p) with $A\in\mathbb{F}^{n\times n}$ and $0<k<n$ then we use the following partitioning:

$$A=\begin{bmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{bmatrix} \quad B=\begin{bmatrix}B_1\\ B_2\end{bmatrix} \quad C=[C_1\ C_2] \tag{3.1.1a}$$

where $A_{11}\in\mathbb{F}^{k\times k}$, $B_1\in\mathbb{F}^{k\times m}$, $C_1\in\mathbb{F}^{p\times k}$. If this realization is balanced w.r.t. a matrix Σ , i.e $\Sigma>0$ and diagonal, then we also partition

$$\Sigma = \begin{bmatrix}\Sigma_1 & 0\\ 0 & \Sigma_2\end{bmatrix} \quad \text{where } \Sigma_1\in\mathbb{R}^{k\times k} \tag{3.1.1b}$$

Remark 3.1.1. Note that by convention 2.6.5 we restricted balanced realizations to be minimal realizations and we imposed a decreasing order on the diagonal elements of Σ: $\Sigma=\mathrm{diag}(\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_n)$ with $\sigma_n>0$. o

3.2. CONTINUOUS TIME BALANCED MODEL REDUCTION

In this section we consider two standard model reduction methods for continuous time systems, the so called balanced truncation and the singular perturbational reduction for balanced models. The section is only meant to give a short review of these methods and is therefore restricted to the main ideas and properties of these approaches. These methods will turn out to be special cases of the general method which is presented in section 3.4.

Based on the concept of balancing (section 2.6), Moore (1981) proposed a model reduction method for continuous time systems, which eliminates the states that are weakly observable and controllable. The singular values of the system provide a measure for determining how observable and controllable a certain state is (see section 2.6), resulting in neglecting the states that correspond to the smallest singular values. This results in the following model reduction procedure.

Definition 3.2.1. Let G(s) be a FDLTS system and let [A,B,C,D] be a balanced realization of G(s) w.r.t $\Sigma>0$, partitioned according to 3.1.1. Then $\hat{G}(s)=\mathcal{CB}_k(G)$, the *Continuous Balanced reduced model of order k*, is defined as

$$\hat{G}(s) := D + C_1[sI-A_{11}]^{-1}B_1 \tag{3.2.1}$$

o

The rationale behind this procedure is to replace σ_i by 0, i=k+1,$\cdots$,n, and to retain the resulting system; in other words, the states that correspond with these singular values are replaced with states that are neither controllable nor observable and can therefore be removed. This will generally lead to satisfactory results if the discarded singular values are relatively small. The next proposition gives the condition to retain stability and minimality.

Proposition 3.2.2. [Moore, 1981; Pernebo and Silverman, 1983]. If in definition 3.3.1 $\sigma_k > \sigma_{k+1}$, then $[A_{11},B_1,C_1,D]$ is balanced w.r.t. Σ_1 and is a stable, minimal realization. o

The restriction $\sigma_k > \sigma_{k+1}$ can even be removed if a special canonical form is used for the realization (Ober, 1987), but we will not pursue this. An obvious question concerns the performance of this reduced order model or if this reduction method might even be optimal with respect to a certain measure. Unfortunately there is no such measure known. However, there exists a bound for the L_∞ norm of the error.

Proposition 3.2.3. [Glover 1984; Enns 1984]. Let G(s) be a FDLTS system with McMillan degree n and Hankel singular values $\{\sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} \geq \cdots \geq \sigma_n\}$. Let $\mathcal{CB}_k(G(s))$ be defined by definition 3.2.1, then the error of the approximation in the L_∞ norm is bounded according to

$$\|G(s)-\mathcal{CB}_k(G(s))\|_\infty \leq 2\cdot(\sigma_{k+1}+\cdots+\sigma_n) \qquad (3.2.2)$$

and for k=n–1, this bound is tight. o

From the examples in section 3.5 it will follow that this error bound can be quite conservative if k<n–1. The reason for this is that the proof is based on the fact that $\|G(s)-\mathcal{CB}_{n-1}(G(s))\|_\infty = 2\sigma_n$ and application of the triangle inequality: Denote $G_k := \mathcal{CB}_k(G(s))$ then proposition 3.3.2 shows that $\|G_k-G_{k-1}\|_\infty = 2\sigma_k$. This implies that $\|G-G_k\|_\infty = \|(G-G_{n-1})+(G_{n-1}-G_{n-2})\cdots+(G_{k+1}-G_k)\|_\infty = \leq 2\cdot(\sigma_n+\cdots+\sigma_{k+1})$.

The procedure to calculate this reduced order model is thus:

1. Calculate a balanced realization (Laub, 1980; Sveinsson and Fairman, 1985; Yeh and Yang, 1987; Yang–Yeh, 1989);
2. Decide on the order of the reduced order model from the singular values;
3. Calculate the reduced order model with definition 3.2.1.

The question which criterion should be used in the second step is for instance addressed by Moore (1981), but in practice it will be a trial and error procedure depending on the designers demands and goals.

In general this model reduction method produces satisfactory results, as well in the time domain (compare impulse or step responses) as in the frequency domain (Bode plots). The method is numerically efficient and stable and only if the poles of the original system G(s) are close to the imaginary axis, then the balancing procedure tends to have numerical problems. A favorable feature of the method is the stability and minimality of the approximations.
A problem we only touched upon is the non–uniqueness of the balanced realizations. In (Kabamba ,1985; Ober, 1987; Ober and McFarlane ,1988) canonical forms are derived for balanced realizations and this may be used to get a unique procedure.
Another favorable property of this method is the **consistency**, which means that $\mathcal{CB}_r(\mathcal{CB}_k(G(s)) = \mathcal{CB}_r(G(s))$, if $r \leq k$; in other words, once we have a k^{th} order reduced model, we can use this to construct lower order approximations. This is a situation which will often occur in practical applications, where one is searching the lowest order approximation that will meet the designers qualifications.

Fernando and Nicholson (1982), Al–Saggaf and Franklin (1988) and Liu and Anderson (1989) combined the balanced model reduction method with the method of singular pertur– bational approximations (Kokotovic, 1984), resulting in the following reduction method.

Definition 3.2.4. Let G(s) be a FDLTS system and let [A,B,C,D] be a balanced realization of G w.r.t $\Sigma>0$, with $\sigma_k>\sigma_{k+1}$. Partition A,B,C compatibly as in (3.1.1). We define $\hat{G}(s)=\mathcal{CSB}_k(G)$, the *Continuous Singular perturbational Balanced reduced model of order k*, by:

$$\hat{G}(s) := \hat{D} + \hat{C}[sI-\hat{A}]^{-1}\hat{B} \tag{3.2.3a}$$

where

$$\hat{A} = A_{11}-A_{12}A_{22}^{-1}A_{21} \qquad \hat{B} = B_1 - A_{12}A_{22}^{-1}B_2 \tag{3.2.3b,c}$$

$$\hat{C} = C_1 - C_2A_{22}^{-1}A_{21} \qquad \hat{D} = D - C_2A_{22}^{-1}B_2 \tag{3.2.3d,e}$$

o

The rationale behind this approximation method can in a heuristic way be described as follows (Fernando and Nicholson, 1982):
Let x(t),u(t) and y(t) respectively be the state–, input– and output vector of the realization [A,B,C,D] and let x(t) be partitioned compatibly as $x(t)=\begin{bmatrix}x_1(t)\\x_2(t)\end{bmatrix}$. Assume furthermore that A_{22} is 'fast', i.e. has eigenvalues with large (negative) real parts, while A_{12} and A_{21} are small. The state space equations for the original system can be decomposed as

$$\dot{x}_1(t) = A_{11}x_1(t) + A_{12}x_2(t) + B_1u(t) \tag{3.2.4a}$$

$$\dot{x}_2(t) = A_{21}x_1(t) + A_{22}x_2(t) + B_2u(t) \tag{3.2.4b}$$

$$y(t) = C_1x_1(t) + C_2x_2(t) + Du(t) \tag{3.2.4c}$$

The assumptions justify the approximation $\dot{x}_2(t) = 0$. Then (3.2.4b) results in an algebraic state equation, which can be transformed to:

$$x_2(t) = -A_{22}^{-1}\left[A_{21}x_1(t) + B_2u(t)\right] \tag{3.2.5a}$$

Substitution of this equation in the other two state space equations leads to:

$$\dot{x}_1(t) = \hat{A}x_1(t) + \hat{B}u(t) \tag{3.2.5b}$$

$$y(t) = \hat{C}x_1(t) + \hat{D}u(t) \tag{3.2.5c}$$

with $[\hat{A},\hat{B},\hat{C},\hat{D}]$ defined by (3.2.3).

Note that in a static situation, where $\dot{x}(t)=0$, the equation $\dot{x}_2(t)=0$ is exact and this implies that the static gain of $\hat{G}(z)$ is equal to the static gain of $G(z)$.

The feasibility of this method is shown by the following two propositions, which are the counterparts of propositions 3.2.2 and 3.2.3 for *CB*.

Proposition 3.2.5. [Fernando and Nicholson, 1982; Liu and Anderson, 1989]. The realization (3.2.3) of $\mathcal{CSB}_k(G(s))$ is balanced w.r.t. Σ_1 and is a stable, minimal realization. ○

Proposition 3.2.6. [All–Saggaf and Franklin, 1988; Liu and Anderson, 1989].
The error of the approximation (3.2.3) is bounded in the the L_∞ – norm:

$$\|G(s)-\mathcal{CSB}_k(G(s))\|_\infty \leq 2\cdot(\sigma_{k+1}+\cdots+\sigma_n) \tag{3.2.6}$$

and for k=n–1, this bound is tight. ○

Hence we find exactly the same error bound as for *CB*, but again we note that this bound can be very conservative if k<n–1.

Remark 3.2.7. Note that in definition 3.2.4 we explicitly pose the condition $\sigma_k \neq \sigma_{k+1}$. This is necessary to guarantee the stability of A_{22} and thus the existence of A_{22}^{-1} (Pernebo and Silverman, 1983).

This method replaces the 'fast' dynamical equation with an algebraic one, restricting the static gain of $\mathcal{CSB}_k(G)$ to be equal to the static gain of G. Again this method is consistent, so $\mathcal{CSB}_r(\mathcal{CSB}_k(G(s))) = \mathcal{CSB}_r(G(s))$, if $r \leq k$.

It should be stressed that propositions 3.2.5 and 3.2.6 are valid without any of the assumptions we used to give an heuristic interpretation of this method. This shows that *CSB* will be a good reduction method if the discarded singular values are small. An important difference with *CB* is the better approximation of the low frequency components of the original system, in terms of the exact fit for the frequency s=0, so the approximation has the same static gain as the original system. ○

3.3. DISCRETE TIME BALANCED MODEL REDUCTION

So far we dealt with model reduction of continuous time systems, based on balanced realizations. In this section we deal with the discrete time version, where we make a distinction between the balanced truncation directly in the discrete time domain, as proposed among others by Pernebo and Silverman (1983) and the result of combining *CB* and the ω–transformation, proposed by Al–Saggaf and Franklin (1988), which method consist of transformation of a discrete system to the continuous time with ω, truncation in the continuous time domain (*CB*) and transformation of this result to the discrete time domain. Next to these two methods we consider singular perturbational model reduction of discrete time balanced systems. These three methods will again be special cases of the generalized model reduction method, presented in the next section.

The discrete truncation is performed in exactly the same way as is the case with continuous time systems (*CB*) :

DEFINITION 3.3.1. Let G(z) be a FDLTS system and let [A,B,C,D] be a balanced realization of G(z) w.r.t $\Sigma>0$. Partition A,B,C according to (3.1.1). Then $\hat{G}(s)=\mathcal{TB}_k(G)$, the *Discrete Truncated Balanced reduced model of order k*, is defined by

$$\hat{G}(z) = D + C_1[zI-A_{11}]^{-1}B_1 \qquad (3.3.1)$$

o

Pernebo and Silverman (1983) show that –if $\sigma_k>\sigma_{k+1}$– this approximation is again minimal and stable, but contrary to the continuous time case this does not apply for the other subsystem $[A_{22},B_2,C_2,D]$. Also the given approximation will in general not be balanced, nor have $\{\sigma_1,\cdots,\sigma_k\}$ as its singular values. This is explained with the difference in the Lyapunov equations for continuous time and discrete time systems. Consider the Lyapunov equations of a balanced system in continuous time:

$$A\Sigma+\Sigma A^*+BB^*=0 \qquad \Sigma A+A^*\Sigma+C^*C=0$$

Pre–multiplying these equations with $T = [I_k\ 0]$ and post–multiplying with T^* shows that the reduced model is balanced with respect to Σ_1:

$$A_{11}\Sigma_1+\Sigma_1A_{11}^*+B_1B_1^*=0 \qquad \Sigma_1A_{11}+A_{11}^*\Sigma_1+C_1^*C_1=0$$

However, in the discrete time the Lyapunov equations have triple matrix products

$$A\Sigma A^*+BB^*=\Sigma \qquad A^*\Sigma A+C^*C=\Sigma$$

and application of the same multiplications results in:

$$A_{11}\Sigma_1A_{11}^*+A_{12}\Sigma_2A_{12}^*+B_1B_1^*=\Sigma_1 \qquad A_{11}^*\Sigma_1A_{11}+A_{21}^*\Sigma_2A_{21}+C_1^*C_1=\Sigma_1$$

The Gramians P_1 respectively Q_1 of $[A_{11},B_1,C_1,D]$ obey the equations:

$$A_{11}P_1A_{11}^*+B_1B_1^*=P_1 \qquad A_{11}^*Q_1A_{11}+C_1^*C_1=Q_1$$

so $A_{11}(\Sigma_1-P_1)A_{11}^*+A_{12}\Sigma_2A_{12}^*=(\Sigma_1-P_1)$ $\quad A_{11}^*(\Sigma_1-Q_1)A_{11}+A_{21}^*\Sigma_2A_{21}=(\Sigma_1-Q_1)$

which shows that in general ($A_{12}\neq 0\neq A_{21}$) we may conclude that $\Sigma_1 \geq P_1$ and $\Sigma_1 \geq Q_1$. This explains that the realization $[A_{11},B_1,C_1,D]$ is in general not balanced w.r.t. Σ_1 and that the Hankel singular values of this realization are different. Nevertheless an even stronger bound for the L∞–norm of the approximation error holds true.

Proposition 3.3.2. [Pernebo and Silverman, 1983; Al–Saggaf and Franklin, 1986].
1. If in definition 3.3.1 $\sigma_k > \sigma_{k+1}$ then $[A_{11},B_1,C_1,D]$ is a minimal stable realization.
2. The error of the approximation $\mathcal{MB}_k(G)$ is bounded in the L∞–norm:

$$\| G(z) - \mathcal{MB}_k(G(z)) \|_\infty \leq 2\cdot(\sigma_{k+1}+\cdots+\sigma_n) \qquad (3.3.2)$$

with strict inequality if $\sigma_k \neq \sigma_n$. ○

Notice that the error bound (3.3.2) is essentially different from the error bounds, which were derived for *CB* and *CSB*: the only situation under which equality in (3.3.2) is possible is if $\sigma_k=\sigma_{k+1}=\cdots=\sigma_n$. Thus this proposition implies that we have a strict bound (<) in (3.3.2) if $\sigma_n<\sigma_{n-1}$, while propositions 3.2.3 and 3.2.6 shows that in this case the bounds for *CB* and *CSB* are exact (=).

Al–Saggaf and Franklin (1987) propose a method that is slightly different from $\mathcal{MB}$, but which has the same properties as the continuous time method *CB*, by applying the ω–transformation, defined in (2.4.1). We know that under this transformation Gramians are invariant (Glover, 1984), which shows that the transformation of a continuous realization, which is balanced w.r.t Σ, is a discrete realization, balanced with respect to Σ. Since we implicitly assume stability we are assured that the ω–transformation is well defined. The reduction method mentioned above consists of the following steps:

Procedure 3.3.3. Let G(s) be a FDLTS system.
1. Create a balanced realization $[A_d,B_d,C_d,D_d]$ of G(z);
2. Transform this realization with the ω–transformation to $[A_c,B_c,C_c,D_c]$, which is a balanced realization of $G_c(s):=\omega^{-1}(G(z))$, using (2.5.3);
3. Calculate $[\hat{A}_c,\hat{B}_c,\hat{C}_c,\hat{D}_c]$, the realization of $CB(G_c(s))$, as defined by definition 3.2.1;
4. Transform the realization $[\hat{A}_c,\hat{B}_c,\hat{C}_c,\hat{D}_c]$ with the ω–transformation to a realization $[\hat{A}_d,\hat{B}_d,\hat{C}_d,\hat{D}_d]$ of a discrete time system $\hat{G}(z)$. ○

Hence the approximation $\hat{G}(z)$ can be described as $\hat{G}(z)=\omega(CB(\omega^{-1}(G(z))))$. Figure 3.3.1 shows the scheme of procedure 3.3.3:

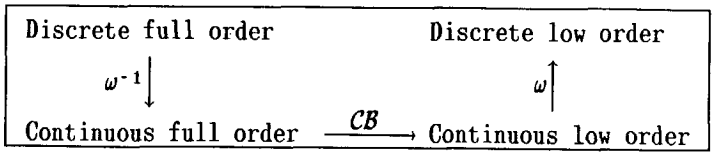


Figure 3.3.1 Discrete time model reduction, based on the ω–transformation and $\mathcal{CB}$

Clearly this procedure guarantees that the properties of $\mathcal{CB}$ are valid for this approach, so the approximation is minimal, stable and balanced w.r.t Σ_1, if $\sigma_k > \sigma_{k+1}$. Moreover, the method is consistent. In actually making the calculations that are required for this procedure we do not have to go through all these steps separately. The following proposition shows how the calculation can be done without actually using the ω–transformation.

Proposition 3.3.4. [Al–Saggaf and Franklin, 1987]. Let G(z) be a FDLTS system with realization [A,B,C,D], partitioned as in (3.1.1) and balanced w.r.t. $\Sigma>0$, with $\sigma_k > \sigma_{k+1}$. Further, let $[\hat{A},\hat{B},\hat{C},\hat{D}]$ be the k^{th} order realization, that results from applying procedure 3.3.3. Then:

$$\hat{A} = A_{11} - A_{12}[I+A_{22}]^{-1}A_{21} \qquad \hat{B} = B_1 - A_{12}[I+A_{22}]^{-1}B_2 \qquad (3.3.3a,b)$$

$$\hat{C} = C_1 - C_2[I+A_{22}]^{-1}A_{21} \qquad \hat{D} = D - C_2[I+A_{22}]^{-1}B_2 \qquad (3.3.3c,d)$$

o

We will refer to this procedure as discrete balanced model reduction:

Definition 3.3.5. Let G(z) be a FDLTS system and let [A,B,C,D] be a realization of G(z) that is balanced w.r.t $\Sigma>0$, with $\sigma_k > \sigma_{k+1}$. Let this realization be partitioned according to (3.1.1). Then $\hat{G}(s)=\mathcal{DB}_k(G)$, the *Discrete Balanced reduced model of order k*, is defined as:

$$\hat{G}(z) = \hat{D} + \hat{C}[zI-\hat{A}]^{-1}\hat{B} \qquad (3.3.4)$$

with $[\hat{A},\hat{B},\hat{C},\hat{D}]$ defined by (3.3.3). o

Remark 3.3.6. The fact that $[\hat{A},\hat{B},\hat{C},\hat{D}]$ is again balanced w.r.t. Σ_1, can be shown in a similar way as for the continuous time case, by multiplication of the different terms of the discrete time Lyapunov equation of a balanced realization with a specific matrix T. Consider the controllability equation:

$$A\Sigma A^* + BB^* = \Sigma \quad \text{and the matrix} \quad T = \left[I_k \mid -A_{12}[I+A_{22}]^{-1}\right]$$

It can be readily verified that pre– respectively post multiplication with T and T^* leads to

$$\hat{A}\Sigma_1\hat{A}^* + \hat{B}\hat{B}^* = \Sigma_1$$

The same argument can be used for the observability Lyapunov equation if we use $T = \left[I_k \mid -A_{21}^*[I+A_{22}^*]^{-1}\right]$ and it follows that $\hat{A}^*\Sigma_1\hat{A} + \hat{C}^*\hat{C} = \Sigma_1$ o

The discrete analog of ***CSB*** has been reported by Fernando and Nicholson (1983), however without adaptation of the feedthrough matrix, which was added in (Al–Saggaf and Franklin, 1988; Liu and Anderson, 1989). It is in fact the result of the previous procedure 3.3.3 if ***CB*** is replaced by ***CSB***. It is again a combination of balancing and singular perturbational model reduction.

DEFINITION 3.3.7. Let G(z) be a FDLTS system and let [A,B,C,D] be a balanced realization of G(z) w.r.t. $\Sigma>0$, with $\sigma_k>\sigma_{k+1}$, partitioned according to (3.1.1). We define $\hat{G}(z)=\mathcal{DSB}_k(G)$, the *Discrete Singular perturbational Balanced reduced model of order k*, by:

$$\hat{G}(z) := \hat{D} + \hat{C}[zI-\hat{A}]^{-1}\hat{B} \tag{3.3.5a}$$

where,
$$\hat{A} = A_{11}+A_{12}[I-A_{22}]^{-1}A_{21} \qquad \hat{B} = B_1 + A_{12}[I-A_{22}]^{-1}B_2 \tag{3.3.5b,c}$$
$$\hat{C} = C_1 + C_2[I-A_{22}]^{-1}A_{21} \qquad \hat{D} = D + C_2[I-A_{22}]^{-1}B_2. \tag{3.3.5d,e}$$

o

All the properties of ***CSB*** carry over to $\mathcal{DSB}$ as stated in the following corollary.

COROLLARY 3.3.8. [Al–Saggaf and Franklin, 1987]. Let G(z) be a FDLTS system and let [A,B,C,D] be a balanced realization of G(z) w.r.t. $\Sigma>0$, with $\sigma_k>\sigma_{k+1}$, partitioned according to (3.3.1). Let $\hat{G}(z)=\mathcal{DB}_k(G)$ or $\hat{G}(z)=\mathcal{DSB}_k(G)$, with a realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ of $\hat{G}(z)$ given by (3.3.3) respectively by (3.3.5).

1. The realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ is stable, minimal and balanced w.r.t. Σ_1.
2. The approximation error is bounded in the L_∞–norm by

$$\|G(z)-\hat{G}(z)\|_\infty \le 2\cdot(\sigma_{k+1}+\cdots+\sigma_n) \tag{3.3.6}$$

and if k=n–1 the bound is achieved. o

Liu and Anderson (1989) present a procedure that consist in using combinations of the standard methods and they show that this can improve the performance of the reduced order models in the frequency domain and decrease the DC–errors (static gain). Such a combination consists of two or more steps, for instance using ***CB*** to reduce from order n to k_1 and ***CSB*** to reduce further to order $k_2(<k_1)$. The choice of an optimal combination is the result of a trial and error procedure. In the next section we present a more generalized method, with which we can improve the frequency behavior considerably, if compared with the results of Liu and Anderson (1989). This method is accomplished in only one reduction step, but does also require a trial and error procedure.

3.4. A FAMILY OF MODEL REDUCTION METHODS BASED ON BALANCED REALIZATIONS

In this section we combine the results of the previous two sections and present a new generalized model reduction method, that has the five methods –*CB*, *CSB*, *DB*, *DTB* and *DSB*– as special cases. Therefore a so called reduction parameter α is introduced which induces a family of reduced order models in the sense that different values of α lead to different models. This parameter can thus be considered as a design variable and specific choices of α lead to the standard methods. We commence with the rationale of the method, after which we will present a formal definition.

The idea behind this framework is due to, among others, Santiago and Jamshidi (1986) and is based on a general partitioning of a transfer function matrix. Let G(p) be a FDLT system where we use no specific time domain, implying that we consider both the continuous (p=s) and the discrete (p=z) time case. Let [A,B,C,D] be a realization of G(p), $G(p)=D+C[pI-A]^{-1}B$ with $A\in\mathbb{F}^{n\times n}$. Let $0<k<n$ and let A,B,C be partitioned according to (3.1.1). We can rewrite G(p) in the following decomposition:

$$G(p) = \tilde{D}(p) + \tilde{C}(p)\,[pI-\tilde{A}(p)]^{-1}\tilde{B}(p) \qquad (3.4.1a)$$

with $\tilde{A}(p):=A_{11}+A_{12}[pI-A_{22}]^{-1}A_{21}$ $\quad \tilde{B}(p):= B_1 + A_{12}[pI - A_{22}]^{-1}B_2$ (3.4.1b,c)

$\tilde{C}(p):=C_1 + C_2[pI-A_{22}]^{-1}A_{21}$ $\quad \tilde{D}(p):= D + C_2[pI - A_{22}]^{-1}B_2$ (3.4.1d,e)

All the model reduction methods that we considered in the previous sections can in fact directly be obtained from the description (3.4.1) if we approximate $[\tilde{A}(p),\tilde{B}(p),\tilde{C}(p),\tilde{D}(p)]$ by $[\tilde{A}(p^o),\tilde{B}(p^o),\tilde{C}(p^o),\tilde{D}(p^o)]$ with p^o a fixed scalar constant. A choice of $p^o=\infty$ and p=s leads to *CB*; $p^o=1$ and p=z results in *DSB* etc. The approach presented here is to define the family of reduced order models by letting p^o vary over $\mathbb{R}$ and to find the restrictions that have to be satisfied in order to guarantee stable and minimal reduced order models. Note that from the decomposition (3.4.1) one would expect that p^o should be chosen on the imaginary axis (p=s) or on the unit circle (p=z). In general such a choice will lead to complex valued reduced order systems with complex valued outputs even if the original system and the inputs are real valued. However, we will show that is does make sense to choose p^o real.

Santiago and Jamshidi (1986) use real values of p^o to define a model reduction method for systems with unstable poles, which in continuous time comes down to:

find a p^o such that $A-p^oI$ is stable
use *CB* on $[A-p^oI,B,C,D]$
shift the resulting $\hat{A}$ back to $\hat{A}+p^oI$.

It will be clear that the result depends highly on the choice of p^o, and can in general

change the number of unstable poles, which in applications as control design is unfavorable. They also implicate that different values of p^o might lead to better results for systems with more than one time scale.

If we let p^o vary over $\mathbb{R}$ then (3.4.1) defines a one parameter family of reduced order models. In this section we will derive conditions on this parameter under which stability and minimality of the reduced order models are guaranteed. Furthermore, we shall give a bound for the L_∞ norm of the error transfer functions.
In the next definition we formalize the reduction method that we propose.

Definition 3.4.1. Let G(p) be a FDLTS system and let [A,B,C,D] be a realization of G(p), that is balanced w.r.t. $\Sigma>0$, with $\sigma_k>\sigma_{k+1}$. Partition A,B,C,Σ according to (3.1.1) and let $\alpha\in\mathbb{R}\cup\{\pm\infty\}$ such that $\alpha\notin\sigma(A_{22})$, were $\sigma()$ denotes the spectrum of a matrix. We define $\hat{G}(p)=\mathcal{GB}_k^\alpha(G)$, the *General Balanced reduced model with order k and reduction-parameter* α, by:

$$\hat{G}(p) = \hat{D} + \hat{C}[pI-\hat{A}]^{-1}\hat{B} \qquad (3.4.2a)$$

where
$$\hat{A}=A_{11}+A_{12}[\alpha I-A_{22}]^{-1}A_{21} \qquad \hat{B} = B_1 + A_{12}[\alpha I-A_{22}]^{-1}B_2 \qquad (3.4.2b,c)$$
$$\hat{C}=C_1 + C_2[\alpha I-A_{22}]^{-1}A_{21} \qquad \hat{D} = D + C_2[\alpha I-A_{22}]^{-1}B_2 \qquad (3.4.2d,e)$$

and where $[\alpha I-A_{22}]^{-1}:=0$ for $\alpha=\pm\infty$ ○

As stated before,we did not specify the time domain, writing G(p) where p can be both p=s or p=z. The condition $\alpha\notin\sigma(A_{22})$ is necessary to assure the existence of $[\alpha I-A_{22}]^{-1}$. The following proposition shows how definition 3.4.1 covers the model reduction methods from the previous sections.

Proposition 3.4.2. Let G(p) be a FDLTS system. Then:

p=s: $\mathcal{CB}_k=\mathcal{GB}_k^\infty$, $\mathcal{CSB}_k=\mathcal{GB}_k^0$ p=z: $\mathcal{DB}_k=\mathcal{GB}_k^{-1}$, $\mathcal{DSB}_k=\mathcal{GB}_k^1$, $\mathcal{DTB}_k=\mathcal{GB}_k^\infty$ ○

Proof: Substitute the values of α in definition 3.4.1 and compare the results with the definitions of the 'standard' model reduction methods. □

The next lemma shows the effect of the ω-transformation on the different reduction methods.

Lemma 3.4.3. Let G(p) be a FDLTS system.

If p=s and $G_d(z)=\omega(G(s))$ then $\omega(\mathcal{GB}_k^\alpha(G)) = \mathcal{GB}_k^\beta(G_d)$ with $\beta=\frac{1+\alpha}{1-\alpha}$

or equivalently

If p=z and $G_c(z)=\omega^{-1}(G(z))$ then $\omega^{-1}(\mathcal{GB}_k^\beta(G)) = \mathcal{GB}_k^\alpha(G_c)$ with $\alpha=\frac{\beta-1}{\beta+1}$ ○

Proof: See appendix 3A.

In the next theorem we present the main result of this section. It gives the conditions under which $\mathcal{GB}$ will lead to stable and minimal reduced order models and it gives a bound for the L∞–norm of the approximation error. First we define the notion of an admissible region.

Definition 3.4.4. Let G(p) be a FDLTS system. We define $AR \subset \mathbb{R} \cup \{\pm\infty\}$, the *admissible region* for G(p), by:

$$AR=[0,\infty] \text{ if G is continuous} \tag{3.4.3a}$$

$$AR=[-\infty,-1]\cup[1,\infty] \text{ if G is discrete.} \tag{3.4.3b}$$

where we use the notation $[a,\infty]$ for $[a,\infty]:= \{\mathbb{R}\cap[a,\infty)\} \cup \{\infty\}$.

Furthermore, we define the *interior* and *boundary* of AR as

	interior:	boundary:
G continuous:	$(0,\infty)$	$\{0,\infty\}$
G discrete:	$[-\infty,-1)\cup(1,\infty]$	$\{-1,1\}$

o

Theorem 3.4.5. Let G(p) be a FDLTS system and let [A,B,C,D] be a realization of G(p), that is balanced w.r.t. $\Sigma>0$, with $\sigma_k>\sigma_{k+1}$. Let $\alpha\in AR$, as defined by definition 3.4.4 and partition A,B,C,Σ according to (3.1.1). Further, let $\hat{G}=\mathcal{GB}_k^{\alpha}(G)$ and let $[\hat{A},\hat{B},\hat{C},\hat{D}]$ be given by (3.4.2). Then:

1. $[\hat{A},\hat{B},\hat{C},\hat{D}]$ is a stable and minimal realization.
2. The L∞ norm of the approximation error is bounded by:

$$\|G-\hat{G}\|_{\infty} \leq 2(\sigma_{k+1}+\cdots\sigma_n)$$

with strict inequality if α is in the interior of AR.
and strict equality if k=n–1 and α is on the boundary of AR. o

Proof: See appendix 3B.

This theorem implies that if k=n–1, a value of α in the interior of AR will lead to a smaller frequency error than a value of α on the boundary. In section 3.5 we will show that the difference can be quite large and that this property also holds true for other values of k.

Remark 3.4.6.

1. In definition 3.4.1 and theorem 3.4.5 we used a parameter $\alpha \in \mathbb{R}$, which in fact indexes the family of reduced order models. It is straightforward to show that one can get a similar result as theorem 3.4.5 if α is allowed to be complex valued. In this case the admissible region AR, as defined by (3.4.3) is redefined as $\{\alpha\in\mathbb{C}, \text{real}(\alpha)\geq 0\}\cup\{\infty\}$ for continuous time systems and $\{\alpha\in\mathbb{C}, |\alpha|\geq 1\}\cup\{\infty\}$ for discrete time systems. Note that in general this leads to reduced order systems, with complex valued state space matrices,

though the original system was real valued. Such systems will produce complex valued outputs, even if the inputs are real valued, which is the reason we did not focus on this.

2. The reduction parameter α connects the standard methods in a continuous way. This is understood best if we consider the continuous time case, where $\alpha=\infty$ coincides with *CB* and $\alpha=0$ with *CSB*. Thus *CB* will result in an exact fit for $s=\infty$ and *CSB* produces an exact fit for $s=0$. Variation of α from 0 to ∞ gives a continuous transition from a match on the very low frequencies to a match on the very high frequencies, with the result that we expect that in the interval $(0,\infty)$ these two goals are more or less weighted against each other with weighting factors determined by the choice of α. Hence we expect that the freedom in the choice of α can be used to optimize the frequency characteristics of the approximant according to the designers specifications, in the bandwidth which is of importance. This expectation is confirmed by practical experiences, but we were not able to substantiate them with analytical results. However, the introduction of the reduction parameter presents a freedom in the design of reduced order models, which is a major advantage over the standard methods that only leave the choice between matching either the very high or very low frequency behavior.

3. From practical experiments we have the very strong impression that only one value of α exists for which the L_∞–norm of the error transfer function reaches a minimum. If one would define a function $f(\alpha) := \|G(p)-\mathcal{GB}_k^\alpha(G)\|_\infty$ then our experiences show that this function has only one global minimum $f(\alpha_{min})$ and no local minima. If we consider the continuous time then $f(\alpha)$ will reach 2 maxima on the boundary of the admissible region, i.e. $\alpha=0$ and $\alpha=\infty$, and have no other local maxima. However we did not succeed in finding values for α_{min} and $f(\alpha_{min})$ or in proving that $f(\alpha)$ has only one minimum.

4. Liu and Anderson (1989) propose to combine the standard methods in order to improve the frequency characteristics of the reduced order model. They use for instance the combination of *CB* and *CSB* and show through examples how this can improve the error. We strongly believe that a 'good' choice of α can do an even better job in just one reduction step without using several 'one at a step' reductions. As mentioned before we have not yet succeeded in finding rules for the optimal value of α, but the improvement can be quite impressive, as will be shown in the next section, where we compare the model reduction methods based on balanced realizations by means of some examples. o

3.5 EXAMPLES OF BALANCED REDUCTIONS

In this section we present three examples for the application of the model reduction techniques that were defined in the previous sections. We will show the influence of the extra freedom which is created by the reduction parameter α, introduced in section 3.4, where we emphasize the influence on the frequency behavior. The first example is merely

used for explanatory reasons; the other examples both are taken from literature on model reduction.

EXAMPLE 1

We consider a first order stable SISO system: $G(s) = \frac{1}{s-b} + d$ (3.5.1)

where b<0. This system has a realization [b,1,1,d] that is balanced with respect to the Hankel singular value $\sigma_1=\frac{-1}{2b}$. This can be checked with the Lyapunov equations

$$A\Sigma+\Sigma A^*+BB^* = A^*\Sigma+\Sigma A+C^*C = 0$$

Approximation of this system with a 0^{th} order system by applying $\mathcal{GB}$, results in:

$$\mathcal{GB}_0^{\alpha}(G) = d + \frac{1}{\alpha-b} \tag{3.5.2}$$

and the error transfer function $E(s):=G(s)-\mathcal{GB}_0^{\alpha}(G)(s)$ is given by

$$E(s) = \frac{1}{s-b} - \frac{1}{\alpha-b} \tag{3.5.3}$$

A first order system reaches the maximum of its frequency amplitude in s=0 or in s=∞, depending on the value of the feedthrough term. This shows that

$$\|E(j\omega)\|_{\infty} = \max(|\frac{1}{b} + \frac{1}{\alpha-b}| \ , \ |\frac{1}{\alpha-b}|) = \max(|\frac{\alpha}{b}| \ , \ 1)\times|\frac{1}{\alpha-b}| \tag{3.5.4}$$

Since $\alpha\geq 0$ and b<0 we find: $\|E(j\omega)\|_{\infty} = |\frac{1}{\alpha-b}|$ if $\alpha+b\leq 0$ (3.5.5a)

$$= |\frac{\alpha}{b}|\cdot|\frac{1}{\alpha-b}| \quad \alpha+b\geq 0 \tag{3.5.5b}$$

The standard methods $\mathcal{CB}$ ($\alpha=\infty$) and $\mathcal{CSB}$ ($\alpha=0$) lead to the same error $\frac{-1}{b}$, while a choice of $\alpha=-b$ results in an error $\frac{-1}{2b}$. This turns out to be the minimal value of the error, so in this simple case the frequency errors of the standard methods can be reduced with a factor 2. Figure 3.5.1 shows the plot of $\|G-\mathcal{GB}_k^{\alpha}(G)\|_{\infty}$ as a function of α (b=–10).

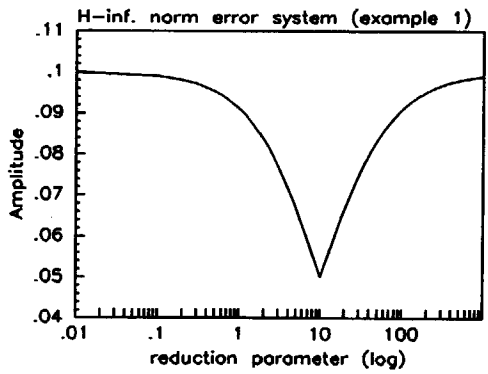


Fig. 3.5.1 $f(\alpha)=\|G-\mathcal{GB}_k^{\alpha}(G)\|_{\infty}$ for example 1.

EXAMPLE 2

As a second example showing the influence of the parameter α, we consider a simple third order system, that was used by Enns (1984). The transfer function is:

$$G(s) = \frac{(s+0.8)(s+2)}{(s+1.5)(s^2+1.4s+1)} \tag{3.5.6}$$

The Hankel singular values of this system are

$$\{\sigma_1,\sigma_2,\sigma_3\} = \{0.6985,\ 0.1599,\ 0.0053\} \tag{3.5.7}$$

We approximate G(s) with first order models, applying different values of α. As to be expected the result shows that for α=0 ($\mathcal{CSB}$) the approximation has the same static gain as G(s), while for α=∞ ($\mathcal{CB}$) the high frequency behavior is matched. This is shown in Fig. 3.5.2 and Fig 3.5.3. Figure 3.5.2 shows the Bode amplitude plot of the original model and the approximations with α=0,1,∞ ($\mathcal{CSB}$, $\mathcal{GB}$, $\mathcal{CB}$). In Fig. 3.5.3 the frequency errors are shown for the same values of α. It is clear that the response for α=1 lies more or less in the middle of the responses for the approximations with α=0 and α=∞. Figure 3.5.3 leads to the impression that the response for α=1 is a weighted sum of the other two responses. If we make α smaller then the weighting of α=0 will become more important, while an increase of α leads to more influence of the response for α=∞. This presents a freedom to design the reduced order model according to specific demands.

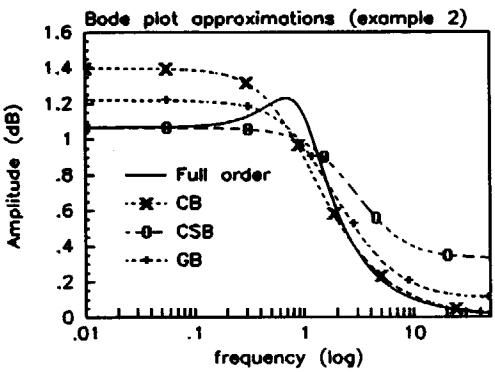


Fig. 3.5.2 Bode amplitude plot of G and $\mathcal{GB}_k^{\alpha}$(G) for α=0,1,∞

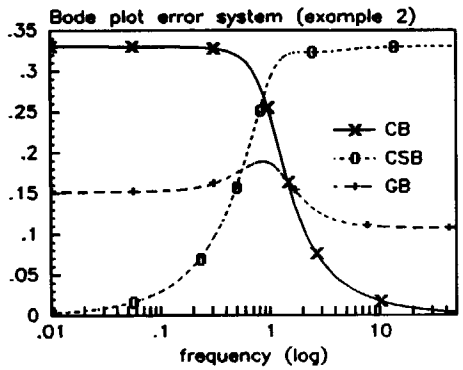


Fig. 3.5.3 Bode amplitude plot of G-$\mathcal{GB}_k^{\alpha}$(G) for α=0,1,∞

(example 2)

Figure 3.5.4 depicts the L∞–norm of the error transfer function as a function of α It is a plot of the function $f(\alpha) = \|G(s)-\mathcal{GB}_1^{\alpha}(G)\|_{\infty}$. The form of this function is typical for what we found with all kind of different systems. This lead to the impression (remark 3.4.6–3) that such functions have only one global minimum.

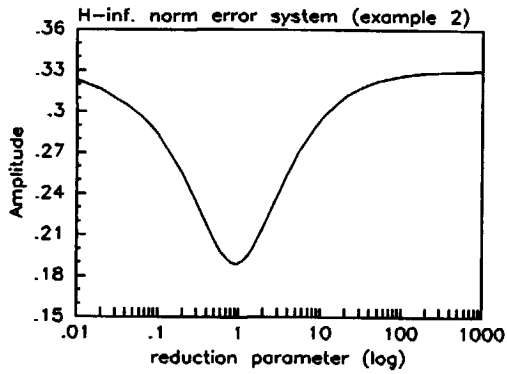


Fig. 3.5.4 $f(\alpha)=\|G-\mathcal{GB}_k^{\alpha}(G)\|_{\infty}$ for example 2.

EXAMPLE 3

We consider the example used by Liu and Anderson (1989) and create second order approximations of

$$G(s) = \frac{(s+4)}{(s+1)(s+3)(s+5)(s+10)} \tag{3.5.7}$$

which has Hankel singular values:

$$\{\sigma_1,\sigma_2,\sigma_3,\sigma_4\} = \{1.5938\text{x}10^{-2},\ 2.7243\text{x}10^{-3},\ 1.272\text{x}10^{-4},\ 8.006\text{x}10^{-6}\} \tag{3.5.8}$$

The theoretical upper bound for the L∞ norm of the error transfer functions is $2\times(\sigma_3+\sigma_4) = 2.7024\text{x}10^{-4}$. Liu and Anderson use a mixture of one at a step standard reductions (*CB* and *CSB*) to compare the maximal frequency errors and the errors at DC (s=0). This means they first reduce to order 3 and then from order 3 to 2, which we denote by *CB/CSB* if the first method used is *CB* and the second method is *CSB*. By using an iterative search procedure we calculated the optimal α with respect to the L∞ norm to be α_{opt}:=11.83. This results in a much smaller error, as can be seen in Table 3.5.1. We copied this table from the paper by Liu and Anderson and added the errors for $\mathcal{GB}^{\alpha_{opt}}$.

TABLE 3.5.1. Frequency Errors of the Reductions (example 3)

$\times 10^{-4}$

	CB	*CSB*	*CB/CSB*	*CSB/CB*	$\mathcal{GB}^{\alpha}$opt
$\|G-\hat{G}\|_\infty$	2.4802	2.3692	2.5248	2.6402	1.3415
DC–err	2.384	0.0	0.1601	2.5441	0.9810

While the DC–error is still acceptable, the L∞ norm of the frequency error is nearly half of what is achieved by the other methods. In Fig. 3.5.5 the amplitude plots of the error transfer functions for the approximations are shown on the whole frequency scale, and it shows that $\mathcal{GB}$ makes a trade–off between matching high and low frequencies. However, it should be pointed out that the frequency error of *CSB* is only large for the high frequencies, which may be of no interest for certain applications. In Fig. 3.5.6 we depict the maximal frequency error as a function of α, and it shows a similar curve as Fig. 3.5.4 , with only one global minimum.

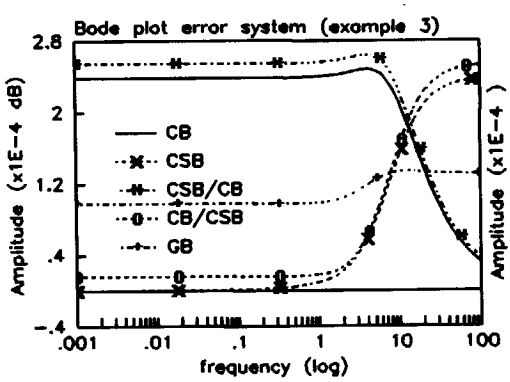


Fig. 3.5.5 Bode amplitude plot of G–Ĝ

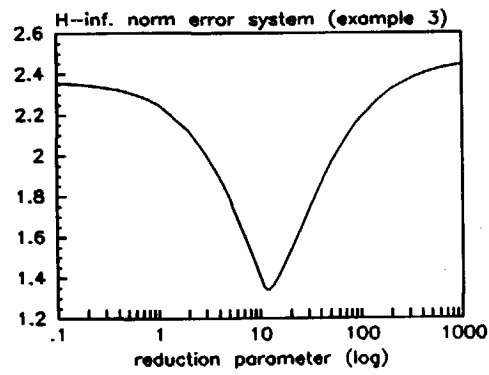


Fig. 3.5.6 $f(\alpha)=\|G-\mathcal{GB}_k^{\alpha}(G)\|_\infty$ (example 3)

We concentrated the error analysis to the L∞ norm of the error system, and though no bounds are known for other measures of the error, it is interesting to consider an other measure that is widely used in the field of model reduction, the Hankel norm of the error system. Glover (1984) shows that the Hankel norm for a k^{th} order approximation is bounded from below by σ_{k+1}, thus in this case the theoretical bound is $\sigma_3=1.272\text{x}10^{-4}$. We calculated the Hankel norm of the error system for the approximations that were used for table 3.5.1. and present them in Table 3.5.2, from which we conclude that for this example $\mathcal{GB}$ is also superior to the other methods in the Hankel norm.

TABLE 3.5.2. Hankel norm of Reduction Errors

x 10^{-4}

	$\mathcal{CB}$	$\mathcal{CSB}$	$\mathcal{CB}/\mathcal{CSB}$	$\mathcal{CSB}/\mathcal{CB}$	$\mathcal{GB}^{\alpha}$opt
$\|G-\hat{G}\|_H$	2.4291	1.8646	2.5874	1.9722	1.3177

In Table 3.5.2 we used again α_{opt}=11.84 for $\mathcal{GB}$, but this is not the optimal value of α for the Hankel norm. In Fig. 3.5.7 the Hankel norm of the reduction error is shown as a function of the reduction parameter; it reaches a minimum 1.2931×10^{-4} in α_{han}:=13.28, which is near the theoretical underbound. This shows that, for the optimal value of α, $\mathcal{GB}$ results in a very good approximation with respect to the Hankel norm. In Fig. 3.5.7 we see again that there is only one global minimum, but, again, we are not able to present analytical proof for this property.

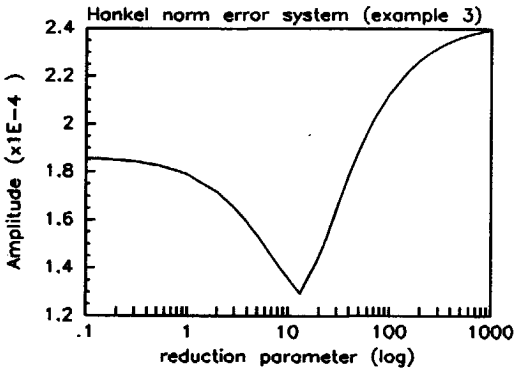


Fig. 3.5.7 $f(\alpha)=\|G-\mathcal{GB}_k^{\alpha}(G)\|_H$ for example 3.

Remark 3.5.1. The presented examples show that the introduction of the design variable α has two major consequences. First, the L_∞ norm of the error transfer function can be decreased considerably, if compared to the standard methods. Secondly it offers the opportunity to incorporate pre–defined demands and qualifications such that the reduced order model can be shaped according to the designers wishes. This freedom in design can be compared with the possibilities of model reduction methods which use frequency weighting functions (Enns, 1984; All Saggaf and Franklin, 1987,1988). We expect that the combination of the generalized model reduction method with the application of frequency weighting will open the way to improve the results of these authors.

3.6 HANKEL NORM APPROXIMATION

3.6.1. PROBLEM DEFINITION

In this section we consider the problem of approximating a stable system with a lower order stable system, such that the Hankel norm of the difference transfer function is minimized, the so called Hankel norm approximation problem. This problem was first shown to be solvable (in the discrete time domain) by Adamyan, Arov and Krein (1968, 1971, 1978) and the first reliable state space algorithm (for the continuous time domain) was developed by Glover (1984). Many other authors have contributed to this field, among which Kung and Lin (1981[ab]), Ball and Ran (1987[ab]) for the discrete time domain, and Savonof et.al. (1987,1990) for the continuous time domain. For applications the algorithms of Savonof and Glover are commonly used. These are given for continuous time systems whereas for discrete time systems the problem is solved by using the ω–transformation (2.5) to transform the problem into a continuous time setting and to transform the continuous time solution back to the discrete time domain. Glover (1984) showed that the latter procedure results in an optimal discrete time solution. As shown in section 2.8 the use of the ω–transformation involves the inverse of the system matrix and this requires much computational effort if the system has a large McMillan degree. We believe that an algorithm in terms of state space matrices would be an important contribution and such an algorithm does not exist in the literature to our knowledge. Furthermore we are interested in the structure of this problem with respect to system approximation. Thereto we follow the line of reasoning of Glover (1984), applied to discrete time systems. Thus we aim at a direct discrete time solution without using the ω–transformation (2.4). This does not result in nice closed forms for a solution but leads to a set of sufficient equations. First we state the problem, independent of a time domain, and show that there exist optimal solutions. In section 3.6.2 we sketch the line of reasoning for the continuous time case and in section 3.6.3 we derive the new formulas for the discrete time case. We concentrate on stable systems, because the Hankel norm has no influence on the unstable part of a system, as pointed out before in section 2.10. We begin with a proper definition of the problem.

DEFINITION 3.6.1 THE HANKEL NORM APPROXIMATION PROBLEM

Let G(p) be a p×m FDLTS system with McMillan degree n. The *Hankel norm approximation problem* is to find for k<n, an approximating p×m FDLTS system $\hat{G}_k(p)$ with McMillan degree k, such that $\|G(p)-\hat{G}_k(p)\|_H$ is minimal with respect to all FDLTS systems with McMillan degree k. Any such $\hat{G}(p)$ is called a *Hankel norm approximation* of G(p). ○

Note that formally stability is not required, but we use this definition since the Hankel norm is only influenced by the stable part of the systems. This problem is solvable, which is in fact quite surprising because there are no norms known other than the Hankel norm for which the equivalent problem has a closed form solution. The following lemma states the solvability of the problem and gives the value of the minimal norm of the error system.

LEMMA 3.6.2 [Adamyan, Arov and Krein, 1978; Glover, 1984]. Let G(p) be a p×m FDLTS system with McMillan degree n and Hankel singular values $\{\sigma_1 \geq \cdots \geq \sigma_n\}$. If k<n and $\sigma_k > \sigma_{k+1}$ then for every p×m FDLTS system $F_k(p)$ with McMillan degree k the Hankel norm of the error transfer function is bounded by:

$$\|G(p)-F_k(p)\|_H \geq \sigma_{k+1} \tag{3.6.1}$$

Furthermore there exist a $\hat{F}_k(p)$, such that this bound is achieved,

$$\|G(p)-\hat{F}_k(p)\|_H = \sigma_{k+1} \tag{3.6.2}$$

and if p=m=1, the strictly proper part of this system $\hat{F}_k(p)$ is unique. o

Another error bound that is widely used is the bound on the L_∞ norm of the error transfer function in the frequency domain, which we used to compare the reduced order models based on balancing (theorem 3.4.5). The next lemma shows that the same bound, as for the balanced reductions, applies for Hankel norm approximations, but that it can be decreased if a feedthrough matrix is added.

LEMMA 3.6.3 [Glover, 1984] Let G(p) be a p×m FDLTS system with McMillan degree n and Hankel singular values $\{\sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} \geq \cdots \geq \sigma_n\}$ and let $\hat{G}_k(p)$ be an Hankel norm approximation of G(p) with McMillan degree k. Then the L_∞ norm of the error transfer function is bounded by:

$$\|G(p)-\hat{G}(p)\|_\infty \leq 2(\sigma_{k+1} + \sigma_{k+2} + \cdots + \sigma_n) \tag{3.6.3}$$

Furthermore there exists a feedthrough matrix $D \in \mathbb{F}^{p \times m}$ such that

$$\|G(p)+D-\hat{G}(p)\|_\infty \leq (\sigma_{k+1} + \sigma_{k+2} + \cdots + \sigma_n) \tag{3.6.4}$$

o

REMARK 3.6.4. The D–matrix in (3.6.4) has no influence on the Hankel norm, which only depends on the strictly proper stable part of a transfer function (see remark 2.10.2), but it does influence the L_∞–norm. This is easy to see if we consider SISO systems, where the L_∞–norm is the same as the supremum of the Bode amplitude plot. o

First we sketch a way to solve this problem for continuous time systems and use the same approach in the discrete time domain. On beforehand we can state that though the

first does lead to a closed form for the solution the latter will not, but it does have some interesting features.

3.6.2. THE CONTINUOUS TIME CASE

In this section we sketch how the result on all pass functions as stated in theorem 2.7.2 leads to a solution of the continuous time Hankel norm approximation problem for square transfer functions; we follow the line of Glover (1984) and refer to that paper for the presented results. We can distinguish six main steps that yield the solution:

1. Let G(s) be a square FDLTS system with McMillan degree n and let [A,B,C,D] be a realization of G, that is balanced w.r.t. $\Sigma = \begin{bmatrix} \sigma_1 I_{m_1} & & & \\ & \sigma_2 I_{m_2} & & \\ & & \ddots & \\ & & & \sigma_r I_{m_r} \end{bmatrix}$

   where m_i is the multiplicity of σ_i and $\sigma_i > \sigma_{i+1}$. Thus $n = m_1 + m_2 + \cdots + m_r$.
2. Glover considers the case that the order n_r of a reduced order model is equal to $n_k = \sum_{i=1}^{k} m_i$ $(k<r)$. Let $\sigma := \sigma_{k+1}$ and transform $\{A,B,C,\Sigma\}$ such that this singular value is replaced to the last block entry of Σ. In the sequel we therefore assume that [A,B,C,D] is a balanced realization w.r.t Σ, (non standard with respect to convention 2.6.5) with $\Sigma = \begin{bmatrix} \Sigma_1 & \\ & \sigma I \end{bmatrix}$ (3.6.5)

   where Σ_1 contains the other singular values: $(\Sigma_1 - \sigma I)$ is non–singular and we denote the dimension of Σ_1 by n_1 $(n_1 = n - m_{k+1})$.
3. The approach taken by Glover is motivated by the work of Adamyan, Arov and Krein (1978), who show that there exists a FDLT system F(s) (not necessarily stable) such that $\sigma^{-1} \times (G(s)-F(s))$ is an all–pass function. This shows immediately that $\|G-F\|_\infty = \|G-F\|_H = \sigma$. Remember that the Hankel norm is not influenced by the unstable part of the system and let $\hat{G}(s)$ be the stable part of F(s), then it follows that $\|G-\hat{G}\|_H = \sigma$. This $\hat{G}(s)$ will indeed have McMillan degree n_k.

ASSUMPTION: We assume that there exists an FDLT system F(s) with Mcmillan degree n_1 such that E(s) = G(s)–F(s) is $\sigma \times$ all–pass. ○

This assumption restricts the set of solutions considerably, but will lead to a solution. Glover (1984) (chapter 8) shows that all solutions are obtained if we allow F(s) to have also other McMillan degrees.

4. Let $[\hat{A},\hat{B},\hat{C},\hat{D}]$ be a realization of F(s) and $[A_e,B_e,C_e,D_e]$ the corresponding realization of E(s):
$$A_e = \begin{bmatrix} A & \\ & \hat{A} \end{bmatrix} \quad B_e = \begin{bmatrix} B \\ \hat{B} \end{bmatrix} \quad C_e = [C \;\; -\hat{C}] \quad D_e = D - \hat{D} \tag{3.6.6}$$
Apply theorem 2.7.2 on E(s) to show that $[A_e,B_e,C_e,D_e]$ must have solutions to the Lyapunov equations P_e and Q_e with
$$A_eP_e + P_eA_e^* + B_eB_e^* = 0 \qquad A_e^*Q_e + Q_eA_e + C_e^*C_e^* = 0 \tag{3.6.7a,b}$$
$$P_eQ_e = \sigma^2 I \qquad D_e^*D_e = \sigma^2 I \tag{3.6.8a,b}$$
$$D_e^*C + B_e^*Q_e = 0 \qquad D_eB_e^* + C_eP_e = 0 \tag{3.6.9a,b}$$
5. Define $\Gamma := \Sigma_1^2 - \sigma^2 I$ (non singular) (3.6.10)
then Glover shows we can assume that:
$$P_e = \begin{bmatrix} \Sigma_1 & 0 & I \\ 0 & \sigma & 0 \\ I & 0 & \Sigma_1\Gamma^{-1} \end{bmatrix} \qquad Q_e = \begin{bmatrix} \Sigma_1 & 0 & -\Gamma \\ 0 & \sigma & 0 \\ -\Gamma & 0 & \Sigma_1\Gamma \end{bmatrix} \tag{3.6.11}$$
6. Partition the matrices A,B,C according to Σ in
$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad B = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad C = [C_1 \; C_2] \tag{3.6.12}$$
Note that the balancedness implies that $-\sigma(A_{22}+A_{22}^*) = B_2B_2^* = C_2^*C_2$ and hence
$$\exists \text{ a unitary matrix U with } B_2U = C_2^*. \tag{3.6.13}$$
Now consider the equation (3.6.9a) $D_eB_e^* + C_eP_e = 0$:
$$(1)\; D_eB_1^* + C_1\Sigma_1 - \hat{C} = 0 \tag{3.6.14a}$$
$$(2)\; D_eB_2^* + \sigma C_2 = 0 \tag{3.6.14b}$$
$$(3)\; D_e\hat{B}^* + C_1 - \hat{C}\Sigma_1\Gamma^{-1} = 0 \tag{3.6.14c}$$
With these equations (3.6.14) we can calculate $\{\hat{B},\hat{C},\hat{D}\}$ as follows.
$$(3.6.14b) + (3.6.13) \Rightarrow \qquad D_e = -\sigma U^*, \;\; \hat{D} = D + \sigma U^* \tag{3.6.15a}$$
$$(3.6.14a) + (3.6.15a) \Rightarrow \qquad \hat{C} = -\sigma U^*B_1^* + C_1\Sigma_1 \tag{3.6.15b}$$
$$(3.6.14c) + (3.6.15b) \Rightarrow \qquad \hat{B}^* = (\sigma UC_1 + B_1^*)\Gamma^{-1} \tag{3.6.15c}$$
For the evaluation of $\hat{A}$ we consider the (1,3)–block of $A_eP_e + P_eA_e^* + B_eB_e^* = 0$, which gives immediately an explicit form : $A_{11} + \hat{A}^* + B_1\hat{B}^* = 0$ (3.6.15d)

The equations (3.6.13) and (3.6.15) thus give an explicit solution for the realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ and it remains to proof that the stable part of $[\hat{A},\hat{B},\hat{C},\hat{D}]$ has the correct McMillan degree. This is proven in theorem 6.3 of Glover (1984), based on the inertia of $\Sigma_1\Gamma^{-1}$ and the fact that $\hat{A}$ has no purely imaginary poles.

The main steps can be summarized as follows:
- Given a FDLTS system G(z), create a realization [A,B,C,D] of G(z), that is (non standard) balanced w.r.t Σ as given by (3.6.5); (3.6.16a)
- find solutions P_e an Q_e (3.6.11) to the Lyapunov equations for the realization of the error system as given by (3.6.6); (3.6.16b)

- Calculate the (possibly unstable) realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ of the system F(z) from the equations (3.6.7/8/9); (3.6.16c)
- Calculate $\hat{G}(z)$, the stable part of F(z). (3.6.16d)

We wish to emphasize again that the above is only meant to be a rough sketch of a solution, with only the intention to show the line of reasoning. For a complete and thorough treatment we refer to Glover (1984).

3.6.3 THE DISCRETE TIME CASE

In this section we use the same line of reasoning as in the previous one, based on the fact that the discrete time Hankel norm approximation problem can be solved by applying the ω–transformation to create a continuous time system, solving the problem as in section 3.6.2 and using again the ω–transformation to yield the discrete time solution. This will indeed produce an optimal solution because the Hankel singular values are invariant under the ω–transformation. The question arises if it would be possible to solve the problem 3.6.1 with p=z, directly in the discrete time domain, which would save the calculation of the ω–transformation. Especially if the system order n is large this would save quite some computation time, since the ω–transformation requires the inversion of an n×n–matrix.
Moreover for the problem of approximate system identification we are mainly interested in the estimation of discrete time models. Since we wish to get a better understanding of Hankel norm approximation for a possible application in an identification context, we are interested in the structure of this problem in the discrete time case. Using the approach of Glover leads to a set of new sufficient equations, which are hard to solve.

We follow the steps as given by (3.6.16). Thus we assume that we have a square stable transfer function matrix G(z), with a non–standard balanced realization [A,B,C,D] w.r.t $\Sigma = \begin{bmatrix}\Sigma_1 \\ & \sigma I\end{bmatrix}$ as in (3.6.5) and define $\Gamma := \Sigma_1^2 - \sigma^2 I$, non–singular. We search for a F(z), such that E(z)=G(z)–F(z) has the property that $\sigma^{-1}E(z)$ is an all–pass function. Let $[\hat{A},\hat{B},\hat{C},\hat{D}]$ and $[A_e,B_e,C_e,D_e]$ be realizations of G(z) and E(z) respectively as given by (3.6.6). We apply theorem 2.7.4 to show that there must exist 'Gramians' P_e,Q_e for the realization of E(z) and the invariance of Gramians under the ω–transformation implies that we may assume, as in (3.6.11), that:

$$P_e = \begin{bmatrix}\Sigma_1 & 0 & I \\ 0 & \sigma & 0 \\ I & 0 & \Sigma_1\Gamma^{-1}\end{bmatrix} \qquad Q_e = \begin{bmatrix}\Sigma_1 & 0 & -\Gamma \\ 0 & \sigma & 0 \\ -\Gamma & 0 & \Sigma_1\Gamma\end{bmatrix} \tag{3.6.17}$$

The next step is the calculation of $[\hat{A},\hat{B},\hat{C},\hat{D}]$ from the equations for P_e, Q_e and $[A_e,B_e,C_e,D_e]$ which are given in theorem 2.7.4 on discrete time all–pass functions.
At this point the derivation deviates from the continuous time case, because of the different form of the discrete time Lyapunov equations. The next lemma shows the result of applying theorem 2.7.4 to this situation.

LEMMA 3.6.5. Let G(z) be a m×m FDLTS system with McMillan degree n and let $\{\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_k>\sigma_{k+1}=\sigma_{k+2}=\cdots\sigma_{k+r}>\sigma_{k+r+1}\geq\cdots\sigma_n\}$ be the set of Hankel singular values of G(z). Define

$$\Sigma_1:=\mathrm{diag}(\sigma_1,\cdots,\sigma_k,\sigma_{k+r+1},\cdots,\sigma_n) \tag{3.6.18a}$$

$$\sigma:=\sigma_{k+1} \text{ and } \Gamma:=\Sigma_1^2-\sigma^2 I_r \tag{3.6.18b,c}$$

and let [A,B,C,D] be a minimal realization of G(z) that is (non standard, convention 2.6.5) balanced with respect to $\Sigma = \begin{bmatrix}\Sigma_1 & 0\\ 0 & \sigma I_r\end{bmatrix}$. Assume that there exists a m×m FDLT system F(z) with McMillan degree $n_1:=(n-r)$, such that

$$E(z):= \sigma^{-1}\times(G(z)-F(z)) \text{ is all–pass with McMillan degree } 2n-r. \tag{3.6.19}$$

Then any such F(z) has a minimal realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ that obeys the equations (3.6.20/21). Furthermore, any set of matrices $\{\hat{A}\in\mathbb{F}^{n_1\times n_1},\hat{B}\in\mathbb{F}^{n_1\times m},\hat{C}\in\mathbb{F}^{m\times n_1},\hat{D}\in\mathbb{F}^{m\times m}\}$ that obeys these equations constitutes a minimal realization of an F(z) with the property (3.6.19).

$$\left[\begin{bmatrix}I\\0\end{bmatrix}\,\middle|\, A\begin{bmatrix}\Sigma_1\\ & \sigma I\end{bmatrix}C^* + BD^*\right] = \begin{bmatrix}A_{11} & B_1\\ A_{21} & B_2\end{bmatrix}\begin{bmatrix}\hat{A}^* & \hat{C}^*\\ \hat{B}^* & \hat{D}^*\end{bmatrix} \tag{3.6.20a}$$

$$\hat{A}\left[\Sigma_1\Gamma^{-1}\hat{C}^* - C_1^*\right] = \hat{B}D_e^* \tag{3.6.20b}$$

$$\hat{A}\Sigma_1\Gamma^{-1}\hat{A}^* + \hat{B}\hat{B}^* = \Sigma\Gamma^{-1} \tag{3.6.20c}$$

and the dual relations:

$$\left[\begin{bmatrix}\Gamma\\0\end{bmatrix}\,\middle|\, A^*\begin{bmatrix}\Sigma_1\\ & \sigma I\end{bmatrix}B + C^*D\right] = \begin{bmatrix}A_{11}^*\Gamma & C_1^*\\ A_{12}^*\Gamma & C_2^*\end{bmatrix}\begin{bmatrix}\hat{A} & \hat{B}\\ \hat{C} & \hat{D}\end{bmatrix} \tag{3.6.21a}$$

$$\hat{A}^*\left[\Gamma B_1 - \Sigma_1\Gamma\hat{B}\right] = -\hat{C}^*D_e \tag{3.6.21b}$$

$$\hat{A}^*\Sigma_1\Gamma\hat{A} + \hat{C}^*\hat{C} = \Sigma_1\Gamma \tag{3.6.21c}$$

o

PROOF: See appendix 3C

Once we have found a realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ for F(z), the stable part of this system will be a stable Hankel norm approximation to G(z) as in (3.6.16d).

REMARK 3.6.6. It should be stressed here that with the above equations we will only find those Hankel norm approximations that fulfil the all–pass condition (3.6.19); there may exist other candidates, for instance if $E(z)\tilde{E}^T(\frac{1}{z})=\begin{bmatrix}0 & 0\\ 0 & I\end{bmatrix}$ then we also would find a

Hankel norm approximation, but not one that obeys the conditions of lemma 3.6.5, where we used theorem 2.6.4 and the condition $E(z)\bar{E}^T(\frac{1}{z})=I$. An example of this is the situation $G(z)=z^{-1}\Sigma^2$, with $\Sigma=\begin{bmatrix}\Sigma_1 & \\ & \sigma\end{bmatrix}$ square and diagonal with $\sigma I<\Sigma_1$, i.e. a system with one Markov parameter. It is immediate that $\hat{G}(z)=z^{-1}\begin{bmatrix}\Sigma_1^2 & \\ & 0\end{bmatrix}$ is a Hankel norm approximation, but it does not fit in the framework of lemma 3.6.5. o

Investigation of the equations (3.6.20/21) has shown that in general it will not be easy to find a closed form for the realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$. A close look at these equations reveals that (3.6.20a) and (3.21a) are equations that are linear in the unknown parameters. This implies that a solution can readily be found in the case that these equations are overdetermined, in the sense that there are more independent relations than unknown parameters. This situation occurs if we are dealing with a SISO system, for which the realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ is completely determined by these equations. This is stated in the following proposition.

Proposition 3.6.7. Under the conditions of lemma 3.6.5 with m=1, i.e. if G(z) is a SISO system, then equation (3.6.20a) (or equivalently (3.6.21a)) completely determines the realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ and this is given by:

$$\begin{bmatrix}\hat{A}^* & \hat{C}^* \\ \hat{B}^* & \hat{D}^*\end{bmatrix} = \begin{bmatrix}A_{11} & B_1 \\ A_{12} & B_2\end{bmatrix}^{\dagger} \left[\begin{bmatrix}I \\ 0\end{bmatrix} \;\middle|\; A\begin{bmatrix}\Sigma_1 & \\ & \sigma I\end{bmatrix}C^* + BD^*\right] \tag{3.6.22}$$

where † denotes the left inverse. o

Proof: See appendix 3D

We conclude that in the SISO case equation (3.6.20a) solves the Hankel norm approximation problem. Without actually proving it, we state that equation (3.6.22) and its dual equation can also be used to solve the Hankel norm approximation problem for all MISO, respectively SIMO systems. This property is not a new result, since the Hankel norm approximation problem for these systems was already solved by, among others, Kung and Lin (1981). However we believe that the structure of the solutions as presented here is new. Furthermore, we know that any solution to the equations (3.6.20/21) leads to an optimal Hankel norm approximation for square systems and if we could find a solution then the non square case could be solved in the same way.

The reason why in the MIMO case the equation (3.6.20a) is not sufficient stems from the fact that in those situations the matrix $\begin{bmatrix}A_{11} & B_1 \\ A_{12} & B_2\end{bmatrix}$ is not left invertible. This matrix has dimensions $n\times(n_1+m) = n\times(n-r+m)$, where r is the multiplicity of σ. Generically r=1, so for MIMO systems this matrix has more columns than rows and will not be left

invertible. In other words, the number of equations, defined by (3.6.20a) is smaller then the number of unknown parameters and is thus not sufficient to determine the approximation completely. This was to be expected, since we know from Adamyan, Arov and Krein (1978) and Glover (1984) that in the multivariable case there is a lot of freedom left and there exist many different Hankel norm approximations.
We tried to overcome this problem by following the line of thought of Kung and Lin (1981) for multivariable systems. The idea is to tackle the problem by means of order extension and to create a situation in which the multiplicity of σ is large enough to allow the matrix $\begin{bmatrix} A_{11} & B_1 \\ A_{12} & B_2 \end{bmatrix}$ to be left invertible. Kung and Lin (1981) applied this idea for suboptimal Hankel norm approximation; we use it for the optimal case.

LEMMA 3.6.8. [Kung and Lin, 1981b]. Let G(z) be a FDLTS system with strictly proper part $G^P(z)$ and let $G_1(z):=z^{-1}(G^P(z)+D)$, where D is a constant matrix with appropriate dimensions. If $\hat{G}_1(z)$ is an optimal Hankel norm approximation to $G_1(z)$, with McMillan degree k, and $\hat{G}_1^P(z)$ is the strictly proper part of $\hat{G}_1(z)$, then $\hat{G}(z):=z\hat{G}_1^P(z)$ is an optimal Hankel norm approximation to G(z), with McMillan degree k. o

The proof of this lemma is based on the fact that the spectral norm of a sub matrix of a Hankel matrix $\mathscr{H}$ is smaller than or equal to the spectral norm of $\mathscr{H}$.

Let G(z) be a m×m FDLTS system and let σ be the Hankel singular value of G that we wish to discard, with multiplicity r<m. We try to use lemma 3.6.8 in order to find a D matrix such that $G_1(z)$ in lemma 3.6.8 has a singular value σ with multiplicity r_1 which is larger than r. Our goal is thus to find a D–matrix such that application of (3.6.20a) to the system $G_1(z)$ leads to an optimal Hankel norm approximation to $G_1(z)$. Hence we have to investigate the Hankel singular values of $G_1(z)$. Thereto we use the following proposition, that gives the Gramians of a realization of $G_1(z)$ and conditions under which σ is a Hankel singular value of $G_1(z)$ with a larger McMillan degree:

PROPOSITION 3.6.9. Let G(z) be a m×m strictly proper FDLTS system and let [A,B,C,0] be a minimal realization of G(z) that is (not necessarily standard) balanced w.r.t. $\Sigma>0$. Let $D\in\mathbb{F}^{m\times m}$ and $G_1(z):=z^{-1}(G(z)+D)$, then

1. $G_1(z)$ has a realization $[A_1,B_1,C_1,D_1]$ with:

$$A_1=\begin{bmatrix} A & B \\ 0 & 0 \end{bmatrix} \quad B_1=\begin{bmatrix} 0 \\ I \end{bmatrix} \quad C_1=[C \;\; D] \quad D_1=0 \qquad (3.6.23a)$$

with Gramians: $P_1=\begin{bmatrix} \Sigma & 0 \\ 0 & I \end{bmatrix}$ $\quad Q_1=\left[\begin{array}{c|c} \Sigma & A^*\Sigma B+C^*D \\ \hline D^*C+B^*\Sigma A & D^*D+B^*\Sigma B \end{array}\right]$ (3.6.23b)

Further $[A_1,B_1,C_1,D_1]$ is a minimal realization if and only if the system G(z)+D has no invariant zeros in z=0.

2. The Hankel singular values of $G_1(z)$ are the square roots of the non zero eigenvalues of the matrix $M:=\left[\begin{array}{c|c} \Sigma^2 & \Sigma^{\frac{1}{2}}(A^*\Sigma B+C^*D) \\ \hline (D^*C+B^*\Sigma A)\Sigma^{\frac{1}{2}} & D^*D+B^*\Sigma B \end{array}\right]$ (3.6.23c)

3. If $\Sigma=\begin{bmatrix}\Sigma_1 & 0\\ 0 & \sigma I_r\end{bmatrix}$, where $\Gamma:=(\Sigma_1^2-\sigma^2 I)$ is non singular, then $G_1(z)$ has a Hankel singular value σ with multiplicity $r_1\geq 0$, where:

   a. $r_1\geq r$ if $[0\ I_r](A^*\Sigma B+C^*D)=0$ (3.6.23d)

   b. $r_1\geq 2r$ if next to (3.6.23d) $[0\ I_r](A\Sigma C^*+BD^*)=0$ (3.6.23e)

   c. $r_1\geq r+m$ if next to (3.6.23d) $D^*WD+D^*V+V^*D+Z=0$ (3.6.23f)

   where with $Y:=\begin{bmatrix}-\Sigma_1\Gamma^{-1} & 0\\ 0 & 0\end{bmatrix}$

   $W:=I_m+CYC^*,\ V:=CYA^*\Sigma B,\quad Z:=B^*(\Sigma AYA^*\Sigma+\Sigma)B-\sigma^2 I_m$ (3.6.23g)

PROOF: See Appendix 3E ○

This proposition shows two ways to enlarge the multiplicity of σ. The most efficient method will be the last one (3c.), since this ensures that the system $G_1(z)$, which has McMillan degree $\leq(n+m)$ has a singular value σ with multiplicity $r_1\geq r+m$. Equation (3.6.20a), applied to this system, then results in (n+m) equations in (n+m–r) unknown parameters. However, this involves the solution of the Riccati like equation (3.6.23f), with the condition (3.6.23d), and this can be a problem, because for instance the matrix W in (3.6.23f,g) will in general be indefinite. This Riccati equation is equivalent with the Riccati equation from Kung and Lin (1981[b]). The method of (3b) is numerically superior, but it may be necessary to repeat it several times in order to enlarge the multiplicity such that equation (3.6.20a) could provide a solution.
Note however that we have offered no proof that a solution will actually be found if the multiplicity of σ is larger than m. However, practical experience indicates that if we find a D–matrix such that (3c) is valid then we will always come up with a solution.
Both methods require the existence of a D matrix such that equation (3.6.23d) is fulfilled and unfortunately there are systems, for which this is impossible and for which therefore procedure 3.6.9 cannot offer a method to solve the Hankel norm approximation problem. One such counter example is adapted from Pernebo and Silverman (1983):

EXAMPLE 3.6.10. Let $G(z)=\mathrm{diag}(\frac{z}{z^2-0.25},\frac{1}{z-0.5})$. $G(z)$ has a balanced realization:

$$A=\frac{1}{2}\cdot\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}\quad B=C^*=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\quad D=0\quad \text{w.r.t. } \Sigma=\frac{1}{15}\begin{bmatrix}16&0&0\\0&20&0\\0&0&4\end{bmatrix}.$$

We try to find a Hankel norm approximation with McMillan degree two with procedure 3.6.9. The problem is that $B_2=C_2^*=0$, while $[0\ 0\ 1]A^*\Sigma B\neq 0\neq[0\ 0\ 1]A\Sigma C^*$. Hence it will be impossible to find a D–matrix which fulfils (3.6.23d). It can be verified that there exists no D–matrix, such that $G_1(z):=z^{-1}(G(z)+D)$ has $\sigma=\frac{4}{15}$ as singular value with

multiplicity larger than 1 and therefore the extension procedure cannot be applied to this system. We can still solve the Hankel norm problem for this particular system, by observing that $G(z)=\text{diag}(\ G_{11}(z),G_{22}(z)\ )$, where $G_{11}(z)$ has Hankel singular values $\{\frac{16}{15}, \frac{4}{15}\}$ and $G_{22}(z)$ has a Hankel singular value $\{\frac{4}{3}\}$. Apply lemma 3.6.5 on $G_{11}(z)$ to find a first order Hankel norm approximation $\hat{G}_{11}(z)$, i.e $\|G_{11}(z)-\hat{G}_{11}(z)\|_H=\frac{4}{15}$ and define $\hat{G}(z)=\text{diag}(\ \hat{G}_{11}(z),G_{22}(z)\ )$. This is clearly a Hankel norm approximation to $G(z)$: $\|G(z)-\hat{G}(z)\|_H=\frac{4}{15}$. This latter procedure is transparent because $G(z)$ has the simple diagonal structure, but there exist many systems which do not have such a structure and still have $B_2=C_2^*=0$. Consider for instance the system $H(z):=UG(z)V$, with U and V unitary matrices. This system has a balanced realization [A,BV,UC,0] and while the diagonal structure may disappear, we still find $(BV)_2=(UC)_2^*=0$. o

Remark 3.6.11. We suspect that the reason for this problem is the occurrence of zeros in $z=0$ for the strictly proper system $G(z)$, but we have not been able to get a complete grip on the subject. As shown by proposition 3.6.9(1) a zero in $z=0$ implies that $z^{-1}G(z)$ has McMillan degree smaller than n+m. This has the consequence that the choice of a D–matrix, such that the system $z^{-1}(G(z)+D)$ has the desired singular values, is restricted. For the system of example 3.6.10 it is straightforward to show that for any D matrix the extended system $G_1(z)$ might have a singular value 4/15, but never with multiplicity larger that one. We will return to this problem in section 4.7, where it will be shown that under a certain transformation we may overcome this problem. It is not more than honest to admit at this point that a possible solution along these lines would very probable cause much more numeric problems then the procedure through the ω–transformation and the continuous time solution. o

3.7 FRACTIONAL BALANCED REDUCTION

In this section we extend the model reduction methods based on coprime factorizations for continuous time systems (Liu and Anderson, 1986; Meyer 1988) to the discrete time case. As in section 3.4 we will use the reduction factor α as a new design variable, that can be used to improve the quality of the reduced order models. A major advantage of these methods is that they can be applied to both stable and unstable plants. Furthermore, the error which is created can be interpreted in a close loop sense. The essential part of the methods is the existence of normalized coprime factorizations for discrete time transfer functions, which was established in section 2.8. The main idea is that one does not aim directly at reduction of the original plant, but – instead – at reduction of the coprime factors of the plant after which the reduced factors give rise to

a reduced order plant. This is often referred to as reduction of the graph of the plant, which is defined as follows,

Definition 3.7.1. [Vidyasagar, 1985]. Let G(z) be a p×m FDLT system. The graph $\mathcal{G}\{G(z)\}$ is defined as the set of input/output pairs of G(z) that are of finite energy:

$$\mathcal{G}\{G(z)\} = \{(u,y)\in\ell_2^{p+m} \mid y=Gu\} \qquad (3.7.1)$$

o

The relation y=Gu in this definition must be interpreted in the same way as in the definition 2.4.3 of the behavior of a system. Thus the graph is controlled by those inputs, which are bounded in energy and that lead to outputs that are bounded in energy. For stable plants the latter is immediately assured, but this of course not the case with unstable plants. However there will always be bounded inputs that lead to bounded outputs. Note that the graph of a system is exactly the same as the ℓ_2 behavior of the system, as defined by definition 2.4.5, but we use this terminology because it is custom in this context. Note that a plant characterizes its graph and vice versa. The following proposition shows how this notion is connected with the coprime factorization of the system G(z).

Proposition 3.7.2. [Vidyasagar, 1985]. Let G(z) be a p×m FDLT system and let (N,M) be a right coprime factorization of G(z) as in definition 2.8.2. Then the graph of G(z) equals:

$$\mathcal{G}\{G(z)\} = \{\begin{bmatrix}M\\N\end{bmatrix}w \mid w\in\ell_2^m\} \qquad (3.7.2)$$

o

The reduction problem we wish to address can be seen as approximation of $\mathcal{G}\{G(z)\}$ by $\mathcal{G}\{G_r(z)\}$ or equivalently as approximation of $\begin{bmatrix}M\\N\end{bmatrix}$ by $\begin{bmatrix}M_r\\N_r\end{bmatrix}$, where $G_r(z)$ is defined by $G_r=N_rM_r^{-1}$. This is motivated by the observation that such a procedure can be applied to both stable and unstable plants, which was not possible with the method that we considered before in this chapter. In the behavioral context of Willems (1986,1988), where systems are described by all possible signal trajectories, we can consider this method as an approximation of this behavior contrary to the methods of section 3.4 that in fact approximate a representation of the behavior of a system.

Because the coprime factors of a FDLT system G(z) are always stable, we can use any model reduction method of the previous sections to perform the reduction step. However it is important to note that it is not clear whether the approximant $\begin{bmatrix}M_r\\N_r\end{bmatrix}$ can be seen as a factorization of a causal FDLT system $G_r:=N_rM_r^{-1}$; for instance if M_r is strictly proper we may encounter the problem that G_r is not causal. We will show that, under weak

conditions, the pair (M_r,N_r) constitutes a factorization of a causal FDLT system $G_r(z)$ and though we are not able to proof that this pair is coprime, a realization of $G_r(z)$ can be obtained without major calculations.

We concentrate on the reduction methods based on balancing as defined in section 3.4. Assume therefore that we are given a FDLT system G(z) with a realization [A,B,C,D] and denote this by:

$$G(z) \approx \left[\begin{array}{c:c} A & B \\ \hdashline C & D \end{array}\right] \tag{3.7.3}$$

With the theory of section 2.8 we can create a normalized right coprime factorization $G = NM^{-1}$ with:

$$\begin{bmatrix} M \\ N \end{bmatrix} \approx \left[\begin{array}{c:c} A-BF & BH \\ \hdashline -F & H \\ \hdashline C-DF & DH \end{array}\right] \tag{3.7.4}$$

Assume that this realization is balanced with respect to a matrix $\Sigma = \mathrm{diag}\{\sigma_i\}$ where $\sigma_k > \sigma_{k+1}$. These Hankel singular values are referred to as the **Graph Hankel singular values** and they are all smaller than or equal to one, which was shown by Meyer (1988) for the continuous time case.

LEMMA 3.7.3. Let G(z) be a FDLT system and let (N,M) be a right normalized coprime factorization for G(z). The Hankel singular values of the FDLTS system $\begin{bmatrix} M(z) \\ N(z) \end{bmatrix}$ are smaller than or equal to one ○

Proof: Use the ω–transformation to transform the system (3.7.4) to the continuous time domain and apply theorem 5.2 of Glover (1984) to show that this system is a sub–system of an all–pass system with the same system matrix, which shows that the Hankel singular values are smaller or equal to one. Since the Hankel singular values are invariant under the ω–transformation it follows that $\sigma \leq 1$. □

The next step is the creation of the reduced order model:

$$\begin{bmatrix} M_r(z) \\ N_r(z) \end{bmatrix} := \mathcal{GB}_k^{\alpha}\left\{\begin{bmatrix} M(z) \\ N(z) \end{bmatrix}\right\} \tag{3.7.5}$$

with $\mathcal{GB}_k^{\alpha}$ defined by definition 3.4.1.

The question arises if (N_r,M_r) can be seen as a factorization of a causal system $G_r(z) := N_r(z)M_r(z)^{-1}$. This is true if the system (3.7.5) has again a realization of the form

$$\begin{bmatrix} M_r \\ N_r \end{bmatrix} \approx \left[\begin{array}{c:c} A_r-B_rF_r & B_rH_r \\ \hdashline -F_r & H_r \\ \hdashline C_r-D_rF_r & D_rH_r \end{array}\right] \tag{3.7.6}$$

with H_r a full rank matrix, and in that case we can retrieve a realization $[A_r,B_r,C_r,D_r]$ of $G_r(z)$ with minor calculations. The following lemma shows that under a weak condition the matrix H_r will always be of full rank.

LEMMA 3.7.4. Let G(z) be a p×m FDLT system with McMillan degree n and let [A,B,C,D] be a minimal realization of G(z). Let (N,M) be a right normalized coprime factorization of G(z) with a realization (3.7.4). Assume that the realization [A,B,C,D] is such that the realization (3.7.4) is balanced w.r.t. Σ, with $\sigma_k > \sigma_{k+1}$. Partition [A,B,C,D] according to (3.1.1) and let $\alpha \in \mathbb{R} \cup \{\pm\infty\}$ with $|\alpha| \geq 1$ and $\alpha \notin \sigma(A_{22})$. Define $\begin{bmatrix} M_r(z) \\ N_r(z) \end{bmatrix} := \mathcal{CB}_k^{\alpha}\{\begin{bmatrix} M(z) \\ N(z) \end{bmatrix}$, then this system has a realization of the form (3.7.6), where H_r is non singular and the system $G_r(z) := N_r(z)M_r(z)^{-1}$ has a realization $[A_r,B_r,C_r,D_r]$, which is not necessarily minimal. ○

PROOF: See appendix 3F

This lemma shows that we have to check that the value of α is not an eigenvalue of A_{22}. This matrix can only be calculated after the construction of (N,M) and can very well have unstable eigenvalues. If α is an eigenvalue of A_{22} then H_r will not be of full rank, hence M_r will not have a causal inverse and G_r will not causal.

REMARK 3.7.5. As a direct consequence of this reduction procedure we can give a bound for the error transfer function, created by using the presented method, in the Graph metric as used by Vidyasagar (1985) and Meyer (1988). Meyer shows that the error in the Graph metric is bounded by the L_∞ norm of the error transfer function $\begin{bmatrix} M(z) \\ N(z) \end{bmatrix} - \begin{bmatrix} M_r(z) \\ N_r(z) \end{bmatrix}$. Theorem 3.4.4 then shows that a bound for the graph metric is $2\times(\sigma_{k+1} + \cdots + \sigma_n)$ with strict inequality if $|\alpha| > 1$. Furthermore the experiences with the generalized balanced reduction, which were described in section 3.5, show that there will exist an optimal value of α, which can reduce the L_∞-norm of the error and hence the bound for the Graph metric considerably. We refer to Vidyasagar and Meyer for the consequences on the closed loop behavior and controller design. ○

REMARK 3.7.6. Note that in lemma 3.7.4 we did not state that $[A_r,B_r,C_r,D_r]$ is a minimal realization. This conclusion would be valid iff we could prove that (N_r,M_r) defines a coprime factorization. Definition 2.8.2 shows that this can be accomplished by showing that the Bezout-identity is valid and this requires that (A_r,C_r) is a stabilizable pair. We were not able to show this. Hence the McMillan degree of $G_r(z)$ is smaller or equal to k. Meyer (1988) shows that for the continuous time case the application of $\mathcal{CB}$ leads to a coprime factorization, which is even normalized. This can be used to show that the method we presented will have the same feature if we use $\alpha = -1$, by application of the ω-transformation. It should be noted that in the latter paper an extra condition is imposed, which states that the graph singular values are all distinct. We believe that this is a superfluous condition, which could be relaxed to $\sigma_k > \sigma_{k+1}$. We also believe that every α which obeys the conditions of lemma 3.7.4 leads to a coprime factorization,

which is normalized if $\alpha=\pm 1$, but we can not present proof for this hypothesis. Thus we can state that the above procedure leads to a reduced order model with McMillan degree smaller then or equal to k, with a not guaranteed minimal realization $[A_r,B_r,C_r,D_r]$ and our hypothesis is that these realizations are indeed minimal. o

HYPOTHESIS 3.7.7. The McMillan degree of G_r in lemma 3.7.4 is equal to k. o

REMARK 3.7.8. In this section we focused on discrete time systems; the same procedure can also be applied to continuous time systems. The results of this section are valid for the continuous time case, if the condition $|\alpha|\geq 1$ in lemma 3.7.4 is replaced by $\alpha\geq 0$. o

3.8 DISCUSSION

In this chapter we presented three new approaches to create low order models, as approximations of high order models. The first method, referred to as general balanced reduction, unifies well known model reduction methods, that are based on balanced realizations. This method creates a freedom in the design of low order approximations through the introduction of the reduction parameter. This design parameter can be used to optimize the frequency behavior of the approximation and we showed with three examples that such an optimization can yield a considerable improvement if compared to the standard methods. Furthermore, we presented conditions under which the realizations of the created approximations are guaranteed to be stable and minimal.
The second method is aimed at a solution to the optimal Hankel norm problem for discrete time systems, without applying solution methods for continuous time systems. We derived a set of sufficient equations, that determine a state space realization of a solution and showed that these equations yield the solution for MISO and SIMO systems. Furthermore, we presented a procedure that can be applied to solve the Hankel norm approximation problem for a broad class of MIMO systems. We did not succeed in solving the discrete Hankel norm approximation in terms of a closed form solution; however we derived a set of sufficient conditions explaining the structure of the solution method.
The third method combines the general balanced reductions method with the concept of coprime factorizations and we showed that this combination yields a model reduction procedure that can be applied to both stable and unstable systems and can be interpreted as an approximation of the behavior of a system rather than an approximation of a representation of a system.

3.9. APPENDICES OF CHAPTER 3.

Appendix 3A proof of lemma 3.4.3.

We consider the case p=z, i.e G(p)=G(z), a discrete time transfer function. Let $[A^d,B^d,C^d,D^d]$ be a realization of G(z), balanced w.r.t Σ and let $[A^c,B^c,C^c,D^c]$ the transform (2.5.3) under the ω–transformation. This realization of a continuous transfer function is again balanced w.r.t Σ (Glover, 1984). Let further $[\hat{A}^d,\hat{B}^d,\hat{C}^d,\hat{D}^d]$ be the realization of $\mathcal{GB}_k^{\alpha}(G)$, as in definition 3.4.1 and $[\hat{A}^c,\hat{B}^c,\hat{C}^c,\hat{D}^c]$ the transform of the latter realization under the ω–transformation. We have the following equations:

$$A^c = [I+A^d]^{-1}[A^d-I] = I-2[I+A^d]^{-1} \qquad \hat{A}^c = I-2[I+\hat{A}^d]^{-1} \qquad (3A.1a,b)$$

$$B^c = \sqrt{2}\,[I+A^d]^{-1}B^d \qquad \hat{B}^c = \sqrt{2}\,[I+\hat{A}^d]^{-1}\hat{B}^d \qquad (3A.1c.d)$$

$$C^c = \sqrt{2}\,C^d[I+A^d]^{-1} \qquad \hat{C}^c = \sqrt{2}\,\hat{C}^d[I+\hat{A}^d]^{-1} \qquad (3A.1e,f)$$

$$D^c = D^d-C^d[I+A^d]^{-1}B^d \qquad \hat{D}^c = \hat{D}^d-\hat{C}^d[I+\hat{A}^d]^{-1}\hat{B}^d \qquad (3A.1g,h)$$

We must show that $[\hat{A}^c,\hat{B}^c,\hat{C}^c,\hat{D}^c]$ is a realization of $\mathcal{GB}_k^{\beta}(H)$, where H(s) is the continuous system with realization $[A^c,B^c,C^c,D^c]$. This is achieved with pure matrix manipulation and we concentrate on the system matrices.

$\hat{A}^d = A_{11}^d + A_{12}^d[\alpha I-A_{22}^d]^{-1}A_{21}^d$ (from 3.4.4) and with this we calculate

$$[I+\hat{A}^d]^{-1} = \left[I + A_{11}^d + A_{12}^d[\alpha I-A_{22}^d]^{-1}A_{21}^d\right]^{-1}$$

$$= [I+A_{11}^d]^{-1}\left[I- A_{12}^d\left[A_{21}^d[I+A_{11}^d]^{-1}A_{12}^d + \alpha I-A_{22}^d\right]^{-1}A_{21}^d[I+A_{11}^d]^{-1}\right] \qquad (3A.2a)$$

Here we applied the matrix inversion lemma (Kailath,1980), which we also use for:

$$[I+A^d]^{-1} = \begin{bmatrix} \left[I+A_{11}^d-A_{12}^d\,[\,I+A_{22}^d]^{-1}A_{21}^d\right]^{-1} & -[I+A_{11}^d]^{-1}A_{12}^dS^{-1} \\ -S^{-1}A_{21}^d\,[\,I+A_{22}^d\,]^{-1} & S^{-1} \end{bmatrix} \qquad (3A.2b)$$

$$\text{where } S = I+A_{22}^d-A_{21}^d[I+A_{11}^d]^{-1}A_{12}^d \qquad (3A.2c)$$

Substitute (3A.2b) in the definition (3A.1a) of A^c:

$$A_{11}^c=I-2\left[(I+A_{11}^d)-A_{12}^d[I+A_{22}^d]^{-1}A_{21}^d\right]^{-1} \qquad A_{12}^c=2[I+A_{11}^d]^{-1}A_{12}^dS^{-1} \qquad (3A.3a)$$

$$A_{21}^c=2S^{-1}A_{21}^d[I+A_{11}^d]^{-1} \qquad A_{22}^c=I-2S^{-1} \qquad (3A.3b)$$

Let $\beta\in\mathbb{R}$ and consider the system matrix of $\mathcal{GB}_k^{\beta}(H)$: $A_{11}^c + A_{12}^c\left[\beta I-A_{22}^c\right]^{-1}A_{21}^c$, and substitute (3A.3) and (3A.2c):

$$= I - 2\left[(I+A_{11}^d)-A_{12}^d[I+A_{22}^d]^{-1}A_{21}^d\right]^{-1} + 4[I+A_{11}^d]^{-1}A_{12}^dS^{-1}\left[(\beta-1)I+2S^{-1}\right]^{-1}S^{-1}A_{21}^d[I+A_{11}^d]^{-1}$$

$$= I - 2[I+A_{11}^d]^{-1} -2[I+A_{11}^d]^{-1}A_{12}^d\left[-A_{21}^d[I+A_{11}^d]^{-1}A_{12}^d + I + A_{22}^d\right]^{-1}A_{21}^d[I+A_{11}^d]^{-1} +$$

$$+ 4[I+A_{11}^d]^{-1}A_{12}^dS^{-1}[(\beta-1)S+2I]^{-1}A_{21}^d[I+A_{11}^d]^{-1}$$

$$= I - 2[I+A_{11}^d]^{-1} + 2[I+A_{11}^d]^{-1}A_{12}^dKA_{21}^d[I+A_{11}^d]^{-1} \qquad (3A.4)$$

where we define K as:

$$\begin{aligned} K &= -\left[-A_{21}^{d}[I+A_{11}^{d}]^{-1}A_{12}^{d} + I + A_{22}^{d}\right]^{-1} + 2S^{-1}[(\beta-1)S+2I]^{-1} \\ &= S^{-1}\left[-I + 2\left[(\beta-1)S+2I\right]^{-1}\right] = (1-\beta)\left[(\beta-1)S+2I\right]^{-1} \\ &= \left[-S + 2(1-\beta)^{-1}I\right]^{-1} = \left[\frac{\beta+1}{1-\beta}I - A_{22}^{d} + A_{21}^{d}[I+A_{11}^{d}]^{-1}A_{12}^{d}\right]^{-1} \end{aligned} \tag{3A.5}$$

Now we have already seen (3A.1b) that $\hat{A}^c = I - 2[I+\hat{A}^d]^{-1}$ or:

$$\hat{A}^c = I - 2\left[I+A_{11}^{d}\right]^{-1}\left[I - A_{12}^{d}\left[A_{21}^{d}\left[I+A_{11}^{d}\right]^{-1}A_{12}^{d} + \alpha I - A_{22}^{d}\right]^{-1}A_{21}^{d}\left[I+A_{11}^{d}\right]^{-1}\right] \tag{3A.6}$$

A close look at the equations (3A.4–6) reveals that:

$$\hat{A}^c = A_{11}^{c} + A_{12}^{c}\left[\beta I - A_{22}^{c}\right]^{-1}A_{21}^{c} \quad \text{for } \alpha = \frac{\beta+1}{1-\beta} \text{ or } \beta = \frac{\alpha-1}{\alpha+1} \tag{3A.7}$$

Analogous calculations show that

$$\hat{B}^c = B_1^c + A_{12}^{c}\left[\beta I - A_{22}^{c}\right]^{-1}B_2^c \qquad \hat{C}^c = C_1^c + C_2^{c}\left[\beta I - A_{22}^{c}\right]^{-1}A_{21}^{c} \tag{3A.8a}$$

$$\hat{D}^c = D^c + C_2^{c}\left[\beta I - A_{22}^{c}\right]^{-1}B_2^c \qquad \text{for } \alpha = \frac{\beta+1}{1-\beta} \text{ or } \beta = \frac{\alpha-1}{\alpha+1} \tag{3A.8b}$$

Conclusion from (3A.7,8): $[\hat{A}^c,\hat{B}^c,\hat{C}^c,\hat{D}^c]$ is a realization of $\mathcal{GB}_k^{\beta}\left[\omega(G)\right]$ if $\beta = \frac{\alpha-1}{\alpha+1}$

which proofs the first statement of the lemmma. The proof of the second part is a direct result the above since the ω-transformation is bijective and $\beta = \frac{\alpha-1}{\alpha+1}$ implies $\alpha = \frac{\beta+1}{1-\beta}$ and reversing α and β gives the result. □

Appendix 3B. proof of theorem 3.4.4.

Part 1 – Stability

Let G be discrete. Corollary 3.3.8 shows that the case $|\alpha|=1$ is one of the standard methods for which stability was already proven. So let $|\alpha|>1$. From Pernebo and Silverman (1983) we know that A_{22} is stable, so $\alpha \notin \sigma(A_{22})$ and thus $\hat{A}$ is well defined. Now suppose that $\hat{G}$ is not stable, so:

$$\exists x \in \mathbb{F}^n,\ \lambda \in \mathbb{C},\ x \neq 0,\ |\lambda| \geq 1 \text{ with } \hat{A}x = \lambda x. \tag{3B.1}$$

We will show that this leads to a contradiction.

$$(3B.1) \Rightarrow \left[A_{11} + A_{12}[\alpha I - A_{22}]^{-1}A_{21}\right]x = \lambda x \Rightarrow \left[A_{11} \mid A_{12}\right]\begin{bmatrix} I \\ [\alpha I - A_{22}]^{-1}A_{21} \end{bmatrix}x = \lambda x \tag{3B.2}$$

$$\left[A_{21} \mid A_{22}\right]\begin{bmatrix} I \\ [\alpha I - A_{22}]^{-1}A_{21} \end{bmatrix} = \left[I + A_{22}[\alpha I - A_{22}]^{-1}\right]A_{21} = \alpha[\alpha I - A_{22}]^{-1}A_{21} \tag{3B.3}$$

Combine (3B.2) and (3B.3): $A\begin{bmatrix} I \\ [\alpha I - A_{22}]^{-1}A_{21} \end{bmatrix}x = \begin{bmatrix} \lambda I \\ \alpha[\alpha I - A_{22}]^{-1}A_{21} \end{bmatrix}x$ (3B.4)

Let $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} := \begin{bmatrix} I \\ [\alpha I - A_{22}]^{-1}A_{21} \end{bmatrix}x$ then (3B.4) reads: $Ay = \begin{bmatrix} \lambda & 0 \\ 0 & \alpha \end{bmatrix}y.$ (3B.5)

Note that $y_1 \neq 0 \neq y_2$ since $y_1 = x \neq 0$ (3B.1) and if $y_2 = 0$ then (3B.2) shows $A_{11}x = \lambda x$ but A_{11} is stable (Pernebo and Silverman ,1983), so $y_2 \neq 0$.

Apply the result of Pernebo and Silverman (1983) that $\|A\|_s \leq 1$:

$$\|Ay\|_2 \leq \|A\|_s \|y\|_2 \leq \|y\|_2 \tag{3B.6}$$

and note that $$\left\| \begin{bmatrix} \lambda & 0 \\ 0 & \alpha \end{bmatrix} y \right\|_2 \geq \|y\|_2 \tag{3B.7}$$

with equality iff $|\lambda| = |\alpha| = 1$, since $y_1 \neq 0 \neq y_2$ and $|\alpha|, |\lambda| \geq 1$.

Combine (3B.5,6,7) and conclude that $|\alpha| = 1$, which is in contradiction with the assumption $|\alpha| > 1$. This shows that $\hat{G}(z)$ is stable for $|\alpha| \geq 1$. The stability of the continuous time equivalent follows from proposition 2.5.1 and lemma 3.4.3. This proves the stability of $\hat{G}$ for $\alpha \in AR$. (stability)□

Part 1 – Minimality

Consider the continuous time case. The case $\alpha = 0$ is covered in proposition 3.2.5. Pernebo and Silverman (1983) proved the minimality of $\{\hat{A},\hat{B},\hat{C}\}$ for the discrete time case, with $\alpha = \infty$, which with lemma 3.4.3 shows the correctness for the continuous time case for $\alpha = 1$. The correctness for $\alpha = \infty$ (continuous time) is given in proposition 3.2.2. Thus we have minimality for $\alpha = 0,1,\infty$ (in he continuous time case).

Now let $0 < \alpha < \infty$ and define: $[\tilde{A},\tilde{B},\tilde{C},\tilde{D}] := [\alpha^{-1}A, \alpha^{-\frac{1}{2}}B, \alpha^{-\frac{1}{2}}C, D]$. (3B.8)

It is easy to see that this realization is still balanced w.r.t. Σ and stable. We just showed that the reduction of such a system with $\alpha = 1$ gives a stable minimal approximation of (3B.8) with a realization $[\hat{\tilde{A}},\hat{\tilde{B}},\hat{\tilde{C}},\hat{\tilde{D}}]$ where:

$$\hat{\tilde{A}} = \tilde{A}_{11} + \tilde{A}_{12}[I - \tilde{A}_{22}]^{-1}\tilde{A}_{21} = \alpha^{-1}\left[A_{11} + A_{12}[\alpha I - A_{22}]^{-1}A_{21}\right] = \alpha^{-1}\hat{A} \tag{3B.9a}$$

$$\hat{\tilde{B}} = \tilde{B}_1 + \tilde{A}_{12}[I - \tilde{A}_{22}]^{-1}\tilde{B}_2 = \alpha^{-\frac{1}{2}}\left[B_1 + A_{12}[\alpha I - A_{22}]^{-1}B_2\right] = \alpha^{-\frac{1}{2}}\hat{B} \tag{3B.9b}$$

$$\hat{\tilde{C}} = \tilde{C}_1 + \tilde{C}_2[I - \tilde{A}_{22}]^{-1}\tilde{A}_{21} = \alpha^{-\frac{1}{2}}\left[C_1 + C_2[\alpha I - A_{22}]^{-1}A_{21}\right] = \alpha^{-\frac{1}{2}}\hat{C} \tag{3B.9c}$$

Because $\{\alpha^{-1}\hat{A}, \alpha^{-\frac{1}{2}}\hat{B}, \alpha^{-\frac{1}{2}}\hat{C}\}$ is minimal the Popov–Belevitch–Hautus test (Kailath, 1980) shows the minimality of $[\hat{A},\hat{B},\hat{C},\hat{D}]$. Consequently we have proven the minimality for continuous time systems for $0 \leq \alpha \leq \infty$. The minimality of the discrete counterpart follows from proposition 2.5.1 and lemma 3.4.3. (minimality)□

Part 2.

Let E(p) be the error transfer function: $E(p) := G(p) - \hat{G}(p)$, with $\hat{G}(p) = \mathcal{GB}_k^{\alpha}(G)$.

Our aim is to proof that $\|E(p)\|_\infty \leq 2\sum_{k+1}^{n} \sigma_i$ with strict inequality if α is in the interior of *AR*. The cases with G discrete and $\alpha = -1,1,\infty$ are proven by Al-Saggaf and Franklin (1987,1988). The ω-transformation then gives the corresponding bounds for G continuous and $\alpha = 0,1,\infty$. Now let G(s) be a continuous time system,

$0<\alpha<\infty$ and let $\hat{G}(s):=\mathcal{GB}_k^{\alpha}(G)$. Define $\tilde{G}(s):=G(\alpha s)$ and $\tilde{\hat{G}}(s)=\hat{G}(\alpha s)$. Note that (3B.8) defines a stable realization of $\tilde{G}$, balanced with respect to Σ and that $\tilde{\hat{G}}(s)$ has a realization $[\alpha^{-1}\hat{A},\alpha^{-\frac{1}{2}}\hat{B},\alpha^{-\frac{1}{2}}\hat{C},\hat{D}]$. It is straightforward that $\tilde{\hat{G}}(s)=GB_k^{\downarrow}(\tilde{G})$ and hence:

$$\|G(s)-\hat{G}(s)\|_\infty = \|G(\alpha s)-\hat{G}(\alpha s)\|_\infty = \|\tilde{G}(s)-\tilde{\hat{G}}(s)\|_\infty < 2\sum_{k+1}^{n}\sigma_i.$$

This completes the proof for G continuous and $0<\alpha<\infty$ and thus also for $0\leq\alpha\leq\infty$. Lemma 3.4.3 and the properties of the ω-transformation now give the proof for G discrete and $\alpha\leq -1$ or $\alpha\geq 1$ and hence we have proven part 2 which completes the proof of theorem 3.5.4. □

Appendix 3C proof of lemma 3.6.5.

Apply theorem 2.7.4 on the realization $[A_e,B_e,C_e,D_e]$ and obtain the following equations:

$$A_eP_eA_e^* + B_eB_e^* = P_e \qquad A_eP_eC_e^* + B_eD_e^* = 0 \qquad \text{(3C.1a,b)}$$

$$A_e^*Q_eA_e + C_e^*C_e = Q_e \qquad A_e^*Q_eB_e + C_e^*D_e = 0 \qquad \text{(3C.2a,b)}$$

$$D_eD_e^* + C_eP_eC_e^* = \sigma^2 I \qquad D_e^*D_e + B_e^*Q_eB_e = \sigma^2 I \qquad \text{(3C.3a,b)}$$

We concentrate on (3C.1) and evaluate both equations:

(3C.1a): $$\begin{bmatrix} A\Sigma A^* & * \\ \hat{A}[\,I\;\;0]A^* & \hat{A}\Sigma_1\Gamma^{-1}\hat{A}^* \end{bmatrix} + \begin{bmatrix} BB^* & * \\ \hat{B}B^* & \hat{B}\hat{B}^* \end{bmatrix} = \begin{bmatrix} \Sigma & * \\ [I\;\;0] & \Sigma_1\Gamma^{-1} \end{bmatrix}$$

The lower block row results in two equations:

$$\hat{A}[I\;0]A^*+\hat{B}B^* = [I\;0] \qquad \hat{A}\Sigma_1\Gamma^{-1}\hat{A}^*+\hat{B}\hat{B}^*=\Sigma_1\Gamma^{-1} \qquad \text{(3C.4a,b)}$$

(3C.1b): $$\begin{bmatrix} A\Sigma C^*+A\begin{bmatrix} I \\ 0 \end{bmatrix}\hat{C}^* \\ \hat{A}[I\;\;0\,]\,C^*+\hat{A}\Sigma_1\Gamma^{-1}\hat{C}^* \end{bmatrix} + \begin{bmatrix} BD_e^* \\ \hat{B}D_e^* \end{bmatrix} = 0 \qquad \begin{matrix} \text{(3C.5a)} \\ \text{(3C.5b)} \end{matrix}$$

Combine (3C.4a) and (3C.5a) and conclude that they are equivalent to (3.6.20a). The equations (3C.5b) and (3C.4b) are the same as respectively (3.6.20b) and (3.6.20c). A similar argument shows that (3C.3a,b) are equivalent with (3.6.21).

It remains to deal with the equations (3C.3a,b). The following argument shows that they follow from the other equations and the rank conditions on B and C:

(3C.1b): $A_eP_eC_e^*+B_eD_e^*=0 \quad \Rightarrow \quad A_eP_eC_e^*D_e+B_eD_e^*D_e=0$

$\xRightarrow{\text{3C.2b}}$ $-A_eP_eA_e^*Q_eB_e+B_eD_e^*D_e=0$ $\xRightarrow{\text{3C.2a}}$ $B_eB_e^*Q_eB_e-\sigma^2B_e+B_eD_e^*D_e=0$

$\Rightarrow$ $B_e(B_e^*Q_eB_e-\sigma^2I+D_e^*D_e)=0$ and the rank condition on B implies that B_e has full column rank and hence $B_e^*Q_eB_e+D_e^*D_e=\sigma^2I$ which is (3C.3a). Equation (3C.3b) follows from the same line of reasoning. This completes the proof. □

Appendix 3D. proof of proposition 3.6.7.

We denote equation (3.6.20a) as $L=R\Theta$ and let n_1 be the dimension of A_{11} and Σ_1 and r be the multiplicity of σ, thus $n_1=n-r$. The dimensions of $L=R\Theta$ are

$$L: n\times(n_1+1) \qquad R: n\times(n_1+1) \qquad \Theta: (n_1+1)\times(n_1+1)$$

This shows that it is sufficient to show that R has rank (n_1+1), because $n_1+1=n-r+1\leq n$ and in that case we find that R^*R is invertible. This is proven if we can show that L has rank (n_1+1). The identity matrix in the left upper part of L has dimensions $n_1\times n_1$, and therefore it is sufficient to show that the last r rows of the right part of L are not identically zero. This will be proven by contradiction. Since R does not depend on D, we may choose D freely and this shows that – under the condition that $B_2\neq 0$ – we can always find a D such that L has rank (n_1+1). Hence we only have to consider the case that $B_2=0$ and we must contradict that $[0\ I_r]A\Sigma C^*=0$.

Hypothesis: $B_2=0$ and $[0\ I_r]A\Sigma C^*=0$ (3D.1)

$$\Rightarrow [0\ I_r]\ A\Sigma C^*C = 0 \qquad \Rightarrow [0\ I_r]\ A\Sigma\ (A^*\Sigma A-\Sigma) = 0$$

$$\Rightarrow [0\ I_r]\left[(A\Sigma A^*)\Sigma A-A\Sigma^2\right] = 0 \qquad \Rightarrow [0\ I_r]\ \left[(\Sigma-BB^*)\Sigma A-A\Sigma^2\right] = 0$$

$$\Rightarrow [0\ I_r]\ (\Sigma^2A-A\Sigma^2) = 0 \qquad (\text{ since } [0\ I_r]B=B_2=0\ )$$

$$\Rightarrow [\sigma^2A_{21}\ \ \sigma^2A_{22}] - [A_{21}\Sigma_1\ \ \sigma^2A_{22}] = 0 \tag{3D.2}$$

$$\Rightarrow \sigma^2A_{21}-A_{12}\Sigma_1=0 \Rightarrow A_{21}\Gamma = 0 \Rightarrow A_{21} = 0 \tag{3D.3}$$

where we used the regularity of Γ. But now (3D.1) and (3D.3) show that $A_{21}=B_2=0$ which is in contradiction with the controllability of $\{A,B\}$. Hence we conclude that the hypothesis (3D.1) is incorrect and that L has rank (n_1+1). This completes the proof. □

Appendix 3E proof of proposition 3.6.9.

1. That $[A_1,B_1,C_1,D_1]$ is a realization of $G_1(z)$ is the result of straightforward matrix calculation. If $G(z)+D$ has no invariant zeros in $z=0$ then we know that the matrix $\begin{bmatrix}A & B\\ C & D\end{bmatrix}$ must be regular. Now it is easy to see that $Q_1=\begin{bmatrix}A & B\\ C & D\end{bmatrix}\begin{bmatrix}\Sigma & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}A & B\\ C & D\end{bmatrix}^*$ and this is thus a regular matrix. P_1 and Q_1 full rank implies that we are dealing with a minimal realization.
2. Follows from (1.) by application of definition 2.4.2.
3. For the proof of (3.) we introduce the matrix M:

$$M:=\begin{bmatrix}\Sigma_1^2 & 0 & K_1\\ 0 & \sigma^2 I_r & K_2\\ K_1^* & K_2^* & L\end{bmatrix} \quad \text{with:} \quad \begin{aligned} K_1&= [\ \Sigma_1^{\frac{1}{2}} \ \ |\ \ 0_r\ ](A^*\Sigma B+C^*D)\\ K_2&= [\ 0_{n-r}\ |\ \ \sigma^{\frac{1}{2}}I_r\ ](A^*\Sigma B+C^*D)\\ L&= D^*D+B^*\Sigma B\end{aligned} \tag{3.6.24}$$

3a.

If $K_2=0$ then it is immediate that M has an eigenvalue σ^2 with multiplicity $r_1 \geq r$ and that the r columns of the matrix $Y:=[0_{n-r}, I_r, 0_{n-r}]^*$ (3E.1) are corresponding eigenvectors.

3b.

We will show that M has r eigenvectors, corresponding with σ^2, that are independent of Y (3E.1). Define the matrix $X:=\begin{bmatrix} X_1 \\ 0 \\ X_2 \end{bmatrix}:=\begin{bmatrix} \Sigma_1^{\frac{1}{2}} A_{21}^* \\ 0 \\ B_2^* \end{bmatrix}$ (3E.2)

where we use the partitioning (3.6.12)). Note that X is linearly independent of Y. We want to show that $MX=\sigma^2 X$ and that X has rank r.

Evaluate

$$(A^*\Sigma B+C^*D)X_2 = (A^*\Sigma BB^*+C^*DB^*)\begin{bmatrix} 0 \\ I_r \end{bmatrix} = (A^*\Sigma^2-A^*\Sigma A\Sigma A^*+C^*DB^*)\begin{bmatrix} 0 \\ I_r \end{bmatrix}$$

$$= (A^*\Sigma^2-\Sigma^2 A^*+C^*C\Sigma A^*+C^*DB^*)\begin{bmatrix} 0 \\ I_r \end{bmatrix} = (\text{apply (3.6.23e)}) = (A^*\Sigma^2-\Sigma^2 A^*)\begin{bmatrix} 0 \\ I_r \end{bmatrix}$$

This implies that $[0\ I_r](A^*\Sigma B+C^*D)X_2 = (\sigma^2 A_{21}^* - \Sigma_1^2 A_{21}^*)$ (3E.3)

With (3E.3) we can deduce that $\Sigma_1^2 X_1 + K_1 X_2 = \sigma^2 X_1$ (3E.4)

A similar argument shows that $K_1^* X_1 + L X_2 = \sigma^2 X_2$ (3E.5)

and the equations (3E.4,5) imply that $MX=\sigma^2 X$. From Pernebo and Silverman (1983) we deduct that X has rank r: $X^*X = A_{21}\Sigma_1 A_{21}^* + B_2 B_2^* = \sigma I_r - \sigma A_{22} A_{22}^* > 0$. Hence the 2r columns of [Y X] are independent eigenvectors that correspond to the eigenvalue σ^2.

3c.

From part (3a.) we know that the matrix M has an eigenvalue σ^2 with multiplicity $r_1 \geq r$. If $r_1 \geq (r+m)$ then the matrix N must have an eigenvalue σ^2 with Mcmillan degree m, where $N:=\begin{bmatrix} \Sigma_1^2 & K_1 \\ K_1^* & L \end{bmatrix}$. This is fulfilled if there exists $\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$ with rank m such that

$$\begin{cases} \Sigma_1^2 X_1 + K_1 X_2 = \sigma^2 X_1 \\ K_1^* X_1 + L X_2 = \sigma^2 X_2 \end{cases} \Leftrightarrow \begin{cases} X_1 = -\Gamma^{-1} K_1 X_2 \\ K_1^* X_1 = (\sigma^2 - L) X_2 \end{cases} \Leftrightarrow (K_1\Gamma^{-1}K_1 + \sigma^2 - L)X_2 = 0 \quad (3E.6)$$

We choose $X_2=I_m$, which assures that the rank condition is fullfiled. A straightforward calculation now shows that the resulting condition (3E.6) $(K_1\Gamma^{-1}K_1+\sigma^2-L)=0$ is equivalent with the Riccati equation (3.6.23g). □

Appendix 3F proof of lemma 3.7.4.

First we show that H_r is of full rank. The feedthrough matrix $\hat{D}$ of $\begin{bmatrix} M_r(z) \\ N_r(z) \end{bmatrix}$ follows from (3.4.1e): $\hat{D} := \begin{bmatrix} \hat{D}_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} H \\ DH \end{bmatrix} + \begin{bmatrix} -F \\ C-DF \end{bmatrix}_2 [\alpha I-(A-BF)_{22}]^{-1} (BH)_2$

where the index refers to the partitioning of the matrices as in 3.1.1. We want to show that $\hat{D}_1$ is nonsingular:

$$\hat{D}_1 = H - F_2 [\alpha I-A_{22}+B_2F_2]^{-1}B_2H$$

Since H is of full rank this is equivalent with $\hat{D}_1H^{-1}$ full rank:

$$\hat{D}_1H^{-1} = I - F_2 [\alpha I-A_{22}+B_2F_2]^{-1}B_2$$

Suppose $\hat{D}_1H^{-1}$ is singular:

$\Rightarrow \exists x \neq 0,\ x=F_2[\alpha I-A_{22}+B_2F_2]^{-1}B_2x$, and thus $B_2x \neq 0$. Pre-multiply with B_2:

$\Rightarrow \quad (B_2x)=B_2F_2[\alpha I-A_{22}+B_2F_2]^{-1}(B_2x)$

$\Rightarrow I-B_2F_2 [\alpha I-A_{22}+B_2F_2]^{-1}$ singular

$\Rightarrow [\alpha I-A_{22}][\alpha I-A_{22}+B_2F_2]^{-1}$ has not full rank.

If $|\alpha| \geq 1$ then the matrix on the right is always full rank (theorem 3.4.4) and thus this expression can only be singular if $[\alpha I-A_{22}]$ is singular. This shows that $\hat{D}_1$ is non singular and thus we can calculate the matrices $\{A_r,B_r,C_r,D_r\}$ from (3.7.6). Use the same argument as in in the proof of theorem 2.8.3 (Appendix 2C, part 1, step 2) to show that $[A_r,B_r,C_r,D_r]$ is a realization of $G_r(z)$. □

4. APPROXIMATIONS WITH LAGUERRE POLYNOMIALS

4.1 INTRODUCTION

As explained in chapter one, the aim of this thesis is the development of theory and methods to deal with the problem of approximate modeling. In the previous chapter we dealt with the approximation on the basis of model descriptions. For approximate system identification we have to consider the modeling on the basis of data, in general measured input/output series of a plant or system. The most applied (approximate) identification methods aim at the minimization of the one step prediction error. For these PE methods few results are known on the character of the approximation (Ljung, 1987, 1989; Wahlberg and Ljung, 1985; Swaanenburg and co–workers, 1985; Van den Hof and Janssen, 1987). These methods have the drawback that the algorithms are often based on the use of complex non–linear optimization techniques, without a guaranty that an optimal solution will be found. Furthermore these methods tend to emphasize the high frequency characteristics of the system. This problem can be overcome by the application of prefilters, but the optimal prefilter is implicitly connected with the optimal solution in the sense that the quality of the prefilter can only be evaluated after the identification experiment has been performed (Ljung, 1987).

We aim at alternative methods with the property that satisfactory results can be obtained with the use of pre–specified prefilters and application of simple optimization techniques. From the many possible approaches towards this goal we choose to study the use of orthogonal functions in the context of approximation and identification. This is motivated by the fact that the concepts of orthogonality and approximation are strongly connected. An example of this is the approximation with the aid of projection methods, which for instance is the basis of Kalman filtering theory (Anderson and Moore, 1971). Another reason is that with a specific set of orthogonal functions we may be able to decompose the mapping of a system (think of impulse responses) into basic components and that it will be possible to retrieve those components that have the largest influence on the system behavior. A third motivation comes from the observation that orthogonality is intrinsic to various model approximation methods; consider for instance balanced realizations and the singular value decomposition of a Hankel matrix.

The use of discrete signals and digital computers in identification processes motivates the restriction of attention to discrete time orthogonal functions. There are many sets of such functions known and especially the various sets of orthogonal polynomials have been applied extensively in the context of system identification (Paraskevopoulos, 1985). We choose to study the set of Laguerre polynomials, because they are defined on an infinite interval which implies that we can study the mapping of time series that in principle also deal with time series of infinite length, such as the impulse response. Furthermore these functions can be generalized to a broad class of orthogonal functions, that are directly based on linear systems. This generalization will be the subject of chapter five.

The application of Laguerre functions to system identification has been considered in the early work of Lee (1933) and Wiener (1949) and more recently it has been applied by King and Paraskevopoulos (1977, 1979), Nurges and Yaaksoo (1981), Nurges (1987) and Wahlberg (1989). The Laguerre polynomials constitute an orthonormal basis for the space of ℓ_2 functions and can thus be used for the purpose of data transformation as a generalization of the Fourier transformation. Therefore we will talk about the Laguerre domain as an equivalent of the frequency domain. This leads to alternative descriptions of ℓ_2 time series in the Laguerre domain and if this is applied to the input/output series of a FDLTS system then it induces alternative system descriptions in the orthogonal domain.
These properties show that next to identification in the frequency domain and in the time domain we can also consider identification in the Laguerre domain. The application of this idea leads to alternative approximate identification methods with new design variables. It will follow that prefiltering is an intrinsic part of these methods. In chapter five we will extend the ideas and concepts of this chapter to a more general class of orthonormal functions, which will lead to a better understanding of the use of these design variables.

In section 4.2 we will define the Laguerre polynomials, both in time and frequency domain terms. The transformation of time series is the subject of section 4.3 and in section 4.4 it is shown how the calculations for such a transformation can actually be performed. It is explained that in general this will have to be an approximation. If the transformation of time series is applied to the input/output variables of a linear system then this gives rise to an alternative description of the system in the Laguerre domain. This Laguerre system transformation is the subject of section 4.5. In section 4.6 the properties of these transformations are used to derive three different approaches to deal with the problem of approximate system identification, which will be illustrated with examples. In section 4.7 we derive a continuous time analog of the Laguerre system

transformation and present a commutation diagram that connects the Laguerre transformations, the ω–transformations and two model reduction methods. The results, presented in this chapter will be reviewed in section 4.8 and section 4.9 contains appendices with the proofs of some of the results.

4.2 LAGUERRE POLYNOMIALS

In this section we formally define the finite difference Laguerre polynomials and show how they are derived form their continuous counterpart. Furthermore we present a set of a recursion formulas that enable us to calculate these functions. A third way to describe these functions is by considering them as impulse responses of certain transfer functions. First we show how the continuous time Laguerre polynomials are the solutions to a specific set of differential equations. The discrete time polynomials are then defined as the discretized version of the latter functions and in the remainder of this chapter we will only deal with the discrete time Laguerre polynomials.

We consider a function P: $\mathbb{R}\to\mathbb{R}$ to be polynomial of degree n if P(t) can be expressed as:

$$P(t) = p_n t^n + p_{n-1}t^{n-1} + \ldots + p_1 t + p_0 \qquad (4.2.1)$$

The *continuous time Laguerre polynomials* are solutions of the following differential equation (Nurges, 1987):

$$t\ddot{L}_k(t) + (1-t)\dot{L}_k(t) + kL_k(t) = 0 \qquad k\in\mathbb{N}^0 \qquad (4.2.2a)$$

with solutions [1]:

$$L_k(t) := \frac{e^t}{k!}\frac{d^k}{dt^k}(t^k e^{-t}) \qquad k\in\mathbb{N}^0 \qquad (4.2.2b)$$

These functions have been modified (Chang,Chen and Wang, 1985) with an extra time scaling parameter $\beta>0$ to the following functions:

$$L_k(t,\beta) := \frac{e^{\beta t}}{k!}\frac{d^k}{dt^k}(t^k e^{-\beta t}) \qquad k\in\mathbb{N}^0 \qquad (4.2.3a)$$

Hence for $k\in\mathbb{N}^0$ this definition results in a function L_k: $\mathbb{R}\times\mathbb{R}_+ \to \mathbb{R}$, with for instance

$$L_0(t,\beta)=1 \qquad L_1(t,\beta)=1-\beta t \qquad L_2(t,\beta)=1-2\beta t+\tfrac{1}{2}\beta^2 t^2 \qquad (4.2.3b)$$

and it can readily be shown that $L_k(t,\beta)$ is a polynomial in t of degree k.
For a fixed value of β these functions are mutually orthogonal on the interval $[0,\infty)$ with respect to the weighted L_2–inner product with weighting function $e^{-\beta t}$:

[1] Under specific initial conditions.

$$K_{ij}^{\beta} \int_0^{\infty} e^{-\beta t} L_i(t,\beta) L_j(t,\beta)\, dt = \begin{cases} 0 & i \neq j \\ 1 & i=j \end{cases} \tag{4.2.4}$$

$$\text{where } K_{ij}^{\beta} \in \mathbb{R},\ K_{ij}^{\beta} \neq 0$$

The *finite difference Laguerre polynomials* are derived from expression (4.2.3a) and in the sequel we will deal with the normalized version of these discrete time polynomials. However, in order to define them we need a definition of the unnormalized version.

Definition 4.2.1 [Nurges and Yaaksoo, 1981] The *unnormalized discrete time Laguerre polynomials* γ_k: $\mathbb{N}^0 \times \mathbb{R}_+ \to \mathbb{R}$ are defined by:

$$\gamma_k(t,\beta) := e^{\beta t} \Delta^k [\, \binom{t}{k} e^{-\beta t} \,] \tag{4.2.5}$$

where $k,t \in \mathbb{N}^0$ Δ is the difference operator, $(\Delta f)(t) := f(t) - f(t-1)$

$\beta \in \mathbb{R}_+$ $\binom{t}{k}$ is the binomial coefficient. ○

If β is kept constant then these functions are again mutually (weighted) orthogonal on $\mathbb{N}^0$; in this thesis we will work with the normalized version, which will be called the Laguerre polynomials hereafter. These normalized polynomials and a more explicit form are given in the following definition and proposition.

Definition 4.2.2 [Nurges and Yaaksoo, 1981] Let $\{\gamma_k, k \in \mathbb{N}^0\}$ be the set of unnormalized Laguerre polynomials (4.2.5). The (*discrete time normalized*) *Laguerre polynomials* with *discount factor* ξ are defined as ψ_k^{ξ}: $\mathbb{N}^0 \to \mathbb{R}$, with

$$\psi_k^{\xi}(t) := (-1)^k \sqrt{(1-\xi^2)} \cdot \frac{\gamma_k(t,\beta)}{\xi^{2k-t}} \qquad k \in \mathbb{N}^0 \tag{4.2.6a}$$

$$\text{where } \beta = -\ln(\xi^2) \quad (\xi = \sqrt{e^{-\beta}}) \tag{4.2.6b}$$

○

Remark 4.2.3. Note the difference between the definitions of γ_k and ψ_k^{ξ}; while γ_k is a function of t and β, we let ψ_k^{ξ} be only a function of t, emphasizing that this definition generates for every $\xi \in (0,1)$ a set of functions. In the sequel we will often omit the index ξ, writing ψ_k instead of ψ_k^{ξ}, with the implication that we assume this parameter ξ to be fixed. Whenever we refer to *the set of Laguerre polynomials*, it must be understood that we consider the set of functions $\{\psi_k^{\xi} : \mathbb{N}^0 \to \mathbb{R},\ k \in \mathbb{N}^0\}$ for a specific value of ξ. It follows from (4.2.6b) that $\xi \in (0,1)$. ○

PROPOSITION 4.2.4. [Nurges and Yaaksoo, 1981]. An explicit form for the normalized Laguerre polynomials (4.2.6a) is given by:

$$\psi_k^\xi(t) = \sqrt{(1-\xi^2)} \sum_{j=0}^{k} \left[(-1)^{k+j} \binom{k}{j} \binom{t+k-j}{k} \xi^{t+k-2j} \right] \tag{4.2.7}$$

which leads to the following recursion formulas:

$$\psi_0(0) = \sqrt{(1-\xi^2)} \tag{4.2.8a}$$

$$\psi_{k+1}(0) = -\xi\psi_k(0) \tag{4.2.8b}$$

$$\psi_0(t+1) = \xi\psi_0(t) \tag{4.2.8c}$$

$$\psi_k(t+1) = \xi\psi_k(t) + (1-\xi^2)\cdot \sum_{i=0}^{k-1} (-\xi)^{k-i-1}\psi_i(t) \tag{4.2.8d}$$

○

With the equations (4.2.8) we can easily calculate the Laguerre polynomials, which will be necessary for the application to system identification that will be discussed in section 4.6. For these calculations we will also use an alternative way to define these polynomials, that fits in nicely within the framework of system theory; this alternative is based on considering $\{\psi_k^\xi(t), t\in\mathbb{N}^0\}$ as the set of Markov parameters (2.4.4) of a specific transfer function:

PROPOSITION 4.2.5. [King and Paraskevopoulos, 1977,1979]. Let $\{\psi_k^\xi\}$ be defined by definition 4.2.2. Then ψ_k^ξ is the impulse response of the *generating transfer function* $\Psi_k^\xi(z)$, or $\{\psi_k^\xi(t), t\in\mathbb{N}^0\}$ is the set of Markov parameters of $\Psi_k^\xi(z)$ [2], where

$$\Psi_k^\xi(z) = \sqrt{(1-\xi^2)}\cdot z\frac{(1-z\xi)^k}{(z-\xi)^{k+1}} \tag{4.2.9}$$

○

Again we will often write $\Psi_k(z)$ or Ψ_k instead of $\Psi_k^\xi(z)$. It is straightforward to show that ψ_k is the impulse response of Ψ_k. Note that the equations (4.2.8) and (4.2.9) show a certain recursive structure in the sequence of Laguerre polynomials. This will be of intrinsic importance in the remainder of this chapter.

Because of the normalization these Laguerre polynomials are mutually *orthonormal* on the discrete interval $[0,\infty)$ w.r.t the ℓ_2–inner product:

$$\sum_{t=0}^{\infty} \psi_i(t)\psi_j(t) = \delta_{ij} \tag{4.2.10}$$

where δ_{ij} is the Kronecker delta. They form a complete basis of $\ell_{2[0,\infty)}$ (Gotlieb, 1938), i.e. every ℓ_2–function f: $\mathbb{N}^0\to\mathbb{R}$ can be written as a unique, converging sum of Laguerre

---

[2] This is equivalent with stating that $\{\psi_k^\xi(t)\}$ are the polynomial coefficients of the stable Laurent expansion around $z=\infty$ of a proper rational function $\Psi_k(z)$.

polynomials; hence there exist parameters $\alpha_i \in \mathbb{R}$, with $i \in \mathbb{N}^0$, such that

$$\forall t \in \mathbb{N}^0, \quad f(t) = \sum_{i=0}^{\infty} \alpha_i \psi_i(t) \tag{4.2.11}$$

In the next section we will show how to calculate the coefficients α_i and thus how we can transform an ℓ_2 signal $\{f(t)\}$ into a signal $\{\alpha_i\}$ in what could be called the orthonormal or Laguerre domain. Whenever we talk about ℓ_2 in the sequel we use this as an abbreviation for the signal space $\ell_{2[0,\infty)}$.

4.3 TRANSFORMATION OF TIME-SERIES

In section 4.2 we noticed that the Laguerre polynomials can be used as a basis of ℓ_2. In this section we show how we can explicitly calculate the coefficients of an expansion of an ℓ_2 function and explain that we can consider the sequence of these coefficients as another ℓ_2 function. Furthermore we elaborate on the special recursive structure in the sequence of the polynomials and show how this can be expressed in terms of vectors and matrices.

Definition 4.3.1 Let $\{y(t), t \in \mathbb{N}^0\}$, be a ℓ_2-sequence and let $\{\psi_k, k \in \mathbb{N}^0\}$ be the set of Laguerre polynomials, defined by definition 4.2.2. We define the k^{th} *Laguerre coefficient* Y_k of y by:

$$Y_k := \sum_{t=0}^{\infty} y(t)\psi_k(t) \qquad k \in \mathbb{N}^0 \tag{4.3.1}$$

and we define the set of these coefficients $\{Y_k, k \in \mathbb{N}^0\}$ as the *Laguerre transform* of the sequence y. ○

Note that $y, \psi_k \in \ell_2$ implies that $Y_k < \infty$ for every $k \in \mathbb{N}^0$. The mapping from $\{y(t), t \in \mathbb{N}^0\}$ to $\{Y_k, k \in \mathbb{N}^0\}$ will be referred to as the *Laguerre transformation*; it is in fact a bijective mapping, as is shown by the next proposition.

Proposition 4.3.2. [King & Paraskevopoulos, 1979]. Let y be an ℓ_2 function and let $\{Y_k, k \in \mathbb{N}^0\}$ be the set of Laguerre coefficients of y as defined by (4.3.1). Then,

$$y(t) = \sum_{k=0}^{\infty} Y_k \, \psi_k(t) \tag{4.3.2a}$$

Let the function $Y: \mathbb{N}^0 \to \mathbb{R}$ be defined by $Y(k) := Y_k$. Then $Y \in \ell_2$ and the Parseval identity holds:

$$\|y\|_2 = \|Y\|_2 \tag{4.3.2b}$$

○

Remark 4.3.3. We can compare the relations (4.3.1) and (4.3.2a) with for instance the well known Fourier transformation, so we can consider $(\{y(t),t\in\mathbb{N}^0\},\{Y_k,k\in\mathbb{N}^0\})$ as a transform pair. We call (4.3.1) the *forward Laguerre transformation* and (4.3.2a) the *backward* or *inverse Laguerre transformation.* With Fourier transforms it is common to consider the transformed signal as a representation in the frequency domain, we will use the term *orthogonal domain* or *Laguerre domain* for signals that are transformed with Laguerre polynomials. o

An important property of the Laguerre polynomials, but also of various other sets of orthogonal polynomials (Paraskevopoulos, 1985) is the recursive structure that follows from (4.2.8) and (4.2.9). Because of this structure it is possible to design efficients algorithms to calculate Laguerre coefficients (section 4.4). Furthermore it will be of use for system identification (section 4.6). This recursive property is often called the *shift structure.* By this we mean that the first N polynomials can be written in the form of a controllability matrix (2.4.5). This is formalized in the following proposition.

Proposition 4.3.4. [King & Paraskevopoulos, 1979]. Let $\{\psi_k^\xi\}$ be the set of Laguerre polynomials as defined in definition 4.2.2. Denote $\eta:=1-\xi^2$ and let $j\in\mathbb{N}$, then:

$$\begin{bmatrix}\psi_0(0)\\ \psi_1(0)\\ \vdots\\ \psi_n(0)\end{bmatrix} = \sqrt{\eta}\begin{bmatrix}1\\ -\xi\\ \vdots\\ (-\xi)^n\end{bmatrix} \quad\text{and}\quad \begin{bmatrix}\psi_0(k+j)\\ \psi_1(k+j)\\ \vdots\\ \psi_n(k+j)\end{bmatrix} = T_n^j\begin{bmatrix}\psi_0(k)\\ \psi_1(k)\\ \vdots\\ \psi_n(k)\end{bmatrix} \tag{4.3.3a}$$

where

$$T_n:=\begin{bmatrix} \xi & 0 & & & \cdots & 0\\ \eta & \xi & 0 & & \cdots & 0\\ -\xi\eta & \eta & \xi & 0 & \cdots & 0\\ \vdots & & \ddots & \ddots & \ddots & \vdots\\ (-\xi)^{n-1}\eta & & & \ddots & \ddots & \xi \end{bmatrix} \tag{4.3.3b}$$

o

The matrix T_n is often referred to as the *Shift Matrix.* This is in fact the same property as relation 4.2.8, however now expressed in matrix form and it shows that we can consider the matrix of the first (n+1) Laguerre polynomials as a controllability matrix:

Denote the vector of the first (n+1) polynomials at time t by $\underline{\psi_n}(t)$:

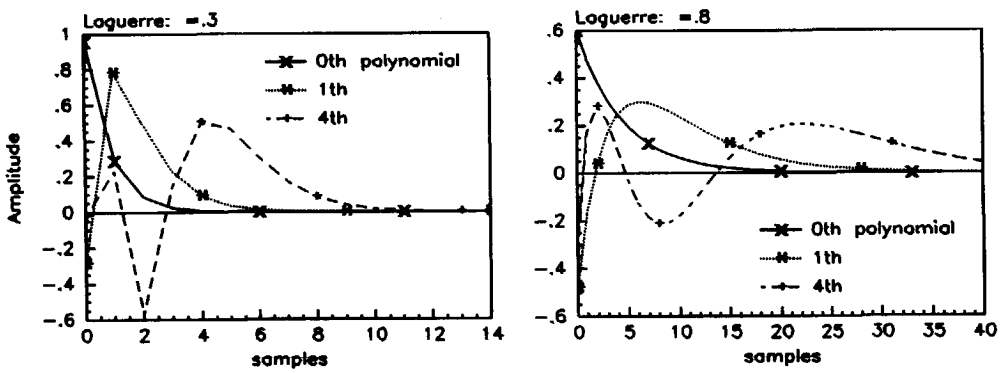
$$\underline{\psi_n}(t):=[\psi_0(t)\ \psi_1(t)\quad \psi_n(t)]^T \tag{4.3.4a}$$

$$\text{then}\quad [\ \underline{\psi_n}(0)\ |\ \underline{\psi_n}(1)\ |\cdots] = [\ \underline{\psi_n}(0)\ |\ T_n\underline{\psi_n}(0)\ |\ T_n^2\underline{\psi_n}(0)\ |\cdots] \tag{4.3.4b}$$

This shift structure will be used in section 4.6 to derive an identification method.

Remark 4.3.5. The parameter $\xi \in (0,1)$, often called the discount factor, is a degree of freedom, which can be understood to be a time scaling parameter. Consider the first generating transfer function as defined by (4.2.9): $\Psi_0(z) = \sqrt{(1-\xi^2)}\ \frac{z}{(z-\xi)}$ then we recognize a first order system with a pole in $z=\xi$. This has the result that changing ξ is equivalent with transformation of the time axis. At the same time the amplitude axis of the functions will be transformed in order to retain the orthonormality. This can be recognized from (4.3.4b), since a change of ξ will change the poles of T_n (4.3.3b) and hence also the decay rate in (4.3.4b). This effect is visualized in the following example. o

Example 4.3.6. In Fig. 4.3.1 an example of some of theses polynomials is given. It shows that changing the discount factor leads to different initial values, which corresponds to the initial conditions in 4.3.3. The second effect is a change of the time scaling, the functions are dragged along the time axis.
In the diagram on the left in Fig. 4.3.1 the first 15 samples are shown of ψ_0, ψ_1 and ψ_4 for $\xi=0.3$. The diagram on the right depicts the first 40 samples of the same Laguerre functions for $\xi=0.8$. o



$\xi=0.3$ $\xi=0.8$

Fig. 4.3.1. Example of Laguerre polynomials: ψ_0, ψ_1 and ψ_4.

4.4 FINITE TIME SERIES: APPROXIMATIONS AND CALCULATION.

We will use the Laguerre polynomials in a context of system identification and therefore we actually will have to calculate the Laguerre coefficients (4.3.1) of signals. In this section it is shown how this can be done in an efficient way for the special class of ℓ_2

functions that are zero for t>N. This results in an algorithm that makes full use of the recursive structure. For the more general class of functions, where we only have knowledge of a finite number of samples, we cannot calculate the Laguerre coefficients but we have to approximate these by choosing a specific extension. We consider two different approximations and show how a finite number of samples can be described in an exact manner with a finite number of coefficients. Furthermore we show the influence of the discount factor on such approximations.

In a practical situation we will be dealing with time–series of finite length, for instance the sampled input/output signals of a system. We consider the problem of transforming such a signal to the orthogonal domain. First we concentrate on the case that we are dealing with a time series $f\in\ell_2$ with $f(t)\equiv 0$, $t>N$, $N\in\mathbb{N}$. King and Paraskevopoulos (1979) show how the shift structure 4.3.3 can be used to calculate the transformation coefficients for such a finite sequence in an efficient way:

Proposition 4.4.1. [King and Paraskevopoulos, 1979] Let $f\in\ell_2$ with $f(t)\equiv 0$ for $t>N\in\mathbb{N}$ and let $\{F_k, k\in\mathbb{N}^0\}$ be the set of Laguerre coefficients (4.3.1) of f with corresponding generating transfer functions $\{\Psi_k(z)\}$. Define the signals $u, y_k\in\ell_{2\mathbb{Z}}$ by

$$u(t) := f(N-t) \quad t\in[0,N] \tag{4.4.1a}$$

$$u(t) := 0 \quad t\notin[0,N] \tag{4.4.1b}$$

$$y_k := \Psi_k(z)u \ ^{(3)} \tag{4.4.1c}$$

Then $y_k(N) = F_k$ o

Proof: The k^{th} Laguerre coefficient (4.3.1) of f(t) is defined as:

$$F_k := \sum_{t=0}^{\infty} f(t)\psi_k(t) = \sum_{t=0}^{N} f(t)\psi_k(t) \tag{4.4.2}$$

Now we use proposition 4.2.4, that shows that we can consider $\{\psi_k(t)\}$ as the impulse response parameters of $\Psi_k(z)$. With (4.4.1) this implies that

$$y_k(N) = \sum_{t=0}^{N} \psi_k(t)u(N-t) \tag{4.4.3}$$

Now combine (4.4.2) and (4.4.3), then it follows that:

$$u(N-t)=f(t) \Rightarrow y_k(N)=F_k \tag{4.4.3a}$$

or

$$u(t)=f(N-t) \Rightarrow y_k(N)=F_k \tag{4.4.3b}$$

□

This proposition shows that the calculation of F_k can be performed by leading the inverse sequence $[f(N),f(N-1),\cdots,f(0)]$ through the generating transfer function $\Psi_k(z)$ and that F_k will be the output upon the last input entry f(0). Now we use the recursive structure that is given by (4.2.9b):

---

[3] In the sense that $y(-\infty)=0$, i.e. $\exists M<0$ with $y(t)=0$ for $t<M$.

$$\Psi_k(z) = \frac{(1-z\xi)}{(z-\xi)}\,\Psi_{k-1}(z) \tag{4.4.4}$$

to build a simple network for calculating the first M Laguerre coefficients. Let

$$P(z) := \frac{(1-z\xi)}{(z-\xi)} \quad \text{and} \quad Q(z) := \Psi_0(z) = \sqrt{1-\xi^2}\cdot\frac{z}{(z-\xi)} \tag{4.4.5}$$

then $\Psi_k(z)=P^k(z)\cdot Q(z)$. It is straightforward that for the signals y_k (4.4.1c):

$$y_{k+1}=P(z)y_k \qquad k\in\mathbb{N}^0 \tag{4.4.6}$$

This shows that we only have to repeat the process (4.4.6) in order to calculate a new Laguerre coefficient. In Fig. 4.4.1 the scheme of the calculation program is drawn for the calculation of the first M coefficients, which can be programmed efficiently (P(z) is a first order system).

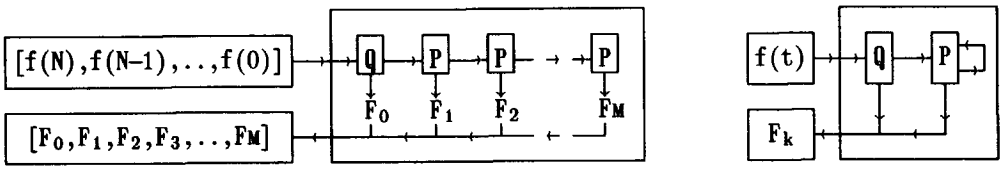


Fig. 4.4.1 Laguerre transformation network

When the time series under consideration are the result of sampling, for instance data obtained through an identification experiment, then the number of available data points will be finite. In general it will not be realistic to assume that these signals have a zero extension, which implies that we cannot calculate the Laguerre coefficients with the method of proposition 4.4.1. In fact it is not possible at all to calculate the 'correct' coefficients, since every different extension of the time series will lead to another set of Laguerre coefficients. A similar situation occurs with the Fourier transformation. In the calculation of Fourier coefficients of a given function f: $[0,N-1]\to\mathbb{R}$ it is implicitly assumed that this function is periodic, i.e. $f(t+kN)=f(t)$ for $k\in\mathbb{N}$. Hence a specific extension of the function is chosen. A nice feature of this choice is that only a finite number of Fourier coefficients (N) is needed in order to exactly describe the function f on the interval [0,N–1]. If the method of proposition 4.4.1 is applied to calculate Laguerre coefficients then it can be concluded that in general the number of coefficients will be infinite. Since the series of Laguerre coefficients is in ℓ_2 we know that they will tend to zero, but it is not known in which way. The question arises if it is possible to describe a function f exactly on a finite interval with a finite number of Laguerre coefficients or equivalently, does there exist an extension of f such that the extended function has only a finite number of Laguerre coefficients. The answer to this question is affirmative as is shown by the next proposition.

PROPOSITION 4.4.2. Let $\{\psi_k\}$ be the set of Laguerre polynomials and let Γ be the matrix with these functions: $(\Gamma)_{ij} := \psi_{i-1}(j-1)$. Let f_1: $[0,n-1] \to \mathbb{R}$ be a given ℓ_2 function. Then a unique extension f_2 of f_1 exists, f_2: $[n,\infty) \to \mathbb{R}$, such that the function $f: \mathbb{N}^0 \to \mathbb{R}$, defined by the concatenation of f_1 and f_2 is an ℓ_2 function that has only a finite number of non zero Laguerre coefficients $\{F_k\}$, in the sense that $F_i=0$, $i \geq n$. In particular if we denote

$$\Gamma = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \text{ with } \Gamma_{11} \in \mathbb{R}^{n \times n} \tag{4.4.7a}$$

$$\underline{f}_1 := \begin{bmatrix} f_1(0) \\ \vdots \\ f_1(n-1) \end{bmatrix} \quad \underline{f}_2 := \begin{bmatrix} f_2(n) \\ f_2(n+1) \\ \vdots \end{bmatrix} \quad \underline{F}_1 := \begin{bmatrix} F_0 \\ \vdots \\ F_{n-1} \end{bmatrix} \quad \underline{F}_2 := \begin{bmatrix} F_n \\ F_{n+1} \\ \vdots \end{bmatrix} \tag{4.4.7b}$$

$$\text{then:} \quad \underline{F}_1 = \Gamma_{11}^{-T} \underline{f}_1 \qquad \underline{F}_2 = 0 \qquad \text{and} \qquad \underline{f}_2 = \Gamma_{12}^T \Gamma_{11}^{-T} \underline{f}_1 \tag{4.4.7c}$$

o

PROOF: First note that the matrix Γ_{11} consists of the first n columns of the controllability matrix (4.3.4b). The controllability of the pair $\{T_n, \underline{\psi}_n(0)\}$ as defined by (4.3.4a, 4.3.3b) implies that Γ_{11} is non–singular. Γ is unitary, $\Gamma^T\Gamma = I$. The vector $\underline{F}$ with Laguerre coefficients of the function f obeys the equation

$$\begin{bmatrix} \underline{F}_1 \\ \underline{F}_2 \end{bmatrix} = \begin{bmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{bmatrix} \begin{bmatrix} \underline{f}_1 \\ \underline{f}_2 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} \underline{f}_1 \\ \underline{f}_2 \end{bmatrix} = \begin{bmatrix} \Gamma_{11}^T & \Gamma_{21}^T \\ \Gamma_{12}^T & \Gamma_{22}^T \end{bmatrix} \begin{bmatrix} \underline{F}_1 \\ \underline{F}_2 \end{bmatrix} \tag{4.4.8}$$

Because Γ_{11} is square and regular and because $\underline{F}_2 \equiv 0$ it follows that $\underline{F}_1 = \Gamma_{11}^{-T} \underline{f}_1$ is unique which implies that also $\underline{f}_2 = \Gamma_{12}^T \underline{F}_1 = \Gamma_{12}^T \Gamma_{11}^{-T} \underline{f}_1$ is unique. The fact that $f \in \ell_2$ is a direct result of $\|f\|_2 = \|F_1\|_2$. □

REMARK 4.4.3.

1. The extension, defined by proposition 4.4.2, depends on the first part of the signal. In the case of Fourier transformations often faders are applied to omit abrupt transitions. A similar procedure may be necessary for Laguerre transformations, but we have not pursued the research on this subject.

2. One reason to apply Laguerre transformations to data is the so called datareduction property (King and Paraskevopoulos, 1979), i.e the number of Laguerre coefficients to approximate a signal on a given interval is often much smaller then the number of samples. However it must be understood that for such an approximation we must implicitly assume an extension of the function and different extensions will lead to different Laguerre polynomials. If we are merely interested in the approximation on a finite interval with a given number of coefficients then we are dealing with the following problem:

Let $f \in \ell_2$ and let $n,m \in \mathbb{N}$. Find coefficients $\{\hat{F}_0, \cdots, \hat{F}_{m-1}\}$ such that

$$\{\hat{F}_0, \cdots, \hat{F}_{m-1}\} = \text{Arg Min} \left[\sum_{t=0}^{n-1} \left(f(t) - \sum_{k=0}^{m-1} F_k \psi_k(t) \right)^2 \right] \tag{4.4.9a}$$

where we assume that we want to approximate in ℓ_2 sense. If $n \leq m$ then this problem is solved by proposition 4.4.2, that shows that we can describe f exactly with n coefficients.

If $m<n$ then the solution to this least squares problem can be expressed in terms of the matrix Γ with Laguerre polynomials as follows:
Let Γ be partitioned according to (4.4.7a) with $\Gamma_{11}\in\mathbb{R}^{mxn}$ and denote by $\underline{F}\in\mathbb{R}^m$ and $\underline{f}\in\mathbb{R}^n$ the vectors of coefficients and samples as in (4.4.7b). Then we can reformulate (4.4.9a) as

$$\hat{\underline{F}} = \text{Arg Min } \|\underline{f} - \Gamma_{11}^T\underline{F}\|_2 \tag{4.4.9b}$$

which implies

$$\hat{\underline{F}} = (\Gamma_{11}\Gamma_{11}^T)^{-1}\Gamma_{11}\underline{f} \tag{4.4.9c}$$

Hence we have defined three different ways to calculate Laguerre coefficients:

Laguerre coefficients	Motivation
$\hat{\underline{F}}_1=\Gamma_{11}\underline{f}$	approximation based on the assumption of a 0–extension
$\hat{\underline{F}}_2=(\Gamma_{11}^T)^{-1}\underline{f}$	exact description on a finite length interval
$\hat{\underline{F}}_3=(\Gamma_{11}\Gamma_{11}^T)^{-1}\Gamma_{11}\underline{f}$	optimal least squares approximation

From a numerical point of view the first method ($\hat{\underline{F}}_1$) is to be preferred, since an efficient algorithm exists to calculate the coefficients. Apparently the calculations for the other two sets of coefficients are more involved. Note that $\lim_{n\to\infty}\Gamma_{11}\Gamma_{11}^T=I_m$ ($\Gamma_{11}\in\mathbb{R}^{mxn}$), which implies that for large values of n the approximations $\hat{\underline{F}}_1$ and $\hat{\underline{F}}_2$ are almost identical. The rate of convergence of $\Gamma_{11}\Gamma_{11}^T$ depends on the value of ξ; for small ξ this expression will converge rapidly. In example 4.4.8 we will illustrate the differences between the approximations $\hat{\underline{F}}_1$ and $\hat{\underline{F}}_2$. o

Remark 4.4.4. If we want to examine how well a signal is approximated with a finite number of Laguerre coefficients then we will have to perform the inverse transformation. This will be a necessary step for the identification methods of section 4.6. The following proposition shows that this can be done in exactly the same way as the forward transformation and that we can calculate the inverse Laguerre transformation in the same way as the forward transformation by using $-\xi$ instead of ξ. Therefore we point out that replacing ξ by $-\xi$ does not alter any of the properties of the Laguerre polynomials. A formal proof is omitted here, but it is a direct consequence of the theory which we will develop in chapter five. Let $(\{f(t),t\in\mathbb{N}^0\},\{F_k,k\in\mathbb{N}^0\})$ be a transform pair and denote the vectors with function values by $\underline{F}$ and $\underline{f}$ and the matrix with Laguerre polynomials by $\Gamma(\xi)$:

$$\Gamma(\xi):=\begin{bmatrix}\psi_0(0) & \psi_0(1) & \cdots\\ \psi_1(0) & \psi_1(1) & \cdots\\ \vdots & \ddots & \vdots\end{bmatrix} \qquad \underline{F}=\begin{bmatrix}F_0\\ F_1\\ \vdots\end{bmatrix} \qquad \underline{f}=\begin{bmatrix}f(0)\\ f(1)\\ \vdots\end{bmatrix} \tag{4.4.10}$$

Then we know that $\Gamma(\xi)\Gamma^T(\xi)=I$ and that $\underline{f}$ and $\underline{F}$ are related by: $\underline{F}=\Gamma(\xi)\underline{f}$ and $\underline{f}=\Gamma^T(\xi)\underline{F}$. Now a straightforward calculation shows that $\Gamma^T(\xi)=\Gamma(-\xi)$:

Proposition 4.4.5. Let $\Gamma(\xi)$ be given by (4.4.10) then $\Gamma^T(\xi)=\Gamma(-\xi)$. o

Proof: See appendix 4A.

In practical applications one might be interested in data–reduction or in approximating signals or sequences with a small number of basis functions. This is one of the reasons why orthogonal functions are widely used in a large number of fields (King and Paraskevopoulos, 1979; Paraskevopoulos, 1985). The orthogonal polynomials serve as an appropriate tool for this purpose and the previous proposition shows how one can check the approximation error in case of approximation with a finite number of Laguerre polynomials. Note that in this case (Given $\underline{F}$ with $F_i\equiv 0$ for $i>N$, calculate the corresponding $\underline{f}$) we can use the method of proposition 4.4.1 to calculate the approximation $\underline{f}$, because the finiteness of the set of Laguerre coefficients is equivalent with stating that the ℓ_2 series defined by these coefficients has a zero extension.

Remark 4.4.6. Until now we have assumed that the discount factor ξ is constant, but we know that every ξ defines another set of Laguerre polynomials. Now suppose that we want to approximate a function f on the interval [0,n–1] in ℓ_2 sense with m Laguerre functions. In remark 4.4.3 an expression for the optimal coefficients was derived for a specific value of ξ. We can ask which choice of ξ will lead to the smallest approximation error. There is no analytical way known to calculate the optimal discount factor, but in (Clowes, 1965; King and O'Canain, 1969; King and Paraskevopoulos, 1979; Maione and Turchiano, 1985) some methods are proposed to approximate the optimal ξ. These methods are basically Fibonacci search methods, but one should be careful using these:

Let $f\in\ell_2$ and let $\xi\in(0,1)$. We want to approximate f on the interval [0,n–1] in ℓ_2 sense with m Laguerre coefficients, as in (4.4.9a). The solution $\{\hat{F}_k^{\xi},\ k=0\cdots m-1\}$ to this problem is given by (4.4.9c). Denote the approximation error by $J(\xi)$:

$$J(\xi):=\left[\sum_{t=0}^{n-1}\left[f(t)-\sum_{k=0}^{m-1}\hat{F}_k^{\xi}\psi_k^{\xi}(t)\right]^2\right]^{\frac{1}{2}} \tag{4.4.11}$$

where we indexed F_k with ξ to emphasize its dependence on ξ. Furthermore we want to find the value of ξ that minimizes the function $J(\xi)$.

To apply Fibonacci search algorithms successfully $J(\xi)$ should not have local minima in the interval (0,1), but experiments show this is not always the case. The best value of ξ depends also on the number of orthonormal coefficients, that are used, i.e if ξ_1 is the

optimal value for the approximation with m polynomials then it is not true in general that ξ_1 is also optimal for approximation with (m+p) polynomials. An example of this is the approximation with two and three Laguerre polynomials of the first 10 samples of a time series f, given by:

f(0)= 1.2085 f(2)=-1.2152 f(4)=-0.0548 f(6)= 1.3583 f(8)= 2.1130
f(1)= 1.7887 f(3)=-1.4315 f(5)= 0.5919 f(7)= 0.9228 f(9)= 0.0343

In Fig. 4.4.2 the functions $J(\xi)$ are depicted for the approximation with one and two Laguerre functions. It shows that $J(\xi)$ has local minima and that the optimal value of ξ depends on the number of Laguerre polynomials that is used. This shows that the optimal ξ may only be found with a trial and error procedure. o

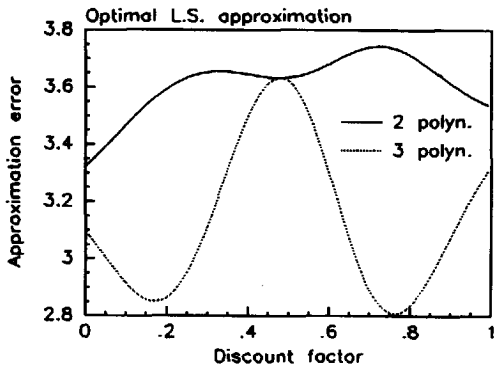


Fig. 4.4.2 Approximation error $J(\xi)$.

Remark 4.4.7. Contrary to the Fourier transformation the index of the first sample that is used for a transformation has a large influence on the resulting Laguerre coefficients. If $f_1 \in \ell_2$ then the Laguerre coefficients of f_1 can be completely different from the coefficients of the shifted function $f_2 \in \ell_2$, defined by $f_2(t) := f_1(t+1)$. However they are related: denote the vectors of the functions and the vectors with coefficients of both functions with $\underline{f}_i$ and $\underline{F}_i$ (i=1,2) respectively, then: $\underline{F}_1 := \Gamma \underline{f}_1 \quad \underline{F}_2 := \Gamma \underline{f}_2$
Let T and $\underline{\psi}$ be the infinite versions of (4.3.3), then we can write $\Gamma = [\, \underline{\psi} \mid T\Gamma \,]$
and hence $\underline{F}_1 = \underline{\psi} \times \underline{f}_1(0) + T \times \Gamma \underline{f}_2 = \underline{\psi} \times \underline{f}_1(0) + T \times \underline{F}_2$
In the next example it is shown that this will change the coefficients. o

Example 4.4.8. This example is meant to illustrate the effect of the different approximation methods with a finite number of Laguerre polynomials and to show that

the choice of the discount factor is very important. The approximated signal, the solid line, is a measured impulse response and we consider the first 80 samples of this signal. Figure 4.4.3 depicts the approximation of this signal with:

Upper left: ξ=0.5, 10 polynomials Upper right: ξ=0.8, 5 polynomials

Lower left: ξ=0.5, 20 polynomials Lower right: ξ=0.8, 10 polynomials

The coefficients are calculated with the method of proposition 4.4.1, i.e they are the correct coefficients if the approximated signal has a zero extension.

In Fig. 4.4.4a we illustrate the difference between the approximation based on a zero extension and the optimal least squares approximation. The picture shows the resulting approximations when 10 Laguerre polynomials are used with a discount factor ξ=0.9. As noted in remark 4.4.3 the difference will get smaller if more functions or smaller values of ξ are used. In Fig. 4.4.4b we show the effect on the Laguerre coefficients of shifting the time series (see remark 4.4.7). The diagram depicts the first 20 Laguerre coefficients (ξ=0.5) for the signals f(t) and f(t+1). The coefficients are calculated under the assumption that the signals have a zero extension.

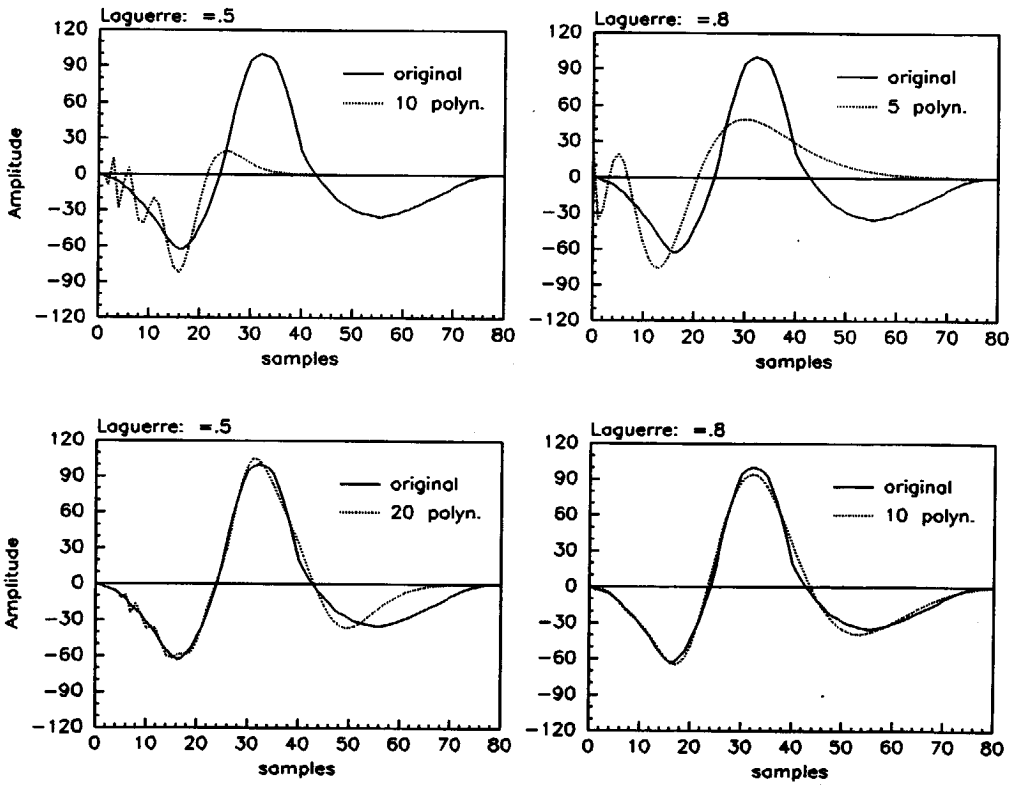


Fig. 4.4.3 Approximation in ℓ_2 of a measured impulse response with Laguerre polynomials with ξ=0.5 and ξ=0.8.

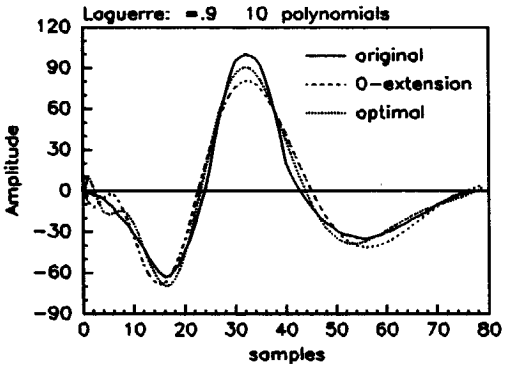


Fig. 4.4.4a. Different types of Laguerre approximations.

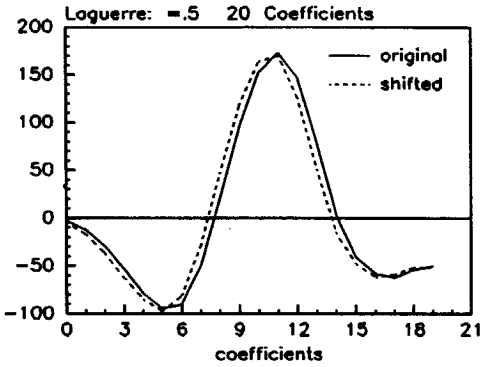


Fig. 4.4.b. Effect of time shifting on the Laguerre coefficients

4.5 TRANSFORMATION OF LINEAR SYSTEMS

In the previous sections we explained the notion of finite difference Laguerre polynomials and showed how they can be considered as an orthonormal basis for $\ell_2[0,\infty)$. Further, it was shown how ℓ_2 signals can be transformed to the so called Laguerre domain, by considering the sequence of expansion coeffients as a new signal in itself. In this section we will apply this transformation to the input and output variables of a FDLTS system and show that the resulting signals in the Laguerre domain can again be considered as the input and output variables of a FDLTS system. This new system can be expressed elegantly in terms of the original system both on transfer function level as in terms of state space descriptions. Thus we can consider this procedure as the Laguerre transformation of a FDLTS system. This procedure is covered by the following theorem.

THEOREM 4.5.1. Let G(z) be a (p×m) FDLTS system and let [A,B,C,D] be a stable realization of G. Let $(u,y)\in\mathcal{B}_2^{\infty}(G)$ be an input output pair of G and let x_0 be the vector with initial conditions, $\|x_0\|_2<\infty$, such that there exists a vector proces $x\in\ell_2^n$ with:

$$x(t+1) = Ax(t) + Bu(t) \tag{4.5.1a}$$

$$y(t) = Cx(t) + Du(t) \tag{4.5.1b}$$

$$x(0) = x_0 \tag{4.5.1c}$$

Let $\{\psi_k^{\xi}(t),\ k,t\in\mathbb{N}^0\}$ be the set of Laguerre polynomials as defined by definition 4.2.2 with discount factor ξ and let $\{U_k,\ k\in\mathbb{N}^0\}$ and $\{Y_k,\ k\in\mathbb{N}^0\}$ be the sets with Laguerre coefficients (4.3.1) of u(t) respectively y(t).

$$U_k = \sum_{k=0}^{\infty} u(t)\ \psi_k(t) \qquad u(t) = \sum_{k=0}^{\infty} U_k\ \psi_k(t) \tag{4.5.2a}$$

$$Y_k = \sum_{k=0}^{\infty} y(t)\ \psi_k(t) \qquad y(t) = \sum_{k=0}^{\infty} Y_k\ \psi_k(t) \tag{4.5.2a}$$

1. **State space description** [Nurges and Yaaksoo, 1981; Nurges, 1987]. Let $\{F,G,H,J,z_0\}$ be defined by

$$F = [A-\xi I][I-\xi A]^{-1} \qquad G = \sqrt{1-\xi^2}[I-\xi A]^{-1}B \tag{4.5.3a}$$

$$H = \sqrt{1-\xi^2}\ C[I-\xi A]^{-1} \qquad J = D + \xi C[I-\xi A]^{-1}B \tag{4.5.3b}$$

$$z_0 = \sqrt{1-\xi^2}\ x_0 \tag{4.5.3c}$$

Then $\{U,Y\}$ obey the equations:

$$z_{k+1} = Fz_k + GU_k \tag{4.5.4a}$$

$$Y_k = Hz_k + JU_k \quad k\in\mathbb{N}^0 \tag{4.5.4b}$$

In other words, $\{U_k,Y_k\}$ is an input/output pair of a system $L(z)$ with realization $[F,G,H,J]$, $(U,Y)\in\mathcal{B}_2^+(L)$.

2. **Transfer functions** [Maione and Turchiano, 1985]. Let $\pi_\xi(z)$: $\mathbb{C}\to\mathbb{C}$ be defined by

$$\pi_\xi(z) = \frac{\xi+z}{1+\xi z} \tag{4.5.5a}$$

then $[F,G,H,J]$ (4.5.4) is a realization of a system $L(z)=G(\pi_\xi(z))$ (4.5.5b)

Proof: See appendix 4B. o

This theorem shows that the transformation of time series induces a transformation of systems, that is referred to as the *Laguerre system transformation.* The equations (4.5.4) with the matrices (4.5.3) are known as the *Laguerre state equations* of the system G. This transformation on state space level has very nice properties, as stated in the next proposition.

Proposition 4.5.2. The transformation defined by the equations (4.5.3) leaves the following properties invariant:

1. Controllability, observability, minimality, stability (Nurges and Yaaksoo, 1981)
2. Gramians, Hankel singular values, balancedness and static gain. o

Proof: See appendix 4C.

Remark 4.5.3.

1. As we have seen before in proposition 4.4.5 the inverse of the Laguerre signal transformation is the same as the forward Laguerre transformation if we replace ξ by $-\xi$. It is straightforward that this is also valid for the Laguerre system transformation. Consider the inverse of the transformation (4.5.5):

$$\pi_\xi(z) = \frac{\xi+z}{1+\xi z} \Leftrightarrow z=\frac{\pi_\xi(z)-\xi}{1-\pi_\xi(z)} = \pi_{-\xi}\Big[\pi_\xi(z)\Big] \Rightarrow \pi_{-\xi}(z) = \pi_\xi^{-1}(z) \tag{4.5.6}$$

2. These transformations now have been defined for discrete time systems, but it is also possible to consider the transformation of continuous time systems which is induced by the combination of the (discrete) Laguerre system transformation with the ω-transformation. We will consider this (continuous) system transformation in detail in section 4.7, but mention the result at this point because it gives a clear insight in the underlying mechanism. If G(s) is a FDLTS system then the induced Laguerre transform of G(s) is $G_\ell(s)$ with

$$G_\ell(s) := G(\beta s) \quad \text{where} \quad \beta = \frac{1-\xi}{1+\xi} \tag{4.5.7}$$

Hence in continous time the Laguerre transformation comes down to a rescaling of the frequency scale. ○

In theorem 4.5.1 we have given expressions for the transfer function and a state space realization of the Laguerre transform of a FDLTS system. Another representation, that is widely used in system identification, is the so called autoregressive (AR) representation. Nurges (1987) translated the result of theorem 4.5.1 in terms of the latter representation and applied this to develop a system identification method based on Laguerre transformations. The result of the Laguerre system transformation in AR terms is presented in the following proposition. The application of this result to system identification will be considered in section 4.6.

Proposition 4.5.4. [Nurges, 1987] Let a (pxm) FDLTS system be given by the following autoregressive representation:

$$\forall t \in \mathbb{N}^0 \qquad \sum_{j=0}^{n} A_j y(t+j) + \sum_{i=0}^{n-1} B_i u(t+i) = 0 \tag{4.5.8}$$

$y(s)=0, u(s)=0,\ s=0,\cdots,n-1. \qquad A_k \in \mathbb{R}^{p\times p},\ B_k \in \mathbb{R}^{p\times m}.$

Let $u,y \in \ell_2$ and define Y_k, U_k as the k^{th} Laguerre coefficients (4.3.1) of y and u. The sequences $\{Y_k\}$ and $\{U_k\}$ are subject to the *Laguerre difference equation*:

$$\forall k \in \mathbb{N}^0 \qquad \sum_{j=0}^{n} F_j Y_{k+j} + \sum_{i=0}^{n} G_i U_{k+j} = 0 \tag{4.5.9a}$$

$$\text{where for } 0 \le i \le n \qquad F_i = \sum_{j=0}^{n} \sum_{r=0}^{j} \binom{n-j}{i-r} \binom{j}{r} \xi^{i+j-2r} A_j \tag{4.5.9b}$$

$$G_i = \sum_{j=0}^{n-1} \sum_{r=0}^{j} \binom{n-j}{i-r} \binom{j}{r} \xi^{i+j-2r} B_j \tag{4.5.9c}$$

○

Remark 4.5.5.

1. In remark 4.5.3 we mentioned the continuous time analog of Laguerre system transformations, induced by the discrete Laguerre system transformation and the ω-transformation. Next to that transformation there exists a transformation of continous time systems which is directly based on the transformation of time functions (Kukk, 1971). In this case the continuous time Laguerre polynomials are used as a basis of the

space L_2, so every function is written as a converging sum of continuous time Laguerre polynomials (4.2.3), $f(t)=\Sigma F_k L_k(t)$. If this is applied to the input/output signals of a FDLTS system G(s) then this leads to 2 'discrete' sequences $\{U_k\}$ and $\{Y_k\}$. Kukk (1971) shows that these signals are the input/output variables of a discrete time system. This idea can be applied for identification of continuous time systems with discrete type algorithms (Chang, Cheng and Wang, 1985; Unbehauen and Rao, 1988 and references therein), but we will not pursue this line of reasoning. Another application for continuous time systems is derived by Steiglitz (1965), where the continuous time transfer function is approximated by its Laguerre spectrum. Mäkilä (1990) applies Laguerre functions and filters to approximate infinite dimensional stable systems with finite dimensional systems.

2. In the definition of the Laguerre polynomials it was stated that the discount factor $\xi\in(0,1)$ and we showed that the inverse transformation is induced by using $-\xi$. In fact we may allow the case $\xi=0$ too, in which case $\psi_k(t)=\delta_{kt}$ (Kronecker delta) and the Laguerre transformation is the identity map. Hence we may conclude that the Laguerre transformation is well defined for $\xi\in(-1,1)$. Note that $\xi\in(-1,1) \Leftrightarrow T_n^{\xi}$ is stable, where T_n is the shift matrix (4.3.3b). In chapter 5, where we present a more general class of orthonormal functions, it will be shown that this is an essential property. o

In the remainder of this chapter we will frequently use the Laguerre system transformation defined by theorem 4.5.1. and in order to make a precise distinction between the original system and its transform we give the following definition.

Definition 4.5.6 Let G(z) be a FDLTS system with input/output pairs $\{u(t),y(t)\}$ and let L(z) be the Laguerre transformation of G(z) as defined by theorem 4.5.1 with input/output pairs $\{U_k,Y_k\}$. We say that G, u(t) and y(t) *work* in the *time domain* and L, U_k and Y_k in the *Laguerre domain* or *orthogonal domain.* o

4.6 LAGUERRE TRANSFORMATIONS AND SYSTEM IDENTIFICATION

4.6.1 INTRODUCTION

In this thesis we aim at the development of a new theory and alternative methods to approach the problem of approximate system identification. For this purpose we consider the use of orthogonal polynomials. In the previous sections it was shown that the Laguerre polynomials constitute a basis for ℓ_2 and that we can transform time domain signals into sequences of Laguerre coefficients, that are the representations of the former signals in the Laguerre basis. Furthermore it has been shown that this transformation of

signals induces a transformation of FDLTS system. In this section we show how these concepts can be used to develop alternative identification methods. These methods are based on respectively the transformation of data (section 4.6.2), the transformation of systems (section 4.6.3) and the description of systems in terms of the generating Laguerre transfer functions (section 4.6.4).

The use of orthogonal functions for identification purposes has a long history, see for instance the work of Lee (1932) and Wiener (1949). During the past decade continuous time orthogonal functions have been used by many authors in the context of continuous time system identification, (Chang, Cheng and Wang, 1985; Paraskevopoulos, 1985, Unbehauen and Rao, 1988). We restrict attention to the discrete time case, which is motivated by the fact that modern computing technology enforces the use of digital calculations. There are many different ways to use orthogonal functions in this context; in this section we consider three different approaches that are based on the theory of the previous sections. A more detailed introduction and motivation on this concept is given in section 5.5.

4.6.2 DIRECT APPROACH

In this section we consider an identification method, that was presented by King and Paraskevopoulos (1979). It is called the direct approach because in essence it does not alter the estimation procedure. It is mainly a method to compress the number of calculations and is based on the data reduction properties of the Laguerre functions. Hence it may be considered to be motivated by numerical reasons. We explain the underlying idea by applying the method to the case of a simple model structure and criterion.

Suppose we have sampled time–series y(t) and u(t), t=0,1,$\cdots$N, respectively the output and the input of a SISO–system and we want to fit the following equation error model or ARX model (Ljung, 1987) to the data in a least squares sense:

$$y(t) = ay(t-1) + bu(t-1) + e(t) \qquad (4.6.1a)$$

where (a,b) are unknown parameters and the equation error
e(t) is the result of disturbances and modeling errors.

We wish to estimate values $\{\hat{a},\hat{b}\}$ for the parameters such that the equation error is minimized in least squares sense, hence we want to minimize the criterion

$$\mathcal{J}(a,b) := \sum_{t=0}^{N} e^2(t) \qquad (4.6.1b)$$

Let $\{\psi_k, k\in\mathbb{N}^0\}$ be the set of Laguerre polynomials, for a fixed value of ξ.
We denote by $\underline{\psi}_n(t)$ the vector of the first (n+1) polynomials at time t:

$$\underline{\psi}_n(t) := [\psi_0(t), \psi_1(t), \cdots, \psi_n(t)]^T \tag{4.6.2a}$$

and we will use the shift property which was defined by (4.3.3):

$$\underline{\psi}_n(t) = T_n \underline{\psi}_n(t-1) \tag{4.6.2b}$$

The parameter n can be considered as a design variable. It is implicitly assumed that y, u and e are ℓ_2 time series, which will always be the case if we assume that these signals are zero for t>N. This implies that the Laguerre coefficients (4.3.1) of these signals are well defined and we denote by $\underline{Y}_n$, $\underline{U}_n$, $\underline{E}_n$ the vectors of the first (n+1) Laguerre coefficients of respectively y(t), u(t) and e(t):

$$\underline{Y}_n := \sum_t \underline{\psi}_n(t)y(t) \qquad \underline{U}_n := \sum_t \underline{\psi}_n(t)u(t) \qquad \underline{E}_n := \sum_t \underline{\psi}_n(t)e(t) \tag{4.6.2c}$$

With these preliminaries we can explain the idea of the direct approach:
Multiply each term in (4.6.1a) with $\underline{\psi}_n(t)$:

$$\underline{\psi}_n(t)y(t) = a\underline{\psi}_n(t)y(t) + b\underline{\psi}_n(t)u(t-1) + \underline{\psi}_n(t)e(t) \tag{4.6.3a}$$

and use the shift property to rewrite this to:

$$\underline{\psi}_n(t)y(t) = aT_n\underline{\psi}_n(t-1)y(t-1) + bT_n\underline{\psi}_n(t-1)u(t-1) + \underline{\psi}_n(t)e(t) \tag{4.6.3b}$$

where we assume $y(-1)=u(-1)=\psi_n(-1)=0$.
Sum each term in (4.3.6b) over all $t\in\mathbb{N}^0$ and use the definition of $\underline{Y}_n$, $\underline{U}_n$, $\underline{E}_n$:

$$\underline{Y}_n = a\, T_n\underline{Y}_n + b\, T_n\underline{U}_n + \underline{E}_n \tag{4.6.4}$$

If n is large enough (here $n\geq 1$) then (4.6.4) results in an overdetermined set of equations from which we can estimate [a,b] by minimizing $\langle \underline{E}_n, \underline{E}_n\rangle$, leading to estimates $[\hat{a},\hat{b}]_n$. Calculation of an extra Laguerre coefficient leads to an "updated" equation

$$\underline{Y}_{n+1} = a\, T_{n+1}\underline{Y}_{n+1} + b\, T_{n+1}\underline{U}_{n+1} + \underline{E}_{n+1} \tag{4.6.5}$$

King and Paraskevopoulos (1979) show that the updated estimate can be calculated in a recursive way from the previous one, leading to an algorithm of the form:

$$[\hat{a},\hat{b}]_{n+1} = f\left[[\hat{a},\hat{b}]_n,\ \underline{Y}_{n+1},\ \underline{U}_{n+1}\right] \tag{4.6.6}$$

As was explained in section 4.4 the calculation of a new Laguerre coefficient can be performed in an efficient way with the network of Fig. 4.4.1 (under the assumption that the signals have a zero extension; if this is not true then we can only approximate the Laguerre coefficients). From proposition 4.3.2 we know that the Parseval identity is valid, hence we know that

$$\lim_{n\to\infty} \langle \underline{E}_n, \underline{E}_n\rangle = \sum_{t=0}^{N} e^2(t) \tag{4.6.7}$$

so if n is taken large enough we will have minimized $\sum_{t=0}^{N} e^2(t)$, which was the objective (it is assumed that $e(t) \in \ell_2$, which is satisfied by the assumption that $y(t)=u(t)=0$, $t>N$).

The motivation of the authors for this approach is that the transformation of the data will lead to a considerable data reduction, such that the number of Laguerre coefficients to be calculated is much smaller than the length of the original time series. Thus they aim at a reduction the number of least squares equations. Note that the estimated parameters are the same as the parameters of the original problem. In the other approaches, treated in the next sections, this will not be the case.

Remark 4.6.1.

1. This method must be considered as an approximate method at two different levels; first we have to assume a specific extension for the signals u and y where the use of zero extensions seems the best choice from a numerical point of view; secondly the signals are approximated with a finite number of coefficients.
2. The multivariable extension of this method is easy to derive from the above. King and Paraskevopoulos (1979) use this method for rapidly converging input sequences and they state (without proof) that this algorithm is less sensitive to disturbances on the data than the "normal" least squares approach. They further propose the use of Laguerre functions as inputs for the identification experiment, in which case the calculation of the coefficients of the input sequence is unnecessary. o

4.6.3 IDENTIFICATION IN THE LAGUERRE DOMAIN.

In this section we consider a quite general idea, that is based on the notion that every FDLTS system has an equivalent in the Laguerre domain, as was explained in section 4.5. The main idea is to identify this 'Laguerre system' instead of the original 'time' system and to use the transformation properties (4.5.3/5) to derive the 'time' system from the estimated 'Laguerre' system. Hence it is required that we calculate the Laguerre coefficients of the available signals and use an identification technique on these data. We will examine what this procedure actually means in terms of the original data sequences. This will lead to a restriction on the kind of model class that is used in the procedure.

A rough sketch of the idea is that, given time series of inputs and outputs, the following procedure is carried out:

1– Calculate Laguerre coefficients of input– and output series.
2– Use any identification technique to find a model explaining these data.
3– Transform this model to the time domain.

If we consider the first step as a pre–filter operation (in the time domain) then this fits in nicely within the framework of standard identification methods as decribed by Ljung (1985,1987,1989) and Wahlberg and Ljung (1986). The question arises if it is possible to determine what this means in time domain terms, or how can the resulting model estimate be considered as a function of the original data, the model class and the criterion. In order to get an idea we use the transformation of transfer functions described in theorem 4.5.1 and we adopt the general model description for linear systems as proposed by Ljung (1987). We assume that such a model is derived from the data in the Laguerre domain:

$$A\ell(z^{-1})Y_k = \frac{B\ell(z^{-1})}{F\ell(z^{-1})} U_k + \frac{C\ell(z^{-1})}{D\ell(z^{-1})} E_k \qquad (4.6.8)$$

with $A\ell$, $B\ell$, $F\ell$, $C\ell$ and $D\ell$ polynomials in the shift operator z^{-1} with degree n_a, n_b etc, and Y_k, U_k, E_k the k^{th} Laguerre coefficients of output,input and disturbance. We indexed the polynomials with index "ℓ" to emphasize that they constitute a model for the Laguerre data.

For this analysis we consider the asymptotic case, i.e we assume that we are dealing with complete infinite sequences of Laguerre coefficients and ℓ_2 time functions samples.

Transformation of a model (4.6.8) to the time domain leads to transformation of the polynomials in this expression to the time domain according to the transformation defined in (4.5.5): $z^{-1} \to \pi_\xi^{-1}(z^{-1}) = \frac{z^{-1}-\xi}{1-\xi z^{-1}}$ (see remark 4.5.3(1)).

PROPOSITION 4.6.2. Let $P(z^{-1}) := \sum_{i=0}^{n_P} P_i z^{-i}$ be a polynomial in z^{-1} with degree n_p, where $\{P_i\}$is a set of constant matrices. Let the Laguerre transformation $\pi_\xi(z)$ be defined by (4.5.5a). Then there exist a polynomial $Q(z^{-1}) := \sum_{i=0}^{n_P} Q_i z^{-i}$ with degree n_p such that

$$P(\pi^{-1}(z^{-1})) = (1-\xi z^{-1})^{-n_P} Q(z^{-1}) \qquad \circ$$

PROOF: Evaluate the result of replacing z^{-1} by $\pi^{-1}(z^{-1})$:

$$P(\pi^{-1}(z^{-1})) = \sum_{i=0}^{n_P} P_i \left[\frac{z^{-1}-\xi}{1-\xi z^{-1}}\right]^i = \frac{1}{(1-\xi z^{-1})^{n_P}} \sum_{i=0}^{n_P} P_i \times (1-\xi z^{-1})^{n_P-i}(z^{-1}-\xi)^i \qquad (4.6.9a)$$

Define

$$Q(z^{-1}) := \sum_{i=0}^{n_P} P_i \times (1-\xi z^{-1})^{n_P-i}(z^{-1}-\xi)^i \qquad (4.6.9b)$$

For $0 \le i \le n_p$ the function $(1-\xi z^{-1})^{n_P-i}(z^{-1}-\xi)^i$ is a polynomial in z^{-1} with degree n_p, and hence $Q(z^{-1})$ is a polynomial with degree n_p. □

Remark 4.6.3. If in proposition 4.6.2 the polynomial P is monic ($P_0=I$) then the polynomial Q will in general not be monic. It can be verified from (4.6.9b) that

$$Q_0 = \sum_{i=0}^{n_P} (-\xi)^i P_i. \qquad \circ$$

Now we apply the inverse Laguerre system transformation on (4.6.8). From theorem 4.5.1 we can deduce that this results in the transformation of each of the polynomials. Proposition 4.6.2 shows that there exist polynomials A_t,B_t,C_t,D_t,F_t, with degree n_a,n_b etc, such that

$$A\ell(\pi^{-1}(z^{-1})) = \frac{A_t(z^{-1})}{(1-\xi z^{-1})^{n_a}} \tag{4.6.10}$$

Transformation of the Laguerre model (4.6.8) to the time domain results in transformation of all polynomials in that expression to the time domain and the series of Laguerre coefficients of the signals are transformed into the original time domain signals:

$$\frac{A_t(z^{-1})}{(1-\xi z^{-1})^{n_a}} y(t) = \frac{B_t(z^{-1})(1-\xi z^{-1})^{n_f}}{F_t(z^{-1})(1-\xi z^{-1})^{n_b}} u(t) + \frac{C_t(z^{-1})(1-\xi z^{-1})^{n_d}}{D_t(z^{-1})(1-\xi z^{-1})^{n_c}} e(t) \tag{4.6.11}$$

where the signals y(t), u(t) and e(t) represent the output, input and disturbance of the time domain model. For reasons of simplicity we restrict attention to the model structures that are most common in the context of system identification and we will investigate how a model structure in the 'Laguerre' domain imposes a model structure in the 'time' domain. The model structures that are considered for the Laguerre domain are summarized in table 4.6.1, that is based on Ljung (1987).

Table 4.6.1. Common black–box models

Polynomials	Name of Model structure
B	FIR (Finite Impulse Response)
AB	ARX
ABC	ARMAX
BF	OE (Output Error)
BFCD	BJ (Box–Jenkins)

If in the Laguerre model a polynomial is equal to the identity, then the transformation $z^{-1} \rightarrow \pi^{-1}(z^{-1})$ shows that the time domain equivalent of this polynomial is again equal to the identity. We now consider the application in the Laguerre domain of each of the model structures in table 4.6.1 and derive the time domain equivalent (4.6.11).

1. ARMAX in the Laguerre domain. It is common to use $n_a=n_b$. In the time domain this leads to another ARMAX model with a special C–polynomial:

$$A_t y(t) = B_t u(t) + C_t(1-\xi z^{-1})^{n_a-n_c} e(t) \tag{4.6.12a}$$

If also $n_a=n_c$ the result is a usual ARMAX model.

2. ARX in the Laguerre domain. Assume $n_a=n_b$, then we find for the time domain an ARMAX model with a fixed C–polynomial:

$$A_t y(t) = B_t u(t) + (1-\xi z^{-1})^{n_a} e(t) \tag{4.6.12b}$$

3. Box–Jenkins leads to Box–Jenkins, we assume $n_f=n_b$ and $n_d=n_c$:

$$y(t) = \frac{B_t(z^{-1})}{F_t(z^{-1})} u(t) + \frac{C_t(z^{-1})}{D_t(z^{-1})} e(t) \tag{4.6.12c}$$

4. FIR gives output error with a fixed F polynomial:

$$y(t) = \frac{B_t(z^{-1})}{(1-\xi z^{-1})^{n_b}} u(t) + e(t) \tag{4.6.12d}$$

5. OE yields OE, where we assume that $n_b=n_f$:

$$y(t) = \frac{B_t(z^{-1})}{F_t(z^{-1})} u(t) + e(t) \tag{4.6.12e}$$

Remark 4.6.3.

1. Notice that the occurence of $(1-\xi z^{-1})$ in (4.6.12) can in each case be considered as a pre–filter operation with a stable, minimum phase filter.

2. When any of the model structures of table 4.6.1 is used for identification, one has to restrict the polynomials in a certain sense in order to ensure the solvability of the problem (Ljung, 1987; Van den Hof, 1989). Consider the general model structure (4.6.8); if we do not impose any restrictions on the polynomials and we aim for instance at the minimization of $\|E\|$ then for all data the trivial solution will be $A\equiv B\equiv 0$, $C\equiv D\equiv F=I$. Another reason to use specific restrictions on the polynomials is to ensure identifiability or discriminablity, i.e. to make sure that different elements of the model class can be distinguished through applying the identification criterion. A commonly applied restriction is that the A– and C–polynomial in (4.6.8) are monic, i.e the leading coefficient matrix (A_0) is equal to the identity. In general different types of restrictions result in different model estimates (Van den Hof, 1989). From proposition 4.6.2 it follows that if polynomials in the model for the Laguerre domain are restricted to be monic, then the corresponding polynomials in the time domain model will not have this

property. This implies that specific polynomial restrictions in the Laguerre domain induce different polynomial restrictions in the time domain.

3. If we assume that the identification criterion aims at the least squares minimization of the (Laguerre) residual signal E, then the Parseval identity implies that this is equivalent with the least squares minimization of the time domain residual e. It follows that identification in the time domain, using a model structure of table 4.6.1 and a least squares criterion, is equivalent with applying the same criterion to pre–filtered time domain data, with the same model structure but different polynomial restrictions. This implies that some of these cases (4.6.12) are not advisable since the prefilter operation might remove too much information from the signal. As an example we consider case 4. Let $u_f=(1-\xi z^{-1})^{-n_b}u$, then (4.6.12d) can be written as a classic FIR model: $y(t)=B(z^{-1})u_f(t)+e(t)$. The order of the B polynomial is usually large. Figure 4.6.1 shows the (normalized) amplitude plot of the filter $(1-\xi z^{-1})^{-n_b}$ for n_b=1, 5, 10 and ξ=0.3, 0.7.

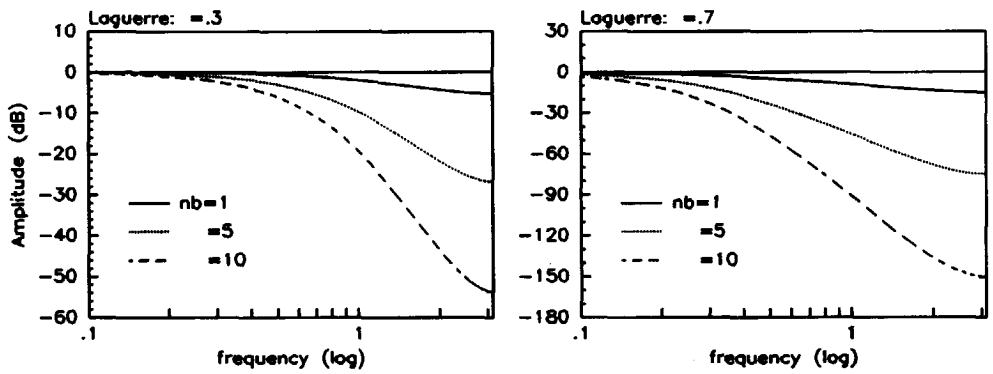


Fig. 4.6.1. Bode amplitude plot of prefilter $(1-\xi z^{-1})^{-n_b}$

This shows that a high value of n_b results in a complete cut–off at a specific frequency. Hence the high frequent information is removed from the input but will still be present in the output. This has the result that the model estimate cannot contain reliable information for the higher frequencies and that the input u_f might not be persistently exciting. This explains that the resulting model estimate can have a poor performance.

4. In the cases 1,3 and 5 the structure of the model does not alter, so the estimation procedure will be the same, but the polynomial restrictions will be different.

5. The application of case 2, the estimation of an ARX model in the Laguerre domain, can lead to satisfactory results. First of all it should be noted that the estimation of an ARX model with the common least squares criterion leads to a unique, analytically

solvable solution. However this model structure is not very well motivated from a practical point of vieuw, because it imposes a very restrictive structure on the noise model. Still this model structure is widely used because of the nice estimation properties (Swaanenburg and co–workers, 1985; Ljung, 1987). From (4.6.12b) we deduct that the time domain equivalent is a kind of ARMAX system, with a restricted noise polynomial or an ARX model with prefiltered input and output signals. As explained by Ljung (1987) this is equivalent with using a weighted least squares criterion and the filter characteristic, as shown in Fig. 4.6.1 shows that this weighting will improve the low frequent characteristics of the model estimate. An ideal situation will occur if $A_t(z^{-1})=(1-\xi z^{-1})^{n_a}$; in that case (4.6.12b) is equivalent with an output error structure:

$$y(t)=\frac{B_t(z^{-1})}{A_t(z^{-1})}\,u(t) + e(t) \qquad (4.6.13)$$

Practical experiences show that this method can lead to considerably better results than application of standard ARX techniques in the time domain. As mentioned above the use of this approach can also be motivated by the data–reduction property of the Laguerre transformation. This is emphasized by Nurges (1987), who claims that the optimal discount factor should be taken at the center of gravity of the poles of the original system. He proposes an iterative procedure, using the Laguerre difference equations as given by proposition 4.5.5, with the following steps:

a. Assign an initial value ξ_0 and transform the data to the Laguerre domain.
b. Estimate the parameters of the Laguerre difference equation and calculate the corresponding time difference equation.
c. Let ξ_i:= the center of gravity of the poles of the model.
d. Repeat the procedure till $\{\xi_i\}$ converges.

This procedure is in fact an iterative version of case 2 and has a close connection with the method of Steiglitz and McBride (1965), who propose to use the A_t polynomial that results from an ARX estimation as a new prefilter and to repeat the estimation with the filtered data. The aim of this procedure is to converge to an output error situation (4.6.13). The optimal $\hat{\xi}$ for the application of ARX in the Laguerre domain can be seen as the value of ξ with the property that $(1-\hat{\xi}z^{-1})^{n_a}$ approximates the resulting A polynomial in the best way.

6. One reason to use the Laguerre approach for these model structures can be the data reduction property of the Laguerre transformation. In general the data sequences are finite and they will not have a zero extension. As was explained in section 4.4 it is therefore not possible to calculate the correct Laguerre coefficients and thus we will have to work with approximations. o

EXAMPLE 4.6.4. We present an example of the application of case 2. The considered system is adapted from (Backx, 1987). The system has 2 inputs, 2 outputs and 6 states. It was excited with white noise and extra disturbance (white noise + offset) was added to the outputs. The signal to noise ratio was +1 dB. for each output. With these data we estimated ARX models in the time domain and in the Laguerre domain (ξ=.5), using a standard least squares criterion. Both models have McMillan degree 6. After the transformation of the Laguerre model to the time domain this resulted in two model estimates. In Fig. 4.6.2a,b we depict the step responses respectively the Bode amplitude plots of the original system and of both approximations. Since the data are produced in an output error configuration it is to be expected that the time domain model will only result in a fair approximation of the first 6 Markov parameters (Swaanenburg and co–workers, 1985; Van den Hof and Jansen, 1987). The performance of the Laguerre model is clearly much better.

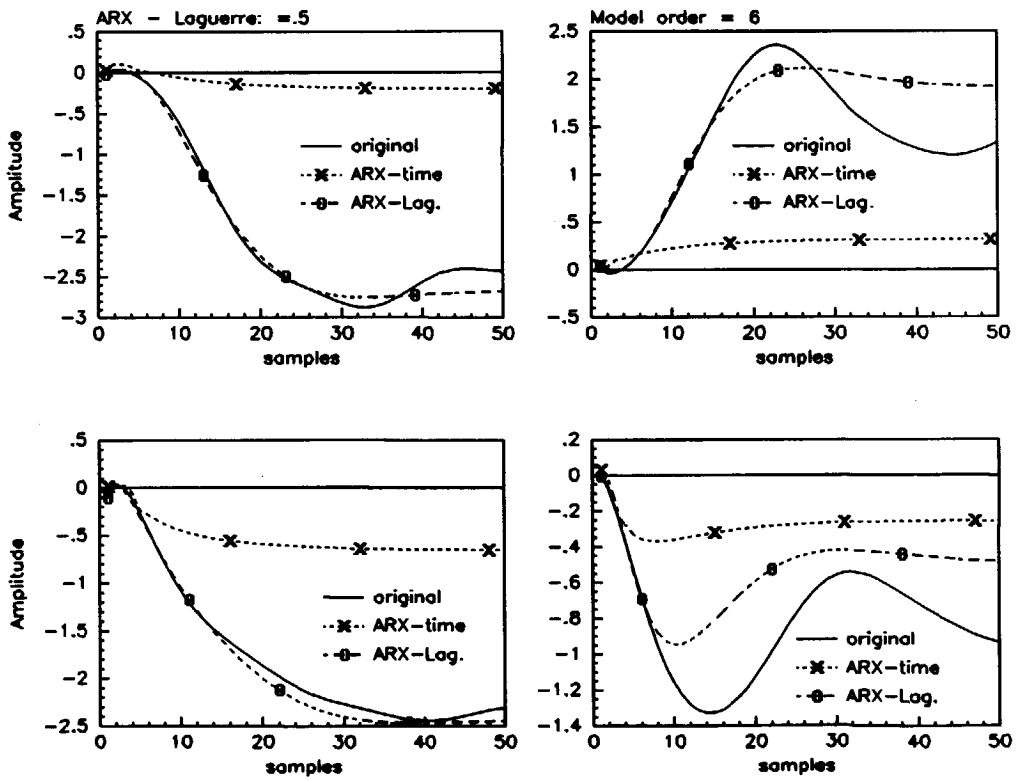


Fig. 4.6.2a. Estimation of ARX models in time and Laguerre domain: Step responses.

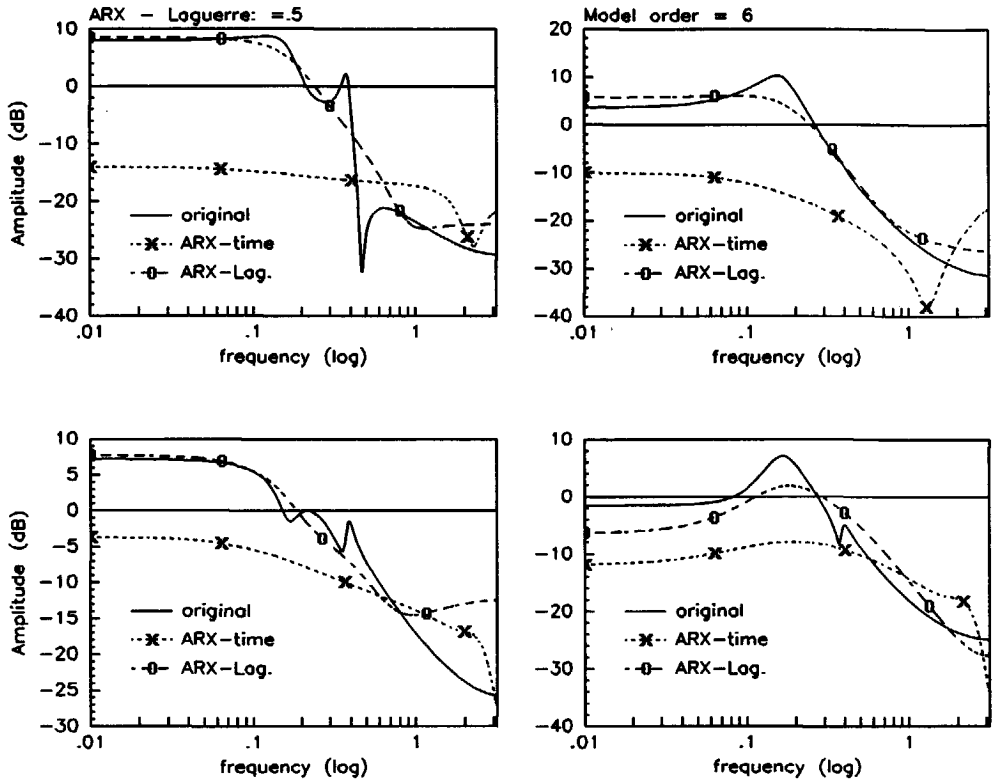


Fig. 4.6.2b. Estimation of ARX models in time and Laguerre domain: Bode amplitudes

4.6.4 IMPULSE RESPONSE APPROXIMATION

In the previous section we transformed the input/output data to the Laguerre domain and estimated a model for the Laguerre domain data. In this section we consider a method, that is based on the transformation of the impulse response of the system, which is equivalent with an expansion of the transfer function in the frequency domain in terms of the Laguerre transfer functions. This results in an estimation procedure that approximates a system on the basis of time–domain data. The main idea is adopted from Zervos and co–workers (1985,1988), who use the expansion of a transfer function in Laguerre transfer functions for control purposes. We apply this idea to identification and present a method to estimate the expansion coefficients; furthermore we propose three methods to derive low order models from the estimated coefficients.

The approach has a strong connection with the estimation of Markov–parameters of a transfer function. We use the expansion (2.4.4a) of a proper rational transfer function

G(z) in z^{-1}: $$G(z) = \sum_{i=0}^{\infty} M_i z^{-i} \tag{4.6.14}$$

with $\{M_i\}$ the set of Markov parameters (2.4.4b). A well known identification method (FIR, see table 4.6.1) is to estimate the first N Markov parameters, and to use an approximate realization method to create a state space model (Ho and Kalman, 1966; Silverman, 1971; Kung, 1978; Swaanenburg and co–workers, 1985; Backx, 1987; Van Helmont, Van der Weiden and Annevelt, 1990). The use of an FIR model with a least squares criterion has the favorable property that the optimal parameters can be calculated analytically. Furthermore an FIR model is an output error model, which is considered to be a resonable structure from a practical point of view. A drawback of such a structure is that often a large number of parameters must be estimated to obtain satisfactory results. This results in an over–parametrization, which is the reason why approximate realization methods are applied to obtain low order state space models. We will encounter the same problem with this approach and we present three different methods for the realization step.

Note that $\{z^{-i}\}_{i=0}^{\infty}$ represents the set of Laguerre transfer functions for ξ=0. From Zervos, Belanger and Dumont (1985,1988) we adopt the idea to generalize this method to $\xi \in [0,1)$ and we consider the expansion of the strictly proper part of G(z), $G_p(z)$, in terms of Laguerre transfer functions: $G_p(z) = \sum_{i=0}^{\infty} L_i \Psi_i(z)$

$$\text{or} \qquad G(z) = D + \sum_{i=0}^{\infty} L_i z^{-1} \Psi_i(z) \tag{4.6.15}$$

Zervos and co–workers use this approach for the fine tuning of PID controllers, showing that the coefficients L_i can be estimated without bias under white noise conditions. We use this concept for identification purposes by approximating the expansion (4.6.15) with a finite number of Laguerre functions and the (approximate) model is thus:

$$\hat{G}(z) = D + \sum_{i=0}^{n} L_i z^{-1} \Psi_i(z) \tag{4.6.16}$$

The identification method we propose consists of the estimation of the constant matrices $\{D,L_i\}$ with respect to a certain criterion. The parameters $\{D,L_i\}$ are thus estimated from the relation $$y(t) = \left[D + \sum_{i=0}^{n} L_i z^{-1} \Psi_i(z) \right] u(t) + e(t) \tag{4.6.17a}$$

which by replacing the brackets can also be formulated as

$$y(t) = Du(t) + \sum_{i=0}^{n} L_i \left[\Psi_i(z) z^{-1} u(t) \right] + e(t) \tag{4.6.17b}$$

and if we define the *orthonormal states* [4] $X_k(t)$ by

$$X_k(t) := \Psi_k(z) z^{-1} u(t) \quad k \in [0,n] \tag{4.6.18}$$

[4]This does not imply that the functions $X_k(t)$ are orthonormal functions, but that they can be considered as the state vectors of Laguerre systems $\Psi_k(z)$.

then substitution of these states in equation (4.6.17b) yields

$$y(t) = Du(t) + \sum_{i=0}^{n} L_i X_i(t) + e(t) \qquad (4.6.19)$$

These orthonormal states can be calculated directly from the input in an efficient way by using the structure of the generating Laguerre transfer functions (4.2.9); we are left with a regression model (4.6.19) where $\{D,L_i\}$ are the unknown parameters. For identification these parameters must be estimated with respect to a specific criterion; we propose to use a standard least squares criterion, which leads to a unique solution.
This procedure is equivalent with the Laguerre approximation of the functions defined by the sequence of Markov parameters; if G(z) has Markov parameters $\{M_t, t\in\mathbb{N}^0\}$ and if we consider the functions $f_{ij}(t):=(M_{t+1})_{ij}$ $(f_{ij}\in\ell_2)$ then the parameters $\{(L_k)_{ij}\}$ are the Laguerre coefficients of the functions f_{ij}. This is explained in detail in chapter five.

After the estimation of D and a finite number L_i's we have to consider the fact that the model (4.6.16) might be of a high order and will probably be over–parametrized. This observation leads to the use of realization algorithms if FIR models are considered and we discriminate between three different approaches that can be used for this approximate realization, or equivalently model reduction.

1. Realize an exact state space model and use a model reduction method

Assume for simplicity that we are dealing with a SISO system. First we consider the case $\xi=0$. Suppose we have a sequence of Markov parameters $\{M_i, i=0,1,\cdots n\}$ and we wish to find a state space model of a system that has $\{M_i\}$ as its first (n+1) Markov parameters. Such a system is given by the realization [A,B,C,D] with

$$A=\begin{bmatrix}0 & & \\ 1 & \ddots & 0 \\ & \ddots & 1\end{bmatrix} \quad B=\begin{bmatrix}1\\0\\0\end{bmatrix} \quad C=[M_1|M_2|\cdots M_n] \quad D=M_0 \qquad (4.6.20)$$

It can readily be verified that $CA^{i-1}B=M_i$ for $i=1,\cdots,n$. For the Laguerre case we can define an exact model of (4.6.16) with the realization $[A_1,B_1,C_1,D_1]$ where:

$$A_1=T_{n-1} \quad B_1=\underline{\psi}_{n-1}(0) \quad C_1=[L_0|L_1|\cdots|L_n] \quad D_1=D \qquad (4.6.21)$$

where $\underline{\psi}_{n-1}$ and T_{n-1} are defined by (4.3.3,4). Consider therefore the transfer function:

$$z[zI-T_{n-1}]^{-1}\underline{\psi}_{n-1}(0) \qquad (4.6.22)$$

The Markov parameters of (4.6.22) are $M_i=T_{n-1}^{i}\underline{\psi}_{n-1}(0)$ $i\in\mathbb{N}^0$ and propositions 4.2.4 and 4.3.4 show that these are also the Markov parameters of the transfer function

$$\underline{\Psi}_n(z):= [\Psi_0(z)|\Psi_1(z)|\cdots|\Psi_{n-1}(z)]^T \qquad (4.6.23)$$

Therefore we can identify both transfer functions (4.6.22) and (4.6.23). We conclude that $D_1+C_1[zI-A_1]^{-1}B_1 = D + z^{-1}[L_0|L_1|\cdots|L_n]\times \underline{\Psi}_n(z)$ and this implies

$$D_1+C_1[zI-A_1]^{-1}B_1 = D+\sum_{i=0}^{n} L_i z^{-1}\Psi_i(z)$$

This proves that (4.6.21) is a state space realization for (4.6.16). This procedure results in a high order model with all poles in $z=\xi$ and this can be reduced to a low order approximation with model reduction techniques, as described in chapter three. A nice feature is that the matrices $\{A_1,B_1\}$ are already in input balanced form (the controllability Gramian is I); this saves computing time if one of the reduction methods based on balanced realizations is used. o

2. Approximate realization with the impulse response

Calculate the Markov parameters of the system (4.6.16). Proposition 4.2.4 shows that these parameters are given by: $M_0=D$ and $M_i=\sum_{k=0}^{n} L_k\psi_k(i)$. This calculation can be performed with the network of Fig. 4.4.1. The Markov parameters can be used as the input to a standard approximate realization method and this results in a low order state space approximation. For $\xi=0$ this coincides with the identification method mentioned above (estimation of Markov parameters and realization of a state space model). o

3. Approximate realization of $\{L_i\}$

We use the transfer function transformation property (4.5.5,6), through the following observation:

$$G(z)= D + \sum_{i=0}^{\infty} L_i z^{-1}\Psi_i(z) = D + \sum_{i=0}^{\infty} L_i\sqrt{(1-\xi^2)}\,\frac{(1-z\xi)^i}{(z-\xi)^{i+1}}$$

so

$$G(z) = D + \frac{\sqrt{(1-\xi^2)}}{(z-\xi)}\sum_{i=0}^{\infty} L_i\,\frac{(1-z\xi)^i}{(z-\xi)^i} = D + \frac{\sqrt{(1-\xi^2)}}{(z-\xi)}\sum_{i=0}^{\infty} L_i\,[\pi_\xi^{-1}(z)]^i \qquad (4.6.24)$$

where $\pi_\xi(z)$ is defined by (4.5.5b). Now let $W(z)$ be the system that has $\{L_i\}$ as its Markov parameters:

$$W(z) = \sum_{i=0}^{\infty} L_i z^{-i} \qquad (4.6.25)$$

Apply the inverse Laguerre system transformation on $W(z)$: $\tilde{W}(z)=\pi_\xi^{-1}\big[W(z)\big]$, then it follows that

$$\tilde{W}(z):=\sum_{i=0}^{\infty} L_i\,[\pi_\xi^{-1}(z^{-1})]^i \qquad (4.6.26)$$

and we may conclude that the combination of (4.6.24) and (4.6.26) yields

$$G(z) = D + \frac{\sqrt{(1-\xi^2)}}{(z-\xi)}\,\tilde{W}(z) \qquad (4.6.27)$$

Since we will have estimated only a finite number of Laguerre parameters $\{L_i\}$, the realization of the system $W(z)$ (4.6.25) is an approximation. The reduction step can either be the use of approximate realization methods to create $W(z)$, or the application of model reduction procedures to reduce the order of $G(z)$. o

Remark 4.6.5. An interesting feature of this method (estimation of Laguerre parameters) is that it not only estimates the first part of the impulse response, as is the case with FIR ($\xi=0$), but because each $\Psi_i(z)$ has an extension the method makes an estimate over

a larger horizon. This implies that the error, caused by the finiteness of the number of parameters, will be more or less smeared out over a larger frequency range then in the case of FIR estimation, where the error is mostly low frequent. Experiments with this method show promising results, one of the main features being that the number of Markov parameters necessary to get good models is considerably larger than the number of Laguerre parameters. A direct consequence of this observation is that the third realization method which was proposed will not always be applicable, because the number of estimated L_i's may be too small for realization of a state space model. o

EXAMPLE 4.6.6. To illustrate this method we used the same input/output data as in example 4.6.4 and we compare the results of estimating 40 Markov parameters and of estimating 15 Laguerre parameters (ξ=0.7). We used an approximate realization method (Kung, 1979) to create a 6th order state space model from the Markov parameters. With the Laguerre parameters we created a 6th order model by application of the first realization method that we proposed (exact realization and model reduction, in this case

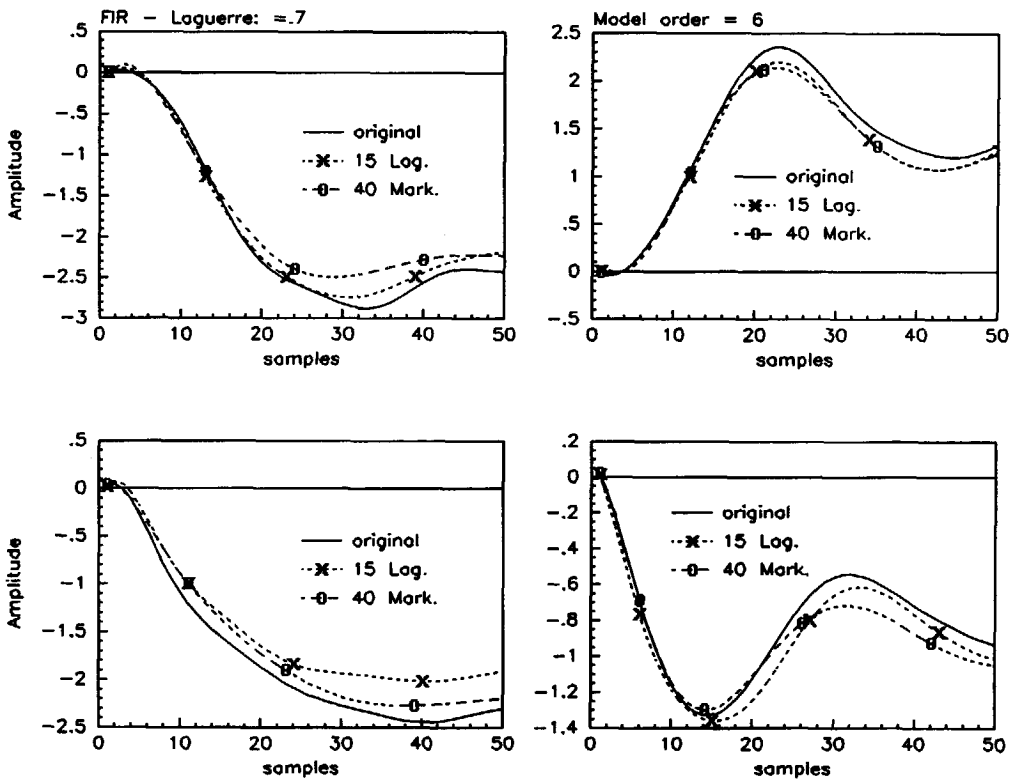


Fig. 4.6.3a. Estimation of 15 Laguerre and 40 Markov parameters: Step responses

balanced reduction). Fig. 4.6.3a and 4.6.3b depict the step responses and Bode amplitudes of the resulting estimates.

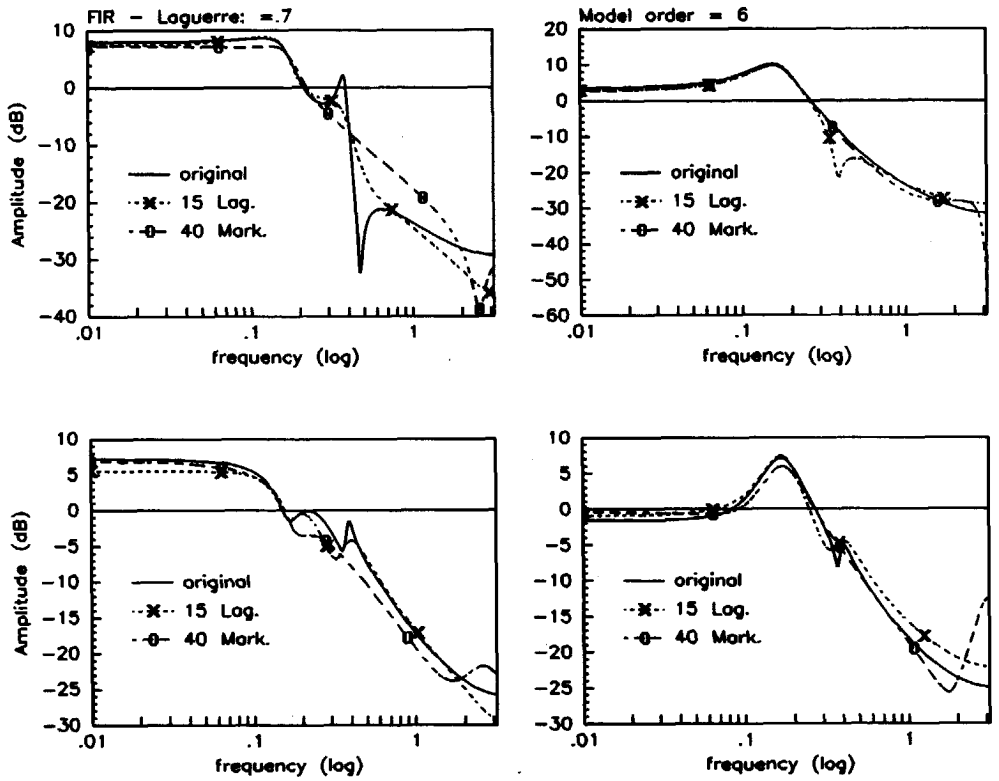


Fig. 4.6.3b. Estimation of 15 Laguerre and 40 Markov parameters: Bode amplitudes

It is apparent that both approximations are very alike, which shows that this method leads to a considerable reduction of the number of parameters that have to be estimated. We note here that the number of 40 Markov parameters was necessary to obtain reasonable results. o

4.7 MODEL REDUCTION.

In this section we explain the relation between the Laguerre system transformation and two model reduction methods, the reduction based on balanced realizations (section 3.5) and Hankel norm approximation (section 3.6). We present this relation in the form of a

commutation diagram. First we consider the procedure that results from combining the ω–transformation with the Laguerre system transformation. This results in the continuous time equivalent of the (discrete) Laguerre transformation and we will use it for the relation of Laguerre transformations and model reduction.

We have defined the Laguerre system transformation for discrete time systems; it is also possible to consider the transformation of continuous time system which is induced by the combination of the discrete Laguerre transformation with the ω–transformation. This turns out to be a very simple transformation, which gives a clear insight in the underlying mechanism. The following definition and proposition explain what we mean by this, where we use the following indexes:

t:(time) c:(continuous)
ℓ:(Laguerre) d:(discrete) to distinguish between different domains.

Definition 4.7.1. Let $G_t^c(s)$ be a FDLTS system. The continuous time Laguerre transform $G_\ell^c(s)$ of $G_t^c(s)$ is defined as the result of the following procedure:

Let $G_t^d(z)$ be the (discrete time) ω–transform of $G_t^c(s)$ and $G_\ell^d(z)$ the Laguerre transform of $G_t^d(z)$. Then $G_\ell^c(s)$ is the ω–transform of $G_\ell^d(z)$. o

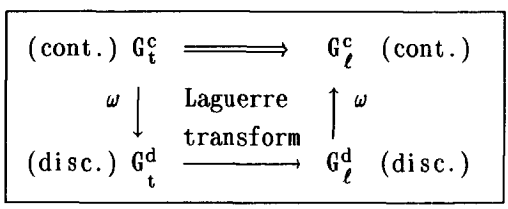


Fig. 4.7.1. Continous time Laguerre system transformation.

This transformation is visualized in Fig. 4.7.1 and the next proposition shows that this transformation can be performed with a simple calculation:

Proposition 4.7.2. Let $G_t^c(s)$ be a FDLTS system and let $G_\ell^c(s)$ be the Laguerre transform of $G_t^c(s)$ as defined by definition 4.7.1, where ξ is the degree of freedom of the Laguerre polynomials. Then $G_\ell^c(s) = G_t^c(\beta s)$ with $\beta = \frac{1-\xi}{1+\xi}$ (4.7.1) o

Proof: From definition 2.6.1 and theorem 4.5.1: $G_\ell^c(s)=G_\ell^d(\frac{1+s}{1-s})$ and $G_\ell^d(z)=G_t^d(\frac{\xi+z}{1+\xi z})$

$$\Rightarrow G_\ell^c(s) = G_t^d\left(\frac{\xi + \frac{1+s}{1-s}}{1+\xi\times\frac{1+s}{1-s}}\right) = G_t^d\left(\frac{1 + \frac{1-\xi}{1+\xi}\times s}{1 - \frac{1-\xi}{1+\xi}\times s}\right) \overset{\text{(definition 2.6.1)}}{=} G_t^c\left(\frac{1-\xi}{1+\xi}\times s\right) \qquad \square$$

This proposition gives a clear view of the underlying mechanism and if we define the obvious transformation of a state space realization [A,B,C,D] of $G_t(s)$ by

$$A \to \beta^{-1}A \qquad B \to \beta^{-\frac{1}{2}}B \qquad C \to \beta^{-\frac{1}{2}}C \qquad D \to D \tag{4.5.8}$$

then it is immediate that under this transformation the properties mentioned in proposition 4.5.2 are invariants (Grammians, static gain, Hankel singular values).
If we consider the application of a model reduction method to a FDLTS system G(p) as a system transformation, then we have defined three different transformations in this thesis:

a. the ω–transformation
b. the Laguerre transformation
c. a model reduction method

The following theorem shows how these transformations are connected if we consider for (c) either the Hankel norm approximation or the reduction method based on balanced realizations. In order to distinguish between different domains and McMillan degrees we index a system by $G_b^a(p)$ where:

a is the McMillan degree of the system
b=t (time domain) or b=ℓ (Laguerre domain)
p=s (continuous) or p=z (discrete)

THEOREM 4.7.3. Let $G_t^n(s)$ be a FDLTS system and let $\mathcal{L}^\xi$ denote the Laguerre system transformation induced by Laguerre polynomials with discount factor ξ. Furthermore let $\mathcal{R}_i$ denote a model reduction method and let $\alpha \geq 0$. Then the following diagram holds:

$$\begin{array}{ccccccccc}
G_t^n(s) & \underset{\mathcal{L}^{-\xi}}{\overset{\mathcal{L}^\xi}{\rightleftarrows}} & G_\ell^n(s) & \underset{\omega^{-1}}{\overset{\omega}{\rightleftarrows}} & G_\ell^n(z) & \underset{\mathcal{L}^{-\xi}}{\overset{\mathcal{L}^\xi}{\rightleftarrows}} & G_t^n(z) & \underset{\omega}{\overset{\omega^{-1}}{\rightleftarrows}} & G_t^n(s) \\
\mathcal{R}_1 \downarrow & & \mathcal{R}_2 \downarrow & & \mathcal{R}_3 \downarrow & & \mathcal{R}_4 \downarrow & & \mathcal{R}_1 \downarrow \\
G_t^k(s) & \underset{\mathcal{L}^{-\xi}}{\overset{\mathcal{L}^\xi}{\rightleftarrows}} & G_\ell^k(s) & \underset{\omega^{-1}}{\overset{\omega}{\rightleftarrows}} & G_\ell^k(z) & \underset{\mathcal{L}^{-\xi}}{\overset{\mathcal{L}^\xi}{\rightleftarrows}} & G_t^k(z) & \underset{\omega}{\overset{\omega^{-1}}{\rightleftarrows}} & G_t^k(s)
\end{array}$$

With:

either 1. $\mathcal{R}_1=\mathcal{R}_2=\mathcal{R}_3=\mathcal{R}_4=$ optimal Hankelnorm approximation

or 2. $\mathcal{R}_1=\mathcal{GB}_k^\alpha$; $\mathcal{R}_2=\mathcal{GB}_k^\gamma$; $\mathcal{R}_3=\mathcal{GB}_k^\delta$; $\mathcal{R}_4=\mathcal{GB}_k^\nu$;

where $\beta=\frac{1-\xi}{1+\xi}$; $\gamma=\alpha\beta^{-1}$; $\delta=\frac{\beta+\alpha}{\beta-\alpha}$; $\nu=\frac{1+\alpha}{1-\alpha}$. $\circ$

Proof: See appendix 4D.

Remark 4.7.4. This theorem explains how all these transformations are connected. It shows for instance that an optimal Hankel norm approximation of a given FDLTS system can be calculated along various lines. A well known example is the solution of the discrete Hankel norm approximation problem by transforming the problem to a continuous time problem and transforming the solution back to the discrete time domain (Glover, 1984). In section 3.6 we remarked that a direct solution of the discrete Hankel norm approximation problem is not available. We proposed several approaches to solve this problem, but (see example 3.6.10 and remark 3.6.11) noted that there are systems for which these approaches cannot be succesfully applied. We mentioned the suspicion that those systems have zeros in z=0. If this suspicion is correct then theorem 2.7.3 opens the way to solve the problem using the Laguerre transformation, because there always exists a value of ξ such that the Laguerre transformation of this system has no zeros in z=0. The following example illustrates that the application of this idea for the system of example 3.6.10 makes it possible to use the approach of proposition 3.6.9. o

Example 4.7.5. In example 3.6.10 we presented a system G(z) for which the extension procedure of proposition 3.6.9 could not be applied. We will show that there exists a ξ such that this proposition can be applied to the Laguerre transformation of G(z). This system G(z) has a balanced realization [A,B,C,D]

$$A=\frac{1}{2}\cdot\begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix}\qquad B=C^*=\begin{bmatrix}1&0\\0&1\\0&0\end{bmatrix}\qquad D=0\qquad \text{w.r.t. } \Sigma=\frac{1}{15}\begin{bmatrix}16&0&0\\0&20&0\\0&0&4\end{bmatrix}$$

Let L(z) be the Laguerre transformation of G(z); a balanced realization of L(z) is given by [F,G,H,J] (apply the formulas of theorem 4.5.1):

where

$$F=\frac{1}{4-\xi^2}\begin{bmatrix}-3\xi & 0 & 2-2\xi^2\\ 0 & (2+\xi)(1-2\xi) & 0\\ 2-2\xi^2 & 0 & -3\xi\end{bmatrix}\qquad G=H^*=\frac{\sqrt{1-\xi^2}}{4-\xi^2}\begin{bmatrix}4&0\\0&4+2\xi\\2\xi&0\end{bmatrix}$$

A necessary condition to apply proposition 3.6.9 is the existence of a matrix $\tilde{D}$ such that (3.6.23d) is fulfilled: $[0\ 0\ 1][F^*\Sigma G + H^*\tilde{D}]=0$.

It is straightforward that this results in an equation of the form:

$$[c_1\ 0] + [\xi c_2\ 0]\tilde{D} = 0$$

where c_1 and c_2 are constants with $c_2\neq 0$. If $\xi\neq 0$ then such a $\tilde{D}$ will always exist. Furthermore if $\tilde{D}$ is chosen symmetric, then it can be verified that also (3.6.23e) is fulfilled. Proposition 3.6.9 shows that with such a matrix $\tilde{D}$ the system $L_e(z)$ defined by:

$$L_e(z):=z^{-1}(L(z)-J+\tilde{D})$$

has a Hankel singular value $\sigma_3=\frac{4}{15}$ with multiplicity 2. It would lead to far to present all te calculations, but it can be verified that it is possible to calculate an optimal Hankel norm solution $\hat{L}_e(z)$ to $L_e(z)$ with formula (3.6.22): $\|L_e-\hat{L}_e\|_H = \sigma_3$ and that $\hat{L}(z):=z\hat{L}_e(z)$ is an optimal Hankel norm to L(z) with McMillan degree 2. Create $\hat{G}(z)$, the inverse

Laguerre transform of $\hat{L}(z)$, then it follows that $\hat{G}$ is an optimal Hankel norm approximation to G. Hence we showed that for this example the Laguerre transformation can be used to solve the optimal Hankel norm approximation. o

Remark 4.7.6. We do not suggest that the procedure described in example 4.7.5 should be used to calculate Hankel norm approximations. We merely wish to point out that the Laguerre transformation is quite powerful. It must be concluded that for the solution of the problem we still need a transformation and that a direct solution is not available. o

4.8 DISCUSSION

In this chapter it has been shown that the set of finite difference Laguerre polynomials, which form an orthonormal basis for $\ell_2[0,\infty)$, can be used with simple computations to transform time series and finite dimensional linear discrete time systems to the so called Laguerre domain. These calculations can be performed in an efficient way because of the shift structure of these polynomials. While the transformation of time series may may be used to obtain a considerable data reduction, the system transformation does not alter the main characteristics of the systems. We derived the continuous time analog of the Laguerre system transformation which shows clearly the underlying mechanism of these transformations and showed that the inverse Laguerre transformation is a dual version of the forward transformation.

These properties led to 3 different types of algorithms to deal with the problem of approximate system identification, of which the Laguerre ARX method and the impulse approximation method show promising results. The connection between the different time and Laguerre domains and the application of model reduction techniques has been explained with the use of a commutative diagram. It has been explained that this connection might be used to solve the discrete Hankel norm approximation problem.

An open problem is the choice of the discount factor, which determines the Laguerre polynomials. It was made clear that this choice is of major importance if the Laguerre polynomials are applied for data transformation and for the purpose of system identification. A second open problem is the influence of the choice for the extension of data, that have to be transformed to the Laguerre domain.

The discount factor can be considered as a degree of freedom when Laguerre polynomials are applied to approximate identification; different values lead to different approximate models. For the algorithm based on the estimation of ARX models in the Laguerre domain this can partially be explained as the use of different prefilters and different polynomial restrictions. The exact influence of this degree of freedom on the resulting model estimates, especially in the case of finite data, remains unclear.

4.9 APPENDICES OF CHAPTER 4

Appendix 4A. proof of proposition 4.4.2.

Denote the functions from $\Gamma(\xi)$ with ψ and those of $\Gamma(-\xi)$ with φ. It is our aim to show that $\psi_k(t)=\varphi_t(k)$, $k,t\in\mathbb{N}^0$. We use expression 4.2.7. with $\eta = 1-\xi^2$

$$\psi_k(t)= \sqrt{\eta}\sum_{j=0}^{k}\left[(-1)^{k+j}\binom{k}{j}\binom{t+k-j}{k}\xi^{t+k-2j}\right]$$

$$\varphi_t(k)= \sqrt{\eta}\sum_{j=0}^{t}\left[(-1)^{t+j}\binom{t}{j}\binom{t+k-j}{t}(-\xi)^{t+k-2j}\right] = \sqrt{\eta}\sum_{j=0}^{t}\left[(-1)^{k+j}\binom{t}{j}\binom{t+k-j}{t}\xi^{t+k-2j}\right]$$

Suppose that $k\geq t$:

$$\psi_k(t) = \sqrt{\eta}\sum_{j=0}^{t}\left[(-1)^{k+j}\binom{k}{j}\binom{t+k-j}{k}\xi^{t+k-2j}\right] + \sqrt{\eta}\sum_{j=t+1}^{k}\left[(-1)^{k+j}\binom{k}{j}\binom{t+k-j}{k}\xi^{t+k-2j}\right]$$

Now for $j>t$ we know that by definition of the binomial coefficient $\binom{t+k-j}{k} = 0$ so:

$$\begin{aligned}\psi_k(t)-\varphi_t(k) &= \sqrt{\eta}\sum_{j=0}^{t}\left[(-1)^{k+j}\binom{k}{j}\binom{t+k-j}{k}\xi^{t+k-2j} - (-1)^{k+j}\binom{t}{j}\binom{t+k-j}{t}\xi^{t+k-2j}\right]\\ &= \sqrt{\eta}\sum_{j=0}^{t}\left[(-1)^{k+j}\xi^{t+k-2j}\left[\binom{k}{j}\binom{t+k-j}{k}-\binom{t}{j}\binom{t+k-j}{t}\right]\right]. \qquad (4A.1)\end{aligned}$$

Now examine the product of the binomial coefficients:

$$B_j := \binom{k}{j}\binom{t+k-j}{k} = \frac{k!}{(k-j)!\ j!}\ \frac{(t+k-j)!}{(t-j)!\ k!} = \frac{t!}{(t-j)!\ j!}\ \frac{(t+k-j)!}{(k-j)!\ t!} = \binom{t}{j}\binom{t+k-j}{t}$$

This shows that each term in the right hand side of (4A.1)is identical to zero and we may conclude that $\psi_k(t)-\varphi_t(k) = 0$ for $k\geq t$. The symmetric argument shows the same for $k\leq t$ and hence $\Gamma^T(\xi) = \Gamma(-\xi)$ □

Appendix 4B proof of theorem 4.5.1.

part 1: Denote by X_k,U_k,Y_k the Laguerre coefficients of resp. $x(t),u(t)$ and $y(t)$.

$$\begin{aligned}x(t+1) &= Ax(t) + Bu(t)\\ y(t) &= Cx(t) + Du(t)\\ x(0) &= x_0\end{aligned}$$

Multiply with $\psi_k(t)$ and sum over t:

$$\tilde{X}_k = AX_k + BU_k \qquad (4B.1)$$

$$\text{where } \tilde{X}_k := \sum_{t=0}^{\infty} x(t+1)\psi_k(t) \qquad (4B.2)$$

$$Y_k = CX_k + DU_k \qquad (4B.3)$$

We have to evaluate $\tilde{X}_k$, for which we express $x(t+1)$ with (4.3.2):

$$x(t+1) = \sum_{k=0}^{\infty} X_k\psi_k(t+1) \overset{4.2.8}{=} \sum_{k=0}^{\infty} X_k\left[\xi\psi_k(t) + (1-\xi^2)\sum_{t=0}^{k-1}(-\xi)^{k-i-1}\psi_i(t)\right]$$

Substitute this expression in (4B.2): $\tilde{X}_k = \xi X_k + (1-\xi^2)\sum_{i=0}^{\infty}(-\xi)^{i-1} X_{k+i}$ so

$$\tilde{X}_k + \xi\tilde{X}_{k+1} = \xi X_k + X_{k+1} \qquad (4B.4)$$

Add (4B.1) with $\xi\times$(4B.1) with k→k+1:

$$\tilde{X}_k + \xi\tilde{X}_{k+1} = AX_k + BU_k + \xi AX_{k+1} + \xi BU_{k+1} \qquad (4B.5)$$

Substract (4B.4).from (4B.5):

$$X_{k+1} - \xi AX_{k+1} - \xi BU_{k+1} = AX_k + BU_k - \xi X_k \qquad (4B.6)$$

Let $$z_k = (I - \xi A)X_k - \xi BU_k \qquad (4B.7)$$

we obtain $z_{k+1} = Fz_k + GU_k$

$Y_k = Hz_k + JU_k$ with {F,G,H,J} defined in 4.5.3.

These equations are referred to as the **Laguerre state equations** of the system.
The initial conditions for the 'Laguerre system' follow from (4B.7):

$$z_0 = (I-\xi A)X_0 - \xi BU_0 = \sum_{t=0}^{\infty}[\, x(t) - \xi x(t+1)\,]\psi_0(t)$$

$$= \sqrt{(1-\xi^2)}\,[x(0) - \xi x(1) + \xi x(1) - \xi^2 x(2) + \xi^2 x(2) \cdots] = \sqrt{(1-\xi^2)}x(0) \qquad \text{(part 2)}\square$$

Part 1: Use the state space expression 4.5.2, then $L(z) = H[Iz-F]^{-1}G + J$

$$L(z) = (1-\xi^2)C\,[I-\xi A]^{-1}\,\{\, Iz-[A-\xi I][I-\xi A]^{-1}\,\}^{-1}[I-\xi A]^{-1}B + J$$

$$= (1-\xi^2)\,C\,[\, I(z+\xi) - A(1+\xi z)\,]^{-1}\,[I-\xi A]^{-1}B + J$$

$$= C\Big[(1-\xi^2)\,[\, I(z+\xi)-A(1+\xi z)\,]^{-1} + \xi I\Big]\,[I-\xi A]^{-1}B + D \qquad (4B.8)$$

We can simplify in (4B.8) the expression between large brackets:

$$[*] = (1-\xi^2)\,\{\, I(z+\xi)-A(1+\xi z)\,\}^{-1} + \xi I$$

$$= [I(z+\xi)-A(1+\xi z)]^{-1}\,[\, I-\xi^2 I+\xi zI+\xi^2 I-\xi A-\xi^2 zA]$$

$$= [I(z+\xi)-A(1+\xi z)]^{-1}\,(1+\xi z)\,[I-\xi A]$$

$$= [I(z+\xi)(1+\xi z)^{-1}-A]^{-1}[I-\xi A] \qquad (4B.9)$$

Substitution of (4B.9) in (4B.8) leads to:

$$L(z) = C\,[\, I(z+\xi)(1+\xi z)^{-1}-A\,]^{-1}B + D \quad \text{or} \quad L(z) = g(\pi(z)) \qquad \text{(part 1)}\square$$

Appendix 4C proof of proposition 4.5.2.

a. The invariance of stability follows from 4.5.3: Let λ be an eigenvalue of A, then $\mu = \frac{\lambda-\xi}{1-\xi\lambda}$ is an eigenvalue of F, and for $\xi \in (0,1)$ this is a one to one function mapping the unit disc into itself.

b. We will first prove that the controllability Grammian P is invariant. From (2.3.14) it follows that –if A is stable– P is the positive definite, symmetric solution to the Lyapunov equation $APA^T + BB^T = P$. Evaluate $FPF^T + GG^T$:

$$
\begin{aligned}
&FPF^T + GG^T \\
&\quad = (A-\xi I)(I-\xi A)^{-1}P(I-\xi A^T)^{-1}(A^T-\xi I) + (1-\xi^2)(I-\xi A)^{-1}BB^T(I-\xi A^T)^{-1} \\
&\quad = (I-\xi A)^{-1}\Big[\ (A-\xi I)P(A^T-\xi I) + (1-\xi^2)BB^T\ \Big](I-\xi A^T)^{-1} \\
&\quad = (I-\xi A)^{-1}\Big[\ APA^T + \xi^2 P - \xi AP - \xi PA^T + BB^T - \xi^2 BB^T\ \Big](I-\xi A^T)^{-1} \\
&\quad = (I-\xi A)^{-1}\Big[\ (APA^T + BB^T) + \xi^2(P-BB^T) - \xi AP - \xi PA^T\ \Big](I-\xi A^T)^{-1} \\
&\quad = (I-\xi A)^{-1}\Big[\ P + \xi^2 APA^T - \xi AP - \xi PA^T\ \Big](I-\xi A^T)^{-1} \\
&\quad = (I-\xi A)^{-1}\Big[\ (I-\xi A)P(I-\xi A^T)\ \Big](I-\xi A^T)^{-1} \\
&\quad = P
\end{aligned}
$$

This shows that P is the controllability Gramian of the Laguerre system and thus the controllability grammian is invariant. This proves immediately that controllability is invariant (the system is controllable iff P has full rank).

There is also another way to prove this. Consider the controllability matrices:
$R = [B\ AB\ A^2B\ \cdots]$ $S = [G\ FG\ F^2G\ \cdots]$, We know that $P=R^TR$
Define the following m flow systems: $v^i(t+1) = Av^i(t)$ with $[v^1(0)\cdots v^m(0)] = B$.
Calculate Laguerre coefficients V_k^i of $v^i(t)$ using (4B.8):
$(I-\xi A)V_{k+1}^i = (A-\xi I)V_k^i$ or $V_{k+1}^i = FV_k^i$.
Let $v(t) = [v^1(t)\ v^2(t)\cdots v^m(t)]$ and $V_k = [V_k^1\ V_k^2\cdots V_k^m]$
$V_0 = \sum_{t=0}^{\infty} v(t)\psi_0(t) = \sqrt{(1-\xi^2)}\sum_{t=0}^{\infty}(-\xi A)^T B = \sqrt{(1-\xi^2)}\,[I-\xi A]^{-1}B = G.$
Now we combine $V_{k+1}^i = FV_k^i$ and $V_0 = G$: $[V_0\ V_1\ V_2\cdots] = [G\ FG\ F^2G\ \cdots] = S.$
and this shows that $S = R\,\Gamma(\xi)$ (4C.1)
(Where Γ is defined by (4.4.1):. $\Gamma\Gamma^T = I$).
$\Rightarrow \quad SS^T = RR^T = P.$

c. Replacing A by A^T and B by C^T in (ii) gives the prove for the observability Grammian Q and observability and hence with b, also for minimality.

d. Because the singular values are the square roots of the eigenvalues of PQ, these must be invariant too.

e. Balancedness implies $P=Q=\Sigma$, the diagonal matrix with singular values. So the above proves immediately the invariance of this property.

f. The static gains of the systems are respectively g(1) and L(1). We know from (4.5.5): $L(z)=g(\pi(z)),\quad \pi(z) = \frac{\xi + z}{1 + \xi z}$ So $L(1)=g(\pi(1))=g(1)$ □

Appendix 4D proof of theorem 4.7.1

Part 1

Let [A,B,C,D] and [F,G,H,J] be state space models of $G_t^n(z)$ and $G_\ell^n(z)$. We use a property, used in the proof of proposition 4.5.2 (appendix 4C, equation (4C.1)):

if $R_t = [B\ AB\ A^2B\ \cdots]$ and $R_\ell = [G\ FG\ F^2G\ \cdots]$

$\Gamma(\xi)$ matrix with polynomials, $\Gamma(\xi)\Gamma^T(\xi) = I$ then $R_\ell = R_t \cdot \Gamma(\xi)$.

The same can be shown for the observability matrices: $Q_t = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix}$ and $Q_\ell = \begin{bmatrix} H \\ HF \\ HF^2 \\ \vdots \end{bmatrix}$ obey $Q_\ell = \Gamma^T(\xi)\, Q_t$. The corresponding Hankel matrices can be written as $\mathcal{H}_t = Q_t R_t$ and $\mathcal{H}_\ell = Q_\ell R_\ell$. So we get as a direct result: $\mathcal{H}_\ell = \Gamma^T(\xi)\ \mathcal{H}_t\ \Gamma(\xi)$. Let $\hat{\mathcal{H}_t}$ be the Hankel matrix of the optimal Hankel norm approximation $G_t(k)$, and $\{\sigma_1,\sigma_2,\cdots,\sigma_n\}$ be the ordered singular values of $\mathcal{H}_t$ and $\mathcal{H}_\ell$ ($\sigma_{i-1} \geq \sigma_i$). Then $\| \mathcal{H}_t - \hat{\mathcal{H}_t} \|_s = \sigma_{k+1}$, where $\|\cdot\|_s$ is the spectral norm. Now the spectral norm is invariant under unitary transformations,

$$\| \mathcal{H}_t - \hat{\mathcal{H}_t} \|_s = \| \Gamma^T(\xi)(\mathcal{H}_t - \hat{\mathcal{H}_t})\Gamma(\xi) \|_s = \| \mathcal{H}_\ell - \Gamma^T(\xi)\ \hat{\mathcal{H}_t} \Gamma(\xi) \|_s$$

This shows that $\Gamma^T(\xi)\ \hat{\mathcal{H}_t}\Gamma(\xi)$ is an optimal Hankel norm approximation to $\mathcal{H}_\ell$.

Because $\Gamma^T(\xi) = \Gamma(-\xi)$, the proof is the same the other way around. The ω-transformation also transforms Hankel norm reductions (Glover, 1984) and thus we showed the validity of part 1. (part 1)□

Part 2: From proposition 4.5.4 : $G_\ell^c(s) = G_t^c(\beta s)$ with $\beta = \frac{1-\xi}{1+\xi}$. Now let [A,B,C,D] be a balanced realization of $G_t^c(s)$ then it is immediate that $G_\ell^c(s)$ has a balanced realization $[\beta^{-1}A, \beta^{-\frac{1}{2}}B, \beta^{-\frac{1}{2}}C, D]$. $GB_k^\alpha(G_t^c)(s)$ has a realization $[A_r,B_r,C_r,D_r]$ with

$$A_r = A_{11} + A_{12}[\alpha I - A_{22}]^{-1}A_{21} \qquad B_r = B_1 + A_{12}[\alpha I - A_{22}]^{-1}B_2$$
$$C_r = C_1 + C_2[\alpha I - A_{22}]^{-1}A_{21} \qquad D_r D + C_2[\alpha I - A_{22}]^{-1}B_2]$$

and $GB_k^\gamma(G_\ell^c)(s)$ has a realization $[\hat{A},\hat{B},\hat{C},\hat{D}]$ with

$$\hat{A} = \beta^{-1}(A_{11} + A_{12}[\gamma\beta I - A_{22}]^{-1}A_{21}) \qquad \hat{B} = \beta^{-\frac{1}{2}}(B_1 + A_{12}[\gamma\beta I - A_{22}]^{-1}B_2)$$
$$\hat{C} = \beta^{-\frac{1}{2}}(C_1 + C_2[\gamma\beta I - A_{22}]^{-1}A_{21}) \qquad \hat{D} = D + C_2[\gamma\beta I - A_{22}]^{-1}B_2]$$

This shows – under the condition $\alpha = \gamma\beta$ – that $[\hat{A},\hat{B},\hat{C},\hat{D}] = [\beta^{-1}A_r, \beta^{-\frac{1}{2}}B_r, \beta^{-\frac{1}{2}}C_r, D_r]$ or $GB_k^{\gamma\beta}(G_\ell^c)(s) = GB_k^\alpha(G_t^c)(\beta s)$. Hence $\gamma = \alpha\beta^{-1}$

We use lemma 3.4.3 to check the other values of the reduction parameter:

$$\delta = \frac{1+\gamma}{1-\gamma} = \frac{\alpha+\beta}{\alpha-\beta} \quad \text{and} \quad \nu = \frac{\alpha+1}{1-\alpha}$$

(part 2)□

(theorem)□

5. TRANSFORMATION AND APPROXIMATION WITH SYSTEM BASED ORTHONORMAL FUNCTIONS

5.1 INTRODUCTION

In this chapter we elaborate further on the use of orthonormal or orthogonal functions in the context of linear systems. In the previous chapter it was shown that the Laguerre polynomials have very nice properties, which can be used in an efficient way to approach the system identification problem through an orthogonalization of the system behavior. The choice of the discount factor however is still somewhat arbitrary, and the choice of Laguerre functions, over for instance Walsh– Block–pulse functions or other orthogonal polynomials, like Legendre, Hermite or Tchebychev, is also an arbitrary one. Furthermore the results of application depend strongly on the set of Laguerre functions which were used. With system identification in mind and the idea to use orthonormal functions to approach this issue one might expect that for a specific problem there will be a specific set of orthonormal functions that is marked out for this purpose.
The question arises if linear systems give rise to orthonormal functions in a natural way. The answer to this question is affirmative and in this chapter we will derive the theory involved. We will show that every finite dimensional stable discrete time system gives rise to a complete set of orthonormal functions, based on input (or output) balanced realizations, or equivalently on the singular value decomposition of the Hankel matrix of the system. These functions are a generalization of the Laguerre polynomials and the discrete pulse functions. The theory is more or less based on the properties of discrete all–pass functions, treated in chapter two. In analogy with the Laguerre polynomials we can use these functions to transform time–series and linear systems to a so called orthogonal domain, leading to new system descriptions. In this setting new possibilities arise for the construction of approximate identification methods, with favorable properties as the use of simple estimation techniques and a systematic choice of prefilters.
In section 5.2 it will be shown that all–pass functions in a natural way lead to infinitely large sets of orthonormal functions, which in section 5.3 will be used in to derive a similar result for arbitrary FDLTS systems. It is demonstrated that under very weak conditions these sets of functions constitute an orthonormal basis of ℓ_2. Based on this property we can transform ℓ_2 time–series in these new basis functions, which is considered in section 5.4.

These signal transformations then again induce a transformation of FDLTS system, with the interesting property that the 'optimal' transformation more or less removes the dynamics of the system, thus leading to a system with only two Markov parameters. Section 5.5 is devoted to the use of these concepts in the context of approximate system identification. The chapter is reviewed section 5.6. The proofs of the theoretical results presented in this chapter are not so much complicated but very lengthy. They are mostly given in appendices at the end of the chapter.

5.2 ORTHONORMAL FUNCTIONS GENERATED BY ALL–PASS TRANSFER FUNCTIONS.

In this section we will show that a square stable all–pass function as defined by definition 2.7.3 gives rise to two sets of orthonormal functions. The main tool for this purpose is theorem 2.7.4, where we gave a characterization of square all–pass functions. The derivation is based on the fact that for a minimal realization of a FDLTS system with controllability matrix M_c the controllability Gramian P obeys the equality $P=M_cM_c^*$. If we consider the rows of M_c as discrete time functions, f_i: $\mathbb{N}^0 \to \mathbb{F}$, then the entries of P are the ℓ_2 inner products of these functions, which are well defined since the stability of the realization ensures that these functions are indeed in ℓ_2. This shows that in case $P=I$, these functions are mutually orthonormal in ℓ_2–sense. The next step is to construct an embedding of an all–pass functions with McMillan degree n within an all–pass function with McMillan degree k×n. The latter has a controllability matrix with k×n rows and these will again be orthonormal. If we let $k \to \infty$ this leads to an infinite number of rows or orthonormal functions.

Let $G(z):=D+C[zI-A]^{-1}B$ be a square stable all–pass function with McMillan degree n and let [A,B,C,D] be a minimal realization of G. Then theorem 2.7.4 shows that the Gramians P and Q of such a realization have the property that $PQ=I$. The stability assures that $P,Q>0$. We can always find a minimal realization with $P=Q=I$, using well known balancing techniques (Laub, 1980; Moore, 1981; Enns, 1984). We assume in the sequel that we are dealing with such a realization, satisfying:

$$AA^* + BB^* = A^*A + C^*C = I \tag{5.2.1}$$

The stability and minimality of the realization imply that the controllability and observability matrix, M_c respectively M_o, have an orthonormality property.

$$M_cM_c^* = P = I \tag{5.2.2a}$$

$$M_o^*M_o = Q = I \tag{5.2.2b}$$

Hence we can consider the rows of M_c (and the columns of M_o) as n mutually

orthonormal discrete time functions $\mathbb{N}^0 \to \mathbb{F}$. The following remark shows how this is connected with another well known orthonormality concept.

Remark 5.2.1. For a realization, as described above, we can show that (5.2.2) also provides a singular value decomposition of the Hankel matrix $\mathscr{H}$ that corresponds with G(z). We know that $\mathscr{H} = M_o M_c$ and theorem 2.7.4 implies that the singular values of this matrix are all equal to one. This shows that a singular value decomposition of $\mathscr{H}$ is given by:

$$\mathscr{H} = U\Sigma V^*, \quad U=M_o, \quad \Sigma=I, \quad V=M_c^* \qquad (5.2.3)$$

o

These properties thus show that a stable square all–pass system in a natural way gives rise to two sets of n orthonormal functions, where n is the McMillan degree of the system. The next step is to enlarge the number of such functions, by embedding the realization of G in a realization of a stable all–pass function with larger McMillan degree. Note that if G(z) is a square stable all–pass function then also $G(z)G^*(z)$ is all pass, but if G was stable, GG^* will not be. Since G is square we can also consider $G^2(z) := \left[G(z)\right]^2$, which is clearly stable and all–pass. The following lemma shows that we can easily find a realization of $G^2(z)$ satisfying the property (5.2.1).

Lemma 5.2.2. Let G(z) be a square stable all–pass function with McMillan degree n and let [A,B,C,D] be a realization of G(z), balanced w.r.t. $\Sigma=I_n$. Then $G^2(z)$ is a square stable all–pass function which has a realization $[A_2,B_2,C_2,D_2]$ that is balanced w.r.t. $\Sigma_2=I_{2n}$, where

$$A_2=\begin{bmatrix} A & 0 \\ BC & A \end{bmatrix} \quad B_2=\begin{bmatrix} B \\ D \end{bmatrix} \quad C_2=[DC\ C] \quad D_2=D^2 \qquad (5.2.4)$$

o

Proof: See appendix 5A.

Again we can, as a result of lemma 5.2.2, consider the rows (columns) of the controllability (observability) matrix of the realization $[A_2,B_2,C_2,D_2]$ as 2×n orthonormal functions. Note that the structure of A_2 and B_2 (C_2) implies that the first (last) n controllability (observability) functions are exactly the controllability (observability) functions of the original realization [A,B,C,D]. In other words we extended the set of n orthonormal functions to a set of 2×n orthonormal functions. The next theorem generalizes this procedure to arbitrary powers of G(z).

Theorem 5.2.3. Let G(z) be a square stable all–pass function with McMillan degree n and let [A,B,C,D] be a realization of G(z), balanced w.r.t. $\Sigma=I_n$. Then $G^k(z)$ is a square stable all–pass function which has a realization $[A_k,B_k,C_k,D_k]$ that is balanced w.r.t. $\Sigma_k=I_{kn}$, where,

$$A_k = \begin{bmatrix} A & 0 & \cdots & \\ BC & A & 0 & \cdots \\ BDC & BC & A & \\ \vdots & \vdots & \ddots & \ddots \\ BD^{k-2}C & \cdots & & BC \; A \end{bmatrix} \qquad B_k = \begin{bmatrix} B \\ BD \\ BD^2 \\ \vdots \\ BD^{k-1} \end{bmatrix} \tag{5.2.5a}$$

$$C_k = [D^{k-1}C \cdots DC \; C] \qquad D_k = D^k \tag{5.2.5b}$$

o

Proof: See appendix 5B.

Again we conclude that the controllability (observability) functions of the realization $[A_k,B_k,C_k,D_k]$ constitute a set of k×n orthonormal functions. The structure of A_k makes it easy to see that the first (last) controllability (observability) functions induced by this realization are exactly the controllability (observability) functions induced by $[A_{k-1},B_{k-1},C_{k-1},D_{k-1}]$. The consequence is that, by letting $k\to\infty$, theorem 5.2.3 shows how we can create an infinite set of orthonormal functions.

Remark 5.2.4. As was mentioned in remark 5.2.1, the controllability and observability matrices of $[A_k,B_k,C_k,D_k]$ generate a singular value decomposition of the Hankel matrix, corresponding to $G^k(z)$. The structure of the realization (5.2.5) and the decomposition (5.2.3) show that we have actually extended the matrices of singular vectors U and V in (5.2.3) to U_k, V_k by adding extra columns, such that these extended matrices are still unitary and their product still has the Hankel property (i.e. $U_kV_k^*$ is a block Hankel matrix). o

Remark 5.2.5. It will be clear that instead of considering $G^k(z)$ we could also have chosen other extensions, for instance products of powers of G(z) and G(–z), which would also result in stable all–pass functions. It is straightforward to show that this will result in the same controllability (observability) matrix as the one we presented, pre–(post–) multiplied with a sign–matrix, and hence such extensions may be considered to be equivalent to the one presented. Extensions that include powers of $G^*(\pm z)$ will also lead to all–pass functions, but these will not necessarily be stable and proper. Furthermore the McMillan degree of such extensions cannot be pre–specified, because the all–pass property $G(z)G^*=(z)$ shows that each occurrence of G(z) will cancel an occurrence of $G^*(z)$. We will need stability and properness to ensure that the resulting orthonormal functions are in ℓ_2 and to be able to express the extensions in state space form. Let for instance $G(z)=\frac{1}{z}$, which is the simplest example of a stable all–pass function, then $G^*(z)=z$ which is not proper. Another example is $G(z)=\frac{z-2}{2z-1}$, which is stable and all–pass, but $G^*(z)=\frac{1-2z}{2-z}$ is all–pass but unstable. Since the stability and properness of the extensions is of central importance in the theory presented, we use the extension as presented in this section. o

EXAMPLE 5.2.6. When theorem 5.2.3 is applied to $G(z)=\frac{1}{z}$, then it becomes clear that the extension is indeed quite natural. A balanced realization w.r.t. 1 of G is [0,1,1,0] with controllability matrix $M_c = [1\ 0\ 0\cdots]$. The matrices A_k, B_k and C_k are deduced from (5.2.5): $A_k=\begin{bmatrix}0 & & & \\ 1 & 0 & & 0 \\ & \ddots & \ddots & \\ & & 1 & 1\end{bmatrix}$, $B_k^* = [1\ 0\ \cdots\ 0]$, $C_k = [0\cdots 0\ 1]$ and the controllability matrix is $M_{k_c} = [I_k\ 0_k\cdots]$, a natural extension of M_c. o

5.3 ORTHONORMAL FUNCTIONS GENERATED BY GENERAL TRANSFER FUNCTIONS.

In this section we concentrate on general FDLTS systems and we use the results of the previous section concerning extensions of all–pass functions to define sets of orthonormal functions which are induced by an arbitrary stable transfer function G(z). This will be accomplished by splitting off an all–pass function and to use the theory developed in section 5.2 on this all–pass part. The line of reasoning is best understood by considering the Hankel matrix $\mathscr{H}$ of G(z). Let this matrix have a singular value decomposition,

$$\mathscr{H} = U\Sigma V^* \tag{5.3.1}$$

with $U^*U=V^*V=I$ and Σ the diagonal matrix with singular values. The unitarity of U and V implies that the columns of U and V can be seen as two sets of orthonormal discrete time functions. With the theory of section 5.2 we can extend such a set to an infinite number of mutually orthonormal functions, with the recursive structure which was implied by theorem 5.2.3. In general it will not be possible to extend U and V simultaneously, if we aim at this recursive structure, since the Hankel singular values of the system are not equal in general. This means we are not able to extend (5.3.1) to an extended Hankel matrix $\mathscr{H}_e = U_e\Sigma_e V_e^*$ such that $U_e=[\cdots,U,\cdots]$, $V_e=[\cdots,V,\cdots]$ and $\Sigma_e=\mathrm{diag}(\cdots,\Sigma,\cdots)$. We will concentrate on the extension of V, but everything can also be done for U. The orthonormal functions that are generated in this way represent dynamic behavior of the system G(z). It will follow that under weak conditions these functions constitute an orthonormal basis for ℓ_2 and that the procedure which is presented generalizes two well known sets of orthonormal functions.

Let G(z) be a p×m FDLTS system and let [A,B,C,D] be a (minimal) input balanced realization of G(z). Recall from definition 2.6.7 that such a realization has the property:

$$AA^*+BB^* = I \tag{5.3.2a}$$

$$A^*\Sigma^2A+C^*C = \Sigma^2 \tag{5.3.2b}$$

where Σ is the matrix with Hankel singular values. Let M_c and M_o be the controllability

respectively observability matrix of this realization, then we deduce from (2.4.7) that

$$M_c M_c^* = I \qquad M_o^* M_o = \Sigma^2 \tag{5.3.3}$$

This can be used to find a singular value decomposition of the Hankel matrix $\mathscr{H}$ of the system G(z):

$$\mathscr{H} = U\Sigma V^* \text{ with } U=M_o\Sigma^{-1} \text{ and } V=M_c^* \tag{5.3.4}$$

since $\mathscr{H} = M_o M_c = (M_o\Sigma^{-1})\Sigma M_c$ and $U^*U = V^*V = I$.

As stated we aim at an extension of the matrix V of right hand singular vectors, which in this case is the same as the controllability matrix. The controllability matrix $M_c = [B \mid AB \mid A^2B \cdots]$ is independent of the output matrices C and D and the extension of this matrix can be performed with application of the theory of section 5.2 if we can consider M_c (or V^*) as the controllability matrix of a realization of a square stable all–pass system. Hence we want to expand the matrices {A,B} with two new matrices $\{\tilde{C},\tilde{D}\}$, such that $[A,B,\tilde{C},\tilde{D}]$ is a (minimal) balanced realization of a square all–pass transfer function $\tilde{G}(z) := \tilde{D} + \tilde{C}[zI-A]^{-1}B$. Theorem 2.7.4 shows that for this purpose it is sufficient to require that $A^*A+\tilde{C}^*\tilde{C}=I$. The following lemma shows that this can be achieved through the singular value decomposition of the matrix A. It also presents two equations in terms of these matrices, which will be used for the proof of theorem 5.3.3.

Lemma 5.3.1. Let $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$, with A stable and $AA^*+BB^*=I$. Further let $A=U_aS_aV_a^*$ be a singular value decomposition of A and define:

$$F := U_aV_a^* \tag{5.3.5a}$$

$$\tilde{C} := B^*F \tag{5.3.5b}$$

then:

1. $A^*A + \tilde{C}^*\tilde{C} = I$ (5.3.6a)
2. $\exists\, \tilde{D} = \tilde{D}^* \in \mathbb{C}^{m\times m}$ such that
   a. $\tilde{G}(z) = \tilde{D} + \tilde{C}[zI-A]^{-1}B$ is all–pass (5.3.6b)
   b. $B\tilde{D} = -FA^*B$ (5.3.6c)
   c. $\tilde{D}\tilde{C} = -\tilde{C}AF$ (5.3.6d)

o

Proof: See appendix 5C.

Remark 5.3.2. Note that in lemma 5.3.1 we did not require that {A,B} are matrices that belong to an input balanced realization of a FDLTS system G(z), since there may exist many realizations [A,B,C,D] that obey $AA^*+BB^*=I$ but not $A^*\Sigma^2A+C^*C=\Sigma^2$ with Σ a real–valued diagonal matrix. However if this would be the case, then we can give the interpretation that the matrix $[B \mid AB \mid A^2B \cdots]$ is exactly the matrix with right hand side singular vectors of the Hankel matrix that corresponds to G(z). o

An interesting consequence of this lemma is that for a given controllability matrix (of a stable controllable pair) we can always define an all–pass system such that a realization of this system exists that has the same controllability matrix:

Let $M_c=[B|AB|A^2B\cdots]$ be given and let $P=P^*>0$ be the solution to the Lyapunov equation $APA^*+BB^*=P$ then the pair $\{P^{-\frac{1}{2}}AP^{\frac{1}{2}},P^{-\frac{1}{2}}B\}$ is in input balanced form. Lemma 5.3.1 shows the existence of a pair $\{C,D\}$ such that $G(z):=D+C[zI-P^{-\frac{1}{2}}AP^{\frac{1}{2}}]^{-1}P^{-\frac{1}{2}}B$ is all pass. This $G(z)$ can also be expressed by $G(z):=D+CP^{-\frac{1}{2}}[zI-A]^{-1}B$. We conclude that the all–pass system $G(z)$ has a realization $[A,B,CP^{-\frac{1}{2}},D]$ and this realization has the controllability matrix M_c. o

With lemma 5.3.1 we have explained how an all–pass function can be 'split off' from an arbitrary FDLTS system. The obvious next step is the combination of the results of theorem 5.2.3 and lemma 5.3.1, and derive how we can extend the unitary matrix V in (5.3.4) in a recursive way to an infinitely large unitary matrix. In other words, we have found the means to create an infinite set of mutually orthonormal functions, based on arbitrary FDLTS systems. This result is specified in the following theorem.

Theorem 5.3.3. Let $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$, with A stable and $AA^*+BB^*=I$. Further let $A=U_aS_aV_a^*$ [1] be a singular value decomposition of A and define:

$$F:= U_aV_a^* \tag{5.3.7a}$$

$$P:= -FA^* = -U_aS_aU_a^* \tag{5.3.7b}$$

$$X:= F(I-A^*A) = U_aV_a^* - U_aS_a^2V_a^* \tag{5.3.7c}$$

$$A_e:=\begin{bmatrix} A & 0 & \cdot & \cdot & \cdot \\ X & A & 0 & \cdot & \cdot \\ PX & X & A & 0 & \cdot \\ P^2X & PX & X & A & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \quad B_e:=\begin{bmatrix} B \\ PB \\ P^2B \\ \vdots \\ \vdots \end{bmatrix} \tag{5.3.7d}$$

$$M_{c_e}:= [B_e \mid A_eB_e \mid A_e^2B_e \cdots] \tag{5.3.8}$$

Then: 1. A_e is stable and $A_eA_e^*+B_eB_e^*=I_e$

2. The rows of M_{c_e} are mutually orthonormal o

Proof: Lemma 5.3.1 shows that there exist matrices C and D such that $G(z):=D+C[zI-A]^{-1}+B$ is all–pass with $BD=-FA^*B=PB$. Therefore $BD^k=P^kB$. Further $C=B^*F$ so $BC=BB^*F=[I-AA^*]F=F[I-A^*A]=X$. Substitution of these expressions for BD^k and BC in theorem 5.2.3 gives A_e and B_e as in (5.3.7d). The stability of A_e is a direct consequence of the stability of A. The orthonormality of M_{c_e} follows from the fact that $M_{c_e}M_{c_e}^* = B_eB_e^* + A_eM_{c_e}M_{c_e}^*A_e^*$ which with (1) implies $M_{c_e}M_{c_e}^*=I_e$. □

[1] This represents a freedom since the svd of a matrix is not necessarily unique. However it is straightforward that the matrices (5.3.7d) are unique if A is non–singular.

This theorem thus shows how a pair $\{A,B\}$ with A stable and $AA^*+BB^*=I$ gives rise to an infinite set of (discrete) orthonormal functions, that are defined by the rows of the controllability matrix M_{c_e}. With this result we can generalize the extension concept to arbitrary controllable pairs $\{A,B\}$ with A stable. The controllability and stability imply that such a pair has a positive definite Gramian and therefore a similarity transformation exists which transforms this Gramian into an identity matrix. Application of theorem 5.3.3 then again defines a set of orthonormal functions. In the sequel we will denote that set with $\Psi_e(A,B)$, which is formally defined in the next definition.

DEFINITION 5.3.4. Extension Procedure Let $A\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$ be a stable controllable pair. Let $P>0$ be the solution of the Lyapunov equation $APA^*+BB^*=P$ .Define $W:=P^{\frac{1}{2}}$ [2], $\tilde{A}=W^{-1}AW$ and $\tilde{B}=W^{-1}B$, which implies $\tilde{A}\tilde{A}^*+\tilde{B}\tilde{B}^*=P$. Create with $\{\tilde{A},\tilde{B}\}$ the matrices A_e,B_e as in theorem 5.3.3 and let M_{c_e} be defined by (5.3.8). We define $\psi_k\{A,B\}$ as the discrete time function that is induced by the $(k+1)^{th}$ row of M_{c_e} and denote the set of these functions by:

$$\Psi_e\{A,B\}:= \{\psi_k\{A,B\},k\geq 0\} \tag{5.3.9}$$

With slight abuse of notation we also use $\Psi_e\{A,B\}$ to denote M_{c_e} of (5.3.8). o

Note that we consider $\psi_k\{A,B\}$ as a function $\mathbb{N}^0\to\mathbb{F}$. These functions are all in ℓ_2 because of the stability of A. Moreover they are orthonormal because the extended matrices have a Gramian I_e. Again we make the remark that we never required that $\{A,B\}$ is part of an input balanced realization of a FDLTS system, but in that case one can identify the procedure with the extension of unitary matrices that result from a singular value decomposition of the Hankel matrix.

The set of functions $\Psi_e\{A,B\}$ can be interpreted as a sequence of Markov parameters $G_e(z)$ with realization $[A_e,B_e,A_e,B_e]$ as follows: Let $M_i:=A_e^{i-1}B_e$. $i\in\mathbb{N}^0$, be the Markov parameters of $G_e(z)$ then the sequence $[M_o|M_1|M_2\cdots]$ is exactly the sequence (5.3.8). There exists also a more compact way to describe the orthonormal functions as responses of linear systems, which will be of use in later sections of this chapter where these functions have to be actually calculated for the purposes of signal approximation and system identification. This description will be referred to as the *transfer function description* and has a close relationship with the transfer function description (4.2.9) of the Laguerre polynomials.

---

[2]This represents a freedom since there exist many matrices W with $W^2=P$. This freedom can be removed if we restrict W to be positive symmetric (Golub and Van Loan, 1989).

PROPOSITION 5.3.5. Transfer function description Let {A,B} and F be as in theorem 5.3.3 and let M_{c_e} be the extended controllability matrix (5.3.8). Define for $k \in \mathbb{N}^0$ the

transfer function: $$H_k(z) := \left[[zI-A]^{-1}F[I-zA^*]\right]^k z[zI-A]^{-1}B \qquad (5.3.10a)$$

Let M_i^k , $i \in \mathbb{N}^0$, be the Markov parameters of $H_k(z)$ and define the Markov parameter matrix

$$\mathscr{M}_k := [M_0^k | M_1^k | M_2^k \cdots] \qquad (5.3.10b)$$

Then the rows of $\mathscr{M}_k$ are rows of M_{c_e} with the following ordering

$$M_{c_e} = [\mathscr{M}_0^* | \mathscr{M}_1^* | \mathscr{M}_2^* \cdots]^* \qquad (5.3.10c)$$

o

PROOF: See appendix 5D.

This proposition shows that we do not have to create the large matrices $\{A_e,B_e\}$ in order to calculate the orthonormal functions. This can be done by calculation of the first N (block) Markov parameters of $H_0(z) = z[zI-A]^{-1}B$, which creates the first N×m samples of the first n orthonormal functions. If we use these as inputs for the transfer function $[zI-A]^{-1}F[I-zA]$, then this results in the first N×m samples of the second n orthonormal functions. Using these as inputs for the latter transfer functions results in the third n functions etc. This property will be useful when transforming a finite time series in terms of the elements of Ψ_e; this will be discussed in section 5.4.

REMARK 5.3.6. In this section we focused on the input side of FDLTS systems and created orthonormal functions based on a transformed controllability matrix. An analogous procedure can be carried out on the output side, starting with a realization [A,B,C,D] that obeys $A^*A+C^*C=I$. If the system is output balanced, we can again interpret the extension in terms of the matrix with left hand side singular vectors of the Hankel matrix.
Interpreted in this context, we defined in this section a method that, given a FDLTS system G(z) with Hankel matrix $\mathscr{H}=U\Sigma V^*$, extends V or U in a recursive way to an infinite unitary matrix V_e or U_e, by adding new columns to the original columns. o

REMARK 5.3.7. We emphasize that the orthonormal functions, that we presented in this section have a dynamical character, in the sense that they are equivalent with sequences of Markov parameters of a dynamical system. This property will be applicable for the approximation of signals and for system approximation, if this dynamical character is shaped according to the underlying dynamics of the signals and systems, that are approximated. o

COMPLETENESS

We have presented a method to create an infinite sequence of orthonormal functions based on a FDLTS system. It is our goal to use these functions to give alternative descriptions of linear systems and to apply them for the purpose of system identification as was done in chapter four with the Laguerre polynomials. A necessary condition for further development of such a theory will be that these functions form a basis for the function space we wish to consider, which in our case is ℓ_2. In other words, we must consider the question if, under appropriate conditions on a pair $\{A,B\}$, the set of functions $\Psi_e\{A,B\}$ forms a complete orthonormal basis for ℓ_2. The answer to this question is affirmative and the conditions on the matrices $\{A,B\}$ are the same as the conditions posed for the extension in theorem 5.3.3, with a full column rank condition on the B matrix.

For the proof of this completeness property we need two 2 lemmas. They will be presented here, instead of in an appendix, because they will be used again later on. The first lemma states that we can transform a pair $\{A,B\}$, which is already in input balanced form with a similarity matrix, that is unitary, without changing the characteristics of the resulting sets of orthonormal functions.

LEMMA 5.3.8. Let $A,T\in\mathbb{C}^{n\times n}$, $B\in\mathbb{C}^{n\times m}$, with A stable, $AA^*+BB^*=I$ and T unitary, $TT^*=T^*T=I$. Let $\Psi_e\{A,B\}$ be defined by definition 5.3.4. Define $\tilde{A}:=TAT^*$, $\tilde{B}:=TA$ and T_e as the infinite block diagonal matrix $T_e:=\begin{bmatrix} T & & \\ & T & \\ & & \ddots \end{bmatrix}$. Then the following two statements are valid:

1. $\tilde{A}\tilde{A}^* + \tilde{B}\tilde{B}^* = I$ (5.3.11a)
2. There exists an extension $\Psi_e\{\tilde{A},\tilde{B}\}$ such that
$\Psi_e\{A,B\} = T_e\Psi_e\{\tilde{A},\tilde{B}\}$ (5.3.11b)

where we identify Ψ_e with the controllability matrix M_{c_e} (5.3.8). ○

PROOF: See appendix 5E

It is apparent that the matrix T in lemma 5.3.8 defines a bijection between both sets of functions $\Psi_e\{\tilde{A},\tilde{B}\}$ and $\Psi_e\{A,B\}$. Therefore we can conclude on beforehand:

$$\Psi_e\{\tilde{A},\tilde{B}\} \text{ complete} \Leftrightarrow \Psi_e\{A,B\} \text{ complete.}$$

The second lemma we need shows that under a special similarity transformation the structure of $\{A,B\}$ becomes very simple and can be linked to the singular value decomposition of A. This is a pure technical result, used to proof the completeness.

LEMMA 5.3.9. Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$ with rank(B)=m,

1. $\exists$ unitary matrix T such that TAT^* has a svd $TAT^*=I_nSV^*$.
2. If $AA^*+BB^*=I$ then A has (n–m) singular values equal to one and m singular values smaller than one.
3. If $AA^*+BB^*=I$ and A has a svd $A=I_nSV^*$, where S is in decreasing order, so $S(i,i)\geq S(i+1,i+1)$, then
   a. $S=\begin{bmatrix} I_{n-m} & 0 \\ 0 & S_a \end{bmatrix}$, $S_a<I_m$ an mxm diagonal matrix
   b. $B=\begin{bmatrix} 0 \\ B_1 \end{bmatrix}$ where $B_1 \in \mathbb{C}^{m\times m}$ has a svd $B_1=I_mS_bV_b^*$ with $S_b^2=I_m-S_a^2$. ○

PROOF: See appendix 5F.

Note that in this lemma we introduced a rank condition on B, which is used to show that $S_a<I$. This will be used in the proof of the next theorem. These two lemmas enable us to prove a main result of this section, which is the completeness of the infinite set of orthonormal functions, generated by an arbitrary stable controllable pair {A,B}.

THEOREM 5.3.10. Let $A \in \mathbb{C}^{n\times n}$, $B \in \mathbb{C}^{n\times m}$, with A stable, rank(B)=m and {A,B} a controllable pair. Let $\Psi_e\{A,B\}$ be defined as in definition 5.3.4. Then this set of functions forms a complete orthonormal basis for ℓ_2. ○

PROOF: See appendix 5G.

REMARK 5.3.11. The only restriction that is made in theorem 5.3.10 is the rank condition on the matrix B. The reason for this can be made plausible as follows:

The proof of theorem 5.3.10 is based on the fact that $\Psi_e\Psi_e^*=\Psi_e^*\Psi_e=I_e$. A simple example shows why the property that $\Psi_e\Psi_e^*=I_e$ is not sufficient for completeness and why we need $\Psi_e^*\Psi_e=I_e$. Therefore we consider the matrix Γ:

$$\Gamma=\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix}, \text{ then } \Gamma^*\Gamma=\begin{bmatrix} 0 & & & \\ & 1 & & \\ & & 1 & \\ & & & \ddots \end{bmatrix} \text{ and } \Gamma\Gamma^* = I_e.$$

Consider the rows of Γ as discrete time functions. It is clear that this constitutes an orthonormal set of ℓ_2–functions, but we do not have a basis for ℓ_2 since the function $(1\ 0\ 0\ 0\cdots)$ cannot be written as a converging sum of the other functions. This can be translated to the fact that $\Gamma^*\Gamma\neq I_e$. The rank condition on B is necessary to omit situations like the one above. Assume for instance that B is of the form $B=[B_1\ 0]$. Any similarity transformation will leave this form unaltered and since B_e is a column matrix with only entries of the form P^kB, the whole matrix B_e and all products $A_e^kB_e$ will be of this form, causing Ψ_e to have zero columns. In that case it is clear that we have no completeness. ○

Theorem 5.3.10 shows that the sets of orthonormal functions, based on FDLTS systems, which were introduced in this section, form orthonormal bases for ℓ_2. Hence any ℓ_2 time series can be written as a converging sum of these functions. This will in section 5.4 be applied to ℓ_2 input/output pairs {u,y} of a linear system. We will show how we can use these results in order to define alternative descriptions of linear systems. First we present two applications of the extension procedure, showing how well known sets of orthonormal functions can be derived with the developed theory, by choosing a specific FDLTS system as 'generator' of these functions.

COROLLARY 5.3.12. Laguerre polynomials. Let $\xi\in\mathbb{R}$, with $|\xi|<1$ and let $\eta:=1-\xi^2$. Let $\{\psi_k^\xi\}$ be the set of Laguerre polynomials, as defined by definition 4.2.2. Then the set of orthogonal functions $\Psi_e\{\xi,\eta^{\frac{1}{2}}\} = \{\psi_k^\xi\}$ and the generating transfer functions as defined by propositions 4.2.4 and 5.3.5 are equal. Hence every stable first order system G(z) with a pole in ξ can be considered as the system that generates the Laguerre polynomials. ○

PROOF: Let G(z) be a first order stable SISO–system and let [ξ,b,c,d] be a realization of G(z), then $[\xi,\eta^{\frac{1}{2}},\eta^{-\frac{1}{2}}bc,d]$ is an input balanced realization of G(z). So with $A=\xi$ and $B=\sqrt{\eta}$ we have $AA^*+BB^*=1$. Now we follow the procedure outlined in theorem 5.3.3; a singular value decomposition of A is $A=U_aS_aV_a^*$ with $U_a=V_a=1$ and $S_a=\xi$. Substitution of these values in (5.3.7) yields:

$$\begin{matrix} F=1 \\ X=\eta, \\ P=-\xi \end{matrix} \qquad A_e = \begin{bmatrix} \xi & 0 & & & \\ \eta & \xi & 0 & & \\ -\xi\eta & \eta & \xi & 0 & \\ \xi^2\eta & -\xi\eta & \eta & \xi & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \qquad B_e = \begin{bmatrix} \sqrt{\eta} \\ -\xi\sqrt{\eta} \\ \xi^2\sqrt{\eta} \\ \vdots \end{bmatrix} \qquad (5.3.12a)$$

and these are exactly the matrices we found in proposition 4.4.4 as the matrices that constitute the finite difference Laguerre polynomials. If we look at the generating transfer functions defined in proposition 5.3.5, $H_k(z)=\left[[zI-A]^{-1}F[I-zA^*]\right]^k z[zI-A]^{-1}B$ and substitute $A=\xi$, $B=\sqrt{\eta}$ and $F=1$ then we get the Laguerre transfer functions

$$H_k(z) = \sqrt{\eta}\, z[1-z\xi]^k[z-\xi]^{-k-1} \qquad (5.3.12b)$$

that were presented in proposition 4.2.4. □

This corollary shows that with the extension procedure we generalized the construction of the finite difference Laguerre polynomials. Considering the fact that this application shows that Laguerre polynomials are constructed from a system with one pole, we might consider the general orthonormal functions as a multivariable extension of these polynomials, in the sense that we have more degrees of freedom in the definition of the orthogonal functions than the one degree of freedom (the discount factor) of the Laguerre polynomials.

COROLLARY 5.3.13. Pulse functions. Let $A,B \in \mathbb{R}^{n\times n}$, with $A=0$ and $B=I$. Let $\{p_i, i\in\mathbb{N}^0\}$ be the set of discrete pulse functions $\mathbb{N}^0 \to \mathbb{R}$, defined by $p_i(t) := \delta_{it}$. Then the set of orthogonal functions $\Psi_e\{A,B\} = \{p_i, i\in\mathbb{N}^0\}$ and the generating transfer functions $H_k(z)$, corresponding with $\Psi_e\{A,B\}$ as defined by proposition 5.3.5 are equal to $z^{-k}I$. Furthermore, every stable system $G(z)$ with a finite number of non–zero Markov parameters can be considered as a system that generates the discrete pulse functions. o

PROOF: A singular value decomposition of A is $A=U_aS_aV_a^*$ with $U_a=V_a=I$ and $S_a=0$. Substitution in (5.3.7) results in

$$\begin{array}{l} F:= I \\ P:= 0 \\ X:= I \end{array} \quad A_e = \begin{bmatrix} 0 & 0 & 0 & \cdots \\ I & 0 & 0 & \cdots \\ 0 & I & 0 & \cdots \\ & \ddots & \ddots & \ddots \end{bmatrix} \quad B_e = \begin{bmatrix} I \\ 0 \\ 0 \\ \vdots \end{bmatrix} \quad \Psi_e = \begin{bmatrix} I & 0 & 0 & \cdots \\ 0 & I & 0 & \cdots \\ 0 & 0 & I & \cdots \\ & \ddots & \ddots & \ddots \end{bmatrix} \qquad (5.3.13)$$

So the extended set of functions are the pulse functions $p_i(t)=\delta_{it}$, which is the usual basis for ℓ_2. The generating transfer functions follow from substitution in proposition 5.3.5. If a FDLTS system $G(z)$ has only a finite number of Markov parameters $\{M_i, i=0\cdots k\}$, then we can always find a (not necessarily minimal) realization $[A_1,B_1,C_1,D_1]$ of $G(z)$ with A_1 and B_1 finite versions of A_e respectively B_e, $C_1=[M_1|M_2|\cdots|M_k]$ and $D_1=M_0$. The matrices A_1 and B_1 obey the conditions of theorem 5.3.3. An svd of A_1 is $A=U_1S_1V_1$ with $V_1=I$ and

$$U_1 = \begin{bmatrix} 0 & 0 & 0 & I \\ I & 0 & 0 & \cdot \\ 0 & I & 0 & \cdot \\ & 0 & I & 0 \end{bmatrix} \quad S_1 = \begin{bmatrix} I & 0 & 0 & \cdot \\ 0 & I & 0 & \cdot \\ 0 & 0 & I & \cdot \\ & \ddots & \ddots & 0 \end{bmatrix} \Rightarrow F_1=U_1 \quad P_1 = -\begin{bmatrix} 0 & 0 & 0 & \cdot \\ 0 & I & 0 & \cdot \\ 0 & 0 & I & \cdot \\ & \ddots & \ddots & I \end{bmatrix} \quad X_1 = \begin{bmatrix} 0 & 0 & \cdot & I \\ 0 & 0 & \cdot & 0 \\ 0 & 0 & \cdot & \cdot \\ & \ddots & \ddots & 0 \end{bmatrix}$$

Substitution of these variables in (5.3.7) yields $A_{1_e}=A_e$, $B_{1_e}=B_e$. This shows that

$$\Psi_e\{A_1,B_1\} = \Psi_e\{A,B\}. \qquad \square$$

REMARK 5.3.14. The applications, that were considered in the corollaries 5.3.12 and 5.3.13, show that the extension procedure 5.3.4 is quite natural and leads to a generalization of well known orthonormal bases for ℓ_2. o

5.4. TRANSFORMATIONS

5.4.1 INTRODUCTION

In section 5.4 we will use the set of orthonormal functions of the previous sections as a basis of ℓ_2 and we will define the transformation of ℓ_2 time series in these terms (section 5.4.2). When this transformation is applied to the ℓ_2 input/output variables of a FDLTS system, this leads to another system description in terms of the coefficients of the expansions (section 5.4.3). These new descriptions have several interesting features,

depending on the choice of the orthonormal basis. Another application of the transformation of time–series, the transformation of impulse responses of a FDLTS system, results in an alternative frequency domain description (section 5.4.4). This section is the main body of chapter five and many results will be presented and various problems will emerge from the theory. For this reason we present an overview in section 5.4.5. To omit tenuous notation we will in the remainder only consider the case where $\mathbb{F}=\mathbb{R}$, i.e. we only consider systems and signals with real–valued entries and parameters. This implies that for a matrix or function A we have $A^*=A^T$.

We assume for the time being that an orthonormal basis Ψ_e is given, so let $\{A,B\}$ be a stable, controllable pair, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times m}$, rank(B)=m≤n and let $\Psi_e\{A,B\}$ be defined by definition 5.3.4.

5.4.2. TRANSFORMATION OF TIME SERIES

The set of functions Ψ_e is complete in ℓ_2 so we can expand in a unique way any $\ell_{2[0,\infty)}^p$ time series f: $\mathbb{N}^0\to\mathbb{R}^p$ in these functions, as follows:

$$\forall t\in\mathbb{N}^0 \qquad f(t) = \sum_{k=0}^{\infty} F_k\psi_k(t) \tag{5.4.1a}$$

where the $F_k\in\mathbb{R}^p$, are the *expansion coefficients* of f, given by

$$\forall k\in\mathbb{N}^0 \qquad F_k = \sum_{t=0}^{\infty} f(t)\psi_k(t) \tag{5.4.1b}$$

Since $f,\psi_k\in\ell_2$ we are assured that F_k is well defined. As in the case of the Laguerre polynomials we may compare this procedure with the Fourier transformation and we can call $\{\ \{f(t)\},\ \{F_k\}\ \}$ the transform pair. The orthonormality of the basis functions shows that the Parseval identity is valid for this transform pair:

$$\langle f,f\rangle_{\ell_2} = \langle F,F\rangle_{\ell_2} \tag{5.4.2a}$$

or

$$\sum_{t=0}^{\infty} f(t)^*f(t) = \sum_{k=0}^{\infty} F_k^*F_k \tag{5.4.2b}$$

In order to make full use of the shift structure of $\Psi_e\{A,B\}$ we will not proceed with the transformation in this form, but instead we first group the orthonormal functions in groups of n functions each.

DEFINITION 5.4.1. Let $\{A,B\}$ be a stable, controllable pair, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times m}$, rank(B)=m≤n and let $\Psi_e\{A,B\}=\{\psi_k,\ k\in\mathbb{N}^0\}$ be defined by definition 5.3.4. We define the *grouped orthonormal functions* $\varphi_k\{A,B\}$: $\mathbb{N}^0\to\mathbb{R}^n$ by

$$\varphi_k\{A,B\}(t) := [\psi^*_{kn}(t), \psi^*_{kn+1}(t), \cdots \psi^*_{kn+n-1}(t)]^* \quad k\in\mathbb{N}^0 \tag{5.4.3a}$$

and denote the set of these functions by

$$\Phi_e\{A,B\} := \{\ \varphi_k\{A,B\},\ k\in\mathbb{N}^0\} \tag{5.4.3b}$$

○

Hence $\varphi_0(t)$ is the column vector with the first n orthonormal functions, $\varphi_1(t)$ the vector with the second n and so further. This definition fits in nicely with the generating transfer functions $H_k(z)$ (5.3.10a), since $[\varphi_k(0)|\ \varphi_k(1)|\ \cdots]$ coincides with the Markov parameter matrix $\mathcal{M}_k$ (5.3.10b) of $H_k(z)$. Whenever it is permitted by the context we will use the notation $\varphi_k(t)$ instead of $\varphi_k\{A,B\}(t)$. We redefine the orthogonal coefficients in terms of the grouped orthonormal functions.

Definition 5.4.2. Let $\{A,B\}$ be a stable, controllable pair, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times m}$, $\text{rank}(B)=m\leq n$ and let $\Phi_e\{A,B\} = \{\varphi_k, k\in\mathbb{N}^0\}$ be defined by definition 5.4.1. Let f be a ℓ_2^p time-series. We define the *k^{th} grouped orthogonal coefficient* of f by $L_k\in\mathbb{R}^{p\times n}$,

$$L_k = \sum_{t=0}^{\infty} f(t)\varphi^*_k(t) \tag{5.4.4a}$$

and we denote

$$\{L_k, k\in\mathbb{N}^0\}_\Phi = \Phi_e(f) \tag{5.4.4b}$$

○

It will be clear that $L_k = [F_{kn}|F_{kn+1}|\cdots|F_{kn+n-1}]$ with F_k defined by (5.4.1b) and that the Parseval identity is again valid. The reversed process, the backward transformation is covered by the next proposition.

Proposition 5.4.3 Let $\{A,B\}$,f(t), $\varphi_k(t)$ and L_k be as in definition 5.4.2. Then $\{f(t)\}$ and $\{L_k\}$ are subject to the following equation:

$$f(t) = \sum_{k=0}^{\infty} L_k\varphi_k(t) \tag{5.4.5}$$

○

Proof: In the proof of theorem 5.3.3 we established that $\Psi_e^*\Psi_e=I$. It follows that $\Phi_e^*\Phi_e=I$ and thus $\sum_{k=0}^{\infty} \varphi^*_k(s)\varphi_k(t)= \delta_{st}$. Now evaluate the right hand side of (5.4.5):

$$\sum_{k=0}^{\infty} L_k\varphi_k(t) = \sum_{k=0}^{\infty}\left[\sum_{s=0}^{\infty} f(s)\varphi^*_k(s)\right]\varphi_k(t) = \sum_{s=0}^{\infty} f(s)\left[\sum_{k=0}^{\infty}\varphi^*_k(s)\varphi_k(t)\right] = \sum_{s=0}^{\infty} f(s)\delta_{st} = f(t) \qquad \square$$

As in chapter four we will use the orthogonal functions in a context of system identification and therefore we will perform calculations with the coefficients (5.4.4) of signals. We can also use the orthonormal basis to approximate given signals with a small number of functions. As explained in section 4.4 it will in general not be possible to calculate the orthonormal coefficients, if the time series is of infinite length. However in practical situations, as is the case with identification problems which are considered in

section 5.5, the time series under consideration will be of finite length. In the identification setting we will actually have to calculate the coefficients $\{L_k\}$ (5.4.4b). In section 4.4 we considered the underlying approximation character in detail and the same analysis is valid for the generalized orthonormal functions (see the table in remark 4.4.3). We refrain from repeating those arguments. If the number of data points of the signal, that must be transformed or approximated, is large, then the numerically most efficient approximation of the orthonormal coefficients is based on the assumption that the signal has a zero extension. The following proposition explains how these calculations can be performed.

PROPOSITION 5.4.4 Let $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times m}$, $f\in\ell_2^p$, $\{\varphi_k\}$ and $\{L_k\}$ be as in definition 5.4.2 with p=1. Let $\{H_k(z)\}$ be the set of generating transfer functions, that corresponds with $\{\varphi_k\}$ and let $N\in\mathbb{N}$ such that $f(t)\equiv 0$, $t\geq Nm$. Define the *m−reversed time series of f*,

$$r_f\colon \mathbb{Z}\to\mathbb{F}^m, \text{ by} \qquad r_f(t):=\begin{cases}\begin{bmatrix} f\big[(N-t-1)m\big] \\ \vdots \\ f\big[(N-t)m-1\big]\end{bmatrix} & t\in[0,N-1] \\ 0 & t\notin[0,N-1]\end{cases} \tag{5.4.6a}$$

Define the time series $y_k\colon \mathbb{Z}\to\mathbb{F}^n$ by $y_k:=H_k(z)r_f$, $y_k(t)=0 \quad t<0$ (5.4.6b)

Then $L_k=y_k^*(N-1)$ ○

PROOF: Let $\{A_e,B_e\}$ be defined by (5.3.7) and (5.3.8). It is immediate that:

$$\begin{bmatrix}L_0^* \\ L_1^* \\ \vdots\end{bmatrix} = [B_e|A_eB_e|A_e^2B_e|\cdots]\times\begin{bmatrix}f^*(0) \\ f^*(1) \\ \vdots\end{bmatrix}$$

$$= B_e\begin{bmatrix}f(0) \\ \vdots \\ f(m-1)\end{bmatrix} + A_eB_e\begin{bmatrix}f(m) \\ \vdots \\ f(2m-1)\end{bmatrix} + \cdots + A_e^{N-1}B_e\begin{bmatrix}f\big[(N-1)m\big] \\ \vdots \\ f(Nm-1)\end{bmatrix}$$

$$= B_er_f(N-1) + A_eB_er_f(N-2) + \cdots + A_e^{N-1}B_er_f(0)$$

Now apply proposition 5.3.5 to show that also:

$$\begin{bmatrix}y_0(N-1) \\ y_1(N-1) \\ \vdots\end{bmatrix} = B_er_f(N-1) + A_eB_er_f(N-2) + \cdots + A_e^{N-1}B_er_f(0)$$

This shows that $L_k^*=y_k(N-1)$, and completes the proof. □

REMARK 5.4.5. Because of the iterative structure of the H_k's the calculations of proposition 5.4.4 can be done using a simple cascade like network. For time series that are of dimension higher than one, $f\in\ell_2^p$ with p>1, we can apply proposition 5.4.4 on each of the scalar time series f_i. This proposition is analogous to the procedure we presented for the transformation of time series, that have a zero extension, with Laguerre polynomials, as depicted in Fig. 4.4.1.

It should be emphasized that for approximation of a signal with a finite number of orthonormal functions, the procedure of proposition 5.4.4 is not optimal in least squares sense and that the difference between a zero–extension and a least squares approximation can be considerable if the number of samples is relatively small compared to the decay rate of the orthonormal functions (see remark 4.4.3 and example 4.4.8). o

EXAMPLE 5.4.6. We will show how knowledge of the character of a signal can be employed to define orthonormal functions for the approximation of that signal. Furthermore we illustrate the difference between the approximation based on the assumption of a zero–extension and a least squares approximation. For this purpose we consider the same signal that was used in the same context with Laguerre polynomials in example 4.4.8 (Fig. 4.4.3 and 4.4.4). This signal is a measured impulse response. With this 'knowledge' we used an approximate realization method (Van Helmont and co–workers, 1990) to create a 2nd order approximation of the underlying system. We intentionally choose this order, that is too low to describe the impulse response sufficiently well, because we wish to show how this little knowledge can be used to get satisfactory results. We created the orthonormal functions generated by the 2nd order approximation and approximated the original signal with ten functions in least squares sense. The result is presented in Fig. 5.4.1a, where the original signal, the impulse response of the low order model and the least squares approximation are depicted. We see that the signal is well approximated with ten functions. If a better low order system was created then the number of polynomials would be smaller. In Fig. 5.4.1b the least squares approximation is compared with the approximation based on a zero extension. It shows that there is a clear difference. This difference will get smaller if more functions are used or if more samples of the impulse response are approximated. o

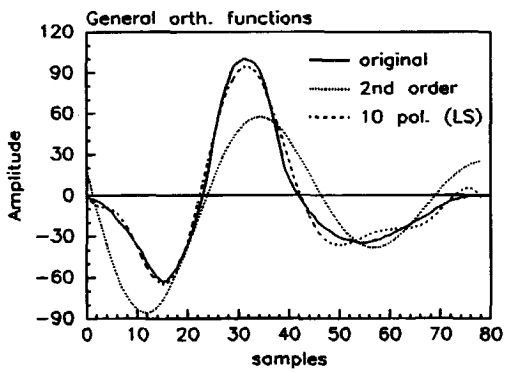


Fig. 5.4.1a Approximations with response of second order model and 8 system based orthonormal functions.

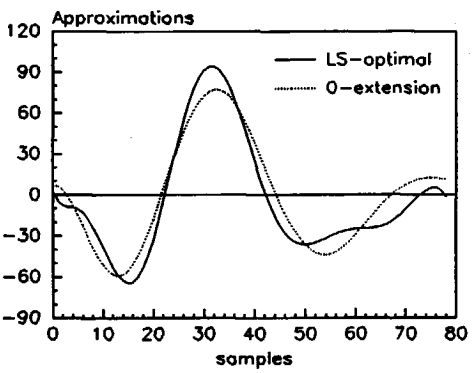


Fig. 5.4.1b. Approximations with 8 orthonormal functions: least squares and zero–extension.

5.4.3. TRANSFORMATION OF SYSTEMS

We will now proceed with the transformation of linear systems, as an extension of the previous chapter. This is defined by considering the input output variables of a FDLTS system, transformation of these signals to an orthonormal domain and considering the system that interrelates these orthonormal coefficients. The calculations, that are required, are more tedious then in the Laguerre case and we will use Kronecker matrix calculation, which is reviewed in section 2.9.

Suppose we have at hand an arbitrary FDLTS system G(z) with McMillan degree n and dimensions pxm. Let $(u,y)\in\mathcal{B}_2^{\perp}(G)$, then we can transform these time–series with any orthonormal set $\Psi_e\{A,B\}$ to an orthonormal domain. We do not assume any relation between {A,B} and G(z), so {A,B} might be matrices originating from a totally different system. However we do need to impose the restriction that B has only one column, so $B\in\mathbb{R}^{n\times 1}$. The reason for that restriction will become clear later on. Let $\{U_k\}_\Phi$ and $\{Y_k\}_\Phi$ denote the orthonormal coefficients (5.4.4) of the time series u and y, $U_k\in\mathbb{R}^{m\times n}$,$Y_k\in\mathbb{R}^{p\times n}$. First we define the transformation of a restricted ℓ_2–behavior, as defined by definition 2.4.5.

DEFINITION 5.4.7. Let {A,B} be a stable controllable pair, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times 1}$ with rank(B)=1 and let G(z) be a FDLTS system. Let $\mathcal{B}_2^{\perp}(G)$ and $\Phi_e\{A,B\}$ be defined by definition 2.4.5 and definition 5.4.1. We define the transform $\Phi(\mathcal{B}_2^{\perp}(G))$ of $\mathcal{B}_2^{\perp}(G)$ by:

$$\Phi\{\mathcal{B}_2^{\perp}(G)\}:=\Big\{\big[\mathrm{Vec}(Y_k),\mathrm{Vec}(U_k)\big]\in(\mathbb{R}^{np+nm})^{\mathbb{Z}_+} \mid \exists\ (y,u)\in\mathcal{B}_2^{\perp}(G) \text{ with } \{Y_k\}_\Phi=\Phi_e(y) \text{ and } \{U_k\}_\Phi=\Phi_e(u)\Big\} \qquad (5.4.7a)$$

o

Note that the completeness of Ψ_e implies that this defines a bijective transformation between $\mathcal{B}_2^{\perp}(G)$ and $\Phi(\mathcal{B}_2^{\perp}(G))$. Our aim is to show that $\Phi\{\mathcal{B}_2^{\perp}(G))$ is in it self a restricted ℓ_2–behavior of a system $G_o(z)$ in the orthonormal domain, so we shall show that

$$\exists\ G_o(z) \text{ such that } \mathcal{B}_2^{\perp}(G_o) = \Phi(\mathcal{B}_2^{\perp}(G)) \qquad (5.4.7b)$$

Moreover we want to establish that the system $G_o(z)$ is stable and has the same McMillan degree as G. In order to attain this result, which in the Laguerre case was relatively straightforward (see theorem 4.5.1), we have to deal with complicated calculations and to go through lengthy proofs before we have proven the desired result completely. Therefore we will follow a step by step procedure which leads us to the desired result. In these steps we show successively:

1. The existence of a FDLT system $G_o(z)$ that obeys (5.4.7b) with a realization of $G_o(z)$ (theorem 5.4.8);
2. The non–minimality of this realization (lemma 5.4.9);
3. The controllability and observability of the stable poles (lemma 5.4.10);
4. The non–controllability and non–observability of the unstable poles (lemma 5.4.11)

We begin by showing that the transformed signals of an input/output pair are indeed connected through a FDLT system in the orthonormal domain and present a (non–minimal, non–stable) realization of the latter.

THEOREM 5.4.8. Let $\{A,B\}$ be a stable, controllable pair, $A\in\mathbb{R}^{n\times n}$, $B\in\mathbb{R}^{n\times 1}$ and let $\Phi_e\{A,B\}$ be defined by definition 5.4.1. Further let $W_t(z)$ be a p×m FDLTS system with McMillan degree n_w and let $\mathcal{B}_2^\sharp(W_t)$ be the restricted ℓ_2–behavior of $W_t(z)$, as defined by definition 2.4.5. Then:

a. $\exists$ (pn×mn) FDLT system $W_o(z)$ such that:
$$\Big[(y,u)\in\mathcal{B}_2^\sharp(W_t),\ \{Y_k\}_\Phi:=\Phi(y),\ \{U_k\}_\Phi:=\Phi(u)\Big] \Leftrightarrow (\mathrm{Vec}(Y_k),\mathrm{Vec}(U_k))\in\mathcal{B}_2^\sharp(W_o)$$

b. If [R,G,H,L] is a minimal realization of $W_t(z)$ with initial state x(0) then $W_o(z)$ has a realization $[R_o,G_o,H_o,L_o]$ with initial state Z_o where $R_o\in\mathbb{R}^{n_w n\times n_w n}$, $G_o\in\mathbb{R}^{n_w n\times nm}$, $H_o\in\mathbb{R}^{np\times n n_w}$, $L_o\in\mathbb{R}^{np\times nm}$ are given by:

$$\begin{aligned} R_o &:= NM^{-1} && \text{where} && M &:= I\otimes I - A\otimes R \\ G_o &:= W - NM^{-1}V && && N &:= -FA^*\otimes I + F\otimes R \\ H_o &:= (I\otimes H)M^{-1} && && V &:= -A\otimes G \\ L_o &:= I\otimes L - (I\otimes H)M^{-1}V && && W &:= F\otimes G \\ Z_o &:= B\otimes x(0) \end{aligned} \qquad (5.4.8)$$

o

PROOF: See appendix 5H.

The next step in the proof is to consider the realization $[R_o,G_o,H_o,L_o]$ of (5.4.8) and to recognize that this realization of the transformed system $W_o(z)$ has completely uncontrollable poles in z=–1.

LEMMA 5.4.9. There exists a similarity transformation matrix T such that

$$TR_oT^{-1}=\begin{bmatrix}-I_x & 0\\ * & *\end{bmatrix},\quad TG_o=\begin{bmatrix}0\\ *\end{bmatrix}\quad\text{and}\quad TZ_o=\begin{bmatrix}0\\ *\end{bmatrix} \qquad (5.4.9)$$

where $x=(n-1)n_w$ and where T=I if A has a singular value decomposition $A=ISV^*$. Therefore the realization $[R_o,G_o,H_o,L_o]$ of the system $W_o(z)$, as defined by (5.5.8), has $(n-1)n_w$ uncontrollable poles in z=–1. o

PROOF: See appendix 5I.

We can conclude that the McMillan degree of the system $W_o(z)$ is smaller than or equal to $nn_w-(n-1)n_w=n_w$. We will show that it is equal to n_w, the McMillan degree of the original system. For this purpose we will show that every stable pole of W_o is controllable and observable after which we shall establish that W_o has indeed n_w stable poles.

LEMMA 5.4.10. Let λ and μ be stable eigenvalues of R_o, then no vectors $z,y \neq 0$ exist such that

$$z^*R_o=\lambda z^* \qquad R_oy=\mu y$$
$$z^*G_o=0 \qquad H_oy=0$$

and hence the Popov–Hautus–Belevitch test shows that every stable eigenvalue of R_o is a controllable and observable pole of $W_o(z)$. o

PROOF: See appendix 5J

It is obvious that we will have to proof that R_o has n_w stable poles. This is established by the following lemma, which also shows how we can actually calculate the eigenvalues of the system matrix R_o. We recall that for the definition of the orthonormal functions $\Psi_e\{A,B\}$ we used in (5.3.7) an intermediate matrix F, that is created from the svd of A:

$$A=USV^* \Rightarrow F=UV^* \tag{5.4.10}$$

LEMMA 5.4.11. Let A, R and R_o be as in theorem 5.4.8 and let F be the unitary matrix that is created from the svd of A as in (5.4.10). Let R have eigenvalues $\{\mu_1,\ldots,\mu_{n_w}\}$.

Then: 1. R_o has n_w stable eigenvalues and $(n-1)n_w$ eigenvalues in $z=-1$;

2. The eigenvalues of R_o are the eigenvalues of the matrices:

$$F(\mu_i-A^*)(I-\mu_iA)^{-1} \quad i=1,\cdots,n_w \tag{5.4.11}$$

3. For every μ_i that is an eigenvalue of A the matrix R_o has a zero eigenvalue. o

PROOF: See appendix 5K.

At this point we have reached our first goal, which was that we wanted to show the validity of equation (5.5.7b) with the additional property that the transformed system is stable and has the same McMillan degree as the original system.

However there is even more to say about the system $W_o(z)$. Lemma 5.4.11 shows that W_o will have poles in zero if eigenvalues of A coincide with poles of W_t and it is intriguing to explore what happens if there is a connection between the original system W_t and the set of orthonormal polynomials $\Psi_e\{A,B\}$. We know for instance from lemma 5.4.11 that if the set of poles of W_t is a subset of the set of eigenvalues of A then W_o will only have poles in zero, but we have no information about the coupling of these

poles[3]. If W_o has only poles in zero then we can conclude that the system has a finite number of Markov parameters. This number depends on the coupling of these poles. The following proposition gives conditions under which poles in zero are guaranteed to be uncoupled.

Proposition 5.4.12. Let the Jordan matrices J_a and J_r of respectively A and R have q diagonal blocks of dimension $d_i \times d_i$ $(i=1 \cdots q)$ in common. Then the matrix R_o has $n_q := \sum_{i=1}^{q} d_i$ uncoupled eigenvalues equal to zero. o

Proof: See appendix 5L.

Remark 5.4.13. This proposition 5.4.12 is interesting, because it shows that the dynamics is more or less removed from the system. From lemma 5.4.11(3) it follows that if A and R are of equal dimension and have the same eigenvalues then $W_o(z)$ has only poles in z=0. Proposition 5.4.12 implies that in the stronger case that A and R are even similar, i.e they have the same Jordan blocks, then all these poles are uncoupled. This is equivalent with the statement that the system matrix of a minimal realization of $W_o(z)$ is equal to zero and hence the system $W_o(z)$ is described by only two Markov parameters (L_o and H_oG_o). We might say that all dynamic behavior is absorbed by the transformation. o

Corollary 5.4.14. If A and R are similar, then the transformed system $W_o(z)$ only has uncoupled poles in zero and only the first two Markov parameters of $W_o(z)$ are unequal to zero:

$$W_o(z) = L_o + z^{-1}H_oG_o \qquad (5.4.12)$$

o

The next result on the transformation of systems concerns the Hankel singular values of the transformed system. It turns out that these are exactly the same as those of the original system, which is not really surprising because of the orthonormality of the transformation.

Proposition 5.4.15.

a. $W_o(z)$ has the same Hankel singular values as $W_t(z)$.

b. Let the realization [R,G,H,L] of $W_t(z)$ have Gramians P and Q, let A have a svd $A=I_nSV^*$ and let $T:=[0\ I_{n_w}]$. Then $[TR_oT^*,TG_o,H_oT^*,L_o]$ is a minimal realization of $W_o(z)$, with Gramians $(1-\sigma^2)P$ and $(1-\sigma^2)^{-1}Q$, where σ is the only diagonal element of S that is unequal to one (See lemma 5.4.2). o

Proof: See appendix 5M.

---

[3]Coupling of poles is here considered in terms of Jordan blocks of a Jordan decomposition of a square matrix (Stewart, 1973). Two poles are coupled iff they appear in the same Jordan block.

Finally in the next proposition we show that $W_t(z)$ and $W_o(z)$ have the same H_∞–norm. The implication of this result will be commented on in remark 5.4.18.

Proposition 5.4.16. The systems $W_t(z)$ and $W_o(z)$ have the same H_∞ norm. ○

Proof: This is a direct result of the Parseval identity. If $\{U_k\}$ is the transformed signal from an ℓ_2 signal $\{u(t)\}$ then both signals have the same amount of energy. Now the H_∞ norm of a system can be interpreted as the maximal energy at the output which is the result of an ℓ_2 input with norm 1. Hence both systems must have the same H_∞–norm. □

In the following corollary we combine the results, obtained on the transformation of FDLTS systems with system based orthonormal functions.

Corollary 5.4.17. Let $\{A,B\}$ be a stable, controllable pair, $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times 1}$, and let $\Phi_e\{A,B\}$ be defined by definition 5.4.1. Let $W_t(z)$ be a $(p\times m)$ FDLTS system with McMillan degree n_w and let $\mathcal{B}_2^\sharp(W_t)$ and $\Phi(\mathcal{B}_2^\sharp(W_t))$ be defined according to definitions 2.4.5 and 5.4.7. Then there exists a $(pn\times mn)$ FDLTS system $W_o(z)$ with McMillan degree n_w, such that

a. $\mathcal{B}_2^\sharp(W_o) = \Phi(\mathcal{B}_2^\sharp(W_t))$;
b. $W_o(z)$ has a non–minimal, non stable realization (5.4.8);
c. $W_t(z)$ and $W_o(z)$ have the same Hankel singular values and the same H_∞ norm;
d. For every eigenvalue of A, that is a pole of $W_t(z)$, the system $W_o(z)$ has a zero pole and if A is the system matrix of a realization of $W_t(z)$ then the system $W_o(z)$ has only uncoupled poles in zero and only the first two Markov parameters of $W_o(z)$ are unequal to zero. ○

In Fig. 5.4.2 we visualize this system transformation, that is induced by the transformation of ℓ_2 time series:

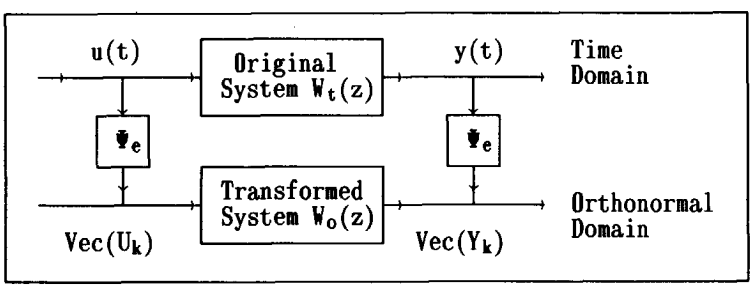


Fig. 5.4.2. System transformation

Remark 5.4.18. It is important to note that the system $W_o(z)$ must belong to a restricted class of FDLTS systems, since the set of (pn×mn) FDLTS systems with McMillan degree n_w is larger than the set of (p×m) FDLTS systems of order n_w. This implies that not every system in the former set can be considered as the transform of a system in the latter set. These restrictions are difficult to find. This is a consequence of the fact that we are not able to find expressions for a minimal realization of $W_o(z)$ other than in the form of matrix multiplications and Kronecker products. If we apply lemma 5.4.9 and assume that A (the matrix that is used to generate the orthonormal functions) has an svd $A=ISV^*$, then a minimal realization of $W_o(z)$ is given by $[R_1,G_1,H_1,L_0]$ where with $T=[0\ I_{n_w}]$:

$$R_1=TR_oT^* \qquad G_1=TG_o \qquad H_1=H_oT^*$$

but we have not yet been able to reduce these expression any further. As a result we are not able to describe the restrictions on $W_o(z)$.

Another consequence of this observation is that it is also difficult to solve the

> *inverse transformation problem:*
> given a set of orthonormal functions Ψ_e and a system $H_o(z)$ with the knowledge that $H_o(z)$ is the transform under Ψ_e of a system $H(z)$, how to calculate $H(z)$.

Lemma 5.4.11 shows for instance how the poles of the original system are related to the poles of the transformed system, but this information is not enough to follow the inverse direction. Even stronger, many systems with different poles exist, whose transformed systems have the same poles [4]. It will be necessary to use information on the zeros of the system as well, which is not available to us at this stage. However, as a result of the following procedure, an analytical expression for the inverse transformation can be given, that also provides a pragmatic way to tackle this problem. o

Procedure 5.4.19. Inverse system transformation.

Let $\Psi_e\{A,B\}$ be a set of orthonormal functions, as defined by definition 5.3.4. Let $W(z)$ be a (p×m) FDLTS system and let $W_o(z)$ be the transform of $W(z)$ under Ψ_e as defined by theorem 5.4.8. The Markov parameters $\{M_t\}$ of $W(z)$ can be calculated with the following procedure:

1. Define the m pulse signals u_i: $\mathbb{N}^o \to \mathbb{F}^m$ $(i=1\cdots m)$ as the pulse on the i^{th} input entry on time 0.
2. Calculate the transform U_i of u_i under Ψ_e with proposition 5.4.4.

---

[4] This can be verified with the formula from lemma 5.4.11(2), where it is stated that:

$$\lambda\in\sigma(R_o)\Leftrightarrow\exists\mu\in\sigma(R),\ \lambda\in\sigma(\ F(\mu-A^*)(I-\mu A)^{-1}\ )$$

Now suppose that A,F and λ are given and we wish to find the solutions for μ. A straightforward calculation shows that the above is equivalent with the following condition:

$$\lambda\in\sigma(R_o)\Leftrightarrow\exists\mu\in\sigma(R),\ \mu\in\sigma(\ (\ F+\lambda A)^{-1}(FA^*+\lambda I)\ )$$

and in general this last condition has n different solutions, if $A\in\mathbb{R}^{n\times n}$.

3. Calculate the m signals Y_i by applying $Vec(U_i)$ as input to $W_o(z)$:

$$Vec(Y_i):=W_o(z)Vec(U_i)$$

4. Calculate the signals y_i: $\mathbb{N}^0 \to \mathbb{F}^p$ with the inverse signal transformation (5.4.5)
5. Let

$$M_t:=[y_1(t),y_2(t),\cdots,y_m(t)]$$

A state space model of the system $W(z)$ can be obtained from the Markov parameters $\{M_t\}$ by application of a realization method (Ho and Kalman, 1966, Silverman, 1971). o

Proof: This is immediate from Fig. 5.4.2. □

It is possible to express the Markov parameters in this procedure by means of products and sums, but we refrain from doing so, because those formulas do not shed more light on the structure. However in a special case it is relatively easy to derive expressions for the the variables U_i and Y_i, and get more insight in the structure of the problem.

Proposition 5.4.20. If in procedure 5.4.19 the matrix A has a singular value decomposition $A=ISV^*$ with S in decreasing order, then $B=[0,\cdots 0,,b]^*$ with $b\in\mathbb{R}$, $|b|\leq 1$. Let $\nu:=(1-b^2)^{\frac{1}{2}}$, $|\nu|<1$, and let the system $W_o(z)$ have Markov parameters $\{M_k^o, k\in\mathbb{N}^0\}$

Then the signals $Vec(U_i)$ and $Vec(Y_i)$ are given by:

$$[Vec(U_1)|\cdots|Vec(U_m)]_k = b\times(-\nu)^k \begin{bmatrix}0\\ I_m\end{bmatrix} \qquad (5.4.13a)$$

$$[Vec(Y_1)|\cdots|Vec(Y_m)]_k = b\times \sum_{i=0}^{k} (-\nu)^{k-i} M_i^o \begin{bmatrix}0\\ I_m\end{bmatrix} \qquad (5.4.13b)$$

o

Proof: The structure of B follows from lemma 5.3.2 and the fact that $B\in\mathbb{R}^{n\times 1}$. Furthermore this lemma shows that $S=diag(1,\cdots 1,\nu)$. Recall that the matrix B_e which defines the first column of $\Psi_e\{A,B\}$ is created in the following way: $\{A=USV^*$, $P=-UV^*A^*=-USU^*$ $B_e^*=[B^*|B^*P^*|\cdots]^*\}$. Because $U=I$ and B has the special structure it follows that $P^tB=(-\nu)^t b\times[0,\cdots 0,1]^*$. Now the signal u_i is a pulse signal, $u_i(t)\equiv 0$ for $t>0$ and $u_i(0)=[0,\cdots 0,1,0,\cdots 0]^*$. We calculate U_i according to definition 5.4.2:

$$Ui_k = \sum_t u_i(t)\varphi_k^*(t) = u_i(0)(P^kB)^* = (-\nu)^k b\times[0|\cdots|0|u_i(0)].$$

which implies $\quad Vec(U_{i_k})=(-\nu)^k b\times[0,\cdots,0,u_i^*(0)]^*$

The latter formula is equivalent with (5.4.13a) and (5.4.13b) is a direct result of the Markov parameter description of the system $W_o(z)$. o

Remark 5.4.21

1. Proposition 5.4.20 shows that a large part of the Markov parameters of $H_o(z)$ does not influence the result of the calculation; it follows from equation (5.4.13b) that only $M_k^o\begin{bmatrix}0\\ I_m\end{bmatrix}$ is of importance. We will not elaborate on this subject, aside from mentioning

that we expect that with a similar argument it can be shown that only p rows of M_k^o influence the calculations. This would account for the fact that not all (pn×mn) systems are the result of the transformation of a (p×m) system (see remark 5.4.18).

2. For every k we can calculate the orthonormal coefficients of the output according to

$$[\mathrm{Vec}(Y_1)|\cdots|\mathrm{Vec}(Y_m)]_k = b\times\left[(-\nu)^k M_0^0 + (-\nu)^{k-1}M_0^1 + \cdots + M_0^k\right]\times\begin{bmatrix}0\\ I_m\end{bmatrix}$$

Hence we can determine on beforehand from the system $H_0(z)$ and the value of ν how many of these coefficients are significant. This implies that the knowledge, obtained by procedure 5.4.19, can be applied to calculate the inverse system transformation in a pragmatic approximating way. o

Remark 5.4.22. Calculation of the infinity norm.

If we combine the results of corollary 5.4.13 and proposition 5.4.16 then we come to the following conclusion regarding the H∞–norm of a system:

> The H∞–norm computation of a general FDLTS system $W_t(z)$ can be reduced to the H∞–norm computation of a system $W_o(z)=M_0 + M_1z^{-1}$

This is the effect of using orthogonal functions that are based on the system W(z), and hence $n=n_w$. The matrices M_0 and M_1 can be calculated from the realization (5.4.8) and are given by: $M_0=L_0$, $M_1=H_oG_o$. If $W_t(z)$ has dimensions (p×m) and McMillan degree n, then $W_o(z)$ has dimensions (pn×mn) and McMillan degree n.

As far as we know there are no analytical methods known to calculate the H∞–norm of general FDLTS systems and the calculation is always an iterative process (Doyle and co–workers, 1989; Bruinsma and Steinbuch, 1990). The question arises if the above result could lead to alternative algorithms. This means that we have to calculate the H∞–norm of a system $W_o(z)$ with two Markov parameters, which is essentially MIMO because there are always multiple inputs and outputs. We have

$$W_o(z) = M_o + M_1z^{-1} \qquad (5.4.14)$$

where M_o and M_1 are (pn×mn) matrices and M_1 has rank n. This rank follows from the fact that the singular values of M_1 are the Hankel singular values of $W_o(z)$ and $W_t(z)$. We search for the minimal value of $\gamma\in\mathbb{R}_+$ that obeys

$$\forall\ z=e^{j\omega}\quad W_o(z)W_o^*(z)\leq\gamma I \qquad (5.4.15)$$

which is equivalent with

$$\forall\ z=e^{j\omega}\quad (M_oM_o^*+M_1M_1^*) + M_oM_1^*z + M_1M_o^*z^{-1} \leq\gamma I \qquad (5.4.16)$$

This problem would be trivial if $M_0=0$, but in general this will never be the case, as follows from evaluating (5.4.8). We have not succeeded in finding an analytical method to calculate this γ, except for the case that $\min(pn,mn)<3$. In the latter case it is fairly simple to derive an expression for the eigenvalues of the left hand side of (5.4.16) and

also for the derivatives of these expressions. Note that the eigenvalues are real valued and non–negative. Setting the derivatives equal to zero results in two or less values of z for which the eigenvalues reach an optimum. If we calculate these optima and the values at the boundary ($z=\pm 1$) then we know that the maximum of all these values is equal to the infinity norm γ. For the general case however, we feel that it might be possible to obtain results pursuing this line of thought, but that it might be necessary to use the information on the matrices M_0 and M_1 that can be retrieved from (5.4.8).

The literature on H∞ theory focuses mainly on continuous time systems. The above can be translated to the continuous time domain, by application of the ω–transformation. This results in:

Given a (p×m) FDLTS system $G_t(s)$ with McMillan degree n then there exists a (pn×mn) FDLTS system $G_o(s)$ with McMillan degree n, such that

$$\|G_t\|_\infty = \|G_o\|_\infty \quad \text{with} \quad G_o = M_0 + \frac{1-s}{1+s}M_1 = (M_0-M_1) + \frac{2}{1+s}M_1$$

where M_0 and M_1 are (pn×mn) matrices and M_1 has rank n and the non–zero singular values of M_1 are equal to the Hankel singular values of $G_t(s)$ and $G_o(s)$.

As stated, we believe that these transformed problems could lead to alternative methods to calculate the H∞ norm of a system. If this norm could be calculated analytically then new possibilities may arise for system approximation with respect to this norm. Hence more attention and research on this problem is highly recommended. o

Remark 5.4.23. Discrete Hankel norm approximation.
The transformation of linear systems as defined in this section has implications for the discrete Hankel norm approximation problem: Let G(z) be a FDLTS system with McMillan degree n and let $\hat{G}(z)$ be an optimal Hankel norm approximation of order $k<n$ to G(z), as defined by definition 3.6.1. Let Ψ_e be the set of orthonormal functions, generated by G(z), and denote the transformed systems of G(z) and $\hat{G}(z)$, under the transformation with these functions, by $G_o(z)$ and $\hat{G}_o(z)$ respectively. We know from proposition 5.4.15 and the linearity of the orthonormal transformations that the following systems have mutually the same Hankel singular values:

$$G(z) \text{ and } G_o(z) \qquad \hat{G}(z) \text{ and } \hat{G}_o(z) \qquad G(z)–\hat{G}(z) \text{ and } G_o(z)–\hat{G}_o(z)$$

The following reasoning shows that $\hat{G}_o(z)$ is an optimal Hankel approximation to $G_o(z)$:

a. G(z) and $G_o(z)$ have Hankel singular values $\{\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_k > \sigma_{k+1} \geq \cdots \geq \sigma_n\}$
b. $G(z)–\hat{G}(z)$ has a maximum Hankel singular value σ_{k+1}
c. $G_o(z)–\hat{G}_o(z)$ has a maximum Hankel singular value σ_{k+1}
d. $\hat{G}_o(z)$ is an optimal Hankel norm approximation to $G_o(z)$

Now observe that $G_o(z)$ is a system with only two Markov parameters and that it is very easy to find an optimal Hankel norm approximation to $G_o(z)$. However $G_o(z)$ is an essentially MIMO system (with dimensions pn×mn) and therefore there exist many different Hankel norm approximations to $G_o(z)$ (Glover, 1984). It is not true that each of these is the transform of an Hankel norm approximation to G(z). This is easy to understand by the following argument:

> Let G(z) be a SISO system, then Glover (1984) shows that there exists a unique (aside from addition with constant terms) Hankel norm approximation $\hat{G}(z)$. $G_o(z)$ is a MIMO system with many Hankel norm approximations, but there can be only one such approximation, that is the transform of $\hat{G}(z)$.

The reason for this phenomenon is that not every G_o–$\hat{G}_o$ can be considered as the transformation of a system G–$\hat{G}$, as described in remark 5.4.18. Under the assumption that it is known which optimal Hankel norm approximation to $G_o(z)$ is the transform of $\hat{G}(z)$ we can solve the optimal Hankel norm approximation problem for discrete time systems, by using the inverse system transformation. We have not succeeded in determining which Hankel norm approximations to $G_o(z)$ correspond to Hankel norm approximations to G(z), but from experience we know that in general the trivial approximations to $G_o(z)$, created by setting Hankel singular values equal to zero [5], are not corresponding to approximations to G(z). However we have the strong impression that this problem is solvable. A solution might give more insight in the approximation of systems and therefore further research is recommended. o

5.4.4. FREQUENCY DOMAIN EXPANSIONS

The orthonormal functions and the corresponding generating transfer functions can also be used to describe systems in the frequency domain by using these transfer functions as a basis for the frequency domain. In section 5.5 this concept will be applied in the context of system identification. The main idea can be compared with the expansion of a transfer function in the basis functions z^{-i}, as is done with the Markov parameter description (2.4.4). We generalize this idea by using the orthonormal transfer functions as a basis. If $\{H_k(z), k\in\mathbb{N}^0\}$ is a set of these transfer functions then we want to expand the transfer function of a FDLTS system in these terms, $G(z)=\sum_k W_k H_k(z)$ with $\{W_k, k\in\mathbb{N}^0\}$ a set of constant matrices. It will follow that these coefficients matrices are

---

[5] G_o is given by $G_o(z)=M_o+M_1z^{-1}$. If M_1 has svd USV* with $S=\mathrm{diag}(s_1,s_2,\ldots,s_n)$ and we define S_1 by $S_1=\mathrm{diag}(s_1,\ldots,s_k,0,0..0)$ then the system $\hat{G}_o(z):=M_o+US_1V^*z^{-1}$ is a trivial Hankel norm approximation because $\|G_o-\hat{G}_o\|_H = \|S-S_1\|_S = s_{k+1}$

the same as the orthonormal coefficients of the time series defined by the Markov parameters of G(z). Thus the matrices $\{W_k\}$ can be considered as of an expansion in both the frequency domain and the time domain. First we will consider the expansion in the time domain; the expansion in the frequency domain will be a direct consequence.

Let W(z) be a (p×m) FDLTS system with Markov parameters $\{w(t), t\in\mathbb{N}^0\}$. For the moment we discard the first Markov parameter w(0), the reason for that will be clarified later on. Define the vector time series $w^v:\mathbb{N}^0\to\mathbb{R}^p$, that is induced by the (shifted) Markov parameter matrix $[w(1)|\ w(2)\ |\cdots]$:

$$[w^v(0)|\ w^v(1)\ |\cdots] = [w(1)|\ w(2)\ |\cdots] \tag{5.4.17}$$

Note that $w^v(t)\in\mathbb{R}^p$ and $w(t)\in\mathbb{R}^{p\times m}$. The stability of W(z) assures that w^v is an ℓ_2^p–function. Furthermore we assume that a set $\Phi_e\{A,B\}$ of orthonormal functions is given, with $B\in\mathbb{R}^{n\times m}$, so B and W(z) have equal column dimension m. We can expand w^v in terms of the orthonormal functions with the formulas of section 5.4.2 and we will show that this leads to an expansion of the transfer function W(z) in terms of the generating transfer functions of $\Phi_e\{A,B\}$.

Expand w^v in terms of $\Phi_e\{A,B\}=\{\varphi_k, k\in\mathbb{N}^0\}$ as defined by definition 5.4.2:

$$w^v(t) = \sum_{k=0}^{\infty} W_k\varphi_k(t) \tag{5.4.18a}$$

so

$$[w^v(0)|w^v(1)|\cdots] = [W_0|W_1|\cdots] \times \begin{bmatrix} \varphi_0(0) & \varphi_0(1) & \cdots \\ \varphi_1(0) & \varphi_1(1) & \cdots \\ \vdots & \vdots & \end{bmatrix} \tag{5.4.18b}$$

We can describe the system W(z) by $\{w(t)\}$ in the sense that any input/output pair $(y,u)\in\mathcal{B}(W)$, with $y(t)=0$ for $t<M<0$, obeys the following relation:

$$y(t) = \sum_{s=0}^{\infty} w(s)u(t-s) = w(0)u(t) + \sum_{s=1}^{\infty} w(s)u(t-s) \tag{5.4.19}$$

or in matrix notation:

$$y(t)-w(0)u(t) = [w(1)|w(2)|\cdots] \times \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \end{bmatrix} \tag{5.4.20}$$

Substitution of (5.4.17) and (5.4.18b) in this equation yields:

$$y(t)-w(0)u(t) = [W_0|W_1|\cdots] \times \begin{bmatrix} \varphi_0(0) & \varphi_0(1) & \cdots \\ \varphi_1(0) & \varphi_1(1) & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \end{bmatrix} \tag{5.4.21}$$

Now we define the **orthonormal states** X_k by:

$$X_k(t) := [\varphi_k(0)\ \ \varphi_k(1)\ \cdots] \times \begin{bmatrix} u(t-1) \\ u(t-2) \\ \vdots \end{bmatrix} \tag{5.4.22}$$

which shows that we can describe the relation between u and y by

$$y(t) = w(0)u(t) + \sum_{k=0}^{\infty} W_k X_k(t) \tag{5.4.23}$$

Consider the orthonormal states X_k (5.4.22), then it is apparent that X_k is the output of the generating transfer function $H_k(z)$, as defined by proposition 5.3.5, when the shifted signal u(t–1) is applied as input. Thus we can express the orthonormal state X_k by

$$X_k(t) = z^{-1}H_k(z)u(t) \tag{5.4.24}$$

and hence (5.4.23) can be reformulated as

$$y(t) = w(0)u(t) + z^{-1}\sum_{k=0}^{\infty} W_k H_k(z)u(t) \tag{5.4.25}$$

This leads immediately to the expansion of W(z) in terms of the orthonormal functions, since we also know that y(t)=W(z)u(t),

$$W(z) = w(0) + z^{-1}\sum_{k=0}^{\infty} W_k H_k(z) \tag{5.4.26}$$

and this is the expression that we aimed at.

The reason for using the shifted impulse response w_s to derive this result (5.4.26) is that we prefer to have as few coefficients $\{W_k\}$ as possible in this representation. The minimum number of non–zero coefficients will occur if the set of orthonormal functions was taken from the transfer function W(z), i.e there exist matrices D and C such that W(z) could be written as

$$W(z) = D + C[zI-A]^{-1}B \tag{5.4.27a}$$

for in that case we have only p×(n+m) parameters (W(z) is p×m).

$$w(0)=D\ ,\ W_0=C,\ W_k=0 \text{ for } k>0 \tag{5.4.27b}$$

Another way of putting this is to say that we only wish to approximate the strictly proper part of W(z) in terms of the orthonormal transfer functions.
In the following proposition we formalize the result obtained.

Proposition 5.4.24. Let W(z) be a p×m FDLTS system and $\Psi_e\{A,B\}$ with $B\in\mathbb{C}^{n\times m}$ be a set of orthonormal functions with corresponding generating transfer functions $H_k(z)$, as defined by definition 5.3.4 and proposition 5.3.5. Then there exist unique matrices $D\in\mathbb{R}^{p\times m}$, $W_k\in\mathbb{R}^{p\times n}$, such that $W(z) = D + z^{-1}\sum_{k=0}^{\infty} W_k H_k(z)$ (5.4.28) o

As might be expected we can show that the Markov parameter description of W(z) is a special case of proposition 5.4.24. In corollary 5.3.13 it was shown that in the case that $\{A,B\}=\{0,I\}$ the corresponding orthonormal functions are the discrete pulse functions and if we substitute $\{A,B\}=\{0,I\}$ in the definition of $H_k(z)$ (5.3.10a) then it follows that $H_k(z) = z^{-k}I$. After substitution in (5.4.28) this results in

$$W(z) = D + \sum_{k=0}^{\infty} W_k z^{-k-1} \tag{5.4.29}$$

which is indeed the Markov parameter description (2.4.4) of W(z).

In the case of Markov parameters it is straightforward to find explicit formulas for the expansion coefficients $\{W_k\}$ in proposition 5.4.24. We now investigate if an analogous result can be derived for the general case. This turns out to have a close connection with the formulas we established in section 5.4.3 on the transformation of systems. The problem will again be approached through the state space description and we assume that W(z) is a FDLTS system and that [R,G,H,J] is a minimal realization of W(z),

$$W(z) := J + H[zI-R]^{-1}G \tag{5.4.30}$$

Furthermore we assume that we have a set of orthonormal polynomials $\Psi_e\{A,B\}$, where B and G have equal column dimension, and the corresponding set of generating transfer functions

$$H_k(z) = \left[[zI-A]^{-1}F[I-zA^*]\right]^k z[zI-A]^{-1}B \tag{5.4.31}$$

We wish to find formulas to express the matrices D and $\{W_k\}$ of the expansion (5.4.28). It is clear that D=J and we are left with the expression

$$H[zI-R]^{-1}G = \sum_{k=0}^{\infty} W_k \left[[zI-A]^{-1}F[I-zA^*]\right]^k [zI-A]^{-1}B \tag{5.4.32}$$

The following proposition gives an expression for the coefficients $\{W_k\}$ in terms of coupled Sylvester equations.

Proposition 5.4.25. Let W(z) be a FDLTS system and let [R,G,H,J] be a minimal realization of W(z). Further let $\Psi_e\{A,B\}$ be a set of orthonormal polynomials with corresponding generating transfer functions $H_k(z)$, as defined by definition 5.3.4 and proposition 5.3.5, where we assume that the column dimensions of the matrices G and B are equal. Then W(z) has a unique expansion

$$W(z) = D + z^{-1}\sum_{k=0}^{\infty} W_k H_k(z) \tag{5.4.33a}$$

with $D=J$ and $W_k = HQ_k$ (5.4.33b)

where $Q_0 = RQ_0A^* + GB^*$ (5.4.33c)

and $Q_{n+1} = RQ_{n+1}A^* + RQ_nF^* - Q_nAF^*$ (5.4.33d)

o

Proof: See appendix 5N.

Example 5.4.26. We show again how the Markov parameter description fits into this proposition, and therefore we assume that A=0 and B=I, then also F=I.
The equations (5.4.33) then reduce to: $W_k=HQ_k$ with $Q_0=G$ and $Q_{n+1}=RQ_n$ and thus $W_k=HR^kG$, the Markov parameters (2.4.4b) of W(z).

Another example is the case where W(z) is used to generate the orthonormal functions, i.e A=R and B=G. In that case we get: $Q_0=AQ_0A^*+BB^*$ and the

normality of {A,B} implies that $Q_0=I$. Further $Q_{n+1}=AQ_{n+1}A^*+(AQ_n-Q_nA)$ and since $Q_0=I$ this implies $Q_1=AQ_1A^*$ and the stability of A shows that $Q_1=0$

(use Kronecker notation: $(I-A\otimes A)\mathrm{Vec}(Q_1)=0$, $(I-A\otimes A)$ regular and thus $Q_1=0$).
The same argument yields that $Q_k=0$ for $k>0$. This shows that we are left with $W_0=H$ and $W_k=0$ for $k>0$, which is exactly what was to be expected. o

When Kronecker calculation is applied to the formulas (5.4.33) then we find more explicit formulas for the coefficients W_k.

Proposition 5.4.27. The coefficients $\{W_k\}$ as defined by (5.4.33) are determined by,

$$\mathrm{Vec}(W_k) = (I\otimes H)M^{-1}(NM^{-1})^k\mathrm{Vec}(GB^*) \tag{5.4.34a}$$

where

$$M = (I\otimes I)-(A\otimes R) \tag{5.4.34b}$$

$$N = (F\otimes R)-(FA^*\otimes I) \tag{5.4.34c}$$

o

Proof: This is immediate from the propositions 2.9.2 and 5.4.25. □

Remark 5.4.28. This proposition shows directly the close connection with the transformation of systems (section 5.4.3), since the matrices N and M (5.4.34) are the same as those defined by (5.4.8). In section 5.4.3 we showed that the transformed system had a realization $[R_o,G_o,H_o,L_o]$ with $R_o=NM^{-1}$ and $H_o=(I\otimes H)M^{-1}$. If we use the same definition of R_o and H_o for the equation (5.4.34a) then we can write:

$$\mathrm{Vec}(W_k) = H_oR_o^k\mathrm{Vec}(GB^*) \tag{5.4.35}$$

In other words we can consider $\mathrm{Vec}(W_k)$ as a series of Markov parameters of a system with realization $[R_o,\mathrm{Vec}(GB^*),H_o,0]$. With the same arguments that were used to proof lemma 5.4.9 it can be shown that this realization has $(n-m)n_w$ uncontrollable poles in $z=-1$. o

We would like to find an analog of the standard realization methods (Ho–Kalman, 1966; Silverman, 1971) to retrieve the system W(z) of the sequence of parameters $\{W_k\}$, but we have not succeeded to find a satisfactory result. We can of course consider the sequence $\{\mathrm{Vec}(W_k)\}$ and retrieve a system which realizes this sequence which according to (5.4.35) can be seen as a sequence of Markov parameters in the classical sense. Under the condition that we have enough of these parameters, it will thus be possible to find matrices R_1,G_1 and H_1, such that $W_k = H_1R_1^kG_1$, but the problem remains how to retrieve from these the matrices R,G and H. The realization of W(z) from $\{W_k\}$ remains an open problem and in section 5.5 we will discuss it again in an identification context.

5.4.5 OVERVIEW AND DISCUSSION

In section 5.4 we presented a series of concepts and results and raised many questions. In this sub–section we give a short overview of the main issues. We assume that $\Psi_e\{A,B\}$ is a set of orthonormal functions, based on a dynamical system and that $\{H_k(z),k\in\mathbb{N}^0\}$ is the corresponding set of generating transfer functions.

In section 5.4.2 we defined the expansion of an ℓ_2 function f in terms of $\Psi_e\{A,B\}$, showing that this results in a set of coefficients $\{F_k,k\in\mathbb{N}^0\}$, that can be considered as a new ℓ_2 function. We commented on the approximation character of such an expansion if only a finite number of samples of f is available.
In section 5.4.3 we restricted the matrix B to have only one column and applied the signal transformation to the input/output signals of a (p×m) FDLTS system $W_t(z)$ with McMillan degree n_w: $\{u(t),y(t),t\in\mathbb{N}^0\} \rightarrow \{U_k,Y_k,k\in\mathbb{N}^0\}$.

We proved that the vectors with these new signals $\{Vec(U_k),Vec(Y_k),k\in\mathbb{N}^0\}$ are the input/output variables of a FDLTS system $W_o(z)$ in the orthonormal domain with the same McMillan degree, but larger dimension (pn×mn). Since this induced system transformation is bijective we conclude that only a restricted class of (pn×mn) can be considered as the transform of a (p×m) system.
We presented a realization $[R_o,G_o,,H_o,L_o]$ (stable nor minimal) of $W_o(z)$ in terms of Kronecker products and we proved that if $A=I_nSV^*$ (svd) then:

$$R_o=\begin{bmatrix}-I & 0\\ * & *\end{bmatrix} \qquad G_o=\begin{bmatrix}0\\ *\end{bmatrix}$$

For general A we showed that R_o and G_o can be brought to this form with a similarity transformation. The latter expression provides a minimal realization of $W_o(z)$ in a straightforward way.
For the special case that A and R are similar it followed that $W_o(z)$ has only uncoupled poles in zero and is determined by its first two Markov parameters:

$$W_o(z)=L_o+z^{-1}(H_oG_o)$$

Furthermore we established that the systems $W_t(z)$ and $W_o(z)$ have both the same Hankel singular values and $H\infty$ norm and we commented on the implications of this property for Hankel norm approximation and $H\infty$ norm computation.
In section 5.4.4 we only posed the restriction that B has full column rank m and we considered another application of expansions in terms of orthonormal functions. We transformed the p–dimensional vector time series that is defined by the Markov parameter matrix of a (p×m) FDLTS system W(z) and showed that this is equivalent with a description of W(z) in terms of the transfer functions $\{H_k(z)\}$:

$$W(z) = D + \sum_{k=0}^{\infty} W_k z^{-1} H_k(z)$$

with $\{D,W_k,k\in\mathbb{N}^0\}$ constant coefficient matrices. This expansion can be understood as a generalization of the Laurent expansion of W(z) round $z=\infty$, also know as the Markov parameter description. We elaborated on the generalized realization problem, that emerges from this representation, i.e. given $\{D,W_k,k\in\mathbb{N}^0\}$ find a realization of the system W(z). This problem was shown to be related to the system transformation, discussed in section 5.4.3, because the $\{W_k\}$ can be interpreted as the Markov parameters of a system with realization $[R_o,*,H_o,*]$.

We discussed a number of resulting problems, that are recommended for further research:

- Determination of the restrictions on the structure of systems $W_o(z)$, that result from the system transformation of orthonormal functions.
- Derivation of expressions for a minimal realization of $W_o(z)$ that give more insight in the pole–zero structure of this system.
- Development of a more direct method to perform the inverse system transformation.
- Considering the property that $W_o(z)$ is given by the simple equation $W_o(z)=M_0+z^{-1}M_1$ if A and R are similar:
  - which Hankel norm approximations to $W_o(z)$ correspond with Hankel norm approximations to $W_t(z)$? This would solve the Hankel norm approximation problem.
  - The same question with the H_∞ norm, this might lead to optimal H_∞ norm approximation.
  - Is it possible to analytically calculate the H_∞ norm of a MIMO system with two non–zero Markov parameters? This would result in an analytical H_∞ norm computation for general FDLTS systems.
- Solution of the generalized realization problem, which will be useful for the application of the expansion $W(z) = D + \sum_{k=0}^{\infty} W_k z^{-1} H_k(z)$ for the purpose of approximate system identification.

Aside from these issues that were raised in the previous sections, we can also pose a number of more fundamental questions:

- Does a model reduction method exist that corresponds with truncation of the number of orthonormal functions in a system description.
- Do model reduction methods give rise to alternative sets of orthonormal or orthogonal functions, that can be applied for system descriptions, with the property reduction ≈ truncation (to a smaller number of orthonormal functions).

- What is the implication of the use of orthonormal functions based on input balanced realization versus output balanced realizations. Is there a method to incorporate both sets in a transformation context.
- The definition of system based orthonormal functions can be generalized to functions defined on $\mathbb{Z}$, for instance by allowing unstable systems. The stable part is used as in this chapter, the unstable part is used to define functions on $\mathbb{Z}_-$. What are the implications of such a generalization.

In the scope of these thesis we did not consider the latter problems, but as explained in chapter one they are intrinsically related to the general problem of system approximation.

5.5. SYSTEM IDENTIFICATION

5.5.1 INTRODUCTION

In this section we use the orthonormal functions based on linear systems to address the problem of approximate system identification, so it will not be assumed that we aim at finding a 'true' system with identification techniques. As described in chapter one this is motivated by various reasons such as:

- in general it will be impossible to avoid modeling errors;
- the model classes that are commonly considered are very restricted;
- the resulting models are often required to be of a restricted complexity in order to be of any use for practical applications.
- the number of samples of the measurement data is finite and often small.

Hence, while the 'true' system might be of a high order, non–linear and/or time varying, we wish to approximate its behavior with low order, linear and time invariant models. Within the field of system identification the methods that are most commonly used are prediction error (PE) methods (Ljung, 1987), and this applies also for the field of approximate identification. There are a number of results on the properties of prediction error methods in approximate identification (Ljung, 1985,1987) and Wahlberg and Ljung (1986) derived several expressions on the variance and bias of the resulting estimates for PE methods, all of which are given in the frequency domain. In the time–domain Swaanenburg and co–workers (1985) and Van den Hof and Janssen (1987) presented results on the fit of Markov parameters for equation error methods. In Wahlberg and Ljung (1986), Ljung, 1987, 1989) it was shown that the results of approximate identification depend highly on what is known as the design variables, such as the choice

of sampling intervals, the length of the prediction horizons and probably most important the choice of prefilters. In general model structures (4.6.8) the use of a prediction error method leads to algorithms with a high complexity, due to the need to use non–linear optimization techniques to solve the problem. In our opinion it would be preferable if it would be possible to construct and apply prefilters, such that the use of relatively simple model structures and estimation schemes in combination with the filtered data would lead to satisfactory results. In this section we consider the use of orthonormal functions to construct such prefilters. As mentioned in chapter four this has been considered before in the early work of Lee (1933) and Wiener (1949) and more recently it was applied by King and Paraskevopoulos (1979), Paraskevopoulos (1985), Nurges and Yaaksoo (1981), Nurges (1987) and Wahlberg (1989), where Laguerre polynomials and connected functions where used in identification settings. These applications have partially been described in chapter four and in this section we aim at a generalization by using the system based orthonormal functions instead of Laguerre polynomials. Another application is the use of orthogonal functions to address the problem of identification of continuous time systems, see the papers of Paraskevopoulos (1985) and Unbehauen and Rao (1988) and the references therein. Though the theory which we developed in the previous sections has a straightforward analog for continuous time systems, we will not go into this.

In chapter one we explained that the problems of model reduction and approximate system identification are closely connected. If the orthonormal functions are based on a model reduction procedure, for instance in the sense of truncate = reduce, then the approximation of a system behavior with these orthonormal functions would be based on the same criterion as the model reduction method. We stated in chapter one that in an ideal situation a variety of such criteria would be available. We have not been able to show that the system based orthonormal functions, presented in the previous sections, are intrinsically based on a model reduction procedure, but it is apparent that they are closely related to the reduction methods based on balanced realizations.

The motivation to consider the system based orthonormal functions in the context of approximate system identification is twofold:

1. Approximate identification on the basis of measurement data is equivalent with distinguishing a specific structure that is present in these data and with creating a model that accounts for this structure. In this context the search for an optimal set of orthonormal functions can be considered as an operation with the same objective. Both applications of the orthonormal functions on system level, that are presented in section 5.4, i.e the system transformation and the transfer function expansion, can be considered to perform in an optimal way if the set of orthonormal functions is generated by the system that creates the data, in the sense that the number of coefficients necessary to

describe the result of the transformation is minimal (see corollary 5.4.14 and example 5.4.26). In this situation the transformation has removed most of the structure, that was present in the data. We conclude that the search for an optimal orthonormal domain is complementary to the identification problem.

2. Within the set of general model structures (4.6.8), as proposed by Ljung (1987) two model classes exist that, in combination with a least squares criterion, lead to an analytically solvable estimation problem. These model classes, referred to as ARX and FIR, both have the drawback that high order models have to be used in order to get unbiased estimates of the underlying system. However a larger number of parameters enlarges the variance of the estimated parameters and especially in the case of finite data this can cause bad estimation results. The use of low order models improves the variance, but results in an unbiased estimate of only a small number of Markov parameters and this number is often too small to describe the underlying system completely. This problem can partly be overcome by the use of prefilters (Ljung, 1987,1989) but the optimal prefilter is unknown. We propose to use the transformation of signals with the system orthonormal functions as a type of prefilter operation and to use the ARX model structure in combination with a least squares criterion to estimate a model in the orthonormal domain. The motivation is that a good choice of the orthonormal functions will absorb the dynamic behavior in the orthonormal domain. A second method that we present is a generalization of the estimation of FIR models. We propose to use the expansion of transfer functions in terms of the orthonormal functions, as described in section 5.4.4 and to estimate a finite number of the expansion coefficients with a least squares criterion. Since every basis function has an infinite horizon this procedure will not result in an estimate of the first n Markov parameters, but will estimate as many parameters as available data points. An appropriate choice of orthonormal functions will result in a few significant parameters. Both methods are based on a least squares estimation in a model structure that is linear in the parameters, generally leading to a unique solution.

These two methods are the subject of section 5.5.2 and 5.5.3 and they are illustrated with examples in section 5.5.4.

5.5.2 AN IDENTIFICATION METHOD BASED ON THE TRANSFORMATION OF SYSTEMS

It is a well known fact that the use of prefilters on data in approximate identification methods can have an important influence on the resulting models. Ljung (1987) showed how a prefilter can be used to improve the frequency behavior of the model estimate.

However in general it is not known a priori which prefilter should be used to obtain the desired result. In fact this can only be calculated if the underlying system is known. We propose to use a method that can be explained in this context. It is a generalization of the identification method that we considered in section 4.6.3 (ARX in the Laguerre domain) and the methods proposed by Nurges and Yaaksoo (1981) and Nurges (1987). The basic principle is explained in the following three step procedure:

PROCEDURE 5.5.1. ARX in the orthonormal domain

Let $\Psi_e\{A,B\}$ be a set of system based orthonormal functions and let $\{y(t),u(t),t\in\mathbb{N}^0\}$ be a set of measurement data. We define the model estimate, *based on ARX in the orthonormal domain* as the result of the following procedure:

1. Transform the measurement data to the orthonormal domain
$$\{y(t),u(t),\ t\in\mathbb{N}^0\} \rightarrow \{Y_k,U_k,k\in\mathbb{N}^0\}$$
2. Estimate an ARX model on the vectorized transformed data $\{Vec(Y_k),Vec(U_k),k\in\mathbb{N}^0\}$ with a least squares identification criterion
3. Transform the ARX model to the time domain. o

This procedure in visualized in Fig. 5.5.1.

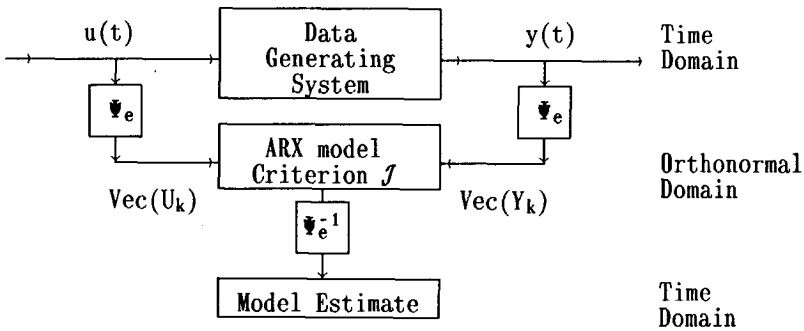


Fig. 5.5.1 ARX in the orthonormal domain

We explain the details and concepts of procedure 5.5.1 in the next four remarks, where we will comment on each of the steps separately (in reversed order), and on the influence of finite data sequences in this procedure. It will follow that the procedure can only partially be explained and has a somewhat adhoc character.

REMARK 5.5.2 Comment on procedure 5.5.1. ad 3: inverse transformation.
Assume that the first two steps of procedure 5.5.1 have been performed and that an ARX model has been estimated in the orthonormal domain. The inverse system transformation can be applied to transform this system back to the time–domain, under a specific restriction. The latter is the result of an observation that was made in remark 5.4.18. If $y \in \mathbb{R}^p$, $u \in \mathbb{R}^m$ and $A \in \mathbb{R}^{n \times n}$ then $\mathrm{Vec}(Y_k) \in \mathbb{R}^{pn}$ and $\mathrm{Vec}(U_k) \in \mathbb{R}^{mn}$ and the ARX model in the orthonormal domain is a model with output–input dimension (pn×mn). In remark 5.4.18 we showed that only a restricted class of the set of (pn×mn) systems can be considered as the result of the system transformation with orthonormal functions. A complete characterization of these restrictions is not known. This implies that specific restrictions must be made concerning the structure of the ARX model, in order to guarantee that a unique time domain equivalent can be calculated. If such restrictions are not imposed, for instance because they are not known, then it follows that the result of the inverse system transformation, as described by procedure 5.4.19, will be an approximation. o

REMARK 5.5.3 Comment on procedure 5.5.1. ad 2: ARX models
Suppose that we are given input/output data $\{u(t),y(t)\}$ of a plant or system and we wish to approximate the plant by a FDLTS model. Let $\Psi_e\{A,B\}$ be a set of system based orthonormal functions and apply the transformation of time series, defined in section 5.4.2, to obtain the orthonormal coefficients $\{U_k,Y_k\}$ and consider the (Kronecker) vectors of these coefficients $\{\mathrm{Vec}(U_k),\mathrm{Vec}(Y_k)\}$. We propose to fit a fairly simple model to these data with a least squares criterion. For the model we choose an ARX (Auto Regressive eXogenous) model (Ljung, 1987):

$$\mathrm{Vec}(Y_k) = \sum_{i=1}^{n} P_i \mathrm{Vec}(Y_{k-i}) + \sum_{i=0}^{m} Q_i \mathrm{Vec}(Y_{k-i}) + E_k \qquad (5.5.1)$$

where E_k is the equation error of the model and the matrices $\{P_i,Q_i\}$ are assumed to be constant and of appropriate dimensions. We consider these matrices as the unknown parameters and we estimate these parameters with the well known least squares criterion

$$\mathcal{J}\{P_i,Q_i\} := \sum_k E_k^* E_k \qquad (5.5.2)$$

It is well known that the minimization of this criterion function with respect to the parameters can be calculated analytically and leads to a unique solution $\{\hat{P}_i,\hat{Q}_i\}$, provided that we choose some kind of normalization to exclude the solution to be zero and to assure identifiability or discriminability (Ljung, 1987; Van den Hof, 1989) This procedure thus leads to a model in what we call the orthonormal domain. o

REMARK 5.5.4 Comment on procedure 5.5.1. ad 1: signal transformation.
In section 4.6.3 it was shown that transformation of measurement data with Laguerre

polynomials is (in the asymptotic case) equivalent with the application of a prefilter and the choice of different polynomial restrictions for the model structure.
For the generalized orthonormal functions such an analysis cannot be given because we have not enough insight in the system transformation with orthonormal functions. However the Laguerre case leads to the strong impression that something alike must be valid for the general case. The following argument makes this idea more plausible:

Suppose we have transformed the measurement data to the orthonormal domain and fit the data in the following ARX model:

$$P_0\mathrm{Vec}(Y_k) + P_1\mathrm{Vec}(Y_{k-1}) + Q_0\mathrm{Vec}(U_k) + Q_1\mathrm{Vec}(U_{k-1}) = \mathrm{Vec}(E_k) \qquad (5.5.3)$$

For this explanation we must assume that all these sequences are ℓ_2 functions and have time domain equivalents. Equation (5.5.3) can be rewritten as:

$$E_k = \left[\, [P_0\ Q_0] + z^{-1}[P_1\ Q_1] \,\right] \begin{bmatrix}\mathrm{Vec}(Y_k)\\ \mathrm{Vec}(U_k)\end{bmatrix} \qquad (5.5.4a)$$

From the definition of Vec(X) it follows that a simple permutation matrix T exists such that

$$T\begin{bmatrix}\mathrm{Vec}(Y_k)\\ \mathrm{Vec}(U_k)\end{bmatrix} = \mathrm{Vec}\begin{bmatrix}Y_k\\ U_k\end{bmatrix}$$

$$E_k = \left[\, [P_0\ Q_0]T^* + z^{-1}[P_1\ Q_1]T^* \,\right] \mathrm{Vec}\begin{bmatrix}Y_k\\ U_k\end{bmatrix} \qquad (5.5.4b)$$

This describes a system $W_o(z)$ with two Markov parameters and in section 5.4.3 we showed that such systems are the result of the orthonormal system transformation if the correct system matrix was used to generate the orthonormal functions $\Psi_e\{A,B\}$. Note that for this conclusion we must assume that the system (5.5.4) obeys the restrictions that are commented on in remark 5.5.2. Under that assumption we conclude that a system $W_t(z)$ exists such that

$$W_t(z) = L + H[zI-R]^{-1}G$$

$$e = W_t(z)\begin{bmatrix}y\\ u\end{bmatrix} \qquad (5.5.5)$$

where e is the inverse transform of E. The only restriction on R is that it is similar with A, thus R has the same characteristic polynomial D(z) as A. The transfer function $W_t(z)$ can always be decomposed as a ratio of D(z) and a numerator polynomial M(z):

$$e = \frac{M(z^{-1})}{D(z^{-1})}\begin{bmatrix}y\\ u\end{bmatrix} \qquad (5.5.6)$$

which shows that we are dealing with an ARX model in filtered data:

$$e = M_1(z^{-1})y_f + M_2(z^{-1})u_f \qquad (5.5.7a)$$

where $M=[M_1\ M_2]$

$$\begin{bmatrix}y_f\\ u_f\end{bmatrix} = \frac{I}{D(z^{-1})}\begin{bmatrix}y\\ u\end{bmatrix} \qquad (5.5.7b)$$

It follows, under the assumptions made above, that the model (5.5.3) is equivalent with the model (5.5.7b). The Parseval identity implies $\|E\|_2=\|e\|_2$ and we conclude that least squares estimation of the parameters in both models is equivalent.
In the case that the order of the ARX model (5.5.3) is n>2 a similar argument can be used to show that the corresponding model (5.5.4) has only n non–zero Markov parameters, which shows that it is the result of an orthogonal transformation of a system whose system matrix R has the same poles as A, but is no longer similar with A (see lemma 5.4.11). However both matrices have the same characteristic polynomial and hence the same line of reasoning applies for this situation.
In chapter four we were able to show that ARX in the Laguerre domain is, in the asymptotic case, equivalent with ARX in the time domain on prefiltered data with different polynomial restrictions for the ARX model. This latter phenomenon is likely to be present in the generalized case, but we cannot find expressions for the transformed restrictions, since the structure of the model (5.5.3) is not thoroughly understood.
We repeat that the above is by no means a proof of the fact that an ARX model on the transformed data is equivalent with an ARX model on filtered data, but it is only meant to make this statement plausible. In the remainder we will call the transformation operation a prefilter operation.
Note that if the data are generated by an output error model: $y=\frac{M(z^{-1})}{D(z^{-1})}u + \epsilon$ then the estimation in the model structure $P(z^{-1})y_f = Q(z^{-1})u_f$ will result (under appropriate conditions on ϵ, u and polynomial degrees) in the solution P=D and Q=M. This concept is used by Steiglitz and McBride (1965) who propose an iterative ARX estimation, using the estimated denominator as prefilter. The aim of this procedure is that the resulting model estimates converge to an optimal output error model. This motivates us to propose the same concept for ARX in the orthonormal domain, i.e to use the resulting model estimate for the generation of new orthonormal functions and to repeat procedure 5.5.1 with this new basis.
If, for explanatory reasons, we assume that the data generating system is a FDLTS system G(z) and there is no noise involved in the data generation, then we learn from section 5.4.3 that the set of orthonormal functions, which is induced by G(z), leads to an orthonormal system with only two Markov parameters. This motivates the idea that one should be searching for that set of orthonormal functions that minimizes the dynamic behavior in the orthonormal domain.
To start such an iterative process an initial set of orthonormal functions can be chosen on various grounds, such as a priori knowledge about important modes of the system or information obtained from previous modeling. An alternative choice is the set of Laguerre polynomials. o

REMARK 5.5.5 Comment on procedure 5.5.1. : finite data and experiment setup.
For application of procedure 5.5.1 to an actual set of measurement data we must recognize that the available data will be of finite length. As explained in section 5.4.2 this implies that the orthonormal coefficients cannot be computed exactly, but can only be approximated. The reason for this phenomenon is that for the calculation of the coefficients we must implicitly assume a specific extension of the data series towards infinity. This problem can partially be solved if the experiment with which the data are created is organized in such a way that the tail of the input and output signals tends to zero. Practically this means: set the input to zero for t>N and keep recording the output till it has decayed to zero as well. Another important issue concerns the question of the optimal input signals. Since the actual estimation is performed in the orthonormal domain, the orthonormal input signal U_k should be persistently exciting. We have not studied this issue in detail, but do note that it should be investigated. o

We conclude with stating that we presented an approach to approximate identification which has a close relationship with equation error identification combined with the application of prefilters. The precise character of these prefilters is unknown but it was made plausible that they are connected with the characteristic polynomial of the matrix A that was used to generate the orthonormal polynomials. The approximation character of the procedure is threefold:

1. the data transformation (remark 5.5.5)
2. the modeling process (the true model does not exist, section 5.5.1)
3. the inverse system transformation (remark 5.5.2)

In section 5.5.4 we will present examples of the application of this approach.

5.5.3 AN IDENTIFICATION METHOD BASED ON THE EXPANSION OF TRANSFER FUNCTIONS

This second approach towards approximate identification is based on the theory of section 5.4.4. There it was explained how the transfer function of a FDLTS system can be described as an expansion in terms of the generating transfer functions of a set of system based orthonormal functions. We propose the estimation of a finite number of the expansion coefficients and show that this is a generalization of the estimation of a finite number of Markov parameters (FIR models). In combination with a least squares criterion the optimal coefficients can be calculated analytically. The procedure is a generalization of the method described in section 4.6.4, FIR in the Laguerre domain.

PROCEDURE 5.5.6. FIR in the orthonormal domain

Let $\Psi_e\{A,B\}$ be a set of system based orthonormal functions with corresponding transfer functions $\{H_k(z),k\in\mathbb{N}^0\}$, as defined by definition 5.3.4 and proposition 5.3.5. Let $\{y(t),u(t),t\in\mathbb{N}^0\}$ be a set of measurement data. We define the model estimate $\hat{G}(z)$, ***based on FIR in the orthonormal domain*** as the result of the following procedure:

1. Estimate the parameters $\{\hat{D},\hat{C}_k,k=0\cdot\cdot N\}$ such that

$$\{\hat{D},\hat{C}_k\} = \text{Arg Min } \mathcal{J}_Q\{D,C_k\} \tag{5.5.8a}$$

where
$$\mathcal{J}_Q\{D,C_k\} := \sum_{t=0}^{\infty} e(t)^* Q e(t) \tag{5.5.8b}$$

with
$$e := y - \left[D + \frac{1}{z}\sum_{k=0}^{N} C_k H_k(z)\right]u \tag{5.5.8c}$$

2. Define
$$\hat{G}(z) := \hat{D} + \frac{1}{z}\sum_{k=0}^{N} \hat{C}_k H_k(z) \tag{5.5.8d}$$

3. Reduce $\hat{G}(z)$ with a model reduction method or an approximate realization procedure. ○

In the next three remarks we consider respectively the motivation, the computational aspects and the reduction step of procedure 5.5.6.

REMARK 5.5.7. Motivation for procedure 5.5.6.

In section 5.4.4. we showed that every FDLTS system G(z) can be described by an expansion in terms of $\{H_k(z)$:

$$G(z) = D + \frac{1}{z}\sum_{k=0}^{\infty} C_k H_k(z) \tag{5.5.9}$$

and we showed that this expansion is equivalent with the expansion in terms of the orthonormal function $\Psi_e\{A,B\}$ of the function that is defined by the columns of the Markov parameter matrix of G(z). The idea is to replace this exact description of the Markov parameter matrix by an approximation with a finite number of orthonormal functions, optimal in least squares sense. This results in a new set of coefficients $\{\hat{D},\hat{C}_k$, $k\in\mathbb{N}^0$, $\hat{C}_j\equiv 0$ $j>N\}$. This approximation then corresponds with a new system

$$\hat{G}(z) = \hat{D} + \frac{1}{z}\sum_{k=0}^{N} \hat{C}_k H_k(z) \tag{5.5.10}$$

In the context of approximate identification we assume that the input/output behavior of a plant can be approximated with a Markov parameter description

$$y(t) = \sum_{i=0}^{\infty} M_i u(t-i) + E(t) \tag{5.5.11}$$

or
$$y = G(z)u + E$$

and we approximate the Markov parameters with a finite number of orthonormal functions. In section 5.4.4 we showed (with $E\equiv 0$) that if the orthonormal functions are generated by G(z), then only the first two coefficient matrices are unequal to zero. A

well known application of this method is the least squares estimation of FIR models, i.e. the estimation of a finite number of Markov parameters. This follows from example 5.4.26, where it was explained that if $\{A,B\}=\{0,I\}$ then the parameters $\{D,W_k\}$ are the Markov parameters of G(z). o

Remark 5.5.8. Computational aspects of procedure 5.5.6.

Recall the concept of so called orthonormal states $X_k(t)$, that was used in section 5.4.4. These are vector time series defined by:

$$X_k := z^{-1}H_k(z)u \tag{5.5.12}$$

This implies that the residue or error signal e(t) (5.5.8c) obeys

$$e(t) = y(t) - Du(t) - \sum_{k=0}^{N} C_k X_k(t) \tag{5.5.13}$$

Note that the orthonormal states can be calculated directly from u(t) and $H_k(z)$. Thus the residual e is linear in the parameters $\{D,C_k\}$ and an analytical solution exists to the minimization of the criterion $J_Q\{D,C_k\}$, under the condition that the input u is sufficiently exciting.

The recursive structure of the generating transfer functions $H_k(z)$ can be used for efficient computation of the orthonormal states; we present a method that is less efficient from a computational point of view but that explains the underlying structure of the problem:

Let A_e and B_e be the matrices that constitute the orthonormal functions. Consider the state equation

$$X(t+1) = A_eX(t) + B_eu(t)\ ,\ X(0)=0 \tag{5.5.14}$$

Then

$$X(t) = \begin{bmatrix} X_0(t) \\ X_1(t) \\ \vdots \end{bmatrix} \tag{5.5.15}$$

where $\{X_k(t)\}$ are defined by (5.5.12).

This follows from proposition 5.3.5. The lower triangular structure of A_e shows that the vector X_N with the first n×N elements of X(t) (n=dim(A)) obeys a state space equation

$$X_N(t+1) = A_NX(t) + B_Nu(t)\ \ ,\ X_N(0)=0 \tag{5.5.16a}$$

where A_N and B_N are the corresponding (left) upper submatrices of A_e and B_e. This shows that we can present the model in a state space form by adding the output equation

$$y(t) = CX_N(t) + Du(t) + e(t) \tag{5.5.16b}$$

Where

$$C:=[C_1|C_2|\cdots|C_N].$$

This realization can be used for model reduction, as shown in the next remark. o

REMARK 5.5.9. Model reduction aspects of procedure 5.5.6.
Consider the model (5.5.8d), which has a state space realization

$$[A_N, B_N, C, D] \tag{5.5.17}$$

where A_N and B_N are defined by (5.5.16), $C=[\hat{C}_1|\hat{C}_2|\cdots|\hat{C}_N]$ and $D=\hat{D}$

The McMillan degree of this model depends on the estimated parameters $\{\hat{C}_k\}$ but can be as high as n×N. For practical applications this order will often be too high and this raises the question how we can reduce the order. We consider three different approaches to this problem:

1. Generalized realization
   In section 5.4.4 we commented on the generalized realization problem that is induced by the expansion of a transfer function in orthonormal transfer functions. A solution to this problem would yield one method to determine an underlying system of lower order.

2. Approximate realization based on Markov parameters
   We can calculate the Markov parameters of the system (5.5.8d) and with these any approximate realization method (Kung, 1979; Backx, 1987; Van Helmont and co–workers, 1990) can be applied to derive lower order models.

3. Model reduction of the state space model
   In chapter three we considered several model reduction methods, that are all based on state space descriptions. Most of these procedures require a balanced realization of the system. For this purpose we can use the structure of the orthonormal functions, because we know that the matrices A_N and B_N are normalized:

   $$A_N A_N^T + B_N B_N^T = I \tag{5.5.18}$$

   This implies that we only have to calculate the observability Gramian of the system in order to find the similarity matrix that transforms the matrices into a balanced realization.

Experiences with the latter two methods show that the third method leads to better results. Since the general realization problem is unsolved we have no experience with the first method. o

This method thus generalizes the estimation of Markov parameters in FIR models and the Laguerre approach of Zervos, Belanger and Dumont (1985, 1988), which was discussed in chapter 4.

Figure 5.5.2 depicts the structure of this approach.

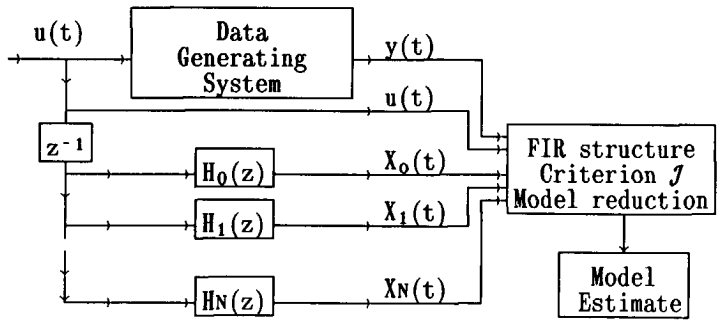


Fig. 5.5.2. FIR in the orthonormal domain

In the next section we present examples of the two approaches which were discussed, i.e ARX and FIR in orthonormal terms..

5.5.4. EXAMPLES

In this section we demonstrate the approximate identification methods wich have been discussed in sections 5.5.2 and 5.5.3 by means of two examples. The results are compared with the results of the standard identification methods, of which the proposed methods are generalizations, so we compare the result of ARX in the orthonormal domain with the result of ARX in the time domain and the result of the decomposition of a transfer function in orthonormal elements is compared with the result of the estimation of Markov parameters.

EXAMPLE 1

We simulated a SISO system with McMillan degree 3, by using a PRBS (Pseudo Random Binary Sequence) as input, and we corrupted the resulting output with additive noise, such that the signal to noise ratio on the output is 0 dB. The system is given by the following transfer function,

$$G(z) = \frac{3z^2 - 5.7z + 2.75}{z^3 - 2.75z^2 + 2.56z - 0.8075}$$

with poles in $0.9 \pm 0.2i$ and 0.95 and zeros in $0.95 \pm 0.119i$.

The system has important high and low frequency behavior. This may be deducted from Fig. 5.5.4, where the solid lines depict the step response respectively the Bode amplitude of this system. We use the following notation to describe the data:

u(t): input signal (PRBS) y(t) : deterministic output $y(t) = G(z)\,u(t)$

n(t): additive noise $y_n(t)$: corrupted output $y_n(t) = y(t) + n(t)$

Method 1 ARX in the orthonormal domain

The following procedure was followed

1. Use Laguerre polynomials with discount factor ξ=0.5 and transform u(t) and $y_n(t)$ to U_k and Y_k, signals in the Laguerre domain.
2. Estimate with U_k and Y_k an ARX model $\hat{G}_\ell(z)$ of McMillan degree 5.
3. Transform $\hat{G}_\ell(z)$ to a time domain model estimate $\hat{G}_t$
4. Estimate with u(t) and $y_n(t)$ an ARX model $\hat{G}_d(z)$ directly in the time domain, with McMillan degree 8.

Figure 5.5.3 shows next to the deterministic output and the additive noise, the Laguerre coefficients Y_k, where 1100 samples of $y_n(t)$ were used to create 500 coefficients. Coefficients Y_k with k>500 are negligible, which shows that the transformation leads to a considerable data reduction.

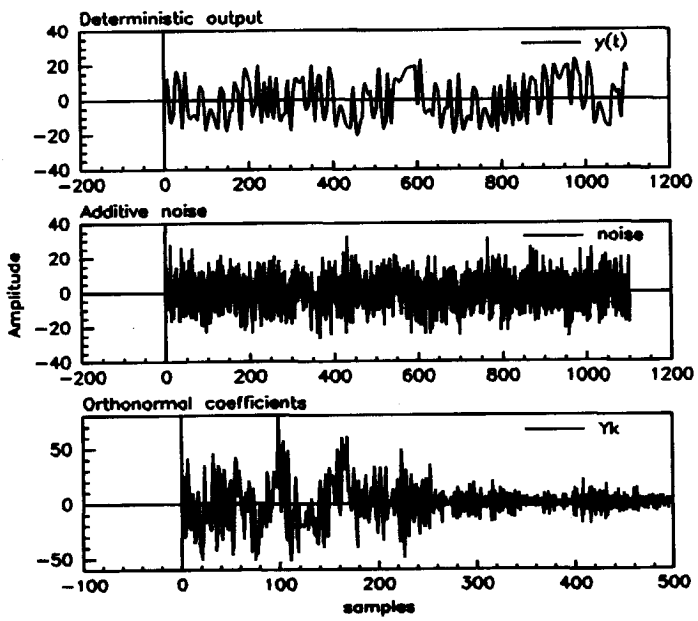


Fig. 5.5.3 Simulated output y(t) of the system, additive noise n(t) and the transformed (corrupted) output.

In Fig. 5.5.4 the step responses and the Bode amplitudes are depicted of the original system G(z) and the approximations $\hat{G}_t(z)$, $\hat{G}_d(z)$. As to be expected the first 8 Markov parameters of $\hat{G}_d(z)$ fit the original Markov parameters (Swaanenburg and co–workers, 1985; Van den Hof and Janssen, 1987), but the Bode amplitude plot shows that this model is only satisfactory for the very high frequencies and that the orthonormal method gives a much better approximation over the whole frequency range.

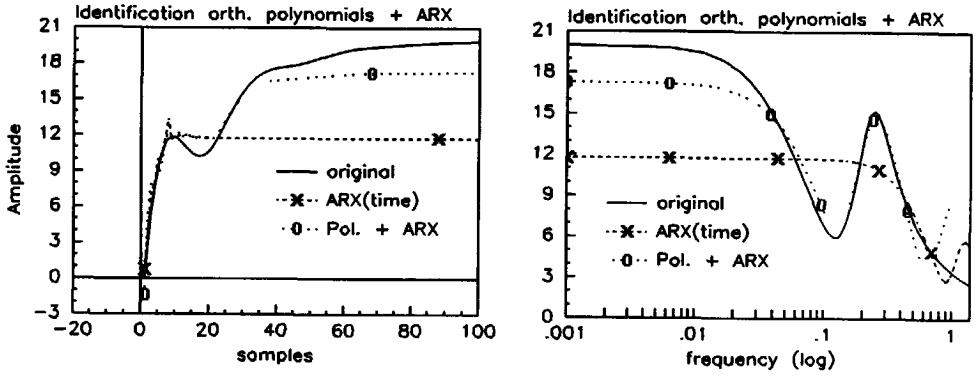


Fig. 5.5.4 Step responses and Bode amplitude plots of the original system G(z), the (time) ARX model $\hat{G}_d(z)$ and the (orthonormal) ARX model $\hat{G}_t(z)$

Method 2 FIR in the orthonormal domain

For this method we used the same data as for method 1, so u(t) and $y_n(t)$.
We estimated 50 Markov parameters, from which a model $\hat{G}_m(z)$ was created with an approximate realization method with McMillan degree 13 (Van Helmont and co–workers, 1990). This large order is the result of the large amount of noise on the data, which leads to a large variance in the estimated parameters. For the generation of the orthonormal functions we used the result of method 1, this is $G_t(z)$. We estimated 15 parameters and realized a model $\hat{G}_i(z)$ with McMillan degree 5. Note that both models are created with realization methods. The resulting responses are shown in Fig. 5.5.5.
Since the data are produced in an output error setting it is to be expected that an output error model like the FIR–model $G_m(z)$ will be a better approximation than the ARX model $G_d(z)$. Comparison of Fig. 5.5.4 and 5.5.6 shows that this is indeed the case. The result $\hat{G}_i(z)$ is clearly superior to $\hat{G}_m(z)$ and it is slightly better then $\hat{G}_t(z)$.

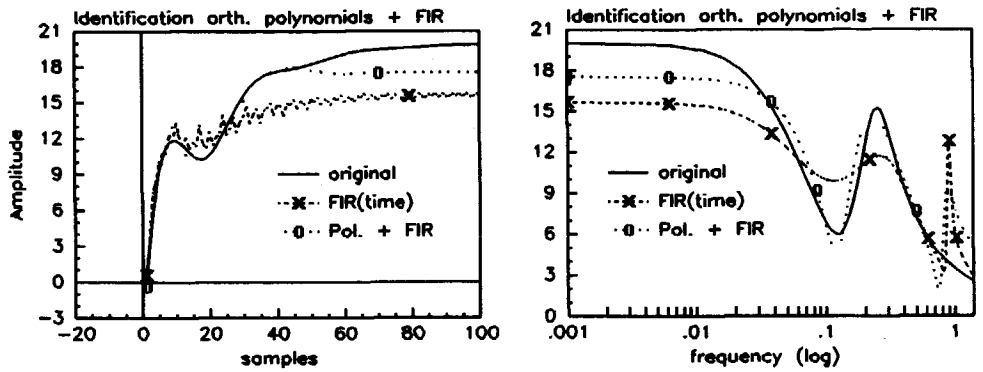


Fig. 5.5.5 Step responses and Bode amplitude plots of the original system G(z), the (time) FIR model $\hat{G}_m(z)$ and the (orthonormal) FIR model $\hat{G}_i(z)$

EXAMPLE 2

As a second example we consider a SISO system, which is the (1,1) part of the multivariable system, which was used in examples 4.6.4 and 4.6.6. The solid lines in Fig. 5.5.6a depict respectively the step response and the Bode amplitude of the system. We use this system to show the result of an iterative estimation in the orthonormal domain. We excited this 6^{th} order system with PRBS (u) and we corrupted the output y with additive white noise n, such that the signal to noise ration is 3 dB. We used 400 samples of the input u and padded this signal with 100 zeros to ensure that the signals u and y tend to zero.

Method 1. ARX in the orthonormal domain

With the data (500 samples) u and y+n we first estimated the following approximative models with McMillan degree 4:

1. a model G_t with ARX directly in the time domain
2. a model G_P, based on ARX with prefiltered data, where the prefilter P is derived from the system $G_t=P^{-1}Q$.
3. a model G_o^1 with ARX in the orthonormal domain (section 5.5.2), where the orthonormal functions are derived from G_t
4. an updated model G_o^2 with ARX in the orthonormal domain, where the orthonormal functions are derived from G_o^1.

Hence we iterate as follows: $G_t \rightarrow G_o^1 \rightarrow G_o^2$. In Fig. 5.5.6a the step responses and Bode amplitude of the original system G, G_t and G_p are shown. It is apparent that the prefiltering improves the poor estimation G_t considerably. In Fig. 5.5.6b the same responses are shown of the original system and the orthonormal estimates G_o^1 and G_o^2. The first orthonormal estimate G_o^1 is better than the prefilter estimate G_p for the lower frequencies. The update G_o^2 improves the estimation further for the low frequencies.

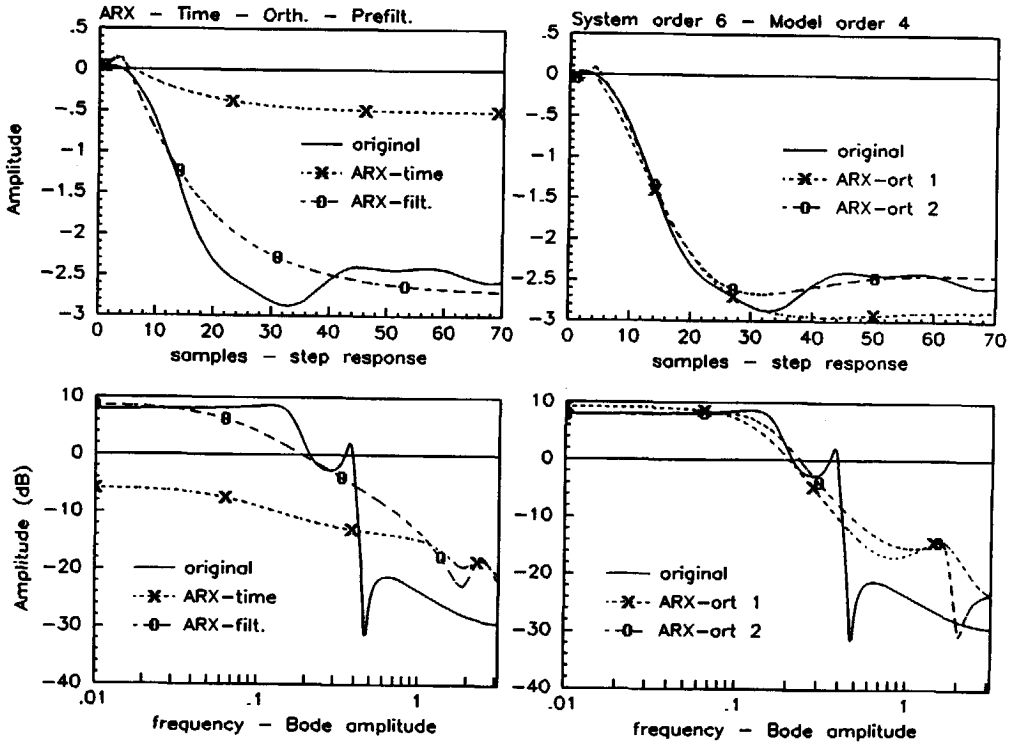


Fig. 5.5.6a Step responses and Bode amplitudes of the systems G, G_t and G_p.

Fig. 5.5.6b Step responses and Bode amplitudes of the systems G, G_o^1 and G_o^2.

example 2

Method 2. FIR in the orthonormal domain

With the same data we estimated models, based on FIR in the orthonormal domain (section 5.5.3). We estimated 12 coefficients of the expansion in generating transfer functions and realized models with McMillan degree 4. This realization consisted in creation of an exact 12^{th} order state space model and model reduction with $\mathcal{TSB}$ (definition 3.3.7).

As initial system to generate orthonormal polynomials we used the system $\frac{1}{z}$, hence in the first step we estimated 13 Markov parameters. We used the resulting model estimates to generate new orthonormal functions. In Fig. 5.5.7 the step responses and Bode amplitudes are depicted of the first three models that were created in this iterative setup. Clearly this procedure shows a convergence towards a fair estimate.

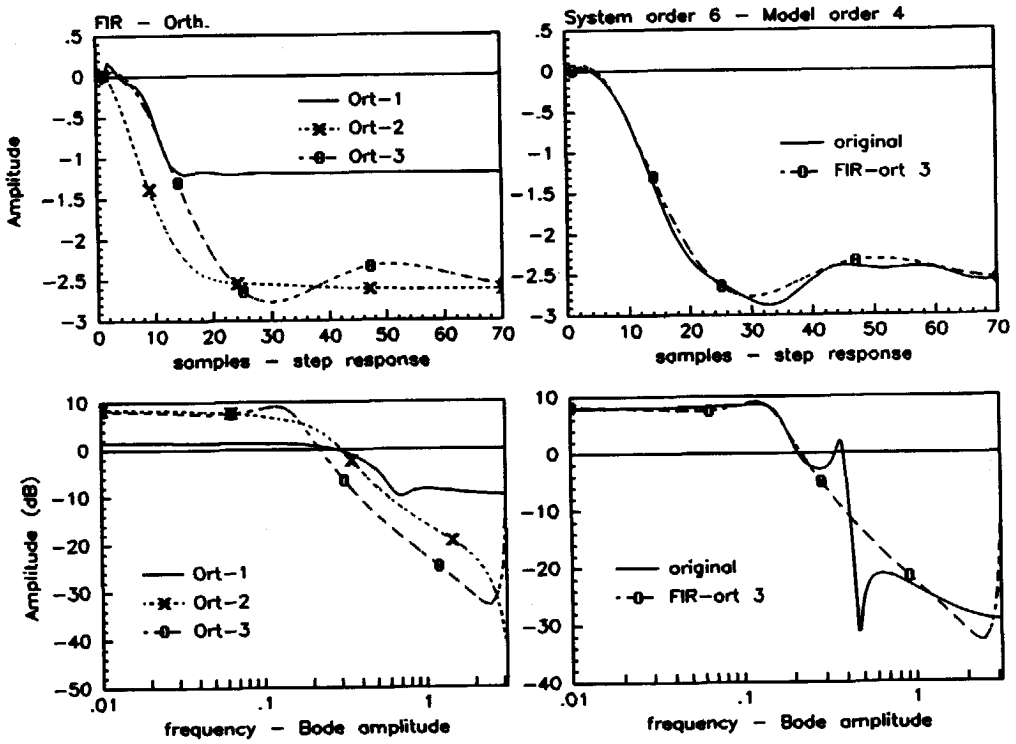


Fig. 5.5.7a Step response and Bode amplitude of the approximations based on FIR in the orthonormal domain.

Fig. 5.5.7b Step response and Bode amplitude of the systems G and the third orthonormal FIR approximation

example 2

These examples illustrate that the methods that we proposed can successfully be applied and that they can improve the result of their classical counterparts considerably. The ARX method improves the approximation over a large frequency scale and the FIR method improves the behavior for especially the lower frequencies and requires the estimation of less parameters than the classical FIR method.

5.6 DISCUSSION

In this chapter we presented a theory on what is referred to as system based orthonormal functions. It was shown that every FDLTS system gives rise to two sets of such functions that can be considered as a completion of the matrices with singular vectors of the Hankel matrix of the system. An important property of these functions is the recursive structure, expressed by the so–called shift property and the representation with generating transfer functions. Specific cases of these functions are the Laguerre polynomials and the discrete pulse functions, which shows that the presented procedure generalizes the concept of these well known orthonormal bases.
Similar to the Laguerre case we applied the generalized orthonormal functions in a transformation and approximation context. Expansion of ℓ_2 functions in the new basis leads to new ℓ_2 functions in the so called orthonormal domain. If this transformation is applied to the elements of the restricted ℓ_2 behavior of a system then the transformed signals are elements of such a behavior of a new FDLTS system in the orthonormal domain. Another application of the orthonormal functions on system level is the expansion of a transfer function in terms of the generating transfer functions. It was shown that this expansion is a generalization of the Laurent expansion or the Markov parameter description of a transfer function. Both the transformation issue and the expansion concept have been discussed in detail in section 5.4.5, where also the unsolved problems and open questions are listed. The most important property is the loss of dynamics in the orthonormal domain if the set of orthonormal functions are generated by the system itself.
Both applications of orthonormal function on system level have in section 5.5 been used for the definition of alternative approximate identification methods. These methods are based on the use of simple model structures (ARX and FIR) in combination with a least squares criterion. This has the merit that the resulting estimation problem has a straightforward solution, under the condition that the input is sufficiently exciting. It was made plausible that the alternative ARX method is equivalent with the use of classical equation error methods in combination with prefilters. The generalized FIR method is based on an approximation of all the Markov parameters of a system, instead of the estimation of a finite number of these Markov parameters, which is the objective

of a classical FIR method. This feature suggests that the alternative method will result in better estimates of low frequent components of systems.
We showed that both new methods will perform better if the set of orthonormal functions is based on a system whose behavior is close to the actual dynamic behavior contained in the measurement data. This implies that these methods should be used in an iterative setup, updating the functions with new model estimates. We note that many other applications in an identification setting are possible. For the methods, that we considered, further study is required on the optimal experiment design and on the analysis of the influence of various types of disturbances.

In the discussion of section 5.4.5 a number of open problem was mentioned and several fundamental question were raised, concerning the relation between model reduction methods and orthogonal functions. Answers to these questions might lead to other alternative approaches to system approximation in general.

5.7 APPENDICES OF CHAPTER 5

Appendix 5A Proof of lemma 5.2.2

First we have to show that $[A_2,B_2,C_2,D_2]$ is indeed a realization of $G^2(z)$:

$$C_2[zI-A_2]^{-1}B_2 + D_2 = [DC\ C]\begin{bmatrix} [zI-A]^{-1} & 0 \\ [zI-A]^{-1}BC[zI-A]^{-1} & [zI-A]^{-1} \end{bmatrix}\begin{bmatrix} B \\ BD \end{bmatrix} + D^2$$

$$= \{DC + C[zI-A]^{-1}BC\}\,[zI-A]^{-1}B + C[zI-A]^{-1}BD + D^2$$

$$= \{D + C[zI-A]^{-1}B\}\,C[zI-A]^{-1}B + \{C[zI-A]^{-1}B + D\}\,D$$

$$= \{D + C[zI-A]^{-1}B\}\,\{C[zI-A]^{-1}B + D\} = G(z)G(z) = G^2(z).$$

Since A is stable this shows that A_2 and thus G_2 is stable. To proof that this realization is again balanced we evaluate: $A_2A_2^* + B_2B_2^*$

$$= \begin{bmatrix} AA^*+BB^* & AC^*B^*+BD^*B^* \\ BCA^*+BDB^* & BCC^*B+AA^*+BDD^*B^* \end{bmatrix} = \begin{bmatrix} AA^*+BB^* & (AC^*+BD^*)B^* \\ B(CA^*+DB^*) & B(CC+DD^*)B^*+AA^* \end{bmatrix}$$

now apply theorem 2.7.4(2) with $P=I, \sigma=1$: $AC^*+BD^*=0$ and $CC^*+DD^*=I$. Substitution in the above equation yields: $A_2A_2^* + B_2B_2^* = I_{2n}$.

Analogous calculation shows that $A_2^*A_2 + C_2^*C_2 = I_{2n}$ and hence we showed that $[A_2,B_2,C_2,D_2]$ is balanced w.r.t. I_{2n} □

Appendix 5B Proof of theorem 5.2.3.

We use induction on k to proof this. Lemma 5.2.2 shows the correctness for k=2. Now suppose the theorem is valid for k=n–1 and consider the case k=n Notice that we can write:

$$A_k=\begin{bmatrix} A_{k-1} & 0 \\ BC_{k-1} & A \end{bmatrix} \quad B_k=\begin{bmatrix} B_{k-1} \\ BD^{k-1} \end{bmatrix} \quad C_k=[DC_{k-1}\ C]$$

First we show that $[A_k,B_k,C_k,D_k]$ is indeed a realization of $G^k(z)$:

$$C_k[zI-A_k]^{-1}B_k+D_k = [DC_{k-1}\ C]\begin{bmatrix} [zI-A_{k-1}]^{-1} & 0 \\ [zI-A]^{-1}BC_{k-1}[zI-A_{k-1}]^{-1} & [zI-A]^{-1} \end{bmatrix}\begin{bmatrix} B_{k-1} \\ BD^{k-1} \end{bmatrix} + D^k$$

$$= \{DC_{k-1} + C[zI-A]^{-1}BC_{k-1}\}\,[zI-A_{k-1}]^{-1}B_{k-1} + C[zI-A]^{-1}BD^{k-1} + D^k$$

$$= \{D + C[zI-A]^{-1}B\}\,C_{k-1}[zI-A_{k-1}]^{-1}B_{k-1} + \{C[zI-A]^{-1}B + D\}\,D^{k-1}$$

$$= \{D + C[zI-A]^{-1}B\}\ \{C_{k-1}[zI-A_{k-1}]^{-1}B_{k-1} + D_{k-1}\} = G(z)G^{k-1}(z) = G^k(z).$$

To proof the balancedness we consider $A_kA_k^*+B_kB_k^*=$

$$= \begin{bmatrix} A_{k-1}A_{k-1}^*+B_{k-1}B_{k-1}^* & B(A_{k-1}C_{k-1}^*+D_{k-1}D_{k-1}^*)B^* \\ B(C_{k-1}A_{k-1}^*+D_{k-1}D_{k-1}^*)B^* & B(C_{k-1}C_{k-1}^*+D_{k-1}D_{k-1}^*)B^* + AA^* \end{bmatrix}$$

Apply theorem 2.7.4(2): $A_{k-1}C_{k-1}^*+B_{k-1}D_{k-1}^*=0$ and $C_{k-1}C_{k-1}^*+D_{k-1}D_{k-1}^*=I$

Substitution shows that $A_kA_k^*+B_kB_k^* = I$. Analogous calculations shows that $A_k^*A_k + C_k^*C_k = I$ and hence $[A_k,B_k,C_k,D_k]$ is balanced w.r.t. I_{kn} □

Appendix 5C Proof of lemma 5.3.1

In the proof we omit the index $_a$ for U_a etc.

1. $A=USV^* \Rightarrow AA^*=US^2U^*$ and $A^*A=VS^2V^*$. $AA^*+BB^*=I$ so $BB^*=I-AA^*$.
   $\tilde{C}^*\tilde{C} = F^*BB^*F = F^*\left[I-AA^*\right]F = VU^*\left[I-US^2U^*\right]UV^* = VV^*-VS^2V^* = I-A^*A$
2. i: First we assume that B has full column rank m. From (1.) and theorem 2.7.4 we now conclude there exists a $\tilde{D}$ such that $\tilde{G}(z) = \tilde{D} + \tilde{C}[zI-A]^{-1}B$ is square all-pass. We can even choose $\tilde{D}$ Hermitian because the conditions of theorem 2.7.4(2) yield:
   $\tilde{C}^*\tilde{D}= -A^*B= -VSU^*B= -VU^*USV^*VU^*B= -F^*A^*F^*B= -F^*A\tilde{C}^*= F^*B\tilde{D}^*= \tilde{C}^*\tilde{D}^*$
   Now $\text{rank}(C^*)=\text{rank}(B)=m\leq n$ or C^* is left invertible and hence we have $\tilde{D}=\tilde{D}^*$. Furthermore we have from the previous calculation:
   $$\tilde{D}\tilde{C} = \tilde{D}^*\tilde{C} = -\tilde{C}A^*F \quad \text{and} \quad B\tilde{D} = B\tilde{D}^* = -A\tilde{C}^* = -AF^*B = -FA^*B$$
2. ii: Now assume that $\text{rank}(B)<m$. Let $B=[B_1\ 0]$, with B_1 full rank, then $\tilde{C}^*=[\tilde{C}_1^*\ 0]$ and $\tilde{C}[zI-A]^{-1}B = \begin{bmatrix}\tilde{C}_1[zI-A]^{-1}B_1 & 0\\ 0 & 0\end{bmatrix}$. Let $\tilde{D}:= \begin{bmatrix}\tilde{D}_1 & 0\\ 0 & I\end{bmatrix}$ where $\tilde{D}_1$ is Hermitian as in (2i). It is immediate that (5.3.6) remains valid. Now observe that $\text{rank}(B)<m \Rightarrow \exists$ a unitary matrix T such that $BT=[B_1\ 0]$. Apply the previous on (A,BT).
   Postmultiplying the resulting $\tilde{C}$ and $\tilde{D}$ with T^* gives the desired result. □

Appendix 5D Proof of proposition 5.3.5

Denote $M(z):= [zI-A]^{-1}F[zI-A^*]$, so $H_k(z)=zM^k(z)[zI-A]^{-1}B$. We have to show that

$$z[zI-A_e]^{-1}B_e = \begin{bmatrix}H_0(z)\\ H_1(z)\\ \vdots\end{bmatrix} = z\begin{bmatrix}M^0(z)\\ M^1(z)\\ \vdots\end{bmatrix}[zI-A]^{-1}B$$

or equivalently

$$\begin{bmatrix}P^0\\ P^1\\ \vdots\end{bmatrix}B = B_e = [zI-A_e]\begin{bmatrix}M^0(z)\\ M^1(z)\\ \vdots\end{bmatrix}[zI-A]^{-1}B \qquad (5D.1)$$

For this purpose we use the property: $(\ -X+[zI-A]M(z)\ )[zI-A]^{-1} = P$ (5D.2)

(Proof: $-X+[zI-A]M(z) = F(-I+A^*A+zI-A^*) = -FA^*(zI-A)$)

We proof (5D.1) with induction on the number k of the block rows.

k=0: $[zI-A][zI-A]^{-1}B = B$ is correct.

k=1: $(\ -X + [zI-A]M(z)\ )[zI-A]^{-1}B = PB$ because of (5D.2).

Suppose the hypothesis is correct for the $(k+1)^{\text{th}}$ block row:

$$(\ -P^{k-1}X\ -P^{k-2}XM(z)\cdots -XM^{k-1}(z) +[zI-A]M^k(z)\ )\ [zI-A]^{-1}B = P^kB \qquad (5D.3)$$

and consider the next block row:

$$\begin{aligned}&(\ -P^kX\ -P^{k-1}XM(z)\ \cdots\ -PXM^{k-1}(z)\ -XM^k(z)\ +[zI-A]M^{k+1}(z)\ )\ [zI-A]^{-1}B\\ &= (\ -P^kX\ -P^{k-1}XM(z)\cdots\ -PXM^{k-1}(z)\ +P[zI-A]M^k(z)\ )\ [zI-A]^{-1}B \quad \text{(use (5D.2))}\\ &= P\cdot P^kB = P^{k+1}B\ , \quad \text{because of (5D.3).}\end{aligned}$$

This shows the correctness of (5D.1) for all block rows and completes the proof. □

Appendix 5E Proof of lemma 5.3.8.

1. $\tilde{A}\tilde{A}^* + \tilde{B}\tilde{B}^* = T(AA^* + BB^*)T^* = TT^* = I$.

2. We follow the construction of Ψ_e as defined by theorem 5.3.3. From the singular value decomposition of the matrix A and $\tilde{A}$ we derive the matrices F,P,X and $\tilde{F},\tilde{P},\tilde{X}$. If A has an svd $A=U_aS_aV_a^*$ then it is immediate that $\tilde{A}$ has an svd $\tilde{A}=(TU_a)S_a(TV_a)^*$. We use this decomposition to define $\tilde{F},\tilde{P},\tilde{X}$ and to express these in F,P,X. This is done by evaluation of the formulas (5.3.7):

$$\tilde{F} = (TU_a)(TV_a)^* = TU_aV_a^*T^* = TFT^*$$
$$\tilde{P} = -\tilde{F}\tilde{A}^* = -TFT^*TA^*T^* = -TFA^*T^* = -TPT^*$$
$$\tilde{X} = \tilde{F}(I-\tilde{A}^*\tilde{A}) = TFT^*(I-TA^*T^*TAT^*) = TF(I-A^*A)T^* = TXT^*$$

This shows that $\tilde{P}^k\tilde{X} = TP^kXT^*$ and $\tilde{P}^k\tilde{B} = TP^kB$ $k\geq 0$. If we combine this with $\tilde{A}=TAT^*$ and $\tilde{B}=TB$ it follows that $\tilde{A}_e=T_eA_eT_e^*$, $\tilde{B}_e=T_eB_e$.

Now it is obvious that the unitarity of T implies that T_e is unitary and thus we have:

$$[\tilde{B}_e\ \tilde{A}_e\tilde{B}_e\cdots] = T_e[B_e\ A_eB_e\cdots] \quad \text{or} \quad \Psi_e\{\tilde{A},\tilde{B}\} = T_e\Psi_e\{A,B\}$$ □

Appendix 5F proof of lemma 5.3.9.

1. Let A have svd $A=USV^*$ and let $T=U^*$ then $TAT^*=I_nS(UV)^*$.
2. $AA^*+BB^*=I$, $AA^*=S^2 \Rightarrow S^2+BB^*=I \Rightarrow BB^*$ is diagonal with rank m. This proves that at least (n–m) elements of S are equal to 1 ($S\geq 0$ and diagonal).
3. $S^2+BB^*=I$ implies $S\leq I$. From (2) and the ordering of S it follows that $S=\begin{bmatrix} I & 0 \\ 0 & S_a \end{bmatrix}$ and $BB^*=\begin{bmatrix} 0 \\ B_1B_1^* \end{bmatrix}$, with $S_a,B_1B_1^*\in\mathbb{C}^{m\times m}$, diagonal and $S_a^2+B_1B_1^* =I$. B has rank m $\Rightarrow$ $B_1B_1^*$ has full rank $\Rightarrow$ $(B_1B_1^*)_{ii}>0 \Rightarrow S_a<I$. B now must have the form $B=\begin{bmatrix} 0 \\ B_1 \end{bmatrix}$ and since $B_1B_1^*$ is diagonal B_1 must have a svd $B_1=IS_bV_b^*$. $S_a^2+B_1B_1^*=I$ then implies $S_b^2=I-S_a^2$. □

Appendix 5G proof of theorem 5.3.10

First note that every function $\psi_k\{A,B\}$ k=0,1,..is an ℓ_2 function, since proposition 5.3.5 shows it is the response of a stable linear system on a ℓ_2 input signal. From definition 5.3.4 we deduct that without loss of generality we may assume that $AA^*+BB^*=I$. Let A have svd $A=USV^*$ and apply lemma 5.3.8 with $T=U^*$. This gives new matrices $\{\tilde{A},\tilde{B}\}$ with $\tilde{A}=ISV^*U=ISV_1^*$, $V_1^*V_1=I$. Lemma 5.3.8 shows that without loss of generality we thus can assume that {A,B} already is in this form: $A=ISV^*$. Lemma 5.3.9 then gives information on the structure of S and B.

Define F,X,P for {A,B} as in (5.3.7) and note that P=–S. Since ℓ_2 is an Hilbert space it is necessary and sufficient (Kreyszig, 1978) for the proof of the completeness of $\Psi_e\{A,B\}$ that the following equivalence relation holds:

$$Y = [y(0),y(1),\cdots] = 0 \;\Leftrightarrow\; \sum_{t=0}^{\infty} \psi_k(t)y(t) = 0 \quad \forall k \geq 0$$

or
$$Y = 0 \Leftrightarrow [B_e | A_eB_e | A_e^2B_e \cdots]Y^* = 0 \qquad (5G.1)$$

It is clear we only have to prove "$\Leftarrow$" and this will be proven if we can show that $\Psi_e^*\{A,B\}\Psi_e\{A,B\}=I_e$, for in that case we have: $\Psi_eX^*=0 \Rightarrow \Psi_e^*\Psi_eX^*=0 \Rightarrow X^*=0$.
We will proof this in 3 steps: 1: $B_e^*B_e=I$, 2:$A_e^*B_e=0$, 3:$A_e^*A_e=I_e$.

Note : $A=SV^*$ $F=V^*$ $P=-S$ $X=(I-S^2)V^*$

Step 1: $B_e^*B_e=I$.

Consider the left upper mxm block matrix of $\Psi_e^*\Psi_e$. This matrix is equal to (use (5.3.7d) and lemma 5.3.9 and $B^*=[0\ b^*]$):

$$B_e^*B_e = \sum_{i=0}^{\infty} B(P^*)^iP^iB \overset{P=-S}{=} \sum_{i=0}^{\infty} BS^{2i}B = \sum_{i=0}^{\infty} [0\ b^*]\begin{bmatrix} I \\ & S_a \end{bmatrix}^{2i}\begin{bmatrix} 0 \\ b \end{bmatrix} = \sum_{i=0}^{\infty} b^*S_a^{2i}b \qquad (5G.2)$$

Since $S_a<I$ and diagonal (5G.2) is equivalent with $B_e^*B_e = b^*[I-S_a^2]^{-1}b$. Recall from lemma 5.3.9 that $b=S_bV_b^*$ with $S_b^2=I-S_a^2$ and conclude that $B_e^*B_e=I$. (step 1)□

Step 2: $A_e^*B_e=0$.

This can be proven as step 1, but an alternative proof uses step 1 and $A_eA_e^*+B_eB_e^*=I$:
$(B_e^*A_e)(A_e^*B_e) = B_e^*(I_e-B_eB_e^*)B_e = B_e^*B_e - (B_e^*B_e)(B_e^*B_e) = I-I = 0$ (step 2)□

Step 3: $A_e^*A_e=I$.

First consider the diagonal blocks of this product. They have all the same form:

$$A^*A+\sum_{i=0}^{\infty}(P^iX)^*P^iX = A^*A + \sum_{i=0}^{\infty}X^*P^{2i}X = VS^2V^* + \sum_{i=0}^{\infty}V[I-S^2]S^{2i}[I-S^2]V^*$$

$$= V\begin{bmatrix} I \\ & S_a^2 \end{bmatrix}V^* + V\left[\sum_{i=0}^{\infty}\begin{bmatrix} 0 \\ & I-S_a^2 \end{bmatrix}\begin{bmatrix} I \\ & S_a^2 \end{bmatrix}\begin{bmatrix} 0 \\ & I-S_a^2 \end{bmatrix}\right]V^* = V\begin{bmatrix} I \\ & S_a^2 \end{bmatrix}V^* + V\begin{bmatrix} 0 \\ & I-S_a^2 \end{bmatrix}V^* = I.$$

Next we consider the (i,j) block of the product with i<j. This has the form:

$$(P^{j-i-1}X)^*A+(P^{j-i}X)^*X+(P^{j-i+1}X)^*PX+\cdots = X^*P^{j-i-1}\left[A+PX+P^3X+\cdots\right] \qquad (5G.3)$$

Evaluate $\sum_{i=0}^{\infty}P^{2i}X = \sum_{i=0}^{\infty}S^{2i}[I-S^2]V^* = \begin{bmatrix} 0 \\ I \end{bmatrix}V^*$ with the same argument as before, and substitute this in (5G.3). This shows that the (i,j)–block is

$$X^*P^{j-i-1}\left[A + P\begin{bmatrix} 0 \\ I \end{bmatrix}V^*\right] = X^*(-S)^{j-i-1}S\left[\begin{bmatrix} I \\ & I \end{bmatrix}-\begin{bmatrix} 0 \\ & I \end{bmatrix}\right]V^*$$

$$= -X^*(-S)^{j-i}\begin{bmatrix} I \\ 0 \end{bmatrix}V^* = -V[I-S^2]\begin{bmatrix} (-I)^{j-i} \\ & 0 \end{bmatrix}V^* = -V\begin{bmatrix} 0 \\ & I-S_a^2 \end{bmatrix}\begin{bmatrix} (-I)^{j-i} \\ & 0 \end{bmatrix}V^* = 0.$$

Since $A_e^*A_e$ is symmetric we get that the (i,j) blocks for i>j are also zero and hence we can conclude that $A_e^*A_e = I_e$. (step 3)□

Now we combine these results and calculate $\Psi_e^*\Psi_e = \begin{bmatrix} B_e^* \\ B_e^*A_e^* \\ \vdots \end{bmatrix}\left[B_e | A_eB_e | A_e^2B_e \cdots\right]$

The (i,j)-block of this product is: $B_e^*(A_e^*)^{i-1}(A_e)^{j-1}B_e$ (5G.4)
If i=j we find from step 3 that (5G.4) is $B_e^*B_e$ and step 1 shows this is I.
If i>j we find from step 3 that (5G.4) is $B_e^*(A_e^*)^{i-j}B_e$ and step 2 shows this is 0.
If i<j then symmetry shows that (5G.4) is again 0.
We may therefore conclude that $\Psi_e^*\Psi_e=I_e$. and as stated before this proofs that $\Psi_e\{A,B\}$ forms a complete orthonormal basis for ℓ_2. (theorem 5.3.10)□

Appendix 5H proof of theorem 5.4.8

Before we proof the theorem we need the following, purely technical lemma.

LEMMA 5H.1. Let A,F,P,X,U,S,V be as in theorem 5.3.3 and define $Q=A^*F=VSV^*=F^*A$.
Then: 1. $XQ=FA^*X=-PX$. 2. $X+AQ=F$ ○

PROOF:
1. $XQ= UV^*[I-VS^2V^*]VSV^*= USV^*[I-VS^2V^*]= UV^*VSU^*UV^*[I-VS^2V^*]= FA^*X$
2. $X+AQ= UV^*[I-VS^2V^*] + USV^*VSV^*= UV^*-US^2V^* + US^2V^*= UV^*= F$ □

So Q plays a similar role as $P=-USU^*$.

Proof of theorem 5.4.8:
1. '⇒' Let $(u,y)\in\mathcal{B}_2^+(W_t(z)$ and let x be the state proces such that:

$$x(t+1) = Rx(t)+Gu(t) \tag{5H.1}$$
$$y(t) = Hx(t)+Lu(t) \tag{5H.2}$$

We denote the orthonormal coefficients of u(t),y(t),x(t) by U_k,Y_k,X_k k=0,1,...
Multiply all time functions in (5H.1,2) with $\varphi_k^*(t)$ and sum over all $t\in[0,\infty]$

$$\hat{X}_k = R\,X_k + G\,U_k \tag{5H.3}$$
$$Y_k = H\,X_k + L\,U_k \tag{5H.4}$$
$$\text{with}\quad \hat{X}_k := \sum_{t=0}^{\infty} x(t+1)\varphi_k^*(t) \tag{5H.5}$$

From proposition 5.4.3 $$x(t+1) = \sum_{k=0}^{\infty} X_k\varphi_k(t+1) \tag{5H.6}$$

(5H.6) in (5H.5) gives $$\hat{X}_k = \sum_{t=0}^{\infty}\left[\sum_{p=0}^{\infty} X_p\varphi_p(t+1)\right]\varphi_k^*(t) \tag{5H.7}$$

The restriction that B has only one column shows that the shift structure leads to

$$\varphi_p(t+1) = A\varphi_p(t) + \sum_{j=1}^{p-1}(-FA^*)^{p-j-1}X\varphi_j(t) \tag{5H.8}$$

(5H.8) in (5H.7) gives $$\hat{X}_k = X_kA + \sum_{j=1}^{\infty} X_{k+j}(-FA^*)^{j-1}X \tag{5H.9}$$

where we used the orthonormality of $\Phi_e\{A,B\}$.

(5H.9) in (5H.3) $$\sum_{j=1}^{\infty} X_{k+j}(-FA^*)^{j-1}X = -X_kA+RX_k+GU_k \tag{5H.10}$$

Introduce $Q=F^*A$ as in lemma 5H.1 and rewrite (5H.10) with $k\to k+1$ and multiply both sides with Q:

$$\sum_{j=1}^{\infty} X_{k+j+1}(-FA^*)^{j-1}XQ = -X_{k+1}AQ+RX_{k+1}Q+GU_{k+1}Q \qquad (5H.11)$$

lemma 5H.1 gives

$$-\sum_{j=1}^{\infty} X_{k+j+1}(-FA^*)^{j}X = -X_{k+1}AQ+RX_{k+1}Q+GU_{k+1}Q \qquad (5H.12)$$

rewrite the index

$$-\sum_{j=2}^{\infty} X_{k+j}(-FA^*)^{j-1}X = -X_{k+1}AQ+RX_{k+1}Q+GU_{k+1}Q \qquad (5H.13)$$

Now we get rid of the infinite sums by adding (5H.10) and (5H.13):

$$X_{k+1}X = -X_kA + RX_k + GU_k -X_{k+1}AQ + RX_{k+1}Q + GU_{k+1}Q \qquad (5H.14)$$

Move the (k+1)–terms to the left and apply lemma 5G.1(2):

$$X_{k+1}F - RX_{k+1}Q + GU_{k+1}Q = -X_kA + RX_k + GU_k \qquad (5H.15)$$

F is a unitary matrix and $QF^* = A^*$ so multiplication of (5H.15) with F^* gives

$$X_{k+1} - RX_{k+1}A^* - GU_{k+1}A^* = -X_kAF^* + RX_kF^* + GU_kF^* \qquad (5H.16)$$

At this point we introduce Kronecker notation (section 2.9) to rewrite (5H.16):

$$(I\otimes I-A\otimes R)\mathrm{Vec}(X_{k+1}) - (A\otimes G)\mathrm{Vec}(U_{k+1}) =$$
$$= (-FA^*\otimes I+F\otimes R)\mathrm{Vec}(X_k) + (F\otimes G)\mathrm{Vec}(U_{k+1}) \qquad (5H.17)$$

or

$$M\cdot\mathrm{Vec}(X_{k+1}) + V\cdot\mathrm{Vec}(U_{k+1}) = N\cdot\mathrm{Vec}(X_k) + W\cdot\mathrm{Vec}(U_{k+1}) \qquad (5H.18)$$

This leads to a linear system if we define the state vector Z_k as:

$$Z_k = M\cdot\mathrm{Vec}(X_k) + V\cdot\mathrm{Vec}(U_k) \qquad (5H.19)$$

Under the restriction that M is non–singular we can substitute (5H.19) in (5H.18):

$$Z_{k+1} = NM^{-1}Z_k + (W-NM^{-1}V)\mathrm{Vec}(U_k) \qquad (5H.20)$$

M = (I⊗I–A⊗R)is indeed non–singular, because proposition 2.9.2 shows this matrix has eigenvalues $1-\lambda_i\mu_j$, where λ_i, μ_j are the eigenvalues of A respectively R. Since both matrices are stable we have that $|1-\lambda_i\mu_j|>0$.

For the output equation we transform (5H.4) to Kronecker notation:

$$\mathrm{Vec}(Y_k) = (I\otimes H)\mathrm{Vec}(X_k) + (I\otimes L)\mathrm{Vec}(U_k) \qquad (5H.21)$$

or

$$\mathrm{Vec}(Y_k)= (I\otimes H)M^{-1}Z_k + [I\otimes L-(I\otimes H)M^{-1}V]\mathrm{Vec}(U_k) \qquad (5H.22)$$

In order to complete part b we have to define the initial state for the transformed system. We use the definition of X_k,U_k and write:

$$RX_0 + GU_0 = R\sum_{t=0}^{\infty} x(t)\varphi_0^*(t) + G\sum_{t=0}^{\infty} u(t)\varphi_0^*(t) \qquad (5H.23)$$

Remember that $\varphi_0(j) = A^jB$ and multiply (5H.23) with A^*

$$RX_0A^* + GU_0A^* = R\sum_{t=0}^{\infty} x(t)\varphi_0^*(t+1) + G\sum_{t=0}^{\infty} u(t)\varphi_0^*(t+1) \qquad (5H.24)$$

(5H.1):

$$RX_0A^*+GU_0A^*= \sum_{t=0}^{\infty} x(t+1)\varphi_0^*(t+1) =X_0-x(0)\varphi_0^*(0) = X_0-x(0)B^* \qquad (5H.25)$$

(5H.19):

$$Z_0=M\cdot\mathrm{Vec}(X_0)+V\cdot\mathrm{Vec}(U_0) = (I\otimes I-A\otimes R)\mathrm{Vec}(X_0)-(A\otimes G)\mathrm{Vec}(U_0) \qquad (5H.26)$$

Now write (5H.25) in Kronecker form and substitute in (5H.26):

$$Z_0 = \mathrm{Vec}(x(0)B^*) = B \otimes x(0) \tag{5H.27}$$

With (5H.20,22,27) we have proven that $\{U_k,Y_k\}$ is an input/output pair of a FDLT system $W_o(z)$ with a realization (5.4.8). $(\Rightarrow)\square$

2. '$\Leftarrow$ ' Let ($\mathrm{Vec}(Y_k)$,$\mathrm{Vec}(U_k)$) $\in \mathcal{B}_2^+(W_o(z))$. We have to show that the corresponding signal y and u obey $(y,u)\in\mathcal{B}_2^+(W_t(z))$. Therefore we have to use a result that will be proven by lemma 5.4.9, that states that the realization $[R_o,G_o,H_o,L_o]$ has $n_w(n-1)$ uncontrollable poles in z=–1. Hence a state process Z_k exists, that relates $\mathrm{Vec}(Y_k)$ and,$\mathrm{Vec}(U_k)$ in the state space description and obeys:

$$Z_0 = B \otimes x_o \quad x_o \in \mathbb{R}^{n_w} \tag{5H.28}$$

Let (y,u) be the time signals and define the processes π and ω by:

$$\pi(t+1) = R\pi(t) + Gu(t) \tag{5H.29}$$

$$\omega(t) = H\pi(t) + Lu(t) \tag{5H.30}$$

$$\pi(0) = x_o \tag{5H.31}$$

Let $\{\Pi_k\}$,$\{\Omega_k\}$ be the sets of coefficients of π and ω and define the vector process

$$\tilde{Z}_k = M \cdot \mathrm{Vec}(\Pi_k) + V \cdot \mathrm{Vec}(U_k) \tag{5H.32}$$

As in formula (5H.20): $\tilde{Z}_{k+1} = R_o\tilde{Z}_k + H_o\mathrm{Vec}(U_k)$ (5H.33)

and the same goes for $\{Z_k\}$ $Z_{k+1} = R_oZ_k + H_o\mathrm{Vec}(U_k)$ (5H.34)

As in the previous part (5H.27) it folllows that:

$$\tilde{Z}_0 = B \otimes \pi(0) \tag{5H.35}$$

Combine (5H.35) and (5H.28): $\tilde{Z}_0 = Z_0$ (5H.36)

From the equations (5H.33,34,36) we can therefore conclude that :

$$\forall k \quad \tilde{Z}_k = Z_k \tag{5H.37}$$

The analog for the output process with (5H.22):

$$\mathrm{Vec}(\Omega_k) = H_o\tilde{Z}_k + L_o\mathrm{Vec}(U_k) \tag{5H.38}$$

$$\mathrm{Vec}(Y_k) = R_oZ_k + H_o\mathrm{Vec}(U_k) \tag{5H.39}$$

(5H.37,38,39) imply that $\Omega_k = Y_k$ (5H.40)

and we may conclude that $\omega = y$ (5H.41)

Substition of (5H.41) in (5H.30) shows with (5H.29) that $(y,u)\in\mathcal{B}_2^+(W_t(z))$.

This completes the proof of the theorem $\square$

Appendix 5I Proof of lemma 5.4.9.

$R_o = NM^{-1}$ with $M = I\otimes I - A\otimes R$ and $N = -FA^*\otimes I + F\otimes R$. Let $A=USV^*$ be the svd of A, so $F=UV^*$. Let $T=U^*\otimes I$ then T is unitary, $T^*=U\otimes I$.

$TMT^* = I\otimes I - (SV^*U)\otimes R$ and $TNT^* = -(V^*A^*U)\otimes I + V^*U\otimes R = -S\otimes I + V^*U\otimes R$

We recall from lemma 5.3.9 that S has only one ($B\in\mathbb{R}^{n\times 1}$) diagonal element unequal to 1.

The following then shows that $TNM^{-1}T^*= TNT^*(TMT^*)^{-1}=\begin{bmatrix}-I & 0\\ * & *\end{bmatrix}$. Therefore we consider the upper rows of $\begin{bmatrix}-I_x & 0\\ * & *\end{bmatrix}(TMT^*)= -I+([I_x\ 0]V^*U)\otimes R$ and conclude that these are exactly the upper rows of TNT^*. T is unitary so $TNM^{-1}T^*$ defines a similarity transformation and thus we may conclude that $TNM^{-1}T^*$ has $(n-1)n_w$ poles in $z=-1$.

$G_0= W-NM^{-1}V$ with $V=-A\otimes G$ and $W=F\otimes G$. Consider $TV=-(SV^*)\otimes G$ and $TW=V^*\otimes G$ then it is clear that the upper rows of $TV+TW$ are zero. Since $TNM^{-1}T^*=\begin{bmatrix}-I & 0\\ * & *\end{bmatrix}$ we know that the upper rows of $TNM^{-1}T^*TW$ are minus the upper rows of TW, or the upper rows of $TV-TNM^{-1}T^*TW$ are zero: $TG_0 = \begin{bmatrix}0\\ *\end{bmatrix}$.
This shows that the poles in $z=-1$ are not controllable.

As a last step we show that under this similarity transformation also the initial condition will not excite these poles. $Z_0 = B\otimes x(0)$, again we recall lemma 5.3.9, which states that under similarity transformation of $\{A,B\}$ with U^* we get that $U^*B=\begin{bmatrix}0\\ b\end{bmatrix}$, $TZ_0=TB\otimes x(0) = (U^*B)\otimes x(0)$ is therefore also of the form $\begin{bmatrix}0\\ b\otimes x(0)\end{bmatrix}$, which shows that the poles in $z=-1$ are never excited by the initial conditions. This concludes the proof of the uncontrollability of the poles in $z=-1$. □

Appendix 5J proof of lemma 5.4.10.

We first consider the controllability:

$z^*R_0=\lambda z^*$	$z^*G_0=0$
$z^*N=\lambda z^*M$	$z^*W= z^*R_0V = \lambda z^*V$
$z^*[-FA^*\otimes I + F\otimes R] = \lambda z^*[I\otimes I - A\otimes R]$	$z^*[F\otimes G + \lambda A\otimes G] = 0$
$z^*[(F+\lambda A)\otimes R] = z^*[(\lambda I+FA^*)\otimes I]$	$z^*[(F+\lambda A)\otimes G] = 0]$

Create a matrix Z from z such that $Vec(Z)=z$, then proposition 2.9.2 shows that:

$R^*Z(F+\lambda A) = Z(\lambda I+FA^*)$	$G^*Z(F+\lambda A)=0$

Observe that $(F+\lambda A) = (UV^*+\lambda USV^*)= U(I+\lambda S)V^*$ and that $(I+\lambda S)$ is invertible since $S\leq I$ by lemma 5.3.9 and $|\lambda|<1$. Thus $(F+\lambda A)$ is invertible.

$R^*Z = Z(\lambda I+FA^*)(F+\lambda A)^{-1}$	$G^*Z=0$

However, this latter equation contradicts with the controllability of (R,G) and hence $Z=0$ and $z=0$. In other words, (R_0,G_0) has no stable uncontrollable poles.

The argument to show the observability of stable poles follows the same line of reasoning and we sketch the main line:

$$R_o y=\lambda y \text{ and } H_o y=0 \Rightarrow NM^{-1}y=\lambda y\ ,\ (I\otimes H)M^{-1}y=0$$

let $z=M^{-1}y$ and $\mathrm{Vec}(Z)=z$

$\Rightarrow Nz=Mz$, $(I\otimes H)z=0 \Rightarrow RZ=Z((\lambda I+AF^*)(F^*+\lambda A^*)^{-1}$ and $HZ=0$

this shows again that the observability of (H,R) implies Z=0 and hence Y=0 so y=0. We may conclude that every stable pole is observable. □

Appendix 5K proof of lemma 5.4.11.

First we observe that if λ is an eigenvalue of R_o, then λ is also an eigenvalue of $R_1:=N_1M_1^{-1}=[-I\otimes FA^*+R\otimes F]\ [I\otimes I-R\otimes A]^{-1}$. This proven by the following argument:

Let $R_o z=\lambda z$ and $\mathrm{Vec}(Z)=z$ then we get:

$-ZAF^* + RZF^* = \lambda Z - \lambda RZA^* \Rightarrow -FA^*Z^* + FZ^*R^* = \lambda Z^* - \lambda AZ^*R^* \Rightarrow$

$[-I\otimes FA^*+ R\otimes F]\ \mathrm{Vec}(Z^*) = \lambda[I\otimes I - R\otimes A]\mathrm{Vec}(Z^*)$ or $N_1\mathrm{Vec}(Z^*)=\lambda M_1\ \mathrm{Vec}(Z^*)$

We use R_1 to proof the lemma. Without loss of generality we assume that R is in complex Jordan form, so R is block diagonal with diagonal blocks $J_i=\begin{bmatrix}\mu_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \mu_i\end{bmatrix}$. It is easy to see that this imposes a block–diagonal structure on M_1^{-1} with diagonal blocks:

(with $K_i:=[I-\mu_iA]^{-1}$)

$$[I\otimes I-J_i\otimes A]^{-1}=\begin{bmatrix}I-\mu_iA & -A & 0 & & \\ & I-\mu_iA & -A & \ddots & \\ & & \ddots & \ddots & 0 \\ & & & \ddots & -A \\ & & & & I-\mu_iA\end{bmatrix}^{-1}=\begin{bmatrix}K_i & AK_i^2 & AK_i^3 & \cdots & \\ & K_i & AK_i^2 & \cdots & \\ & & \ddots & \ddots & \\ & & & \ddots & AK_i^2 \\ & & & & K_i\end{bmatrix}$$

and in the same way N_1 is block diagonal with diagonal blocks

$$[J_i\otimes F-I\otimes FA^*]=\begin{bmatrix}F & & & \\ & F & & \\ & & \ddots & \\ & & & F\end{bmatrix}\begin{bmatrix}\mu_iI-A^* & I & & 0 \\ & \mu_iI-A^* & \ddots & \\ & & \ddots & I \\ & & & \mu_iI-A^*\end{bmatrix}$$

This implies that R_1 is block diagonal with diagonal blocks: (with $X=F(I-A^*A)$)

$$\begin{bmatrix}F(\mu_iI-A^*)K_i & XK_i^2 & & XAK_i^3 & XA^2K_i^4 & \\ & F(\mu_iI-A^*)K_i & XK_i^2 & \ddots & XAK_i^3 & \\ & & \ddots & \ddots & \ddots & XK_i^2 \\ & & & \ddots & \ddots & \\ & & & & & F(\mu_iI-A^*)K_i\end{bmatrix}$$

This shows immediately that all eigenvalues of R_1 and thus also of R_o are the eigenvalues of the matrices $F(\mu_i-A^*)(I-\mu_iA)^{-1}$ and it remains to proof that each of these matrices generates (n–1) poles in z=–1 and 1 stable pole.

Let μ be an eigenvalue of R ($|\mu|<1$) and consider the matrices

$$\Delta := F(\mu - A^*)(I-\mu A)^{-1}$$
$$Z := \mu F((I-\mu A)^{-1}$$

Then it is easy to show that $\Delta = -FA^* + (I-AA^*)Z$:

proof: $-FA^* + (I-AA^*)\mu F((I-\mu A)^{-1} = (-FA^* + \mu FA^*A + \mu F - \mu AA^*F)\,(I-\mu A)^{-1}$
$= F(\mu - A^*)(I-\mu A)^{-1}$ because $FA^*A = AA^*F$

Hence we are looking for the eigenvalue of the matrix $\Delta = -FA^* + (I-AA^*)Z$

Express Δ in terms of U, S and V ($A=USV^*$, $S=\mathrm{diag}(1,\cdots 1,\sigma)$ $0\leq\sigma<1$):

$$\Delta = -USU^* + U(I-S^2)U^*Z = U(-S + (I-S^2)U^*ZU)U^*$$

hence
$$U^*\Delta U = \begin{bmatrix} -I & \\ & -\sigma \end{bmatrix} + \begin{bmatrix} 0 & \\ & 1-\sigma^2 \end{bmatrix} U^*ZU$$

and this expression shows that $[I_{n-1}\ 0]U^*\Delta U = -[I_{n-1}\ 0]$. So $U^*\Delta U$ (and Δ) has (n–1) poles in z=–1.

For the remaining pole of Δ, we consider the expression $\Delta := F(\mu - A^*)(I-\mu A)^{-1}$.

Let $\{\lambda_i\}$ be the eigenvalues of Δ then we know that $\det(\Delta) = \prod_{i=1}^{n} \lambda_i$:

Now let $\lambda_1=\cdots=\lambda_{n-1}=-1$ then: $(-1)^{n-1}\lambda_n = \det(F)\det(\mu - A^*)\det((I-\mu A)^{-1})$

and because F is unitary $|\lambda_n| = |\det(\mu - A)\det((I-\mu A)^{-1})|$

Let A have eigenevalues $\{\rho_i\}$: $|\lambda_n| = \prod_{i=1}^{n} \left|\frac{\mu-\rho_i}{1-\mu\rho_i}\right|$

The stability of A and R now implies that each of these terms $\left|\frac{\mu-\rho_i}{1-\mu\rho_i}\right|<1$ and we conclude that λ_n is a stable pole of Δ, and thus a stable pole of R_1 and R_0.

Finally we must proof that for every μ_i which is an eigenvalue of both R and A the matrix R_0 will have a pole in 0. This is equivalent with the statement that λ_n above is zero. Let z be the corresponding eigenvector of A^*: $A^*z=\mu_i z$. Then $(I-\mu_i A)^{-1}F(\mu_i - A^*)z=0$ or $\lambda_n=0$.

This completes the proof of the lemma. □

Appendix 5L proof of proposition 5.4.12.

$R_0=NM^{-1}$ has n_q uncoupled eigenvalues in zero if there exist a matrix Γ with dimensions $(n_w n \times n_q)$ and $\mathrm{rank}(\Gamma)=n_q$ such that $R_0\Gamma=0$. This is equivalent with the statement that $Ny=0$, with y a vector, has n_q independent solutions.

$Ny=0 \Leftrightarrow (-A^*\otimes I+I\otimes R)y=0 \Leftrightarrow AY=YR$ where Y is such that $\mathrm{Vec}(Y)=y$.

Now let A and R have complex Jordan decompositions $A=T_a^{-1}J_aT_a$ and $R=T_rJ_rT_r^{-1}$ where J_a and J_r are ordered such that their $n_q \times n_q$ left upper blocks are the same. We denote this block by J_c and substitute these decompositions in the above equality,

$T_a^{-1}J_aT_aY = YT_rJ_rT_r^{-1} \Leftrightarrow J_aZ = ZJ_r$ where $Z=T_aYT_r$

This last equation has clearly n_q indepent solution Z_k for $k=1\cdot\cdot n_q$: Let $(Z_k)_{ij}=0$ except for the element $(Z_k)_{kk}=1$. This generates n_q independent solutions y of the equation $Ny=0$, which completes the proof of the proposition. □

Appendix 5M proof of proposition 5.4.15.

For the proof we assume that U=I, hence A has a svd $A=ISV^*$ with $S=\begin{bmatrix}1 & \\ & \sigma\end{bmatrix}$ with $0\leq\sigma<1$. In that case $FA^*=S$. This assumption is valid since the proof of lemma 5.4.9 shows that there exist a similarity transformation on $[R_o,G_o,H_o,L_o]$ which is equivalent with making U=I. From lemma 5.4.9 we learn that the minimal part of $[R_o,G_o,H_o,L_o]$ is given by $[\ [0\ I]R_o\begin{bmatrix}0\\I\end{bmatrix},\ [0\ I]G_o,\ H_o\begin{bmatrix}0\\I\end{bmatrix},\ L_o]$. We will show that this system has the proposed Gramians. The proof is not complicated but quite tricky and therefore we first establish some equations that will shorten the final calculations.

a. $VV^* = S^2\otimes P -(I-M)(I\otimes P)(I-M^*)$

Proof $VV^* = (A\otimes G)(A^*\otimes G^*) = (AA^*)\otimes(GG^*) =(AA^*)\otimes(P-RPR^*)$

$=(AA^*)\otimes P - (A\otimes R)(I\otimes P)(A^*\otimes R^*) = S^2\otimes P -(I-M)(I\otimes P)(I-M^*)$

b. $N(I\otimes P)-WV^*= (F\otimes R)(I\otimes P)M^*$

Proof $N(I\otimes P)-WV^* = (-S\otimes I + F\otimes R)(I\otimes P) + (F\otimes G)(A^*\otimes G) =$

$= -(S\otimes P) + (F\otimes RP) + (S\otimes P) - (S\otimes RPR^*) = (F\otimes R)(I\otimes P)(I- F^*S\otimes R^*)$

$= (F\otimes R)(I\otimes P)(I- A^*\otimes R^*) = (F\otimes R)(I\otimes P)M^*$

The following calculation now shows that (R_o,G_o) has a Gramian $(I-S^2)(I\otimes P)$:

$R_o(I-S^2)(I\otimes P)R_o^* + G_oG_o^* =$

$= NM^{-1}(I-S^2)(I\otimes P)M^{-*}N^* + WW^* + NM^{-1}VV^*M^{-*}N^* - WV^*M^{-*}N^* -NM^{-1}VW^*$

$\underline{a}$ $NM^{-1}\Big[(I-S^2)(I\otimes P) + S^2\otimes P -(I-M)(I\otimes P)(I-M^*)\Big]M^{-*}N^* + WW^* - WV^*M^{-*}N^*$

$-NM^{-1}VW^*$

$= NM^{-1}\Big[-M(I\otimes P)M^*+(I\otimes P)M^*+ M(I\otimes P)\Big]M^{-*}N^* + WW^* - WV^*M^{-*}N^* -NM^{-1}VW^*$

$= -N(I\otimes P)N^* + NM^{-1}\Big[(I\otimes P)N^*- VW^*\Big] + \Big[N(I\otimes P)- WV^*\Big]M^{-*}N^* + WW^*$

$\underline{b}$ $-N(I\otimes P)N^* + N(I\otimes P)(F^*\otimes R^*) + (F\otimes R)(I\otimes P)N^* + (I\otimes GG^*)$

$= -N(I\otimes P)N^* + N(I\otimes P)(N^* + S\otimes I) + (N + S\otimes I)(I\otimes P)N^* + (I\otimes P) - (I\otimes RPR^*)$

$= N(I\otimes P)N^* + N(I\otimes P)(S\otimes I) + (S\otimes I)(I\otimes P)N^* + (I\otimes P) - (I\otimes RPR^*)$

$= (S^2\otimes I)(I\otimes P) + (I\otimes P)(I\otimes RPR^*) -(SF^*\otimes I)(I\otimes PR^*) - (FS\otimes I)(I\otimes RP) - (S^2\otimes I)(I\otimes P) +$

$(SF^*\otimes I)(I\otimes PR^*) - (S^2\otimes I)(I\otimes P) + (FS\otimes I)(I\otimes RP) + (I\otimes P) - (I\otimes RPR^*)$

$= -(S^2\otimes I)(I\otimes P) + (I\otimes P) = (I-S^2)(I\otimes P)$

A close look at this Gramian shows that it has the form $\begin{bmatrix} 0 \\ (1-\sigma^2)P \end{bmatrix}$ which shows that the minimal part of (R_o,G_o) has Gramian $(1-\sigma^2)P$.

We proceed with the observability Gramian, which can not be calculated in the same manner, since the equation $R_o^*Q_oR_o + H_o^*H_o = Q_o$ has no solution because of the poles in z=–1. This implies that we must proof in a direct way that

$$[0\ \mathrm{I}]R_o^*\begin{bmatrix}0\\ \mathrm{I}\end{bmatrix}Q[0\ \mathrm{I}]R_o\begin{bmatrix}0\\ \mathrm{I}\end{bmatrix} + (1-\sigma^2)[0\ \mathrm{I}]H_o^*H_o\begin{bmatrix}0\\ \mathrm{I}\end{bmatrix} = Q$$

We first proof the following identity, which shortens the calculations:

c: $N^*\left[\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes Q\right]N + (1-\sigma^2)(\mathrm{I}\otimes H^*H)$

$= M^*\left[\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}\otimes Q\right]M + M^*\left[\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q\right] + \left[\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q\right]M$

Proof:

Left side identity=

$= (F^*\otimes R^*-S\otimes \mathrm{I})\left[\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes Q\right](F\otimes R-S\otimes \mathrm{I}) + (1-\sigma^2)(\mathrm{I}\otimes Q) - (1-\sigma^2)(\mathrm{I}\otimes R^*QR)$

$= F^*\begin{bmatrix}0&0\\0&1\end{bmatrix}F\otimes R^*QR + S\begin{bmatrix}0&0\\0&1\end{bmatrix}S\otimes Q - F^*\begin{bmatrix}0&0\\0&1\end{bmatrix}S\otimes R^*Q - S\begin{bmatrix}0&0\\0&1\end{bmatrix}F\otimes QR + (1-\sigma^2)(\mathrm{I}\otimes Q) -$

$-(1-\sigma^2)(F^*F\otimes R^*QR)$

$= F^*\begin{bmatrix}\sigma^{2-1}&0\\0&\sigma^2\end{bmatrix}F\otimes R^*QR + \begin{bmatrix}1-\sigma^2&0\\0&1\end{bmatrix}\otimes Q - F^*S\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes R^*Q - \begin{bmatrix}0&0\\0&1\end{bmatrix}SF\otimes QR$

Right side identity=

$= (\mathrm{I}\otimes \mathrm{I}-A^*\otimes R^*)\left[\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}\otimes Q\right](\mathrm{I}\otimes \mathrm{I}-A\otimes R) + (\mathrm{I}\otimes \mathrm{I}-A^*\otimes R^*)\left[\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q\right]$

$+ \left[\begin{bmatrix}1-\sigma&0\\0&0\end{bmatrix}\otimes Q\right](\mathrm{I}\otimes \mathrm{I}-A\otimes R)$

$= \begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}\otimes Q + A^*\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}A\otimes R^*QR - A^*\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}\otimes R^*Q - \begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}A\otimes QR$

$+ 2\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q - A^*\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes R^*Q - \begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}A\otimes QR$

$= \begin{bmatrix}1-\sigma^2&0\\0&1\end{bmatrix}\otimes Q + A^*\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}A\otimes R^*QR - A^*\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes R^*Q - \begin{bmatrix}0&0\\0&1\end{bmatrix}A\otimes QR$

$= \begin{bmatrix}1-\sigma^2&0\\0&1\end{bmatrix}\otimes Q + F^*S\begin{bmatrix}\sigma^{2-1}&0\\0&1\end{bmatrix}SF\otimes R^*QR - F^*S\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes R^*Q - \begin{bmatrix}0&0\\0&1\end{bmatrix}SF\otimes QR$

$= \begin{bmatrix}1-\sigma^2&0\\0&1\end{bmatrix}\otimes Q + F^*\begin{bmatrix}\sigma^{2-1}&0\\0&\sigma^2\end{bmatrix}F\otimes R^*QR - F^*S\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes R^*Q - \begin{bmatrix}0&0\\0&1\end{bmatrix}SF\otimes QR$

and this shows the validity of equation c.

Now we evaluate the equation for the observability Gramian:

$$[0\ I]R_o^*\begin{bmatrix}0\\I\end{bmatrix}Q[0\ I]R_o\begin{bmatrix}0\\I\end{bmatrix} + (1-\sigma^2)[0\ I]H_o^*H_o\begin{bmatrix}0\\I\end{bmatrix} =$$

$$= [0\ I]M^{-*}\left[\ N^*\begin{bmatrix}0\\I\end{bmatrix}Q[0\ I]N + (1-\sigma^2)(I\otimes H^*)(I\otimes H)\ \right]M^{-1}\begin{bmatrix}0\\I\end{bmatrix}$$

$$= [0\ I]M^{-*}\left[\ N^*\left(\begin{bmatrix}0&0\\0&1\end{bmatrix}\otimes Q\right)N + (1-\sigma^2)(I\otimes H^*H)\ \right]M^{-1}\begin{bmatrix}0\\I\end{bmatrix}$$

$$\stackrel{c}{=} [0\ I]M^{-*}\left[\ M^*\left(\begin{bmatrix}\sigma^2-1&0\\0&1\end{bmatrix}\otimes Q\right)M + M^*\left(\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q\right) + \left(\begin{bmatrix}1-\sigma^2&0\\0&0\end{bmatrix}\otimes Q\right)M\ \right]M^{-1}\begin{bmatrix}0\\I\end{bmatrix}$$

$$= [0\ I]\begin{bmatrix}\sigma^2-1&0\\0&1\end{bmatrix}\otimes Q\begin{bmatrix}0\\I\end{bmatrix} = Q$$

This proves that the minimal part of (H_o,R_o) has Gramian $(1-\sigma^2)^{-1}Q$.

Hence we have shown that if U=I the minimal part of $[R_o,G_o,H_o,L_o]$ has Gramians $(1-\sigma^2)P$ and $(1-\sigma^2)^{-1}Q$. The Hankel singular values of the system are the square roots of the eigenvalues of the product of the Gramians, i.e. of PQ, which shows that [R,G,H,L] and $[R_o,G_o,H_o,L_o]$ have the same Hankel singular values. This completes the proof of the proposition. □

Appendix N proof of proposition 5.4.25

The proof is based on comparison of the responses of the expressions on both sides of the equal sign in (5.4.33) on special input sequences. Therefore we write (5.4.33) in a Markov parameter form, by assuming that $\{u(t),y(t)\}$ is an input/output pair of both expressions. Then it follows that:

$$y(t) = [HG\ HRG\ HR^2G\ \cdots]\begin{bmatrix}u(t)\\u(t-1)\\\vdots\end{bmatrix} = [W_o\ W_1\ \cdots][B_e\ A_eB_e\ A_e^2B_e\ \cdots]\begin{bmatrix}u(t)\\u(t-1)\\\vdots\end{bmatrix}$$

and from this we may conclude that for any set of inputs $\{u_1(t),\cdots,u_n(t)\}$ we have the identity

$$[HG\ HRG\ \cdots]\begin{bmatrix}u_1(t)&\cdots u_n(t)\\u_1(t-1)&\cdots u_n(t-1)\\\vdots&\vdots\end{bmatrix} = [W_o\ W_1\ \cdots][B_e\ A_eB_e\ \cdots]\begin{bmatrix}u_1(t)&\cdots u_n(t)\\u_1(t-1)&\cdots u_n(t-1)\\\vdots&\vdots\end{bmatrix}$$

Now let $\{u_1(t),u_2(t),\cdots\}$ be defined by $\begin{bmatrix}u_1(t)&u_2(t)&\cdots\\u_1(t-1)&u_2(t-1)&\cdots\\\vdots&\vdots\end{bmatrix} = \begin{bmatrix}B_e^*\\B_eA_e^*\\\vdots\end{bmatrix}$ then the orthonormality of $[B_e\ A_eB_e\ \cdots]$ implies that $[W_o\ W_1\ \cdots] = H\sum_{k=0}^{\infty} R^kGB_e^*(A_e^*)^k$

We denote $Q:=\sum_{k=0}^{\infty} R^kGB_e^*(A_e^*)^k$ $(=[Q_0\ Q_1\cdots])$, so $W_k=HQ_k$.

It follows that $RQA_e^* = Q - GB_e^*$ and if we recall the form of A_e this results in:

$$R[Q_0\ Q_1\cdots]\begin{bmatrix} A^* & X^* & X^*P^* & \cdot\cdot \\ 0 & A^* & X^* & \cdot\cdot \\ 0 & 0 & A^* & X^*\cdot\cdot \\ 0 & 0 & 0 & A^*\cdot\cdot \\ \cdot & \cdot & \cdot & \cdot\ \cdot\cdot \end{bmatrix} = [Q_0\ Q_1\cdots] - G[B^*\ B^*P^*\ B^*(P^*)^2\cdots]$$

Examine the first element, it follows immediately that

$$RQ_0A^* = Q_0 - GB^*$$

which is (5.4.33c) and post multiplication with P^* ($P = -FA^*$) shows that

$$-RQ_0A^*AF^* = -Q_0AF^* - GB^*P$$

which is the same as ($X = F(I - A^*A)$)

$$RQ_0X^* - RQ_0F^* = -Q_0AF^* - GB^*P$$

Examine the second element:

$$RQ_0X^* + RQ_1A^* = Q_1 - GB^*P^*$$

Subtract the previous equation:

$$RQ_1A^* + RQ_0F^* = Q_1 + Q_0AF^*$$

and conclude that

$$Q_1 = RQ_1A^* + RQ_0F^* - Q_0AF^*$$

which is (5.4.33d) for n=0.

The other equations follow from the same line of reasoning. □

6. REVIEW, CONCLUSIONS AND PROSPECT

In this thesis we considered the problem of system approximation and we focused on two major subjects that are intrinsic to this problem:

- Model reduction, i.e. the approximation based on a model description
- The application of orthonormal functions to system descriptions and to approximate system identification

The study of these concepts was motivated by the fact that in general exact modeling of dynamic processes is not feasible, not possible or even not desirable. This led to the observation that the problem of system identification should be considered as an approximation issue and this links identification to the problem of model reduction. Both fields deal essentially with same question:

> "How can we approximate the behavior of a dynamic process with the behavior of a model from a specific model class, with respect to a certain criterion?"

We noted an existing mismatch between the common approaches to both problems, in the sense that the underlying criteria are not compatible and sometimes not well understood. The ideal situation, in which a variety of criteria could be applied to both problems with insight and understanding of the character of the approximation, has not been created yet. The aim of this thesis is to contribute to establishing such a situation, that in our point of view is necessary to approach the problem of system approximation in a promising way. Though some people ventilate the opinion that the research on system identification has reached all it can achieve, we strongly believe that this research has only just started.

We are fully aware that it will take a lot of time and effort to reach a situation, described above, and that at first it is necessary to achieve better understanding of the principles of model reduction, approximate identification and approximation in itself. This motivated us to focus on the concepts of model reduction and orthogonality. We concentrated on discrete time systems and signals, because the use of digital computers enforces the use of these, both for data manipulations and for the application of models, such as control engineering and prediction.

The concept of orthogonality is widely used in mathematics and systems theory and it is closely connected with the principle of approximations. We confined ourselves to the use of orthonormal functions in the context of signal and system descriptions. The underlying idea is that if we are able to establish an orthogonal decomposition of a systems behavior, then this enables the approximation by means of projection in this orthogonal basis and create approximating models from a finite number of these orthogonal components. The aim is the reconstruction of the most important part of the behavior along this way.

Model reduction has a long history as well, and many different methods have been proposed and applied. Though many of these methods have been used successfully or with satisfactory results, the nature of the approximation remains obscure for most procedures. In fact only one method is known to be optimal with respect to a well defined criterion, the so called Hankel norm approximation. Another popular method, which is closely connected with the latter is balanced truncation. We studied both methods in order to get a better understanding of the approximation principle, especially for discrete time systems. Another motivation to study these methods is the fact that both are based on an orthogonality principle. We suspected that more insight in these methods might lead to the use of this specific orthogonality concept for the general problem of system approximation.

The main part of this thesis, dealing with the issues described before, is embodied in chapters two, three, four and five. Chapter two deals in principle with standard mathematical and system theoretical tools. We choose to include a new result on state space representations of discrete all–pass functions, because this representation of a specific class of orthonormal functions is basic to the theory of the remaining chapters. The new feature of this theorem 2.7.4 is that no restrictions are imposed on the poles of the all–pass system. This implies that the occurrence of time delays will not cause problems. It is well known that in practice time delays are often present. A second new result that we presented in chapter two deals with state space representations of coprime factors of discrete time systems, again without restrictions on the poles of the system. This theorem 2.8.3 enables the description of the signal behavior of a system in more general terms than is possible with other system representations. The result enables the construction of alternative model reduction procedures in chapter three.

The problem of model reduction is addressed in chapter three. We defined a generalized reduction method based on balanced realizations by introduction of a design variable, the so called reduction factor. It was shown that the standard methods based on balanced realizations are special cases of this more general procedure. Conditions were

derived under which stability and minimality of the reduced order models is guaranteed and we established a bound for the frequency characteristics of the error systems. It was conjectured that the H_∞ norm of the error transfer function, considered as a function of the reduction factor, is a function with only one minimum. This feature was illustrated with some examples which led to the conclusion that an optimal choice of the reduction factor can improve the frequency behavior considerably when compared to the standard methods. A complete formulation of the underlying criterion of these approximations has not yet been obtained.
For optimal Hankel norm approximation of discrete time systems we presented a set of sufficient equations in terms of state space matrices and showed that with these equations a solution can be calculated for the SIMO/MISO case. We presented a method to solve the problem for a restricted class of MIMO systems. In chapter four we presented a method that can reduce these restrictions, but we need a specific transformation for this purpose. We must conclude that the solution of the discrete time Hankel norm problem directly in the time domain is much more involved than the continuous time equivalent. The goal of understanding the orthogonality principle of Hankel norm approximation for the application in terms of orthogonal functions was not completely achieved.
The combination of any reduction method with the state space description of coprime factors, as derived in chapter two, results in a model reduction method that is referred to as fractional model reduction. Such a method, that contrary to the other model reduction procedures does not distinguish between stable or unstable systems, can be explained to be a method that approximates the signal behavior of a system and therefore has a close connection with the problem of approximate system identification. Therefore further study of these methods, preferably in combination with a well understood reduction methods, is highly recommended. In this thesis we have not pursued this line.

In the chapters four and five we considered the application of orthonormal functions to the description of signals and systems. In chapter four we restricted attention to a special class of functions, the discrete Laguerre polynomials and elaborated on the use of these functions to represent signals. The fact that these functions constitute a basis for ℓ_2 enabled the definition of the Laguerre transformation of time series to a sequence of expansion coefficients. It was shown how the recursive structure of these polynomials, often called the shift structure, can be applied for efficient calculation of the coefficients. Application of this transformation to the input/output signals of finite dimensional linear time-invariant systems induced a transformation on system level, the Laguerre system transformation, which leaves important system properties invariant. This bijective transformation has been expressed on transfer function level, as well as in state

space form. Furthermore we explained that the forward and backward Laguerre transformations are symmetric operations.
We considered three different approaches to apply these transformations in the context of approximate system identification, which resulted in two procedures with promising results. These methods are referred to as ARX respectively FIR in the Laguerre domain. For ARX we showed that the procedure is (in the asymptotic case) equivalent with ARX in the time domain on prefiltered data, where the prefilter is induced by the specific choice of the orthogonal polynomials. The orthogonal FIR method is a generalization of the estimation of FIR models in the time domain in the sense that the approximation with pulse functions is generalized to the approximation with impulse functions. The character of the prefilters and impulse responses in both methods is determined by the so called discount factor, the one degree of freedom in the definition of Laguerre polynomials. It has been shown that both methods can improve the results of their time domain equivalent considerably, as well in respect to the resulting approximations as to the number of parameters to be estimated. The exact nature of the approximations remains unknown, but examples show that along this way the combination of simple model structures with least squares estimation techniques can provide results that are superior to the time domain equivalents.
The combination of the Laguerre and the ω–transformation yielded the Laguerre transformation of continuous time systems and we presented a commutation diagram that links these transformations with the model reduction methods based on balanced realizations and with Hankel norm approximation. This enabled us to show that problems with the solution of the discrete Hankel norm approximations, which were detected in chapter three, can be overcome with the application of Laguerre transformation.

The ambiguity in the choice of the discount factor of the Laguerre polynomials and the somewhat unclear reason to use specifically these polynomials as orthogonal functions motivated the search for other orthonormal functions, that can be found and defined on system level.
This resulted in chapter five in the derivation of the so called system based orthonormal functions, an infinite set of functions, directly based on (input) balanced realizations of finite dimensional linear time invariant systems. These functions are induced by the (orthonormal) matrices with singular vectors of a Hankel matrix and the theory is based on the description of discrete all–pass functions, which was presented in chapter two. We showed that under weak conditions such a set of functions forms a basis for the function space ℓ_2 and hence that every system defines its own basis. These new functions represent the dynamic behavior of the system that generated them. It was shown that the sets of Laguerre polynomials are the special cases of these general functions if the

generating systems are first order siso systems. Furthermore the new orthonormal functions have the same type of recursive structure as the Laguerre polynomials. Extending the results of chapter four on transformations we defined signal and system transformations, based on these system based orthonormal functions. The recursive shift structure served for a compact description of the functions as Markov parameter matrices of the so called generating transfer functions, which enabled an efficient algorithm for the actual calculations of a signal transformation. The induced system transformation is much more involved than in the Laguerre case, but it was shown that the important properties of McMillan degree, stability, Hankel norm and H_∞ norm are invariant under the system transformation with system based orthonormal functions. An important feature of this transformation is that, if the orthonormal functions are obtained from the system which must be transformed, then the resulting new system has no dynamics in the sense that all poles are zero and the new system has only two non–zero Markov parameters. The inverse of this transformation is well defined but we have not been able to present but an adhoc method to perform the calculations that are involved in this step.
For the application of approximate system identification we generalized the methods of chapter four to ARX respectively FIR in the orthonormal domain. In a classical sense these methods can be seen as the application of ARX in the time domain with the use of a special class of prefilters respectively as a generalization of the estimation of FIR models in the time domain. Especially the latter approach can be seen as an orthonormal decomposition of the system behavior. This new FIR method raised a problem, that is a generalization of the approximate realization problem.

An iterative application of these orthonormal methods, updating the orthonormal basis with new model estimates, can be seen as a search for that set of orthonormal functions, that minimizes the dynamic behavior of the system in the orthonormal domain. This shows that the search for an optimal orthonormal set is complementary to approximate identification in the sense that both methods aim at the characterization of the dynamic structure, that is contained in the data. While normally the identification method has to account for the dynamics, with the use of orthonormal functions part of this dynamics is covered by the orthonormal transformation. This observation motivates that identification should more be considered as grey box modeling. The knowledge of the system can be used for the definition of orthonormal functions and the identification has to deal with the unknown aspects of the system.

If we summarize the approach towards approximate system identification, that we proposed and considered in this thesis then we distinguish the following main steps and elements:

1. Choice of an initial set of orthonormal functions $\{\Psi_k\}$ (4.2.7, 5.3.9) with corresponding generating transfer functions (4.2.9, 5.3.10). These functions can either be a set of Laguerre polynomials or a set of system based orthonormal functions, based on an initial model guess.

↓

2. Choice of the domain in which the estimation has to be performed:

a. <u>The orthonormal domain.</u>	b. <u>The time domain.</u>
↓	↓
3. a. Transformation of input/output data to the orthonormal domain (5.4.4.), $\{u,y\} \rightarrow \{U,Y\}$	b. Calculation of the orthonormal states $\{X_k\}$ (5.4.24).
↓	↓
4. a. Estimation of a model on the data $\{U,Y\}$, for instance an ARX–model (section 5.5.2). (Choice of model–structure)	b. Estimation of a FIR–model $y = D \cdot u + \sum_{i=0}^{n} C_i X_i$ (Choice of n, the number of orthonormal states)
↓	↓
5. a. Transformation of the model to a time domain model (procedure 5.4.19, remark 5.5.2).	b. Realization of a low order model from the estimated parameters $\{D,C_i\}$ (remark 5.5.9). (Choice of realization method)

↓

6. Validation of the resulting model. If the result is not satisfactory then repeat previous steps with different settings:
   - another estimation domain in 2.
   - a different model structure in 4a.
   - another number of orthonormal states in 4b.
   - another realization method or a higher model order in 5b.

↓

7. Use the obtained model estimate to define a new set of system based orthonormal functions with corresponding generating transfer functions and repeat the steps described above.

If we consider the problem statement and the related issues that were considered in chapter one, then we may conclude that we have shown that there are indeed promising possibilities in the combination of prediction error techniques with the use of orthogonal functions and descriptions. We showed the existence of a specific class of system based orthonormal functions that lead to alternative descriptions of linear systems, preserving the important properties of the system. We have not established a transparent link between model reduction and orthogonal functions, but the functions that we presented are closely related to balanced realizations and presumably also with the model reduction methods that are based on the representation. Experiences with the alternative identification methods show that these methods can be applied successfully and can improve the frequency characteristics of the model estimates, if compared to there classical counterparts. These observations show that the approach of approximate system identification along these lines is promising and that further research on this subject offers many opportunities to improve the results of existing methods.

We are aware that various critical remarks can be made concerning the methods and ideas that we proposed and that we left many loose ends, which shows that further study into the possibilities and limitations of this approach is necessary. In this light we want to propose a number of items, which we believe deserve attention in future research.

- Further research into the orthogonality concept of Hankel norm approximation and the application of the obtained knowledge to the definition of orthogonalsystems.

- Evaluation of the approximation character of reduction methods based on balanced realizations.

- System identification in terms of coprime factor descriptions and the application of the system based orthonormal functions to such descriptions.

- The study of system based functions that are orthogonal on the two–sided infinite interval or on finite intervals.

- Further research on the transformed systems, with better understanding of the imposed restrictions. This may lead to the solution of the inverse system transformation and the generalized realization problem. Furthermore it will yielbetter understanding of the orthonormal ARX method.

- Further study on the implications of the orthonormal transformation for model reduction, based on the observation that reduction in time domain and orthonormal domain are intrinsically related, while the description of the system in the optimal orthonormal domain is given by only two Markov parameters.

- Research on the optimal experiment design for methods that are based on orthogonal functions, If the identification or estimation is performed in an orthogonal domain then the input signals in the orthogonal domain should be persistently exciting.

- Analysis of the influence of different types of disturbances on the results of the proposed identification approaches and of the question whether an iterative version of these procedures will converge to an optimal set of orthonormal functions.

We acknowledge that this list is by no means complete. We expect that the answers to the above items will lead to new contributions to and developments in the fields of approximate system identification and model reduction.

REFERENCES

Adamyan, V.M., D.Z. Arov and M.G. Krein (1968). Infinite Hankel matrices and generalized Caratheodory–Fejer and Riesz problems. *Funct. Anal. Appl. 2*, 1–18.

Adamyan, V.M., D.Z. Arov and M.G. Krein (1971). Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur–Takagi problem. *Math. USSR Sbornik, vol.15, no.1,* 31–73.

Adamyan, V.M., D.Z. Arov and M.G. Krein (1978). Infinite block Hankel matrices and related extension problems. *Amer.Math.Soc.Transl.(2), vol.111,* 133–156.

Al–Saggaf, U.M. and G.F. Franklin (1986). On model reduction. *Proc. 25th Conf. on Decision and Control,* Athens, Greece, 1064–1069.

Al–Saggaf, U.M. and G.F. Franklin (1987). An error bound for a discrete reduced model of a linear multivariable system. *IEEE Trans. Automat. Contr. AC–32,* 815–819.

Al–Saggaf, U.M. and G.F. Franklin (1988). Model reduction via balanced realizations: an extension and frequency weighting techniques. *IEEE Trans. Automat. Contr., vol. AC–33,* 687–692.

Anderson, B.D.O and J.B. Moore (1971). *Linear Optimal Control.* Prentice–Hall, Englewood Cliffs N.J.

Backx, A.C.P.M. (1987). *Identification of an Industrial Process: A Markov Parameter Approach.* Dr.–Dissertation, Dep. Electr. Eng., Eindhoven Univ. Technology, The Netherlands.

Ball, J.A. and A.C.M. Ran (1987). Optimal Hankel norm model reductions and Wiener–Hopf factorization I: The canonical case. *Siam J.Contr. and Optim.* vol. 25–2, 362–382. II: The non–canonical case. *Integral Equations and Operator Theory, vol 10,* Birkhäuser Verlag, Basel, 416–436.

Bartels, R.H. and G.W. Stewart (1972). Solution of the matrix equation AX+XB=C. *Communs. Ass. Comput. Mach. 15,* 820–826.

Bellman, R. (1970). *Introduction to Matrix Computations.* McGraw–Hill, New York.

Bongers, P.M.M. and P.S.C. Heuberger (1990[a]). Discrete normalized coprime factorization and fractional balanced reduction. *Selected Topics in Identification, Modelling and Control, Vol. 1,* 57–60, Delft University Press, The Netherlands.

Bongers, P.M.M. and P.S.C. Heuberger (1990[b]). Discrete normalized coprime factorization. *Proc. 9th Int. Conf. Analysis and Optimization of Systems, Antibes, France, June 12–15.* 307–313.

Brockett, R.W. (1970). *Finite Dimensional Linear Systems.* John Wiley and Sons Inc. New York.

Bruinsma, N.A. and M. Steinbuch (1990). A fast algorithm to compute the H_∞ norm of a transfer function matrix. *Syst. Control Letters 14,* 287–293.

Chang, R., K. Chen and M. Wang (1985). A new approach to the parameter estimation of linear time–invariant delayed systems via modified Laguerre polynomials. *Int. J. Systems Sci., vol. 16, no. 12,* 1505–1515.

Chu, C–C. (1988). On discrete inner–outer and spectral factorizations. *Proc. Amer. Contr. Conf., Atlanta, Georgia,* 1699–1700.

Clowes, G.J. (1965). Choice of the time–scaling factor for linear system approximations using orthonormal Laguerre functions. *IEEE Trans. Automat. Contr., vol. AC–10,* 487–489.

Desoer, C.A, R–W Liu, J. Murray and R. Seaks (1980), Feedback System Design: The Fractional Representation Approach to Analysis and Synthesis. *IEEE Trans. Automat. Contr., vol. AC–25,* 399–412.

Doyle, J.C., K. Glover , P.P. Khargonekar and B.A. Francis (1989). State–space solutions to standard $\mathcal{H}_2$ and $\mathcal{H}_\infty$ control problems. *IEEE Trans. Automat. Contr., vol. AC–34,* 831–847.

Enns, D.F. (1984). *Model Reduction for Control System Design.* PhD. diss., Dept. of Aeronautics and Astronautics, Stanford University.

Eykhoff, P. (1974). *System Identification: Parameter and State Estimation.* Wiley & Sons, London.

Fernando, K.V. and H. Nicholson (1982). Singular perturbational model reduction of balanced systems. *IEEE Trans. Automat. Contr., AC–27,* 466–468.

Fernando, K.V. and H. Nicholson (1983). On the structure of balanced and other principal representations of SISO systems. *IEEE Trans. Automat. Contr., AC–28,* 228–231.

Fernando, K.V. and H. Nicholson (1983). Singular perturbational approximations for discrete time balanced systems. *IEEE Trans. Automat. Contr., AC–28,* 240–242.

Francis, B.A. (1987). *A Course in H–infinity Control Theory.* Lect. Notes Cont. Inf. Sc., vol. 88. Springer Verlag, Berlin.

Glover, K. (1984). All optimal Hankel norm approximations of linear multivariable systems and their L_∞ – error bound. *Int. J. Contr. 39–6,* 1115–1193.

Glover, K. and D. McFarlane (1988). Robust stabilization of normalized coprime factors: an explicit H_∞ solution. *Proc. Amer. Contr. Conf.. Atlanta,* Georgia.

Golub, G.H. and C.F. Van Loan (1989). *Matrix Computations (Sec.Ed).* The John Hopkins University Press, Baltimore and London.

Gottlieb, M. (1938) Polynomials orthogonal on a finite or enumerable set of points. *Am. J. Math., 60,* 453–458.

Gray, W.S, and E.I. Verriest (1987). Optimality properties of balanced realizations: Minimum Sensitivity. *Proc. 26th Conf. on Decision and Control, Los Angeles,* 124–128.

Helmont, H. van, A.J.J. van der Weiden and H. Anneveld (1990). Design of optimal controller for a coal fired Benson boiler based on a modified approximate realisation algorithm. In: Whalley R. (Ed.) *Application of Multivariable System Techniques. Proc. Int. Symp. Appl. Multivariable System Techn. (AMST),* Bradford, UK, April 1990. Elsevier Sc. Publ. Ltd., London.

Heuberger, P.S.C. (1990). A family of reduced order models, based on open loop balancing. *Selected Topics in Identification, Modelling and Control, Vol. 1,* 1–10, Delft University Press, The Netherlands.

Heuberger, P.S.C. and O.H. Bosgra (1990[a]). Complete orthonormal sets, based on linear systems. *Selected Topics in Identification, Modelling and Control, Vol. 1,* 27–36, Delft University Press, The Netherlands.

Heuberger, P.S.C. and O.H. Bosgra (1990[b]). Approximate system identification using system based orthonormal functions. In *Proc. 29th. Conf. on Decision and Control, Hawaii, december 1990, section WP–XIV–2.*

Ho B.L. and R.E. Kalman (1966). Effective construction of linear state–variable models from input–output functions. *Regelungstechnik 14,* 545–548.

Hoffman, K. (1962). *Banach Spaces of Analytic Functions.* Prentice Hall.

Huang, Q. and R. Liu (1987). A necessary and sufficient condition for stability of a perturbed system. *IEEE Trans. Automat. Contr., AC–32,* 337–340.

Kabamba, P.T. (1985). Balanced forms: Canonicity and Parametrization. *IEEE Trans. Automat. Contr., AC–*30, 1106–1109.

Kailath, T. (1980). *Linear Systems,* Prentice Hall, New Jersey.

King, J.J. and T. O'Canainn (1969). Optimum pole positions for Laguerre– function models. *Electr.Letters 5,* 601–602.

King, R.E. and P.N. Paraskevopoulos (1977). Digital Laguerre filters. *Circ.Th. and Appl., vol. 5,* 81–91.

King, R.E. and P.N. Paraskevopoulos (1979). Parametric identification of discrete time SISO systems. *Int.J.Contr., vol.30,* 1023–1029.

Kreyszig, E. (1978). *Introductory Functional Analysis with Applications.* John Wiley & Sons, New York.

Kukk, K. (1971). State variables and Laguerre series. *Electr. Letters, vol. 7, no.10,* 269–270.

Kung, S. (1978). A new identification and model reduction algorithm via singular value decomposition. In: *Proc. 12th Amer. Asilomar Conf. Circuits. Syst. and Computers,* Pacific Grove, Ca., 705–714.

Kung, S–Y and D.W. Lin (1981). Optimal Hankel–norm model reductions: multivariable systems. *IEEE Trans. Automat. Contr., vol. AC–26,* 832–852.

Kung, S–Y and D.W. Lin (1981). A state space formulation for optimal Hankel norm approximations. *IEEE Trans. Automat. Contr., vol. AC–26,* 942–946.

Laub, A.J. (1980). Computation of balancing transformations. *Proc. 1980 Joint Amer. Contr. Conf. San Francisco.*

Lee, Y.W. (1933). Synthesis of electrical networks by means of the Fourier transforms of Laguerre functions. *J. Math. Physics, vol.11,* 83–113.

Liu, Y. and B.D.O. Anderson (1989). Singular perturbation approximation of balanced systems. In: *Proc. 28th Conf. on Decision and Control, Tampa, Florida,* 1355–1360.

Ljung, L. (1985). Asymptotic variance expressions for identified black box transfer function models. *IEEE Trans. Automat. Contr., AC–30,* 834–844.

Ljung L. (1987). *System Identification : Theory for the User.* Prentice–Hall Inc., Englewood Cliffs, New Jersey.

Ljung, L. (1989). Choices of prefilters and forgetting factors in noise free identification. *Proc. Amer. Control Conf. Pittsburgh, PA.* 1409–1415.

Maciejowski, J.M. and R.J Ober (1988). Balanced parametrizations and canonical forms for system identification. In: *Preprints of the 8th IFAC/IFORS Symp. on Identification and System Parameter Estimation,* Beijing, China, 989–996.

Maione, B. and B. Turchiano (1985). Laguerre z–transfer function representation of linear discrete–time systems. *Int. J. Control, vol 41, no.1,* 245–257.

Mäkilä, P.M. (1990). Approximation of stable systems by Laguerre Filters. *Automatica, Vol 26,* 333–345.

McFarlane, D. (1988). *Robust Controller Design using Normalized Coprime Factor Plant Descriptions.* PhD.–thesis, University of Cambridge.

Meyer, D.G. and G.F. Franklin (1988). A connection between normalized coprime factorization and linear quadratic optimal regulator theory. *IEEE Trans. Automat. Contr., vol. AC–32,* 1041–1047.

Meyer, D.G. (1988). A fractional approach to model reduction. *Proc. Amer. Control Conf. Atlanta, Georgia,* 1041–1047.

Moore, B.C. (1981). Principal component analysis in linear systems: controllability, observability and model reduction. *IEEE Trans. Automat. Contr., vol. AC–26,* 17–32.

Mullis, C.T. and R.A. Roberts (1976). Synthesis of minimum roundoff noise fixed point digital filters. *IEEE Trans. Circuits Syst., CAS–23,* 551–561.

Nett, C.N., C.A. Jacobson and M.J. Balas (1984). A connection between state space and doubly coprime fractional representations. *IEEE Trans. Automat. Contr., vol. AC–29,* 831–832.

Nurges, Y. (1987). Laguerre models in problems of approximation and identification of discrete systems. *Automation and Remote Control 48, no.3,* 346–352.

Nurges, Y. and Y. Yaaksoo (1981). Laguerre state equations for a multivariable discrete system. *Automation and Remote Control, vol 42,* 1601–1603.

Ober, R.J. (1987). Balanced Realizations: canonical form, parametrization, model reduction. *Int. J. Control 46,* 643–670.

Ober, R. and D. McFarlane (1988). Balanced Canonical Forms for Minimal Systems: A normalized coprime factor approach. *Report CUED/F – INFENG/TR11,* University of Cambridge, UK.

Paraskevopoulos, P.N. (1985). System analysis and synthesis via orthogonal polynomial series and Fourier series. *Mathematics and Computers in Simulation 27,* 453–469.

Pernebo, L. and L.M. Silverman (1983). Model reduction via balanced state space representation. *IEEE Trans. Automat. Contr., vol. AC–27,* 382–387.

Prabhakara, Rao C.V.K. and G. Maskarinec (1989). Implementation of arbitrary realizations in orthogonal systems with applications to principle axis realizations. In: M.A Kaashoek et. al.(Ed.) *Progress in System and Control Theory*, Birkhauser Boston Inc. 1990. Proc. Int. Symp. MTNS Amsterdam, 1989.

Rosenbrock, H.H. (1970). *State Space and Multivariable Theory.* Thomas Nelson and Sons Ltd., London, Great Britain.

Rudin, W. (1966). *Real and Complex Analysis.* McGraw–Hill.

Santiago, J.M and M. Jamshidi (1986). On the extensions of balanced approach for model reduction. *Contr. Theory and Adv. Techn. Vol.2, no.2,* 207–226.

Savonov, M.G., R.Y. Chiang and D.J.N. Limebeer (1987). Hankel model reduction without balancing, – A descriptor approach–. *Proc. 26th Conf. on Decision and Control, Los Angeles,* 112–117.

Silverman L.M. (1971). Realization of linear dynamical systems. *IEEE Trans. Automat. Contr., AC–16,* 554–567.

Steiglitz, K. (1965). Rational transform approximation via the Laguerre spectrum. *Journal of the Franklin Institute, vol.5, no.5,* 387–394.

Steiglitz K. and L.E. McBride (1965). A technique for the identification of linear systems. *IEEE Trans. Automat. Contr., AC–10,* 461–464.

Stewart, G.W. (1973). *Introduction to Matrix Computations.* Academic Press Inc.

Sveinsson, J.R. and F.W. Fairman (1985). Minimal balanced realization of transfer function matrices using Markov parameters. *IEEE Trans. Automat. Contr., vol. AC–30,* 1014–1016.

Swaanenburg, H.A.C., W.M.M. Schinkel, G.A. Van Zee and O.H. Bosgra (1985). Practical aspects of industrial multivariable process identification. In: Barker H.A. and P.C. Young (Eds.) (1985). *Identification and System Parameter Estimation,* IFAC Proc. Series 1985, York, UK, 201–206.

Unbehauen, H. and G.P. Rao (1988). Continuous–Time Approaches to System Identification. In: *Preprints of the 8th IFAC/IFORS Symp. on Identification and System Parameter Estimation,* Beijing, China, 60–68.

Van den Hof, P.M.J. and P.H.M. Janssen (1987). Some asymptotic properties of multivariable models identified by equation error techniques. *IEEE Trans. Automat. Contr. AC–32,* 89–92.

Van den Hof, P.M.J. (1989). *On Residual–Based Parametrization and Identification of Multivariable Systems.* Dr–thesis, Dept. Electr. Eng. Eindhoven Univ. Techn., The Netherlands.

Vidyasagar, M. (1985). *Control System Synthesis: A factorization approach.* MIT Press.

Wahlberg, B. (1984). Connections between system identification and model reduction. *Report LiTH–ISY–I–0746, Dept.Electr.Eng.,* Linköping Univ., Sweden.

Wahlberg, B. (1989). System identification using high order models, revisited. In: *Proc. 28th Conf. on Decision and Control, Tampa, Florida,* 634–639.

Wahlberg, B. and L. Ljung (1986) Design variables for bias distribution in transfer function estimation. *IEEE Trans. Automat. Control, AC–32,* 89–92.

Wiener, N. (1949) *Extrapolation, Interpolation and Smoothing of Stationary Time Series.* MIT Press, Cambridge, Ma.

Willems, J.C. (1986). From time series to linear system – Part I. Finite dimensional linear time invariant systems. *Automatica 22,* 561–580.

Willems, J.C. (1988). Models for dynamics. In: Kirchhaber, U. and H.O. Walther (Eds.), *Dynamics Reported.* Vol 2, 171–269. Wiley and Teubner, 1988.

Yang, C–D. and F–B. Yeh (1989). A simple algorithm on minimal balanced realization for transfer function matrices. *IEEE Trans. Automat. Contr., vol. AC–34,* 879–882.

Yeh, F–B. and C–D. Yang (1987). A new algorithm for the minimal balanced realization of a transfer function matrix. *System and Contr. Letters. 9,* 25–31.

Youla, D.C. (1961). On the factorization of rational matrices. *IRE Trans. Information Theory,* 172–189.

Zervos, C., P.R. Belanger and G.A. Dumont G.A. (1985). On PID controller tuning using orthonormal series identification. *IFAC Adapt. Contr. Chem. Processes. Frankfurt am Main,* 13–20.; Idem (1988) *Automatica, vol.24–2,* 165–175.

NOTATION

GENERAL

$\circ$	end of statement (theorem, lemma, proposition or remark)
$\square$	end of proof
A^T	transposed of the matrix A
A^{-T}	$(A^T)^{-1}$
A^*	$\bar{A}^T$, conjugate transpose of the matrix A
FDLT(S)	Finite Dimensional, Linear, Time invariant (, Stable)
$G^*(z)$	$\bar{G}^T(z^{-1})$
$G^*(s)$	$\bar{G}^T(-s)$
$y(-\infty)=0$	$\exists\, M\in\mathbb{R},\ t<M \Rightarrow y(t)=0$
$\sigma(A)$	the set of eigenvalues of the matrix A
$\lambda(A),\mu(A)$	an eigenvalue of the matrix A
σ	singular value
$\sigma_{max}(A)$	maximal singular value of the matrix A
$\sigma(G(p))$	Hankel singular value of the system G
δ_{ij}	Kronecker delta, $\delta_{ij}=1$ if i=j, $\delta_{ij}=0$ if i≠j
$\mathbb{Z}$	set of integers {..,–1,0,1,1,..}
$\mathbb{Z}_+$	set of non–negative integers; $\mathbb{Z}_+=\mathbb{Z}\cap[0,\infty)$
$\mathbb{Z}_-$	set of negative integers; $\mathbb{Z}_-=\mathbb{Z}\cap(-\infty,-1]$
$\mathbb{R}$	set of real numbers
$\mathbb{R}_+$	set of non–negative real numbers; $\mathbb{R}_+=\mathbb{R}\cap[0,\infty)$
$\mathbb{R}^q$	set of q–dimensional real vectors
$\mathbb{R}^{p\times q}$	set of p×q real matrices
$\mathbb{R}\cup\{\infty\}$	set of real numbers extended with infinity
$\mathbb{N}$	set of natural numbers; $\mathbb{N}=\mathbb{Z}\cap[1,\infty)$
$\mathbb{N}^0$	set of natural numbers extended with zero; $\mathbb{N}^0=\mathbb{N}\cup\{0\}$
$\mathbb{C}$	set of complex numbers
$\mathbb{C}^q$	set of q–dimensional complex vectors
$\mathbb{C}^{p\times q}$	set of p×q complex matrices
$\mathbb{F}$	a field of numbers, $\mathbb{F}=\mathbb{C}$ or $\mathbb{F}=\mathbb{R}$
$\{a_1,a_2,\cdots\}$	set consisting of the elements a_1,a_2 etc.
$[a,\infty]$	$[a,\infty)\cup\{\infty\}$

I_p	(p×p) identity matrix
A⊗B	the Kronecker product of the matrices A and B
det(A)	determinant of the matrix A
vec(A)	vector with the columns of the matrix A, stacked on top of each other
$\mathbb{F}[p]$	the set of polynomials with indeterminate variable p (p=z or p=s) and coefficients in $\mathbb{F}$
$\mathbb{F}(p)$	the set of rational functions with indeterminate variable p (p=z or p=s) and coefficients in $\mathbb{F}$
M_c	controllability matrix
M_o	observability matrix
P	controllability Gramian
Q	observability Gramian
$\mathcal{H}$	Hankel operator/matrix
Σ	diagonal matrix with Hankel singular values
$\mathcal{B}(G)$	behavior of the dynamical system G on $\mathbb{Z}$
$\mathcal{B}^+(G)$	behavior of G restricted to $\mathbb{Z}_+$
$\mathcal{B}_2(G)$	ℓ_2 behavior of the system G
$\mathcal{B}_2^+(G)$	ℓ_2 behavior of G restricted to $\mathbb{Z}_+$
ω	bilinear transformation
A_d	matrix from a realization of a discrete time system
B_c	matrix from a realization of a continuous time system
ARX	model structure, AutoRegressive eXogenous model
FIR	model structure, Finite Impulse Response model
$\mathcal{J}$	identification criterion

CHAPTER 3

$\mathcal{CB}(G)$	continuous time reduced order model, based on balancing and truncation
$\mathcal{CSB}$	continuous time reduced order model, based on balancing and singular perturbational truncation
$\mathcal{DTB}$	discrete time reduced order model, based on balancing and truncation in the discrete time domain
$\mathcal{DB}$	discrete time reduced order model, based on balancing and truncation in the continuous time domain
$\mathcal{DSB}$	discrete time reduced order model, based on balancing and singular perturbational truncation
$\mathcal{GB}^\alpha$	reduced order model, based on balancing and the reduction factor α

α	reduction parameter
AR	admissible region for $\mathcal{GB}$
$\mathcal{G}(G)$	graph of the system G

CHAPTER 4

$\gamma_k(t,\beta)$	unnormalized discrete time Laguerre polynomial
ξ	discount factor of Laguerre polynomials
η	$1-\xi^2$
$\overset{\xi}{\psi_k}(t),\psi_k(t)$	normalized discrete time Laguerre polynomial
$\overset{\xi}{\Psi_k}(z),\Psi_k(z)$	generating transfer function of $\overset{\xi}{\psi_k}(t)$
Γ, $\Gamma(\xi)$	matrix with Laguerre polynomials
$\{Y_k\}$	set of orthonormal coefficients of a ℓ_2 function y
T_n	shift matrix of Laguerre polynomials
$\underline{\psi_n}(t)$	vector of the first (n+1) Laguerre polynomials at time t.
$\pi_\xi(z)$	Laguerre system transformation
$A_t(z^{-1})$	polynomial in z^{-1}, working on time domain functions
$A\ell(z^{-1})$	polynomial in z^{-1}, working on Laguerre domain functions

CHAPTER 5

A_e, B_e	extended matrices, induced by the matrices A and B
F, X, P	matrices induced by a matrix A
A_N, B_N	finite versions of A_e and B_e
$\Psi_e\{A,B\}$	set of orthogonal functions, induced by the matrices A and B
$\psi_i\{A,B\}$	an element of $\Psi_e\{A,B\}$
$\Phi_e\{A,B\}$	set of grouped orthogonal functions, induced by the matrices A and B
$\varphi_i\{A,B\}$	an element of $\Phi_e\{A,B\}$
$H_k(z)$	generating transfer function of $\varphi_k\{A,B\}$
$\Phi\{\mathcal{B}_2^+(G)\}$	transform of $\mathcal{B}_2^+(G)$ under $\Phi_e\{A,B\}$
$W_t(z)$	time domain system
$W_o(z)$	system in the orthonormal domain
R_o	system matrix in the orthonormal domain
H_o	output matrix in the orthonormal domain
G_o	input matrix in the orthonormal domain
L_o	feedthrough matrix in the orthonormal domain

STELLINGEN

behorende bij het proefschrift

ON APPROXIMATE SYSTEM IDENTIFICATION WITH SYSTEM BASED ORTHONORMAL FUNCTIONS

van

Peter Heuberger

1. Gezien de veelbelovende mogelijkheden van het gebruik van orthonormale functies bij het benaderend modelleren van dynamische processen verdient dit onderwerp meer aandacht en onderzoek.

   Dit proefschrift.

2. Voor elk causaal, eindig dimensionaal, lineair, tijdinvariant, discrete tijd systeem bestaat er een basis van de signaalruimte, dusdanig dat in deze basis alleen de eerste twee Markov–parameters van het systeem ongelijk aan nul zijn.

   Dit proefschrift.

3. Bij de ontwikkeling van benaderende modellen dient in grote mate rekening te worden gehouden met het gebruiksdoel van deze modellen. De geschiktheid van een model hangt sterk af van de toepassing. Het verdient daarom aanbeveling om doelgerichte identificatie–criteria te ontwikkelen.

4. Het ontwerp van een observer voor een stochastisch proces op basis van een H_∞ criterium, zoals dat in [1] gepresenteerd wordt, is afhankelijk van de gebruikte realisatie van het stochastisch proces in die zin dat verschillende realisaties van hetzelfde proces tot essentieel verschillende observers leiden.

   [1] Bernstein D.S. and W.M. Haddad. Steady–state Kalman filtering with an H_∞ error bound. *System & Control Letters 12* (1989), 9–16.

5. De door het ministerie van WVC uitgevaardigde richtlijn dat nieuw aan te stellen ambtenaren, resorterend onder dit ministerie, niet meer verhuisplichtig gesteld worden, strookt niet met het overheidsbeleid om mensen dichter bij hun werk te laten wonen en om op die wijze het milieu minder zwaar te belasten. Gezien het feit dat het Rijks Instituut voor Volkgezondheid en Milieuhygiene onder dit ministerie resorteert wordt daardoor de geloofwaardigheid van dit instituut in diskrediet gebracht.

6. Gezien de voortdurende verwarring over de interpretatie van de reclameregels die door het Commissariaat van de Media gehanteerd worden, is het interessant om de reactie van deze instellling af te wachten indien sponsors van sportverenigingen hun produkten benoemen met de naam van de vereniging.

7. De publicatie van een stelling uit een proefschrift in een dagblad is geen aanwijzing voor de kwaliteit van het proefschrift noch voor die van de krant.

8. De dreigende ontwrichting van de wereldeconomie ten gevolge van de inval van Irak in Koeweit motiveert de toekenning van uitgebreide financiele en technische hulp aan landen met grote olievoorraden, zoals de Sovjet–Unie, met als doel het stimuleren van de olieproductie in die landen en het verminderen van de afhankelijkheid van de olieproductie in het Midden–Oosten.

9. Het fileprobleem in Nederland kan deels worden opgelost door (toekomstige) arbeidstijdverkorting niet meer te vertalen in vrije (ATV) dagen, maar in een daadwerkelijke vermindering van het aantal werkuren per week, waarbij de resulterende vrije uren evenredig over werknemers en spitsuren verdeeld dienen te worden.

10. Het groeiende aantal promovendi en assistenten in opleiding bij de universiteiten vraagt om de aanstelling van meer medewerkers om deze mensen te begeleiden en om de invoering van een eenvoudigere promotie–procedure.

11. Shell hielp.

Dit proefschrift.

SAMENVATTING

In diverse takken van wetenschap wordt gebruik gemaakt van wiskundige modellen voor het beschrijven van dynamische processen. In dit proefschrift wordt nader ingegaan op de problematiek van het ontwikkelen van dergelijke modellen. Er wordt daarbij onderscheid gemaakt tussen de situatie dat een bestaand model vereenvoudigd dient te worden (modelvereenvoudiging) en de situatie dat een model gecreëerd dient te worden aan de hand van meetgegevens van het proces (systeemidentificatie). Terwijl bij modelvereenvoudiging altijd sprake zal zijn van een benadering, wordt er heden ten dage van uitgegaan dat dit ook geldt voor identificatie: de werkelijkheid is veel ingewikkelder dan beschreven kan worden door een eenvoudig lineair, tijdinvariant, eindig dimensionaal model. Dit toont aan dat de problemen van modelvereenvoudiging en systeemidentificatie duidelijk aan elkaar gerelateerd zijn; beide zijn gericht op het vinden van een (optimaal) benaderend model van een dynamisch proces. De methodieken die in de literatuur worden aangedragen om deze problemen aan te pakken zijn vrijwel niet aan elkaar gerelateerd en vaak is het niet geheel en al duidelijk wat of hoe er benaderd wordt. In onze optiek zou het wenselijk zijn om een situatie te creëren, waarbij een veelheid van methodieken beschikbaar is om beide problemen op vergelijkbare wijze te behandelen, dusdanig dat de grondslag en het resultaat van de benadering op inzichtelijke wijze te doorgronden is.

Met de resultaten, als gepresenteerd in dit proefschrift, wordt een bijdrage geleverd aan de totstandkoming van voornoemde situatie. Daartoe wordt enerzijds aandacht besteed aan het probleem van de modelvereenvoudiging en anderzijds aan het gebruik van orthogonale functies binnen modelbeschrijvingen en systeemidentificatie. Het gebruik van orthogonaliteit is gemotiveerd door het feit dat dit begrip sterk verbonden is met het principe van benadering, zoals bijvoorbeeld bij projecties. Het achterliggende idee is dat, indien het mogelijk is om te ontdekken of een modelvereenvoudigingsmethodiek gebaseerd is op een bepaald orthogonaliteitsprincipe, dit aanleiding kan geven tot specifieke orthogonale functies, die op hun beurt weer aangewend kunnen worden voor benaderende systeemidentificatie.

Op het gebied van modelvereenvoudiging wordt een methode gepresenteerd die een generalisatie is van bestaande vereenvoudigingsmethoden, gebaseerd op gebalanceerde realisaties. Met deze methode kan het frequentie–gedrag van de benaderende modellen beduidend worden verbeterd. Tevens wordt deze methode gecombineerd met fractionele beschrijvingen van systemen.

Op het gebied van orthogonale functies wordt eerst het gebruik van Laguerre polynomen voor beschrijvingen van signalen en systemen geanalyseerd. Vervolgens wordt deze klasse van polynomen uitgebreid tot een klasse van orthonormale functies die geinduceerd worden door lineaire systemen. Deze functies vormen een basis van de signaal ruimte ℓ_2 en kunnen worden toegepast voor een alternatieve beschrijving van signalen en systemen in deze basis. Indien een systeem wordt beschreven in de basis van orthonormale functies die door dat systeem zelf geïnduceerd wordt, dan heeft de alternatieve systeembeschrijving slechts twee Markovparameters. In dit geval wordt het dynamisch gedrag als het ware geabsorbeerd door de orthonormale functies. Deze 'systeem gebaseerde' functies worden geinduceerd door zogenaamde gebalanceerde realisaties, welke nauw verbonden zijn met diverse modelvereenvoudigings methodieken.
Het gebruik van klassieke identificatiemethodieken binnen deze alternatieve beschrijvingen leidt tot nieuwe methoden om het probleem van benaderend modeleren aan te pakken. Van deze methodes worden het schatten van ARX en FIR modellen in orthogonale termen nader uitgewerkt en er wordt aangetoond dat deze alternatieve methodes de resultaten van hun klassieke tegenhangers substantieel kunnen verbeteren. Het zoeken naar een optimale verzameling orthonormale functies binnen deze context is complementair aan het doel van de systeemidentificatie, i.e. het bepalen van de dynamische structuur in de meetgegevens.
We mogen concluderen dat het toepassen van orthogonale functies in combinatie met klassieke identificatiemethodieken een veelbelovende aanpak voor het systeembenaderingsprobleem oplevert.

CURRICULUM VITAE

30 april 1957	Geboren te Maastricht.
1969 – 1975	Gymnasium–B aan het Henric van Veldeke College te Maastricht.
1975 – 1983	Studie Wiskunde aan de Rijksuniversiteit Groningen. Afgestudeerd met specialisatie Systeemtheorie.
1983 – 1990	Wetenschappelijk assistent aan de Technische Universiteit Delft, Faculteit der Werktuigbouwkunde en Maritieme Techniek, vakgroep Meet– en Regeltechniek, sectie Technische Systemen.
per 1–1–1991	Werkzaam bij het Centrum voor Wiskundige Methoden (C.W.M) van het Rijks Instituut voor Volksgezondheid en Milieuhygiene (R.I.V.M) te Bilthoven.