CWI Tracts

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science
P.O. Box 4079, 1009 AB Amsterdam, The Netherlands

The CWI is a research institute of the Stichting Mathematisch Centrum, which was founded on February 11, 1946, as a nonprofit institution aiming at the promotion of mathematics, computer science, and their applications. It is sponsored by the Dutch Government through the Netherlands Organization for the Advancement of Research (N.W.O).

CWI Tract 63

Identifiability, recursive identification and spaces of linear dynamical systems part I

B. Hanzon

Centrum voor Wiskunde en Informatica
Centre for Mathematics and Computer Science

1980 Mathematics Subject Classification: 49E, 49F, 93B, 93D, 93E.
ISBN 90 6196 371 0
NUGI-code: 811

PREFACE.

In the course of the study that led to this book I have had the help of many people, of which I can name only a few.

In 1979 it all started at the Department of Mathematics of the Econometric Institute, Erasmus University Rotterdam, where I worked with Michiel Hazewinkel. We started a seminar on Time Series, which resulted, among other things, in discussions and research concerning the so-called 'finite identifiability problem'. The results can be found in Chapter 3.

In 1980 Michiel Hazewinkel and Jan Willems organized a conference/summer school at Les Arcs in the French Alps, about stochastic systems, in particular the mathematical aspects of filtering and identification. There I met many important researchers in the field of systems theory. Among them Steve Marcus.

In 1982 my wife and I spent almost a year in Austin, where I worked together with Steve Marcus at the University of Texas. I investigated a Riemannian metric for the space of stable linear systems. This is described in Chapter 5. During the latter part of our stay in Austin, the question was raised whether the Riemannian metric could also be used in an algorithm for recursive identification and if so, what the properties of such an algorithm would be. At first this did not seem to give many difficulties and some months after our return from Texas, I sent a sketch of the construction of the algorithm, the asymptotic properties and their proof to Steve Marcus. But working this out turned out to be a constantly expanding task. This can be retraced in the length of Chapter 6 in which the algorithm and the results about the asymptotic properties are described.

In 1984 I moved to the Department of Mathematics and Informatics of the Technical University of Delft, to work with Geert-Jan Olsder. There I was enthusiastically supported to finish this study.

I want to thank all the people who have helped me, those mentioned and all the others. Special thanks go to Mrs. J.G.E.G. Oosthout and Mrs. C.E. Hamerslag of the Econometric Institute of the Erasmus University for all the work that they have done, and the 'Vakgroep Wiskunde' for their support. Special thanks also go to Mr. P.J. van der Meer for making the figures. Finally I want to thank my wife and my family for their constant moral and concrete support.

TABLE OF CONTENTS

Part I

Part II

CHAPTER 1

INTRODUCTION

In many sciences, like econometrics, engineering science and many others, model identification is an important problem. It can be sketched as follows. Using a priori knowledge, theoretical considerations and often some amount of speculation, the investigator proposes a 'model' for the phenomenon under consideration. In most cases this 'model' is a rough structure containing a number of unknown quantities. This means that one has in fact not one model, but a whole set of models; with each choice for the unknown quantities corresponds a fully specified model. This set of models will simply be called the model set. The problem of model identification is to identify the correct model from the model set, using observations concerning the phenomenon. The question arises whether the model can be identified from the observations. If so we will say that we have model identifiability (cf. chapter 3). If not, then there are several possible reasons.

The first one is that the correct model is not in the model set, i.e. the phenomenon under consideration is not well-described by any of the models in the model set. One should always be aware of this possibility of falsification of the model set. A good model identification method should include some kind of falsification test. The second possible reason is that the available observations do not suffice; one needs more observations of some quantities and/or observations of other quantities. The third possible reason is that several models are observationally equivalent. This means that although the models are apparently different, they describe exactly the same relations between all observable variables of the phenomenon. The approach taken here will be to consider the relationships between the observable variables as essential and observationally equivalent models will be considered as different representations of the same model. In other words, a model will be considered to be an abstract object that describes the relations between (in principle) observable variables, and that can be represented in many different ways.)* This is in fact one of the basic themes in this tract. This can be seen by considering the main points of this work.

(1) In chapter 2 (among other things) various different representations of linear dynamical systems, deterministic and stochastic, are presented and

)* A fourth possible reason is that several models are 'partially observationally equivalent'. By redefining the model set this possibility can be ruled out (cf. p.103 esp. (3.2.1.1-14)).

their interrelations are described.

(2) In chapter 3 a problem that is originally stated in terms of an a.r.m.a. representation of a linear dynamical system namely the socalled 'finite identifiability problem' (also called 'the informative sample size problem'), is treated by using i.a. a state-space representation, a Hankel matrix representation and a transfer matrix representation of the system. In this way, results are found that appear difficult to obtain directly from the a.r.m.a. representation.

(3) Because models are considered as abstract objects, if one needs structure on the model set, like a topological structure and a metric structure, one has to impose this. This means that one has to decide when models are close and when far apart, in a qualitative sense (a topology) and in a quantitative sense (a metric). For the classes of linear dynamical systems that are considered (cf. chapter 2) these subjects are treated in chapter 4, where an overview of some of the existing results of the topology of spaces of linear systems is given and in chapter 5 where for several spaces of linear systems a Riemannian metric is presented and analyzed.

(4) To describe the model set, parameters will be used. In our set-up they just play the role of coordinates. They should not be confused with what could be called the 'reaction coefficients' of a model. To give an example of this, consider the following situation. Suppose that in some model there are variables x and y and if x changes, then according to the model, y changes with some factor times the change of x. This factor ia a 'reaction coefficient'. Such coefficients may of course be important in the analysis of the behaviour of the phenomenon. Therefore, often they are used as parameters. However, one has to be careful in doing this. For example, (a) perhaps in some situations x cannot be changed, while y can; then mathematically speaking the reaction coefficient may be $\pm \infty$; (b) there may be an infinite number of relevant reaction coefficients in the model; (c) often there are restricting relations between the various reaction coefficients, which makes that they cannot be identified (or estimated) independently; (d) there may be problems with the causality, for example the model $y = 2x$ may have a perfect fit in some situation, but if x is controlled to increase by one, y may increase not by two, but by one, the reason being that there is a third variable, z, and the true model is

$$y = x+z$$
$$x = z+u$$

where u is the control (which is supposed to be zero if x is not controlled). This all shows that one has to be very careful with using the reaction coefficients as parameters. To describe the model set with the aid of parameters, two approaches can be distinguished. The first one is to use some canonical form for the model. This means that the model is written down in such a (canonical) way that any two models which are (observationally) equivalent have the same canonical form (and of course models which are not equivalent have different canonical forms). One typically has a number of discrete variables (often their only possible values are one and zero) and a number of continuous variables (e.g. taking their values in some open subset of a Euclidean space) that describe the canonical form. The advantage of a canonical form is that each model is described in a unique way. The disadvantage in many cases (including the case of multivariable linear dynamical systems) is that it is discontinuous, i.e. a small change in the model may lead to a completely different canonical form. This difficulty is avoided in the second approach, which consists of using a number of overlapping parametrizations. In this case the model space is covered by a number of overlapping open sets, called charts and each of them is provided with (continuous) coordinates (the 'local parameters'). (As a simple example of a covering with charts - from which in fact the terminology stems - one can think of a complete atlas of the earth, inclusing a chart which contains the north pole and one that contains the south pole). To cover a space in such a way is a standard method in differential geometry, leading to the theory of differentiable manifolds. This way of parametrizing the model set will be used in chapter 6. There a generalization of the well-known prediction error algorithm for recursive identification of stochastic linear systems, to the case in which the model set is a differentiable manifold (with certain properties) is constructed and its asymptotic behaviour is analyzed.

Now let us go through the contents in some more detail. In chapter 2 an introduction is given to deterministic and stochastic linear (dynamical) systems. The chapter starts with a general definition of a deterministic dynamical system with initial conditions. It is shown that under certain conditions, up to scaling, there are only five possibilities for the 'time axis', namely $\mathbb{R}, \mathbb{Z}, [0,\infty)$, $\mathbb{N}_0 (\equiv \mathbb{N} \cup \{0\})$ and $\{0\}$. Special care has been taken to give a good treatment of 'initial conditions at $-\infty$'. Stochastic systems will be represented as deterministic systems with stochastic inputs (and possibly stochastic initial

conditions; in case of 'initial conditions at $-\infty$' this would lead to non-ergodicity, which will be ruled out). Starting with the general definition of a dynamical system, and adding assumptions, the concept of a deterministic, finite dimensional linear system is obtained. The various representations of such a system are treated, and their interrelations investigated. An important role in the theory of multivariable linear dynamical systems is played by the Kronecker indices. An important partial order on these indices is the specialization order. It is defined and it is shown to have the property, that in limit points of the model space the Kronecker indices of the system can only 'drop' (i.e. become 'more special') or remain the same. The results given here are somewhat more general than in the literature.

For stochastic linear systems various representations are treated also. Special care has been taken to show that they are well-defined with probability one. The spectral factorization theorem is presented; it plays a role in chapter 3. The same holds for the Hankel matrices of covariances.

In chapter 3 arma(p,q) models are investigated. To obtain results about the unique partial realization problem for such models, a characterization of such models in terms of the rank structure of the corresponding Hankel matrix is given. The results and proofs in the deterministic case are for a great deal parallel to those in the stochastic case. The results that are obtained for the unique partial realization problem in the stochastic case are applied to the finite identifiability problem, which stems from econometrics. However, to be able to get a good interpretation of the results, a concise treatment is given of the methodology of the identifiability problem. Among other things two concepts, called system identifiability and parameter identifiability are introduced which hopefully will contribute to a somewhat better understanding of identifiability problems. The results that are obtained about the finite identifiability problem are for a large part stronger than those in the literature.

Chapters 4, 5 and 6 have a different flavour - compared to chapters 2 and 3 - due to the fact that they make use of concepts and results of topology and differential geometry. In chapter 4 an overview is given of some of the known results about the differential geometric and topological structure of certain families of linear systems. A central role is played by the socalled state (fibre) bundle. This is a fibre bundle, of which each fibre - as a whole - corresponds to a linear dynamical system, and each point of the fibre in turn

corresponds to some state of that linear dynamical system. Some of the connectedness properties that are presented are shown here for the first time.
In chapter 5 a new element is added to the structure of the model space, namely a metric. Or, in fact, two metrics. One that will be called the outer metric, and a corresponding one, the inner metric. If a space is embedded in a metric space, then it inherits its metric. This will be called the outer metric. (Think e.g. of a unit sphere in three dimensional Euclidean space. In the outer metric the distance between north pole and south pole is two). However, if one considers the 'travelling distances within the space' one obtains another metric, the socalled inner metric. (In our example of the unit sphere, in the inner metric the distance between north pole and south pole, i.e. the 'travelling distance' between those points, is π). In general to compute the inner metric globally is a difficult problem. Locally on a differentiable manifold, however, it can very often be computed. One then works with a so-called Riemannian metric, which is in fact an (inner product-) metric on each tangent space of the differentiable manifold. The advantage of the outer metric/inner metric approach is that one can choose an intuitively natural embedding of the model set in a metric space (often a Euclidean space or, more generally, a Hilbert space) and compute the corresponding inner metric. In the case of deterministic asymptotically stable linear time-invariant systems one can embed the model set in the set of all asymptotically stable 'm.a.(∞)' systems, and use the square-root of the sum of squares of all m.a. (∞)-coefficients as norm of a system. This approach is worked out in chapter 5. The corresponding Riemannian metric is derived and for several cases it is explicitly computed. For the simplest case (s.i.s.o., order one) the inner metric is computed. This turns out to be (isometric to) a double infinite-sheeted Riemann surface. Special attention is being paid to the 'short time-interval' case. A representation is found for discrete-time systems with time-interval length Δ such that for $\Delta \downarrow 0$ the representation becomes the continuous time representation, while for $\Delta = 1$ one has the usual discrete-time representation. The representation has elegant properties concerning its stability region in the complex plane. Surprisingly enough, as a by-product of this analysis, an interesting group of isometries of the model space is found. In the simplest case, alluded to above, it corresponds to the group of isometries of the Riemann surface!
Especially for practical purposes it is important to know at which points the Riemannian metric tensor degenerates. This is not a property of the space but

of the chosen parametrization! It is shown that for a usual parametrization of linear systems, the Riemannian metric tensor degenerates at points where the McMillan degree drops.
The Riemannian metric on a manifold of linear systems can be extended to a Riemannian metric on the state bundle in such a way that if one uses socalled parallel displacement of the state the resulting (inner) distance between two systems does not depend on the state, and is equal to the (inner) distance defined before.
Turning to the stochastic case, it is first shown that the outer metric/inner metric/Riemannian metric structure can also be found for the Fisher metric (which is a Riemannian metric). The corresponding outer metric is the socalled Hellinger distance, which is defined between any pair of measures on the same space. The Fisher metric on spaces of stationary stochastic systems is defined in a standard way, namely by dividing the Fisher information matrix for any time interval by the time interval length T and taking the limit for $T \to \infty$. The relationship between Fisher metric and Hellinger distance is lost in the limit.
The chapter is concluded with some remarks on the relationship between local identifiability and nondegeneracy of the Riemannian metric tensor.

Due to the size of the work it comes in two parts, part 1 consisting of chapters 1-5 and part 2 consisting of the long chapter 6 (and the references, etc.).
Chapter 6 is about Riemannian gradient algorithms for recursive identification. As mentioned before it contains a generalization of the prediction error algorithm for recursive identification of stochastic linear systems to the case in which the model set is a differentiable manifold. This differentiable manifold is required to be a compact submanifold of the manifold of stationary stochastic systems of order n, which have an innovations representation with asymptotically stable inverse. In a long analysis, which i.a. makes use of the compactness of certain sets of interpolation curves, it is shown that the parameter sequence that is generated by the algorithm converges to a critical point of the criterion function, or, at least, to a connected set of critical points (on which the criterion function is constant). A more detailed introduction to chapter 6 is given in section 6.1. at the beginning of part 2. Chapter 6 ends with a number of remarks and possible subjects for further research.

CHAPTER 2

INTRODUCTION TO DETERMINISTIC AND STOCHASTIC LINEAR SYSTEMS

2.1. Definitions in the deterministic case

In this section we want to define the class of deterministic, discrete time, time invariant, finite dimensional, linear dynamical systems. To do that we start with a general definition of dynamical systems and then specialize to the desired case. The following definitions are inspired by those in [Wi 79]. We will consider a system to be a set of input-output mappings parametrized by the possible "initial conditions". To clarify the role of continuous and discrete time we will show that discrete time and continuous time systems arise as the only possibilities from certain axioms with respect to the time axis of the system.

2.1-1. Definition. A dynamical system is defined by $\Sigma = \{T, U, \overset{\circ}{U}, Y, \overset{\circ}{Y}, \overset{\circ}{F}\}$, where

(i) $T \subseteq \mathbb{R}$ is the time axis

(ii) U is an arbitrary set, the set of input values called the input alphabet,

(iii) $\overset{\circ}{U}$ is a set of functions, the input functions, $\omega: T \to U$, with the following property: if $\omega_1, \omega_2 \in U$, $t' \in T$, and ω is the concatenation of ω_1 and ω_2, defined by

$$\omega(t) := \begin{cases} \omega_1(t) & \text{if } t < t', \\ \omega_2(t) & \text{if } t \geq t', \end{cases}$$

then $\omega \in \overset{\circ}{U}$;

(iv) Y is an arbitrary set, the set of output values, called the output alphabet;

(v) $\overset{\circ}{Y}$ is a set of functions, the output functions, $\eta: T \to Y$,

(vi) $\overset{\circ}{F}$ is a set of nonanticipative input-output mappings. This means the following: the set can be described as $\overset{\circ}{F} = \{F_b \mid b \in B\}$, B an index set, called the set of initial conditions and

$\forall b \in B: F_b: \overset{\circ}{U} \to \overset{\circ}{Y}$ is a nonanticipative mapping, i.e.

if (a) $\omega_1, \omega_2 \in \overset{\circ}{U}$, $t' \in T$ and if

(b) $\forall t \leq t'$, $t \in T$: $\omega_1(t) = \omega_2(t)$,

then

(c) $(F_b\omega_1)(t') = (F_b\omega_2)(t')$

F_b is called an input-output mapping corresponding to the initial condition $b \in B$. The function $F: B \times \overset{\circ}{U} \to \overset{\circ}{Y}$, $(b,w) \longmapsto F_b(w)$ is called the system function.

2.1-2. Remarks.

(i) This definition resembles the definition of a dynamical system in input-output form as given by [Wi 79].

(ii) Often a system can be represented in many different ways. This holds especially for the system function F. In many cases it is defined implicitly by some representation of the system, usually a recursive equation like a difference or differential equation. This fact will play a central role in this work. If two representations correspond to the same system Σ (in the sense that they generate the same set of input-output mappings $\overset{\circ}{F} = \{F_b | b \in B\}$, while their sets $T,U,\overset{\circ}{U},Y,\overset{\circ}{Y}$ are also equal) they will be called equivalent.

Often it is convenient to have an explicit representation of the initial conditions in the definition of a system:

2.1-3. Definition. A dynamical system with set of initial conditions (s.i.c.) is defined by $\Sigma = \{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},B,F\}$, B the set of initial conditions, F the system function.

It should be clear from the foregoing that two systems with s.i.c. correspond to the same system, if they generate the same set $\overset{\circ}{F} = \{F_b | b \in B\}$, while their sets $T,U,\overset{\circ}{U},Y,\overset{\circ}{Y}$ are equal.

Let us now introduce linearity.

2.1-4. Definition. Consider the system with s.i.c. $\Sigma = \{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},B,F\}$. If $U,\overset{\circ}{U},Y,\overset{\circ}{Y}$ and B are vector spaces over some field k (with pointwise addition and scalar multiplication in $\overset{\circ}{U}$ and $\overset{\circ}{Y}$) and if F is a linear mapping, then Σ is called a linear system with s.i.c. A system $\Sigma = \{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},\overset{\circ}{F}\}$ is called a linear system if there exists a corresponding linear system with s.i.c.

To define time-invariance of systems, we need a shift-invariance structure on the time axis:

2.1-5. Definition. Suppose $\delta \in \mathbb{R}$ is such that $T+\delta \subseteq T$, i.e. $\forall t \in T$: $t+\delta \in T$. Then the shift S_δ is defined by $(S_\delta\omega)(t) = \omega(t+\delta)$ for any function ω with domain T.

2.1-6. Definition. Let $\Delta^+ := \{\delta \mid \delta = t_1-t_2 > 0;\ t_1,t_2 \in T\}$ be the set of positive time differences in T. T will be called forward shift invariant if

$$T + \Delta^+ \subseteq T.$$

(Notation: $A+B = \{a+b \mid a\in A, b\in B\}$).
Also, if $\overset{\circ}{V}$ is any set of functions on T, and $\forall\delta \in \Delta^+$: $S_\delta\overset{\circ}{V} \subseteq \overset{\circ}{V}$, then $\overset{\circ}{V}$ is called forward shift invariant.
Let

$\Delta := \{\delta \mid \delta = t_1-t_2; t_1,t_2 \in T\}$ be the set of all time differences in T. T will be called shift-invariant if $T+\Delta \subseteq T$. ($\leftrightarrow T+\Delta=T$). Also if $\overset{\circ}{V}$ is a set of functions on T and $\forall\delta \in \Delta$: $S_\delta\overset{\circ}{V} \subseteq \overset{\circ}{V}$ ($\leftrightarrow S_\delta\overset{\circ}{V} = \overset{\circ}{V}$), then $\overset{\circ}{V}$ is called shift invariant.

2.1-7. Remark. Shift invariance of the time axis implies that for a space W, the set of all functions W^T is shift invariant.

Now we can introduce the concepts of forward time invariance and time invariance:

2.1-8. Definition. Let the system Σ (with s.i.c.), be as before. Suppose T is forward shift invariant and $\overset{\circ}{U}$ and $\overset{\circ}{Y}$ are forward shift invariant. If there exists a one parameter semi-group of transformations $\{T_\delta : B \to B \mid \delta \in \Delta^+\}$ on B such that $T_{\delta_1} \circ T_{\delta_2} = T_{\delta_1+\delta_2}$ and

$$\forall\delta \in \Delta^+\ \forall b \in B: S_{-\delta} \circ F_b = F_{T_\delta(b)} \circ S_{-\delta}$$

then Σ is called forward time invariant. (A system is called forward time invariant if there exists a corresponding system with s.i.c. that is forward time invariant).

A similar definition can be given for a time invariant system:

2.1-9. Definition. Let Σ be as before. Suppose T is shift invariant, and $\overset{\circ}{U}$ and $\overset{\circ}{Y}$ are shift invariant. If there exists a one parameter group of transformations $\{T_\delta: B \to B \mid \delta \in \Delta\}$ on B such that ($T_{\delta_1} \circ T_{\delta_2} = T_{\delta_1+\delta_2}$ and)

$$\forall \delta \in \Delta, \forall b \in B: S_\delta \circ F_b = F_{T_\delta(b)} \circ S_\delta$$

then Σ is called time invariant. (A system is called time invariant if there exists a corresponding system with s.i.c. that is time invariant).

2.1-10. Remark. Usually the distinction between forward time invariance and time invariance is not made explicitly, and only the term time invariance is used. If there is no chance of confusion we will follow that practice.

Next we want to make the distinction between continuous time and discrete time systems and address the problem of the representation of T, assuming T is forward shift invariant and closed.

2.1-11. Definition. Let Σ be as before and let $T \neq \emptyset$ be forward shift invariant and closed in $\mathbb{R}$. The system Σ is called
(a) a static system if $\Delta^+ = \emptyset$, i.e. $|T| = 1$,
(b) a continuous time system if $\inf \Delta^+ = 0$, and
(c) a discrete time system if $\inf \Delta^+ > 0$.

We will now show that by a translation and a positive scaling factor, T can be brought into one of five standard forms.

2.1-12. Theorem. Let Σ be as before and $T \neq \emptyset$ forward shift invariant and closed. Then there is a real number t_o and a positive real number λ such that $\lambda.(T-t_o)$ is one of the following five sets

(a) $\{0\}$
(b.1) $\mathbb{R}$
(b.2) $[0,\infty)$
(c.1) $\mathbb{Z}$
(c.2) $\mathbb{N}_o$ ($:= \mathbb{N} \cup \{0\}$);

(a) corresponds to the static case,

(b) to the continuous time case and (c) to the discrete time case.

Proof. Without loss of generality we can assume $T \subseteq \Delta$, if not then take $t_o \in T$ and translate the time axis to $T-t_o \subseteq \Delta$. In the static case, this already leads to the standard form: $T = \{0\}$. From now on assume $\Delta^+ \neq \emptyset$. First let us show that $\Delta + \Delta = \Delta$:

(1) $0 \in \Delta$, so $\Delta + \Delta \supseteq \Delta$

(2) $T+\Delta^+ \subseteq T$, so if $\delta_1,\delta_2 \in \Delta$ and $t_1,t_2 \in T$ such that $\delta_1 = t_1-t_2$ (such t_1,t_2 can of course always be found), then $\delta_1+\delta_2 = t_1-t_2+\delta_2$; there are now three possibilities:

(i) $\delta_2 = 0$

(ii) $\delta_2 \in \Delta^+$

(iii) $-\delta_2 \in \Delta^+$;

in case (i), it follows that $\delta_1+\delta_2 = t_1-t_2 \in \Delta$; in case (ii), $\delta_1+\delta_2 = (t_1+\delta_2)-t_2$, and $t_1+\delta_2 \in T$, so $(t_1+\delta_2)-t_2 \in \Delta$, and in case (iii), $\delta_1+\delta_2 = t_1-(t_2-\delta_2)$, and $t_2-\delta_2 \in T+\Delta^+ \subseteq T$, so $t_1-(t_2-\delta_2) \in \Delta$. In all three cases it follows that $\delta_1+\delta_2 \in \Delta$. So $\Delta + \Delta \subseteq \Delta$. (1) and (2) imply $\Delta + \Delta = \Delta$.

If $t \in T$ then $\forall t' > t$ with $t' \in \Delta$, $t'-t \in \Delta^+$. It follows that $t' \in T$. So if $t \in T$, then $\Delta \cap [t,\infty) \subseteq T$. Consider inf T. If inf $T = -\infty$, then it follows that $T = \Delta$. If inf $T = t_o > -\infty$, then $T \neq \Delta$. Note that if $\delta \in \Delta \backslash T$, then $T \subseteq \Delta \cap (\delta,\infty)$. Because T is closed, $t_o \in T$. By the translation $t \to t-t_o$, T is mapped to $T-t_o = \Delta \cap [0,\infty) = \{0\} \cup \Delta^+$. So by applying a translation $t \to t-t_o$ one can map T to Δ or $\{0\} \cup \Delta^+$. Because T is closed, so is $\{0\} \cup \Delta^+$ and so is Δ $(= -\Delta^+ \cup \{0\} \cup \Delta^+)$. If inf $\Delta^+ = 0$ then it follows that Δ is dense in $\mathbb{R}$ and so $\Delta = \mathbb{R}$. This gives rise to two possibilities (b.1) $T-t_o = \mathbb{R}$; (b.2) $T-t_o = [0,\infty)$. If inf $\Delta^+ > 0$, let $\lambda := (\text{inf } \Delta^+)^{-1} > 0$, and multiply the elements of $(T-t_o)$ with λ. If $T-t_o = \Delta$ it follows that $\lambda(T-t_o) = \mathbb{Z}$ (c.1) and if $T-t_o = \{0\} \cup \Delta^+$ then it follows that $\lambda(T-t_o) = \mathbb{N}_o = \mathbb{N} \cup \{0\}$.

Q.E.D.

2.1-13. Convention. From now on we will exclude the static case; i.e. $|T| > 1$ is assumed, unless the opposite is stated explicitly.

One of the fundamental notions in system theory is the concept of the state of

a system, and of a state space representation of a system. This will be defined next. It will also lead us to the definition of a finite dimensional (finite order) system. The main difference with a similar definition in [Wi 79] is that more emphasis is put on the role of the initial conditions. To be able to cope with all forms of T, especially $T = \mathbb{R}$ and $T = \mathbb{Z}$, the following constructions have to be made: $\omega_1, \omega_2 \in \overset{\circ}{U}$ are called 'equal in the past' if $\exists t \in T$ such that $\omega_1|_{<t} = \omega_2|_{<t}$. Clearly 'equality in the past' is an equivalence relation. If $\inf T = t_o > -\infty$, then the equivalence relation is trivial in the sense that all inputs $\omega \in \overset{\circ}{U}$ are equivalent; if $\inf T = -\infty$ then the equivalence classes are germs of functions $\omega \in \overset{\circ}{U}$ 'in $-\infty$'.
Let $\overset{\circ}{U}_{\inf T}$:= the set of equivalence classes in $\overset{\circ}{U}$ under this equivalence relation.
Let X be a set and let $\overset{\circ}{X}$:= the set of time functions $\xi: T \to X$; $\xi_1, \xi_2 \in \overset{\circ}{X}$ are called 'initially equal' if $\exists t \in T$ such that $\xi_1|_{\leq t} = \xi_2|_{\leq t}$. Clearly 'initial equality' is an

equivalence relation. Let $\overset{\circ}{X}_{\inf T}$ = the set of equivalence classes in $\overset{\circ}{X}$ under this equivalence relation.
So if $\inf T = t_o > -\infty$, then $\overset{\circ}{X}_{\inf T} \simeq \{\xi(t_o) \in X\} = X$, the state space; and if $\inf T = -\infty$, then $\overset{\circ}{X}_{\inf T}$ is the set of germs of functions in $\overset{\circ}{X}$ 'at $-\infty$'.
Notation: If A denotes a set of sets: $A = \{a | a \in A\}$, then $x \in\in A$ denotes $x \in \cup_{a \in A} a$; i.e. x is an element of an element of A. In the case that A is a set of equivalence classes, $x \in\in A$ means that x is a representative of an equivalence class $[x] \in A$.

2.1-14. Definition. A state space representation of a system with s.i.c. $\Sigma = \{T, U, \overset{\circ}{U}, Y, \overset{\circ}{Y}, B, F\}$ is given by $\{\Sigma, X, \phi, r, \beta\}$, where

(i) X is a set called the state space,

(ii) ϕ is called the state evolution function; it is a function $\phi: \{(t_1, t_o) \in T^2 | t_1 \geq t_o\} \times X \times \overset{\circ}{U} \to X$, and satisfies the following conditions:

(ii.1) $\phi(t, t, x, \omega) = x$ (consistency),

(ii.2) $\phi(t_2, t_1, \phi(t_1, t_o, x_o, \omega), \omega) = \phi(t_2, t_0, x_o, \omega)$ (semigroup property),

(ii.3) if $\omega_1, \omega_2 \in \overset{\circ}{U}$ and for some t_o, t_1 with $t_o < t_1$:
$\forall t \in [t_o, t_1) \cap T: \omega_1(t) = \omega_2(t)$, then
$\forall x_o \in X: \phi(t_1, t_o, x_o, \omega_1) = \phi(t_1, t_o, x_o, \omega_2)$ (determinism),

(iii) r is called the read-out map, it is a map $r: X \times U \times T \to Y$

(iv) β is called the initial condition mapping.
The initial condition mapping is a
mapping $\beta: B \times \overset{\circ}{U}_{\text{inf } T} \to \overset{\circ}{X}_{\text{inf } T}$ with the following property:

(iv.1) if $\omega \in\in \overset{\circ}{U}_{\text{inf } T}, \xi \in \beta(b,[\omega])\left(\in \overset{\circ}{X}_{\text{inf } T}\right)$, then $\exists t \in T$ such that

$\forall t_2,t_1$ with $t \geq t_2 > t_1$ and $t_2,t_1 \in T$:

$$\phi(t_2,t_1, \xi(t_1), \omega) = \xi(t_2)$$

(compare (ii.3).

(v) the system mapping

$F: \overset{\circ}{U} \times B \to \overset{\circ}{Y}$, $(\omega,b) \longmapsto \eta = F(\omega,b)$, is determined by the following condition

(v.1) $\exists \xi \in \overset{\circ}{X}$ such that
(I) $\forall t_1,t_2$ with $t_2 > t_1$ and $t_2,t_1 \in T$: $\phi(t_2,t_1,\xi(t_1),\omega) = \xi(t_2)$,
(II) $\forall t \in T$: $\eta(t) = r(\xi(t),\omega(t),t)$ and
(III) $\xi \in \beta(b,[\omega])$.

(End of definition)

2.1-15. Proof of well-definedness of (2.1-14).
It has to be shown that F is well-defined by (2.1-14) (v). Let $b \in B$, $\omega \in \overset{\circ}{U}$ be arbitrary. Choose any $\xi' \in \beta(b,[\omega])(\in \overset{\circ}{X}_{\text{inf } T})$. Then $\exists t' \in T$ such that $\forall t_2,t_1$
with $t_2 < t_1$ and $t_2,t_1 \in T$ and $t_2 \leq t'$: $\phi(t_2,t_1,\xi'(t_1),\omega) = \xi'(t_2)$.
Define $\xi(t)$ as follows:

$$\xi(t) = \begin{cases} \xi'(t) \text{ if } t \leq t', \\ \phi(t,t',\xi'(t'),\omega) \text{ if } t \geq t'. \end{cases}$$

This shows the existence of ξ, and therefore of η (see (v.1)II).
Now let us prove uniqueness. Let ξ'',ξ''' both satisfy (v.1). Then $\exists t''$ such that $\forall t \leq t''$, $t \in T$:

$$\xi''(t) = \xi'''(t).$$

Furthermore $\forall t \geq t''$:
$\xi''(t) = \phi(t,t'',\xi''(t''),\omega) = \phi(t,t'',\xi'''(t''),\omega) = \xi'''(t)$ because of (v.1)I.
Therefore $\xi'' = \xi'''$. So ξ and therefore η, is uniquely determined
by $b \in B$, $\omega \in \overset{\circ}{U}$ and (v). So F is well-defined. Q.E.D.

2.1-16 Remark. If $T = [0,\infty)$ or $\mathbb{N}_0$, then $\overset{\circ}{U}_{\text{inf } T} = \{\overset{\circ}{U}\}$, a one-element set, and $\overset{\circ}{X}_{\text{inf } T}$ is in one-to-one correspondence with the set $\{\xi(0) \in X\} = X$, the state space. Therefore in that case β: $B \times \overset{\circ}{U}_{\text{inf } T} \to \overset{\circ}{X}_{\text{inf } T}$ can be represented by a mapping from B to X. In other words: β indicates the state $x \in X$ that corresponds to a choice $b \in B$ of initial conditions.

2.1-17. Definition. The set of initial conditions B in the system with initial conditions $\Sigma = \{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},B,F\}$ is called minimal (for the corresponding system $\{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},F\}$) if the following implication holds:

$$(b_1,b_2 \in B,\ b_1 \neq b_2) \Rightarrow (\exists \omega \in \overset{\circ}{U}\colon F(\omega,b_1) \neq F(\omega,b_2)).$$

2.1-18. Remarks. (i) It is very well possible that the minimal set B is a one-element set.
(ii) Different state space representations of the same dynamical system Σ will be called equivalent (compare remark 1.1.2. (ii)).
(iii) It is easy to see how one can obtain a minimal set of initial conditions abstractly. Define an equivalence relation $\sim$ on B by the following rule:
$b_1 \sim b_2$ iff $F_{b_1} = F_{b_2}$. The set of equivalence classes $\{[b] \mid b \in B\}$ forms a minimal set of initial conditions.

The question arises whether a state space representation always exists. We will show that this is indeed so, by constructing one, abstractly. The construction consists of three steps. In the first one a state is defined as a nonanticipative mapping of present and future inputs to present and future outputs. This first step provides us with a 'state space' X_t at each time t. In the second step the state spaces at different times are linked together by telling which states at one time should be identified with which states at another time. In the third step the corresponding state evolution function, the initial conditions mapping and the read out map are constructed. The

following notation will be used: if f and g are mappings and Dom f, Dom g their domains, and $f|_{\text{Domf}\cap\text{Domg}} = g|_{\text{Domf}\cap\text{Domg}}$, then $f \cup g$ denotes the mapping

$$x \to \begin{cases} f(x) \text{ if } x \in \text{Dom } f \\ g(x) \text{ if } x \in \text{Dom } g \end{cases}$$

1. First note that if a state space representation exists, then each state x_o at some arbitrary time $t_o \in T$ induces a mapping x of present and future inputs to present and future outputs:

   $x: \overset{\circ}{U}|_{[t_o,\infty)\cap T} \to \overset{\circ}{Y}|_{[t_o,\infty)\cap T}$, $\omega'|_{[t_o,\infty)\cap T} \longmapsto f$,

   where $f \in \overset{\circ}{Y}|_{[t_o,\infty)\cap T}$ is given by

   $\forall t_1 \in [t_o,\infty) \cap T: f(t_1) = r \circ \phi(t_1,t_o,x_o,\omega')$. The function f does not depend on $\omega'|_{(-\infty,t_o)\cap T}$, but only on $\omega'|_{[t_o,\infty)\cap T}$ because of property (ii.3) (determinism) of the definition (2.1.14) of a state space representation. Now we will turn this around and <u>define</u> a state at time t_o as such a mapping x, as follows. If the initial condition b and the inputs $\omega'|_{(-\infty,t_o)\cap T}$ up till time t_o, excluding t_o itself, are fixed, then the system function F determines a nonanticipative mapping

   $$x: \overset{\circ}{U}|_{[t_o,\infty)\cap T} \to \overset{\circ}{Y}|_{[t_o,\infty)\cap T}$$

   by

   $$\omega''|_{[t_o,\infty)\cap T} \longmapsto F[\omega'|_{(-\infty,t_o)\cap T} \cup \omega''|_{[t_o,\infty)\cap T}, b]|_{[t_o,\infty)\cap T} \in \overset{\circ}{Y}|_{[t_o,\infty)\cap T}.$$

   The set of all possible mappings x that can be obtained in this way will be denoted by X_{t_o}, and called 'the state space at time t_o'. Note that the state at time t_o is completely determined by the past inputs and the initial conditions!

2. Abstractly one can take the disjoint union $\cup_{t\in T} X_t$ as the state space, but usually this is very redundant. From now on we will concentrate on the case in which T is forward shift invariant. We will assume T is in one of the five standard forms of theorem (2.1-12). If $x \in X_t$, we can shift it to the

origin of the time axis, by considering $S_t \circ x \circ S_{-t}$, where

$S_{-t}: \overset{\circ}{U}|_{[0,\infty)\cap T} \to \overset{\circ}{U}|_{[t,\infty)\cap T}$ and $S_{+t}: \overset{\circ}{Y}|_{[t,\infty)\cap T} \to \overset{\circ}{Y}|_{[0,\infty)\cap T}$ are the shift operators on the corresponding spaces of functions (compare definition (2.1-5); if T is only forward shift invariant, then S_{-t} is formally not yet defined, but its definition is obvious). We can now define the state space as

$$X = \bigcup_{t\in T} S_t \, X_t \, S_{-t}.$$

X has the property that $X_t \subseteq S_{-t} X \, S_t$ for all $t \in T$.

3. Now the state evolution function, the initial condition mapping, and the read out map can be constructed, as follows:

$\phi(t_1,t_o,x_{t_o},\omega) = x_{t_1}$, where x_{t_1} is defined as follows:

$$x_{t_1}: \overset{\circ}{U}|_{[t_1,\infty)\cap T} \to \overset{\circ}{Y}|_{[t_1,\infty)\cap T},$$

$$\omega'|_{[t_1,\infty)\cap T} \longmapsto x_{t_o}\left[\omega|_{[t_o,t_1)\cap T} \cup \omega'|_{[t_1,\infty)\cap T}\right]\Big|_{[t_1,\infty)}.$$

The initial condition mapping is obtained as follows: For each initial condition $b \in B$ one has a nonanticipative input-output mapping $F_b: \overset{\circ}{U} \to \overset{\circ}{Y}$. Each $\omega|_{(-\infty,t)}, \omega \in \overset{\circ}{U}$, determines a state x_t at time t, as before, in 1. It follows that each $\omega \in \overset{\circ}{U}$ determines a sequence(or time function) $\{x_t\}_{t\in T}$, and each germ $[\omega]$, under the equivalence of 'equality in the past' determines a germ $\xi = [\{x_t\}]$. In this way one obtains a mapping $\beta: (b,[\omega]) \longmapsto \xi = [\{x_t\}]$. It is straightforward to show that β is an initial conditions mapping for the system.

Finally we construct the read out map $r(x_{t_o},u_{t_o},t_o)$, as follows: x_{t_o} is a nonanticipative mapping

$x_{t_o}: \overset{\circ}{U}|_{[t_o,\infty)\cap T} \to \overset{\circ}{Y}|_{[t_o,\infty)\cap T}$, so $y(t_o) := x_{t_o}(\omega)|_{t_o}$ depends only on $\omega(t_o)$. Now take $\omega(t_o) = u_{t_o}$, then $r(x_{t_o},u_{t_o},t_o) = x_{t_o}(\omega)|_{t_o}$.

It is straightforward to show that $\{\Sigma,X,\phi,r,\beta\}$ satisfies all the conditions to be a state space representation of the system Σ.

In many cases the spaces $U,Y,B,X,\overset{\circ}{U}$ and $\overset{\circ}{Y}$ are topological spaces. If ϕ,r,β and F are continuous mappings we will call the state space representation continuous and if a continuous state space representation exists, we will call the dynamical system continuous.

2.1-19. Definition. A continuous dynamical system Σ is called finite dimensional (or, of finite order) if there exists a continuous state space representation (Σ,X,ϕ,r) with $\dim X < \infty$, and $\dim U$, $\dim Y < \infty$.
In that case $\dim X \in \mathbb{N} \cup \{-1,0\}$, and therefore there exists a minimal state space dimension for the given Σ. This number is called the order of the system.

2.1-20. Remarks.

(i) For the formal definition of the dimension of a topological space, see e.g.[Hu-W]. This definition coincides with the well-known definition in special cases like, e.g. X is a vector space or X is a differentiable manifold etc.

(ii) Similar to the minimal dimension of the state space one can define the minimal dimension of the initial conditions space B. Let us call this the 'initial conditions order'. In a continuous dynamical system we expect the initial conditions order to be smaller or equal to the (state space) order of the system. However, we won't investigate this further, as it is not used in this generality in this work.

2.2. Definitions in the stochastic case

If the inputs and/or the initial conditions of a dynamical system are of a stochastic nature, one obtains a socalled stochastic system. If the initial conditions are deterministic and the inputs are stochastic but observed, then there will be no change in the definition of the dynamical system, compared to the deterministic definition. However, if (part of) the inputs are stochastic and unobserved, and/or the initial conditions are stochastic, then one has to use a different definition of a dynamical system. To prepare for the definition of a stochastic dynamical system, consider a σ-algebra $\overset{\circ}{G}$ of subsets of $\overset{\circ}{Y} = Y^T$, and let $\forall t \in T$: $\overset{\circ}{G}_t$ be the sub-σ-algebra of all subsets of $\overset{\circ}{Y}$ in $\overset{\circ}{G}$ that are of the form $G'_t \times Y^{(t,\infty)\cap T}$, with $G'_t \subseteq Y^{(-\infty,t]\cap T}$. It is clear

that if $t_1 < t_2$, $t_1, t_2 \in T$ then $\overset{\circ}{G}_{t_1} \subseteq \overset{\circ}{G}_{t_2}$, and so $\{\overset{\circ}{G}_t\}_{t\in T}$ forms a socalled filtration of σ-algebras.

2.2-1. Definition. Let $T, U, \overset{\circ}{U}, Y, \overset{\circ}{Y}, B$ be as before and $\overset{\circ}{G} = \bigcup_{t\in T} \overset{\circ}{G}_t$ as above. A stochastic dynamical system with s.i.c. is $\{T, U, \overset{\circ}{U}, Y, \overset{\circ}{Y}, \overset{\circ}{G}, B, P\}$, where

P is a mapping $P: B \times \overset{\circ}{U} \to \mathrm{Prob}(\overset{\circ}{Y}, \overset{\circ}{G})$, $(b,\omega) \longmapsto P_{b\omega}$, and where $\mathrm{Prob}(\overset{\circ}{Y}, \overset{\circ}{G})$ denotes the set of all probability measures on $(\overset{\circ}{Y}, \overset{\circ}{G})$, P such that for each $b \in B$, if $\omega_1|_{\leq t} = \omega_2|_{\leq t}$ then

$$P_{b\omega_1}(\eta|\overset{\circ}{G}_t) = P_{b\omega_2}(\eta|\overset{\circ}{G}_t) \qquad \text{(nonanticipation)}.$$

A stochastic dynamical system is $\{T, U, \overset{\circ}{U}, Y, \overset{\circ}{Y}, \overset{\circ}{G}, \overset{\circ}{P}\}$, where $\overset{\circ}{P} = \{P_b | b\in B\}$ is the set of all mappings $P_b: \overset{\circ}{U} \to \mathrm{Prob}(\overset{\circ}{Y}, \overset{\circ}{G})$, $\omega \longmapsto P_{b\omega}$.

2.2-2. Remarks. In this definition the inputs $\omega \in \overset{\circ}{U}$ should be interpreted as being observed and/or as being not (explicitly) stochastic. The same holds for the initial condition.

If all inputs are stochastic and unobserved and the same holds for the initial conditions then a stochastic dynamical system is just a stochastic process (i.e. a probability measure P on $(\overset{\circ}{Y}, \overset{\circ}{G})$).

Just as we had a state space representation (also called a (state space) realization) for a deterministic dynamical system, one can define (several) representations (also called realizations) for stochastic dynamical systems. The idea is that in many cases it may be possible to find a deterministic dynamical system together with (partly) stochastic inputs of a certain (simple) character and (partly) stochastic initial conditions, such that the outputs are random variables, which have a probability distribution that is exactly the one from the definition of the stochastic system considered. (Alternatively one might consider a deterministic dynamical system that has the probability distribution as its output, as e.g. in quantum mechanical models (cf. [Ma]).)

In this work we will be concerned with stochastic processes. Therefore in the remaining definitions we will consider only that case.

2.2-3. Definition. Consider the stochastic process $\{T,Y,\overset{\circ}{Y},G,P\}$. A stochastic realization is a deterministic dynamical system with initial conditions $\Sigma = \{T,U,\overset{\circ}{U},Y,\overset{\circ}{Y},B,F\}$ together with the σ-algebra $\overset{\circ}{G}$ on $\overset{\circ}{Y}$ and a σ-algebra $\overset{\circ}{H}$ on $B \times \overset{\circ}{U}$, and a probability measure Q on $(B\times\overset{\circ}{U},\overset{\circ}{H})$, such that $F: (B\times\overset{\circ}{U},\overset{\circ}{H}) \to (\overset{\circ}{Y},\overset{\circ}{G})$ is measurable, and the induced probability measure on $(\overset{\circ}{Y},\overset{\circ}{G})$ is P.
A stochastic state space realization is a stochastic realization in which the deterministic system is given by a state space representation.

2.2-4. Remarks. (i) Stochastic realization as defined here is also called weak stochastic realization, because it is only required that the probability measure on $(\overset{\circ}{Y},\overset{\circ}{G})$ is the same as in the stochastic system that is to be represented. It is not required here that the output of the deterministic system with stochastic inputs and initial conditions gives exactly the output of the stochastic dynamical system, which is the requirement for socalled strong stochastic realization.
(ii) The definition of a stochastic realization is very weak, e.g. it allows for a trivial solution y = u, with the stochastic inputs u having exactly the properties that are required for y. However, in practice one looks for stochastic realizations with certain properties; for example one requires the inputs to be an independent sequence (if $T = \mathbb{Z}$) or a Brownian motion process (if $T = \mathbb{R}$) etc.
(iii) Any two realizations of the same stochastic process will be called observationally equivalent. Observational equivalence is an equivalence relation clearly. (Switching from one realization to another will be one of the main tools in this work).

2.2-5. Definition. Suppose T is shift invariant. A stochastic process $\{\eta(t)\,|\,t\in T\}$ is stationary if for any finite subset $\{t_1,\ldots,t_n\} \subseteq T$, (n arbitrary) the joint probability distribution of

$$\eta(t_1+t),\ \eta(t_2+t),\ \ldots,\ \eta(t_n+t)$$

is independent of t.

This definition is taken from [Gu-S], p. 68, I-5. For a stationary stochastic process one has the ergodic theorem of Birkhoff-Khintchine (see e.g. [Gu-S]). To state it we first need some definitions.

2.2.6. Definition. Consider a stationary stochastic process with corresponding probability measure P on $(\overset{\circ}{Y},\overset{\circ}{G})$. A set $G \in \overset{\circ}{G}$ is called δ-shift -invariant if $\delta \in \Delta \setminus \{0\}$ and $P(S_\delta(G) \underline{\Delta} G) = 0$, where $\underline{\Delta}$ stands for the symmetric difference between two sets. A set $G \in \overset{\circ}{G}$ is called shift-invariant if it is δ-shift-invariant $\forall \delta \in \Delta \setminus \{0\}$. For each $\delta \in \Delta \setminus \{0\}$ all δ-shift-invariant sets together form a sub-σ-algebra of $\overset{\circ}{G}$, denoted by $\overset{\circ}{I}_\delta$ and all shift-invariant sets together form the sub-σ-algebra $\overset{\circ}{I} = \bigcap_{\delta \in \Delta \setminus \{0\}} \overset{\circ}{I}_\delta \subseteq \overset{\circ}{G}$.

2.2-7. Definition. Let $G \in \overset{\circ}{G}$ be arbitrary, and let χ_G be its indicator function (i.e. $\chi_G(\eta) = 1$ if $\eta \in G$, $\chi_G(\eta) = 0$ if $\eta \notin G$) then the conditional expectation $\pi(G) := E(\chi_G | \overset{\circ}{I})$ is called the empirical probability.

Let $\frac{1}{n}\upsilon_\eta^\delta(G) := \frac{1}{n}\sum_{k=0}^{n-1} \chi_G(\eta(k\delta))$ for some $\delta \in \Delta \setminus \{0\}$ i.e. $\frac{1}{n}\upsilon_n^\delta$ is the relative frequency of the realization of the event G, if $|\delta|$ is used as the length of the time interval, and $\mathrm{sgn}(\delta)$ as the direction of time.

2.2.8. Theorem. (Special case of the ergodic theorem).

$$\forall \delta \in \Delta^+ \cup \Delta^- : \lim_{n\to\infty} \frac{\upsilon_n^\delta(G)}{n} = E(\pi(G) | \overset{\circ}{I}_\delta)$$

Proof. See [Gu-S], p.154.

It follows from this that it is natural to 'split' the probability measure of the stochastic process, by conditioning on the empirical probability measure. The empirical probability measure π itself is a random variable, $\overset{\circ}{I}$-measurable. It is natural to regard the random choice of the empirical probability measure as a random initial condition in a stochastic realization of the process. For each choice of the initial condition one then obtains an ergodic process. The definition of an ergodic process is as follows:

2.2-9. Definition. A stationary stochastic process $(\overset{\circ}{Y},\overset{\circ}{G},P)$ is called ergodic if its empirical probability measure is equal to the probability measure P, P-a.s.

Using the above reasoning one can restrict oneself in stochastic realization

problems to stochastic realization problems of ergodic processes, because the non-ergodicity can be treated by using a random initial condition. In this way we arrive at the following.

2.2-10. Convention. In the following we will consider ergodic stationary stochastic processes.

2.2-11. Definition. An ergodic stationary stochastic process will be called linear, resp. finite dimensional if there exists a stochastic realization with a linear resp. finite dimensional, time-invariant deterministic dynamical system, together with independent inputs and initial conditions.

2.2-12. Remark. It will turn out that the initial conditions can be taken fixed in this case, because of the ergodicity.

2.2-13. Definition. A stochastic process is called gaussian if its probability measure P is gaussian.

2.2-14. Convention. We will restrict ourselves to the gaussian case.

Let $\{\eta(t) \mid t \in T\}$ denote an ergodic gaussian stationary process. Then $E\eta(t)$ is independent of t, and it can be estimated by $\frac{1}{n}\sum_{t=0}^{n-1} \eta(t)$ because of the ergodicity. As a slight simplification we will therefore assume $E\eta(t)$ to be known. If it is known one can substract it from the observed data to obtain zero-mean data. In this way we come to the following.

2.2-15. Assumption. We will assume $E\eta(t) = 0$.

2.3. Various representations of deterministic, finite dimensional linear systems

2.3.1. Introduction

In this section we want to present various representations of deterministic finite dimensional linear systems. The motivation is that this should work as an introduction for the stochastic case in the next section (2.4) and for the next chapter on minimal partial realization of arma(p,q) models and finite identifiability. Also it should show that indeed several different representations of the same objects exist. This in itself is a major motivation for this work. Especially as different representations are often used in different scientific fields, it is important to show the relations between the different models, like arma-models, state space models, transfer function (frequency domain) models, Hankelmatrix models, etc. We cannot treat this subject in full detail. Instead we will give an introduction which leads us to the results that are needed in later chapters. One of the motivations for the first chapters of this work is to show that even to obtain results for arma(p,q)-models, one can profitably use other representations, like state space models and Hankel matrices!

2.3-2. State space models

Recall from section 2.1 the definition of a deterministic, time invariant, finite dimensional, linear, dynamical, discrete time system. To avoid rather complicated topological dimension-problems, we will redefine finite dimensionality to be that the abstract realization of section 2.1 produces a finite dimensional state space.

Let us consider what this abstract realization gives us now. A state in the abstract realization is a mapping

$$x: \overset{\circ}{U}|_{[t_o,\infty)\cap T} \to \overset{\circ}{Y}|_{[t_o,\infty)\cap T}$$

$$\omega''|_{[t_o,\infty)\cap T} \longmapsto F[\omega'|_{(-\infty,t_o)\cap T} \cup \omega''|_{[t_o,\infty)\cap T}, b]\Big|_{[t_o,\infty)\cap T}$$

in the notation of section 2.1. Now F is linear, so we can write

$$F[\omega'|_{(-\infty,t_o)\cap T} \cup \omega''|_{[t_o,\infty)\cap T}, b] =$$

$$F[\omega'|_{(-\infty,t_o)\cap T} \cup 0|_{[t_o,\infty)\cap T}, b] + F[0|_{(-\infty,t_o)} \cup \omega''|_{[t_o,\infty)\cap T}, 0].$$

So x is completely determined, given F of course, by

$$F[\omega'|_{(-\infty,t_o)\cap T} \cup 0|_{[t_o,\infty)\cap T}, b]\Big|_{[t_o,\infty)\cap T}$$

$$= F[\omega'|_{(-\infty,t_o)\cap T} \cup 0|_{[t_o,\infty)\cap T}, 0]\Big|_{[t_o\infty)\cap T} + F[0,b]\Big|_{[t_o,\infty)\cap T}.$$

So the effect of the initial condition b can be separated from the effect of the inputs. Because B is a vector space and F is linear, the set $\{F[0,b]\Big|_{[t_o,\infty)\cap T}; b \in B\}$ is itself a vector space. Because of the time invariance property of the system, the dimension of this space is independent of the choice of $t_o \in T$. For the state space to be finite dimensional this dimension has to be finite. The problem that remains is: when is the image of the linear mapping $F[\omega'|_{(-\infty,t_o)\cap T} \cup 0|_{[t_o,\infty)\cap T}, 0]\Big|_{[t_o,\infty)\cap T}$ finite dimensional?

Because the system is time invariant and discrete time, the corresponding standard form of T is $T = \mathbb{Z}$. Therefore the mapping is a linear mapping of sequences. (To simplify the notation let us denote $\overset{\circ}{U}|_{(-\infty,t_o)\cap\mathbb{Z}}$ by $\overset{\circ}{U}|_{<t_o}$ etc.) We have not specified $\overset{\circ}{U}$, therefore we have to be somewhat careful as to which sequences we consider as input sequences. For all possible choices of $\overset{\circ}{U}$ that satisfy the conditions, $\overset{\circ}{U}|_{<t_o}$ will at least contain the subset of all

sequences that have only a finite number of nonzero components. Let that subset be denoted by $\overset{\circ}{U}_o|_{<t_o}$. The mapping restricted to $\overset{\circ}{U}_o|_{<t_o}$ can be described

completely by a (doubly semi-infinite) blockmatrix

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \cdots \\ H_{21} & H_{22} & H_{23} \cdots \\ H_{31} & H_{32} & H_{33} \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

such that if $\{y_{t_o+j}|t_o-1\}_{j\geq 0} := F[\omega'|_{<t_o} \cup 0|_{\geq t_o}, 0]|_{\geq t_o}$ denotes the

sequence of outputs that results from a sequence of inputs $\omega' = \{u_t\}_{t=-\infty}^{\infty} \in \overset{\circ}{U}_o$ that is concatenated with the zero sequence at t_o, then

$$\begin{bmatrix} y_{t_o|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \cdots \\ H_{21} & H_{22} & H_{23} \cdots \\ \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} u_{t_o-1} \\ u_{t_o-2} \\ \vdots \end{bmatrix}.$$

Now because of the time invariance, one can show quite easily that $H_{12} = H_{21}$; $H_{13} = H_{22} = H_{31}$, and in general that $H_{i,j} = H_{1,i+j-1}$. To simplify notation, one uses $H_k := H_{1,k}$. The matrix is a socalled block-Hankelmatrix, and it is of the form:

$$(2.3.2\text{-}1)\quad H = \begin{bmatrix} H_1 & H_2 & H_3 \cdots \\ H_2 & H_3 & H_4 \cdots \\ H_3 & H_4 & H_5 \cdots \\ \vdots & \vdots & \vdots \end{bmatrix}.$$

The abstract state space can now be identified with the image of this matrix. The state space must be finite dimensional, so the image of the matrix H must be a finite dimensional linear space! Let n denote this dimension. Choose n independent columns of H, (or more generally: n independent vectors in the image of H, but these can always be written as linear combinations of n independent columns). These form a basis of the state space. The state at time $t_o + 1$ is

$$\begin{bmatrix} y_{t_o+1|t_o} \\ y_{t_o+2|t_o} \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 \\ H_2 \\ \vdots \end{bmatrix} u_{t_o} + \begin{bmatrix} H_2 & H_3 \cdots \\ H_3 & H_4 \cdots \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} u_{t_o-1} \\ u_{t_o-2} \\ \vdots \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & I & 0 \cdots \\ 0 & 0 & I \cdots \\ \vdots & \vdots & \vdots \ddots \end{bmatrix} \begin{bmatrix} H_1 & H_2 \cdots \\ H_2 & H_3 \cdots \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} u_{t_o-1} \\ u_{t_o-2} \\ \vdots \end{bmatrix} + \begin{bmatrix} H_1 \\ H_2 \\ \vdots \end{bmatrix} u_{t_o} =$$

$$= \begin{bmatrix} 0 & I & 0 \cdots \\ 0 & 0 & I \cdots \\ \vdots & \vdots & \vdots \ddots \end{bmatrix} \begin{bmatrix} y_{t_o|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix} + \begin{bmatrix} H_1 \\ H_2 \\ \vdots \end{bmatrix} u_{t_o}.$$

With respect to the chosen basis of the image of H, we obtain an n×n matrix A representing the mapping

$$\begin{bmatrix} y_{t_o|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix} \longmapsto \begin{bmatrix} y_{t_o+1|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix}, \quad \text{(note that both vectors are in the image of H);}$$

and a matrix B representing the matrix

$$\begin{bmatrix} H_1 \\ H_2 \\ \vdots \end{bmatrix};$$

if x_{t_o} represents $\begin{bmatrix} y_{t_o|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix}$ with respect to the chosen basis, then we obtain the equation

$$x_{t_o+1} = Ax_{t_o} + Bu_{t_o}.$$

$y_{t_o|t_o-1}$ can be computed from x_{t_o}, say by the matrix C: $y_{t_o|t_o-1} = Cx_{t_o}$.

(C consists of the first block row of a matrix, the columns of which are formed by the chosen n independent vectors of the image of H). To obtain y_{t_o} we have to add the effect of u_{t_o} on y_{t_o}. This can be described by a matrix D,

$$y_{t_o} = Cx_{t_o} + Du_{t_o}.$$

So we obtain the equations

$$(2.3.2\text{-}2) \quad \begin{cases} x_{t+1} = Ax_t + Bu_t, \\ y_t = Cx_t + Du_t \end{cases} \quad t \in \mathbb{Z}.$$

It follows easily that

$$(2.3.2\text{-}3) \qquad H_i = CA^{i-1}B, \; i = 1,2,3,\ldots$$

(and $H_o := D$).

As usual one defines the socalled reachability matrix

(2.3.2-4) $R(A,B) := [B,AB,\dots]$

and the observability matrix

(2.3.2-5) $$Q(A,C) := \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix}.$$

Then $H = QR$, and because $rkH = n$, Q has n columns, and R has n rows, it follows that

(2.3.2-6) $rk(Q) = rk(R) = n.$

(This is the usual condition in linear systems theory for a minimal realization).
To obtain the full description of the system two things remain to be done:
(i) the initial conditions have to be treated
(ii) the set of inputs of the system has to be extended from $\overset{\circ}{U}_o$ to all of $\overset{\circ}{U}$.

ad(i) $\{F(0,b)|_{\geq t_o}\ ;\ b \in B\}$ is a finite dimensional vector space, with

dimension $\tilde{n}$ say, of sequences

$$\{\tilde{y}_j\}_{j \geq t_o}$$

(because of the assumption that the system is finite dimensional), which can again be considered as abstract states of the system. Because the system is time invariant $\tilde{n}$ does not depend on t_o; in fact the set $\{F(0,b)|_{\geq t_o}\ ;\ b \in B\}$ is a shifted version of $\tilde{X} := \{F(0,b)|_{\geq 0}; b \in B\}$ for each $t_o \in \mathbb{Z}$:
$S_{t_o}\{F(0,b)|_{\geq t_o}; b \in B\} = \tilde{X}$ for each $t_o \in \mathbb{Z}$.

Choose a basis for $\tilde{X}$. With respect to this basis the surjective mapping: $\tilde{X} \to \tilde{X}$,

$$\{\tilde{y}_{t_o+j}\}_{j \geq 0} \longmapsto \{\tilde{y}_{t_o+1+j}\}_{j \geq 0}$$

can be represented by an $\tilde{n}\times\tilde{n}$ non-singular (!) matrix $\tilde{A}$. Let $\{\tilde{y}_{t_o+j}\}_{j\geq 0} \in \tilde{X}$ be represented by $\tilde{x}_{t_o}$ with respect to the chosen basis, for each $t_o \in \mathbb{Z}$. Then one obtains the following set of equations

$$(2.3.2\text{-}7)\quad \begin{cases} \tilde{x}_{t+1} = \tilde{A}\tilde{x}_t \\ \tilde{y}_t = \tilde{C}\tilde{x}_t \end{cases}, \quad t \in \mathbb{Z}.$$

Note that b does not appear in these equations. They form the 'initial conditions' of these equations; note that they have an $\tilde{n}$ dimensional space of solutions; the choice of b amounts to the choice of one of these solutions. Because $\tilde{A}$ is nonsingular such a solution is fully specified by the choice of $\tilde{x}_{t_o}$ at some (arbitrary) $t_o \in \mathbb{Z}$. Once $\tilde{x}_{t_o}$ is determined the solution of these equations is fully determined.
The output of the system is the sum $y_t + \tilde{y}_t$ of the outputs y_t obtained with zero initial conditions and $\tilde{y}_t$. We will not go into further details.

ad(ii). The extension of the set of inputs from $\overset{\circ}{U}_o$ to $\overset{\circ}{U}$ is rather subtle.
(a) One very natural possibility is to define the set of inputs U_a by

$$\overset{\circ}{U}_a := \{\omega = \{u_t\} \mid \omega \in U^{\mathbb{Z}},\ \sum_{i=o}^{\infty} H_i u_{t-i} \text{ converges for all } t \in \mathbb{Z}\},$$

where we have assumed U to be a topological vector space. Then $\overset{\circ}{U}$ clearly depends on the parameters of the system. E.g. if A is nilpotent, then $A^n = 0$, so $H_{n+j} = CA^{n+j-1}B = 0\ \forall j \geq 1$; and then $\overset{\circ}{U} = U^{\mathbb{T}}$. Let now our field be $\mathbb{R}$ or $\mathbb{C}$. Let $\sigma(A) = \{\mu\in\mathbb{C} \mid \det(\mu I - A) = 0\}$ be the spectrum of the matrix A. Let

$$(2.3.2\text{-}8)\quad |\lambda| = \max\{|\mu| : \mu \in \sigma(A)\}.$$

If $|\lambda| \leq 1$ and $\sum_{i=o}^{\infty} u_{t-i}$ is absolutely convergent (for some $t \in \mathbb{Z}$ and therefore for all $t \in \mathbb{Z}$), then it can easily be shown that

$$\sum_{i=o}^{\infty} H_i u_{t-i}$$

is absolutely convergent, $\forall t \in \mathbb{Z}$.

More generally

2.3.2-9. Theorem. If $\exists t \in \mathbb{Z}: \sum_{i=o}^{\infty} |\lambda|^i |u_{t-i}| < \infty$ then $\sum_{i=o}^{\infty} H_i u_{t-i}$ is absolutely convergent, $\forall t \in \mathbb{Z}$, and so $\{u_t\} \in \overset{\circ}{U}_a$.

(b) A second possibility is to define the set of inputs as

$$\overset{\circ}{U}_b := \{\omega=\{u_t\} | \omega \in U^{\mathbb{Z}}, \sum_{i=o}^{\infty} H_i u_{t-i} \text{ is Cesaro summable for all } t \in \mathbb{Z}\}.$$

(c) And a third possibility is:

$$\overset{\circ}{U}_c := \{\omega=\{u_t\} | \omega \in U^{\mathbb{Z}}, \sum_{i=o}^{\infty} H_i u_{t-i} \text{ is Abel summable for all } t \in \mathbb{Z}\}.$$

It is wellknown that $\overset{\circ}{U}_a \subseteq \overset{\circ}{U}_b \subseteq \overset{\circ}{U}_c$. Usually one can work with $\overset{\circ}{U}_a$, but sometimes it is of importance to be able to use $\overset{\circ}{U}_b$ or $\overset{\circ}{U}_c$.
From these examples it is clear that there are several possibilities for the choice of $\overset{\circ}{U}$ and the definition of F, even if F restricted to $U_o \times B$, is completely determined.

2.3.3. Arma models

Consider again the abstract state

$$\begin{bmatrix} y_{t_o|t_o-1} \\ y_{t_o+1|t_o-1} \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & \cdots \\ H_2 & H_3 & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} u_{t_o-1} \\ u_{t_o-2} \\ \vdots \end{bmatrix}.$$

Because the rank of the block Hankel-matrix is n, there are dependency relations between its rows. Consider the left kernel of H:

$$(2.3.3\text{-}1) \qquad \ell\ker H := \left\{ \alpha \;\middle|\; \begin{array}{l} \alpha \text{ a row vector with a finite number of} \\ \text{nonzero components, and } \alpha H = 0 \end{array} \right\}.$$

Let us partition $\alpha = (\alpha_1 \;\; \alpha_2 \;\; \alpha_3 \;\; \ldots)$ in accordance with the blockstructure of H. It follows that $\alpha \in \ell\ker H$ iff $\sum_{i=1}^{\infty} \alpha_i H_{i+j} = 0 \;\forall j \geq 0$. Let the shift S be a

linear mapping given by $S\alpha = (0 \quad \alpha_1 \quad \alpha_2 \ \ldots)$. From the block Hankel structure of H it follows that if $\alpha \in \ell\ker H$ then $S\alpha \in \ell\ker(H)$. We now want to find a minimal set of generating vectors $\{\alpha^{(i)}\}$ such that the span $[\{S^j\alpha^{(i)}\}_{j=0}^{\infty}] = \ell\ker H$, (i.e. each element of $\ell\ker H$ can be written as a finite linear combination of shifted versions of $\alpha^{(i)}$).

Such a set can be found as follows:

First define:

$$(2.3.3\text{-}2) \qquad \forall i \geq 1: \ t_i = \dim \operatorname{Im} \begin{bmatrix} H_1 & H_2 & \cdots \\ H_2 & H_3 & \\ \vdots & \vdots & \\ H_i & H_{i+1} & \cdots \end{bmatrix}, \quad t_o = 0,$$

and $\forall i \geq 1: \ s_i = t_i - t_{i-1}; \ s_o = m :=$ the number of rows in each block.

Then $\{s_i\}$ is monotonically nonincreasing (because of the Hankel structure of H) and $\sum_{i \geq 1} s_i = n$. So $s_{n+i} = 0$ for all $\forall i \geq 1$, and hence $t_{n+i} = n$ for all $i \geq 0$!! The i^{th} block row $(H_i \quad H_{i+1} \ldots)$ of H contains s_i rows that are independent of the rows above and $m-s_i$ rows that are dependent of the rows above in H. To obtain the $\alpha^{(i)}$ proceed as follows: Consider the rows of H one by one starting at the top, and check with each one whether it is dependent on the rows above it. If it is, then there is a corresponding element $\alpha \in \ell\ker H$, namely the vector α which describes the linear relation; the coefficient in α of the row we consider is then unequal to zero. We can always express a row in terms of the rows above that are itself linear independent of their previous rows. Now consider the position k that the considered row takes in its block row ($k \in \{1, \ldots m\}$). If the corresponding relation is given by α then $S^j\alpha \in \ell\ker H \ \ \forall j \geq 0$. Therefore for each k we don't have to put in more than one element $\alpha^{(k)}$ (say) in the generating set. We obtain a generating set $\{\alpha^{(1)}, \ldots, \alpha^{(m)}\}$, $\alpha^{(k)}$ corresponding to a row in position k in its block row. We can write $\alpha^{(k)}$ in its partitioned form:

$$\alpha^{(k)} = (\alpha_1^{(k)} \quad \alpha_2^{(k)} \quad \ldots \quad \alpha_{1+p_k}^{(k)} \quad 0 \quad \ldots)$$

with p_k such that $\alpha_{1+p_k}^{(k)} \neq 0$. In fact one has, by construction $\alpha_{1+p_k}^{(k)} e_k \neq 0$, where e_k is the k^{th} standard unit vector, and (2.3.3-3a) $\alpha_{1+p_k}^{(k)} e_{k+j} = 0$ for

$j \geq 1$. We will assume, without loss of generality that $\alpha^{(k)}$ is normalized such that

(2.3.3-3b) $\alpha^{(k)}_{1+p_k} e_k = 1, \quad k = 1,2,\ldots, m$

It is easy to show that $p_k \leq n$ (this follows from the fact that $s_{n+1} = 0$). In fact there is a one-to-one relation between the bag $\{p_k | k=1,\ldots,m\}$ and the sequence $s_1,\ldots,s_n$. (A bag is, so to speak, 'a set in which an element can occur more than once'. A finite (or countable) bag can formally be constructed as the set of all permutations of a corresponding sequence. E.g. the bag $\{1,1,2\}$ corresponds to the set of sequences $\{(1,1,2),(1,2,1),(2,1,1)\}$. A finite bag will be denoted by one of its representative sequences).
The relationship is as follows: The number of k's for which $1+p_k = i \geq 1$, is equal to $s_{i-1} - s_i = (m-s_i) - (m-s_{i-1})$ i.e. the increase in the number of rows that are dependent on the previous rows, if one goes from block row i-1 to block row i.

2.3.3-4. Definition. The bag $\{p_k | k=1,2,\ldots,m\}$ is called the bag of Kronecker indices or observability indices.

2.3.3-5. Theorem. $\sum_{k=1}^{m} p_k = n$.

Proof. $\Sigma p_k = \sum_{i=o}^{n} i \times |\{k|p_k=i\}| = \sum_{i=o}^{n} i(s_i - s_{i+1}) =$

$$\sum_{0}^{n} is_i - \sum_{0}^{n} is_{i+1} = \sum_{1}^{n} is_i - \sum_{1}^{n} (i-1)s_i = \sum_{1}^{n} s_i = n.$$

Q.E.D.

Now consider the abstract state at time $t_o + 1$:

$$\begin{bmatrix} y_{t_o+1|t_o} \\ y_{t_o+2|t_o} \\ \vdots \end{bmatrix} = \begin{bmatrix} H_1 & H_2 & \cdots \\ H_2 & H_3 & \cdots \\ \vdots & \vdots & \end{bmatrix} \begin{bmatrix} u_{t_o} \\ u_{t_o-1} \\ u_{t_o-2} \\ \vdots \end{bmatrix}.$$

It follows that

$$(2.3.3\text{-}6)\qquad \sum_{j=1}^{1+p_k} \alpha_j^{(k)}\, y_{t_o+j|t_o} = 0, \quad k = 1,2,\ldots,m.$$

Because $S^i\alpha^{(k)} \in \ell\ker H$ for each $i = 0,1,2,\ldots$, it follows that

$$(2.3.3\text{-}7)\qquad \sum_{j=1}^{1+p_k} \alpha_j^{(k)} y_{t_o+j+i|t_o} = 0, \quad k = 1,2,\ldots m;\ i = 0,1,2,\ldots$$

Equations (2.3.3-6) and (2.3.3-7) will be called the autoregressive equations for the abstract state. One has the following simple but important observation.

2.3.3-8. Lemma. The abstract state $(y_{t_o+j|t_o})_{j=1}^{\infty}$ (at time t_o+1) is completely determined by the autoregressive equations and the following n quantities:

$$(2.3.3\text{-}9)\qquad y_{t_o+j|t_o}^T e_k,\ j = 1,\ldots,p_k,\ k = 1,\ldots,m$$

Proof. Consider each scalar component of $(y_{t_o+1|t_o}^T, y_{t_o+2|t_o}^T, \ldots)^T$, one by one, starting at the top. For each component there are two possibilities: Either it is one of the variables in (2.3.3-9), or it can be expressed as a linear combination of the previous scalar components using the autoregressive equations. Q.E.D.

2.3.3-10. Remark. One can use the vector of n variables in (2.3.3-9) as a state vector and work out the state space model that results from it.

We now pose the question whether (2.3.3-7) can be generalized to an equation which holds for all $i \in \mathbb{Z}$. The answer is yes, and it can be done as follows. Note that (2.3.3-6) holds for each $t_o \in \mathbb{Z}$. So we have

$$(2.3.3\text{-}11)\qquad \sum_{j=1}^{1+p_k} \alpha_j^{(k)} y_{t+j|t} = 0, \quad k = 1,2,\ldots,m, \quad t \in \mathbb{Z}.$$

Now one can express $y_{t+j|t}$ in terms of y_{t+j} and $u_{t+j}, u_{t+j-1}, \ldots, u_{t+1}$ as follows:

$$(2.3.3\text{-}12)\qquad y_{t+j|t} = y_{t+j} - (H_o u_{t+j} + H_1 u_{t+j-1} + \ldots + H_{j-1} u_{t+1}),\ j = 1,2,\ldots$$

Substituting this in (2.3.3-11) gives the following set of equations:

$$(2.3.3\text{-}13)\quad \sum_{j=1}^{1+p_k} \alpha_j^{(k)} y_{t+j} = \sum_{j=1}^{1+p_k} \alpha_j^{(k)} \sum_{i=o}^{j-1} H_i u_{t+j-i} = \sum_{\ell=1}^{1+p_k} \Big[\sum_{i=o}^{1+p_k-\ell} \alpha_{\ell+i}^{(k)} H_i \Big] u_{t+\ell}.$$

Let

$$(2.3.3\text{-}14)\quad \beta_\ell^{(k)} := \sum_{i=o}^{1+p_k-\ell} \alpha_{\ell+i}^{(k)} H_i,$$

then (2.3.3-13) can be rewritten as

$$(2.3.3\text{-}15)\quad \sum_{j=1}^{1+p_k} \alpha_j^{(k)} y_{t+j} = \sum_{j=1}^{1+p_k} \beta_j^{(k)} u_{t+j},\ t \in \mathbb{Z},\ k = 1,\ldots,m.$$

These are socalled <u>arma-equations</u>. Define

$$(2.3.3\text{-}16)\quad u_{t|t_o} := \begin{cases} u_t & \text{if } t \le t_o \\ 0 & \text{if } t > t_o. \end{cases}$$

Then it follows that

$$(2.3.3\text{-}17)\quad \sum_{j=1}^{1+p_k} \alpha_j^{(k)} y_{t+j|t_o} = \sum_{j=1}^{1+p_k} \beta_j^{(k)} u_{t+j|t_o},$$

which is the generalization of the equation (2.3.3-7) that was sought for. The following questions arise:

(i) Does the arma model (as constructed here) determine the Hankel-matrix completely, i.e. have we lost no information in going from the Hankel-matrix to the arma model?

(ii) Can we compute the abstract state at time t_o using the arma model, and which y_k's and u_k's are needed to do that?

To answer this we will use a different indexation of the coefficients $\alpha_j^{(k)}$ and $\beta_j^{(k)}$ in the arma-equations (2.3.3-15). Let

$$(2.3.3\text{-}18)\quad \begin{cases} a_j^{(k)} := \alpha_{1+p_k-j}^{(k)} \\ b_j^{(k)} := \beta_{1+p_k-j}^{(k)} \end{cases},\ j = 0,\ldots,p_k.$$

Note that from (2.3.3-3) it follows that

$$(2.3.3\text{-}19)\begin{cases} a_o^{(k)} e_k = 1 \quad , \ k = 1,2, \ldots,m, \\ a_o^{(k)} e_{k+j} = 0 \ , \ 1 \le j \le m-k, \ k = 1,2,\ldots,m-1. \end{cases}$$

Therefore the m×m matrix A_o with $a_o^{(k)}$ as its k^{th} row, $k = 1,2,\ldots m$, is nonsingular. Let A_j, resp. B_j be the matrix with $a_j^{(k)}$, resp. $b_j^{(k)}$ as its k^{th} row. Then it is straightforward to show that the equations (2.3.3-14) can be rewritten in the form

$$(2.3.3\text{-}20) \quad \begin{bmatrix} A_o & 0 & \cdots & 0 \\ A_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ A_p & \cdots & A_1 & A_o \end{bmatrix} \begin{bmatrix} H_o \\ H_1 \\ \vdots \\ H_p \end{bmatrix} = \begin{bmatrix} B_o \\ B_1 \\ \vdots \\ B_p \end{bmatrix},$$

where $p := \max_{1 \le k \le m} p_k$. Because A_o is nonsingular the block-lower-triangular matrix in (2.3.3-20) is nonsingular, and so $H_o,\ldots,H_p$ can be computed from $A_o,\ldots,A_p$, $B_o,\ldots,B_p$, i.e. from the coefficients of the arma-equations. So the answer to question (i) is affirmative.

Using the new coefficients the arma-equations become

$$(2.3.3\text{-}21) \quad \sum_{j=o}^{p_k} a_j^{(k)} y_{t-j} = \sum_{j=o}^{p_k} b_j^{(k)} u_{t-j}, \ k = 1,\ldots,m, \ t \in \mathbb{Z}.$$

or

$$(2.3.3\text{-}22) \quad \sum_{j=o}^{p} A_j y_{t-j} = \sum_{j=o}^{p} B_j u_{t-j}, \ t \in \mathbb{Z}.$$

Any set of equations of the form (2.3.3-22), with A_o nonsingular, will be called an arma model, whether or not it has the structure of (2.3.3-21). The following analysis will answer question (ii) and at the same time show that such an arma model defines a (Hankel matrix of a) linear system.

Because A_o is nonsingular, we can write

$$(2.3.3\text{-}23) \quad y_t = \sum_{j=1}^{p} (-A_o^{-1} A_j) y_{t-j} + \sum_{j=o}^{p} A_o^{-1} B_j u_{t-j}.$$

Just as in (2.3.3-17), it follows that

$$(2.3.3\text{-}24)\quad y_{t|t_o-1} = \sum_{j=1}^{p} (-A_o^{-1}A_j)y_{t-j|t_o-1} + \sum_{j=o}^{p} A_o^{-1}B_j u_{t-j|t_o-1}.$$

Therefore, if $y_{t_o-1},\ldots,y_{t_o-p}$, $u_{t_o-1},\ldots, u_{t_o-p}$ are given then, using the fact that $u_{t|t_o-1} = 0$ if $t \geq t_o$, from (2.3.3-24) one can compute $y_{t_o+j|t_o-1}$ recursively, for $j = 0,1,2,\ldots$. This answers question (ii). It shows that the arma-equations (2.3.3-23) can be considered as a state space model, namely by taking $x_{t_o} = (y_{t_o-1},y_{t_o-2},\ldots,y_{t_o-p},u_{t_o-1},\ldots,u_{t_o-p})$ as the state at time t_o. It contains enough information to be a state vector because the abstract state can be computed from it, as we have just shown. The corresponding state equations can be found as follows:

$$x_{t_o+1} = (y_{t_o},\ldots,y_{t_o-p+1},u_{t_o},\ldots,u_{t_o-p+1}),$$

so to obtain x_{t_o+1} from x_{t_o} and u_{t_o} one only has to compute y_{t_o} from x_{t_o} and u_{t_o}, which can be done using (2.3.3-23) with $t = t_o$.

The read-out map is in fact this same map from x_{t_o} and u_{t_o} to y_{t_o}. Because the number of components in x_t is in general more than n, it is clear that the arma-model is in general a nonminimal state space model for the linear system that it defines.

2.3.4. The transfer matrix and polynomial mfd(= arma) models

It will be useful for us to define both a backward and a forward transfer matrix, where the forward transfer matrix coincides with the usual transfer matrix. The definitions are as follows:

2.3.4-1. Definition. $H(z) := \sum_{i=o}^{\infty} H_i z^i$ is called the backward transfer matrix of the system. $T(s) = H(s^{-1})$ is called the (forward) transfer matrix.

2.3.4-2. Remarks. (i) H(z) can be considered as a formal matrix power series or as a matrix of the complex variable z, with (open, nonempty) domain $\mathrm{Dom}(H) = \{z \in \mathbb{C}; |z||\lambda| < 1\}$, where $|\lambda|$ is the maximum modulus of the spectrum of the system, as in (2.3.2-8). T(s) can be considered as a formal Laurent series, or as a function of the complex variable s, with (open, nonempty) domain $\mathrm{Dom}(T) = \{s \in \mathbb{C};\ |s| > |\lambda|\}$.
(ii) It is clear that H(z) and T(s) are analytic within their domain of definition, and there is a one-to-one relation between the complex matrix function H(z) and the sequence of matrices $\{H_0, H_1, H_2, \ldots\}$ and therefore with the formal power series. A similar result holds for T(s).
(iii) That the domain of H and therefore the domain of T is as described follows easily from the following theorem.

2.3.4-3. Theorem. If (A,B,C)[1] is a minimal realization then $\lim_{k\to\infty} CA^kB = 0$ iff $\sigma(A) \subseteq D(0,1) := \{z \in \mathbb{C};\ |z| < 1\}$.

Proof. See appendix 2A.

We leave it to the reader to conclude from this theorem that the domains of H(z) and T(s) are indeed as claimed (using formula (2.3.2-3)).
Now let us see how the transfer matrix can be obtained from a state space realization and from an arma model.

2.3.4-4. Theorem. Suppose a linear system is given by a state space model (2.3.2-2) (not necessarily a minimal realization). Then the backward and forward transfer matrices of the system are given by, respectively,

$$H(z) = D + zC(I-zA)^{-1}B,\ z \in \mathrm{Dom}(H), z^{-1} \notin \sigma(A) \tag{2.3.4-5}$$

and

$$T(s) = D + C(sI-A)^{-1}B,\ s \in \mathrm{Dom}(T),\ s \notin \sigma(A). \tag{2.3.4-6}$$

---

1) (A,B,C) means $(A,B,C,0)$, i.e. $D = 0$.

Proof. If $z \in \mathrm{Dom}(H)$ and $z^{-1} \in \sigma(A)$ then both sides of (2.3.4-5) are well-defined and analytic. Therefore it is enough to show equality on an open nonempty subset of $\{z \mid z \in \mathrm{Dom}(H)$ and $z^{-1} \notin \sigma(A),\ z \in \mathbb{C}\}$. This subset will be $\{z \in \mathbb{C}; |z||\lambda| < 1\}$, where $|\lambda|$ = maximum modulus of the spectrum $\sigma(A)$ of A. (Note that (A,B,C,D) is not necessarily a minimal realization. Therefore this set does not necessarily coincide with Dom(H). Within this set one has

$$(I-zA)^{-1} = \sum_{k=o}^{\infty} z^k A^k \text{ (sum of a geometric series), so}$$

$$(2.3.4\text{-}7) \qquad D+zC(I-zA)^{-1}B = D + \sum_{k=1}^{\infty} CA^{k-1}Bz^k = \sum_{k=o}^{\infty} H_k z^k = H(z).$$

The equality (2.3.4-6) follows simply by substitution of $z = s^{-1}$ in (2.3.4-5).

Q.E.D.

Notice that from this theorem it follows that the transfer matrix of a linear system is a matrix of rational functions. Furthermore

$$(2.3.4\text{-}8) \qquad H(0) = D = \lim_{s\to\infty} T(s), \text{ so } T(s) \text{ consists of proper rational functions}$$

(i.e. if the rational function is written as the quotient of two polynomials, the degree of the numerator polynomial is smaller than or equal to the degree of the denominator polynomial). They are all strictly proper (i.e. degree of numerator polynomial is smaller than the degree of the denominator polynomial) iff D = 0. The question arises whether each matrix of proper rational functions is the transfer matrix of a linear system. That this is indeed the case follows most easily by considering the transfer matrix of an arma model.

2.3.4-9. Theorem. Suppose a linear system is given by an arma model

$$(2.3.4\text{-}10) \qquad \sum_{j=o}^{p} A_j y_{t-j} = \sum_{j=o}^{p} B_j u_{t-j}, \quad t \in \mathbb{Z},$$

with A_o nonsingular. (Compare (2.3.3-22)). Then the transfer matrix of the corresponding linear system is given by:

$$(2.3.4\text{-}11) \qquad H(z) = \Big(\sum_{j=o}^{p} A_j z^j\Big)^{-1}\Big(\sum_{j=o}^{p} B_j z^j\Big),$$

$$z \in \{z \in \mathbb{C} \mid z \in \mathrm{Dom}(H) \text{ and } \det\Big(\sum_{j=o}^{p} A_j z^j\Big) \neq 0\}$$

and

$$T(s) = \Big(\sum_{j=o}^{p} A_j s^{p-j}\Big)^{-1}\Big(\sum_{j=o}^{p} B_j s^{p-j}\Big),$$

(2.3.4-12)

$$s \in \{s \mid s \in \mathrm{Dom}(T) \text{ and } \det\Big(\sum_{j=o}^{p} A_j s^{p-j}\Big) \neq 0\}.$$

Proof. Let $L = S^{-1}$ denote the backward shift or lag operator. The input/output operator of the system acting on $\overset{\circ}{U}_o$, can then be described by

$$(2.3.4\text{-}13)\quad y_t = \sum_{k=o}^{\infty} H_k u_{t-k} = \sum_{k=o}^{\infty} H_k L^k u_t = H(L)u_t.$$

Premultiplication with $\sum_{j=o}^{p} A_j L^j =: A(L)$ gives

$$(2.3.4\text{-}14)\quad A(L)y_t = A(L)H(L)u_t.$$

From (2.3.4-10) one has, using the notation $B(L) := \sum_{j=o}^{p} B_j L^j$,

$$(2.3.4\text{-}15)\quad A(L)y_t = B(L)u_t.$$

In this way one obtains the equation

$$(2.3.4\text{-}16)\quad A(L)H(L)u_t = B(L)u_t, \text{ for all } \{u_t\} \in \overset{\circ}{U}_o.$$

Therefore

$$(2.3.4\text{-}17)\quad A(L)H(L) = B(L).$$

Substituting z for L one gets for all $z \in \mathrm{Dom}(H)$:

$$(2.3.4\text{-}18)\quad A(z)H(z) = B(z).$$

Now $\det A(0) \neq 0$, so $\det A(z) \not\equiv 0$, i.e. $\det A(z)$ is a polynomial, which is not the zero polynomial. Therefore there are only a finite number of values of z for which $\det A(z) = 0$. For all $z \in \{z \in \mathbb{C} \mid z \in \mathrm{Dom}(H) \text{ and } \det A(z) \neq 0\}$, $A(z)$ is invertible and it follows that

(2.3.4-19) $H(z) = A(z)^{-1}B(z)$.

The corresponding formula for T(s) follows by substitution of $s = z^{-1}$ in (2.3.4-19) and multiplying $A(s^{-1})$ and $B(s^{-1})$ by s^p. Q.E.D.

2.3.4-20. Remarks. (i) Suppose that $H(z) = A(z)^{-1}B(z)$ for some pair of polynomial matrices (A(z), B(z)), with A(0) nonsingular, then $A(z)H(z) = B(z)$ and so $A(L)H(L) = B(L)$, i.e. $A(L)y_t = B(L)u_t$; in other words: H(z) is a transfer matrix of a linear system, described by the arma model $A(L)y_t = B(L)u_t$.
(ii) We saw above that the transfer matrix T(s) of a linear system consists of proper rational functions as its components. This can now be turned around. If T(s) is a matrix of proper rational functions, then H(z) consists of rational functions and H(0) is well-defined. Let a(z) be the smallest common polynomial denominator of all components of H(z), then $a(0) \neq 0$ (because H(0) is well-defined). Let $B(z) := a(z)H(z)$, then B(z) is a polynomial matrix and $H(z) = a(z)^{-1}B(z)$. So H(z) is the backward transfer matrix of the arma model $a(L)y_t = B(L)u_t$. This proves that H(z) is the backward transfer matrix of a linear system.

2.3.4-21. Definition. Let H(z) be the backward transfer matrix of a linear system, and T(s) be the corresponding (forward) transfer matrix.
(i) A pair of polynomial matrices (A(z),B(z)) such that $H(z) = A(z)^{-1}B(z)$ for all $z \in \mathrm{Dom}(H)$, except perhaps for a finite number of points, is called a backward mfd (matrix fraction description) pair.
(ii) A pair of polynomial matrices $(A_v(s),B_v(s))$ such that $T(s) = A_v(s)^{-1}B_v(s)$, for all $s \in \mathrm{Dom}T$, except perhaps for a finite number of points, is called a forward mfd pair.

2.3.4-22. Remark. Note the close relationship between arma models and mfd pairs. If (A(z),B(z)) is a backward mfd pair, then

$$A(L)y_t = B(L)u_t$$

forms a 'generalized' arma model. If A(0) is nonsingular, then it is an arma model. If A(0) is singular then still $\{y_t\}$ and $\{u_t\}$ satisfy this set of

equations, which we will call a generalized arma model. (We keep the word arma model for those cases in which A(0) is nonsingular, because in that case one can directly express y_t as a linear combination of its own past p-1 values and the last p input values. That justifies the terminology autoregressive moving average model).
If $(A_v(s),B_v(s))$ form a forward mfd pair, then $\{y_t\},\{u_t\}$ satisfy the equations

(2.3.4-23) $A_v(S)y_t = B_v(S)u_t$.

These equations together will be called a generalized forward arma model.

It is important to note that if (A(z),B(z)) is a backward mfd pair then premultiplying A(z) and B(z) with a nonsingular polynomial matrix C(z) (i.e. detC(z) $\not\equiv$ 0) gives us another mfd pair (C(z)A(z), C(z)B(z)) corresponding to the same linear system: The transfer matrix is equal to

(2.3.4-24) $H(z) = A(z)^{-1}B(z) = [C(z)A(z)]^{-1}C(z)B(z)$.

Even if C(z) is a nonsingular matrix of rational functions, such that C(z)A(z) and C(z)B(z) are polynomial the same conclusion holds.

2.3.4-25. Notation. To simplify the notation we will often write an mfd pair (A(z),B(z)) as a partitioned matrix[A(z) B(z)]. Premultiplication of A(z) and B(z) with C(z) can then be written as C(z)[A(z) B(z)].
We can conclude that there is a collection of mfd pairs all describing the same linear system. If [A(z) B(z)] is one mfd pair describing the model, then

(2.3.4-26) $\{C(z)[A(z)\ B(z)] \mid$ C(z) a matrix of rational functions, det C(z) $\not\equiv$ 0; C(z)[A(z) B(z)] polynomial$\}$

is the set of all backward mfd pairs describing the same system. (Perhaps it is good to stress again at this point that we are only looking at one aspect of the linear system, namely the input/output behaviour on $\mathring{U}_o$ under zero initial conditions. If one considers nonzero initial conditions, then the models described by [A(z) B(z)] and C(z)[A(z) B(z)] are not the same if deg det C(z) $\geq$ 1. The simplest example is C(z) = 1-z, A(z) = B(z) = 1 (scalar). Then [A(z) B(z)] corresponds to the model $y_t = u_t$, while C(z)[A(z) B(z)] corresponds to the model $y_t - y_{t-1} = u_t - u_{t-1}$. Solutions of this second

model are of the form $y_t = u_t + c$, where c is an arbitrary constant. It is clearly determined by the initial conditions. And in case of zero initial conditions it is of course equal to zero: c = 0)
Similar to (2.3.4-26) one obtains the set of all forward mfd pairs of a linear system, given by $[A_v(s)\ B_v(s)]$, to be

(2.3.4-27) $\{C(s)[A_v(s)\ B_v(s)] \mid$ C(s) a matrix of rational functions, det C(s) $\neq$ 0; $C(s)[A_v(s)\ B_v(s)]$ polynomial$\}$.

If one goes one step further and looks at the set of all rational mfd pairs instead of polynomial mfd pairs, then one gets the following sets.
In the backward case

(2.3.4-28)
$$L_B := \{C(z)[A(z)\ B(z)] \mid \text{C(z) matrix of rational functions, detC(z) } \neq 0\}$$
$$= \{C(z)[I\ H(z)] \mid \text{ " " }\}$$

and in the forward case

(2.3.4-29) $L_F := \{C(s)[I\ T(s)] \mid$ C(s) a matrix of rational functions, det C(s) $\neq$ 0$\}$.

L_B and L_F are m-dimensional linear subspaces of the space $k^{m+m'}$ over the field k, where k is the field of rational functions with real coefficients. It follows that an mfd pair [A(z) B(z)] is just a <u>polynomial basis</u> of the k-linear space L_B. And a similar thing can be said for a forward mfd pair. Now a natural question to be posed is how one can obtain a <u>concise</u> description of the system in terms of an mfd pair. In the literature, work has been done on this problem. We will closely follow the work of Forney [Fo 75]. He defines the concept of a minimal base form:

2.3.4-30. <u>Definition</u>. A backward mfd pair [A(z) B(z)] is said to be in <u>minimal base form</u> if the sum of the (polynomial) degrees of the rows of [A(z) B(z)] is minimal among all mfd pairs in (2.3.4-25).
A similar definition holds for forward mfd pairs.

2.3.4-31. Remarks. (i) [Fo 75] gives an algorithm that brings any mfd pair in minimal base form.
(ii) A minimal base form is not unique. For example, it is always possible to add rows of $[A(z)\ B(z)]$ to rows which have a higher polynomial degree, without altering the sum of the row degrees. So if we started with a minimal base form, we will end up with another minimal base form.
(iii) [Fo 75] has shown, however, that the sequence of row degrees $\{r_1,r_2,\ldots,r_m\}$, where r_i := polynomial degree of the i^{th} row of the mfd pair, is the same, up to a permutation, for all minimal base forms of a system.

To treat the relations between minimal base forms in L_B and L_F, the following notation is useful.

2.3.4-32. Notation. (i) If $[A(z)\ B(z)]$ is a backward mfd pair and $\{r_i\}$ are the row degrees, then

$$(2.3.4\text{-}33)\quad [\tilde{A}_v(s)\ \tilde{B}_v(s)] := \mathrm{diag}(s^{r_1},s^{r_2},\ldots,s^{r_m}).[A(s^{-1})\ B(s^{-1})]$$

If $[A_v(s)\ B_v(s)]$ is a forward mfd pair, and $\{\rho_i\}$ are the row degrees, then

$$(2.3.4\text{-}34)\quad [\tilde{A}(z)\ \tilde{B}(z)] := \mathrm{diag}(z^{\rho_1},\ldots,z^{\rho_m})[A_v(z^{-1})\ B_v(z^{-1})]$$

It is clear that if $(A(z),B(z))$ is a backward mfd pair with row degrees $\{r_i\}$, then $(\tilde{A}_v(s),\tilde{B}_v(s))$ is a forward mfd pair with row degrees $\{\rho_i\}$, where $\rho_i \le r_i$, $i = 1,\ldots,m$. Similarly if $(A_v(s),B_v(s))$ is a forward mfd pair with row degrees $\{\rho_i\}$, then $(\tilde{A}(z),\tilde{B}(z))$ is a backward mfd pair with row degrees $\{r_i\}$, where $r_i \le \rho_i$, $i = 1,\ldots,m$.

2.3.4-35. Theorem. (i) If $(A(z),B(z))$ is in minimal base form then $(\tilde{A}_v(s),\tilde{B}_v(s))$ is in minimal base form with the same row degrees.
(ii) If $(A_v(s),B_v(s))$ is in minimal base form then $(\tilde{A}(z),\tilde{B}(z))$ is in minimal base form, with the same row degrees.
Proof. Let us prove (i); (ii) can be proved similarly. Consider $(A(z),B(z))$, with row degrees $\{r_i\}$. Suppose $(\tilde{A}_v(s),\tilde{B}_v(s))$, with row degrees ρ_i is not in minimal base form. Then there exists a forward mfd pair $(C_v(s),D_v(s))$ of the

same system, with row degrees $\{\rho_i'\}$, satisfying $\sum_1^m \rho_i' < \sum_1^m \rho_i$. Consider $(\tilde{C}(z),\tilde{D}(z))$ with row degrees $\{r_i'\}$. Then

$$(2.3.4\text{-}36)\quad \sum_1^m r_i' \leq \sum_1^m \rho_i' < \sum_1^m \rho_i \leq \sum_1^m r_i,$$

so $(A(z),B(z))$ is not in minimal base form.
If $(A(z),B(z))$ is in minimal base form then $(\tilde{\tilde{A}}(z),\tilde{\tilde{B}}(z)) = (A(z),B(z))$ which implies that $(A(z),B(z))$ has the same row degrees as $(A_v(s),B_v(s))$.

Q.E.D.

2.3.4-37. Definition. The high order coefficient matrix (hocm) of an mfd pair $(A_v(s),B_v(s))$ is $(\tilde{A}(0),\tilde{B}(0))$. Similarly the hocm of $(A(z),B(z))$ is $(\tilde{A}_v(0),\tilde{B}_v(0))$.

2.3.4-38. Theorem. If $(A(z),B(z))$ is an mfd pair in minimal base form of a (nonanticipative!) linear system then $A(0)$ is nonsingular.

Proof. Suppose $A(0)$ is singular. Then there exists a nonzero row vector $x = (x_1,\ldots,x_m)$ such that $xA(0) = 0$. Let i be such that $x_i \neq 0$ and for all j with $x_j \neq 0$, $r_i \geq r_j$ ($\{r_i\}$ are the row degrees of $[A(z)\ B(z)]$). Replace the i^{th} row of $[A(z)\ B(z)]$ by $x[A(z)\ B(z)]$. This gives us again an mfd pair, with row degrees equal to or less than the row degrees of $(A(z),B(z))$.
Now notice that because of the nonanticipation property of linear systems, it follows that the i^{th} row of the new mfd pair can be divided by z. This gives us an mfd pair with a smaller sum of the row degrees. (Note that there can be no row identically zero, because $\det A(z) \not\equiv 0$). Therefore $(A(z),B(z))$ is not in minimal base form.

Q.E.D.

2.3.4-39. Corollary. If $(A_v(s),B_v(s))$ is in minimal base form, then the sum of the row degrees is equal to deg $\det A_v(s)$.

Proof. $\det A_v(s) = s^{\Sigma\rho_i} \det \tilde{A}(0)$ + lower degree terms. Because $\tilde{A}(0)$ is nonsingular the result follows.

Q.E.D.

Let $(A_v(s),B_v(s))$ be in minimal base form. $\{y_t\},\{u_t\}$ satisfy the equations

(2.3.4-40) $A_v(S)y_t = B_v(S)u_t$

and therefore (compare section 2.3.3)

(2.3.4-41) $A_v(S)y_{t|t_o-1} = 0 \qquad \forall t \geq t_o.$

It will turn out to be useful to apply a change of basis in the output alphabet Y, such that w.r.t the new basis, the outputs are given by

(2.3.4-42) $w = \tilde{A}(0)y.$

The equations (2.3.4-41) now become

(2.3.4-43) $A_v(S)\tilde{A}(0)^{-1}w_{t|t_o-1} = 0 \qquad \forall t \geq t_o.$

Notice that the row degrees of any mfd pair are unchanged by such a change of coordinates. The hocm of $A_v(S)\tilde{A}(0)^{-1}$ is the identity matrix. Let us now consider the set of equations

(2.3.4-44) $A_v(S)\tilde{A}(0)^{-1}x_t = 0, \; t \geq t_o.$

It now follows easily that the dimension of the linear space of solutions of (2.3.4-44) is equal to the sum of the row degrees of $A_v(S)$. (Compare the discussion in section 2.3.3.) Equation (2.3.4-43) tells us that the abstract state space is a subspace of the space of solutions. Therefore the dimension of this space of solutions is larger than or equal to n, and equality holds iff the space of solutions of (2.3.4-44) is exactly the abstract state space. Now recall the definition (2.3.3-2) of the sequence $\{t_i\}_{i=o}^{\infty}$ and the subsequent definition of $\{s_i\}$. It is quite simple to see that a change of basis in the output alphabet Y, does not change the sequence $\{t_i\}$ and therefore $\{s_i\}$ too remains unchanged, which implies that the observability indices remain the same. The row degrees of $A_v(s)\tilde{A}(0)$ are exactly the observability indices of the system with outputs $w_r = \tilde{A}(0)y_t$. From all this it follows that

2.3.4-45. Theorem. If $(A_v(s),B_v(s))$ is in minimal base form then the bag of row degrees is equal to the bag of observability indices and the sum of these row degrees is equal to the McMillan degree n. The same holds for a backward

minimal base pair $(A(z),B(z))$. Furthermore one has, if $(A_v(s),B_v(s))$ is in minimal base form: $n = \deg \det A_v(s)$.

2.3.4-46. Remark. It is now clear that the arma model constructed in section 2.3.3 corresponds to an mfd pair in minimal base form. This shows immediately the existence of a minimal base.

2.3.5. Kronecker indices and the specialization order

It is well-known that the observability indices of a finite dimensional linear system come about as the Kronecker indices of a related pencil of matrices. (see e.g. [Ka 80], p. 413).

In the following we want to stress the importance of the Kronecker indices in the determination of the possible limit points of a sequence or family of mfd-pair models (or arma models). A major role is played by a partial order on the set of all Kronecker indices called the specialization order. It will be shown that a model that is the limit of a sequence of arma models with fixed Kronecker indices, will itself have Kronecker indices that are equally or more 'special' in the given order.

Recall the definitions of the sequences $\{t_i\},\{\delta_i\}$ and $\{p_k\}_{k=1}^{m}$, cf. (2.3.3-2), (2.3.3-4):

$$(2.3.5\text{-}1)\quad \begin{cases} t_0 = 0 \\ t_i = \operatorname{rk} \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ \vdots & \vdots & \vdots & \\ H_i & H_{i+1} & H_{i+2} & \cdots \end{bmatrix} & , \forall i \geq 1; \end{cases}$$

$$(2.3.5\text{-}2)\quad \begin{cases} s_0 = m, \\ s_i = t_i - t_{i-1} & , \forall i \geq 1. \end{cases}$$

The definition of $\{p_k\}_{k=1}^{m}$ is given by the rule

$$(2.3.5\text{-}3)\quad |\{k \mid p_k = i-1\}| = s_{i-1} - s_i,\ i = 1,2,3,\ldots .$$

So $\{p_k\}$ is determined up to permutation (i.e. it is a bag). Without loss of generality, let us assume that

(2.3.5-4) $p_1 \leq p_2 \leq p_3 \cdots \leq p_m$.

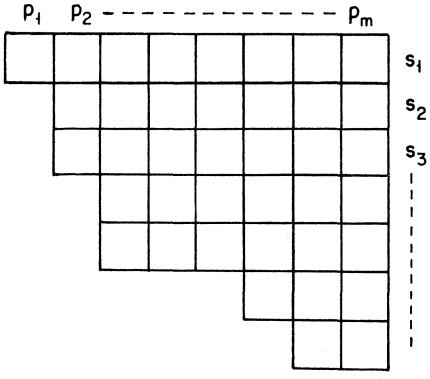
It follows that

(2.3.5-5) $s_i = |\{k | p_k \geq i\}|$, $i = 1,2,\ldots$

It can be shown that

(2.3.5-6) $p_k = |\{i | s_i \geq m-k+1\}|$, $k = 1,2,\ldots,m$.

In fact this can be seen most easily by considering the socalled Young diagram



The length of the i^{th} row is s_i, the length of the j^{th} column is p_j.

2.3.5-7. Definition. $r_k = \sum_{j=1}^{k} p_j$.

We now define a partial order as follows:

2.3.5-8. Definition (a) Let $s = \{s_i\}_{i=o}^{\infty}$, $s' = \{s'_i\}_{i=o}^{\infty}$ be two nonincreasing sequences taking their values in $\{0,1,\ldots,m\}$. Let $t = \{t_i\}_{i=o}^{\infty}$ be defined by $t_i = \sum_{j=o}^{i} s_j$ and $t' = \{t'_i\}_{i=o}^{\infty}$ by $t'_i = \sum_{j=o}^{i} s'_j$. Then we define the partial order $\underset{\sim}{\leq}$ by the rule: $s \underset{\sim}{\leq} s'$ iff $t \leq t'$ i.e. iff $t_i \leq t'_i$ $\forall i$.
(b) Let $p = \{p_k\}_{k=1}^{m}$, $p' = \{p'_k\}_{k=1}^{m}$ be two nondecreasing sequences, taking their values in the set of nonnegative integers $\{0,1,2,\ldots\}$, and let $r = \{r_k\}_1^m$ and $r' = \{r'_k\}_1^m$ be defined by $r_k = \sum_{j=1}^{k} p_j$; $r'_k = \sum_{j=1}^{k} p'_j$. Then we define the partial order $\underset{\sim}{\leq}$ by the rule

$p \lesssim p'$ iff $r \leq r'$, i.e. iff $r_i \leq r'_i$, $i = 1,2,\ldots,m$.

We will now show that the partial orderings defined in (a) and (b) are 'essentially the same', in the sense that if s and p are related by (2.3.5-5) and (2.3.5-6) and s' and p' similarly, then $s \lesssim s'$ iff $p \lesssim p'$.

2.3.5-8. Theorem. If s, s', p, p' are as in the previous definition, and s and p are related as in (2.3.5.5) and (2.3.5.6), and s' and p' similarly, then

(2.3.5-9) $\quad s \lesssim s'$ iff $p \lesssim p'$.

Proof. First we will show that $s \lesssim s'$ implies $p \lesssim p'$. We will show this with induction with respect to $k := |\{i | s_i \geq 1\}|$. If $k = 1$, then $1 \leq s_1 \leq s'_1$ and

$$p_i = \begin{cases} 0 & \text{if } 1 \leq i \leq m-s_1, \\ 1 & \text{if } m-s_1 < i \leq m, \end{cases} \quad \text{and} \quad p'_i = \begin{cases} 0 \text{ if } 1 \leq i \leq m-s'_i, \\ 1 \text{ if } m-s'_i < i \leq m. \end{cases}$$

It follows that

$$\begin{cases} p_i = p'_i = 0 & \text{if } 1 \leq i \leq m-s'_i, \\ 0 = p_i < p'_i = 1 & \text{if } m-s'_i < i \leq m-s_i \text{ and} \\ p_i = p'_i = 1 & \text{if } m-s_i < i \leq m. \end{cases}$$

So $p_i \leq p'_i$ holds for all i. This implies $r_i \leq r'_i$ for all i, and therefore $r \leq r'$. This in turn implies, by definition, $p \lesssim p'$. Now suppose the implication ($s \lesssim s' \Rightarrow p \lesssim p'$) holds for $k = k_o$ (induction hypothesis). Let p(k), r(k) denote the sequences of indices that correspond to the finite sequence $s_1, s_2, \ldots, s_k$; i.e.

$$p_i(k) := |\{j | m-i+1 \leq s_j \text{ and } j \leq k\}|$$

and

$$r_j(k) := \sum_{i=1}^{j} p_i(k).$$

Let $s \lesssim s'$. Then $s(k) \lesssim s'(k)$ for all k, so certainly $s(k_o) \lesssim s'(k_o)$ and

$s(k_o+1) \lesssim s'(k_o+1)$. The induction hypothesis implies that $p(k_o) \lesssim p'(k_o)$. Let us now distinguish two possibilities

(1) $s_{k_o+1} \leq s'_{k_o+1}$,

(2) $s_{k_o+1} > s'_{k_o+1}$.

ad(1). Now $m-s_{k_o+1}+1 \geq m-s'_{k_o+1} + 1$. One has

(i) $$\begin{cases} p_i(k_o+1) = p_i(k_o) \\ p'_i(k_o+1) = p'_i(k_o) \end{cases}, \text{ if } 1 \leq i \leq m-s'_{k_o+1} + 1,$$

(ii) $$\begin{cases} p_i(k_o+1) = p_i(k_o) \\ p'_i(k_o+1) = p'_i(k_o) + 1 \end{cases}, \text{ if } m-s'_{k_o+1}+1 < i \leq m-s_{k_o+1} + 1, \text{ and}$$

(iii) $$\begin{cases} p_i(k_o+1) = p_i(k'_o) + 1 \\ p'_i(k_o+1) = p'_i(k_o) + 1 \end{cases}, \text{ if } m-s_{k_o+1} \leq i \leq m.$$

Using the induction hypothesis, it follows that

(i) $$r_i(k_o+1) = r_i(k_o) \leq r'_i(k_o) = r'_i(k_o+1), \text{ if } 1 \leq i \leq m-s'_{k_o+1} + 1,$$

(ii) $$r_i(k_o+1) = r_i(k_o) \leq r'_i(k_o) \leq r'_i(k_o) + \{i-(m-s'_{k_o+1}+1)\} = r'_i(k_o+1),$$

if $m-s'_{k_o+1}+1 \leq i \leq m-s_{k_o+1} + 1$ and

(iii) $$r_i(k_o+1) = r_i(k_o) + \{i-(m-s_{k_o+1}+1)\} \leq r'_i(k_o) + \{i-(m-s_{k_o+1}+1)\} \leq$$

$$\leq r'_i(k_o) + \{i-(m-s'_{k_o+1}+1)\} = r'_i(k_o+1), \text{ if } m-s_{k_o+1} + 1 < i \leq m.$$

It follows that

$$r_i(k_o+1) \leq r'_i(k_o+1), \; i = 1,2,\ldots,m,$$

i.e. $r(k_o+1) \leq r'(k_o+1)$ and so, by definition $p(k_o+1) \lesssim p'(k_o+1)$.

ad 2. This is the 'nontrivial' possibility. Now $m-s_{k_o+1}+1 < m-s'_{k_o+1} + 1$.

In the following it is used that $p_i(k_o+1) \neq p_i(k_o)$ iff $p_i(k_o+1) = k_o+1$. This follows from the fact that $\{s_k\}$ is monotonically nonincreasing.

(i) $$\begin{cases} p_i(k_o+1) = p_i(k_o) \\ p'_i(k_o+1) = p'_i(k_o) \end{cases}, \quad \text{if } 1 \leq i \leq m-s_{k_o+1} + 1,$$

(ii) $$\begin{cases} p_i(k_o+1) = p_i(k_o) + 1 = k_o + 1 \\ p'_i(k_o+1) = p'_i(k_o) \leq k_o + 1 \end{cases}, \text{ if } m-s_{k_o+1}+1 < i \leq m-s'_{k_o+1} + 1,$$

(iii) $$\begin{cases} p_i(k_o+1) = p_i(k_o) + 1 = k_o+1 \\ p'_i(k_o+1) = p'_i(k_o) + 1 = k_o+1 \end{cases}, \quad \text{if } m-s'_{k_o+1}+1 < i \leq m.$$

Using the induction hypothesis, and the fact, that $r_m(k_o+1) = t_{k_o+1}$ etc, it follows that

(i) $$r_i(k_o+1) = r_i(k_o) \leq r'_i(k_o) = r'_i(k_o+1), \text{ if}$$

$$1 \leq i \leq m-s_{k_o+1} + 1,$$

(ii) and (iii)

$$r_i(k_o+1) = r_m(k_o+1) - (m-i)(k_o+1) = t_{k_o+1} - (m-i)(k_o+1) \leq$$

$$\leq t'_{k_o+1} - (m-i)(k_o+1) \leq r'_m(k_o+1)-(m-i)(k_o+1) \leq$$

$$\leq r'_i(k_o+1), \text{ if } m-s_{k_o+1}+1 \leq i \leq m.$$

This proves one side of the implication in the theorem.
We now want to show that $p \lesssim p'$ implies $s \lesssim s'$. This is done with induction with respect to m.

Let $s_i(m) := \begin{cases} s_i-m \text{ if } s_i > m, \\ 0 \quad \text{if } s_i \leq m, \end{cases}$

and $t_j(m) := \sum_{i=1}^{j} s_i(m)$, etc.

We assume $p \lesssim p'$.
Induction hypothesis: $s(m_o) \lesssim s'(m_o)$.
We distinguish two possibilities:

(1) $p_{m_o+1} \leq p'_{m_o+1}$,

(2) $p_{m_o+1} > p'_{m_o+1}$.

ad(1). In this case

(i) $s_i(m_o+1) = s_i(m_o) + 1$ and $s'_i(m_o+1) = s'_i(m_o) + 1$ if $1 \leq i \leq p_{m_o+1}$,

(ii) $s_i(m_o+1) = s_i(m_o) = 0$ and $s'_i(m_o+1) = s'_i(m_o) + 1$ if $p_{m_o+1} < i \leq p'_{m_o+1}$,

and

(iii) $s_i(m_o+1) = s_i(m_o) = 0$ and $s'_i(m_o+1) = s'_i(m_o) = 0$ if $p_{m_o+1} < i$.

One has

(i) $t_i(m_o+1) = t_i(m_o) + i \leq t'_i(m_o) + i = t'_i(m_o+1)$ if $1 \leq i \leq p_{m_o+1}$,

(ii) $t_i(m_o+1) = t_i(m_o) + p_{m_o+1} \leq t'_i(m_o) + p_{m_o+1} \leq$

$t'_i(m_o) + i = t'_i(m_o+1)$ if $p_{m_o+1} < i \leq p'_{m_o+1}$, and

(iii) $t_i(m_o+1) = t_i(m_o) + p_{m_o+1} \leq t'_i(m_o) + p_{m_o+1} \leq$

$\leq t'_i(m_o) + p'_{m_o+1} = t'_i(m_o+1)$ if $p'_{m_o+1} < i$.

So it follows that $t \leq t'$ iff $s \lesssim s'$.

ad(2) In this case

(i) $s_i(m_o+1) = s_i(m_o) + 1$ and $s'_i(m_o+1) = s'_i(m_o) + 1$ if $1 \leq i \leq p'_{m_o+1}$,

(ii) $s_i(m_o+1) = s_i(m_o) + 1$ and $s'_i(m_o+1) = s'_i(m_o) = 0$ if $p'_{m_o+1} < i \leq p_{m_o+1}$,

(iii) $s_i(m_o+1) = s_i(m_o) = 0$ and $s'_i(m_o+1) = s'_i(m_o) = 0$ if $p_{m_o+1} < i$.

One has

(i) $t_i(m_o+1) = t_i(m_o) + i \leq t'_i(m_o) + i = t'_i(m_o+1)$ if $1 \leq i \leq p'_{m_o+1}$,

(ii) and (iii)

$$t_i(m_o+1) \leq t_\infty(m_o+1) \leq t'_\infty(m_o+1) = t'_{p'_{m_o+1}}(m_o+1) = t'_1(m_o+1) \text{ if } p'_{m_o+1} \leq i.$$

Q.E.D.

2.3.5-10. Remarks. (i) Notice that $m \geq s_1 \geq s_2 \ldots \geq 0$ but $0 \leq p_1 \leq p_2 \ldots \leq p_m$. This is, however, rather natural from our point of view, because if you go through the rows of the Hankel matrix you first find s_1, then s_2, etc.; but also if you go through the rows of the Hankel matrix to determine the observability indices p_k, you first find the smallest, i.e. p_1, then a second one, p_2, greater than or equal to p_1 etc.

(ii) Within the set $\{s \mid t_\infty = \sum_{i \geq 1} s_i = n\}$, n fixed, the partial order $\lesssim$ is called the dominance order.

(iii) Within the set $\{p \mid r_m = n\}$, n fixed, the partial order $\lesssim$ is called the specialization order. (It should be noted that it is apparently more usual to order the partition of n in a nonincreasing way, however, this is rather unnatural from our point of view and unessential. Compare [Haz-M]). In fact we will choose the word specialization order to denote both orders, which are after all essentially the same as was stated in the previous theorem. The reason for choosing this name is given by the following theorem:

2.3.5-11. Theorem. Let $t = \{t_i\}_o^\infty$ be fixed, and let M(t) be the set of all block Hankel-matrices $H = (H_{i+j-1})_{i,j\text{-block}}$ for which

$$\begin{cases} t_o = 0, \\ t_i = \mathrm{rk} \begin{bmatrix} H_1 & H_2 & \cdots \\ H_2 & H_3 & \cdots \\ \vdots & \vdots & \\ H_i & H_{i+1} & \cdots \end{bmatrix}, \quad i = 1,2,\ldots \end{cases}$$

Let the topology on M be the topology of pointwise convergence (componentwise convergence). Then for the closure $\overline{M(t)}$ of M(t) one has

$$\overline{M(t)} \subseteq \bigcup_{t' \leq t} M(t').$$

Proof. Consider a convergent sequence $\{\overset{\circ}{H}(\ell)\}_{\ell=1}^{\infty}$ of Hankelmatrices from M(t). It is wellknown that the limit of a sequence of matrices has rank smaller than or equal to the limsup of the ranks of the matrices. (This can be easily shown using the fact that a matrix has rank $\leq t_i$ iff all $(t_i+1) \times (t_i+1)$ submatrices have determinant equal to zero. It is clear that the limit of a sequence of zero determinants is zero, etc). Therefore if $t(\infty)$ denotes the t-sequence of the limit $\overset{\circ}{H}(\infty)$, then $t(\infty) \leq t$.

Q.E.D.

2.3.5-12. Remark. It is highly probable that no t' for which $t' \leq t$, can be left out of the r.h.s. of the inclusion in the previous theorem. The related statement for the (special) case of a reachable pair has been proved in [Ha-M81]. It would lead us too far to go into that here.

2.3.5-13. Corollary. Let, by some abuse of notation, M(p) denote the set of linear system with observability indices $p = (p_1,p_2,\ldots,p_m)$. Using the same topology as above, namely the topology of pointwise convergence in the corresponding Hankel-matrices, it follows for the closure $\overline{M}(p)$ of M(p) that

$$\overline{M(p)} \subseteq \bigcup_{p' \lesssim p} M(p').$$

This corollary shows the reason why we call $\lesssim$ the specialization order: $p' \lesssim p$ holds for all p' which turn up as limiting cases of systems with observability indices p.

2.3.5-14. Remark. One can give examples in which $\overline{M(p)} \neq M(p)$, so that there are $p' \neq p$, $p' \lesssim p$, which cannot be left out of the inclusion in the

corollary. We will not go into that here (see remark (2.3.5-12) which makes a much stronger statement). It follows that if one specifies the observability indices, (i.e. the maximum lags in the equations of an arma model in minimal base form), then it is recommendable to include in the model space all models which correspond to more special (in the sense of the partial order) observability indices! In fact with respect to this issue the state space model appears to be easier to work with than the arma model, because in the case of arma models, one has to consider the whole set $\{p' \mid p' \lesssim p\}$ of possible lag specifications, and it seems difficult to avoid, that for each lag specification a separate identification algorithm has to be done. In the state space model, the McMillan degree can only remain the same or decrease in a limit point (this also follows from (2.3.5-11)). Of course the McMillan degrees are natural numbers and therefore completely ordered. This makes it simpler than the partial order of the observability indices. Furthermore, degeneration to a smaller McMillan degree can be observed from the observability and reachability matrices of the systems.

2.4. Stochastic linear systems; various representations

2.4.1. Stochastic state space models

From the definition in section 2.2 it follows that a stationary, ergodic, linear finite dimensional, Gaussian, zero mean, stochastic process can be represented by a deterministic, linear finite dimensional time invariant system with Gaussian white noise inputs.

A state space representation of such a system is of the form

$$(2.4.1\text{-}1)\quad \begin{cases} x_{t+1} = Ax_t + Bw_t, & x_t \in \mathbb{R}^n \\ y_t = Cx_t + Dw_t, & y_t \in \mathbb{R}^m, w_t \in \mathbb{R}^{m'}, \end{cases}$$

$\{w_t\}$ Gaussian white noise, $w_t \sim N(0,\overline{\Omega})$, $\overline{\Omega}$ a positive definite matrix (more generally, $\overline{\Omega}$ is positive semi-definite; but if $\overline{\Omega}$ is singular, there exists a w_t' and a rectangular matrix U such that $w_t = Uw_t'$, $w_t' \sim N(0,\overline{\Omega}_1)$ and $\overline{\Omega}_1$ positive definite, rk $\overline{\Omega}_1$ = rk $\overline{\Omega}$. Substituting $w_t = Uw_t'$ shows that by replacing Bw_t by BUw_t' and Dw_t by DUw_t', we can reduce to the case where $\overline{\Omega}$ is positive definite). Without loss of generality we will assume that (A,B,C,D) forms a minimal realization, i.e. n = McMillan degree of the deterministic system given by (A,B,C,D).

From the stationarity of the process, it follows that the variance P_t of the

state vector x_t must be independent of time:

(2.4.1-2) $\exists$ P positive semi-definite, such that $\forall t$: $Ex_t x_t^T = P$.

From (2.4.1-1) it follows that P satisfies the Lyapunov equation:

(2.4.1-3) $P = APA^T + BB^T$.

Here it is used that x_t is measurable with respect to the σ-algebra $\sigma(w_{t-1}, w_{t-2}, \ldots)$ of past input values (nonanticipation property of linear systems). Now suppose $\sigma(A)$ contains an element λ with $|\lambda| \geq 1$; let x^* be a corresponding nonzero left eigenvector of A. Then

(2.4.1-4) $x^*Px = x^*AP(x^*A)^* + x^*BB^*x = |\lambda|^2 x^*Px + x^*BB^*x \geq x^*Px + x^*BB^*x$.

It follows that $x^*B = 0$, and therefore $x^*R(A,B) = 0$ (for the definition of the reachability matrix R(A,B), see (2.3.2-4)). But this is in contradiction with the fact that (A,B,C,D) is a minimal realization (see (2.3.2-6)).
So we have:

2.4.1-5. Lemma. $\sigma(A) \subseteq D(0,1) = \{z \in \mathbb{C}; |z| < 1\}$.

2.4.1-6. Remark. Recall that we work with zero initial conditions. The motivation is the following. Because of our assumption of ergodicity, nonzero initial conditions have to be deterministic. The effects of these can be treated apart from the stochastic part of the model, using superposition. We will not go into that further here.

We now want to show that the necessary condition in the lemma is also sufficient to obtain a well-defined stochastic linear system. Notice that for a stochastic system of the form (2.4.1-1) it is not sufficient to use a deterministic linear system (A,B,C,D) working on the input space $\overset{\circ}{U}_o$, because with probability one, the sequence $\{w_k | k \leq t\}$ has an infinite number of nonzero values, i.e. $\{w_k\} \notin \overset{\circ}{U}_o$. Therefore we have to specify further the input-output mapping, as before denoted by F (see the discussion in ad(ii) at the end of section 2.3.2), and the set of inputs $\overset{\circ}{U}$ that is involved.

2.4.1-7. Definition. (a) The input-output mapping of the system that realizes the stochastic process is given by the formula

$$F: \{w_t\}_t \longmapsto \{\sum_{i=0}^{\infty} H_i w_{t-i}\}_t,$$

where H_i are the impulse response matrices, or Markov matrices, as before.

(b) The set of inputs $\overset{\circ}{U}$ is chosen to be all $\{w_t\}$, for which $\sum_{i=0}^{\infty} H_i w_{t-i}$ converges for all $t \in \mathbb{Z}$.

It will now be shown that if $\{w_t\}$ is a Gaussian white noise process and if $\sigma(A) \subseteq D(0,1)$, then almost surely $\{w_t\} \in \overset{\circ}{U}$.
First we have the following

2.4.1-8. Lemma. Let $\{w_t\}_{t\in\mathbb{Z}}$ be Gaussian white noise (or more generally: $\{w_t\}$ i.i.d., $Ew_t = 0$ and $E\|w_t\|^2 = \sigma^2 > 0$). Then the power series

$$f_{t_o}(z) = \sum_{t\leq t_o} w_t z^{t_o - t}$$

has radius of convergence $R = 1$, for all $t_o \in \mathbb{Z}$, almost surely. (If $f_{t_o}(z)$ has $R = 1$ for one value of t_o, then this holds for all $t_o \in \mathbb{Z}$).

Proof. The proof of this lemma will be given in a sequence of steps.
(i) First we will prove the following: If $\lambda \in (0,1)$ then

$$\lim_{t\to-\infty} w_t \lambda^{t_o - t} = 0 \text{ a.s.}$$

This can be shown as follows: Let $V(t,\varepsilon,\lambda) := \{\omega; \|w_t\|\lambda^{t_o - t} \geq \varepsilon\}$. Then

$$V(\varepsilon,\lambda) := \{\omega \mid \limsup_{t\to-\infty} \|w_t\|\lambda^{t_o - t} \geq \varepsilon\} =$$

$$\lim\sup_{t\to-\infty} V(t,\varepsilon,\lambda)(:= \bigcap_{s=-\infty}^{t_o} \bigcup_{t=-\infty}^{s} V(t,\varepsilon,\lambda)).$$

Then $V(\lambda) := \bigcup_{\varepsilon>0} V(\varepsilon,\lambda)$ is the set of all $\omega \in \Omega$ which correspond to a sequence

$\{\|w_t\|\lambda^{t_o-t}\}_{t \leq t_o}$ that for $t \to -\infty$ diverges, or converges to a nonzero value.

So $V(\lambda)^c := \Omega \backslash V(\lambda)$ is the set of all $\omega \in \Omega$ which correspond to a sequence

$\{w_t\lambda^{t_o-t}\}_{t \leq t_o}$ that converges to zero for $t \to -\infty$. And finally

$V := \bigcup_{\lambda \in (0,1)} V(\lambda)$ is the set of all ω for which a $\lambda \in (0,1)$ exists such that $\{w_t\lambda^{t_o-t}\}_{t \leq t_o}$ does not converge to zero. And V^c is the set of all ω for which $\{w_t\lambda^{t_o-t}\} \to 0$ for all $\lambda \in (0,1)$.

We will show that $P(V) = 0$. Because

$$\varepsilon_1 > \varepsilon_2 \text{ implies } V(\varepsilon_1,\lambda) \subseteq V(\varepsilon_2,\lambda), \quad \forall \lambda \in (0,1)$$

and

$$\lambda_1 > \lambda_2 \text{ implies } V(\lambda_1) \subseteq V(\lambda_2),$$

one has

$$V(\lambda) = \bigcup_{\substack{\varepsilon>0 \\ \varepsilon \in Q}} V(\varepsilon,\lambda)$$

and $V = \bigcup_{\lambda \in (0,1) \cap Q} V(\lambda)$.

Therefore V is a countable union of sets $V(\varepsilon,\lambda)$. This implies that it is sufficient to show

$$\forall \varepsilon > 0, \forall \lambda \in (0,1): P(V(\varepsilon,\lambda)) = 0.$$

This can be shown using the Chebyshev inequality together with the lemma of Borel-Cantelli:

$$(2.4.1\text{-}9) \quad P(V(t,\varepsilon,\lambda)) = P\{\|w_t\| \geq \varepsilon\lambda^{-(t_o-t)}\} \leq \varepsilon^{-2}\lambda^{2(t_o-t)} \quad \text{(Chebyshev)}.$$

Therefore

(2.4.1-10) $\sum_{t\leq t_o} P(V(t,\varepsilon,\lambda)) \leq \varepsilon^{-2}(1-\lambda^2)^{-1} < \infty.$

With the lemma of Borel-Cantelli (see e.g. [Bau], p.168) it follows that $P(V(\varepsilon,\lambda)) = 0$ indeed.

So $P(V^c) = 1$, and this shows (i).

(ii) $\{w_t\}_{t\leq t_o}$ does not converge to zero, almost surely. This follows from the fact that $\exists\varepsilon > 0$ such that $P(\|w_t\| > \varepsilon) = p_\varepsilon > 0$; because the w_t are independent, the second part of the lemma of Borel-Cantelli gives the result.

(iii) Define $W = V^c \setminus \{\omega \mid \lim_{t\to-\infty} w_t = 0\}$. Then (i) and (ii) imply $P(W) = 1$. We will now show that $\forall\omega \in W$, the corresponding power series $\sum_{t\leq t_o} w_t z^{t_o-t}$ has radius of converge $R = 1$ (for each $t_o \in \mathbb{Z}$). Let $\omega \in W$. Then $\{w_t\}$ does not converge to 0, for $t \to -\infty$ so the radius of convergence of the power series must be smaller than or equal to one. It remains to show that if $|z| < 1$, then

$\sum_{t\leq t_o} w_t z^{t_o-t}$ converges.

Let $\lambda \in (|z|,1)$, and $\mu := |z|/\lambda < 1$. Because $\omega \in W \subseteq V^c$ it follows that $\lim_{t\to-\infty} w_t\lambda^{t_o-t} = 0$, so $\{\|w_t\|\lambda^{t_o-t}\}_{t\leq t_o}$ is bounded, say by $M > 0$. It follows that

$$\sum_{t\leq t_o} \|w_t\||z|^{t_o-t} \leq \sum_{t\leq t_o} M\mu^{t_o-t} = M(1-\mu)^{-1} < \infty,$$

so $\sum_{t\leq t_o} w_t z^{t_o-t}$ is convergent. It follows that $R = 1$.

Q.E.D.

2.4.1-11. Notation. Let $\overset{\circ}{U}_1$ denote the set of all sequences $\{u_t\}_{t\in\mathbb{Z}}$ such that the power series $\sum_{t\leq t_o} u_t z^{t_o-t}$ has radius of convergence equal to one (if this holds for one $t_o \in \mathbb{Z}$, then it holds for all $t_o \in \mathbb{Z}$).
From the previous lemma (2.4.1-8), it follows that $P(\overset{\circ}{U}_1) = 1$. We now arrive at the desired result.

2.4.1-12. Theorem. Consider the linear system (2.4.1-1), (2.4.1-7)(a), where $\sigma(A) \subseteq D(0,1)$ (see (2.4.1-5). Then $\sum_{i=o}^{\infty} H_i w_{t-i}$ converges for all $t \in \mathbb{Z}$, with probability one. (To be more specific $\sum_{i=o}^{\infty} H_i w_{t-i}$ converges for all $\{w_t\} \in \overset{\circ}{U}_1$).

Proof. Let $\lambda \in (0,1)$ be such that $\sigma(A) \subseteq D(0,\lambda)$. Then $\lim_{i\to\infty} \lambda^{-i} H_i = \lim_{i\to\infty} C(A\lambda^{-1})^i B = 0$, because $\sigma(A\lambda^{-1}) \subseteq D(0,1)$. Therefore $\exists M > 0$ such that $\sum_{i=o}^{\infty} \|H_i w_{t-i}\| \leq \sum_{i=o}^{\infty} \|H_i\lambda^{-i}\|.\lambda^i \|w_{t-i}\| \leq M \sum_{i=o}^{\infty} \lambda^i \|w_{t-i}\|$ which converges $\forall \{w_t\} \in \overset{\circ}{U}_1$ (this follows from lemma (2.4-18) and the extension to the multivariable case of the well-known fact that a power series converges absolutely within D(0,R), R being the radius of convergence).

Q.E.D.

2.4.1-13. Corollary. The equation (2.4.1-1) and the definition (2.4.1-7) together define a stochastic stationary ergodic system (with zero initial conditions) iff $\sigma(A) \subseteq D(0,1)$.

2.4.1-14. Definition. Let T(s) be a proper transfer function. If T(s) is square, and its inverse $T(s)^{-1}$ exists and is proper, then the system with this transfer function is called the inverse system.

2.4.1-15. Remarks. (i) In this definition T(s) is associated with a deterministic system, and the existence of the inverse of this deterministic system is considered.
(ii) If the system is given by a backward mfd pair (A(z),B(z)) in minimal base

form then the inverse exists if B(z) is square, and B(0) is nonsingular. The inverse is given by the backward mfd pair (B(z),A(z)).
(iii) If the system is given by

$$\begin{cases} x_{t+1} = Ax_t + Bu_t \\ y_t = Cx_t + Du_t, \end{cases}$$

$y_t \in \mathbb{R}^m$, $u_t \in \mathbb{R}^m$, $x_t \in \mathbb{R}^n$, and if D is nonsingular, then the inverse system exists and is given by

$$(2.4.1\text{-}16)\begin{cases} x_{t+1} = (A-BD^{-1}C)x_t + BD^{-1}y_t \\ u_t = -D^{-1}Cx_t + D^{-1}y_t. \end{cases}$$

(Here it is <u>not</u> necessary that $A-BD^{-1}C$ is asymptotically stable, because the inverse system is considered as a deterministic system).
For systems with square transfer matrix, one can define the socalled minimum phase property. Usually it is only done in the context of stochastic systems.

2.4.1-17. <u>Definition</u>. If a system with as many input channels as output channels has an inverse (in the sense of (2.4.1-14)) of which the spectrum lies in the <u>closed</u> unit disk, then it is called minimum phase.

2.4.1-18. <u>Remarks</u>. (i) For systems with nonsquare transfer matrix one can also define the minimum phase property, but we shall not go into that here.
(ii) In the next subsection it will be shown that a stochastic system of the type treated here can always be represented by a <u>minimum phase</u> (and asymptotically stable, see (2.4.1-13)) deterministic system with white noise inputs, i.e. (2.4.1-1) can be taken to be minimum phase without loss of generality.

2.4.2. <u>The spectral density matrix and spectral factorization</u>
Consider a stochastic system of the type treated in the previous subsection 2.4.1. It is a gaussian zero-mean process. Therefore it is completely determined by the covariance matrices

$$(2.4.2\text{-}1) \qquad \Gamma_k := Ey_{t+k}y_t^T, \ \forall k \in \mathbf{Z}, \ \forall t \in \mathbf{Z}.$$

Because of the stationarity of the process, Γ_k does not depend on t. Furthermore it is clear that

(2.4.2-2) $\quad \Gamma_k^T = \Gamma_{-k}, \ \forall k \in \mathbb{Z}.$

Using (2.4.1-1,2 and 3), Γ_k can be computed:

$$\Gamma_k = \begin{cases} CPC^T + D\bar{\Omega}D^T & \text{if } k = 0, \\ CA^{k-1}(B\bar{\Omega}D^T + APC^T) & \text{if } k \geq 1, \end{cases} \tag{2.4.2-3}$$

where P is as in (2.4.1-2) and (2.4.1-3). Because A is asymptotically stable (see previous subsection 2.4.1) it follows that

(2.4.2-4) $\quad \exists \lambda > 1$ such that $\lim_{k\to\infty} \lambda^k \Gamma_k = 0.$

Therefore, if $s \in \mathbb{C}$, $|s| < \lambda$, $\sum_{k=0}^{\infty} \Gamma_k s^k$ converges, and $\sum_{k=1}^{\infty} \Gamma_k s^k$ converges, and so if $\lambda^{-1} < |s| < \lambda$, then $\sum_{k=-\infty}^{\infty} \Gamma_k s^k = \left(\sum_{k=0}^{\infty} \Gamma_k s^k\right) + \left(\sum_{k=1}^{\infty} \Gamma_k s^{-k}\right)^*$ is well-defined.

2.4.2-5. Definition. $\Gamma(s) = \sum_{k\in\mathbb{Z}} \Gamma_k s^{-k}, |s| \in (\lambda^{-1}, \lambda) \neq \emptyset$ is called the spectral density matrix.

2.4.2-6. Remark. $\Gamma(s)$ is defined on the whole unit circle $\{s; |s| = 1\} =: C(0,1)$. Because $\Gamma(s)$ is analytic, $\Gamma(s)$ is completely determined by the restriction of Γ to the unit circle. Therefore, often Γ is only considered on the unit circle. One then considers the mapping $[0,2\pi) \to \mathbb{C}^{m\times m}$, $\lambda \longmapsto \Gamma(e^{i\lambda})$.

2.4.2-7. Theorem. Let T(s) denote the (forward) transfer matrix of (2.4.1-1), then

(2.4.2-8) $\quad \Gamma(s) = T(s)\bar{\Omega}T(s)^*, \quad \forall s \in C(0,1)$

Proof. Let, as before, $T(s) = \sum_{i=0}^{\infty} H_i s^{-i}$ (compare (2.3.4-1)), with $H_0 = D$; $H_i = CA^{i-1}B$; $i = 1,2,\ldots$. By convention, let $H_i = 0, \forall i < 0$. Then

$$\Gamma_k = Ey_{t+k}y_t^T = E\left(\sum_{i=0}^{\infty} H_i w_{t-i+k}\right)\left(\sum_{i=0}^{\infty} H_i w_{t-i}\right)^T =$$

$$= E\Big(\sum_{i=o}^{\infty} H_{i+k} w_{t-i}\Big)\Big(\sum_{i=o}^{\infty} H_i w_{t-i}\Big)^T = \sum_{i=o}^{\infty} H_{i+k} \overline{\Omega} H_i^T.$$

So $\Gamma_k s^{-k} = \sum_{i=o}^{\infty} H_{i+k} s^{-(i+k)} \overline{\Omega} (H_i s^{-i})^*$ and

$$\Gamma(s) = \sum_{k\in\mathbb{Z}} \Gamma_k s^{-k} = \sum_{k\in\mathbb{Z}} \sum_{i=o}^{\infty} H_{i+k} s^{-(i+k)} \overline{\Omega} (H_i s^{-i})^* =$$

$$= \sum_{k\in\mathbb{Z}} \sum_{i=o}^{\infty} H_k s^{-k} \overline{\Omega} (H_i s^{-i})^* = T(s)\, \overline{\Omega}\, T(s)^*.$$

Q.E.D.

2.4.2-9. Remark. We can also define a backward spectral density matrix, by

(2.4.2-10) $\Gamma_b(z) = \Gamma(z^{-1}) = \sum_{k\in\mathbb{Z}} \Gamma_k z^k, |z| \in (\lambda^{-1}, \lambda).$

From the theorem it follows that

(2.4.2-11) $\Gamma_b(z) = H(z)\overline{\Omega}H(z)^*,$

where $H(z) = \sum_{i=o}^{\infty} H_i z^i$, the backward transfer function (see (2.3.4-1)).

2.4.2-12. Corollary. A spectral density matrix $\Gamma(s)$ is positive semi-definite Hermitean for each $s \in C(0,1)$.

Proof. $\forall s \in C(0,1), \Gamma(s) = T(s)\,\overline{\Omega}\,T(s)^*$, and $\overline{\Omega}$ is positive definite symmetric.

Q.E.D.

2.4.2-13. Corollary. A spectral density matrix $\Gamma(s)$ is rational.

Proof. This follows from the fact that $T(s)$ is rational (see theorem (2.3.4-4)).

Q.E.D.

2.4.2-14. Corollary. $\Gamma(s)$ is analytic on $C(0,1)$.

Proof. This follows from the fact that the system described by $T(s)$ is

asymptotically stable. Q.E.D.

If Γ_o is singular and has rank m", one can find m" independent components of y_t, such that the other (m-m") components are linearly dependent on them. If we denote the vector of these m" components as $\tilde{y}_t$, the model (2.4.1-1) can be rewritten as

$$(2.4.2\text{-}15)\quad \begin{cases} x_{t+1} = Ax_t + Bw_t \\ \tilde{y}_t = \tilde{C}x_t + \tilde{D}w_t \\ y_t = E\tilde{y}_t \end{cases}$$

for some m×m" matrix E with m>m",rk(E) =m". We will restrict ourselves to studying the first two equations of (2.4.2-15). In other words we will assume

(2.4.2-16) Γ_o is positive definite.

Because $\Gamma_o = \sum_{k=o}^{\infty} H_k H_k^T$, it is equivalent to say that $rk[H_o \quad H_1 \ldots] = m$. (One could also express this by saying, that the observability indices p_k are all greater than or equal to 1). It is also equivalent to say that T(s) has full rank 'almost everywhere', it has full rank at all but finitely many points s. Now the following standard result on spectral factorization can be applied (see e.g. [An-M], pp. 240-241 or [Roz], Ch.2, 10).

2.4.2-17. Theorem. If $\Gamma(s)$ is a matrix valued function on $s \in C(0,1)$, with the properties:

(i) $\Gamma(s)$ is analytic on C(0,1) and rational and has full rank almost everywhere,

(ii) $\Gamma(s)$ is square, and real (i.e. $\Gamma(s) = \sum_{k\in\mathbb{Z}} \Gamma_k s^{-k}$ with each Γ_k a real matrix)

(iii) $\Gamma(s)$ is nonnegative definite Hermitean on C(0,1) (in fact positive definite Hermitean almost everywhere),

then there is a factorization of $\Gamma(s)$ as

$$(2.4.2\text{-}18)\quad \Gamma(s) = \bar{H}(s)\,\bar{\Sigma}\,\bar{H}(s)^*$$

where $\bar{H}(s)$ is a square, real (i.e. $\bar{H}(s) = \sum_k \bar{H}_k s^{-k}$ with each $\bar{H}_k$ a real matrix),

proper rational transfer matrix, all poles lie in $s \in D(0,1)$, $\bar{H}_o = I$, $\bar{H}(s)^{-1}$ is analytic in $\bar{D}(0,1)^c = \{s; |s| > 1, s \in \mathbb{C}\}$ (or equivalently $\bar{H}(s)$ has constant full rank on $\bar{D}(0,1)^c$), and $\bar{\Sigma}$ is positive definite symmetric. Moreover the factorization is unique.

2.4.2-19. Remarks. (i) One can also say that $\bar{H}(s)$ is the transfer function of an asymptotically stable, minimum phase system of the form (2.4.1-1) with $D = I$. (This defines an innovation representation; see section 2.4.4).
(ii) Note that $\Gamma(s)$ is a complete description of the stochastic process $\{y_t\}$. The theorem (2.4.3-17) implies that without loss of generality we can assume that the system of the form (2.4.1-1) with which we represent the process, is minimum phase. I.e. it is invertible (so a fortiori m=m') and its inverse has spectrum $\sigma(A-BC) \subseteq \bar{D}(0,1)$. From now on we will make this assumption. For later reference let us mention the following corollary.

2.4.2-20. Corollary. Suppose $T(s)$ is an $m \times m$ rational transfer matrix, $T(s) = \sum_{k=o}^{\infty} H_k s^{-k}$, with H_o positive definite symmetric. Then there exists an $\varepsilon > 0$ such that

$$\Gamma_\varepsilon(s) := \sum_{k=o}^{\infty} (H_k \varepsilon^k) s^{-k} + \sum_{k=1}^{\infty} (H_k^T \varepsilon^k) s^k, \quad s \in C(0,1)$$

is a spectral density matrix.

Proof. For all $\varepsilon \in (0,1]$, $\Gamma_\varepsilon(s)$ is analytic on $s \in C(0,1)$ and rational and hermitean. Now $\lim_{\varepsilon \downarrow o} \Gamma_\varepsilon(s) = H_o$, $\forall s \in C(0,1)$. Because H_o is positive definite, $\Gamma_\varepsilon(s)$ is continuous in ε and s, and $C(0,1)$ is compact, it follows that there exists a positive ε with $\Gamma_\varepsilon(s)$ positive definite Hermitean, for all $s \in C(0,1)$. Now apply theorem (2.4.2-17).

Q.E.D.

2.4.3. Hankel matrices of covariances

In the previous subsection we already noted that the sequence of covariances $\Gamma := \{\Gamma_k | k\in\mathbb{Z}\}$ completely determines the stochastic process that we are studying. Because $\Gamma_{-k} = \Gamma_k^*$, $\forall\ k \in \mathbb{Z}$, the sequence $\{\Gamma_k | k=0,1,2,\ldots\}$ determines the stochastic process completely. We will now consider block Hankel matrices of covariances. First we introduce some notation.

2.4.3-1. Notation. Let $H = \{H_k | k\in\mathbb{Z}\}$ be any sequence of equally sized matrices H_k. Then $\overset{\circ}{H}(H,k_o)$ will denote the Hankel matrix with (i,j)-block equal to $H_{k_o+i+j-2}$. I.e.

$$\overset{\circ}{H}(H,k_o) = \begin{bmatrix} H_{k_o} & H_{k_o+1} & H_{k_o+2} & \cdots \\ H_{k_o+1} & H_{k_o+2} & H_{k_o+3} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}. \tag{2.4.3-2}$$

Similarly, let $\overset{\circ}{T}(H,k_o)$ denote the block-Toeplitz matrix with (i,j)-block equal to H_{k_o+i-j}. I.e.

$$\overset{\circ}{T}(H,k_o) = \begin{bmatrix} H_{k_o} & H_{k_o+1} & H_{k_o+2} & \cdots \\ H_{k_o-1} & H_{k_o} & H_{k_o+1} & \cdots \\ H_{k_o-2} & H_{k_o-1} & H_{k_o} & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}. \tag{2.4.3-3}$$

If $H = \{H_k\}_{k\in K}$, where $K \subseteq \mathbb{Z}$, but $K \neq \mathbb{Z}$, then, as a convention, we will use the sequence $\bar{H} = \{\bar{H}_k\}_{k\in\mathbb{Z}}$, defined by $\bar{H}_k := H_k$ if $k \in K$ and $\bar{H}_k := 0$ if $k \in \mathbb{Z}\backslash K$, instead of H, in these definitions.

Consider the Hankel matrix $\overset{\circ}{H}(\Gamma,0)$. It is clear that this Hankel matrix determines the stochastic process completely, because it is in bijective correspondence with Γ. We will investigate its structure. To this end we will start with the fairly general situation of two stochastic systems (2.4.1-1), asymptotically stable and minimum phase with $\Sigma = I$ (without loss of generality), which have the same white noise sequence as input:

system Σ_1:

$$(2.4.3\text{-}4)\quad \begin{cases} x_{1,t+1} = A_1 x_{1,t} + B_1 w_t, & x_{1,t} \in \mathbb{R}^{n_1}, \\ y_t = C_1 x_{1,t} + D_1 w_t, & y_t \in \mathbb{R}^m,\ w_t \in \mathbb{R}^m \end{cases}$$

and

system Σ_2:

$$(2.4.3\text{-}5)\quad \begin{cases} x_{2,t+1} = A_2 x_{2,t} + B_2 w_t, & x_{2,t} \in \mathbb{R}^{n_2}, \\ z_t = C_2 x_{2,t} + D_2 w_t, & z_t \in \mathbb{R}^m. \end{cases}$$

The corresponding sequences of Markov matrices will be denoted by $H_1 := \{H_{1,k}\}_{k=0}^{\infty}$ (system Σ_1) and $H_2 := \{H_{2,k}\}_{k=0}^{\infty}$ (system Σ_2). Note that the dimensions n_1 and n_2 of the state spaces of Σ_1 and Σ_2 are allowed to be different. Now consider the covariances

$$(2.4.3\text{-}6)\quad \Gamma_j^{yz} := Ey_{t+j}z_t^T,\ j \in \mathbb{N}_o$$

and

$$(2.4.3\text{-}7)\quad \Gamma_j^{zy} := Ez_{t+j}y_t^T,\ j \in \mathbb{N}_o.$$

Form the corresponding Hankel matrix $\overset{\circ}{H}(\Gamma^{yz},1)$. The following result is useful:

2.4.3-8. Theorem. Rank $\overset{\circ}{H}(\Gamma^{yz},1) = n_1$. This is still true if Σ_1 is not minimum phase.

Proof. It is straightforward to show that

$$(2.4.3\text{-}9)\quad \overset{\circ}{H}(\Gamma^{yz},1) = \overset{\circ}{H}(H_1,1)\overset{\circ}{T}(H_2,0)^T.$$

(Because of the asymptotic stability of systems Σ_1 and Σ_2, all the series appearing in the computation of the infinite matrix product on the right hand side converge absolutely). Recall our definition (2.3.3-1) of the left kernel $\ell\ker \overset{\circ}{H}$ of a Hankel matrix $\overset{\circ}{H}$. From (2.4.3-9) it follows that

$$(2.4.3\text{-}10)\quad \ell\ker \overset{\circ}{H}(\Gamma^{yz},1) \supseteq \ell\ker \overset{\circ}{H}(H_1,1).$$

This implies that

(2.4.3-11) $\mathrm{rk}\overset{\circ}{H}(\Gamma^{yz},1) \leq \mathrm{rk}\overset{\circ}{H}(H_1,1) = n_1.$

We now set out to show that equality holds in (2.4.3-11), or, equivalently, in (2.4.3-10). Suppose in order to create a contradiction that equality does not hold in (2.4.3-10). I.e.

(2.4.3-12) $\exists\gamma \in \ell\ker\overset{\circ}{H}(\Gamma^{yz},1)\setminus \ell\ker\overset{\circ}{H}(H_1,1).$

Recall (see (2.3.2-4) and (2.3.2-5) etc.) that

(2.4.3-13) $\overset{\circ}{H}(H_1,1) = Q(C_1,A_1)R(A_1,B_1).$

Define

(2.4.3-14) $\xi := Q(C_1,A_1)^T\gamma.$

Then

(2.4.3-15) $\xi^T R(A_1,B_1) = \gamma^T\overset{\circ}{H}(H_1,1) \neq 0,$

and

(2.4.3-16) $\xi^T R(A_1,B_1)\overset{\circ}{T}(H_2,0)^T = \gamma^T\overset{\circ}{H}(\Gamma^{yz},1) = 0.$

Because for all $j = 0,1,2,\ldots$: $A_1^j R(A_1,B_1)$ is a submatrix of $R(A_1,B_1)$, it follows a fortiori that

(2.4.3-17) $\forall j \in \mathbb{N}_o$: $\xi^T A_1^j R(A_1,B_1)\overset{\circ}{T}(H_2,0)^T = 0.$

Consider $V := \mathrm{span}[\xi^T,\xi^T A_1,\xi^T A_1^2,\ldots] \neq 0$. The equation (2.4.3-17) implies

(2.4.3-18) $\forall v^T \in V$: $v^T R(A_1,B_1)\overset{\circ}{T}(H_2,0)^T = 0.$

Furthermore V is A_1-invariant:

(2.4.3-19) $VA_1 \subseteq V,$

and therefore the complexification $V(\mathbb{C})$ of V contains a nonzero eigenvector η^* with left eigenvalue λ(say) in the spectrum of A_1, so

(2.4.3-20) $\eta^* A_1 = \lambda\eta^*$, $\lambda \in D(0,1)$.

Because $R(A_1,B_1)$ has full row rank (this follows from the assumed minimality of (A_1,B_1,C_1)), one obtains:

(2.4.3-21) $0 \neq \eta^* R(A_1,B_1) = [\eta^* B_1 \ \lambda\eta^* B_1 \ \dots \ \lambda^k \eta^* B_1 \dots]$

and therefore

(2.4.3-22) $\mu^* := \eta^* B_1 \neq 0$.

At the other hand, from (2.4.3-18), it follows that

(2.4.3-23) $[\mu^* \ \lambda\mu^* \ \dots \ \lambda^k \mu^* \ \dots]\mathring{T}(H_2,0)^T = 0$

or, equivalently

(2.4.3-24) $\mu^* \sum_{k=0}^{\infty} H_{2,k}^T \lambda^k = 0$.

Let $H_2(z) = \sum_{k=o}^{\infty} H_{2,k} z^k$ denote the backward transfer matrix of Σ_2, then (2.4.3-24) implies that $H_2(\lambda)$ is singular. However, because $\lambda \in D(0,1)$, this is in contradiction with the fact that Σ_2 is minimum phase.

Q.E.D.

Now let us return to the situation of one stochastic process $\{y_t\}$, with sequence of covariances $\Gamma = \{\Gamma_k\}$ and with an asymptotically stable and minimum phase representation of the form (2.4.1-1), with sequence of Markov matrices $\{H_k\}$. (By convention $H_k = 0$ for all negative $k \in \mathbb{Z}$).
Then one can deduce

2.4.3-25. Corollary. $\text{Rk } \mathring{H}(\Gamma,k) = \text{rk } \mathring{H}(H,k)$, $\forall k \in \mathbb{Z}$.

Proof. Comparable to (2.4.3-9), it is straightforward to derive

(2.4.3-26) $\overset{\circ}{H}(\Gamma,k) = \overset{\circ}{H}(H,k)\overset{\circ}{T}(H,0)^T$.

Now apply the previous theorem with the system (A,B,C,D) as Σ_2 and the system with sequence of Markov matrices $\{H_j\}_{j=k}^{\infty}$ as Σ_1. So e.g. if $k \geq 1$, then Σ_1 is the system represented by

$$(2.4.3\text{-}27)\begin{cases} x_{t+1} = Ax_t + Bw_t, \\ y_t = CA^k x_t + CA^{k-1} Bw_t. \end{cases}$$

Then the corollary follows. Q.E.D.

2.4.3-28. Corollary. $\ell\ker \overset{\circ}{H}(\Gamma,k) = \ell\ker \overset{\circ}{H}(H,k)$.

Proof. From (2.4.3-26) it follows that

(2.4.3-29) $\ell\ker \overset{\circ}{H}(\Gamma,k) \supseteq \ell\ker \overset{\circ}{H}(H,k)$; the previous corollary shows that equality must hold (compare (2.4.3-10) and (2.4.3-11)).

Q.E.D.

2.4.3-30. Remarks. (i) This corollary shows that the autoregressive part of the system can be determined directly from the Hankel matrix of covariances (compare section 2.3.3), and so that the observability Kronecker indices, and the McMillan degree can be determined directly from the Hankel matrix of covariances.
(ii) The corollary is one of the most important tools in transferring partial realization results for mfd-models ('deterministic arma models') to stochastic arma models, in the next chapter.
(iii) The case k = 1 of corollary (2.4.3-25) implies that the minimal McMillan degree of a spectral factor can be found to be the rank of the Hankel matrix of covariances $\overset{\circ}{H}(\Gamma,1)$. In fact, this is a well-known result, cf. e.g. [An 73], or [FCG], Prop. 8.8, p. 188.

2.4.4. The innovations representation

Given a stochastic system of the form (2.4.1-1), one can compute the one-step ahead conditional expection of y_t, namely

$$E(y_{t+1}|y_t,y_{t-1},\ldots) = E(y_{t+1}|\sigma(y^t)),$$

where $\sigma(y^t)$ denotes the σ-algebra generated by $y^t = \{y_t,y_{t-1},y_{t-2},\ldots\}$.

2.4.4-1. Definition. $v_t = y_t - E(y_t|\sigma(y^{t-1}))$ is called the innovation at time t of the process $\{y_t\}$.

The computation of $E(y_t|\sigma(y^{t-1}))$ is done by the steady state Kalman filter. This is standard filtering theory, see e.g. [An-M].

It is possible to find a representation of the stochastic process by a system with the innovations as inputs and in which $x_{t+1} = E(x_{t+1}|\sigma(y^t))$:

$$(2.4.4\text{-}2)\quad \begin{cases} x_{t+1} = Ax_t + Bv_t \\ y_t = Cx_t + v_t \end{cases}$$

or equivalently

$$(2.4.4\text{-}3)\quad \begin{cases} x_{t+1} = (A\text{-}BC)x_t + By_t \\ v_t = -Cx_t + y_t \end{cases}.$$

One can show (see [An-M]) that $\mathrm{spec}(A\text{-}BC) \subseteq \bar{D}(0,1)$. ('spec(A-BC)' denotes the spectrum of A-BC).

2.4.4-4. Remark. If (A-BC) is not asymptotically stable, i.e. if (A-BC) has eigenvalues on the unit circle, then the input-output operator corresponding to (2.4.4-3) should be defined as a Cesaro sum (compare section 2.3.2, (ii) (b) after theorem (2.3.2-9)). We will not go into further detail here.

2.4.5. Arma-models

The last representation that we want to mention is the (stochastic) arma-

model.

(2.4.5-1) $A(L)y_t = B(L)w_t$,

$A(z), B(z)$ polynomials, $\{w_t\}$ gaussian white noise, as before. It is possible to choose for $\{w_t\}$ the innovations $\{v_t\}$. It then follows that

(2.4.5-2) $B_0 = I$.

(This follows from the fact that v_t is $\sigma(y^t)$-measurable).

2.4.5-3. Remark. Note that if $\{w_t\}$ is a gaussian white noise, then the same holds for $\{w_{t-k}\}$, for each $k \in \mathbb{Z}$. Therefore (2.4.5-1) is equivalent to

(2.4.5-4) $A(L)y_t = B(L)L^k w_t$,

for each $k \in \mathbb{Z}$. Of course, if $k \geq 0$, then $B(L)L^k$ is still a polynomial. If $B(0) = I$, as in the innovation representation, then $B(L)L^k$ is not a polynomial if $k < 0$. This remark is made for later reference.

Appendix 2A. Proof of theorem (2.3.4-3)

Proof. '$\Leftarrow$' If $\sigma(A) \subseteq D(0,1)$, then $\lim_{k\to\infty} A^k = 0$ (this is wellknown) and therefore $\lim_{k\to\infty} CA^kB = 0$.

'$\Rightarrow$' If there is an element $\lambda \in \sigma(A)$ with $|\lambda| \geq 1$ then $\{CA^kB\}_{k=o}^{\infty}$ does not converge to zero. This can be shown as follows: Let $x \in \mathbb{C}^n$ be a nonzero eigenvector corresponding to λ, and let $\xi_o,\xi_1,\ldots,\xi_{n-1} \in \mathbb{C}$ be such that $\sum_{k=o}^{n-1} A^kB\xi_k = x$. (Because (A,B,C) is minimal, (A,B) is reachable and so $[B\ AB\ \ldots\ A^{n-1}B]$ has full column rank. Therefore such a sequence $\xi_o,\xi_1,\ldots,\xi_{n-1}$ exists). Now consider the sequence of sums

$$\{\sum_{k=o}^{n-1} CA^{k+\ell}B\xi_k\}_{\ell=o}^{\infty} = \{\lambda^\ell Cx\}_{\ell=o}^{\infty}$$

There remain two possibilities, one of which can be ruled out:

(i) $Cx = 0$: this is not possible, because it would imply $CA^kx = 0$ for all $k \in \mathbb{N}_o$, and because (C,A) is observable, this would imply $x = 0$, which is not the case.

(ii) $Cx \neq 0$. Then $\{\lambda^\ell Cx\}_{\ell=o}^{\infty}$ does not converge to zero, because $|\lambda| \geq 1$.

Therefore $\{\sum_{k=o}^{n-1} CA^{k+\ell}B\xi_k\}_{\ell=o}^{\infty}$ does not converge to zero. This implies that CA^kB does not converge to zero for $k \to \infty$.

Q.E.D.

CHAPTER 3

PARTIAL REALIZATION OF ARMA (p,q) MODELS AND THE FINITE IDENTIFIABILITY PROBLEM

3.1. Partial realization of arma(p,q) models

3.1.1. Introduction

First let us describe the socalled partial realization problem. Suppose one has given a finite sequence of m×m' matrices $\{H_i\}_{i=o}^{I}$. If one interprets these as the first I+1 Markov-matrices of a deterministic linear system, one can raise the question how this sequence can be extended to give the sequence of all Markov matrices $\{H_i\}_{i=o}^{\infty}$. Clearly, without further restrictions each extension would do. Let it be required that the resulting sequence has finite McMillan degree, i.e.

$$(3.1.1\text{-}1) \qquad \overset{\circ}{rkH}(H,1) < \infty,$$

then it is possible to find a realization, i.e. a finite dimensional state space description (or, equivalently, an mfd), according to standard realization theory (see e.g. [KFA], Ch.10). Therefore, extension of $\{H_i\}_{i=o}^{I}$ to $\{H_i\}_{i=o}^{\infty}$ which obeys (3.1.1-1) is equivalent to finding a state space model which has $\{H_i\}_{i=o}^{I}$ as its first I+1 Markov matrices (also called impulse-response matrices). The problem to find a state space model of the form (2.3.2-2) with this property, is often called the partial realization problem. However, also the (equivalent) problem of finding an extension with property (3.1.1-1) will be called the partial realization problem, for obvious reasons. Each partial realization has a finite McMillan degree. An obvious partial realization is found by taking $H_{I+j} = 0$, for all $j = 1,2,\ldots$. It follows that there exists a minimal McMillan degree, among all the McMillan degrees of partial realizations of $\{H_i\}_{i=o}^{I}$. Any partial realization which has this minimal McMillan degree will be called a minimal partial realization. An important question is under which circumstances there exists a unique minimal partial realization.

Now consider a system with McMillan degree n, and sequence of Markov matrices $\{H_i\}_{i=o}^{\infty}$. Now suppose an upperbound of the McMillan degree is given, then one can raise the question for what values of I does there exist a unique minimal

partial realization of $\{H_i\}_{i=o}^{I}$. Of course, in that case the partial realization must be equal to the true sequence of Markov matrices $\{H_i\}_{i=o}^{\infty}$. Instead of an upper bound on the McMillan degree, one can also have other structural information about the system, like e.g. upper bounds on the Kronecker indices, etc. In our case the structural information will be (upperbounds of) the degrees p, resp. q of the autoregressive, resp. the moving average part of an arma(p,q) representation of the system. Our main interest here is with the stochastic case, because that gives us a solution to the finite identifiability problem which is treated in section 3.2. However, we will start with the deterministic case, because the techniques are similar and it will clarify the exposition.

3.1.2. The deterministic case

3.1.2.1. The partial realization lemma

In this subsection we will treat the main technical lemma for partial realization. We will start with a simple result, that forms the key of the proof of the partial realization lemma.

3.1.2.1-1. Lemma. Let A be an r×s matrix, b an r×1 column vector, c an 1×s row vector.

(i) If rk A = a = rk[A b] = $\mathrm{rk}\begin{bmatrix}A\\c\end{bmatrix}$, then there is exactly one scalar d, such that $\mathrm{rk}\begin{bmatrix}A & b\\c & d\end{bmatrix}$ = a.

(ii) If rk A = rk[A b] = a-1 and $\mathrm{rk}\begin{bmatrix}A\\c\end{bmatrix}$ = a then for all scalars d, $\mathrm{rk}\begin{bmatrix}A & b\\c & d\end{bmatrix}$ = a.

The proof of this lemma is a simple exercise in linear algebra and is left to the reader. The lemma is called 'Main lemma' in [Kal 79] and 'Extension lemma' in [Mue] which also gives a short proof. In fact the result occurs implicity or explicitly in many papers on realization theory. The idea in the lemma can be applied to block-Hankel-matrices. One has

3.1.2.1-2. Partial realization lemma. [Kalman]

Let $\{H_i\}_{i=o}^{k+j}$ be a finite sequence of m×m' matrices.

(a) If

$$(3.1.2.1\text{-}3)\quad \operatorname{rk}\begin{bmatrix} H_1 & H_2 & \cdots & H_j \\ H_2 & H_3 & \cdots & H_{j+1} \\ \vdots & \vdots & & \vdots \\ H_k & H_{k+1} & \cdots & H_{k+j-1} \end{bmatrix} = \operatorname{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_j & H_{j+1} \\ H_2 & & \vdots & \vdots \\ \vdots & & \vdots & \vdots \\ H_k & \cdots & H_{k+j-1} & H_{k+j} \end{array}\right] =$$

$$\operatorname{rk}\left[\begin{array}{ccc} H_1 & H_2 & \cdots & H_j \\ H_2 & H_3 & \cdots & H_{j+1} \\ \vdots & \vdots & & \vdots \\ H_k & H_{k+1} & \cdots & H_{k+j-1} \\ \hline H_{k+1} & H_{k+2} & \cdots & H_{k+j} \end{array}\right] = n(\text{say}),$$ then there

is a unique sequence $H = \{H_i\}_{i=o}^{\infty}$, extending the finite sequence $\{H_i\}_{i=o}^{k+j}$, such that the rank of the block Hankel-matrix $\overset{\circ}{H}(H;1)$ is n.

(b1) If

$$\operatorname{rk}\begin{bmatrix} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_k & \cdots & H_{k+j-1} \end{bmatrix} = \operatorname{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_j & H_{j+1} \\ \vdots & & \vdots & \vdots \\ H_k & \cdots & H_{k+j-1} & H_{k+j} \end{array}\right] <$$

$$\operatorname{rk}\left[\begin{array}{ccc} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_k & \cdots & H_{k+j-1} \\ \hline H_{k+1} & \cdots & H_{k+j} \end{array}\right] = n,$$ then the choice of H_{k+j+1},

such that $\operatorname{rk}\begin{bmatrix} H_1 & H_2 & \cdots & H_{j+1} \\ \vdots & \vdots & & \vdots \\ H_{k+1} & H_{k+2} & \cdots & H_{k+j+1} \end{bmatrix} = n$

is not unique. In fact there is at least one component of H_{k+j+1} which can be chosen arbitrarily in $\mathbb{R}$ (or $\mathbb{C}$ if one works over $\mathbb{C}$).

(b2) The same conclusion as in (b1) holds, if

$$\operatorname{rk}\begin{bmatrix} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_k & \cdots & H_{k+j-1} \end{bmatrix} = \operatorname{rk}\left[\begin{array}{ccc} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_k & \cdots & H_{k+j-1} \\ \hline H_{k+1} & \cdots & H_{k+j} \end{array}\right] < \operatorname{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_j & H_{j+1} \\ \vdots & & \vdots & \vdots \\ H_k & \cdots & H_{k+j-1} & H_{k+j} \end{array}\right] = n.$$

Proof. The proof of various parts and special cases of this lemma can be found at various places. The scalar case (m=m'=1) is treated in [KFA], Ch.10, App.B, Thm(B1). For a proof of part (a) in the multivariable case, one can consult [Haz 80], esp. lemma 4.9, pp. 377-378. Cf. also [Kal 79], [Mue], [Bo-W].

For the sake of completeness, we will give a proof of part (b1) here. According to the following rules, choose a value for each of the components of H_{k+j+1}, labeled by h(i,g) (= the component in the i^{th} row and g^{th} column of H_{k+j+1}). Use the lexicographical order:
h(1,1), h(1,2), h(1,3),...,h(1,m'),h(2,1)...,h(m,m').
Having arrived at component h(i,g) there are two possible situations, namely possibilities (i) and (ii) from lemma (3.1.2.1-1), if we let $\left[\begin{smallmatrix} A & b \\ c & d \end{smallmatrix}\right]$ consist of the first jm'+g columns and the first km+i rows of

$$\left[\begin{array}{ccc|c} H_1 & H_2 \cdots & H_j & H_{j+1} \\ \vdots & & \vdots & \vdots \\ H_k & \cdots & H_{k+j-1} & H_{k+j} \\ H_{k+1} & \cdots & H_{k+j} & H_{k+j+1} \end{array}\right].$$

So, d corresponds to h(i,g). In case (i) choose d = h(i,g) as in the lemma, i.e. in the unique rank preserving manner. In case (ii) choose d = h(i,g) arbitrarily. In both cases $rk\left(\frac{A}{c}\right) = rk\left(\begin{smallmatrix} A & b \\ c & d \end{smallmatrix}\right)$. From this, it follows by induction that the rank of the submatrix consisting of the first

km+i rows of $\begin{bmatrix} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_{k+1} & \cdots & H_{k+j} \end{bmatrix}$ is equal to the rank of the submatrix of the

first km+i rows of $\left[\begin{array}{ccc|c} H_1 & \cdots & H_j & H_{j+1} \\ \vdots & & \vdots & \vdots \\ H_{k+i} & \cdots & H_{k+j} & H_{k+j+1} \end{array}\right]$, for i = 1,...,m. In the case i = m,

this means that

$$(n=)rk\begin{bmatrix} H_1 & \cdots & H_j \\ \vdots & & \vdots \\ H_{k+1} & \cdots & H_{k+j} \end{bmatrix} = rk\left[\begin{array}{ccc|c} H_1 & \cdots & H_j & H_{j+1} \\ \vdots & & \vdots & \vdots \\ H_{k+1} & \cdots & H_{k+j} & H_{k+j+1} \end{array}\right],$$

as is required in the lemma. If for each h(i,g), i = 1,...,m, g = 1,...,m', case (i) would occur, then the rank would never increase, which contradicts the situation given in (b1). Therefore case (ii) must occur at least once. So there is at least one component h(i,g) of H_{k+j+1} that can be chosen arbitrarily in $\mathbb{R}$ (or $\mathbb{C}$ if one works over the field $\mathbb{C}$).

Q.E.D.

(3.1.2.1-4) Remark. This lemma can be refined, and one can give a count of the number of free parameters. However, we will not need that here. We refer to [Mue], Main Theorem (2.5) esp. part (c).

3.1.2.2. Unique partial realization of state space models

As stated in the introduction we will be concerned with the question what the minimal length of the finite sequence of Markov matrices is, that is sufficient for unique partial realization, given structural information about the system. In this subsection we consider the case in which the available structural information is an upper-bound of the McMillan degree of the system.

3.1.2.2-1. Theorem. (a) Let Σ be a linear, time-invariant deterministic system with McMillan degree $\leq n$, for some given $n \in \mathbb{N}$. The sequence of Markov matrices $\{H_i\}_{i=o}^{I}$ of Σ, with $I = 2n$, uniquely determines Σ. I.e. there is a unique minimal partial realization of $\{H_i\}_{i=o}^{I}$, and this partial realization is (the sequence $\{H_i\}_{i=o}^{\infty}$ of Markov matrices of) Σ.
(b) $I = 2n$ is the minimal value of I for which this holds.

Proof. If $n = 0$ the theorem is trivial. So let $n \geq 1$.
(a) Compare Corollary 4.3 of [Haz 80c]. If (A,B,C,D) is a minimal realization of Σ, and the McMillan degree is $n_1 \leq n$ then rk R(A,B) = rk Q(A,B) = n_1. It follows that

$$\text{(3.1.2.2-2)}\quad \operatorname{rk}[B \;\; AB \;\; \ldots \;\; A^{n_1-1}B] = n_1$$

and

$$\text{(3.1.2.2-3)}\quad \operatorname{rk}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n_1-1} \end{bmatrix} = n_1.$$

Therefore

$$(3.1.2.2\text{-}4)\quad \mathrm{rk}\begin{bmatrix} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_n & \cdots & H_{2n-1} \end{bmatrix} = \mathrm{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_n & H_{n+1} \\ \vdots & & \vdots & \vdots \\ H_n & \cdots & H_{2n-1} & H_{2n} \end{array}\right] =$$

$$\mathrm{rk}\left[\begin{array}{ccc} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_n & \cdots & H_{2n-1} \\ \hline H_{n+1} & \cdots & H_{2n} \end{array}\right] = n_1.$$

From the partial realization lemma in the previous subsection, it follows that there is a unique minimal partial realization with McMillan degree n_1. It is clear that this must be Σ. This proves (a).

(b) First consider the s.i.s.o. case. Suppose $n_1 = n$ and take $I = 2n-1$. We will show that minimal partial realization is now nonunique. Because $m = m' = 1$, and McMillan degree is n, one has

$$(3.1.2.2\text{-}5)\quad \mathrm{rk}[B \quad AB \quad \cdots \quad A^{n-2}B] = n-1 = \mathrm{rk}\begin{bmatrix} C \\ CA \\ \vdots \\ CA^{n-2} \end{bmatrix}.$$

Therefore

$$(3.1.2.2\text{-}6)\quad n-1 = \mathrm{rk}\begin{bmatrix} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_{n-1} & \cdots & H_{2n-2} \end{bmatrix} = \mathrm{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_n & H_{n+1} \\ \vdots & & \vdots & \vdots \\ H_{n-1} & \cdots & H_{2n-2} & H_{2n-1} \end{array}\right] <$$

$$\mathrm{rk}\left[\begin{array}{ccc} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_{n-1} & \cdots & H_{2n-2} \\ \hline H_n & \cdots & H_{2n-1} \end{array}\right] = n.$$

We can apply part (b1) of the partial realization lemma of the previous subsection; it follows that (the scalar) H_{2n} can be chosen arbitrarily, and

$$(3.1.2.2\text{-}7)\quad n=\mathrm{rk}\begin{bmatrix} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_n & \cdots & H_{2n-1} \end{bmatrix} = \mathrm{rk}\left[\begin{array}{ccc|c} H_1 & \cdots & H_n & H_{n+1} \\ \vdots & & \vdots & \vdots \\ H_n & \cdots & H_{2n-1} & H_{2n} \end{array}\right] =$$

$$= \mathrm{rk}\left[\begin{array}{ccc} H_1 & \cdots & H_n \\ \vdots & & \vdots \\ H_n & \cdots & H_{2n-1} \\ \hline H_{n+1} & \cdots & H_{2n} \end{array}\right],$$

i.e. these matrices have maximal rank. Again, from the partial realization lemma, it now follows that - for each choice of H_{2n} - there exists a partial realization with McMillan degree n. So I = 2n-1 is not sufficient for a unique minimal partial realization in the s.i.s.o. case.
The same results holds trivially for the multivariable case: This can be seen as follows. In the case just described nothing is said about D. So the choice of D is still open. D can even be chosen to be an m × m' matrix, in which case one really has an m × m' multivariable system. This shows that also in the multivariable case I = 2n is the minimal value of I, for which one has a guaranteed unique minimal partial realization producing the true system Σ.

Q.E.D.

3.1.2.3. The rank structure of the Hankel matrix of arma(p,q) models

Suppose the system Σ can be described by an arma model (described in section 2.3.3), with backward mfd pair (A(z),B(z)) (compare section 2.3.4), with degrees

(3.1.2.3-1) deg A(z) = p, deg B(z) = q,

or more generally

(3.1.2.3-2) deg A(z) $\leq$ p, deg B(z) $\leq$ q.

In the next section we will answer the question how long a sequence of Markov matrices of Σ is needed, given p and q, to determine Σ completely. To be able to do that, we have to know more about what it means that Σ can be written as an arma(p,q) model. This will be treated in this subsection.

3.1.2.3-3. Warning. If Σ can be written as an arma(p,q) model, it is not necessary that there is an arma(p,q) model of Σ that is in minimal basis form!

(recall section 2.3.4). E.g. consider the following simple example.

$$(3.1.2.3\text{-}4)\quad A(z) = \begin{bmatrix} z^2+\varepsilon z+1 & 0 & 0 \\ 0 & z+1 & 0 \\ 0 & z^2+\alpha z & 1 \end{bmatrix},\quad B(z) = I_3.$$

The system Σ which has $(A(z),B(z))$ as (backward) mfd pair can apparently be written as an arma(2,0) model. To put the mfd pair in minimal basis form, one premultiples $A(z)$ and $B(z)$ with

$$(3.1.2.3\text{-}5)\quad C(z) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -z & 1 \end{bmatrix}.$$

One obtains

$$(3.1.2.3\text{-}6)\quad C(z)A(z) = \begin{bmatrix} z^2+\varepsilon z+1 & 0 & 0 \\ 0 & z+1 & 0 \\ 0 & (\alpha-1)z & 1 \end{bmatrix},\quad C(z)B(z) = C(z).$$

The row degrees, i.e. the observability indices, of $[C(z)A(z)\ C(z)B(z)]$ are 2,1,1 and so the McMillan degree is 4. Note that deg $C(z)A(z)$ = 2 and deg $C(z)B(z)$ = deg $C(z)$ = 1. So one <u>now</u> has an arma(2,1) model! Instead, to keep an arma(2,0) model, C has to be a constant nonsingular matrix. The rank of the coefficient matrix of z^2 will then remain equal to two, so at least two of the rows will have degree two. Therefore one does <u>not</u> obtain a minimal base form in this way!

3.1.2.3-7. <u>Notation</u>. The symbol arma(p,q) will stand for the set of all systems that can be written in the form of an arma(p,q) model:

(3.1.2.3-8)

$$\begin{aligned} \text{arma}(p,q) := \{\Sigma \mid \exists A(z),B(z),\ & \text{polynomial matrices, deg } A(z) = p, \\ & \text{deg } B(z) = q, \text{ and } A(0) \text{ is nonsingular,} \\ & \text{and } A(z)^{-1}B(z) = H(z), \text{ the backward} \\ & \text{transfer matrix of } \Sigma\} \end{aligned}$$

The symbol arma($\bar{p},\bar{q}$) will stand for the set of all systems that can be written in the form of an arma(p',q') model with $p' \leq p, q' \leq q$:

(3.1.2.3-9) $\operatorname{arma}(\bar{p},\bar{q}) = \bigcup_{\substack{0\leq p'\leq p \\ 0\leq q'\leq q}} \operatorname{arma}(p',q')$.

Here $\bar{p}$ can be considered to be the set $\bar{p} = \{0,1,2,\ldots,p\}$, and $\bar{q}$ to be the set $\bar{q} = \{0,1,2,\ldots,q\}$. For later reference it is useful to note that

(3.1.2.3-10) $\operatorname{arma}(p,q) \supseteq \operatorname{arma}(\bar{p},\bar{q}) \setminus \{\operatorname{arma}(\overline{p-1},\bar{q}) \cup \operatorname{arma}(\bar{p},\overline{q-1})\}$,

where we use the convention that $\operatorname{arma}(\bar{p},\bar{q}) = \emptyset$ if $p < 0$ or $q < 0$ (cf. (3.1.2.3-9)).

3.1.2.3-11. Definition. Let Σ be a linear, time-invariant system with backward transfer matrix $H(z)$. The mapping τ which maps Σ to $\tau\Sigma$, where $\tau\Sigma$ has backward transfer matrix $zH(z)$, is called the backward shift. The mapping σ which maps Σ to $\sigma\Sigma$, where $\sigma\Sigma$ has backward transfer matrix $z^{-1}(H(z) - H(0))$ is called the forward shift.

3.1.2.3-12. Remarks. (i) σ is the same as the shift in [KFA] section 10.11, (11.4), p. 289.

(ii) Note that $\sigma \circ \tau$ = identity, but $\tau \circ \sigma \neq$ id. In fact $\tau \circ \sigma(\Sigma)$ has backward transfer matrix $H(z)-H(0) \neq H(z)$ in general.

(iii) The Hankel-matrix of $\sigma^k\Sigma$ is $\overset{\circ}{H}(H,1+k)$, and of $\tau^k\Sigma$ it is $\overset{\circ}{H}(H,1-k)$, where, as before, we use the convention that $H_i = 0$ if $i < 0$.

3.1.2.3-13. Definition. Let $(A(z),B(z))$ be a backward mfd pair. The mapping $\tilde{\tau}: (A(z),B(z)) \longmapsto (A(z),zB(z))$ is called the backward shift. The mapping $\tilde{\sigma}: (A(z),B(z)) \longmapsto (A(z),z^{-1}(B(z) - A(z)A(0)^{-1}B(0)))$ is called the forward shift.

3.1.2.3-14. Lemma. If Σ can be represented by the mfd pair $(A(z),B(z))$, then $\tilde{\tau}(A(z),B(z))$ represents $\tau\Sigma$ and $\tilde{\sigma}(A(z),B(z))$ represents $\sigma\Sigma$.

Proof. First notice that $\tilde{\sigma}(A(z),B(z))$ is indeed an mfd pair of polynomial matrices. The backward transfer matrix of $\tilde{\tau}(A(z),B(z))$ is $zA(z)^{-1}B(z) = zH(z)$. The backward transfer matrix of $\tilde{\sigma}(A(z),B(z))$ is

(3.1.2.3-15) $A(z)^{-1}z^{-1}(B(z)-A(z)A(0)^{-1}B(0)) = z^{-1}(H(z)-H(0))$

Q.E.D.

It is not difficult to derive the following result:

3.1.2.3-16. Theorem. If $\Sigma \in \text{arma}(\bar{p},\bar{q})$, then
(a) $\tau\Sigma \in \text{arma}(\bar{p},\overline{q+1})$
(b) $\sigma\Sigma \in \text{arma}(\bar{p},\overline{\max(p,q)-1})$
Before proving this, let us state the immediate corollary

3.1.2.3-17. Corollary. If $\Sigma \in \text{arma}(\bar{p},\bar{q})$, then
(a) $\tau^k\Sigma \in \text{arma}(\bar{p},\overline{q+k})$
(b) $\sigma^k\Sigma \in \text{arma}(\bar{p},\overline{q-k})$ if $q-k \geq p-1$ and $\sigma^k\Sigma \in \text{arma}(\bar{p},\overline{p-1})$ if $q-k \leq p-1$

Proof of theorem (3.1.2.3-16).
This follows simply from the previous lemma and the definition of $\tilde{\tau}$ and $\tilde{\sigma}$: if $(A(z),B(z))$ is an mfd pair representing Σ, and $\deg A(z) = p' \leq p$, $\deg B(z) = q' \leq q$, then $\deg zB(z) = q'+1 \leq q+1$, so $\tau\Sigma \in \text{arma}(\bar{p},\overline{q+1})$. If one applies $\tilde{\sigma}$ to $(A(z),B(z))$, one obtains $(A(z),z^{-1}(B(z) - A(z)A(0)^{-1}B(0)))$. Now $\deg z^{-1}(B(z)-A(z)A(0)^{-1}B(0)) \leq \max(\deg B(z), \deg A(z)) - 1 = \max(p',q') - 1 \leq \max(p,q) - 1$. So $\sigma\Sigma \in \text{arma}(\bar{p}, \overline{\max(p,q)-1})$.

Q.E.D.

To be able to derive the main result of this subsection we need the following.

3.1.2.3-18. Lemma. Suppose Σ can be represented by the backward mfd pair $(A(z),B(z))$, with row degrees of $[A(z)\ B(z)]$ equal to $p_1' \leq p_2' \leq \ldots \leq p_n'$ (without loss of generality, by permutation of the rows of $[A(z)\ B(z)]$, one can assume this ordering to hold). Let $\kappa_1 \leq \kappa_2 \leq \ldots \leq \kappa_m$ denote the ordered observability indices of Σ. Then

(3.1.2.3-19) $\kappa_i \leq p_i'$, $i = 1,2,\ldots,m$.

Proof. From section 2.3.4 it is known, that $A(0)$ is nonsingular. If one considers $z(t) := A(0)y(t)$ as the output of the linear dynamical system, the observability indices remain the same (compare 2.3.4-42)). The corresponding backward mfd pair is $(A(z)A(0)^{-1},B(z))$. Therefore, without loss of generality

in this proof, one can assume A(0) = I.
Consider the corresponding forward mfd pair $(\tilde{A}_v(s),\tilde{B}_v(s))$. From section 2.3.4 we know that the row degrees of $[\tilde{A}_v(s)\ \tilde{B}_v(s)]$ are the same as those of $[A(z)\ B(z)]$, i.e. they are $p'_1 \leq p'_2 \leq \ldots \leq p'_m$. Furthermore we know that $\tilde{A}_v(s)$ has the <u>same</u> row degrees. For the following, compare equations (2.3.4-40,41 etc.). We know that

(3.1.2.3-20) $\tilde{A}_v(S)y_t = \tilde{B}_v(S)u_t$

and therefore

(3.1.2.3-21) $\tilde{A}_v(S)y_{t|t_o-1} = 0 \quad \forall t \geq t_o.$

Consider the i^{th} equation of this system of equations. It is of the form

$$e_i^T y_{t+p'_i|t_o-1} + \alpha_{p'_i}^{(i)} y_{t+p'_i-1|t_o-1} + \ldots + \alpha_1^{(i)} y_{t|t_o-1} = 0,$$

where the $\alpha_j^{(i)}$ are m-dimensional row vectors of coefficients. It follows that

(3.1.2.3-22) $\alpha^{(i)} := (\alpha_1^{(i)}\ \alpha_2^{(i)}\ \ldots\ \alpha_{p'_i}^{(i)}\ e_i^T\ 0\ 0 \ldots)$

is an element of the left kernel of the Hankel matrix of the system (see (2.3.3-1)). Therefore the i^{th} row of the $(1+p'_i)$-th block row of the Hankel matrix is linearly dependent on the previous rows. In section 2.3.3 an arma model was constructed for Σ by taking the first n linearly independent rows of the Hankel matrix. It is now clear that the i^{th} row of the p'_i-th block row of the Hankel matrix does <u>not</u> belong to those n rows. From this it follows that

(3.1.2.3-23) $p'_i \geq p_i,$

where p_i as in section 2.3.3. This holds for each i = 1,2,...,m.
Now $\{p_1,p_2,\ldots,p_m\}$ is an <u>unordered</u> sequence consisting of a permutation of the m ordered <u>observability indices</u> $\{\kappa_1,\ldots,\kappa_m\}$ of Σ. It follows easily that, because $p'_1 \leq p'_2 \leq \ldots \leq p'_m$,

(3.1.2.3-24) $p'_i \geq \kappa_i, \quad i = 1,2, \ldots, m$

Q.E.D.

3.1.2.3-25. Corollary. Suppose $\Sigma \in \text{arma}(p,q)$. Then (i) the observability indices $\kappa_1 \leq \ldots \leq \kappa_m$ are smaller than or equal to the maximum of p and q:

(3.1.2.3-26) $\kappa_i \leq \max(p,q), \quad i = 1, \ldots, m$

and (ii) for the McMillan degree n of Σ one has

(3.1.2.3-27) $n = \sum_{i=1}^{m} \kappa_i \leq m.\max(p,q).$

It will be useful to extend the notation with respect to Hankel matrices as follows:

3.1.2.3-28. Notation. Let H be a sequence of matrices $\{H_k\}_{k\in\mathbb{Z}}$. Then by $\overset{\circ}{H}(H,k,\ell,m)$ the following block-Hankel matrix will be denoted:

$$\overset{\circ}{H}(H;k,\ell,m) := \begin{bmatrix} H_k & H_{k+1} & \cdots & H_m \\ H_{k+1} & H_{k+2} & \cdots & H_{m+1} \\ \vdots & \vdots & & \vdots \\ H_\ell & H_{\ell+1} & \cdots & H_{m+\ell-k} \end{bmatrix}.$$

We now arrive at the main results of this subsection.

3.1.2.3-29. Theorem. Let Σ be a linear, time-invariant system with sequence of $m\times m'$ Markov matrices H. Then the following statements are equivalent for each $p,q \in \mathbb{N}_o$. We use the notation $a \vee b := \max(a,b)$, $a \wedge b := \min(a,b)$.

(i) $\Sigma \in \text{arma}(\bar{p},\bar{q})$

(ii) $\forall k \in \mathbb{Z}$: $\operatorname{rk} \overset{\circ}{H}(H;k+1,k+(p\vee(q-k)),k+m(p\vee(q-k))) =$
$= \operatorname{rk} \overset{\circ}{H}(H;k+1) \leq m(p\vee(q-k))$

(iii) $\operatorname{rk} \overset{\circ}{H}(H;q-p+1,q,q+(m-1)p) = \operatorname{rk} \overset{\circ}{H}(H;q-p+1) \leq mp.$

Proof. (i) $\Rightarrow$ (ii). This follows from the previous corollary, combined with corollary (3.1.2.3-17) and remark (3.1.2.3-12)(iii).
(ii) $\Rightarrow$ (iii). This follows trivially by substitution of $k = q - p$.
(iii) $\Rightarrow$ (i). Let $r := q-p$, and suppose (iii) holds. From the partial realization lemma it follows that one can realize the backward transfer matrix

(3.1.2.3-30) $H_r(z) := \sum_{i=o}^{\infty} H_{i+r} z^i$

by an arma($\bar{p},\bar{p}$) model with backward mfd-pair $(A_r(z),B_r(z))$, $A_r(0)$ nonsingular; deg $A_r(z) \leq p$, deg $B_r(z) \leq p$.
We distinguish three cases: (a) $r = 0$, (b) $r < 0$ and (c) $r > 0$.

ad(a). If $r = 0$, then $H(z) = H_r(z)$, $p = q$ and Σ can be realized as an arma($\bar{p},\bar{q}$) = arma($\bar{p},\bar{p}$) model, i.e. there remains nothing to show.

ad(b). If $r < 0$, then $H_r(z) = \sum_{i=o}^{\infty} H_{i+r} z^i$ can be divided by $z^{|r|}$, because

$H_j = 0$ if $j < 0$. The same must hold for $B_r(z) = A_r(z)H_r(z)$. Let

(3.1.2.3-31) $B(z) = z^{-|r|} B_r(z)$

and

(3.1.2.3-32) $A(z) = A_r(z)$.

Then deg $B(z) \leq p - |r| = p+r = q$ and deg $A(z) \leq p$. Clearly $(A(z),B(z))$ forms the mfd-pair of the system with backward transfer matrix

(3.1.2.3-33) $A(z)^{-1}B(z) = z^r H_r(z) = \sum_{i=o}^{\infty} H_i z^i = H(z)$.

Conclusion: Σ can be written in the form of an arma($\bar{p},\bar{q}$) model, i.e. $\Sigma \in$ arma($\bar{p},\bar{q}$).
ad(c). If $r > 0$ then from (3.1.2.3-30) it follows that

(3.1.2.3-34) $H_{j-1}(z) = H_j(z)z + H_{j-1}$ for all $j = 1,\ldots,r$.

Let $A(z) = A_r(z)$ and $B_{j-1}(z) := B_j(z)z + A(z)H_{j-1}$, $j = r,r-1,\ldots,1$ then it follows that

$$H_j(z) = A(z)^{-1}B_j(z),\ j = r,r-1,\ \ldots,\ 1.$$

Furthermore it follows easily, that

$$\deg B_j(z) \leq p + (r-j), \quad \deg A(z) \leq p.$$

Therefore $\deg B_o(z) \leq q$ and so Σ can be represented by the arma$(\bar{p},\bar{q})$ model with mfd pair $(A(z),B_o(z))$, i.e. $\Sigma \in$ arma$(\bar{p},\bar{q})$.

Q.E.D.

3.1.2.3-35. Remark. This theorem characterizes the set of all arma$(\bar{p},\bar{q})$ models precisely in terms of the rank structure of an associated Hankelmatrix, namely $\overset{\circ}{H}(H;q-p+1)$. In fact (iii) states nothing more or less than that the observability indices of $\overset{\circ}{H}(H;q-p+1)$ are smaller than or equal to p. From section 2.3.5, on Kronecker indices and the specialization order, it follows that the closure of the set of all systems with observability indices smaller than or equal to p is included in the set of all systems of McMillan degree $\leq$ mp (and very probably these sets are equal - see the discussion in section 2.3.5). Here we find that the closure $\overline{\text{arma}(\bar{p},\bar{q})}$ of arma$(\bar{p},\bar{q})$ is included in (and probably equal to) the set of all systems with sequences of Markov parameters H, where H is such that

(3.1.2.3-36)(i) $H_i = 0$ for all $i < 0$

(ii) $\text{rk}\overset{\circ}{H}(H,q-p+1) \leq mp$

[In a system identification procedure one has to beware of the possibility that one comes close to the topological boundary of arma$(\bar{p},\bar{q})$. To be able to handle such situation, one has to include the boundaries. In general this may be difficult. But if one restricts oneself to the case $\text{rk}\overset{\circ}{H}(H;q-p+1) = n$ for some $n \leq mp$, then this appears to be possible. However, this requires further research. (In fact the topology forces us, even if one starts with mimo arma(p,q), to use a state space model, namely tc realize $\overset{\circ}{H}(H;q-p+1)$. The only exceptions appear to be m = 1 (siso model), q = 0 (ar-model) or p = 0 (ma-model).)]

3.1.2.3-37. Corollary. Let Σ be as in the theorem, and suppose $\Sigma \in$ arma$(\bar{p},\bar{q})$. Then Σ is uniquely determined by its Markov matrices $H_o,H_1,\ldots,H_I$, where

(3.1.2.3-38) $I = mp + q$

Proof. As before, we use the notation $a \vee b = \max(a,b)$. The previous theorem, together with the partial realization lemma, implies that the matrices

$$\overset{\circ}{H}(H;k+1,k+(p\vee(q-k)) + 1,\ k + m(p\vee(q-k))) \text{ and}$$
$$\overset{\circ}{H}(H;k+1,k+(p\vee(q-k)),\ k + m(p\vee(q-k)) + 1)$$

determine the Hankel matrix $\overset{\circ}{H}(H;k+1)$ uniquely. To form these two Hankel-matrices, one needs $H_{k+1},H_{k+2},\ldots,H_{I(k)}$, with $I(k) = k + (m+1)\,(p\vee(q-k))$. This takes its minimum value at $k = q-p$. In that case one has $I(q-p) = (q-p) + (m+1)p = mp + q$. It follows that $I = mp+q$ suffices to determine all of H. Q.E.D.

3.1.2.3-39. Remark. Note that the number m' of input components does not affect $I = mp + q$.

There is a case in which one can improve on $I = mp + q$. This will be especially important for section 3.1.3. The case we have in mind is the one in which the system is invertible, while $p > q$. If an arma(p,q) model with backward mfd-pair $(A(z),B(z))$ is invertible, then $m = m'$ and $B(0)$ is nonsingular, and the inverse system is given by the backward mfd-pair $(B(z),A(z))$, and so it is arma(q,p). So p and q change roles. Now consider the following:

3.1.2.3-40. Lemma. Suppose the linear time-invariant system Σ is invertible, then for each $k \in \mathbb{N}$ the first k Markov matrices of Σ are uniquely determined by the first k Markov matrices of the inverse system and vice versa.

Proof. Let $H(z) = \sum_{i=0}^{\infty} H_i z^i$ be the backward transfer matrix of Σ. Because Σ is invertible, $m = m'$ and H_0 is nonsingular. The inverse system has $H(z)^{-1}$ as its backward transfer matrix. Write

$$(3.1.2.3\text{-}41)\quad H(z)^{-1} = \sum_{i=0}^{\infty} H'_i z^i.$$

Then

$$(3.1.2.3\text{-}42)\quad I_m = \Big(\sum_{i=o}^{\infty} H_i z^i\Big)\Big(\sum_{i=o}^{\infty} H'_i z^i\Big) = \sum_{j=o}^{\infty}\Big(\sum_{i=o}^{j} H_i H'_{j-i}\Big)z^j.$$

It follows that

$$(3.1.2.3\text{-}43)\quad H_o = (H'_o)^{-1}$$

and

$$\forall j \geq 1\colon\ H_j = -\Big(\sum_{i=o}^{j-1} H_i H'_{j-i}\Big)(H'_o)^{-1}$$

from which it follows easily that the first k(=j+1) Markov matrices of Σ are uniquely determined by the first k Markov matrices of the inverse system, for each $k \in \mathbb{N}$.

Q.E.D.

Our first main result on partial realization of deterministic arma($\bar{p},\bar{q}$) models is the following:

3.1.2.3-44. Theorem. Let Σ be a linear, time-invariant system, and suppose Σ is invertible; then if $\Sigma \in \text{arma}(\bar{p},\bar{q})$, Σ is uniquely determined by its first $J = I + 1$ Markov matrices $H_o, H_1, \ldots, H_I$, where

(3.1.2.3-45) $I = m.\min(p,q) + \max(p,q)$, so $J = m.\min(p,q) + \max(p,q) + 1$

3.1.2.3-46. Remark. Equivalent formulas for (3.1.2.3-45) are:

(3.1.2.3-47) $J = p + q + (m-1)\min(p,q) + 1$,

(3.1.2.3-48) $J = \min(mp+q, mq+p) + 1$.

Proof. If $p \leq q$, then corollary (3.1.2.3-31) can be applied. If $p > q$, then corollary (3.1.2.3-31) can be applied to the inverse system: the first mq+p+1 Markov matrices suffice to determine all the Markov matrices. Lemma (3.1.2.3-34) then shows that the first mq+p+1 Markov matrices of the system Σ must determine all the Markov matrices of Σ.

Q.E.D.

There is an alternative proof for this theorem, for the case $p > q$, which doesn't make use of the inverse system. It is this proof that can be adapted to the stochastic case. Therefore we will present it here:

Second proof of theorem (3.1.2.3-44).

Because Σ is invertible H_o is nonsingular. Therefore

$$\overset{\circ}{H}(H;q-p+1,0,0) = \begin{bmatrix} 0 & \cdots & 0 & H_o \\ \vdots & \iddots & & \vdots \\ 0 & \iddots & & \vdots \\ H_o & \cdots & & H_{p-q-1} \end{bmatrix}$$

is clearly nonsingular (it has determinant equal to $\pm (\det H_o)^{p-q} \neq 0$). This implies that the reachability indices of $\overset{\circ}{H}(H;q-p+1)$ are all larger than or equal to $p-q$. The sum of the reachability indices is smaller than or equal to mp. Therefore their maximum is smaller than or equal to $mp-(m-1)(p-q)$. It follows that in this case, theorem (3.1.2.3-29) (iii) can be replaced by

$$\text{(iii)'} \quad \text{rk } \overset{\circ}{H}(H;q-p+1,q,mq) = \text{rk } \overset{\circ}{H}(H;q-p+1) \leq mp.$$

Following similar arguments as in the proof of corollary (3.1.2.3-37), it follows that

$$\overset{\circ}{H}(H;q-p+1,q+1,mq) \text{ and } \overset{\circ}{H}(H;q-p+1,q,mq+1)$$

together determine the Hankelmatrix $\overset{\circ}{H}(H;q-p+1)$. To form these two Hankel matrices one needs $H_o, H_1, \ldots, H_I$, with

$$I = (q+1) + mq - (q-p+1) = mq + p.$$

Q.E.D.

In the remaining part of this subsection we want to show that the number J of Markov matrices of the (invertible) system from which all of them can be determined, found in theorem (3.1.2.3-44) is minimal. Because of lemma (3.1.2.3-40) it suffices to show minimality in the case $p \leq q$. In the pure m.a. case, i.e. $p = o$, it is clear that $J = q+1$ is minimal, because all the

nonzero Markov matrices, $H_o, H_1, \ldots, H_q$ can be chosen arbitrarily. For the remaining cases we make use of the following existence result.

3.1.2.3-50. Lemma. Let $p \geq 1$, $m \geq 1$. There exists an asymptotically stable linear time-invariant system $\tilde{\Sigma}$ which has m observability indices $p,p,\ldots,p$ and m reachability indices $mp,0,\ldots,0$.

Proof. Consider $\tilde{\Sigma}$ with backward transfer matrix

$$\tilde{H}(z) = \begin{bmatrix} (z-a_1)^{-p} & 0 & \cdots & 0 \\ (z-a_2)^{-p} & 1 & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & 1 \\ (z-a_m)^{-p} & 0 & \cdots & 0 \end{bmatrix},$$

where $a_1,\ldots,a_m$ are different real numbers, and $|a_i| > 1$ for all $i = 1,\ldots,m$. Then

(3.1.2.3-51) $$\tilde{H}(z) = A_1(z)^{-1}B_1(z) = B_2(z)^T A_2(z)^{-T}$$

with

(3.1.2.3-52) $$A_1(z) = \mathrm{diag}[(z-a_1)^p, (z-a_2)^p, \ldots, (z-a_m)^p],$$

$$B_1(z) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 1 & (z-a_2)^p & \ddots & \vdots \\ \vdots & 0 & \ddots & 0 \\ 1 & 0 & 0 & (z-a_m)^p \end{bmatrix},$$

$$A_2(z) = \mathrm{diag}\Big[\prod_{i=1}^{m}(z-a_i)^p,\ 1,\ 1,\ldots,1\Big], \text{ and}$$

$$B_2(z) = \begin{bmatrix} \prod_{\substack{j=1 \\ j\neq 1}}^{m}(z-a_j)^p & \prod_{\substack{j=1 \\ j\neq 2}}^{m}(z-a_j)^p & \cdots & \prod_{\substack{j=1 \\ j\neq m}}^{m}(z-a_j)^p \\ 0 & 1 & \cdots & 0 \\ \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Now $(A_1(z),B_1(z))$ is clearly an mfd-pair in minimal base form with row degrees all equal to p. So the observability indices of Σ are all equal to p. At the other hand $(A_2(z),B_2(z))$ is an mfd-pair in minimal base form with row degrees $mp,0,0,\ldots,0$. Therefore Σ has reachability indices $mp,0,\ldots,0$. (The Hankelmatrix of $\tilde{H}^T(z) = A_2(z)^{-1}B_2(z)$ is the transpose of the Hankelmatrix of $\tilde{H}(z)$. Transposition of the Hankelmatrix makes observability indices into reachability indices, and vice versa). Q.E.D.

For such a system $\tilde{\Sigma}$, with backward transfer function $\tilde{H}(z) = \sum_{i=o}^{\infty} \tilde{H}_i z^i$, the following rank inequality holds:

(i) if $p = m = 1$, $rk(\tilde{H}_1) = pm = 1 > 0$.
(ii) if $pm \geq 2$ then

$$rk \begin{bmatrix} \tilde{H}_1 & \cdots & \tilde{H}_{pm-1} \\ \vdots & & \vdots \\ \tilde{H}_p & \cdots & \tilde{H}_{p(m+1)-2} \end{bmatrix} < rk \begin{bmatrix} \tilde{H}_1 & \cdots & \tilde{H}_{pm} \\ \vdots & & \vdots \\ \tilde{H}_p & \cdots & \tilde{H}_{p(m+1)-1} \end{bmatrix} = pm.$$

It follows from the partial realization lemma that $\tilde{\Sigma}$ is not completely determined by $\tilde{H}_o,\ldots,\tilde{H}_{p(m+1)-1}$, and that there is at least one component of $\tilde{H}_{p(m+1)}$ that can be chosen freely: for each such choice there exists a partial realization with McMillan degree mp, and with observability indices $p,p,\ldots,p$, and reachability indices $mp,0,\ldots,0$. This in fact shows the minimality of J in theorem (3.1.2.3-44) in the case $p = q \geq 1$.
Now let us treat the case $q > p \geq 1$. Define $H_{q-p+j} := \tilde{H}_j$ for $j = 1,2,3,\ldots$, and choose $H_o,\ldots,H_{q-p}$ arbitrarily. Then the resulting backward transfer matrix $H(z) = \sum_{i=o}^{\infty} H_i z^i$ represents an arma$(\bar{p},\bar{q})$ model (according to theorem (3.1.2.3-29)) which has the property that (compare (3.1.2.3-29) (iii))

(i) if $p = m = 1$, $rk\, \overset{\circ}{H}(H;q,q,q) = rk\, H_q = rk\, \tilde{H}_1 = 1 > 0$.

(ii) if $pm \geq 2$,

$$rk\, \overset{\circ}{H}(H;q-p+1,q,q+(m-1)p-1) < rk\, \overset{\circ}{H}(H;q-p+1,q,q+(m-1)p) = mp.$$

In the same way as above, one can apply the partial realization lemma to conclude that $H_o,\ldots,H_{q+mp-1}$ do not determine Σ completely, i.e. $I = q+mp$ (and so $J = q+mp+1$) is minimal.
We have shown

3.1.2.3-53. Theorem. Under the condition of theorem (3.1.2.3-44), the first I Markov matrices $H_o,\ldots,H_{I-1}$ do not, in general, determine Σ uniquely. So the number $J = I + 1$ found in theorem (3.1.2.3-44) is minimal.

3.1.2.3-54. Remark. Of course in many cases less than J Markov matrices will determine the system uniquely. However, if J has to be sufficiently large for all possible cases it has to be taken as in theorem (3.1.2.3-44).

3.1.3. The Stochastic case

3.1.3.1. Partial realization of state space models from the covariances

In section 2.4.3 it was shown that the block matrix of covariance matrices has properties similar to the block Hankel matrix of Markov matrices of the stochastic system. This will now be used to obtain results about partial realization of a stochastic system from the first $J = I + 1$ of its covariance matrices $\Gamma_o,\Gamma_1,\ldots,\Gamma_I$. In section 3.2 these results will be applied to the socalled 'finite identifiability problem'.
In many cases the results here will be translations of the results for the deterministic case, treated in section 3.1.2. Therefore, often we will refer for (part of the) proofs to the proofs given in section 3.1.2.

3.1.3.1-1. Theorem. (a) Let Σ be a linear, stationary Gaussian system of the form (2.4.1-1), with McMillan degree $\leq n$, for some known $n \in \mathbb{N}$. The sequence of covariance matrices $\Gamma_0, \Gamma_1, \ldots, \Gamma_I$ of Σ, with $I = 2n$, uniquely determines Σ. (b) $I = 2n$ is the minimal value of I for which this holds.

Proof: (a) Use corollaries (2.4.3-25) and (2.4.3-28), and apply the same proof as for theorem (3.1.2.2-1)(a).
(b) Compare the proof of theorem (3.1.2.2-1) (b) The case $n = 0$ is here again trivial, so let $n \geq 1$. Let $\{H_i\}_{i=1}^{2n-1}$ denote the Markov matrices as constructed in the mentioned proof (with $m = m'$) and let $H_0 = D = I_m$ be the identity matrix. We know that there are several different partial realizations, say $\{\tilde{H}_i\}_{i=0}^{\infty}$ and $\{\bar{H}_i\}_{i=0}^{\infty}$ with

(3.1.3.1-2) $\tilde{H}_i = \bar{H}_i = H_i$ for $i = 0,1,2,\ldots, 2n-1$.

From corollary (2.4.2-20) it follows that there exists an $\varepsilon > 0$ such that

$$(3.1.3.1\text{-}3)\quad \tilde{\Gamma}_\varepsilon(s) := \tilde{H}_0 + \sum_{k=1}^{\infty} (\tilde{H}_k \varepsilon^k) s^{-k} + \sum_{k=1}^{\infty} (\tilde{H}_k^T \varepsilon^k) s^k,\ s \in C(0,1)$$

and

$$\bar{\Gamma}_\varepsilon(s) := \bar{H}_0 + \sum_{k=1}^{\infty} (\bar{H}_k \varepsilon^k) s^{-k} + \sum_{k=1}^{\infty} (\bar{H}_k^T \varepsilon^k) s^k, s \in C(0,1)$$

are both spectral density matrices. They are different, both have McMillan degree n, and $\tilde{\Gamma}_i = \bar{\Gamma}_i$, $i = 0,1,\ldots,2n-1$. So $I = 2n-1$ is not sufficient for unique determination of the system from the first $I+1$ covariance matrices.

Q.E.D.

3.1.3.2. Partial realization in the stochastic arma $(\bar{p},\bar{q})$ case

Combining corollaries (2.4.3-28) and (2.4.3-25) with theorem (3.1.2.3-29), one obtains a completely analogous theorem for the stochastic case:

(3.1.3.2-1). Theorem. Let Σ be a linear, stationary Gaussian system of the form (2.4.1-1), with sequence of $m \times m$ covariance matrices $\Gamma = \{\Gamma_k\}_{k \in \mathbb{Z}}$. Then the following statements are equivalent for each $p,q \in \mathbb{N}_0$:

(i) $\Sigma \in \text{arma}(\bar{p},\bar{q})$

(ii) $\forall k \in \mathbb{Z}$: $\text{rk } \overset{\circ}{H}(\Gamma;k+1,k+(p\vee(q-k)),k+m(p\vee(q-k))) =$
$\text{rk } \overset{\circ}{H}(\Gamma;k+1) \leq m\ (p\vee(q-k))$

(iii) $\text{rk } \overset{\circ}{H}(\Gamma;q-p+1,q,q+(m-1)p) = \text{rk } \overset{\circ}{H}(\Gamma;q-p+1) \leq mp.$

3.1.3.2-2. Remark. By some abuse of notation, (because the notation is also used for the deterministic case) we will denote by $\text{arma}(\bar{p},\bar{q})$ the set of all stochastic systems which can be represented by a ('deterministic') $\text{arma}(\bar{p},\bar{q})$ model with stochastic white noise inputs.

3.1.3.2-3. Corollary. Let Σ be as in the theorem, and suppose $\Sigma \in \text{arma}(\bar{p},\bar{q})$. Then Σ is uniquely determined by its covariance matrices $\Gamma_0,\Gamma_1,\ldots,\Gamma_I$, where

(3.1.3.2-4) $I = mp + q.$

Proof. See corollary (3.1.2.3-37). Q.E.D.

Now recall that a stochastic linear system has an innovations representation (section 2.4.4). Such an innovations representation has an innovations-to-outputs (backward) transfer matrix $H(z)$ (say), with the property that $H(0)$ is nonsingular. I.e. the corresponding deterministic system with this transfer matrix is invertible. Therefore it is perhaps not too surprising that in the stochastic case we obtain a similar result as in the invertible deterministic case, namely:

3.1.3.2-5. Theorem. Let Σ be a linear stationary Gaussian system of the form (2.4.1-1) with sequence of covariance matrices $\Gamma = \{\Gamma_k\}_{k\in\mathbb{Z}}$. Then if $\Sigma \in \text{arma}(\bar{p},\bar{q})$, Σ is uniquely determined by its first $J = I + 1$ covariance matrices $\Gamma_0,\Gamma_1,\ldots,\Gamma_I$, where

(3.1.3.2-6) $I = p+q+(m-1)(p\wedge q)$ so $J = p+q+1+(m-1)(p\wedge q).$

3.1.3.2-7. Remark. Compare theorem (3.1.2.3-44).

Proof. If $p \leq q$, then corollary (3.1.3.2-3) applies. If $p > q$, the second

proof of theorem (3.1.2.3-44) can be applied with one exception: a seperate proof has to be given of the fact that $\overset{\circ}{H}(\Gamma;q-p+1,0,0)$ is nonsingular. Post multiply $\overset{\circ}{H}(\Gamma,q-p+1,0,0)$ with the block matrix $\begin{bmatrix} 0 & \cdots & 0 & I \\ \vdots & & \cdot & 0 \\ 0 & I & & \vdots \\ I & 0 & \cdots & 0 \end{bmatrix}$,

where $I = I_m$ is the m×m identity matrix. One obtains the symmetric Toeplitz matrix

$$\overset{\circ}{T}(\Gamma;0,q+1-p) = \begin{bmatrix} \Gamma_o & \Gamma_1 & \cdots & \Gamma_{|q+1-p|} \\ \Gamma_{-1} & \Gamma_o & & \vdots \\ \vdots & & \ddots & \Gamma_1 \\ \Gamma_{q+1-p} & \cdots & \Gamma_{-1} & \Gamma_o \end{bmatrix}.$$

Let $\overset{\circ}{H}(z) = \sum_{i=o}^{\infty} H_i z^i$ be the innovations-to-outputs (backward) transfer matrix of the system. Then (see e.g. theorem (2.4.2-7) and its proof)

$$(3.1.3.2\text{-}8)\quad \overset{\circ}{T}(\Gamma;0,q+1-p) = \sum_{\ell\in\mathbb{Z}} \begin{bmatrix} H_{\ell+|q+1-p|} \\ H_{\ell+|q+1-p|-1} \\ \vdots \\ H_\ell \end{bmatrix} \begin{bmatrix} H_{\ell+|q+1-p|} \\ H_{\ell+|q+1-p|-1} \\ \vdots \\ H_\ell \end{bmatrix}^T.$$

This is clearly a positive semi-definite Hermitean matrix.
If $x \in \mathbb{R}^{(|q+1-p|+1)m}$ is in the kernel of this matrix then

$$(3.1.3.2\text{-}9)\quad x^T \begin{bmatrix} H_{\ell+|q+p-1|} \\ H_{\ell+|q+p-1|-1} \\ \vdots \\ H_\ell \end{bmatrix} = 0 \text{ for } \underline{\text{all}}\ \ell \in \mathbb{Z}.$$

This can also be written as

$$(3.1.3.2\text{-}10)\quad x^T \begin{bmatrix} H_o & H_1 & H_2 & \cdots & H_{|q+1-p|} & H_{|q+1-p|+1} & \cdots \\ 0 & H_o & & & & \vdots & \\ \vdots & 0 & \ddots & & & \vdots & \\ 0 & 0 & \cdots & 0 & H_o & H_1 & \cdots \end{bmatrix} = 0.$$

Because H_o is nonsingular it follows that x = 0. This proves the nonsingularity of $\overset{\circ}{T}(\Gamma;0,q-p+1)$ and therefore the nonsingularity of $\overset{\circ}{H}(\Gamma;q-p+1,0,0)$. The rest of the proof is completely the same as the second proof of theorem (3.1.2.3-44). Q.E.D.

What remains is the question of minimality. In contradiction to the deterministic case, this problem is not (yet) completely solved. In all cases in which $q \geq p$ minimality can be shown, but in case $q < p$ the question of minimality is still open.

For minimality in the case p = 0, we refer to Theorems 2.2.4 and 3.3.3 in [Tig]. For the case $q \geq p \geq 1$, our proof of minimality in the deterministic case can be adapted. See lemma (3.1.2.3-50) and theorem (3.1.2.3-53); compare the proof of theorem (3.1.3.1-1).
Theorem (3.1.2.3-53) implies that there are (at least) two different sequences of Markov matrices, $\{\tilde{H}_i\}_{i=o}^{\infty}$ and $\{\bar{H}_i\}_{i=o}^{\infty}$, say, corresponding to arma$(\bar{p},\bar{q})$ models (with $q \geq p \geq 1$ fixed), such that

(3.1.3.2-11) $\tilde{H}_i = \bar{H}_i$, i = 0,1,2,...,I-1,

and $\tilde{H}_o = \bar{H}_o$ is positive definite symmetric, with I = q+mp. From the corollary (2.4.2-20) it follows that there exists an $\varepsilon > 0$ such that

$$(3.1.3.2\text{-}12)\quad \tilde{\Gamma}_\varepsilon(s) := \tilde{H}_o + \sum_{k=1}^{\infty} (\tilde{H}_k \varepsilon^k) s^{-k} + \sum_{k=1}^{\infty} (\tilde{H}_k^T \varepsilon^k) s^k,\quad s \in C(0,1),$$

and

$$\bar{\Gamma}_\varepsilon(s) := \bar{H}_o + \sum_{k=1}^{\infty} (\bar{H}_k \varepsilon^k) s^{-k} + \sum_{k=1}^{\infty} (\bar{H}_k^T \varepsilon^k) s^k,\quad s \in C(0,1),$$

are both spectral density matrices.
They are different, both are the spectral density matrices of an arma$(\bar{p},\bar{q})$ model (theorem 3.1.3.2-1), while (in an obvious notation)

$$\tilde{\Gamma}_i = \bar{\Gamma}_i,\quad i = 0,1,2,\ldots,I-1.$$

This shows that I = q+mp is minimal indeed if $q \geq p \geq 1$.
We have shown

3.1.3.2-13. Theorem. Assume the same conditions as in theorem (3.1.3.2-5). If $q \geq p$, then the number $I = q+mp$ $(= p+q+(m-1)(p \wedge q))$ is the minimal number such that the sequence of covariance matrices $\Gamma_0, \Gamma_1, \ldots, \Gamma_I$, uniquely determines the system Σ.

3.2. The finite identifiability problem

Entweder ein Ding hat Eigenschaften, die kein anderes hat, dann kann man es ohne weiteres durch eine Beschreibung aus den anderen herausheben, und darauf hinweisen; oder aber, es gibt mehrere Dinge, die ihre sämtlichen Eigenschaften gemeinsam haben, dann ist es überhaupt unmöglich auf eines von ihnen zu zeigen. Denn ist das Ding durch nichts hervorgehoben, so kann ich es nicht hervorheben, denn sonst ist es eben hervorgehoben.

(L. Wittgenstein, Tractatus Logico-Philosophicus, 2.02331).

3.2.1. On the methodology of the identifiability problem

To explain the precise meaning of finite identifiability it is necessary to be precise with the meaning of identifiability, the role of the observations and the role of parameters. At the same time we will use the opportunity to formulate some definitions that will be important in later chapters. This concerns especially the concept of overlapping parametrizations.

3.2.1.1. A set-theoretic foundation of system identifiability

We take it that system identification (and much more general scientific explanation in the empirical sciences) is basically concerned with the relationships between observable variables. Such a relationship is described by way of a model. Consider the set of all observable variables to which the model is supposed to apply. In many cases, this set will consist of an infinite number of observable variables. E.g. in the case of a stochastic process as described in chapter 2, the set of all variables to which the model is supposed to apply is $\{y_t | t \in \mathbb{Z}\}$. Let the space of joint outcomes of all these observable variables be denoted by D_∞. The subindex ∞ stresses the fact that in most of the cases that we will encounter, D_∞ is an infinite Cartesian product. E.g. in the case of a stochastic process $\{y_t\}_{t \in \mathbb{Z}}$, where $y_t \in \mathbb{R}^m$,

one has

(3.2.1.1-1) $D_\infty = (\mathbb{R}^m)^{\mathbb{Z}}$.

If we want to stress the dependence of D_∞ on the model M, we will denote the set by $D_\infty(M)$.
The procedure in system identification is that a set of possible models $\mathring{M}$ is proposed, on the basis of experience with the phenomenon, theoretical considerations and/or perhaps some speculation. Such a set $\mathring{M}$ is called the model set. All the models in $\mathring{M}$ are supposed to apply to the same set of observable variables and so (without loss of generality) for all $M,M' \in \mathring{M}$: $D_\infty(M) = D_\infty(M')$. It is the task of system identification to identify a true model from the set $\mathring{M}$, if there is one, using measurements of some of the observable variables. It is clear that in the vast majority of the cases it is not possible to measure all the observable variables that are involved, and certainly not, of course, if there are infinitely may of them. It is of course true that in practice the true 'model' does not lie in $\mathring{M}$, and instead one tries to find a model from $\mathring{M}$ that best approximates the truth in some sense. However, we will abstract from that here.
A model M from $\mathring{M}$ is considered to be basically a description of a relationship between the observable variables involved. Therefore it is reasonable to require such a model to describe a relationship between the variables in such a way that one is able to tell whether an element $d_\infty \in D_\infty$, i.e. an ideal observation record of all possible measurements on all observable variables involved, is in accordance with the model, or otherwise falsifies the model. Any model which has this property will be called a (D_∞-) testable model:

3.2.1.1-2. Definition. A (D_∞-) testable model M is a model for which there exists a mapping

$$T_M: D_\infty \to \{\text{True},\text{False}\},$$

which tells whether an element of D_∞ is in accordance with the model (T_M=true) or it is falsifying the model (T_M=false).
We will only consider (D_∞-) testable models. (In the next subsection it will be explained how stochastic models can be made to fit in!).

3.2.1.1-3. Remark. Note that any mapping T: $D_\infty \to$ {True, False} is a testable model by definition.

Let us now introduce the following notation: Let $\mathring{M}$ be a set of (D_∞-testable) models. For each $M \in \mathring{M}$, we define

(3.2.1.1-4) $S(M) := \{d_\infty \in D_\infty | T_M(d_\infty) = \text{true}\} \subseteq D_\infty$.

So S(M) is the set of all ideal observations $d_\infty \in D_\infty$ that are in accordance, i.e. do not falsify, M.

3.2.1.1-5. Definition. A (D_∞-testable) model is called D_∞-falsifiable if $S(M) \neq D_\infty$, i.e. S(M) is a strict subset of D_∞.

3.2.1.1-6. Definition. A (D_∞-testable) model will be called trivial if it is not D_∞-falsifiable, i.e. if $S(M) = D_\infty$

3.2.1.1-7. Definition. A (D_∞-testable) model will be called empty if it is always falsified, i.e. if $S(M) = \emptyset$.

The following concept is very fundamental. If there is no conceivable ideal observation $d_\infty \in D_\infty$ which can distinguish between two models M and M', in the sense that one model is falsified by d_∞, while the other is not, then we will call the models observationally equivalent:

3.2.1.1-8. Definition. Two (D_∞-testable) models M_1 and M_2 are called observationally equivalent if

(3.2.1.1-9) $S(M_1) = S(M_2)$

It is not difficult to see that this induces an equivalence relation on $\mathring{M}$, which will be denoted by: $M_1 \sim M_2$. The set of equivalence classes in $\mathring{M}$ is denoted by $\mathring{M}/\sim$. Note that this is itself a set of D_∞- testable models.

3.2.1.1-10. Remark. Of course, if $M_1 \sim M_2$, M_1 may be far more complex than the other and M_2 may therefore be easier to work with. This usually leads to choosing the simpler model. However, there is no purely logical reason to do

this. A solution to this is to choose $\overset{\circ}{M}/\sim$ as the model set. Then it is allowed, of course, to choose any model from an equivalence class as the representing element of that class.

With (D_∞-) model (or system) identifiability we will mean that there is no ideal observation $d_\infty \in D_\infty$ that is compatible with more than one model from $\overset{\circ}{M}$. Formally

3.2.1.1-11. Definition. A (D_∞-testable) model set $\overset{\circ}{M}$ is called (D_∞-) model (or system) identifiable if

(3.2.1.1-12) $\forall M_1, M_2 \in \overset{\circ}{M},\ M_1 \neq M_2 : S(M_1) \cap S(M_2) = \emptyset$

3.2.1.1-13. Remark. (a) If $\overset{\circ}{M}$ contains observationally equivalent models, $\overset{\circ}{M}$ is not model identifiable.
(b) If a model set $\overset{\circ}{M}/\sim$ is not model identifiable one can consider the 'finer' model set

$$(3.2.1.1\text{-}14)\quad \overset{\circ}{M}_f := \left\{ \left(\bigcap_{\substack{d_\infty \in S(M) \\ M \in \overset{\circ}{M}}} S(M) \right) \cap \left(\bigcap_{\substack{d_\infty \in S(M)^c \\ M \in \overset{\circ}{M}}} S(M)^c \right) \;\middle|\; d_\infty \in \bigcup_{M \in \overset{\circ}{M}} S(M) \right\}$$

which is model identifiable.
Therefore, by considering $\overset{\circ}{M}/\sim$ instead of $\overset{\circ}{M}$, and if necessary $\overset{\circ}{M}_f$, one can obtain formal model identifiability.

From now on except if the opposite is explicitly stated, we will assume that $\overset{\circ}{M}$ is model identifiable and so, a fortiori, $\overset{\circ}{M}/\sim = \overset{\circ}{M}$. For simplicity of notation we will then equate M with $S = S(M)$ and write $S \subseteq D_\infty$, $S \in \overset{\circ}{M}$.
One can conclude that for testable models there is no essential identifiability problem w.r.t. the ideal observations $d_\infty \in D_\infty$, because if there is nonidentifiability, it can be solved by redefining the model space. Especially, in most cases one can use $\overset{\circ}{M}/\sim$. Because the models are only meant to describe relationships in D_∞, nothing essential is lost in going from $\overset{\circ}{M}$ to $\overset{\circ}{M}/\sim$. Going from $\overset{\circ}{M}/\sim$ to $\overset{\circ}{M}_f$, if at all necessary, is a harmless refinement of the model space. Therefore, in this set-up, an essential model identifiability problem can only occur due to incompleteness of the observations. We will clarify that presently.
Let $\Psi: D_\infty \to D$ be a generally noninjective mapping from D_∞ into some set D. Ψ

will be called the experiment mapping. It will represent all the observations that may become available in a given 'experiment', or their limiting behaviour.

3.2.1.1-15. Examples.

(i) Suppose the observables modeled by $M \in \overset{\circ}{M}$ are $\{y_t | t \in \mathbb{Z}\}$, with $y_t \in \mathbb{R}^m$. Then, as before, $D_\infty = (\mathbb{R}^m)^{\mathbb{Z}}$.
Now suppose that the observations that will become available are only taken on even times: $t \in 2\mathbb{Z}$.
Then $\Psi : D_\infty \to D = (\mathbb{R}^m)^{2\mathbb{Z}}$

$$\{y_t\}_{t\in\mathbb{Z}} \longmapsto \{y_t\}_{t\in 2\mathbb{Z}}.$$

Note that Ψ is a noninjective mapping.

(ii) Let $\{y_t\}_{t\in\mathbb{Z}}$ and D_∞ be as in (i).
Now suppose that the only 'observation' that will become 'available' is the infinite-sample mean if it exists:

$$\lim_{\substack{T\to\mathbb{Z} \\ 0<|T|<\infty}} \frac{\sum_{t\in T} Y_t}{|T|}$$

Of course this observation will never become available in practice. However, it is a valid question to ask whether such a piece of information is enough to identify a model in a given model set.

Let the D_∞-model set $\overset{\circ}{M}$ be given. Ψ induces a D-model set $\overset{\circ}{M}_\Psi$ as follows:

(3.2.1.1-16) $$\overset{\circ}{M}_\Psi = \{\Psi(S) | S \in \overset{\circ}{M}, S \subseteq D_\infty\}.$$

$\overset{\circ}{M}_\Psi$ will be called the Ψ-induced model set derived from $\overset{\circ}{M}$.

3.2.1.1-17. Proposition. Each model in $\overset{\circ}{M}_\Psi$ is D-testable.

Now we arrive at a notion which is very important for this chapter, namely system identifiability with respect to an incomplete set of observations, represented by Ψ.

3.2.1.1-18. Definition. (To avoid confusion we state again that $\overset{\circ}{M}$ is supposed to be (D_∞-) model identifiable) $\overset{\circ}{M}$ is called Ψ- model (or system)-identifiable, (or model- or system identifiable with respect to Ψ) if

$$\forall d \in D: \exists\, !\ S \in \overset{\circ}{M} \text{ such that } \Psi^{-1}(d) \subseteq S.$$

(Here $\Psi^{-1}(d) = \{x \in D_\infty | \Psi(x)=d\}$, of course; '$\exists!$' means: 'there exists precisely one').
An equivalent definition is:

(3.2.1.1-19) $\overset{\circ}{M}$ is called Ψ- model (or system-) identifiable if the mapping

$$(3.2.1.1\text{-}20)\quad \bar{\Psi}: \overset{\circ}{M} \to \overset{\circ}{M}_\Psi,\quad S \longmapsto \Psi(S) \subseteq D$$

is injective (and so, by construction, bijective) and $\overset{\circ}{M}_\Psi$ is model identifiable.

3.2.1.1-21. Remark. In the choice of examples, and in the comments we have up till now not considered the possibility of controls in the system. The formalism works the same for that case, $d_\infty \in D_\infty$ should then be interpreted as an ideal observation record of all possible measurements on all observable variables involved, under all possible experimental conditions, i.e. under all possible control inputs. In this case it is perhaps even more clear that it is impossible to obtain $d_\infty \in D_\infty$ completely. The usual condition of persistency of excitation in the case of control system identification is now 'hidden' in the condition of system identifiability with respect to Ψ.

3.2.1.1-22. Remark. The fact that Ψ is called the experiment mapping does not mean there will always be an experiment which corresponds to this mapping Ψ. But in an experiment one usually obtains only partial information, and the situation with partial information can be formalized by using a mapping Ψ.

3.2.1.1-23. Conclusion. System identifiability is always with respect to an 'experiment' Ψ. If Ψ is id: $D_\infty \to D_\infty$, then there is either system identifiability, or, if not, then by a reasonable change in the model space

one obtains system identifiability. This has not been sufficiently recognized in the literature, which has given rise to a lot of confusion about the nature of the identifiability problem.
For a critique on the notion of identifiability as it is used in econometrics (where also the word 'identification' is used, as a synonym to 'identifiability'), see [Kal 80].
Note. A set-theoretic definition of deterministic systems was promoted by [Wi 79a] (see also [Wi 83], [Wi 86] and [Sch]). Our approach is inspired by this.

3.2.1.2. Testable stochastic models

In this section we want to apply the general set-up of the previous subsection to the case of stochastic models, with emphasis on stationary stochastic models.
The difficulty in applying the set-up of the previous subsection to stochastic models is that very often it is not clear how they can be falsified by an element $d_\infty \in D_\infty$. The reason is that if d_∞ has probability zero - under the probability measure on D_∞ induced by the model - d_∞ does not necessarily falsify the model. In fact, often the probability of each $d_\infty \in D_\infty$ is zero. In such a case, if one would declare the model to be false as soon as d_∞ has probability zero, then the model would be empty. The other extreme is to declare the model to be in accordance with each possible $d_\infty \in D_\infty$, then the model would be trivial. Both cases are very unsatisfactory because then the truth or falsehood of the model is independent of the data. How are stochastic models refuted in practice? One computes a (finite) number of test statistics, and compares the outcomes with the theoretical values. If they differ too much, the model will be rejected. Of course there is an element of uncertainty in such a procedure. But in principle, if all data $d_\infty \in D_\infty$ would be available, one would refute the model if one or more of the asymptotic values of the test statistics would be unequal to the theoretical values. The event that such a rejection takes place has probability zero, under the probability measure induced by the stochastic specifications of the model.
Therefore we suggest that, in principle at least, one should add to the stochastic set-up of a model, a (finite or at most) countable number of statistical tests $\{T_i\}$, $T_i: D_\infty \to \{\text{true,false}\}$, which test whether the model should be refuted or not. The tests have to be such that the probability that '$T_i(d_\infty)$ = false', is zero under the probability law induced by the model. It

follows that $N := \{d_\infty \in D_\infty | \exists i: T_i(d_\infty) = \text{false}\}$ has probability zero. It will be called the falsifying nullset of the model. So basically our suggestion is to introduce, besides the stochastic set-up of the model, a falsifying nullset. Models with different falsifying nullsets will be considered to be different. Another way to express our suggestion is that the set in which the random variable d_∞ takes its values, is explicitly defined as $D_\infty \backslash N$, where $P(D_\infty \backslash N) = 1$ under the stochastic conditions defined by the model. If $d_\infty \in D_\infty \backslash N$, then the model is falsified. It is clear that this is formally a different model, if $N \neq \emptyset$, than the model with the same stochastic specifications, but in which d_∞ can take its values in all of D_∞.
We now come to a basic definition:

3.2.1.2-1. Definition. A testable stochastic model is a model for which the set $D_\infty \backslash N$, in which the random variable d_∞ can take its values, is explicitly defined , and which is such that $P(D_\infty \backslash N) = 1$. If $d_\infty \in N$, we say d_∞ falsifies the model.

Note that if M is a testable stochastic model, then

$$S(M) = \{d_\infty \in D_\infty | d_\infty \text{ is in accordance with the model}\} =$$
$$= D_\infty \backslash N.$$

The theory of the previous section can now be applied to the sets $S(M) = D_\infty \backslash N$. This gives us the definitions of identifiability etc. Now let us consider the case of a stationary ergodic stochastic process. As we already noted in section 2.2 it is almost surely possible to compute the empirical probability $\pi(A)$ of a set $A \subseteq D_\infty$ (see e.g. [Gu-S], Ch.III, section 3, Théorèmes Ergodiques). In the usual case, in which the σ-algebra of the process is generated by a countable number of sets $A_i \subseteq D_\infty$, testing all $\pi(A_i)$ is equivalent to testing all $\pi(A)$ for all A in the σ-algebra. In this case one can use $\{d_\infty | \pi(A) \neq P(A)$ for some A in the σ-algebra$\}$, where P(A) is the theoretical value of the probability of A induced by the model - as falsifying nullset, or at least as a part of the falsifying nullset. Other things that can be included are tests on the sample moments, covariances, etc. (as long as the total number of (independent) tests is countable, they can all be included).
In case the model space $\overset{\circ}{M}$ consists of Gaussian stationary ergodic stochastic

processes $\{y_t\}_{t\in\mathbb{Z}}$, with zero mean, it clearly suffices for obtaining system identifiability, to test whether the asymptotic sample covariance matrices are equal to their theoretical values. We get:

$$D_\infty\backslash N = \{d_\infty \mid \lim_{N\to\infty}\frac{1}{N-k+1}\sum_{t=o}^{N-k} y_t y_{t-k}^T = \Gamma_k,\ \forall k \in \mathbb{Z}\}.$$

Let us consider the question of Ψ-system identifiability for various possible choices of Ψ.

3.2.1.2-2. Example. If we model the process $\{y_t\}_{t\in\mathbb{Z}}$ as above, and $\Psi(\{y_t\}_{t\in\mathbb{Z}}) = (y_o, y_1, \ldots, y_I)$, for some finite positive integer I, then the model will not be Ψ- identifiable! The reason is that with any finite number of data it is impossible to test an asymptotic limit. In our case, it is impossible to test with complete certainty whether

$\lim_{N\to\infty}\frac{1}{N-K+1}\sum_{t=o}^{N-k} y_t y_{t-k}^T = \Gamma_k$ holds for all k. We can state the following conclusion, which is intuitively clear, but yet important for our analysis: Such a model is not identifiable from a finite data record.

3.2.1.2-3. Example. Consider the same situation as the previous example, but now with

$$\Psi(\{y_t\}_{t\in\mathbb{Z}}) = (y_{t_j}, y_{t_j+1}, \ldots, y_{t_j+I})_{j=o}^{\infty},$$

for some monotonically increasing sequence $\{t_j\}_{j=o}^{\infty}$. Then (at least) for $k = -I, -I+1, \ldots 0, \ldots I$ it can in principle be tested whether

$$\Gamma_k = \lim_{N\to\infty}\frac{1}{N-k+1}\sum_{t=o}^{N-k} y_{t_j+k} y_{t_j}^T$$

holds. The model is Ψ- identifiable if all covariance matrices can be derived from $\Gamma_o, \ldots, \Gamma_I$, using the properties of the model space. This will only be possible if the model space is restricted, i.e. a strict subset of $\mathring{M}$, introduced above. We will return to this in section 3.2.2.

3.2.1.3. Parameter identifiability

To start with let us quote R.E. Kalman [Kal 80] 'Mathematical common sense requires us to view a system first of all as an abstract object. Parameters, or more accurately, problems of the parametrization of the system, enter the analysis at a later stage. (...) If the abstract object is properly understood, parametrization becomes a well-defined mathematical problem. It is usually a difficult problem (...). Since parametrizations are obtained by abstract mathematical procedures, the resulting 'parameters' seldom have direct, intuitive significance; they are determined by their mathematical properties'.

A very general definition of a parametrization is the following

3.2.1.3-1. Definition. A parametrization of a set $\overset{\circ}{M}$ is a surjective mapping $\phi: \Theta \to \overset{\circ}{M}$, for some arbitrary set Θ, the socalled parameter set.
It follows that any mapping can be considered as a parametrization of its image with its domain as the parameter set. Usually Θ is taken to be some 'numerical' set, e.g. a subset of some $\mathbb{R}^n$, such that it is easy to work with to do computations etc. In many cases it is preferable to work with a covering of the set $\overset{\circ}{M}$ with several parametrized sets.

3.2.1.3-2. Definition. An indexed set $\{\phi_i \,|\, i \in I\}$ of parametrizations

$$\phi_i: \Theta_i \to \phi_i(\Theta_i) \subseteq \overset{\circ}{M}$$

is called a covering of $\overset{\circ}{M}$ with parametrizations, if

$$\overset{\circ}{M} = \bigcup_{i \in I} \phi_i(\Theta_i).$$

We speak of a covering with nonoverlapping parametrizations if $\forall i,j \in I,\ i \neq j: \phi_i(\Theta_i) \cap \phi_j(\Theta_j) = \emptyset$. Otherwise we speak of a covering with overlapping parametrizations. If the index set I is finite, we speak of a finite covering, otherwise of an infinite covering.

3.2.1.3-3. Remark. A covering of $\overset{\circ}{M}$ with parametrizations can formally be considered as one parametrization of $\overset{\circ}{M}$, as follows:
Take $\Theta = \bigcup_{i \in I} \{(\theta,i) \,|\, \theta \in \Theta_i\}$, and

$\phi(\theta,i) = \phi_i(\theta) \in \overset{\circ}{M}$, then $\phi: \theta \to \overset{\circ}{M}$ is a parametrization of $\overset{\circ}{M}$.

From now on, let $\overset{\circ}{M}$ denote a system identifiable model space.

3.2.1.3-4. Definition. A parametrization $\phi: \theta \to \overset{\circ}{M}$ of $\overset{\circ}{M}$ is called parameter-identifiable if ϕ is injective (and therefore bijective).

3.2.1.3-5. Remark. A covering of $\overset{\circ}{M}$ with parameter-identifiable parametrizations is an indexed set $\{\phi_i \mid i \in I\}$ of injective mappings $\phi_i: \theta_i \to \overset{\circ}{M}$, such that $\bigcup_{i \in I} \phi_i(\theta_i) = \overset{\circ}{M}$.

A covering of $\overset{\circ}{M}$ with nonoverlapping (parameter)-identifiable parametrizations forms a (parameter-) identifiable parametrization.

For many purposes, like the interpretation of results from a system identification procedure, or understanding the structure of the model set, or doing computations it is desirable to have a (parameter-) identifiable parametrization or at least a (preferably finite) covering with (parameter-) identifiable parametrizations.
The existence of an identifiable parametrization is trivial, because one can take $\theta := \overset{\circ}{M}$ and $\phi :=$ identity on $\overset{\circ}{M}$. However, this is practically useless, because $\overset{\circ}{M}$ is an abstract object. Apparently, parameter identifiability is not the only requirement that one wants to make for a parametrization. In many cases it is possible to define a natural topology on the model space $\overset{\circ}{M}$ (although there are cases in which there are several competing 'natural topologies'). In such a case it is desirable to have a continuous parametrization, or at least a (preferably finite, or locally finite) covering with continuous parametrizations. If a parametrization is identifiable, then it has an inverse. It is desirable that this inverse is continuous. I.e. it is desirable to have a homeomorphic parametrization, or at least a (preferably finite, or locally finite) covering with homeomorphic parametrizations. And one can go on in making requirements like differentiability etc., depending on the structure of the model space $\overset{\circ}{M}$. We will return to this type of questions for the models under consideration in this work, in chapter 4 (and 5).
Let τ denote a topology on θ. Then we can define the concept of local parameter-identifiability:

3.2.1.3-6. Definition. A parametrization ϕ: $\Theta \to \overset{\circ}{M}$ is called locally (parameter-) identifiable at $\theta \in \Theta$ (w.r.t. to the topology τ on Θ) if there exists an open neighbourhood U of θ, such that $\phi|_U$: $U \to \phi(U) \subseteq \overset{\circ}{M}$ is identifiable. The parametrization ϕ is called locally (parameter-) identifiable if ϕ is locally (parameter-) identifiable at θ for all $\theta \in \Theta$.

3.2.1.3-7. Remark. If ϕ: $\Theta \to \overset{\circ}{M}$ is a locally identifiable parametrization then there exists a covering of overlapping identifiable parametrizations. If Θ is compact, this covering can be taken finite, and if Θ is locally compact, the covering can be taken locally finite. This can be shown as follows: Let for each $\theta \in \Theta$, U_θ be a neighbourhood of θ, as in definition (3.2.1.3-6). Then $\{U_\theta | \theta \in \Theta\}$ is a covering of Θ and $\phi|_{U_\theta}$ is an identifiable parametrization. The rest is standard topology.

3.2.2. Application to the finite identifiability problem

In this section we want to apply the results obtained in section 3.1 on partial realization of covariance sequences to the socalled 'finite identifiability problem'. This problem has been studied in [Tig]. However, there a different terminology is used which we find misleading. Consider a model space $\overset{\circ}{M}$ with parametrization ϕ: $\Theta \to \overset{\circ}{M}$, so Θ is the parameter space. $\overset{\circ}{M}$ consists of stochastic models, and each model is identified, by [Tig] and others, with the probability measure it induces on the observables of the model. At least in the stationary ergodic case the same can be done in the set-up of section 3.2.1, using the empirical probability distribution as test statistic; provided the σ-algebra is countably generated. (However, in general, there might be problems related with contiguity and non-contiguity of time varying stochastic processes. It is then not always sufficient for model identifiability that the probability measures differ). Let P_θ^{I+1} denote the probability measure of a sample $(y_0,\ldots,y_I)$ from a stochastic process $\{y_t\}_{t\in\mathbb{Z}}$. Then $(y_0,\ldots,y_I)$ is called informative by [Tig] (Def. 1.6.2, p.27, with ϕ = id. and n = I+1) for all $\theta \in \Theta$, if for all $\theta_0,\theta_1 \in \Theta$, $\theta_0 \neq \theta_1$, one has

$$(3.2.2\text{-}1) \qquad P_{\theta_0}^{I+1} \neq P_{\theta_1}^{I+1}.$$

We shall call this condition N(I+1). Why is the term 'informative' misleading? To answer this we will use a simple example. Consider a scalar random variable y

with probability law $N(\mu,\sigma^2)$, and let $(y_1,\ldots,y_n)$ be a sample of n independent drawings from this probability distribution. Let $\Theta = \{(\mu,\sigma^2) \mid \mu \in \mathbb{R},\ \sigma^2 \in \mathbb{R}_+\}$. Then what is the smallest n for which $(y_1,\ldots,y_n)$ is informative? According to the definition above, it is n = 1, because if $(\mu_o,\sigma_o^2) \neq (\mu_1,\sigma_1^2)$, then $P^1_{(\mu_o,\sigma_o^2)} \neq P^1_{(\mu_1,\sigma_1^2)}$. This means that a single drawing (y_1) from the normal distribution $N(\mu,\sigma^2)$ is called informative for both μ and σ^2. However, it is intuitively clear that if μ is unknown, one drawing of $N(\mu,\sigma^2)$ can not be informative for σ^2 in any reasonable sense of the word, because from one number (y_1) one can not get any information about the dispersion of the random variable. It is clear that if $P^n_{\theta_o} = P^n_{\theta_1}$, then in the sample $(y_1,\ldots,y_n)$ there cannot be any information that can distinguish between θ_o and θ_1. Therefore we agree with [Tig] that the sample in such a case is not informative. So his definition of 'informative' turns out to be (only) a kind of necessary condition. But as long as there is no good definition of 'informative sample', it is in general impossible to come up with necessary and sufficient conditions. (In fact, if the word 'informative' is interpreted as in ordinary language, then an observation is informative for a model, if our knowledge about the model increases as a result of this observation. In that sense in stochastic models, usually every observation is informative. In deterministic models, there may be a finite number of observations from which the model can be identified (within the model set); any other observations are then no longer informative. Such a finite sample could be called 'fully informative' for the model. However, for stochastic models, a finite sample can not be fully informative in this sense of the word).

One way to save the situation in the models under study, is to define an intuitively appealing sufficient condition for the concept of an informative sample size.

This sufficient condition is

3.2.2-2. Condition. S(I+1): There are parameters $\{\gamma_j\}$, that determine the probability distribution, and there are unbiased estimators $\hat{\gamma}_j(y_o,\ldots,y_I)$ for all j, i.e.

(3.2.2-3) $\quad E\hat{\gamma}_j(y_o,\ldots,y_I) = \gamma_j$ for all j.

If condition S(I+1) holds then we will agree that I+1 is an informative sample size (for the probability distribution). But we do not want to imply that S(I+1) is a necessary condition. This is an open question, especially because, as said above, a good definition of the word informative sample, is still lacking.
It turns out, however, that in several cases N(I+1) is equivalent to S(I+1) for all I. Then N(I+1) (and so S(I+1)) is a necessary and sufficient condition for $(y_o,\ldots,y_I)$ to be informative.
Let us return to our simple example of a sample from $N(\mu,\sigma^2)$. If the sample size is <u>two</u>, one can form unbiased estimators

$$(3.2.2\text{-}4)\quad \begin{cases} \hat{\mu} = (y_o+y_1)/2, \\ \hat{\sigma}^2 = (y_o-\bar{\mu})^2 + (y_1-\bar{\mu})^2 = \frac{1}{2}(y_o-y_1)^2, \end{cases}$$

and μ,σ^2 determine the probability distribution.
In this case N(1) holds, but S(1) does not. So 'N(I+1) iff S(I+1)' does not hold in this case. However, if we put $\mu = 0$, then there exists an unbiased estimator of σ^2 using one observation, namely

$$(3.2.2\text{-}5)\quad \hat{\sigma}^2 = y_o^2$$

So in that case ∀I: N(I+1) iff S(I+1).
A similar situation applies in the case of a stochastic system as treated in section 3.1.3, because there too one has zero mean. To obtain <u>unbiased estimators</u> for the covariance matrices $\Gamma_o,\ldots,\Gamma_I$, one can do with J = I + 1 observations; the unbiased estimators are

$$(3.2.2\text{-}6)\quad \hat{\Gamma}_k = \sum_{t=o}^{I-k} y_{t+k}y_t^T/(I-k+1),$$

which is meaningful for all $k \in \{0,1,\ldots,I\}$.
Because $\Gamma_o,\Gamma_1,\ldots,\Gamma_I$ determine P_θ^{I+1}, it follows that

(3.2.2-7) N(I+1) iff S(I+1) for all $I \geq 0$.

Therefore <u>in this case</u> it suffices to consider N(I+1), I = 0,1,..., i.e. in <u>this</u> case one can say that if N(I+1) holds then I+1 is an informative sample

size (for the model).
There is also another, less discutable interpretation of N(I+1), namely as a (necessary and sufficient) condition for system identifiability in a case of missing observations. If one has a data record $\Psi\{y_k\}$, as described in example (3.2.1.2-3), then there is system identifiability w.r.t. Ψ if N(I+1) holds. (I is the same as in example (3.2.1.2-3)).
For both interpretations of N(I+1) we can apply our results of section 3.1.3.2, to obtain:

3.2.2-8. Theorem. Consider the model space $\overset{\circ}{M}$ = the set of all stochastic arma(p,q) systems (resp. arma($\bar{p}$,$\bar{q}$) models), which satisfy the conditions of (2.4.1-1). Let Θ be an identifiable parametrization of $\overset{\circ}{M}$. Then
(i) if $q \geq p$, $I = q+mp$ is the minimal number such that N(I+1) holds and N(I'+1) holds for all $I' \geq I$.
(ii) if $q \leq p$, N(I'+1) holds for all $I' \geq I = p + mq$.

Proof. This follows from theorems (3.1.3.2-5) and (3.1.3.2-13) and the remarks made in section 3.2.1.2, just before formula (3.2.1.2-2). Q.E.D.

3.2.2-9. Remarks. (i) It follows that $J = I+1 = p+q+(m-1)\min(p,q) + 1$ is an informative sample size, which is proved to be minimal if $q \geq p$.
(ii) In the case of missing observations, such that one has a data record as in example (3.2.1.2-3) then there is system identifiability if $I' \geq I = p+q+(m-1)\min(p,q)$, where I'+1 denotes the number of consecutive measurements (denoted by I in the example). Whether I is minimal or not depends on the sequence $\{t_k\}$. If it is such that no other covariances can be estimated than $\Gamma_0,\ldots,\Gamma_I$, and $q \geq p$ then I is minimal indeed.

(iii) If $\overset{\circ}{M}' \subseteq \overset{\circ}{M}$ is a model space contained in $\overset{\circ}{M}$, and H' an identifiable parametrization of $\overset{\circ}{M}'$, then N(I'+1) holds w.r.t. H' for all $I' \geq I = (p+q) + (m-1)\min(p,q)$. So I+1 is still an informative sample size etc. (But the minimality of I will of course no longer be assured in general).

A few remarks should be made w.r.t. the problem of the parametrization of $\overset{\circ}{M}$. Of course our approach has been to separate the system identifiability problem and the problem of parametrization of the model space, and we have obtained results about system identifiability. This is different in the usual approach

to identifiability, where the two aspects are not separated. (see [Tig]). However, if one wants to apply the results in a practical problem, one has to face the parametrization problem. I.e. how can one construct a parametrization of $\overset{\circ}{M} = \text{arma}(\bar{p},\bar{q})$?

In fact this is not an easy question. The reason is that reduction to a canonical form usually does not leave $\text{arma}(\bar{p},\bar{q})$ invariant, i.e. there are arma(p,q) models that become e.g. arma(p,q+1) - models in the canonical form. Compare the warning (3.1.2.3-3). For standard parametrizations of the arma-model, see e.g. [Han 71], [Gui 75], [Gui 81], [Dei 85], [Fo 75]. None of them seems to be invariant for $\text{arma}(\bar{p},\bar{q})$ if $p \neq q$. The canonical forms based on the (Kronecker) observability indices, i.e. those of [Gui 75], [Fo 75] and others leave $\text{arma}(\bar{p},\bar{p})$ models invariant. (The set $\text{arma}(\bar{p},\bar{p})$ is of course exactly equal to the set of linear stochastic systems with McMillan degree $\leq n = pm$ and observability indices all $\leq p$ so one could even parametrize the set $\text{arma}(\bar{p},\bar{p})$ by parametrizing the set of stochastic state space models of McMillan degree $\leq n = pm$ and observability indices all $\leq p$. How this can be done is described (implicity) in chapter 4.)

One way to obtain an identifiable parametrization is by using theorem (3.1.2.3-29) (and its proof). We distinguish three cases:

(a) If $p = q$, then $\overset{\circ}{M} = \text{arma}(\bar{p},\bar{p})$ and one can use a parametrization as just referred to.

(b) If $p > q$, then define

$$B_r(z) := z^{p-q}B(z), \text{ so } B_r(z) = B_o z^{p-q} + \ldots + B_q z^p.$$

Then $(A(z),B_r(z))$ is an mfd-pair in $\text{arma}(\bar{p},\bar{p})$. So one can use a canonical form for $\text{arma}(\bar{p},\bar{p})$, as was just referred to, with $\deg(A(z)\ B_r(z)) \leq p$ and with A_o nonsingular and such that the coefficient matrices in $B_r(z)$ of all powers of z which are less than p-q, are identically zero. (This forms a system of (linear) restrictions on the parameters). Now $(A(z),z^{-(p-q)}B_r(z))$ can be used as a parametrization of $\text{arma}(\bar{p},\bar{q})$

(c) If $q > p$, $\overset{\circ}{M} = \text{arma}(\bar{p},\bar{q}) \tilde{=} \text{arma}(\bar{p},\bar{p}) \times (\mathbb{R}^{m\times m'})^{(q-p)}$. (see the proof of theorem (3.1.2.3-29)). Here $(A(z),B_r(z)) \in \text{arma}(\bar{p},\bar{p})$ and $(H_o,H_1,\ldots,H_{q-p-1}) \in (\mathbb{R}^{m\times m'})^{(q-p)}$, the first q-p Markov matrices of the system. Now take a parametrization for $\text{arma}(\bar{p},\bar{p})$, giving $(A(z),B_r(z))$,

with $\deg[A(z)\ B_r(z)] \leq p$, as was referred to above. Also use $H_0,\ldots,H_{q-p-1}$ as free parameters. Then $(A(z),B(z))$ is found as described just after (3.1.2.3-34).

This parametrization is surjective, i.e. it parametrizes <u>all</u> elements in $\text{arma}(\bar{p},\bar{q})$ (this is in sharp distinction with e.g. [Han 71], theorem 2 $(iii)_2$). We want to finish this section by a comparison of our results with those of [Tig].

Our general formula concerning informative samples sizes of an arma(p,q) model (or an $\text{arma}(\bar{p},\bar{q})$ model) is that the sample size J is informative if $J \geq I+1 = p + q + (m-1)\min(p,q) + 1$. This number is minimal if $q \geq p$. If $q < p$ minimality has not been proved.

Tigelaars findings coincide with this in the s.i.s.o. case. In the m.i.m.o. case his findings coincide if $p = 0$ or $q = 0$, i.e. in the multivariable ma(q) case and the multivariable ar(p) case. In those cases $I + 1 = q + 1$, resp. $p+1$. In the mixed multivariable arma(p,q) case, his general formula is $\forall m > 1,\ \forall p \geq 1\ \forall q \geq 1$: $J \geq q + (m+1)p$. ([Tig], theorems 3.5.3 and 3.5.4). This is equal to our formula iff $p = 1$. In the case $m = 2$, $p = 1$, $q = 1$ Tigelaar shows the minimality of his result. In <u>all other cases</u> our minimality results are new. And if $m \geq 2$, $p > 1$, our formula is an <u>improvement</u> over Tigelaars formula.

CHAPTER 4

ON THE DIFFERENTIAL GEOMETRICAL AND TOPOLOGICAL STRUCTURE OF FAMILIES OF LINEAR SYSTEMS, AND CANONICAL FORMS

4.1. Short introduction

In the coming chapters we will consider the problem of identification of a stochastic linear system. Research in the past two decades has shown that the space of linear systems of a fixed order is essentially 'nonlinear', i.e. it cannot be parametrized by a Euclidean space $\mathbb{R}^{n(m+p)}$. This has important consequences for identification. Identification of linear systems is connected in a natural way to topology and geometry, because one needs (and there are) natural notions of proximity of models. The problem is, however, that it is not obvious - and in fact not true - that the natural notion of proximity derived from a seemingly obvious parametrization is in fact equivalent to the natural notion of proximity of systems in input-output form. So one has to be careful because depending on the representation of the system and depending on its use, one can come up with different topologies and geometries. Some of these issues will be studied here.

In this chapter, we want to give a concise survey of some of the constructions and results that are known about the topological and geometrical structure of the spaces of linear systems of fixed order. First we start with an introduction in some of the concepts from differential geometry that will be used. (In the next chapter we will treat some concepts from Riemannian geometry).

4.2. On the definitions of a differentiable manifold and a fibre bundle

Of course we cannot do much more here then to give definitions, make some remarks and give some references. The definition of a differentiable manifold is taken from [Boo]. Other references include [Ko-N], [Spi], [Au-M] and [Bi-C].

4.2-1. Definition. A topological space M is called locally Euclidean of dimension n, if each point $p \in M$ has a neighbourhood U which is homeomorphic to an open subset U' of $\mathbb{R}^n$, n fixed.

4.2-2. Definition. A (topological) manifold M of dimension n is a topological

space with the following properties:

(i) M is Hausdorff,

(ii) M is locally Euclidean of dimension n, and

(iii) M has a countable basis of open sets.

4.2-3. Definition. Let M be a topological manifold of dimension n. Each pair (U,ϕ), where U is an open set of M and ϕ is a homeomorphism of U to an open subset of $\mathbb{R}^n$, is called a coordinate neighbourhood.

4.2-4. Definition. Let $R \subseteq \mathbb{R}^n$, $S \subseteq \mathbb{R}^m$ be open subsets. A function $f: R \to \mathbb{R}$ is called (a) C^∞ if f is n times differentiable for each $n \in \mathbb{N}$, (b) C^ω or real analytic , if each $r \in R$ has a neighbourhood in which f can be written as a convergent power series.

A mapping $f: R \to \mathbb{R}^m$ is called C^∞ resp. C^ω if each of its components is C^∞ resp. C^ω.

A mapping $f: R \to S$ is called a (C^∞-) diffeomorphism, resp. real analytic diffeomorphism if f has an inverse $f^{-1}: S \to R$ and f and f^{-1} are both C^∞-(resp. C^ω-) mappings.

4.2-5. Definition. Two coordinate neighbourhoods (U,ϕ) and (V,Ψ) are called C^∞-compatible (resp. C^ω-compatible) if $U \cap V \neq \emptyset$ implies that $\phi \circ \Psi^{-1}: \Psi(U \cap V) \to \phi(U \cap V)$ is a (C^∞-)diffeomorphism (resp. a real analytic diffeomorphism).

4.2-6. Definition. A C^∞-structure (resp. C^ω structure) on a topological manifold M is a family $\mathring{U} = \{(U_\alpha,\phi_\alpha)\}$ of coordinate neighbourhoods such that

(i) $\bigcup_\alpha U_\alpha = M$, i.e. $\{U_\alpha\}$ is an open covering of M,

(ii) for any α,β the neighbourhoods (U_α,ϕ_α) and (U_β,ϕ_β) are C^∞-compatible (resp. C^ω-compatible), and

(iii) any coordinate neighbourhood (V,Ψ) compatible with every $(U_\alpha,\phi_\alpha) \in \mathring{U}$, is itself in $\mathring{U}$.

4.2-7. Definition. A C^∞-differentiable manifold (resp. real analytic manifold) is a topological manifold together with a C^∞-(resp. C^ω-) structure.

The following theorem (see [Boo], p. 54, (1.3)) shows that (i) and (ii) of definition (4.2-6) are sufficient to define a unique C^∞- structure (resp. C^ω-structure).

4.2-8. Theorem. Let M be a Hausdorff space with a countable basis of open sets. If $V = \{V_\beta, \Psi_\beta\}$ is a covering of M by C^∞-compatible (resp. C^ω-compatible) coordinate neighbourhoods, then there is a unique C^∞-structure (resp. C^ω-structure) on M containing these coordinate neighbourhoods. In other words (M,V) defines uniquely a C^∞-differentiable manifold (resp. a real analytic manifold).

Proof: cf [Boo].

Next we want to give the definition of a fibre bundle (what we define here is sometimes called a locally trivial fibre bundle). References are [Ste], [Ko-N]. The following definitions are taken from [Ste].

Definition. A topological group G is a set which has group structure and a topology such that

(a) g^{-1} is continuous for g in G, and

(b) $g_1 g_2$ is continuous simultaneously in g_1 and g_2, i.e. the map $G \times G \to G$ given by $(g_1, g_2) \longmapsto g_1 g_2$ is continuous when $G \times G$ has the usual topology of a product space.

Definition. If G is a topological group, and Y a topological space, we say that G is a topological transformation group of Y relative to a map $\eta: G \times Y \to Y$ if

(i) η is continuous

(ii) $\eta(e,y) = y$, where e is the identity in G, and

(iii) $\eta(g_1 g_2, y) = \eta(g_1, \eta(g_2, y)), \forall g_1, g_2 \in G$, and $\forall y \in Y$.

Notation: $\eta(g,y)$ will be abbreviated by $g \cdot y$.

Definition. G is called effective if $((\forall y \in Y: g.y = y) \Rightarrow g = e)$.

First the definition of a socalled coordinate bundle will be given, and then the definition of a fibre bundle, which can be considered as a coordinate

bundle with an extra property that is comparable to 4.2-6 (iii) for the case of differentiable manifolds.

4.2-9. Definition. A coordinate bundle $\overset{\circ}{B}$ is a collection $(B,X,p,Y,\{(V_j,\bar{\phi}_j)\}_{j\in J},G)$ as follows:

(1) a space B called the bundle space,

(2) a space X called the base space

(3) a map $p: B \to X$ of B onto X called the projection,

(4) a space Y called the fibre,

(5) an effective topological transformation group G of Y called the group of the bundle,

(6) a family $\{(V_j,\bar{\phi}_j)\}_{j\in J}$ of pairs $(V_j,\bar{\phi}_j)$ called coordinate neighbourhoods, where each V_j is an open subset of X, J an index set, $\bigcup_J V_j = X$, and each $\bar{\phi}_j$ is a homeomorphism

(7) $$\bar{\phi}_j: V_j \times Y \to p^{-1}(V_j)$$

called the coordinate function. The coordinate functions are required to satisfy the following conditions:

(8) $$p \circ \bar{\phi}_j(x,y) = x \qquad \forall x \in V_j, \forall y \in Y,$$

(9) if the map $\bar{\phi}_{jx}: Y \to p^{-1}(x)$ is defined by $\bar{\phi}_{j,x}: y \to \bar{\phi}_j(x,y)$, then for each pair i,j with $i,j \in J$ and $\forall x \in V_i \cap V_j$, the homeomorphism

$$(\bar{\phi}_{jx})^{-1}(\bar{\phi}_{ix}): Y \to Y$$

coincides with the operation of an element of G (it is unique since G is effective), and

(10) $\forall i,j \in J$ the map $g_{ji}: V_i \cap V_j \to G$, $x \longmapsto (\bar{\phi}_{jx})^{-1}(\bar{\phi}_{ix})$ (considered as an element of G) is continuous.

4.2-10. Definition. (a) $Y_x := p^{-1}(x)$ is called the fibre over x.

(b) The functions g_{ji} defined in (10) above are called the coordinate transformations of the bundle.

4.2-11. Definition. Two coordinate bundles $\overset{\circ}{B}$ and $\overset{\circ}{B}'$ are said to be equivalent in the strict sense if they have the same bundle space B, the same base space X, the same projection p, the same fibre Y and the same group G, and their families of coordinate neighbourhoods $\{(V_j,\bar{\phi}_j)\}_{j\in J}$ and $\{(V_{j'},\bar{\phi}_{j'})\}_{j'\in J'}$ are such that their union $\{(V_j,\bar{\phi}_j)\}_{j\in J\cup J'}$ is also a set of coordinate neighbourhoods of a coordinate bundle. It can be shown easily that this is a proper equivalence relation (see [Ste], p.9.).

4.2-12. Definition. A fibre bundle is an equivalence class of coordinate bundles.

Remarks. (i) Equivalently one could define a fibre bundle as the coordinate bundle having all possible coordinate functions of an equivalence class. (To do this logically sound we have to restrict the index class J to e.g. being a subset of $2^{X\times Y\times X}$. Note that each pair $(V_j,\bar{\phi}_j) \in 2^{X\times Y\times X}$ if you identify $(V_j,\bar{\phi}_j)$ with the set

$$\{(x,y,\bar{\phi}_j(x,y)) \mid x \in V_j, y \in Y\} \subseteq X \times Y \times X.)$$

(ii) A fibre bundle as defined here is also called a locally trivial fibre bundle in the literature (see e.g. [Hus], section 2.6). In this thesis only locally trivial fibre bundles occur, more general fibre bundles do not occur. Every fibre bundle will be understood to be locally trivial.

4.2-13 Definition. A fibre bundle which has as fibre Y a vector space, and as group G a group of linear transformations of Y is called a vector bundle ([Ste], section 6.6, calls this a bundle of linear spaces).

4.2-14. Definition. A bundle $\overset{\circ}{B} = \{B,X,p,Y, \{(V_j,\tilde{\phi}_j)\},G\}$ is called a principal bundle if Y = G (and G operates on Y = G by left translations) ([Ste], p.35).

4.2-15. Definition. Let $\overset{\circ}{B} = \{B,X,p,Y,\{(V_j,\bar{\phi}_j)\},G\}$ be an arbitrary bundle. The associated principal bundle $\overset{\circ}{\tilde{B}}$ of $\overset{\circ}{B}$ is the bundle with base space X, fibre G, group G and the coordinate transformations $g_{ij}: x \to (\bar{\phi}_{ix})^{-1}\bar{\phi}_{jx} \in G$.

For the existence and the uniqueness under an appropriate notion of equivalence (broader than the notion of strict equivalence introduced above) we refer to [Ste], sections 3.2 and 8.1.

Remark. If $G \subseteq Gl_n(\mathbb{R})$ then to a principal fibre bundle there corresponds an essentially unique (under an appropriate notion of equivalence) vector bundle; i.e. the relation between a vector bundle and its associated principal bundle is essentially one-to-one. See e.g. [Ko-N], vol.I, p.113.

4.2-16. Definition. A coordinate bundle is called a product bundle (or a trivial bundle) if there is just one coordinate neighbourhood $V = X$, and the group G consists of the identity element e alone.

4.2-17. Definition. A cross-section of a bundle is a continuous map $f: X \to B$ such that $p \circ f(x) = x, \forall x \in X$.

An important theorem is the one that gives the relation between the (non-) existence of a cross-section and the (non-) triviality of the bundle. For the meaning of the notion 'equivalent in G' we refer to [Ste], around sections 8.4 and 4.3.

4.2-18. Theorem. A bundle with group G is equivalent in G to a product bundle if and only if the associated principal bundle admits a cross-section.

4.2-19. Remark. If M is a submanifold of the base space X of a fibre bundle $p: B \to X$, then the restriction of p to $p^{-1}(M)$ represents a fibre bundle with base space M. The fibre is the same as in the original bundle. For more details we refer to

4.3. Some general remarks on the fibre structure of families of systems-with-states.

(This section may be skipped by the reader). Before giving the results on the topological fibre structure in the case of linear systems, we want to make some general remarks about why and how such a fibre structure comes up in system theory. We will only consider time-invariant systems, starting at some time $t_o \in \mathbb{R}$. Our basic notion in this section will be a 'system-plus-state'. The notation is as in chapter 2. The definitions given here are only valid

within this subsection.

4.3-1. Definition. A 'system-with-state' is an input-output mapping

$F: \overset{\circ}{U} \to \overset{\circ}{Y}$, $\overset{\circ}{U} \subseteq \{u:[0,\infty) \to U\}$, $\overset{\circ}{U}$ closed under concatenation, and
$\overset{\circ}{Y} = \{y:[0,\infty) \to Y\}$.
In a finite time interval $[0,T]$ a control $u|_{[0,T]}$ maps a 'system-with-state' F to a 'system-with-state' G, which is defined as follows

$$G: \overset{\circ}{U} \to \overset{\circ}{Y},\ y = G(u') \text{ iff } y(t) = y'(t+T) \text{ with } y' = F[u(s)|_{s\in[0,T)} \cup u'(s+T)|_{s\in[0,\infty)}].$$

4.3-2. Assumption. If F is mapped to G by some control function $u: [0,T] \to U$, $u \in \overset{\circ}{U}|_{[0,T]}$ for some T, then there is a T' and

a $u':[0,T'] \to U, u' \in \overset{\circ}{U}|_{[0,T']}$ such that u' maps G to F.

Remark: In the case of linear systems this assumption is satisfied.

4.3-3. Definition. We will call F and G equivalent, notation $F \sim G$, if $\exists T$, $\exists u \in \overset{\circ}{U}|_{[0,T]}$ such that F is mapped to G by u.

This is indeed an equivalence relation:
(i) $F \sim F$; take $T = 0$,
(ii) $F \sim G \Rightarrow G \sim F$; this is assumed (assumption 4.3-2),
(iii) $F \sim G$ and $G \sim H \Rightarrow F \sim H$; just apply the concatenation (u'',T'') of the controls (u,T) and (u',T'), needed for F to be mapped to G resp. G to be mapped to H.

4.3-4. Definition. A system is an equivalence class of 'systems-with-states'.

Let π denote the projection of a system-with-state into its equivalence class, i.e. the system.

4.3-5. Definition. The states of a system [F] are the elements of $\pi^{-1}[F]$, i.e. the elements of the equivalence class.
One can call $\pi^{-1}[F]$ the fibre over F of the projection π.

It seems plausible that this set-up can be generalized and especially that the assumption 4.3-2 can be dropped, but we don't want to go into that any further here. Our only motivation is to show that the fibre structure of the set of systems with states that will be presented in the next paragraph is not something specific for the linear systems case, but is something rather fundamental for system theory.
It also shows clearly that a state is only well-defined with respect to a given system, (i.e. states in different systems cannot be compared directly, one cannot say that two systems are in the same state, without explicitly defining what is meant by that, i.e. what equivalence relation on the states one uses, if any. In many cases no reasonable global equivalence relation exists; this is in fact one of the implications of the result on non-triviality of the 'state bundle' that is treated in the next paragraphs!)

4.4. Nice selections and canonical forms

We now return to deterministic linear systems, of the form (compare (2.3.2-2)):

$$(4.4\text{-}1)\quad \begin{cases} x_{k+1} = Ax_k + Bu_k, & k \in \mathbb{Z},\ x_k \in \mathbb{R}^n,\ u_k \in \mathbb{R}^{m'}, \\ y_k = Cx_k (+ Du_k), & y_k \in \mathbb{R}^m. \end{cases}$$

As the role of D in the matters to be treated is trivial, we take D = 0, or D = I, if this is mentioned explicitly, to simplify the formulas. (The continuous time case is completely analogous. However, in this chapter we will concentrate on the discrete time case). Suppose (A,B,C) is a minimal representation (for more general cases we refer to the literature), i.e. (A,B,C) is observable and reachable. In other words:
(i) the reachability matrix R(A,B) := [B,AB,A^2B ...] has full rank (rank=n) and

(ii) the observability matrix $Q(A,C) := \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix}$, has full rank (rank=n).

Notation: Let

$R_j(A,B) := [B,AB,\ldots,A^{j-1}B]$, and

$$Q_j(A,C) := \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{j-1} \end{bmatrix}.$$

From the Cayley-Hamilton theorem it follows easily that

$$\text{rk } R(A,B) = \text{rk } R_j(A,B) \text{ and } \text{rk } Q(A,C) = \text{rk } Q_j(A,C), \quad \forall j \geq n-1.$$

Now consider the set of all matrix triples (A,B,C) which form a minimal representation and have the sizes n×n, n×m', m×n. This set is denoted by $L^m_{m',n,m}$ ($= L^{co,cr}_{m',n,m}$ in the notation of [Haz 77]), the upper index m stands for minimal. It is well-known that two matrix triples (A,B,C), $(\tilde{A},\tilde{B},\tilde{C}) \in L^m_{m',n,m}$ represent the same system (with respect to input-output behaviour) iff there exists a real nonsingular n×n matrix T, i.e. $T \in Gl_n(\mathbb{R})$, such that $(A,B,C)^T := (TAT^{-1},TB,CT^{-1}) = (\tilde{A},\tilde{B},\tilde{C})$. This means that a change of basis of the state space is allowed without changing the system, and that this is the only kind of change in the entries of (A,B,C) that does not affect the system. We will call two matrix triples (A,B,C), $(\tilde{A},\tilde{B},\tilde{C}) \in L^m_{m',n,m}$ (input-output-) equivalent if they represent the same system. The equivalence classes in $L^m_{m',n,m}$ correspond to the different systems represented by elements of $L^m_{m,n,p}$. An equivalence class is of the form $\{(A,B,C)^T | T \in Gl_n(\mathbb{R})\}$. It is also called an 'orbit under the action of $Gl_n(\mathbb{R})$'.
We define $M^m_{m',n,m}$ ($= M^{co,cr}_{m',n,m}$ in the notation of [Haz 77]) as the quotient space of $L^m_{m',n,m}$ under this equivalence. I.e. the points of $M^m_{m',n,m}$ are the orbits in $L^m_{m',n,m}$ and its topology is the finest topology for which the natural projection $\pi: L^m_{m',n,m} \to M^m_{m',n,m}$ is continuous. (The topology on $L^m_{m',n,m}$ is the standard topology of $L^m_{m',n,m}$ considered as a subset of $\mathbb{R}^{nm'+n^2+mn}$). Let $[(A,B,C)]$ denote the equivalence class $\{(A,B,C)^T | T \in Gl_n(\mathbb{R})\}$ of (A,B,C).

In a similar way one can introduce the corresponding set of systems with states:
Let

$$L^m_{1,m',n,m} := \{(A,B,C,x) | x \in \mathbb{R}^n, (A,B,C) \in L^m_{m',n,m}\}.$$

Let $(A,B,C,x)^T := (TAT^{-1},TB,CT^{-1},Tx)$ and let $[(A,B,C,x)]$ denote the equivalence class $\{(A,B,C,x)^T \mid T \in G\ell_n(\mathbb{R})\}$. Finally, let $M^m_{l,m',n,m} = \{[(A,B,C,x)] \mid (A,B,C) \in L^m_{m',n,m}, x \in \mathbb{R}^n\}$ be the quotient space

$L^m_{l,m',n,m}/G\ell_n(\mathbb{R})$ with the corresponding quotient topology.

Remark. $M^m_{l,m',n,m}$ will turn out to be a vector bundle called the state bundle; see theorem 4.6-1 and remark 4.6-3. It was introduced in [Haz 77], where it is denoted by E or E^u in a somewhat more general setting.

Remark: Note that the action of $T \in G\ell_n(\mathbb{R})$ on (A,B,C,x) can again be interpreted as a change of basis of the state space. From (4.4-1) it follows that

$$(4.4\text{-}2)\quad \begin{cases} Tx_{k+1} = (TAT^{-1})(Tx_k) + (TB)u_k, \\ y_k = (CT^{-1})(Tx_k). \end{cases}$$

So indeed (A,B,C,x) is equivalent to $(A,B,C,x)^T$.
Now we turn to the nice selections. We number the columns of the reachability matrix $R(A,B)$ by pairs of indices (i,j), $i \in \{0,1,2,\ldots\}$, $j \in \{1,2,\ldots,m'\}$; the index (i,j) denotes the $(im'+j)$-th column. The set of all such indices is denoted by J.

4.4-3. Definitions [Haz 77].
(i) A nice selection α is a subset of J of size n such that

$$(i' \leq i \text{ and } (i,j) \in \alpha) \Rightarrow (i',j) \in \alpha.$$

(ii) If α is a nice selection we define the j^{th} successor index of α, $s(\alpha,j)$, $j = 1,2,\ldots,m'$, as the element $(k,j) \in J$ such that $(k',j) \in \alpha$ for all $0 \leq k' < k$ and $(k,j) \notin \alpha$.

4.4-4. Notation. If $\alpha \subseteq J$ we denote with $R(A,B)_\alpha$ the matrix obtained from $R(A,B)$ by removing all columns whose index is not in α.
We now state some relatively simple but important facts.

4.4-5. Lemma. If $(A,B,C) \in L^m_{m',n,m}$ then there is a nice selection α such that $R(A,B)_\alpha$ is a nonsingular (square) matrix.

Proof. Cf. [Haz-K 75], lemma (2.4.1).

4.4-6. Lemma. For each m'-sequence of n-vectors $z_1, z_2, \ldots, z_{m'} \in \mathbb{R}^n$ there is precisely one pair of matrices (A,B) of sizes n×n and n×m' respectively, such that

$$(4.4\text{-}7) \quad R(A,B)_\alpha = I_n,\ R(A,B)_{s(\alpha,j)} = z_j,\ j = 1,2,\ldots,m'.$$

Proof: Cf. [Haz-K 75], lemma (2.3.3).

4.4-8. Notation. Let α be a nice selection. We define

$$(4.4\text{-}9) \quad U_\alpha^m = \{(A,B,C) \in L_{m',n,m}^m \mid R(A,B)_\alpha \text{ nonsingular}\},$$

$$(4.4\text{-}10) \quad W_\alpha^m = \{(A,B,C) \in L_{m',n,m}^m \mid R(A,B)_\alpha = I_n\}.$$

Clearly $W_\alpha^m \subseteq U_\alpha^m$.
Consider the following continuous mapping

$$c_\alpha : U_\alpha^m \to W_\alpha^m,$$

$$(4.4\text{-}11) \quad c_\alpha(A,B,C) = (R(A,B)_\alpha^{-1} A R(A,B)_\alpha, R(A,B)_\alpha^{-1} B,\ C R(A,B)_\alpha).$$

It can now easily be shown that

4.4-12. Lemma. $(A,B,C), (\tilde{A},\tilde{B},\tilde{C}) \in L_{m',n,m}^m$ are (i/o-)equivalent iff

$\exists\alpha$, nice selection, such that

(a) $(A,B,C), (\tilde{A},\tilde{B},\tilde{C}) \in U_\alpha^m$,

(b) $c_\alpha(A,B,C) = c_\alpha(\tilde{A},\tilde{B},\tilde{C})$.

This means that c_α is a continuous canonical form on U_α^m. Because of lemma (4.4-5), each system lies in some U_α^m, so $\{c_\alpha \mid \alpha \text{ nice}\}$ forms a set of local

continuous canonical forms.

The mapping c_α can also be used to obtain local coordinates for $M^m_{m',n,m}$. Lemma (4.4-6) shows that the n-vectors z_j, $j = 1,\ldots,m'+p$, defined by

$$(4.4\text{-}13)\quad z_j := \begin{cases} R(A,B)_{s(\alpha,j)} & \text{for } j = 1,\ldots,m', \\ C_i = C^T e_i, \text{ the } i^{th} \text{column of } C^T, & \text{for } j = m'+i,\ i = 1,\ldots,m, \end{cases}$$

can together serve as a coordinate vector in $\mathbb{R}^{n(m'+m)}$. The choice of $z = \{z_j\}_{j=1}^{m'+m} \in (\mathbb{R}^n)^{m'+m}$ is not completely free, because we require (A,B,C) to be observable, i.e. rk Q(A,C) = n has to hold. However, for an open subset of $(\mathbb{R}^n)^{m'+m}$ this is indeed the case. Let $V^m_\alpha \subseteq (\mathbb{R}^n)^{m'+m}$ denote the set of all admissible choices, and let $(A_\alpha(z), B_\alpha(z), C_\alpha(z))$ denote for all $z \in V^m_\alpha$ the corresponding triple of matrices in W^m_α.

Let $V'^m_\alpha \subseteq M^m_{m',n,m}$ denote the corresponding n(m'+m)-dimensional open subset of $M^m_{m',n,m}$. This notation will be used in the following subsections.

4.5. The differentiable manifold structure of $M^m_{m',n,m}$, $M^{m,a}_{m',n,m}$ and $M^{m,a,f}_{m',n,m}$

The following theorem is due to [Cla].

4.5-1. Theorem. $M^m_{m',n,m}$ is a real analytic manifold of dimension n(m'+m).

This is a well-known result. For a proof one can consult e.g. [Haz 77], theorem 2.5.17, where a direct proof is given, without using an embedding into a space of Hankel matrices. In fact, $M^m_{m',n,m}$ is even known to be an algebraic manifold, however, we will not go into that here.

Notation. Let $M^{m,a}_{m',n,m}$ denote the subset of $M^m_{m',n,m}$ of asymptotically stable systems; i.e.

$$M^{m,a}_{m',n,m} = \{[(A,B,C)] \in M^m_{m',n,m} \mid \forall z \in \sigma(A): |z| < 1\} .$$

where $\sigma(A)$ denotes the spectrum of A. Note that $\sigma(A) = \sigma(TAT^{-1})$, $\forall T \in Gl_n$, therefore it is well-defined on $M^m_{m',n,m}$.

4.5-2. Theorem. $M^{m,a}_{m',n,m}$ is a real analytic manifold of dimension n(m'+m).

Proof. $M^{m,a}_{m',n,m}$ is a subset of $M^{m}_{m',n,m}$. If we can show that for each nice selection α, $M^{m,a}_{m',n,m} \cap V'^{m}_{\alpha}$ is a nonempty open subset of $M^{m}_{m',n,m} \cap V'^{m}_{\alpha}$, then the result follows. An element of V'^{m}_{α} with coordinates $x \in V^{m}_{\alpha}$ can be represented by the triple of matrices $(A_\alpha(x),B_\alpha(x),C_\alpha(x)) \in W^{m}_{\alpha}$, and this matrix triple depends polynomially on x. The condition of asymptotic stability can be written in the form of polynomial inequalities in the entries of $A_\alpha(x)$, and therefore as polynomial inequalities in terms of the components of x. (To be more specific, the inequalities can be derived from the following:

(a) $\det(A+I) \neq 0$

(b) $\det[(z-1)A+(z+1)I]$ must be a Hurwitz polynomial in z, i.e. it must have all its roots in the open left half plane. This implies that certain polynomial inequalities in terms of the coefficients of the polynomial must hold (cf. e.g. [Gan], vol II, Ch.XV; section 6, esp. (36') on p. 195). Here it is used that $z = (w-1)/(w+1)$ maps the open unit disk to the open left half plane in $\mathbb{C}$ and $w = -(z+1)/(z-1)$ is the inverse mapping).

This implies that the set of points x for which these inequalities hold, is open. Therefore each intersection $M^{m,a}_{m',n,m} \cap V'^{m}_{\alpha}$ is open. That each such set is nonempty can be shown easily as follows. If A is multiplied by a nonzero scalar λ then, if $(A,B,C) \in V'^{m}_{\alpha}$, the same holds for $(\lambda A,B,C)$. By choosing $0 < \lambda,\lambda$ small enough an asymptotically stable system is obtained. Such a system is in $M^{m,a}_{m',n,m} \cap V'^{m}_{\alpha}$. Because V'^{m}_{α} is nonempty for each nice selection α, the result follows.

Q.E.D.

Next let us consider the subspace of asymptotically stable, asymptotically stably invertible system:

4.5-3. Definition. A square discrete time linear system $[(A,B,C)] \in M^{m}_{m,n,m}$ with $D = I_m$ is called asymptotically stably invertible if it has all its zeroes in the open unit disk, i.e.

$$(4.5\text{-}4) \quad \{z \in \mathbb{C} \mid \det(C(zI-A)^{-1}B+I) = 0\} \subseteq \{z \in \mathbb{C}: |z| < 1\}$$

4.5-5. Notation: $M^{m,a,f}_{m,n,m}$ is the subset of asymptotically stably invertible systems in $M^{m,a}_{m,n,m}$.

4.5-6. Theorem. $M^{m,a,f}_{m,n,m}$ is a real analytic manifold of dimension 2mn.

Proof. The proof is similar to that of 4.5-2. The main difference is that the inequalities are now different. In fact, assuming (A,B,C) is minimal and asymptotically stable, for the system to be asymptotically stably invertible, the polynomial $p(w) = \{\det(wI-A)\}^m \det[C(wI-A)^{-1}B + I]$ must have all its zeroes within the open unit disk. Applying the transformation $w = -(z+1)/(z-1)$ and multiplying with $(z-1)^{m^2-m}$, we obtain the polynomial

$$\{\det((z+1)I+(z-1)A)\}^m \det(C((z+1)I+(z-1)A)^{-1}B+I).$$

This must be Hurwitz. As before this implies that certain polynomial inequalities in the local coordinates of $M^{m,a}_{m,n,m} \cap V'^m_\alpha$ must hold. This shows that $M^{m,a,f}_{m,n,m}$ is an open subset of $M^{m,a}_{m,n,m}$. For each α, $V'^m_\alpha \cap M^{m,a,f}_{m,n,m}$ is nonempty as can easily be seen by considering $(\lambda A, B, \lambda C)$. The corresponding polynomial is

$$\begin{aligned} p_\lambda(w) &= \{\det(wI-\lambda A)\}^m \det(\lambda C(wI-\lambda A)^{-1}B+I) = \\ &= \lambda^{m^2} \{\det((w\lambda^{-1})I-A)\}^m \det(C(w\lambda^{-1}I-A)^{-1}B+I) = \\ &= \lambda^{m^2} p(w\lambda^{-1}). \end{aligned}$$

It is clear that for $\lambda^{-1} > 0$ large enough $p(w\lambda^{-1})$ has its zeroes for values of w within the open unit disk. The same holds for $\det((w\lambda^{-1})I-A)$, i.e. for the poles. So in each V'^m_α there are points of $M^{m,a,f}_{m,n,m}$.

Q.E.D.

4.6. The state vector bundle and the associated principal fibre bundle

In section 3 we have tried to make clear that one should expect a fibred structure in a family of systems with states, with the state space as the fibre. For linear systems this will presently be treated. To be a fibre bundle (in the sense of [Ste]) the fibres should be homeomorphic. Therefore they must have the same dimension. This explains why we have to fix the McMillan degree (which is the minimal state space dimension of a system) if we want to obtain a fibre bundle structure. This restriction is also sufficient. Recall our notation from section 4: $M^m_{l,m',n,m}$ is the set of 'systems with states', with

m' inputs, m outputs and minimal state space dimension n.

4.6-1. Theorem. $\{M^m_{l,m',n,m}, M^m_{m',n,m}, \pi, \mathbb{R}^n, \{(V_\alpha^{'m}, \bar{\Phi}_\alpha)\}, G\ell_n(\mathbb{R})\}$ is a coordinate bundle and its equivalence class (in the sense of section 4.2) is a fibre bundle, in fact a vector bundle. Here π is the projection

$\pi: M^m_{l,m',n,m} \to M^m_{m',n,m}$ which simply 'forgets' the state. The multi-indices α are the nice selections. The $\bar{\Phi}_\alpha$ are mappings

$$\bar{\Phi}_\alpha: V_\alpha^{'m} \times \mathbb{R}^n \to \Pi^{-1}(V_\alpha^{'m})$$

given by $\bar{\Phi}_\alpha([(A_\alpha(z),B_\alpha(z),C_\alpha(z))],y) = [(A_\alpha(z),B_\alpha(z),C_\alpha(z),y)]$

Proof. Simply check the definition. This is left to the reader.

4.6-2. Remark: In fact the state bundle is a real analytic fibre bundle, i.e. $M^m_{l,m',n,m}$ and $M^m_{m',n,m}$ are real analytic manifolds, π is a real analytic mapping, the ϕ_j are real analytic, the $(\bar{\phi}_{jz})^{-1}\bar{\phi}_{iz}: \mathbb{R}^n \to \mathbb{R}^n$ are linear, hence real analytic, and the mappings $g_{\alpha\beta}: V_\alpha^{'m} \cap V_\beta^{'m} \to G\ell_n(\mathbb{R})$, $z \longmapsto (\bar{\phi}_{jz})^{-1}\bar{\phi}_{iz}$ are real analytic. (Notice that in comparison with section 2, here we use z instead of x).

4.6-3. Remark. The vector bundle of theorem 4.6-1 will be called the state bundle over $M^m_{m',n,m}$. (This terminology is taken from [Del 82], cf. also [Haz 77]).

4.6-4. Notation. To shorten notation we will denote the state bundle over $M^m_{m',n,m}$ by $\pi: M^m_{l,m',n,m} \to M^m_{m',n,m}$ and sometimes simply by $M^m_{l,m',n,m}$, if it is clear from the context what is meant exactly.

4.6-5. Theorem. Let $\beta: L^m_{m',n,m} \to M^m_{m',n,m}$ be given by $\beta(A,B,C) = [(A,B,C)]$. Then $\beta: L^m_{m',n,m} \to M^m_{m',n,m}$ is the associated principal fibre bundle of the state bundle $\pi: M^m_{l,m',n,m} \to M^m_{m',n,m}$.

Proof. [Haz 77], theorem 2.6.6 and remark 4.6; cf.also [Del 82], pp. 18-19.

From the definition of coordinate bundle it is clear that if one restricts the base space to a nonempty open subset, then one obtains another coordinate bundle. Its associated principal bundle is found by restricting the base space of the principal bundle of the original coordinate bundle to the same open subset. Therefore

4.6-6. Corollary. $L^{m,a}_{m',n,m}$ is the associated principal fibre bundle of the state bundle π: $M^{m,a}_{1,m',n,m} \to M^{m,a}_{m',n,m}$.

4.6-7. Remark. A similar result holds for any other nonempty open subset of $M^{m}_{m',n,m}$. In fact it also holds for any submanifold of $M^{m}_{m',n,m}$ (cf. remark 4.2-19)

4.7. On the topological structure of some spaces of linear systems

The first important fact about the topological structure of $M^{m}_{m',n,m}$ is that it is not simply connected (and so it is not homeomorphic to a Euclidean space). Since this became clear by work of Brockett, Glover, Kalman, Hazewinkel and others, the topological structure of these and related spaces has been investigated further (and this research is still going on). To mention a few papers and authors [Haz 77], [Kr-B 80], [Del 82], [Bro 76], [By-D], [By-H], [Glo 73], [Glo 75], [Kri 77], [He], [Seg].
We mention some of the results.

4.7-1. Theorem. [Brockett] $M^{m}_{1,n,1}$ has n+1 connected components.

Proof. cf. [Bro 76].

4.7-2. Remarks. (i) In each of these components the Cauchy index of the transfer functions of the systems in $M^{m}_{1,n,1}$ has a different constant value. Alternatively one can say that the signature of the corresponding Hankel matrices has a different constant value in each component. Because the Hankel matrix has rank n, there are n+1 possibilities for the signature (cf. [Bro 76]).
(ii) Some of these components will be globally diffeomorphic to some Euclidean

space, others, however, are not. [Kri 77] mentions for example that $M^m_{1,2,1}$ has two simply connected components and one component that is homeomorphic to

$S^1 \times \mathbb{R}^3$ (S^1 = circle).

(iii) In this and related results it is essential that one works over $\mathbb{R}$ and not over $\mathbb{C}$.

4.7-3. Theorem. [Glover]. $M^m_{m',n,m}$ is connected if $m' > 1$ or $m > 1$.

Proof. cf. [Glo 73], or [By-D], p. 67.

4.7-4. Theorem. $M^{m,a}_{m',n,m}$ is connected when $\max(m',m) > 1$. It has $n+1$ components when $m' = m = 1$.

Proof. We will use the property of arcwise connectedness, which is equivalent to connectedness for (topological) manifolds (see e.g. [Hu], chapter III, section 5, especially (5.1), (5.9), (5.18) and the remark following (5.20)). If two systems $\Sigma_0, \Sigma_1 \in M^{m,a}_{m',n,m}$ are in the same connected component of $M^m_{m',n,m}$, there exists a continuous curve $\Sigma(t)$, $t \in [0,1]$ with $\Sigma(0) = \Sigma_0$ and $\Sigma(1) = \Sigma_1$ and $\forall t \in [0,1]$: $\Sigma(t) \in M^{m,a}_{m',n,m}$. Consider

(4.7-5) $\quad \sup\{|x|: x \in \sigma(\Sigma(t)),\ t \in [0,1]\}$,

where $\sigma(\Sigma)$ denotes the set of poles of the system Σ. Because $t \longmapsto \max\{|x|: x \in \sigma(\Sigma(t))\}$ is a continuous function of $t \in [0,1]$, it follows that the supremum of (4.75) is attained at some $t_0 \in [0,1]$. Choose $\lambda_0 > 0$ such that λ_0^{-1} is larger than this maximum. Consider the curve $\Sigma'(t)$ defined as follows:

(4.7-6)

$$\begin{cases} \forall t \in [0,\frac{1}{3}]: \Sigma'(t) := (\lambda(t)A_0, B_0, C_0) \text{ with } \lambda(t) = 1-3t(1-\lambda_0) = (1-3t)+3t\lambda_0, \\ \forall t \in (\frac{1}{3},\frac{2}{3}]: \Sigma'(t) := (\lambda_0 A(3t-1), B(3t-1), C(3t-1)), \text{ with } (A(t),B(t),C(t)) \\ \qquad\qquad \text{a representation of } \Sigma(t), \\ \forall t \in (\frac{2}{3},1]: \Sigma'(t) = (\lambda(1-t)A_1, B_1, C_1). \end{cases}$$

Then $\forall t \in [0,1]$: $\Sigma'(t) \in M^{m,a}_{m',n,m}$ and $\Sigma'(t)$ is a continuous arc. It follows that $\Sigma_o, \Sigma_1 \in M^{m,a}_{m',n,m}$ are in the same connected component in $M^{m,a}_{m',n,m}$. It follows that (i) the number of connected components of $M^{m,a}_{m',n,m}$ is smaller than or equal to the number of connected components of $M^{m}_{m',n,m}$. Also, each connected component of $M^{m}_{m',n,m}$ contains an element of $M^{m,a}_{m',n,m}$. This can be seen by considering

(4.7-7) $\forall t \in [0,1] : \Sigma''_o(t) = (\tilde{\lambda}(t)A_o, B_o, C_o)$, $\tilde{\lambda}(t) = 1-t(1-\lambda_o)$,

$\lambda_o > 0$ small enough such that $\lambda_o A_o$ is asymptotically stable. It is clear that if (A_o,B_o,C_o) is a <u>minimal</u> realization, then the same holds for $(\tilde{\lambda}(t)A_o,B_o,C_o)$, $\forall t \in [0,1]$, because $\tilde{\lambda}(t) \neq 0$ on $[0,1]$. It follows that $(\lambda_o A_o,B_o,C_o)$ is in the same connected component as (A_o,B_o,C_o). Because $(A_o,B_o,C_o) \in L^{m}_{m,n,p}$ is chosen arbitrarily, it follows that each connected component of $M^{m}_{m',n,m}$ contains an element of $M^{m,a}_{m',n,m}$. This implies that (ii) the number of connected components of $M^{m,a}_{m',n,m}$ is larger than or equal to the corresponding number of $M^{m}_{m',n,m}$. (i) and (ii) together imply that the number of connected components of $M^{m,a}_{m',n,m}$ is the same as the number of connected components of $M^{m}_{m',n,m}$.

Q.E.D.

4.7-8. <u>Theorem</u>. If m'=m=1 or if m'>1 <u>and</u> m>1, $M^{m}_{m',n,m}$ is not homeomorphic to a Euclidean space)*.

Proof. For the s.i.s.o. case this follows directly from theorem 4.7.1. For the m.i.m.o case this follows from the fact that the first homotopy group of $M^{m}_{m',n,m}$ is nonzero. (see e.g. [Haz 77], section 3.5.7 and further).

Q.E.D.

4.7-9. <u>Theorem</u>. $M^{m}_{m',n,m}$ is noncompact.

)* For the m.i.s.o. and s.i.m.o. case, cf. [Hnz 88a].

Proof. A simple (but somewhat deceiving) proof is the following: Let Σ be a system with a minimal representation (A,B,C). Then $\Lambda := \{\lambda\Sigma := \Sigma(A,\lambda B,C) \mid \lambda > 0\} \subseteq M^m_{m',n,m}$, and Λ is clearly homeomorphic to the real line. Furthermore $\bar{\Lambda} = \Lambda$ clearly, and therefore from the noncompactness of $\mathbb{R}$, the noncompactness of $M^m_{m',n,m}$ follows. Q.E.D.

4.7-10. Remark. This proof is somewhat deceiving, because even if one restricts the matrices B and C to lie in some compact set, noncompactness can still hold.
Although much more is known about the topology of these spaces, we will not go into further details. The properties mentioned here are the ones that are the most important for our purposes.

4.8. Families of stochastic linear systems

In this section we will consider finite dimensional linear systems with Gaussian white noise inputs, i.e. systems of the following form

$$(4.8\text{-}1)\quad \begin{cases} x_{t+1} = Ax_t + B\,v_t, & t \in \mathbb{Z},\ x_t \in \mathbb{R}^n, \\ y_t = Cx_t + D\,v_t, & y_t \in \mathbb{R}^m, \end{cases}$$

where $v_t \in \mathbb{R}^{m'}$ is Gaussian white noise, $Ev_t = 0$, $Ev_t v_{t'}^T = \bar{\Omega}\,\delta_{tt'}$, $\bar{\Omega}$ positive definite. We assume the system is ergodic and stationary. Therefore, as was shown in chapter 2, assuming (A,B) is reachable and (C,A) observable, A must be asymptotically stable. (The following definition is the same as in chapter 2).

4.8-2. Definition. Two stochastic systems with (unobservable) random inputs are called equivalent, if they generate the same probability measure on the outputs.

4.8-3. Definition. The innovations process of a stochastic system (4.8.1) is defined by

$$w_t = y_t - \hat{y}_{t|t-1},$$

where $\hat{y}_{t|t-1} := E\{y_t \mid \sigma(\{y_{t-k}, k=1,2,\ldots\})\}$, and $\sigma(\{y_{t-k},\ k = 1,2,\ldots\})$ the sigma algebra generated by $y_{t-1}, y_{t-2}, y_{t-3}, \ldots$.

Remark: $\hat{y}_{t|t-1}$ is well-defined (see e.g. [Bau], 56.5 and 56.7)

4.8-4. Definition. A model of the form

$$(4.8\text{-}5)\quad \begin{cases} x_{t+1} = Ax_t + Bw_t, \\ y_t = Cx_t + w_t, \end{cases}$$

where $\{w_t\}$ are the innovations, $Ew_t w_t^T = \overline{\Omega}$, is called an innovations representation.

Notation: $(A,B,C,\overline{\Omega})$.

4.8-6. Theorem. (i) An innovations representation exists for each system of the form (4.8-1).
(ii) The transfer function of the innovation representation is uniquely determined.
Proof. (i) See [An-M], chapter 9, especially section 9.4
(ii) The transfer function is $I+C(zI-A)^{-1}B$. The corresponding Hankel matrix is completely determined by the Markov matrices
$Ey_t w_{t-s}^T (Ew_{t-s} w_{t-s}^T)^{-1}$, $s = 0,1,2,\ldots$

Q.E.D.

From now on we will assume an innovations representation to be minimal (i.e. reachable and observable) except if stated explicitly otherwise.

4.8-7. Theorem. A model of the form (4.8-5), (without the explicit requirement that $\{w_t\}$ are the innovations) is an innovations representation iff
$\sigma(A-BC) \subseteq \overline{D(0,1)} \subseteq \mathbb{C}$, $\overline{D(0,1)}$ = the closed unit disk in $\mathbb{C}$).

Proof: cf. [An-M], theorem 4.4.

4.8-8. Theorem. The set of all stochastic systems (4.8-5) with fixed McMillan degree n and fixed number m of output components, which have asymptotically stably invertible innovations representation forms a differentiable manifold diffeomorphic to $M_{m,n,m}^{m,a,f} \times \mathrm{Pos}(m)$, where Pos(m) denotes the manifold of all m×m positive definite matrices.

Proof. The diffeomorphism is given by $[(A,B,C,\bar{\Omega})] \longmapsto ([(A,B,C)],\bar{\Omega})$, $[(A,B,C)] \in M^{m,a,f}_{m,n,m}$; $\bar{\Omega} \in \text{Pos}(m)$. The square brackets stand for 'equivalence class'. $M^{m,a,f}_{m,n,m}$ is a differentiable manifold (proved above), and Pos(p) is a differentiable manifold (standard), therefore the Cartesian product $M^{m,a,f}_{m,n,m} \times \text{Pos}(p)$ is a differentiable manifold.

Q.E.D.

4.8-9. Remark. It follows that the set of stochastic systems of the form (4.8-5) with fixed n,p under the equivalence (4.8-2) forms the Cartesian product of the closure of the submanifold $M^{m,a,f}_{m,n,m}$ of $M^{m,a}_{m,n,m}$ with Pos(m). $\partial M^{m,a,f}_{m,n,m} \times \text{Pos}(m)$ is given by all systems which have $\sigma(A-BC) \cap \{z \in \mathbb{C}: |z| = 1\} \neq 0$, in other words all systems of which the innovations representation is not asymptotically stably invertible. This is the same as saying that the system's steady state Kalman filter is not asymptotically stable. (see [An-M], section 9.4).
It follows that both to be able to work with manifolds and to avoid non asymptotically stable Kalman filters, one has to exclude innovations models which are not asymptotically stably invertible.

4.8-10. Remark. One can give a vector bundle structure to the set of asymptotically stably invertible stochastic systems, with state, with fixed n and p, in an obvious manner (compare section 4.6, esp. (4.6-7)).

CHAPTER 5

RIEMANNIAN GEOMETRY AND FAMILIES OF LINEAR SYSTEMS

5.1. Introduction

For many purposes it is desirable to have a 'natural' metric defined on the model space, which tells us quantitatively how close or far apart, two models are. Especially, in recursive identification if one wants to adapt the model according to the latest data, it is important to have a measure of how large the change is that is going to be made. In particular this plays a role in gradient type algorithm, for which we refer to the next chapter. Another advantage of having a metric, is that one can investigate the geometry of the model space using this metric, and gain insight in the structure of the problem.

If the model space is embedded in a Euclidean space in a natural way, then the metric of the containing Euclidean space induces a metric on the model space, simply by restriction. More generally, the same holds if the model space is embedded in a Hilbert space in a natural way. Under certain conditions there also exists in these cases an induced 'infimum path length metric', i.e. the metric which declares the distance between two points x and y to be the infimum of the lengths of all paths lying completely within the model space, which connect the two points. So intuitively, this is the 'travelling distance' between x and y, if one travels within the model set. A necessary condition for the existence of such a metric (if we do not allow infinite distances) is that the model set is connected (in the induced topology). This metric is called the inner metric ('innere Metrik', cf. [Rin]).

The metrics that will be presented will be constructed in this manner. It has the advantage that one can start with a 'natural' Hilbert space in which the model space is embedded, which makes the interpretation both of the induced metric and the corresponding inner metric, much easier. Furthermore, the metric obtained in this way is completely independent of the parametrization. Therefore it also very well suited for overlapping parametrizations methods.

In most cases it is not possible to obtain the inner metric in closed form. Instead it will be given by a Riemannian metric i.e. a metric on each tangent space to the model set. This Riemannian metric will be expressed in terms of local coordinates or, in other words, in the parameters of a local parametrization. It is not difficult to show that if the Riemannian metric

tensor in terms of the parameters is positive definite, i.e. nonsingular, then the parametrization is locally identifiable. (Cf. section 5.11).
The idea of using Riemannian geometry in system identification is not new. In the stochastic s.i.s.o. case, P.S. Krishnaprasad [Kri 77] has used the Fisher metric, cf. section 5.10; D.F. Delchamps [Del 82] used Riemannian geometry on the state bundle to obtain results about the geometry of the manifold of deterministic m.i.m.o.-systems. In statistics we would like to mention the work of Shun-ichi Amari, cf [Ama 82], [Ama 85], [Ama 86]. For our own previous work, we refer to [Hnz-H], [Hnz-M]. (An important part of the research for this chapter was conducted while the author was visiting the University of Texas at Austin). The contents of this chapter will be as follows. First we will treat the deterministic case in section 5.2, both for continuous and discrete time. In section 5.3, we treat a generalization which shows how the discrete time case and the continuous time case are related. Here we use a new family of difference/differential operators, which has a number of nice properties. The analysis of this family leads to some interesting isometries, which are treated in section 5.4. In section 5.5 we treat some simple cases, among which there is the siso-case with McMillan degree 1. The geodesics can be computed in this case and the inner metric is obtained explicitly. In section 5.6 some remarks are made about the behaviour of the Riemannian metric tensor at points in which the McMillan degree drops, in relation to the parametrization that is used. In section 5.7 we define a fibre metric on the state bundle and combine this with the Riemannian metric on the model space to obtain a Riemannian metric on the whole state bundle. (This section may be skipped by the reader).
In subsection 5.8, the stochastic case is treated, for which there are several possibilities. To compare these with the well-known Fisher metric, we start in section 5.9 to derive the Fisher metric from the socalled Hellinger distance between any pair of probability measures on the same set. The relation between Hellinger distance and Fisher metric is very interesting. It makes it possible to use the same set-up as before, i.e. first embed the model space in a Hilbert space and then derive a Riemannian metric (and a corresponding inner metric). In section 5.10 the Fisher metric for linear stochastic systems is derived. In section 5.11 we treat the relationships between local identifiability and Riemannian metrics. This generalizes some well-known results on the relation between local identifiability and the Fisher information matrix.

5.2. A Riemannian metric for families of stable deterministic systems

In this chapter we will not only consider discrete time models, but also continuous time models. The reason is that the results are tightly related and even if one is mainly interested in the discrete time case, it is still worthwhile to develop the continuous time case. (In section 5.3 it will be shown how the continuous time case can be considered as a limiting case of the discrete time case with small sampling interval).

The models will be as follows:

$$(5.2\text{-}1)\quad \begin{cases} x_{t+1} = Ax_t + Bu_t, & x_t \in \mathbb{R}^n,\ u_t \in \mathbb{R}^{m'},\ y_t \in \mathbb{R}^m \\ y_t = Cx_t + Du_t, & A \in \mathbb{R}^{n\times n},\ B \in \mathbb{R}^{n\times m'},\ C \in \mathbb{R}^{m\times n},\ D \in \mathbb{R}^{m\times m'} \end{cases}$$

in the discrete time case, and

$$(5.2\text{-}2)\quad \begin{cases} \dot{x}_t = Ax_t + Bu_t, \\ y_t = Cx_t + Du_t, \end{cases} \text{ same specification as in (5.2-1),}$$

in the continuous time case. In this case one needs u_t to be integrable with respect to t on each finite interval. We will consider the state variable x_t as not (directly) observable. Only the outputs $\{y_t\}$ and the inputs $\{u_t\}$ are observable. Therefore, as in the set-up of chapter 2, we will consider two quadruples (A,B,C,D) and $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ to be equivalent if they induce the same input-output map

$$(5.2\text{-}3)\quad \begin{aligned} &F: \overset{\circ}{U}_o \to \overset{\circ}{Y}, \\ &\{u_t\} \longmapsto \{y_t\}, \end{aligned}$$

where $\overset{\circ}{U}_o$ = {all input functions u_t such that $\exists t_o$ with $u_t = 0, \forall t < t_o$}.

In the discrete time case F is given by the formula

$$(5.2\text{-}4)\quad y_t = \sum_{k=o}^{\infty} H_k u_{t-k},$$

with H_k the k^{th} Markov matrix (cf. section 2.3.2).

If there is only a finite number q+1 of Markov matrices H_k that do not vanish, i.e. the pure moving average ma(q) case, then it appears to be natural to take as the parameter space the (q+1) × m × m'-dimensional Euclidean space with elements $(H_o,H_1,\ldots,H_q)$. The corresponding Euclidean norm of the system Σ with Markov matrices $(H_o,H_1,\ldots,H_q)$, is

$$(5.2\text{-}5) \qquad \|\Sigma\| = \Big(\operatorname{tr} \sum_{k=o}^{q} H_k H_k^T\Big)^{\frac{1}{2}}.$$

The question arises when this can be extended to the case 'q = ∞'. The answer is: if the model is asymptotically stable. Therefore we will restrict ourselves (mainly) to the asymptotically stable case. The same holds for the continuous time case.

The set of asymptotically stable finite dimensional linear systems with fixed input and output dimensions forms a linear space with the usual addition and scalar multiplication of the input-output maps. In terms of Markov matrices, the sum of two systems, Σ resp $\tilde{\Sigma}$, corresponding to Markov matrix-sequences $(H_o, H_1, \ldots)$ resp. $(\tilde{H}_o, \tilde{H}_1, \ldots)$, is given by the Markov matrix sequence $(H_o+\tilde{H}_o, H_1+\tilde{H}_1, H_2+\tilde{H}_2, \ldots)$. In terms of corresponding matrix quadruples (A,B,C,D) resp. $(\tilde{A},\tilde{B},\tilde{C},\tilde{D})$ the sum of the two systems is given by the matrix quadruple

$$(5.2\text{-}6) \qquad \left[\begin{bmatrix} A & 0 \\ 0 & \tilde{A} \end{bmatrix}, \begin{bmatrix} B \\ \tilde{B} \end{bmatrix}, (C, \tilde{C}), D + \tilde{D}\right].$$

From this formula it is clear that the McMillan degree of the sum system is finite and that the sum system is asymptotically stable. (This holds both for discrete and continuous time).

The linear space of finite dimensional asymptotically stable linear discrete time systems can be regarded as a subspace of the Hilbert space of all square summable sequences of Markov matrices $\{H_k\}_{k=o}^{\infty}$ with inner product

$$(5.2\text{-}7) \qquad \langle \{H_k\}, \{\tilde{H}_k\} \rangle = \operatorname{tr} \sum_{k=o}^{\infty} H_k \tilde{H}_k^T.$$

This is a generalization of the moving average case discussed above. In terms of (A,B,C,D), A asymptotically stable, the formula for the norm of this system Σ in the Hilbert space is

$$\begin{aligned} \|\Sigma\|^2 &= \operatorname{tr}\Big(\sum_{k=o}^{\infty} CA^k BB^T (A^T)^k C^T + DD^T\Big) \\ &= \operatorname{tr}(CL(BB^T)C^T + DD^T) \end{aligned} \qquad (5.2\text{-}8)$$

where L(K) is the solution of the discrete time Lyapunov equation

(5.2-9) $L - ALA^T = K.$

In the continuous time case we associate with an asymptotically stable system Σ, with matrix quadruple (A,B,C,D), the matrix function $G(\sigma) := C \exp(\sigma A)B$, $\sigma \in [0,\infty)$, together with the matrix D. This pair $(G(\sigma),D)$ lies in the Hilbert space of pairs $(G(\sigma),D)$, with inner product

(5.2-10) $$< (G,D), (\tilde{G},\tilde{D}) > = \operatorname{tr} \int_{\sigma=0}^{\infty} G(\sigma)\tilde{G}(\sigma)^T d\sigma + \operatorname{tr} D\tilde{D}^T.$$

In terms of its matrix quadruple (A,B,C,D) the norm of a system Σ in continuous time is equal to

(5.2-11) $$\|\Sigma\|^2 = \operatorname{tr} \int_0^{\infty} C \exp(\sigma A)BB^T \exp(\sigma A^T)C^T d\sigma + \operatorname{tr} DD^T = \operatorname{tr}(CM(BB^T)C^T + DD^T),$$

where $M(K)$ is the solution of the continuous time Lyapunov equation:

(5.2-12) $AM + MA^T = -K.$

Using the Parseval relation these norms can also be written in terms of the transfer matrix $T(s) = D + C(sI-A)^{-1}B$, as follows

(5.2-13) $$\|\Sigma\|^2 = \operatorname{Tr} \frac{1}{2\pi i} \int_{C(0,1)} T(s)T(s)^* \frac{ds}{s},$$

in the discrete time case, and

(5.2-14) $$\|\Sigma\|^2 = \operatorname{Tr} \frac{1}{2\pi i} \int_{-i\infty}^{+i\infty} (T(s)-D)(T(s)-D)^* ds + \operatorname{Tr} DD^*,$$

in the continuous time case, cf. [Bro 70], exercise 6, p.63; cf. also section 5.3. Here $C(0,1)$ denotes the unit circle in the complex plane, while $(-i\infty,i\infty)$ denotes the imaginary axis. For simplicity, suppose $D = 0$. Both formulas can be written as a line integral of $\frac{1}{2\pi}\operatorname{tr} T(s)T(s)^*$ over the boundary of the stability region, as follows. In the discrete time case parametrize the unit circle by $s = e^{i\ell}$, $\ell \in [0,2\pi]$ being the arclength. Then (5.2-13) becomes

(5.2-13a) $$\|\Sigma\|^2 = \operatorname{tr} \frac{1}{2\pi} \int_0^{2\pi} T(e^{i\ell})T(e^{i\ell})^* d\ell = \operatorname{tr} \frac{1}{2\pi} \int_{C(0,1)} TT^* d\ell,$$

in a conventional notation for line integrals.
Parametrizing the imaginary axis by $s = i\ell$, $\ell \in \mathbb{R}$ being again the arclength, one obtains a similar formula:

$$(5.2\text{-}14a)\quad \|\Sigma\|^2 = \mathrm{tr}\frac{1}{2\pi}\int_{-\infty}^{\infty} T(i\ell)T(i\ell)^* d\ell = \mathrm{tr}\frac{1}{2\pi}\int_{\text{Im.axis}} TT^* d\ell.$$

In section 5.3.4 we will find a generalization of these formulas to the case of a general time interval length $\Delta \in [0,1]$.

5.2-15. Remark. Of course other inner products are conceivable. To mention a few possibilities (in the discrete time case)
(i) 'The Frobenius norm of the Hankel matrix'. Suppose for simplicity that $H_o = \tilde{H}_o = 0$. Define

$$(5.2\text{-}16)\quad \langle \{H_k\}_{k=1}^{\infty}, \{\tilde{H}_k\}_{k=1}^{\infty} \rangle_{FH} := \mathrm{tr} \sum_{k=1}^{\infty} kH_k\tilde{H}_k^T.$$

The corresponding norm is (if $D = 0$)

$$(5.2\text{-}17)\quad \|\Sigma\|_{FH} = (\mathrm{tr} \sum_{k=1}^{\infty} kH_kH_k^T)^{\frac{1}{2}} = (\mathrm{tr}\overset{\circ}{H}\overset{\circ}{H}^T)^{\frac{1}{2}},$$

so this is the 'Frobenius' norm of the Hankel matrix $\overset{\circ}{H}$ of the system Σ:

$$(5.2\text{-}18)\quad \overset{\circ}{H} = \begin{bmatrix} H_1 & H_2 & H_3 & \cdots \\ H_2 & H_3 & H_4 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}.$$

Note that in this norm the past gets a higher weight then in the norm (5.2-8). This does not seem to be desirable from an applications point of view, because in applications, usually one wants to put more emphasis on the 'near' past than on the 'remote' past. Therefore we prefer (5.2-7)-(5.2-8). It should be noted, that one can find a closed formula for the norm (and the inner product) using the Lyapunov operator L. This can be done as follows: Σ is represented by (A,B,C), (D = 0). Then

$$(5.2\text{-}19)\quad \|\Sigma\|_{FH}^2 = \mathrm{tr} \sum_{k=1}^{\infty} kCA^{k-1}BB^T(A^T)^{k-1}C^T = \mathrm{tr}\, C\{\sum_{k=1}^{\infty} kA^{k-1}BB^T(A^T)^{k-1}\}C^T.$$

The expression within parentheses can be written as

$$\sum_{k=1}^{\infty} kA^{k-1}BB^T(A^T)^{k-1} = \sum_{\ell=o}^{\infty} A^{\ell}L(BB^T)(A^T)^{\ell} = L[L(BB^T)] = L \circ L\,(BB^T).$$

So

$$(5.2\text{-}20)\quad \|\Sigma\|_{FH}^2 = \mathrm{tr}C.L \circ L(BB^T).C^T.$$

(ii) If one wants to discount the past in the inner product, one can take e.g. (again, take $H_o = \tilde{H}_o = 0$).

$$(5.2\text{-}21)\quad \langle\{H_k\}_{k=1}^{\infty}, \{\tilde{H}_k\}_{k=1}^{\infty}\rangle_e := \mathrm{tr}\sum_{k=1}^{\infty}\frac{1}{(k-1)!}H_k\tilde{H}_k^T.$$

Using (A,B,C) one obtains for the norm of the system Σ:

$$(5.2\text{-}22)\quad \|\Sigma\|_e^2 = \mathrm{tr}C\Big(\sum_{k=1}^{\infty}\frac{1}{(k-1)!}A^{k-1}BB^T(A^T)^{k-1}\Big)C^T.$$

The expression within parenthesis can be evaluated as follows: Let $\mathring{A}$ denote the linear operator that maps BB^T into ABB^TA^T. Then the expression within parenthesis is

$$(5.2\text{-}23)\quad \sum_{k=1}^{\infty}\frac{1}{(k-1)!}\mathring{A}^{k-1}(BB^T) = e^{\mathring{A}}(BB^T).$$

So in this case the norm would be given by

$$(5.2\text{-}24)\quad \|\Sigma\|_e^2 = \mathrm{tr}\ C.e^{\mathring{A}}(BB^T).C^T.$$

The advantage of such a formula is that it strongly discounts the past, and even in such a way that the norm applies not only to asymptotically stable systems, but to <u>all</u> systems. The disadvantage is, however, that it seems computationally more difficult.

(iii) A third possibility is the following. Use the inner product, in the complex case,

$$(5.2\text{-}25)\quad \langle\{H_k\}_{k=1}^{\infty}, \{\tilde{H}_k\}_{k=1}^{\infty}\rangle_p := \mathrm{tr}\sum_{k=1}^{\infty}\frac{1}{k^2}H_k\tilde{H}_k^*$$

where * denotes transposition plus complex conjugation. Note that in this

case, if $\{H_k\}$ is bounded, the norm is finite. I.e. if the corresponding A matrix is stable (i.e. $\{A^k\}_{k=1}^{\infty}$ is bounded), (and not necessarily asymptotically stable), the norm is well-defined. Consider the special case $(A,B,C) = (a,1,1)$, $a \in \mathbb{C}$, $|a| \leq 1$. Then $\{H_k\} = (a^k)_{k=1}^{\infty}$. The norm of this system is

$$(5.2\text{-}26) \quad \|(a^k)_{k=1}^{\infty}\|_P = \left[\sum_{k=1}^{\infty} \frac{|a|^{2k}}{k^2} \right]^{\frac{1}{2}}.$$

If $|a| > 1$ this is infinite, if $|a| \leq 1$ it is finite. In fact, if $|a| = 1$, one has

$$(5.2\text{-}27) \quad \|(a^k)_{k=1}^{\infty}\|_P = \frac{1}{6}(\sqrt{6})\pi.$$

This norm has the same disadvantage as the previous one: it cannot be computed easily. It is mentioned here for two reasons. One is to show that one can obtain a metric on the set of stable system (as opposed to the (sub)set of asymptotically stable systems). The second reason is that the corresponding Riemannian metric has a very nice and well-known structure in the subset of asymptotically stable systems. We will return to this later on, in section 5.5.

The question arises how one can compute the Riemannian metrics that are induced on a manifold of systems with fixed McMillan degree by a given Hilbert space structure. The Riemannian metric can always be computed if the inner product in the Hilbert space can be computed. This will be treated next. First we have to introduce the concept of the derivative of a system with respect to 'a change in the system'.

Let $\overset{\circ}{M}{}^{m,a}_{m',n,m} = M^{m,a}_{m',n,m} \times \mathbb{R}^{m \times m'}$ (in the real case), $M^{m,a}_{m',n,m}$ is as in chapter 4. The difference between $\overset{\circ}{M}{}^{m,a}_{m',n,m}$ and $M^{m,a}_{m',n,m}$ is that the first space contains proper systems, while the second space contains strictly proper systems ("D=0") or alternatively proper systems with $D \neq 0$ fixed.

Consider a curve in $\overset{\circ}{M}{}^{m,a}_{m',n,m}$:

$$(5.2\text{-}28) \quad \begin{array}{l} [0,1] \to \overset{\circ}{M}{}^{m,a}_{m',n,m}, \\ s \longmapsto \Sigma(s), \end{array}$$

where $\Sigma(s)$ is the system that can be represented by $(A(s),B(s),C(s),D(s))$ or,

alternatively by the sequence of Markov matrices $\{H_k(s)\}_{k=o}^{\infty}$. Using the embedding, this can also be considered as a curve in Hilbert space. Suppose that $(A(s),B(s),C(s),D(s))$ is differentiable in $s = s_o$. Then what is the derivative in Hilbert space? Of course, it is

$$(5.2\text{-}29)\quad \lim_{s\to s_o} \frac{1}{s-s_o}(\Sigma(s)-\Sigma(s_o)).$$

How this can be expressed in terms of $A(s_o),B(s_o),C(s_o),D(s_o)$ and $\dot{A}(s_o),\dot{B}(s_o),\dot{C}(s_o),\dot{D}(s_o)$ is given by the following lemma.

5.2-30. Lemma. $\lim_{s\to s_o} \frac{1}{s-s_o}(\Sigma(s)-\Sigma(s_o))$ is itself an asymptotically stable linear system, of McMillan degree smaller than or equal to 2n. It is the system that is represented by

$$\left[\begin{bmatrix} A(s_o) & 0 \\ \dot{A}(s_o) & A(s_o) \end{bmatrix}, \begin{bmatrix} B(s_o) \\ \dot{B}(s_o) \end{bmatrix}, (\dot{C}(s_o),C(s_o)), \dot{D}(s_o)\right].$$

This representation may be nonminimal. The system will be denoted by $\dot{\Sigma}(s_o)$.

Proof. First consider the system $\frac{1}{s-s_o}(\Sigma(s)-\Sigma(s_o))$.

It is represented by (see (5.2-6))

$$(5.2\text{-}31)\quad \left[\begin{bmatrix} A(s) & 0 \\ 0 & A(s_o) \end{bmatrix}, \begin{bmatrix} B(s) \\ B(s_o) \end{bmatrix}, \left[\frac{C(s)}{s-s_o}, \frac{-C(s_o)}{s-s_o}\right], \frac{D(s)-D(s_o)}{s-s_o}\right].$$

Now consider a change of basis in the state space such that the state vector $\begin{bmatrix} x_t(s) \\ x_t(s_o) \end{bmatrix}$ is mapped into $T(s)\begin{bmatrix} x_t(s) \\ x_t(s_o) \end{bmatrix} = \begin{bmatrix} x_t(s_o) \\ \frac{x_t(s)-x_t(s_o)}{s-s_o} \end{bmatrix}$. This means that

$$(5.2\text{-}32)\quad T(s) = \begin{bmatrix} 0 & I \\ \frac{I}{s-s_o} & \frac{-I}{s-s_o} \end{bmatrix} \text{ and } T^{-1}(s) = \begin{bmatrix} I & (s-s_o)I \\ I & 0 \end{bmatrix}.$$

Applying this to the representation (5.2-31) one obtains the representation

$$(5.2\text{-}33)\qquad \left[\begin{bmatrix} A(s_o) & 0 \\ \dfrac{A(s)-A(s_o)}{s-s_o} & A(s) \end{bmatrix}, \begin{bmatrix} B(s_o) \\ \dfrac{B(s)-B(s_o)}{s-s_o} \end{bmatrix}, \begin{bmatrix} \dfrac{C(s)-C(s_o)}{s-s_o}, C(s) \end{bmatrix}, \dfrac{D(s)-D(s_o)}{s-s_o}\right].$$

Taking the limit for $s \to s_o$ one obtains

$$(5.2\text{-}34)\qquad \left[\begin{bmatrix} A(s_o) & 0 \\ \dot{A}(s_o) & A(s_o) \end{bmatrix}, \begin{bmatrix} B(s_o) \\ \dot{B}(s_o) \end{bmatrix}, [\dot{C}(s_o), C(s_o)), \dot{D}(s_o)\right].$$

Taking this limit is allowed)*. One way to see this is to consider the corresponding Markov matrix sequences. For each $k = 0,1,2,\ldots$,

$$\lim_{s\to s_o} \frac{H_k(s)-H_k(s_o)}{s-s_o} = \dot{H}_k(s_o),$$

while $\dot{H}_o(s_o) = \dot{D}(s_o)$ and

$$\dot{H}_k(s_o) = [\dot{C}(s_o), C(s_o)] \begin{bmatrix} A(s_o) & 0 \\ \dot{A}(s_o) & A(s_o) \end{bmatrix}^{k-1} \begin{bmatrix} B(s_o) \\ \dot{B}(s_o) \end{bmatrix}, \quad k = 1,2,\ldots$$

Note also that while this limit exists there is no such thing as a limit of (5.2-31). From the representation of $\dot{\Sigma}(s_o)$ it follows immediately that its McMillan degree does not exceed 2n. Simple examples show that the derivative system of a curve of systems of degree n can indeed have McMillan degree 2n, illustrating again that the condition that the McMillan degree is n is far from linear. From the representation of $\dot{\Sigma}(s_o)$ it also follows that $\dot{\Sigma}(s_o)$ is asymptotically stable, because the eigenvalues of $\begin{bmatrix} A(s_o) & 0 \\ \dot{A}(s_o) & A(s_o) \end{bmatrix}$ are the same as those of $A(s_o)$ (only with double multiplicities). That the representation may be nonminimal can easily be seen as follows. Consider a constant curve $s \longmapsto (A(s),B(s),C(s),D(s)) = (A,B,C,D)$. Then clearly $\dot{\Sigma} = 0$, the zero system, which has McMillan degree zero, and therefore if $\Sigma(s)$ has McMillan degree larger than or equal to one, the representation (5.2-34) is nonminimal.

Q.E.D.

)* cf. remark (iii).

5.2-35. Remarks. (i) The system $\dot{\Sigma}(s_o)$ has the following interpretation.

Consider the input-output mapping $F(s)$: $\overset{\circ}{U} \to \overset{\circ}{Y}$ corresponding to $\Sigma(s)$. (For the meaning of $\overset{\circ}{U},\overset{\circ}{Y},F$, see chapter 2.) Then the output at time t depends on s. Therefore we denote the output by $y_t(s)$. This is differentiable w.r.t. s at $s = s_o$, and the system with (the same inputs and) outputs $\dot{y}_t(s_o) = \frac{d}{ds} y_t(s_o)$

is $\dot{\Sigma}(s_o)$. The corresponding input-output mapping is $\frac{d}{ds}F(s_o)$: $\overset{\circ}{U} \to \overset{\circ}{Y}$.
(ii) Note that here we have considered differentiation along a curve, or in other words, differentiation in one direction only. By taking $n(m+m') + mm'$ independent directions and differentiating in those directions one obtains the Jacobian of the embedding of the differentiable manifold $\overset{\circ}{M}{}^{m,a}_{m',n,m}$ in the Hilbert space. (It is not difficult to show that $\dim \overset{\circ}{M}{}^{m,a}_{m',n,m} = n(m+m')+mm'$).
(iii) In [Hnz 88a] a complete proof of the fact that $\dot{\Sigma}$ is the derivative of $\Sigma(s)$ in the Hilbert space, is given.
Using this lemma it is not difficult to obtain the Riemannian metric that is induced by the Hilbert space structure on the manifold $\overset{\circ}{M}{}^{m,a}_{m',n,m}$. Let T_Σ denote the tangent space at $\overset{\circ}{M}{}^{m,a}_{m',n,m}$ in $\Sigma \in \overset{\circ}{M}{}^{m,a}_{m',n,m}$. A Riemannian metric is an inner product g_Σ on every tangent space T_Σ, which varies smoothly with Σ. (See any book on differential geometry like [Boo], [Ko-N], [Spi], [Au-M], [Bi-C], and also [Ab-M], [Arn]). The following notation will be used:
(i) A tangent vector in T_Σ will be represented by $(\dot{A},\dot{B},\dot{C},\dot{D})$, which is short for $(\dot{A}(s_o),\dot{B}(s_o),\dot{C}(s_o),\dot{D}(s_o))$, and s_o is such that Σ is represented by $(A(s_o),B(s_o),C(s_o),D(s_o))$.
(ii) The inner product g_Σ: $T_\Sigma \times T_\Sigma \to \mathbb{R}$ will be denoted both by

$\langle(\dot{A},\dot{B},\dot{C},\dot{D}), (\tilde{\dot{A}},\tilde{\dot{B}},\tilde{\dot{C}},\tilde{\dot{D}})\rangle_{g_\Sigma}$ and by $g_\Sigma((\dot{A},\dot{B},\dot{C},\dot{D}), (\tilde{\dot{A}},\tilde{\dot{B}},\tilde{\dot{C}},\tilde{\dot{D}}))$. The corresponding

norm will be denoted by $\|(\dot{A},\dot{B},\dot{C},\dot{D})\|_{g_\Sigma}$.

5.2-36. Theorem. The Riemannian metric that is induced by the embedding in Hilbert space is given by

$$(5.2\text{-}37)\quad \langle(\dot{A},\dot{B},\dot{C},\dot{D}),(\tilde{\dot{A}},\tilde{\dot{B}},\tilde{\dot{C}},\tilde{\dot{D}})\rangle_{g_\Sigma} = \langle\dot{\Sigma},\tilde{\dot{\Sigma}}\rangle,$$

where the latter inner product is in Hilbert space. Σ is given by (A,B,C,D),

$\dot{\Sigma}$ by (5.2-30) and similarly for $\tilde{\Sigma}$, $\dot{\tilde{\Sigma}}$.

Proof. This follows from lemma (5.2-30). Q.E.D.

So, using (5.2-37), the Riemannian metric can be computed, if the inner product in Hilbert space can be computed. Now consider the Hilbert space structure induced by the norm (5.2-8), in the discrete time case, and the norm (5.2-11) in the continuous time case. For those cases one can find more detailed formulas for the Riemannian metric. For simplicity of notation we will just present formulas for the norms. (It is wellknown that the inner product can be derived from the norm. In the real case:
$\|x+y\|^2-\|x\|^2-\|y\|^2 = 2\langle x,y\rangle$; in the complex case
$\|x+y\|^2 - \|x\|^2 - \|y\|^2 = 2\mathrm{Re}\langle x,y\rangle$, $\mathrm{Re}\langle x,iy\rangle = \mathrm{Im}\langle x,y\rangle$)

5.2-38. Theorem.

(a) Discrete time case

The norm (5.2-8) induces a Riemannian metric g_Σ on T_Σ at $\overset{\circ}{M}{}^{m,a}_{m',n,m}$ in Σ, where Σ is represented by (A,B,C,D); g_Σ is given by

$$\|(\dot{A},\dot{B},\dot{C},\dot{D})\|^2_{g_\Sigma} = \mathrm{tr}\{\dot{C}.L(BB^T).\dot{C}^T + \dot{C}.L(B\dot{B}^T)C^T + \dot{C}.L(A.L(BB^T).\dot{A}^T)C^T +$$

$$C.L(\dot{B}\dot{B}^T)C^T + C.L(\dot{B}B^T).\dot{C}^T + C.L(A.L(\dot{B}B^T)\dot{A}^T)C^T +$$

(5.2-39)

$$C.L(\dot{A}.L(B\dot{B}^T).A^T)C^T + C.L(\dot{A}.L(BB^T).A^T)\dot{C}^T + C.L[\dot{A}.L(A.L(BB^T).\dot{A}^T).A^T]C^T +$$

$$C.L[A.L(\dot{A}.L(BB^T).A^T)\dot{A}^T]C^T + C.L(\dot{A}.L(BB^T).\dot{A}^T)C^T + \dot{D}\dot{D}^T\},$$

where L(K) is the unique solution of the Lyapunov equation $L-ALA^T = K$, (as before).

(b) Continuous time case

The norm (5.2-11) induces a Riemannian metric g^c_Σ on T_Σ at $\overset{\circ}{M}{}^{m,\ell}_{m,'n,m}$, in the point Σ (the index 'ℓ' in $\overset{\circ}{M}{}^{m,\ell}_{m',n,m}$ stands for the open left half plane in $\mathbb{C}$ and it denotes continuous time asymptotically stable systems). Let Σ be represented by (A,B,C,D) and a tangent vector in T_Σ by $(\dot{A},\dot{B},\dot{C},\dot{D})$. Then g^c_Σ is

given by

$$(5.2\text{-}40)\quad \|(\dot{A},\dot{B},\dot{C},\dot{D})\|^2_{g^c_\Sigma} = \mathrm{tr}\{\dot{C}.M(BB^T).\dot{C}^T + \dot{C}.M(B\dot{B}^T)C^T +$$

$$\dot{C}.M[M(BB^T).\dot{A}^T]C^T + C.M(\dot{B}\dot{B}^T).C^T + C.M(B\dot{B}^T)\dot{C}^T + C.M[M(B\dot{B}^T).\dot{A}^T]C^T +$$

$$C.M[\dot{A}.M(B\dot{B}^T)]C^T + C.M[\dot{A}.M(BB^T)]\dot{C}^T + C.M[\dot{A}.M(M(BB^T).\dot{A}^T)]C^T +$$

$$C.M[M(\dot{A}.M(BB^T))\dot{A}^T]C^T + \dot{D}\dot{D}^T\},$$

where M(K) is the (unique) solution of the continuous time Lyapunov equation $AM + MA^T = -K$, as before.
In the complex case the same formulas hold if the transposition signs are read as 'transposition + conjugation'.

Proof. By definition of g_Σ one has $\|(\dot{A},\dot{B},\dot{C},\dot{D})\|^2_{g_\Sigma} = \|\dot{\Sigma}\|^2$. For simplicity of notation, let $s_o = 0$ and write (A,B,C,D) for $(A(0),B(0),C(0),D(0))$, and $(\dot{A},\dot{B},\dot{C},\dot{D})$ for $(\dot{A}(0),\dot{B}(0),\dot{C}(0),\dot{D}(0))$. Furthermore, let $L_{A,A(s)}(K)$ denote the solution of the Lyapunov equation
$(5.2\text{-}41)\quad L - A.L.A(s)^T = K$, etc.
Then one obtains

$$\|\dot{\Sigma}\|^2 = \|\lim_{s\to o}\frac{\Sigma(s)-\Sigma(0)}{s}\|^2 =$$

$$= \lim_{s\to o}\frac{1}{s^2}\|\Sigma(s) - \Sigma(0)\|^2 =$$

$$= \lim_{s\to o}\frac{1}{s^2}\,\mathrm{tr}\Big[\sum_{k=1}^{\infty}\{C(s)A(s)^{k-1}B(s) - C(0)A(0)^{k-1}B(0)\}\times$$

$$\{C(s)A(s)^{k-1}B(s) - C(0)A(0)^{k-1}B(0)\}^T +$$

$$+ \{D(s)-D(0)\}\{D(s)-D(0)\}^T\Big] =$$

$$= \mathrm{tr}\Big[\sum_{k=1}^{\infty}\frac{d}{ds}\{C(s)A(s)^{k-1}B(s)\}_{s=o}\frac{d}{ds}\{C(s)A(s)^{k-1}B(s)\}^T_{s=o} +$$

$$\dot{D}(0)\dot{D}(0)^T\Big] =$$

$$= \mathrm{tr}\Big[\sum_{k=1}^{\infty}\{\dot{C}A^{k-1}B + C\frac{d}{ds}(A(s)^{k-1}B(s))_{s=o}\} \times$$

$$\{\dot{C}A^{k-1}B + C\frac{d}{dr}(A(r)^{k-1}B(r))_{r=o}\}^T + \ddot{D}\dot{D}^T\Big] =$$

$$= \mathrm{tr}\Big[\dot{C}\{\sum_{k=1}^{\infty}A^{k-1}BB^T.(A^T)^{k-1}\}\dot{C}^T +$$

$$+ \dot{C}\frac{d}{dr}\{\sum_{k=1}^{\infty}A^{k-1}BB(r)^T(A(r)^T)^{k-1}\}_{r=o}C^T +$$

$$+ C\frac{d}{ds}\{\sum_{k=1}^{\infty}A(s)^{k-1}B(s)B^T.(A^T)^{k-1}\}_{s=o}\dot{C}^T +$$

$$+ C\frac{d}{ds}\frac{d}{dr}\{\sum_{k=1}^{\infty}A(s)^{k-1}B(s)B(r)^T(A(r)^T)^{k-1}\}_{s=o,r=o}C^T + \ddot{D}\dot{D}^T\Big] =$$

$$= \mathrm{tr}\Big[\dot{C}.L(BB^T).\dot{C}^T + \dot{C}\frac{d}{dr}L_{A,A(r)}(BB(r)^T)_{r=o}C^T +$$

$$+ C\frac{d}{ds}L_{A(s),A}(B(s)B^T)_{s=o}\dot{C}^T +$$

$$+ C\frac{d}{ds}\frac{d}{dr}L_{A(s),A(r)}(B(s)B(r)^T)_{s=o,r=o}.C^T + \ddot{D}\dot{D}^T\Big].$$

Now we have to compute $\frac{d}{dr}L_{A,A(r)}(BB(r)^T)_{r=o}$, $\frac{d}{ds}L_{A(s),A}(B(s)B^T)_{s=o}$ and $\frac{d}{ds}\frac{d}{dr}L_{A(s),A(r)}(B(s)B(r)^T)_{s=o,r=o}$. Consider the equation

$$L(r) - AL(r)A(r)^T = BB(r)^T.$$

Differentiation with respect to r in r = 0 gives

$$(\frac{dL}{dr}) - A(\frac{dL}{dr})A^T - A.L.\frac{dA^T}{dr} = B(\frac{dB}{dr})^T,$$

or, equivalently,

$$(\frac{dL}{dr}) - A(\frac{dL}{dr})A^T = B\dot{B}^T + AL\dot{A}^T,$$

so $(\frac{dL}{dr})_{r=o} = L[B\dot{B}^T + A.L(BB^T)\dot{A}^T]$, where $(\frac{dL}{dr})_{r=o}$ stands for

$$\frac{d}{dr}L_{A,A(r)}[B.B(r)^T]_{r=o}.$$

Similarly

$$\frac{d}{ds}L_{A(s),A}[B(s)B^T] = L[\dot{B}\dot{B}^T+\dot{A}L(BB^T)A^T].$$

Now consider the equation

$$L(s,r) - A(s)L(s,r)A(r)^T = B(s)B(r)^T.$$

Partial differentiation with respect to s and r gives

$$\frac{\partial^2 L}{\partial r\partial s} - \dot{A}\frac{\partial L}{\partial r}.A^T - \dot{A}L\dot{A}^T - A(\frac{\partial^2 L}{\partial r\partial s})A^T - A\frac{\partial L}{\partial s}.\dot{A}^T = \ddot{B}\dot{B}^T,$$

or, equivalently,

$$(\frac{\partial^2 L}{\partial r\partial s}) - A(\frac{\partial^2 L}{\partial r\partial s})A^T = \ddot{B}\dot{B}^T + \dot{A}L(BB^T)\dot{A}^T + \dot{A}.\frac{\partial L}{\partial r}A^T + A\frac{\partial L}{\partial s}\dot{A}^T =$$

$$= \ddot{B}\dot{B}^T + \dot{A}L(BB^T)\dot{A}^T + \dot{A}.L[\dot{B}\dot{B}^T+\dot{A}.L(BB^T)\dot{A}^T]A^T + A.L[\dot{B}\dot{B}^T+\dot{A}.L(BB^T)A^T]\dot{A}^T.$$

Therefore

$$\frac{\partial^2}{\partial r\partial s}L_{A(s),A(r)}[B(s)B(r)^T]_{s=o,r=o} =$$

$$= L[\ddot{B}\dot{B}^T+\dot{A}L(BB^T)\dot{A}^T+\dot{A}.L\{\dot{B}\dot{B}^T+\dot{A}.L(BB^T)\dot{A}^T\}A^T + A.L\{\dot{B}\dot{B}^T+\dot{A}.L(BB^T)A^T\}\dot{A}^T].$$

Substituting this and using the linearity of the operator L, one obtains

$$\|\dot{\Sigma}\|^2 = tr[\dot{C}.L(BB^T).\dot{C}^T + \dot{C}.L\{\dot{B}\dot{B}^T+\dot{A}L(BB^T)\dot{A}^T\}.C^T + C.L\{\dot{B}\dot{B}^T+\dot{A}L(BB^T)A^T\}\dot{C}^T +$$

$$+ C.L\{\ddot{B}\dot{B}^T+\dot{A}L(BB^T)\dot{A}^T + \dot{A}.L(\dot{B}\dot{B}^T+\dot{A}.L(BB^T)\dot{A}^T).A^T +$$

$$+ A.L(\dot{B}\dot{B}^T+\dot{A}.L(BB^T).A^T)\dot{A}^T\}C^T+\ddot{D}\dot{D}^T] =$$

$$= tr[\dot{C}.L(BB^T)\dot{C}^T+\dot{C}L(\dot{B}\dot{B}^T)C^T+\dot{C}.L\{A.L(BB^T)\dot{A}^T\}C^T +$$

$$+ C.L(\dot{B}\dot{B}^T)\dot{C}^T+C^T.L\{\dot{A}.L(BB^T)A^T\}\dot{C}^T +$$

$$+ C.L(\ddot{B}\dot{B}^T)C^T+C.L\{\dot{A}.L(BB^T)\dot{A}^T\}C^T +$$

$$+ C.L\{\dot{A}.L(B\dot{B}^T)A^T\}C^T + C.L\{\dot{A}.L(A.L(BB^T)\dot{A}^T)A^T\}.C^T +$$

$$+ C.L\{A.L(\dot{B}\dot{B}^T)\dot{A}^T\}C^T + CL\{A.L(\dot{A}.L(BB^T)A^T)\dot{A}^T\}C^T + \ddot{D}\ddot{D}^T],$$

which is the formula (with the 12 terms in a different order) of the theorem. A similar computation can be made for the continuous time case:

$$\|\dot{\Sigma}\|^2 = \|\lim_{s\to o}\frac{\Sigma(s)-\Sigma(0)}{s}\|^2 =$$

$$\lim_{\substack{s\to o\\ r\to o}} \langle\frac{\Sigma(s)-\Sigma(0)}{s},\frac{\Sigma(r)-\Sigma(0)}{r}\rangle =$$

$$\frac{\partial}{\partial s}\frac{\partial}{\partial r}\langle\Sigma(s),\Sigma(r)\rangle_{s=o,r=o} =$$

$$\text{tr}\ \frac{\partial}{ds}\frac{\partial}{\partial r}\{C(s).M_{A(s),A(r)}[B(s)B(r)^T].C(r)^T+D(s)D(r)^T\}_{s=o,r=o} .$$

We will need $\frac{\partial}{\partial r}M_{A,A(r)}[BB(r)^T]_{r=o}$, $\frac{\partial}{\partial s}M_{A(s),A}[B(s)B^T]_{s=o}$ and $\frac{\partial^2}{\partial s\partial r}M_{A(s),A(r)}[B(s)B(r)^T]$. Consider $A(s)M(s,r)+M(s,r)A(r)^T = -B(s)B(r)^T$. Take the partial derivative with respect to r,

$$A(\frac{\partial M}{\partial r}) + (\frac{\partial M}{\partial r})A^T + M\dot{A}^T = - B\dot{B}^T$$

or, equivalently

$$(\frac{\partial M}{\partial r}) = M[B\dot{B}^T+M(BB^T).\dot{A}^T].$$

Similarly

$$(\frac{\partial M}{\partial s}) = M[B\dot{B}^T+\dot{A}M(BB^T)].$$

Now take the second order partial derivative with respect to s and r. This gives

$$\dot{A}\frac{\partial M}{\partial r} + A\frac{\partial^2 M}{\partial r\partial s} + \frac{\partial^2 M}{\partial r\partial s}.A^T + \frac{\partial M}{\partial s}\dot{A}^T = -\ddot{B}\ddot{B}^T.$$

So

$$\left(\frac{\partial^2 M}{\partial r \partial s}\right) = M\left[\ddot{B}B^T + \dot{A}\frac{\partial M}{\partial r} + \frac{\partial M}{\partial s}\dot{A}^T\right] =$$

$$= M\left[\ddot{B}B^T + \dot{A}.M\{\dot{B}B^T + M(BB^T)\dot{A}^T\} + M\{\dot{B}B^T + \dot{A}\,M(BB^T)\}\dot{A}^T\right].$$

Using these formulas, one obtains

$$\begin{aligned}
\|\dot{\Sigma}\|^2 &= \mathrm{tr}\frac{\partial}{\partial s}\frac{\partial}{\partial r}\{C(s)M_{A(s),A(r)}[B(s)B(r)^T].C(r)^T\}_{s=o,r=o} + \mathrm{tr}\ddot{D}D^T = \\
&= \mathrm{tr}\frac{\partial}{\partial s}\{C(s).\frac{\partial M}{\partial r}.C^T + C(s)M(B(s)B^T)\dot{C}^T\}_{s=o} + \mathrm{tr}\ddot{D}D^T = \\
&= \mathrm{tr}\left[\dot{C}.\frac{\partial M}{\partial r}C^T + C\frac{\partial^2 M}{\partial s \partial r}C^T + \dot{C}.M(BB^T)\dot{C}^T + C\frac{\partial M}{\partial s}\dot{C}^T + \ddot{D}D^T\right] = \\
&= \mathrm{tr}[\dot{C}.M(B\dot{B}^T + M(B\dot{B}^T).\dot{A}^T)C^T + \\
&\quad + C.M\{\ddot{B}B^T + \dot{A}.M[\dot{B}B^T + M(B\dot{B}^T).\dot{A}^T] + \\
&\quad + M[\dot{B}B^T + \dot{A}.M(BB^T)].\dot{A}^T\}.C^T + \\
&\quad + \dot{C}M(BB^T)\dot{C}^T + C.M\{\dot{B}B^T + \dot{A}M(BB^T)\}\dot{C}^T + \ddot{D}D^T] = \\
&= \mathrm{tr}[\dot{C}.M(B\dot{B}^T)C^T + \dot{C}.M\{M(BB^T)\dot{A}^T\}C^T + \\
&\quad + C.M(\ddot{B}B^T)C^T + C.M\{\dot{A}.M(B\dot{B}^T)\}C^T + \\
&\quad + C.M\{\dot{A}.M(M(B\dot{B}^T).\dot{A}^T)\}C^T + \\
&\quad + C.M\{M(\dot{B}B^T)\dot{A}^T\}C^T + C.M\{M(\dot{A}.M(BB^T))\dot{A}^T\}C^T + \\
&\quad + \dot{C}.M(BB^T)\dot{C}^T + C.M(\dot{B}B^T)\dot{C}^T + \\
&\quad + C.M(\dot{A}.M(BB^T))\dot{C}^T + \ddot{D}D^T].
\end{aligned}$$

After some rearrangement of the terms, one finds the formula in the theorem.

Q.E.D.

5.2-41. Remarks. (i) It follows directly from the theorem that g_Σ varies smoothly with Σ.

(ii) A similar formula can be found if one uses the inner product on the Hilbert space that corresponds with the Frobenius norm of the Hankel matrix (see (5.2-16) - (5.2-20)).

(iii) If one wants to compute the components of the Riemannian metric tensor with these formulas this can be done as follows. Consider a mapping vec:

$(A,B,C,D) \longmapsto \mathrm{vec}(A,B,C,D)$ which maps (A,B,C,D) into a vector which contains all the components of the matrices A,B,C,D in some specific order. E.g. first the first column of A, then the second etc., then the first column of B, then the second etc., etc. Of course the mapping vec is invertible; vec(A,B,C,D) is an $(n+m).(n+m')$-dimensional vector. Now choose a local canonical form for (A,B,C,D), as described in section 4.4 (for example). In such a canonical form a number of components of vec(A,B,C,D) are fixed (usually they are prescribed to be zero, or one), while the other components can vary freely within (at least) an open neighbourhood. The number of free parameters will be $n(m+m') + mm'$, if there are no restrictions. (E.g. if one requires D to be identically zero, then the number of free parameters will be $n(m+m')$). The tangent space in a point (A,B,C,D) at the manifold $\overset{\circ}{M}{}^{m,a}_{m',n,m}$ can now be described as follows: It is the vector space of all $(n+m)(n+m')$-dimensional vectors, denoted by $\mathrm{vec}(\dot{A},\dot{B},\dot{C},\dot{D})$, the components of which are identically zero iff the corresponding component of vec(A,B,C,D) is fixed in the local canonical form. All other $n(m+m') + mm'$ components can vary freely. So the tangent space is isomorphic to $\mathbb{R}^d$, with $d = n(m+m') + mm'$. Choose a basis of the tangent space, denoted by $e_1,e_2,\ldots,e_d$; $e_i = \mathrm{vec}(\dot{A}_i,\dot{B}_i,\dot{C}_i,\dot{D}_i)$, $i = 1,\ldots,d$, say. With the formulas of the theorem (5.2-38), one can compute

$$g_\Sigma(e_i,e_i) = \|(\dot{A}_i,\dot{B}_i,\dot{C}_i,\dot{D}_i)\|^2_{g_\Sigma}, \text{ and}$$

$$g_\Sigma(e_i+e_j,e_i+e_j) = \|(\dot{A}_i+\dot{A}_j,\dot{B}_i+\dot{B}_j,\dot{C}_i+\dot{C}_j,\dot{D}_i+\dot{D}_j)\|^2_{g_\Sigma}.$$

From this $g_\Sigma(e_i,e_j)$ can be computed by the standard formula

$$2g_\Sigma(e_i,e_j) = g_\Sigma(e_i+e_j,e_i+e_j) - g_\Sigma(e_i,e_i) - g_\Sigma(e_j,e_j).$$

in the real case. (See the remarks just before theorem (5.2-38)). This can be used both for the numerical calculation of the Riemannian metric tensor, as for its algebraic (i.e. symbolic) calculation.

For a number of cases the Riemannian metric tensor was computed symbolically by W. Mak. Some of the results will be presented in Appendix 5A. The formulas become rather big for n = 3 already. In applications, the Riemannian metric tensor will have to be computed numerically in many cases. On the other hand in those cases in which symbolic computation is possible, it is also possible to obtain the answers in Fortran notation, and so one can use the results

directly in a Fortran program without having to type the formulas by hand! Especially if the calculation of the Riemannian metric tensor is needed at each step of an iterative process, the availability of explicit formulas can be a big advantage. Otherwise one has to solve a number of Lyapunov equations at each iteration step. (In the next chapter, an algorithm will be presented in which the Riemannian metric tensor needs to be calculated at each step).

5.3. The short time-interval case

5.3.1. Introduction

If one has a discrete time system with a short time interval, the usual representation of the state equation

$$(5.3.1\text{-}1)\quad x_{t+1} = Ax_t + Bu_t$$

might not be very well suited. The matrix A will be close to the identity matrix in many cases. Furthermore in the limiting case of a zero time interval the equation no longer makes sense. In that case one uses the continuous time state equation

$$(5.3.1\text{-}2)\quad \dot{x}_t = Ax_t + Bu_t.$$

Of course the choice between discrete time and continuous time also affects the norm that was introduced in the previous subsection (formulas (5.2-8) and (5.2-11)). I.e. the norm of a system will be completely different if it is modeled as a discrete time system or (approximately) as a continuous time system. In this paragraph we want to introduce a new way of describing a linear system, using a new δ-operator, which encompasses both the continuous time case and the discrete time case in a mathematically elegant way. With this description we obtain a norm which depends on the length of the time interval. This new norm no longer has the disadvantages mentioned. I.e. it will be continuous in the length of the time interval, and so if the length of the time interval is close to zero, using the continuous time representation will not change the norm very much. This also shows (once more) that the norms that we have defined for the continuous time case and the discrete time case are intimately related.

A mathematically appealing aspect of the δ-operator that we propose is the position of the corresponding stability region. In the discrete time case this is of course the unit disk in while for the continuous time case it is the

left half plane. It is well-known that the unit disk can be mapped onto the left halfplane conformally by a Möbius transformation. This Möbius transformation is in fact just a rotation of 90° of the Riemann sphere. The new δ-operator also corresponds to rotations of the Riemann sphere, but, more generally, to a rotation somewhere between 0° and 90° (and in fact we can formally allow also rotations between -90° and 0°). This means that the stability region will be a hemisphere of the Riemann sphere, just as in the discrete time- and continuous time case. This will be worked out in the following subsections.

5.3.2. A useful family of difference/differential operators

It has been argued that linear systems in discrete time should be represented by a difference operator instead of a delay- or forward-shift-operator (cf. [Goo 85]), especially in the case of short time intervals. We propose an intermediate object, namely the operator δ_Δ, depending on $\Delta \in [-1,1]$, defined by

$$(5.3.2\text{-}1)\quad \delta_\Delta x_t = \begin{cases} \dfrac{x_{t+\Delta} - \sqrt{(1-\Delta^2)}.x_t}{\Delta} & \text{if } 0 < |\Delta| \leq 1, \\ \dot{x}_t & \text{if } \Delta = 0. \end{cases}$$

(We assume x_t to be differentiable). We will usually think of Δ as an element of $[0,1]$, in which case it denotes the time interval length. But formally it can just as well be defined for $\Delta \in [-1,0)$ (the 'backward case'). If $\Delta = 1$, one has the usual forward shift operator $\delta_1 x_t = x_{t+1}$.
At the other hand, for $\Delta \to 0$, one has

$$\delta_\Delta x_t = \frac{x_{t+\Delta} - \sqrt{(1-\Delta^2)}x_t}{\Delta} = \frac{x_{t+\Delta} - x_t}{\Delta} + x_t.\frac{1-\sqrt{(1-\Delta^2)}}{\Delta} = \frac{x_{t+\Delta} - x_t}{\Delta} + x_t O(\Delta).$$

Therefore, for each t, one has

$$(5.3.2\text{-}2)\quad \lim_{\Delta \to o} \delta_\Delta x_t = \lim_{\Delta \to o} \frac{x_{t+\Delta} - x_t}{\Delta} = \dot{x}_t.$$

So it is reasonable to take, as is done here, δ_o to be the differential operator $\delta_o x = \dot{x}$. For Δ close to zero, the δ_Δ operator is close to the differential quotient $(x_{t+\Delta} - x_t)/\Delta$. Now consider a linear system in discrete time, with time interval length $\Delta > 0$:

$$(5.3.2\text{-}3)\begin{cases} x_{t+\Delta} = Ax_t + Bu_t, \\ y_t = Cx_t + Du_t. \end{cases}$$

Suppose (A,B,C,D) forms a minimal representation. Then the system is asymptotically stable iff the spectrum of A is included in the open unit disk in the complex plane, i.e.

$$(5.3.2\text{-}4)\quad \sigma(A) \subseteq D(0,1).$$

Now suppose we describe the same linear system with the δ_Δ-operator

$$(5.3.2\text{-}5)\begin{cases} \delta_\Delta x_t = A_\Delta x_t + B_\Delta u_t, \\ y_t = Cx_t + Du_t, \end{cases}$$

with $A_\Delta = \frac{A-\sqrt{(1-\Delta^2)}I}{\Delta}$, $B_\Delta = \frac{B}{\Delta}$.

Let us introduce the (bijective) mapping

$$(5.3.2\text{-}6)\quad \Psi(z) = \frac{z-\sqrt{(1-\Delta^2)}}{\Delta},$$

then

$$(5.3.2\text{-}7)\quad A_\Delta = \Psi(A)$$

and it is clear that the system (5.3.2-5) is asymptotically stable iff

$$(5.3.2\text{-}8)\quad \sigma(A_\Delta) \subseteq \Psi(D(0,1)).$$

The set $\Psi(D(0,1))$ will be called the '<u>stability region</u> of the δ_Δ-operator'.

5.3.3. <u>The stability region of the δ-operator and rotations of the Riemann sphere</u>

In this section the 'stability region of the δ-operator' will be found, for all $\Delta \in [0,1]$. I.e. the stability of the equation $\delta_\Delta x = Ax$ is discussed in terms of a region of the Riemann sphere such that stability means, that all eigenvalues of A are in this region. It is part of the statement of the result that stability of this equation can be described in this way. It will be shown

that this stability region is a hemisphere of the Riemann sphere, and it can be obtained from the unit disk (which is a hemisphere on the Riemann sphere) by rotation of the Riemann sphere around the axis through $\pm$ i. We will analyze the mapping Ψ, and give a complete geometric description of its action on the Riemann sphere. It will turn out that to test the asymptotic stability of such a system, especially if $\Delta > 0$ is small, it is <u>not</u> preferable to transform back to the standard discrete time representation! Instead one can use a <u>rotation</u> of the Riemann sphere which maps the stability region into the open left half plane, and apply e.g. the Routh-Hurwitz stability criterion. For the theoretical background of this section we refer to [Ne-P].

Let us now introduce the Riemann sphere. We will picture it sligthly different than in [Ne-P]. Let the complex plane be considered as the subspace $\{(x_1,x_2,0)\,|\,x_1+ix_2\in\mathbb{C}\}$ of $\mathbb{R}^3$. Consider the unit ball $S^2 = \{(x_1,x_2,x_3)\,|\,x_1^2+x_2^2+x_3^2 = 1\}$ in $\mathbb{R}^3$. Consider the point $(0,0,-1)$, and project $\mathbb{C}$ stereographically on S^2. I.e. $z \in \mathbb{C}$, $z = x_1 + ix_2$ is mapped to the point of intersection of $S^2\setminus\{(0,0,-1)\}$ and the line through z and $(0,0,-1)$. A little calculation shows that $z = x_1 + ix_2$ is mapped to the point

$$(5.3.3\text{-}1)\qquad \left[\frac{2x_1}{1+|z|^2},\ \frac{2x_2}{1+|z|^2},\ \frac{1-|z|^2}{1+|z|^2}\right] \in S^2\setminus\{(0,0,-1)\}.$$

It is clear that the unit circle, $|z| = 1$, remains fixed under this projection! The mapping $\mathbb{C} \to S^2\setminus\{(0,0,-1)\}$ is one-to-one. Its inverse is given by the mapping $S^2\setminus\{(0,0,-1)\} \to \mathbb{C}$,

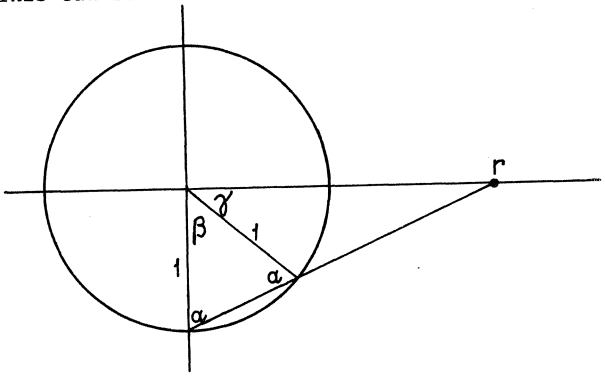
$$(\xi,\eta,\zeta),\ \text{with}\ \xi^2+\eta^2+\zeta^2 = 1,\ (\xi,\eta,\zeta) \neq (0,0,-1)\ \text{is mapped to}$$

$$(5.3.3\text{-}1a)\qquad \frac{\xi+i\eta}{1+\zeta}.$$

The same mapping can also be expressed in polar coordinates for $\mathbb{C}$ and spherical coordinates for S^2. I.e. we denote $z \in \mathbb{C}$ by $z = re^{i\phi}$, $0 \le \phi < 2\pi$, and a point on the sphere S^2 by $(x,y,z) = (\cos\theta\cos\phi,\cos\theta\sin\phi,\sin\theta)$, $-\frac{\pi}{2} \le \theta \le \frac{\pi}{2}$, $0 \le \phi < 2\pi$. In these coordinates, under the mapping $\mathbb{C} \to S^2$, ϕ remains fixed, while r is mapped to

$$(5.3.3\text{-}2)\qquad \theta = \operatorname{arctg}\left(\frac{1-r^2}{2r}\right),\ -\frac{\pi}{2} < \theta \le \frac{\pi}{2}.$$

This can be shown as follows.



Clearly, $r = tg\alpha$. From $2\alpha + \beta = \pi$ and $\beta + \gamma = \frac{1}{2}\pi$, it follows that $\gamma = 2\alpha - \frac{1}{2}\pi$. Now $\theta = -\gamma$, so $\theta = \frac{1}{2}\pi - 2\alpha$, and

$$tg\theta = tg(\tfrac{1}{2}\pi - 2\alpha) = \frac{1}{tg(2\alpha)} = \frac{1-tg^2\alpha}{2tg\alpha} = \frac{1-r^2}{2r}.$$

The inverse is

$$(5.3.3\text{-}3)\quad r = \frac{\cos\theta}{1+\sin\theta},\quad -\frac{\pi}{2} < \theta \leq \frac{\pi}{2}.$$

The Riemann sphere is completed by adding the point $P = (0,0,-1)$, which corresponds to the point ∞. So the Riemann sphere represents $\mathbb{C} \cup \{\infty\}$. Without proof we mention some properties of the Riemann sphere (cf. [Ne-P], chapter 3).

(i) The mapping of the complex plane to the Riemann sphere is angle preserving.

(ii) Circles in the complex plane are mapped to circles not passing through P on the Riemann sphere, and conversely.

(iii) Straight lines in the complex plane are mapped to circles passing through P on the Riemann sphere, and conversely.

(iv) Opposite points on the Riemann sphere correspond to a pair $(z,-(\bar{z})^{-1}) \in \mathbb{C} \cup \{\infty\} \times \mathbb{C} \cup \{\infty\}$.

Fractional linear transformations, or Möbius transformations, are functions of the form

$$w = f(z) = \frac{az+b}{cz+d},$$

where the determinant of the coefficients is assumed to be different from zero: $ad - bc \neq 0$. The linear transformations form a group under composition. A linear transformation furnishes a one-to-one mapping of $\mathbb{C} \cup \{\infty\}$ onto $\mathbb{C} \cup \{\infty\}$, or, equivalently, a one-to-one mapping of the Riemann sphere onto the Riemann sphere. It maps a circle on the Riemann sphere onto a circle on the Riemann sphere (cf. [Ne-P], section 3.4).
Two subgroups of linear transformations will be of special importance to us.
(a) The group of linear transformations of the form

$$(5.3.3\text{-}4)\quad w = \frac{az+b}{-\bar{b}z+\bar{a}},\ (a,b) \in \mathbb{C}^2 \setminus \{0\},$$

represents the group of <u>rotations</u> of the Riemann sphere ([Ne-P], section 3.14)
(b) The group of linear transformations of the form

$$(5.3.3\text{-}5)\quad w = \frac{az+b}{\bar{b}z+\bar{a}},\ a,b \in \mathbb{C},\ |a| > |b|,$$

constitutes the group of conformal mappings of the unit disk onto itself.

We can now state the main theorem of this subsection

5.3.3-6. <u>Theorem</u>. Let $\Delta \in (0,1]$.

(a) The mapping $\Psi(z) = \frac{z-\sqrt{(1-\Delta^2)}}{\Delta}$ can be decomposed as

$$(5.3.3\text{-}7)\quad \Psi = \rho \circ \phi,$$

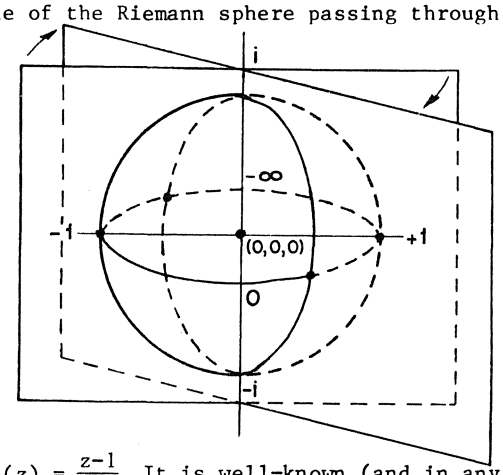
where $\rho(z) = \frac{\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}}{\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}}$, a rotation of the Riemann sphere around the axis through $+i$ and $-i$, and $\phi(z) = \frac{\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}}{-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}}$, a conformal mapping of the unit disk into itself. Its fixed points are ± 1.
(b) The stability region of the δ-operator is equal to $\rho(D(0,1)) = D\left(\frac{-\sqrt{(1-\Delta^2)}}{\Delta}, \frac{1}{\Delta}\right)$, (i.e. the disk in $\mathbb{C}$ with centre $\frac{-\sqrt{(1-\Delta^2)}}{\Delta}$ and radius $1/\Delta$) which is a hemisphere of the Riemann sphere. Even if $\Delta = 0$, $\rho(D(0,1))$ is the stability region of the δ-operator. In that case it is equal to the open left half plane, which is also a Riemann-hemisphere.

Proof. A simple calculation shows that indeed $\rho \circ \phi = \Psi$. Another simple

calculation shows that $-i\rho(i) = 1$, i.e. $\rho(i) = i$, and by conjugation $\rho(-i) = -i$. Because ρ if of the form (5.3.3-4), it is a rotation of the Riemann sphere, around the axis through $\pm i$. $\phi(z)$ is of the form (5.3.3-5), because $\Delta \in (0,1]$, and so ϕ maps the unit disk conformally onto itself. It is clear that $\phi(1) = 1$, and $\phi(-1) = -1$. This proves (a). The stability region of the δ-operator, $\Delta \in (0,1]$, was found to be $\Psi(D(0,1))$. Application of (a) gives

(5.3.3-8) $\Psi(D(0,1)) = \rho \circ \phi\ (D(0,1)) = \rho(D(0,1))$.

Because $D(0,1)$ is a Riemann-hemisphere (it corresponds to the intersection of S^2 with $x_3 > 0$), and ρ is a rotation of the Riemann sphere, around the axis through $\pm i$, $\rho(D(0,1))$ is again a Riemann hemisphere (and the boundary circle is a great circle of the Riemann sphere passing through $\pm i$).



If $\Delta = 0$, then $\rho(z) = \frac{z-1}{z+1}$. It is well-known (and in any case, easily seen if one uses the properties of linear transformation mentioned above) that in this case, ρ maps the open unit disk to the open left half plane (and the unit circle to the imaginary axis). Q.E.D.

5.3.3-9. Remark. It follows that if $z \in$ stability region of the δ_Δ-operator, then $(-\bar{z})^{-1} \notin$ stability region.

In the remainder of this subsection we want to analyse the mappings ρ and ϕ somewhat further. The mapping ρ is a rotation of the Riemann sphere around the axis through $\pm i$, over an angle $-\alpha$, say.

If $\Delta \in [0,1]$, then $\alpha \in [0,\pi/2]$. To be more precise

$$\alpha = 2 \operatorname{arctg}\left(\frac{\sqrt{(1-\Delta)}}{\sqrt{(1+\Delta)}}\right) = \arccos(\Delta).$$

Of course, the set of all rotations of the Riemann sphere around the axis through $\pm i$ forms a group, with addition of the angles, modulo 2π, as the group composition.

5.3.3-10. Example. If $\Delta = \frac{1}{2}\sqrt{2}$, then $\alpha = \pi/4$, i.e. the rotation is halfway between 0 and $\pi/2$. It follows (and it is easy to check directly by computation), that

$$\rho \circ \rho = \frac{z-1}{z+1}$$

if

$$\rho = \frac{\sqrt{(1+\frac{1}{2}\sqrt{2})}z-\sqrt{(1-\frac{1}{2}\sqrt{2})}}{\sqrt{(1-\frac{1}{2}\sqrt{2})}z+\sqrt{(1+\frac{1}{2}\sqrt{2})}}.$$

5.3.3-11. Remark. To find out whether the matrix A_Δ is stable, one can compute its characteristic polynomial $p_\Delta(w)$, and check whether this polynomial has all its zeroes within $D\left(\frac{-\sqrt{(1-\Delta^2)}}{\Delta}, \frac{1}{\Delta}\right)$. To do this one can apply a substitution $w = \rho(z)$, and check whether the resulting rational function in z has its zeroes all within the open unit disk. Of course, one only has to check whether the numerator polynomial has all its zeroes within the open unit disk, for which there are standard procedures. Similarly, one could apply a substitution $w = \rho'(s)$, where ρ' is the rotation of the Riemann sphere over the angle $\frac{\pi}{2} - \alpha$, around the axis through $\pm i$. Here α is the angle found in (5.3.3-9). In this case the resulting numerator polynomial should have all its zeroes in the open left half plane. Of course in both cases numerator and denominator have to be relatively prime as polynomials. This is certainly the case if the denominator is chosen to be $\left(\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\right)^n$, $n = \deg p_\Delta(w)$, in the first case, etc.

5.3.3-12. Example. Let $p_\Delta = aw^2 + bw + c$, then its zeroes are in $D\left(-\frac{\sqrt{(1-\Delta^2)}}{\Delta}, \frac{1}{\Delta}\right)$ iff

$$q(z) := a\left(\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}\right)^2 + b\left(\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}\right)\left(\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\right) + c\left(\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\right)^2 =$$

$$= \{(1+\Delta)a+\sqrt{(1-\Delta^2)}b+(1-\Delta)c\}z^2 + \{-2\sqrt{(1-\Delta^2)}a+2\Delta b+2\sqrt{(1-\Delta^2)}c\}z +$$
$$\{(1-\Delta)a - \sqrt{(1-\Delta^2)}b + (1+\Delta)c\}$$

has its zeroes in the open unit disk. We will not expand on this further.

Let us turn to the mapping ϕ. We already know that ϕ maps the unit disk conformally onto itself. We want to show that this mapping in contradiction to ρ, 'degenerates' for $\Delta \downarrow 0$ in a certain sense. To begin with let us make some preliminary observations for the behaviour of ϕ as $\Delta \downarrow 0$.

5.3.3-13. Proposition. $\lim_{\Delta\downarrow 0} \phi(z) = \begin{cases} -1 \text{ if } z \neq 1, \\ 1 \text{ if } z = 1. \end{cases}$

Proof. If $z \neq 1$, $\lim_{\Delta\downarrow o} \frac{\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}}{-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}} = \frac{z-1}{-z+1} = -1.$

If $z = 1$, $\lim_{\Delta\downarrow o} \frac{\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}}{-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}} = \lim_{\Delta\downarrow o} \frac{\sqrt{(1+\Delta)}-\sqrt{(1-\Delta)}}{-\sqrt{(1-\Delta)}+\sqrt{(1+\Delta)}} = 1.$

Q.E.D.

5.3.3-14. Proposition. If $w = \phi(z)$, then

$$\frac{w-1}{w+1} = \lambda\frac{z-1}{z+1},$$

with $\lambda = \frac{1+\sqrt{(1-\Delta^2)}}{\Delta}$.

Proof.

$$\frac{w-1}{w+1} = \frac{\left(\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}\right)-\left(-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\right)}{\left(\sqrt{(1+\Delta)}z-\sqrt{(1-\Delta)}\right)+\left(-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\right)} =$$

$$= \frac{\left(\sqrt{(1+\Delta)}+\sqrt{(1-\Delta)}\right)z-\left(\sqrt{(1-\Delta)}+\sqrt{(1+\Delta)}\right)}{\left(\sqrt{(1+\Delta)}-\sqrt{(1-\Delta)}\right)z+\left(\sqrt{(1+\Delta)}-\sqrt{(1-\Delta)}\right)} = \left(\frac{\sqrt{(1+\Delta)}+\sqrt{(1-\Delta)}}{\sqrt{(1+\Delta)}-\sqrt{(1-\Delta)}}\right)\cdot\frac{z-1}{z+1}.$$

Therefore $\lambda = \frac{\sqrt{(1+\Delta)}+\sqrt{(1-\Delta)}}{\sqrt{(1+\Delta)}-\sqrt{(1-\Delta)}} = \frac{\left(\sqrt{(1+\Delta)}+\sqrt{(1-\Delta)}\right)^2}{(1+\Delta)-(1-\Delta)} = \frac{2+2\sqrt{(1-\Delta^2)}}{2\Delta} = \frac{1+\sqrt{(1-\Delta^2)}}{\Delta}$.

Q.E.D.

The result that will be shown about ϕ is the following.

5.3.3-15. Theorem. ϕ considered as a mapping of the Riemann sphere onto itself is equal to the mapping that is obtained by the following two-step procedure

(i) Multiply the Riemann sphere with the factor $\lambda = \frac{1+\sqrt{(1-\Delta^2)}}{\Delta}$ in the x_2 and x_3-directions.

(ii) Project stereographically back on the Riemann sphere in the direction of the point $(-1,0,0)$ (which corresponds to $-1 \in \mathbb{C}$ of course).

Proof. The mapping described in (i) is $\beta_1\colon S^2 \to \mathbb{R}^3$, $(x_1,x_2,x_3) \longmapsto (x_1,\lambda x_2,\lambda x_3)$, and the one described by (ii) is

$\beta_2\colon \beta_1(S^2) \to S^2$, $(x_1,\lambda x_2,\lambda x_3) \longmapsto \mu(x_1,\lambda x_2,\lambda x_3) + (1-\mu)(-1,0,0)$, where $\mu \in (0,1]$ is such that the right hand side lies on S^2. If $(x_1,x_2,x_3) = (-1,0,0)$, it follows that $\beta_2 \circ \beta_1(x_1,x_2,x_3) = (-1,0,0)$. If $(x_1,x_2,x_3) \neq (-1,0,0)$, μ can be computed to be

$$(5.3.3\text{-}16) \qquad \mu = \frac{1-\sqrt{(1-\Delta^2)}}{1-x_1\sqrt{(1-\Delta^2)}},$$

using $\lambda = \frac{1+\sqrt{(1-\Delta^2)}}{\Delta}$, $\lambda^{-1} = \frac{1-\sqrt{(1-\Delta^2)}}{\Delta}$.

It follows that $\beta_2 \circ \beta_1$ is equal to

(5.3.3-17) $(x_1,x_2,x_3) \longmapsto (\frac{x_1-\sqrt{(1-\Delta^2)}}{1-x_1\sqrt{(1-\Delta^2)}}, \frac{\Delta x_2}{1-x_1\sqrt{(1-\Delta^2)}}, \frac{\Delta x_3}{1-x_1\sqrt{(1-\Delta^2)}})$.

Let $z = y_1 + iy_2$. Using (5.3.3-1), (5.3.3-17) and then (5.3.3-1a) one obtains a mapping from $\mathbb{C} \to \mathbb{C} \cup \{\infty\}$. Direct computation shows that this is equal to ϕ restricted to $\mathbb{C}$.

Now consider the point at infinity, i.e. (0,0,-1). On the one hand

(5.3.3-18) $\phi(\infty) = -\frac{\sqrt{(1+\Delta)}}{\sqrt{(1-\Delta)}} = -\frac{\sqrt{(1-\Delta^2)}}{(1-\Delta)}$.

On the other hand, from (5.3.3-17) it follows that

$$\beta_2 \circ \beta_1(0,0,-1) = (-\sqrt{(1-\Delta^2)},0,-\Delta),$$

which corresponds, according to (5.3.3-1a), to the point

$$-\frac{\sqrt{(1-\Delta^2)}}{1-\Delta} \in \mathbb{C}.$$

Q.E.D.

Remarks

(i) It follows from the theorem, that ϕ moves points closer to -1, the larger $\lambda = \frac{1+\sqrt{(1-\Delta^2)}}{\Delta}$ is, i.e. the smaller $\Delta \in (0,1]$ is, except for the fixed point +1. It is intuitively clear that for small values of $\Delta > 0$, the inverse of ϕ will behave bad numerically. Therefore it appears to be preferable to use only the rotations mentioned before, to test stability. It is also understandable why in the standard discrete time representation of a stable system with small time interval length, the roots will tend to be close to one, because then ϕ^{-1} is involved. (This can also be seen directly from $\Psi^{-1}(w) = \Delta w + \sqrt{(1-\Delta^2)}$). (Compare [Goo 85]).

(ii) Note that, just as the rotations ρ, the mappings ϕ form part of an abelian group. The group multiplication corresponds to the multiplication of the corresponding factors λ, or equivalently, to the addition of the corresponding terms $\beta := \ell n\lambda$.

Note that

$$(5.3.3\text{-}19)\quad \beta = \ln\left(\Delta^{-1}+\sqrt{(\Delta^{-2}-1)}\right) = \operatorname{arccosh}(\Delta^{-1}) \geq 0.$$

With α and β as defined above, one can write

$$\rho(z) = \rho_\alpha(z) = \frac{z.\cos(\frac{\alpha}{2})-\sin(\frac{\alpha}{2})}{z.\sin(\frac{\alpha}{2})+\cos(\frac{\alpha}{2})}, \text{ with } \frac{\alpha}{2} = \operatorname{arctg}\left(\frac{\sqrt{(1-\Delta)}}{\sqrt{(1+\Delta)}}\right),$$

and

$$\phi(z) = \phi_\beta(z) = \frac{z.\cosh(\frac{\beta}{2}) - \sinh(\frac{\beta}{2})}{-z.\sinh(\frac{\beta}{2})+\cosh(\frac{\beta}{2})}, \text{ with } \frac{\beta}{2} = \operatorname{arctgh}\left(\frac{\sqrt{(1-\Delta)}}{\sqrt{(1+\Delta)}}\right),$$

and one has $\rho_\alpha \circ \rho_{\alpha'} = \rho_{\alpha+\alpha'}$ and $\phi_\beta \circ \phi_{\beta'} = \phi_{\beta+\beta'}$. These properties still hold if we allow $\alpha \in \mathbb{R}$, $\beta \in \mathbb{R}$! In that case the sets $\{\rho_\alpha | \alpha \in \mathbb{R}\}$ and $\{\phi_\beta | \beta \in \mathbb{R}\}$ form abelian groups.

5.3.4. The norm of systems with time interval length Δ

In this section we want to define the norm of a linear system with time interval length $\Delta \in [0,1]$, which corresponds to the norms for the cases $\Delta = 0$ and $\Delta = 1$ already defined. Furthermore we will show how this norm can be computed by a generalized Lyapunov equation that interpolates between the continuous and the discrete time case. And we will give a representation of the norm squared as a line integral of the Frobenius norm squared of the transfer matrix over the boundary of the stability region. This is a generalization of the formulas given before for the continuous and discrete time case.

To derive the norm, we start with a somewhat more general situation. Consider the following question: What happens to the norm of a functional on ℓ^2 if there is a transformation of the argument $t' = \Delta.t$? (a 'time contraction'). To be more specific, let the functional be given by

$$F:\ u \longmapsto y = \langle f,u\rangle = \int f(t)u(t)dt,$$

with f fixed and $f,u \in L^2(\mathbb{R})$. Now consider the mapping $\Phi: L^2 \to L^2$ given by $v(t) \longmapsto w(s) = v(\Delta^{-1}.s)$. This is not an isometry, because

$$\|\Phi(v)\|^2 = \|w\|^2 = \int w(s)^2 ds = \int v(\Delta^{-1}s)^2 ds = \int v(t)^2 \Delta dt = \Delta \|v\|^2.$$

Now consider the functional F. It is mapped to $F \circ \Phi^{-1}$, and so the corresponding element $f' \in L^2$ is determined by

$$\langle f,v\rangle = \langle f,\Phi^{-1}(w)\rangle = \langle f',w\rangle,$$

which implies

$$\int f(t)v(t)dt = \int f'(s)\, v(\Delta^{-1}s)ds = \int f'(\Delta t)\, v(t)\, \Delta\, dt.$$

Therefore $f'(\Delta t) = \Delta^{-1} f(t)$, and so

$$\|f'\|^2 = \int f'(s)^2 ds = \Delta^{-2} \int f(t)^2\, \Delta\, dt = \Delta^{-1} \int f(t)^2 dt,$$

so

(5.3.4-1) $$\|f'\|^2 = \Delta^{-1} \|f\|^2.$$

A linear system with one output can be considered as a collection of L^2-functionals (one for each time instance) on the inputs, both in the continuous and in the discrete time case. In the continuous time case this is clear, in the discrete time case, one has to embed ℓ^2 into L^2 by using step functions: Map each element $\{f(n)\}$ of ℓ^2 to f in L^2, defined by $f(s) = f([s])$, [s] denoting the entier of s. Using similar arguments as those that led to (5.3.4-1) one arrives at the following definition

5.3.4-2. Definition. Let Σ be the system given by (5.3.2-3) and equivalently by (5.3.2-5). Suppose it is asymptotically stable (i.e. (5.3.2-4) holds). Then its norm $\|\Sigma\|$ will be given by

(5.3.4-3) $$\|\Sigma\|^2 = \Delta^{-1} \sum_{k=o}^{\infty} \|CA^k B\|_F^2 + \|D\|_F^2,$$

where $\Delta \in (0,1]$ (and $\|,\|_F$ denotes the Frobenius norm of a matrix). Similarly, the inner product of systems Σ_1 and Σ_2, represented by

(A_1,B_1,C_1,D_1), resp. (A_2,B_2,C_2,D_2), is given by

$$(5.3.4\text{-}3a)\quad \langle \Sigma_1,\Sigma_2\rangle = \frac{1}{\Delta}\,\mathrm{tr}\sum_{k=o}^{\infty} C_1A_1^kB_1B_2^T(A_2^T)^kC_2^T + \mathrm{tr}\ D_1D_2^T.$$

The corresponding distance is given by the formula

$$(5.3.4\text{-}3b)\quad \|\Sigma_1-\Sigma_2\|^2 = \|\Sigma_1\|^2 + \|\Sigma_2\|^2 - 2\langle \Sigma_1,\Sigma_2\rangle.$$

Of course (5.3.4-3a) can be expressed as

$$(5.3.4\text{-}3c)\quad \langle \Sigma_1,\Sigma_2\rangle = \mathrm{tr}\ C_1L_{12}C_2^T,$$

where L_{12} is the solution of the Lyapunov equation

$$(5.3.4\text{-}3d)\quad L_{12} - A_1L_{12}A_2^T = \Delta B_1B_2^T$$

In terms of the matrices $(A_\Delta, B_\Delta, C, D)$ of the δ_Δ-representation of the system, cf. (5.3.2-5), the norm is given by

$$(5.3.4\text{-}4)\quad \|\Sigma\|^2 = \Delta\sum_{k=o}^{\infty} \|C\big(\Delta A_\Delta+\sqrt{(1-\Delta^2)}I\big)^kB_\Delta\|_F^2 + \|D\|_F^2,$$

because $A = \Delta A_\Delta + \sqrt{(1-\Delta^2)}I$, and $B = \Delta B_\Delta$.
We now want to show how (5.3.4-4) can be computed directly in terms of the matrices (A_Δ,B_Δ,C,D) using a generalized Lyapunov equation. For simplicity take $D = 0$. Then one has

$$(5.3.4\text{-}5)\quad \|\Sigma\|^2 = \Delta\,\mathrm{tr}\ C\,L_A(B_\Delta B_\Delta^T)C^T = \mathrm{tr}\ C\,L_A(\Delta B_\Delta B_\Delta^T)C^T.$$

L_A is the solution L of the Lyapunov equation

$$(5.3.4\text{-}6)\quad L - ALA^T = \Delta B_\Delta B_\Delta^T.$$

Substitution of $A = \Delta A_\Delta + \sqrt{(1-\Delta^2)}I$ gives

$$(5.3.4\text{-}7)\quad L - \Delta^2A_\Delta LA_\Delta^T - \Delta A_\Delta L\sqrt{(1-\Delta^2)} - \Delta LA_\Delta^T\sqrt{(1-\Delta^2)} - (1-\Delta^2)L = \Delta B_\Delta B_\Delta^T.$$

Dividing by Δ, and grouping certain terms together, one obtains

(5.3.4-8) $\Delta(L-A_\Delta LA_\Delta^T) + \sqrt{(1-\Delta^2)}(-A_\Delta L-LA_\Delta^T) = B_\Delta B_\Delta^T$

which we will call the generalized (Δ-) Lyapunov equation. If $\Delta = 1$, this is the discrete time Lyapunov equation, while if $\Delta = 0$ it is the continuous time Lyapunov equation.

5.3.4-9. Remark. Note that to obtain a formula for L of the form

$$\alpha(L-\tilde{A}\, L\tilde{A}^T) + \beta(-\tilde{A}\, L-L\tilde{A}^T) = B_\Delta B_\Delta^T$$

using $A = c\tilde{A} + dI$, it is essential to take $\alpha = c = \Delta$, $\beta = d = \sqrt{(1-\Delta^2)}$, i.e. $\tilde{A} = A_\Delta$, so it is essential to use the δ_Δ-operator to obtain such an interpolating result!

If we let x denote the matrix operator, defined by

$$x: \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},\ x(L) = A_\Delta L,$$

and similarly

$$y: \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n},\ y(L) = LA_\Delta^T,$$

then the Lyapunov equation (5.3.4-8) can be written in the following concise form

(5.3.4-10) $\operatorname{Re}\, e^{i\alpha}(1+ix)(1+iy)\,(L) = B_\Delta B_\Delta^T,$

where α is as before (see (5.3.3-9)), $\alpha \in [0,\pi/2]$, the angle of rotation of the Riemann sphere (i.e. $e^{i\alpha} = \Delta + i\sqrt{(1-\Delta^2)}$). (This formula results from a simple calculation). It can be rewritten as

(5.3.4-11) $\operatorname{Re}\{e^{i\frac{\alpha}{2}}(1+ix)\}\{e^{i\frac{\alpha}{2}}(1+iy)\}(L) = B_\Delta B_\Delta^T$

(using $\operatorname{Re}(x) = x$, $\operatorname{Re}(y) = y$). Using $(\sqrt{2})e^{i\frac{\alpha}{2}} = \sqrt{(1+\Delta)}+i\sqrt{(1-\Delta)}$, one obtains:

(5.3.4-12)
$$\begin{aligned}&\operatorname{Re}\{\sqrt{(1+\Delta)}-x\sqrt{(1-\Delta)}+i\big(\sqrt{(1-\Delta)}+x\sqrt{(1+\Delta)}\big)\} \times \\ &\{\sqrt{(1+\Delta)}-y\sqrt{(1-\Delta)}+i\big(\sqrt{(1-\Delta)}+y\sqrt{(1+\Delta)}\big)\}(L) = 2B_\Delta B_\Delta^T\end{aligned}$$

Calculating the real part, and putting the result back in matrix form, instead of using x and y, gives the following form for the Lyapunov equation:

$$
\begin{aligned}
&[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta]L[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta^T] + \\
&- [\sqrt{(1-\Delta)}I+\sqrt{(1+\Delta)}A_\Delta]L[\sqrt{(1-\Delta)}I+\sqrt{(1+\Delta)}A_\Delta^T] = 2B_\Delta B_\Delta^T.
\end{aligned}
\tag{5.3.4-13}
$$

Now let $\tilde{A}_\Delta = \rho^{-1}(A_\Delta) = [\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta]^{-1}[\sqrt{(1-\Delta)}I+\sqrt{(1+\Delta)}A_\Delta]$.

We know that A_Δ is δ_Δ-asymptotically stable iff $\rho^{-1}(A_\Delta)$ has its spectrum in the open unit disk. Assuming A_Δ to be δ_Δ-asymptotically stable it follows from (5.3.4-13) that

(5.3.4-14)

$$
L - \rho^{-1}(A_\Delta)L\rho^{-1}(A_\Delta)^T = 2[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta]^{-1}B_\Delta B_\Delta^T[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta^T]^{-1}.
$$

Remark. In the next subsection this Lyapunov equation will be used to derive results about isometries of the spaces we are studying. Formula (5.3.4-14) can be compared with (5.3.4-6). If $\Delta > 0$ is small $A = \Delta A_\Delta + \sqrt{(1-\Delta^2)}I$ will in practice be close to the identity matrix. Therefore (5.3.4-6) may behave bad numerically, while (5.3.4-14) may not have this problem. In terms of A and B this equation can be written as

$$
L - \phi(A)L\phi(A^T) = 2[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A)]^{-1}BB^T[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A^T]^{-1}. \tag{5.3.4-15}
$$

We know that if $\Delta \downarrow 0$, then $\phi(A) \to -I$ if $\sigma(A) \underline{\in} D(0,1)$. Therefore, if $\Delta > 0$ is taken too small, one should also expect numerical difficulties in this Lyapunov equation. In practice, if there is a freedom in the choice of $\Delta > 0$, it appears to be advisable to choose it such that the real parts of the spectrum of $\phi(A)$ span an interval that lies symmetrically around the origin. In the next subsection this Lyapunov equation (5.3.4-15) will lead to an interesting group that will be used to derive auto-isometries of the metric spaces we are investigating.

To conclude this subsection we will derive a representation of the norm squared as a line integral (for $\delta > 0$ a contour integral) over the boundary

curve of the domain of stability. As before we take D = 0 for simplicity. Let

(5.3.4-16) $T(s) = C(sI-A)^{-1}B$

and

(5.3.4-17) $T_\Delta(s) = C(sI-A_\Delta)^{-1}B_\Delta$,

where A,B,C and A_Δ, B_Δ are as before (see (5.3.2-3) and (5.3.2-5)).

5.3.4-18. Proposition. Let $\Delta \in (0,1]$, and let $z = \Psi(v)$, then $T(v) = T_\Delta(z)$.

Proof.
$$T(v) = C(vI-A)^{-1}B = C(\Psi^{-1}(z)I-A)^{-1}B =$$
$$C[\{\Delta z+\sqrt{(1-\Delta^2)}\}I - \{\Delta A_\Delta+\sqrt{(1-\Delta)^2}I\}]^{-1}B =$$
$$= C(zI-A_\Delta)^{-1}B_\Delta = T_\Delta(z). \qquad \text{Q.E.D.}$$

Let K_Δ be the oriented boundary of the stability region. It is described in detail as follows. It is equal to the counterclockwise oriented circle $K_\Delta = C(-\sqrt{(\Delta^{-2}-1)},\Delta^{-1})$ if $\Delta > 0$, and equal to the oriented imaginary axis $K_o = (-i\infty,+i\infty)$ if $\Delta = 0$. (On the Riemann sphere K_o is also a circle).

5.3.4-19. Theorem. The norm of the system Σ, given by (5.3.2-3), or equivalently by (5.3.2-5), satisfies the following formula

$$\|\Sigma\|^2 = \frac{1}{2\pi i}\,\mathrm{tr}\int_{K_\Delta} T_\Delta(z)T_\Delta(z)^*\frac{dz}{\Delta z+\sqrt{(1-\Delta^2)}}.$$

Proof. Let $\Delta \in (0,1]$. From definition (5.3.4-3) and formula (5.2-13) one has

$$\|\Sigma\|^2 = \frac{\Delta^{-1}}{2\pi i}\,\mathrm{tr}\int_{K_1} T(v)T(v)^*\frac{dv}{v}.$$

Now apply the substitution $v = \Psi^{-1}(z)$, making use of the previous proposition. This gives

$$\|\Sigma\|^2 = \frac{\Delta^{-1}}{2\pi i}\,\mathrm{tr}\int_{\Psi(K_1)} T_\Delta(z)T_\Delta(z)^*\Delta\frac{dz}{\Psi^{-1}(z)}$$

$$= \frac{1}{2\pi i}\mathrm{tr}\int_{K_\Delta} T_\Delta(z)T_\Delta(z)^*\frac{dz}{\Delta z+\sqrt{(1-\Delta^2)}}.$$

For the case $\Delta = 0$, the formula to be proved is equal to formula (5.2-14) (with $D = 0$). Q.E.D.

Just as in the discrete- and continuous-time case, one can also write $\|\Sigma\|^2$ as a line integral over K_Δ:

5.3.4-20. Theorem. The norm of Σ (defined just as in the previous theorem) satisfies the following formula

$$\|\Sigma\|^2 = \frac{1}{2\pi}\mathrm{tr}\int_{K_\Delta} T_\Delta T_\Delta^* d\ell,$$

in the notation of line integrals (compare (5.2-13a) and (5.2-14a)). (Recall K_Δ is oriented).

Proof. For $\Delta = 0$, this is formula (5.2-14a). Now let $\Delta \in (0,1]$. Parametrize K_Δ by arclength ℓ, i.e. let $z(\ell) = \frac{e^{i\Delta\ell} - \sqrt{(1-\Delta^2)}}{\Delta}, \ell \in [0, 2\pi\Delta^{-1}]$. Substituting this in the formula obtained in the previous theorem (5.3.4-19), one obtains:

$$\|\Sigma\|^2 = \frac{1}{2\pi i}\,\mathrm{tr}\int_0^{2\pi/\Delta} T_\Delta(z(\ell))T_\Delta(z(\ell))^* \frac{ie^{i\Delta\ell}d\ell}{e^{i\Delta\ell}} =$$

$$= \frac{1}{2\pi}\,\mathrm{tr}\int_0^{2\pi/\Delta} T_\Delta(z(\ell))T_\Delta(z(\ell))^* d\ell.$$

Because ℓ is the arclength parameter this can be written in the shorter notation of line integrals as

$$\|\Sigma\|^2 = \frac{1}{2\pi}\,\mathrm{tr}\int_{K_\Delta} T_\Delta T_\Delta^* d\ell$$

Q.E.D.

5.3.4-21. Remark. It is perhaps interesting to note that there is also a representation as a contour integral on the Riemann sphere, namely

$$\|\Sigma\|^2 = \frac{1}{4\pi}\int_{K_\Delta} \|\big(\sqrt{(1+\Delta)}z(r)+\sqrt{(1-\Delta)}\big)T_\Delta(z(r))\|_F^2 dr,$$

dr being the line element on the Riemann sphere. This can be shown as follows.

On K_Δ one has $|\sqrt{(1+\Delta)}z+\sqrt{(1-\Delta)}|^2 = 1+|z|^2$. (This can be verified by straightforward calculation, using $z = (v-\sqrt{(1-\Delta^2)})\Delta^{-1}, |v| = 1$).

Furthermore $(\frac{1+|z|^2}{2})dr = ds$, where dr is the line element on the Riemann sphere, and ds the line element in $\mathbb{C}$. (Compare [Ne-P], p.58, ex.30).

5.4. Some results on isometries

Let $M^{m,\Delta}_{m',n,m}$ denote the space of all systems in $M^m_{m',n,m}$, which are δ_Δ-asymptotically stable, i.e. (5.3.2-8) holds. Let

$$L^{m,\Delta}_{m',n,m} := \{(A,B,C) \in \mathbb{R}^{n^2} \times \mathbb{R}^{nm'} \times \mathbb{R}^{mn} | [(A,B,C)] \in M^{m,\Delta}_{m',n,m}\}$$

denote the corresponding principal bundle of the state bundle, as usual (cf. chapter 4). From the Lyapunov equation (5.3.4-14) it will be derived that the metric spaces $M^{m,\Delta}_{m',n,m}$ with m',n,m fixed, are isometric for all values of $\Delta \in [0,1]$. In particular the continuous time case (Δ=0) and the discrete time case (Δ=1) are isometric. This all holds both for the 'outer' metric and for the Riemannian metric. To be unambiguous let $\|,\|_\Delta$ denote the norm defined in (5.3.4-4), and let g_Δ denote the corresponding Riemannian metric of $M^{m,\Delta}_{m',n,m}$. Now consider the mapping

$$\begin{aligned} &P_\Delta : L^{m,1}_{m',n,m} \to L^{m,\Delta}_{m',n,m}, \\ &(A_1,B_1,C_1) \longmapsto (\rho_\Delta(A_1), \sqrt{2}[\sqrt{(1+\Delta)}I+\sqrt{(1-\Delta)}A_1]^{-1}B_1, C_1). \end{aligned} \tag{5.4-1}$$

It has to be shown that this mapping is well-defined, i.e. that

(i) if (A_1,B_1,C_1) is minimal then the same holds for $(\rho_\Delta(A_1), \sqrt{2}[\sqrt{(1+\Delta)}I+\sqrt{(1-\Delta)}A_1]^{-1}B_1, C_1)$, and

(ii) if (A_1,B_1,C_1) is δ_1-asymptotically stable, then $P_\Delta(A_1,B_1,C_1)$ is δ_Δ-asymptotically stable.

ad(i) If (A_1,B_1,C_1) is minimal, then it is reachable and observable. It will be shown that the same holds for $P_\Delta(A_1,B_1,C_1)$. The mapping $A_1 \longmapsto \rho(A_1)$ leaves the set of eigenvectors fixed. The well-known PBH-test can now be applied to show reachability and observability: Suppose $(A_\Delta,B_\Delta,C_\Delta) := P_\Delta(A_1,B_1,C_1)$ would be non-reachable. Then, according to the PBH-test, there would be a nonzero vector q such that $q^T A_\Delta = \lambda q^T$ and $q^T B_\Delta = 0$ for some $\lambda \in \mathbb{C}$. It would then

follow that

$$q^T A_1 = \rho_\Delta^{-1}(\lambda)q$$

and

$$\frac{\sqrt{2}}{\sqrt{(1+\Delta)}+\sqrt{(1-\Delta)}} q^T B_1 = 0, \text{ so } q^T B_1 = 0,$$

i.e. (A_1,B_1) would be non-reachable (PBH), which contradicts the assumption. So (A_Δ,B_Δ) is reachable. Similarly it can be shown that (A_Δ,C_Δ) is observable, and so $(A_\Delta,B_\Delta,C_\Delta)$ is minimal.

ad(ii) From the properties of ρ_Δ and ρ_Δ^{-1} (rotations of the Riemann sphere), it follows that if A_1 is δ_1-asymptotically stable then $A_\Delta = \rho_\Delta(A_1)$ is δ_Δ-asymptotically stable. Also it follows that $[\sqrt{(1+\Delta)}I+\sqrt{(1-\Delta)}A_1]$ and $[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta]$ are invertible.

It is straightforward to show that P_Δ is compatible with base change of the state space. I.e. if T is nonsingular and $(A_\Delta,B_\Delta,C_\Delta) = P_\Delta(A_1,B_1,C_1)$, then

$$P_\Delta(TA_1T^{-1},TB_1,C_1T^{-1}) = (TA_\Delta T^{-1},TB_\Delta,C_\Delta T^{-1}).$$

Therefore P_Δ induces a mapping $\bar{P}_\Delta$ on the orbit space:

$$\bar{P}_\Delta: M^{m,l}_{m',n,m} \to M^{m,\Delta}_{m',n,m'}$$

$$[(A_1,B_1,C_1)] \longmapsto [P_\Delta(A_1,B_1,C_1)].$$

The inverse of P_Δ is given by

$$P_\Delta^{-1}: L^{m,\Delta}_{m',n,m} \to L^{m,l}_{m',n,m'}$$

(5.4-2)

$$(A_\Delta,B_\Delta,C_\Delta) \longmapsto (\rho_\Delta^{-1}(A_\Delta),\sqrt{2}[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_\Delta]^{-1}B_\Delta,C_\Delta).$$

Recall that $\rho_\Delta^{-1}(z) = \big(\sqrt{(1+\Delta)}z+\sqrt{(1-\Delta)}\big)/\big(-\sqrt{(1-\Delta)}z+\sqrt{(1+\Delta)}\big)$. Of course $\bar{P}_\Delta$ is also invertible and its inverse is given by $[(A_\Delta,B_\Delta,C_\Delta)] \longmapsto [P_\Delta^{-1}(A_\Delta,B_\Delta,C_\Delta)]$. Now we can formulate the following theorem.

5.4-3. Theorem [isometry theorem]:
(a) $\forall\Delta \in [0,1]$ the spaces $M^{m,\Delta}_{m',n,m}$ with metric induced by the norm $\|,\|_\Delta$ are isometric. In particular this holds for the continuous and discrete time case: $M^{m,\ell}_{m',n,m}$ and $M^{m,a}_{m',n,m}$, with the 'outer metric' are isometric.
(b) $\forall\Delta \in [0,1]$ the metric spaces with Riemannian metric, $(M^{m,\Delta}_{m',n,m},g_\Delta)$, are isometric. Again, in particular this holds for the continuous- and discrete time case.
(c) $\forall\Delta \in [0,1]$, $\bar{P}_\Delta : M^{m,1}_{m',n,m} \to M^{m,\Delta}_{m',n,m}$ is an isometry, both in the sense of (a) and (b).
Proof. Similarly to the derivation of (5.3.4-14) one can derive that the inner product of two δ_Δ-systems, $\Sigma_{\Delta 1}$ and $\Sigma_{\Delta 2}$, represented by $(A_{\Delta 1},B_{\Delta 1},C_{\Delta 1})$, resp. $(A_{\Delta 2},B_{\Delta 2},C_{\Delta 2})$, is

$$(5.4\text{-}4) \qquad \langle\Sigma_1,\Sigma_2\rangle = \operatorname{tr} C_{\Delta 1}L_{12}C_{\Delta 2}^T$$

where L_{12} is the solution of the Lyapunov equation:

(5.4-5)

$$L_{12} - \rho_\Delta^{-1}(A_{\Delta 1})L_{12}\rho_\Delta^{-1}(A_{\Delta 2}^T) = 2[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_{\Delta 1}]^{-1}B_{\Delta 1}B_{\Delta 2}^T[\sqrt{(1+\Delta)}I-\sqrt{(1-\Delta)}A_{\Delta 2}^T]^{-1}$$

It follows from this equation (which is a generalization of (5.3.4-14)) that P_Δ^{-1} leaves L_{12} invariant, and therefore it leaves the inner product invariant:

$$(5.4\text{-}6) \qquad \langle\Sigma_{\Delta 1},\Sigma_{\Delta 2}\rangle = \langle P_\Delta^{-1}(\Sigma_{\Delta 1}),P_\Delta^{-1}(\Sigma_{\Delta 2})\rangle.$$

Consequently, it leaves the induced outer metric invariant:

$$(5.4\text{-}7) \qquad \|\Sigma_{\Delta 1}-\Sigma_{\Delta 2}\| = \|P_\Delta^{-1}(\Sigma_{\Delta 1}) - P_\Delta^{-1}(\Sigma_{\Delta 2})\|.$$

This proves (c).
But (c) implies (a): an isometry between $M^{m,\Delta}_{m',n,m}$ and $M^{m,\Delta'}_{m',n,m}$ is given by

$$P_{\Delta'}oP_\Delta^{-1} : M^{m,\Delta}_{m',n,m} \to M^{m,\Delta'}_{m',n,m}.$$

Because the inner metric of a set depends solely on the outer metric, it is immediate that also the inner metric remains invariant under these mappings.

Q.E.D.

5.4-8. Remarks. (i) In the notation of Remark (5.3.3-18) (iii), one has that P_Δ is given by

$$(A_1,B_1,C_1)\longmapsto(\rho_\alpha(A_1),(\cos(\tfrac{\alpha}{2})I + \sin(\tfrac{\alpha}{2})A_1)^{-1}B_1,C_1).$$

and P_Δ^{-1} by

$$(A_\Delta,B_\Delta,C_\Delta)\longmapsto(\rho_\alpha^{-1}(A_1),(\cos(-\tfrac{\alpha}{2})I + \sin(\tfrac{-\alpha}{2})A_\Delta)^{-1}B_\Delta,C_\Delta).$$

Therefore, by abuse of notation we will also denote P_Δ by P_α, where Δ is then given by $\Delta = \cos(\alpha)$ (see (5.3.3-9)), and P_Δ^{-1} by $P_{-\alpha}$.

(ii) Because of the isometry between the different spaces for different values of $\Delta \in [0,1]$, it is sufficient to investigate only one of them. In fact we will restrict ourselves to the continuous time case and/or the discrete-time case, and it is to be understood that for every result about the geometry of these cases, a similar result holds for all other cases.

(iii) The fact that the controllability Grammian (i.e. the solution L of the Lyapunov equation (5.3.4-14) for $\Delta = 1$) remains invariant under the mapping P_Δ with $\Delta = 0$ was noted before by [Glo 84], section 2.2, especially (2.23) and (2.24).

To repeat, to study the geometry we can confine ourselves to one case, let us say the discrete time case (for the moment). In the previous theorem we found isometries between pairs of different spaces. We now want to look for isometric maps of the space $M_{m',n,m}^{m,a}$ onto itself. Consider equation (5.3.4-15) and recall that the matrix L in that equation is defined as the solution of (5.3.4-6), or equivalently of the equation

$$L - ALA^T = \Delta^{-1}BB^T. \tag{5.4-9}$$

Let $L_1 := \Delta.L$. Then L_1 is the solution of

$$L_1 - AL_1A^T = BB^T. \tag{5.4-10}$$

By multiplying both sides of equation (5.3.4-15) with Δ, one can see that L_1 is also the solution of

$$
\begin{aligned}
&L_1 - \phi_\Delta(A) L_1 \phi_\Delta(A^T) = \\
&2[\sqrt{(\Delta^{-1}+1)}I - \sqrt{(\Delta^{-1}-1)}A]^{-1} BB^T [\sqrt{(\Delta^{-1}+1)}I - \sqrt{(\Delta^{-1}-1)}A^T]^{-1},
\end{aligned}
\tag{5.4-11}
$$

where we write ϕ_Δ instead of ϕ to stress the dependence of ϕ on $\Delta \in [0,1]$. Similar to the mapping P_Δ in the previous theorem, we now define for each $\Delta \in (0,1]$ the mapping

$F_\Delta: L^{m,a}_{m',n,m} \to L^{m,a}_{m',n,m}$, given by

$$(A,B,C) \longmapsto (\phi_\Delta(A), \sqrt{2}[\sqrt{(\Delta^{-1}+1)}I - \sqrt{(\Delta^{-1}-1)}A]^{-1}B, C).$$

Its inverse is given by

$$F_\Delta^{-1}: (\tilde{A},\tilde{B},\tilde{C}) \longmapsto (\phi_\Delta^{-1}(\tilde{A}), \sqrt{2}[\sqrt{(\Delta^{-1}+1)}I + \sqrt{(\Delta^{-1}-1)}\tilde{A}]^{-1}\tilde{B}, \tilde{C}).$$

There are corresponding mappings $\bar{F}_\Delta, \bar{F}_\Delta^{-1}: M^{m,a}_{m',n,m} \to M^{m,a}_{m',n,m}$.
Now, in this case, the number $\Delta \in (0,1]$ is no longer interpreted as a time interval length. Therefore, we will change the notation: instead of subindex Δ we will use β, where

$$\Delta^{-1} = \cosh(\beta)$$

(compare (5.3.3-19) ff). Then

$$\sqrt{\left(\frac{\Delta^{-1}+1}{2}\right)} = \cosh\left(\frac{\beta}{2}\right), \sqrt{\left(\frac{\Delta^{-1}-1}{2}\right)} = \sinh\left(\frac{\beta}{2}\right).$$

As before, we have

$$\phi_\beta(z) = \frac{z.\cosh\left(\frac{\beta}{2}\right) - \sinh\left(\frac{\beta}{2}\right)}{-z.\sinh\left(\frac{\beta}{2}\right) + \cosh\left(\frac{\beta}{2}\right)},$$

and we will allow negative values of β. The ϕ_β's form an abelian group under composition, with $\phi_\beta \circ \phi_{\beta'} = \phi_{\beta+\beta'}$, and $\phi_\beta^{-1} = \phi_{-\beta}$.
We will now show a similar property for the F_β's:

5.4-12. Lemma. The set $\{F_\beta \mid \beta \in \mathbb{R}\}$ forms an abelian group, with $F_\beta \circ F_{\beta'} = F_{\beta+\beta'}$, and $F_o = id$.

Proof. $F_o = id$. is trivial. Consider

$$F_\beta \circ F_{\beta'}(A,B,C) = F_\beta(\phi_{\beta'}(A),(\cosh(\tfrac{\beta'}{2})I-\sinh(\tfrac{\beta'}{2})A)^{-1}B,C).$$

It will be necessary to compute

$$[\cosh(\tfrac{\beta}{2})I-\sinh(\tfrac{\beta}{2})\phi_{\beta'}(A)]^{-1}(\cosh(\tfrac{\beta'}{2})-\sinh(\tfrac{\beta'}{2})A)^{-1} =$$

$$(\cosh(\tfrac{\beta}{2})\cosh(\tfrac{\beta'}{2})I-\cosh(\tfrac{\beta}{2})\sinh(\tfrac{\beta'}{2})A +$$

$$+ \sinh(\tfrac{\beta}{2})\sinh(\tfrac{\beta'}{2})I-\sinh(\tfrac{\beta}{2})\cosh(\tfrac{\beta}{2})A)^{-1} =$$

$$(\cosh(\tfrac{\beta+\beta'}{2})I-\sinh(\tfrac{\beta+\beta'}{2})A)^{-1}.$$

It follows that

$$F_\beta(\phi_{\beta'}(A),(\cosh(\tfrac{\beta'}{2})I-\sinh(\tfrac{\beta'}{2})A)^{-1}B,C) =$$

$$= (\phi_{\beta+\beta'}(A), [\cosh(\tfrac{\beta+\beta'}{2})I-\sinh(\tfrac{\beta+\beta'}{2})A]^{-1}B,C).$$

Q.E.D.

From the Lyapunov equations (5.4-10) and (5.4-11) it follows that their solution L_1 is invariant under application of F_β, for any $\beta \in \mathbb{R}$. This implies that the F_β's are isometries. There are some other rather obvious isometries, namely

(5.4-13) $(A,B,C) \longmapsto (-A,B,C)$.

and

(5.4-14a) $(A,B,C) \longmapsto (A,BQ,C)$, and

(5.4-14b) $(A,B,C) \longmapsto (A,B,RC)$

with Q and R square orthogonal matrices of the correct sizes, and

(5.4-15) $(A,B,C)\longmapsto(A^T,C^T,B^T)$, 'dualization'.

It is clear that (5.4-13) and (5.4-14a) are isometries, because they leave the controllability Grammian L_1 and C invariant. That (5.4-15) is an isometry follows from the properties of the trace, and (5.2-8). And (5.4-14b) can be obtained by applying (5.4-15), (5.4-14a) and (5.4-15) again.
Formula (5.4-13) corresponds to a conformal mapping of the unit disk onto itself, namely $\phi^-(z) = -z$, just as the ϕ_β. Let

(5.4-16) $\phi_\beta^-(z) := -\phi_\beta(z)$. Then we have

5.4-17. Proposition. (a) $\{\phi_\beta \mid \beta \in \mathbb{R}\} \cup \{\phi_\beta^- \mid \beta \in \mathbb{R}\}$ forms the group of all automorphisms of the unit disk, that leave the real interval [-1,1] invariant.
(b) The group composition is given by

(i) $\phi_\beta \circ \phi_{\beta'} = \phi_{\beta+\beta'}$,

(ii) $\phi_\beta \circ \phi_{\beta'}^- = \phi_{\beta'-\beta}^-$,

(iii) $\phi_\beta^- \circ \phi_{\beta'}^- = \phi_{\beta'-\beta}$,

(iv) $\phi_\beta^- \circ \phi_{\beta'} = \phi_{\beta+\beta'}^-$.

Proof. All automorphisms of the unit disk can be written in the form

(5.4-18) $\phi(z) = \dfrac{az+b}{\bar{b}z+\bar{a}}$, with $|a|^2 - |b|^2 = 1$

(cf. [Ne-P], p. 48).
If it is required that ϕ maps reals into reals, it can be shown easily that a and b must lie both on the real axis, or both on the imaginary axis. In the first case ϕ can be written as

(5.4-19) $\phi(z) = \phi_\beta(z)$ with $a = \cosh(\frac{\beta}{2})$, and $b = -\sinh(\frac{\beta}{2})$,

and in the second case ϕ can be written as

(5.4-20) $\phi(z) = \bar{\phi}_\beta(z)$, with $a = i.\cosh(\frac{\beta}{2})$ and $b = -i.\sinh(\frac{\beta}{2})$.

Part (b) of the proposition follows by straightforward calculation, using the sum formulas for cosh and sinh.

Q.E.D.

5.4-21. Remark. From part (b) of the proposition it follows that the group $\{\phi_\beta\} \cup \{\bar{\phi}_\beta\}$ is not abelian.
Now let $\bar{F}_o$ denote the mapping given by (5.4-13) and let $\bar{F}_\beta := \bar{F}_o \circ F_\beta$, then the following holds:

5.4-22. Theorem. Let m',n,m be fixed.
(a) The set $\{\bar{F}_\beta\} \cup \{\bar{\bar{F}}_\beta\}$ of mappings $M^{m,a}_{m',n,m} \to M^{m,a}_{m',n,m}$ forms a group of isometries (with respect to both the outer metric and the inner metric).
(b) The mapping $\{\bar{F}_\beta\} \cup \{\bar{\bar{F}}_\beta\} \to \{\phi_\beta\} \cup \{\bar{\phi}_\beta\}$ given by $\bar{F}_\beta \longmapsto \phi_\beta$ and $\bar{\bar{F}}_\beta \longmapsto \bar{\phi}_\beta$, for all $\beta \in \mathbb{R}$, is a group isomorphism.

Proof. That the $\bar{F}_\beta$ and the $\bar{\bar{F}}_\beta$ are isometries has already been noted. That $\{\bar{F}_\beta\} \cup \{\bar{F}'_\beta\}$ is group-isomorphic to $\{\phi_\beta\} \cup \{\bar{\phi}_\beta\}$ follows from a simple computation similar to the one given in the proof of lemma (5.4-12).

Q.E.D.

5.4-23. Remarks (i) The isometries of this theorem commute with those of (5.4-14) and (5.4-15) etc. Therefore we have in fact a group of isometries formed by the direct sum of the isometries of (5.4-21), those of (5.4-14) and those of (5.4-15),
(ii) The isometries found extend to isometries on the ℓ^2 space of Markov-matrix-sequences. This can be seen as follows: any such isometry is a linear operator on the Markov matrix sequences, and can therefore be represented by an infinite matrix. The matrix is the same for all McMillan degrees, and by completion it follows that the matrix also represents an isometry on the whole ℓ^2 space of Markov-matrix sequences. (The isometry is in fact an orthogonal linear operator, because it is invertible, unlike for example the shift operator on ℓ^2). The remarkable thing about these ℓ^2 isometries is of course that they leave the McMillan degree invariant!

(iii) All isometries of a space with respect to the outer metric are also isometries of the space with respect to the Riemannian metric, but w.r.t. the Riemannian metric the space may have more isometries. As a simple example, consider a 2-sphere without its equator. The outer metric is the one induced by $\mathbb{R}^3$. W.r.t. this outer metric, the isometries are (a) rotations around the vertical axis; (b) reflection w.r.t. a plane through the vertical axis, and (c) reflection w.r.t. the equator plane. All of these are also isometries w.r.t. the Riemannian metric. But w.r.t. the Riemannian metric there are more isometries,e.g. a rotation of the upper hemisphere w.r.t. the vertical axis, while the lower hemisphere is not rotated but remains fixed etc.
A similar situation occurs in the s.i.s.o. case (m=m'=1), because, as discussed in chapter 4, $M^{m,a}_{l,n,l}$ is not connected.
(iv) A similar group of isometries exists for the Hankel-Frobenius norm and the Hankel (spectral-or operator-) norm. Instead of the group $\{F_\beta\} \cup \{F_\beta^-\}$, one has a group $\{G_\beta\} \cup \{G_\beta^-\}$, where G_β is given by

$$G_\beta: (A,B,C) \longmapsto$$
$$(\phi_\beta(A), (\cosh(\tfrac{\beta}{2})I-\sinh(\tfrac{\beta}{2})A)^{-1}B, C(\cosh(\tfrac{\beta}{2})I-\sinh(\tfrac{\beta}{2})A^T)^{-1}), \tag{5.4-24}$$

and $G_o^- = F_o^-$; $G_\beta^- = G_o^- \circ G_\beta$.
Let, for the moment, L_c denote the controllability Grammian, i.e. the solution of

$$L_c - AL_cA^T = BB^T, \tag{5.4-25}$$

and let L_o be the observability Grammian, i.e. the solution of

$$L_o - A^TL_oA = C^TC. \tag{5.4-26}$$

Then it follows from our previous analysis, that L_c and L_o remain invariant under the mapping G_β! It is not difficult to show (and well-known) that the Hankel-Frobenius norm is given by

$$\|\Sigma\|^2_{HF} = \operatorname{tr} L_cL_o, \tag{5.4-27}$$

and the Hankel (operator) norm is given by

(5.4-28) $\quad \|\Sigma\|_H^2 = \max \sigma(L_c L_o).$

It follows that G_β leaves these norms invariant. As before it then follows that the G_β are isometries.

To conclude this subsection, we want to treat a generalization to the complex case, i.e. we allow the entries of (A,B,C) to be complex numbers. As a rule the definitions for this case are obtained by reading transposition plus conjugation (denoted by upper index *) instead of transposition. We will not work this out in detail; in fact the only point we want to make is, that in the complex case the group of isometries $\{F_\beta\} \cup \{F_\beta^-\}$ can be extended to a group of isometries that is isomorphic to the group of all automorphism of the unit disk in $\mathbb{C}$. (I.e. the requirement that the automorphism should take real numbers to real numbers disappears in this case).
In the complex case, the norm of a system Σ with representation (A,B,C) is given by

(5.4-29) $\quad \|\Sigma\|^2 = \operatorname{tr} C\, L\, C^*,$

where L is the controllability Grammian, i.e. the solution of

(5.4-30) $\quad L - ALA^* = BB^*.$

Just as for the real case one can show that

(5.4-31) $\quad \|\Sigma\|^2 = \int_{C(0,1)} \|T(w)\|_F^2 \frac{dw}{w},$

where $T(w) = C(wI-A)^{-1}B$ is the transfer matrix of the system.
Let

(5.4-32) $\quad \phi(z) = \frac{az+b}{\bar{b}z+\bar{a}},\ a,b \in \mathbb{C},\ |a|^2 - |b|^2 = 1,$

denote an automorphism of the unit disk in $\mathbb{C}$. All such automorphisms can be written in this way. Note that

(5.4-33) $\quad \phi^{-1}(z) = \frac{\bar{a}z-b}{-\bar{b}z+a}.$

Let

(5.4-34) $\quad F_\phi\colon (A,B,C)\longmapsto(\phi(A),(\bar{a}I+\bar{b}A)^{-1}B,C).$

Then

(5.4-35) $\quad F_\phi^{-1} = F_{\phi^{-1}}\colon (A,B,C)\longmapsto(\phi^{-1}(A),(aI-\bar{b}A)^{-1}B,C).$

5.4-36. Lemma. $T_{F_\phi(\Sigma)}(z) = \dfrac{T_\Sigma(\phi^{-1}(z))}{-\bar{b}z+a}.$

Here T_Σ denotes the transfer matrix of the system Σ, etc.

Proof.

$$T_{F_\phi(\Sigma)}(z) = C(zI-\phi(A))^{-1}(\bar{a}I+\bar{b}A)^{-1}B = C(\bar{a}zI+\bar{b}zA-bI-aA)^{-1}B =$$

$$= C\Big(\frac{az-b}{-\bar{b}z+a}\,I - A\Big)^{-1}\frac{B}{(-\bar{b}z+a)} = C(\phi^{-1}(z)I - A)^{-1}B(-\bar{b}z+a) =$$

$$= T_\Sigma(\phi^{-1}(z))/(-\bar{b}z+a).$$

Q.E.D.

5.4-37 Theorem. $\displaystyle\int_{C(0,1)} \|T_\Sigma(w)\|_F^2\,\frac{dw}{w} = \int_{C(0,1)} \|T_{F_\phi(\Sigma)}(z)\|_F^2\,\frac{dz}{z}.$

Proof. Substitute $w = w(z) = \phi^{-1}(z)$ in the integral on the left-hand side. One obtains

$$\int_{C(0,1)} \|T(w)\|_F^2\,\frac{dw}{w} = \int_{C(0,1)} \|T(\phi^{-1}(z))\|_F^2 z\,\frac{w'}{w}\,\frac{dz}{z}.$$

On the unit circle, one has $z = (\bar{z})^{-1}$, so

$$\frac{zw'(z)}{w(z)} = \frac{1}{(-\bar{b}z+a)^2}\,\frac{-\bar{b}z+a}{(\bar{a}z-b)\bar{z}} = \frac{1}{|-\bar{b}z+a|^2}.$$

Now apply the previous lemma. Q.E.D.

So F_ϕ is an isometry indeed.

5.5. Analysis of the Riemannian geometry (inner geometry) in some special, simple cases

In this section the Riemannian geometry (the inner geometry) will be analyzed for some special cases. The most important one (from the point of view of gaining some initial insight) will be the case $M^{m,a}_{1,1,1}$, i.e. the class of asymptotically stable s.i.s.o. systems with McMillan degree 1. For that case we have a complete picture of the Riemannian geometry. Between each pair of points there is one geodesic (if we include the zero system) which can be explicitly computed. Also the length of this geodesic, which is the socalled inner distance of the two points, can be computed explicitly. In more complex cases it is much harder, if not impossible, to find closed form solutions for the geodesic(s) between a pair of points, and the corresponding inner distance. At the end of the section we will also treat two other special cases.

5.5.1. The Riemannian geometry (inner geometry) of $M^{m,a}_{1,1,1}$ and $M^{m,\ell}_{1,1,1}$

It follows from section 5.4 that the continuous time case and the discrete time cases can be treated side by side. In the continuous time case one has

$$(5.5.1\text{-}1)\quad \begin{cases} \dot{x}_t = ax_t + bu_t, \; a < 0, \; b \neq 0, \; c \neq 0, \\ y_t = cx_t \end{cases}$$

with x_t, y_t, u_t, a,b,c all real scalars. We choose a canonical form, by taking

$$(5.5.1\text{-}2)\quad c = 1.$$

So $M^{m,\ell}_{1,1,1}$ is now parametrized by

$\{(a,b)|a < 0, \; b \neq 0\} = \{(a,b)|a < 0, \; b < 0\} \cup \{(a,b)|a < 0, \; b > 0\}$. (This parametrization is parameter identifiable). It consists clearly of two components, each of which is connected (cf. theorem (4.7-1)). Let t denote a tangent vector, represented by $(\dot{a},\dot{b})^T$ in this parametrization. The positive definite quadratic form (i.e. the Riemannian metric tensor) g at the point (a,b) is given by

$$(5.5.1\text{-}3)\qquad g(t,t) = \begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix}^T \begin{bmatrix} \frac{-b^2}{4a^3} & \frac{b}{4a^2} \\ \frac{b}{4a^2} & \frac{-1}{2a} \end{bmatrix} \begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix}.$$

This can be calculated by substitution of $(A,B,C) = (a,b,c)$ and $(\dot{A},\dot{B},\dot{C}) = (\dot{a},\dot{b},o)$ in the formula (5.2-40) obtained in theorem (5.2-38). It turns out to be possible to solve the geodesic differential equations, and to find a closed form expression for the inner distance d_M between any two points of the union of the manifold with the atom set containing the zero system.

The expression is:

$$d_M\big((a_o,b_o),(a_1,b_1)\big) =$$

$$(5.5.1\text{-}4)\qquad = \begin{cases} (r_o^2 - 2r_o r_1 \cos|\theta_o-\theta_1| + r_1^2)^{\frac{1}{2}} & \text{if } |\theta_o-\theta_1| \leq \pi \text{ and sign } r_o = \text{sign } r_1, \\ |r_o| + |r_1| & \text{if } |\theta_o-\theta_1| > \pi \text{ or sign } r_o \neq \text{sign } r_1, \end{cases}$$

where

$$r_k = b_k(2|a_k|)^{-\frac{1}{2}} \qquad , k = 0,1,$$

and

$$\theta_k = \ell n(|a_k|^{\frac{1}{2}}) \qquad , k = 0,1.$$

Instead of showing here how the geodesic differential equations can be solved and how (5.5.1-4) is obtained from the solution, we will use hindsight, and use a transformation of variables, which solves the problem in only a few steps. It also shows the correctness of (5.5.1-4).

Let $r = b(2|a|)^{-\frac{1}{2}}$, $\theta = \ell n(|a|^{\frac{1}{2}})$, then

$$a = -e^{2\theta} < 0, \quad b = \sqrt{(2)}re^{\theta}$$

Using (r,θ) as pair of parameters, $M^{m,\ell}_{1,1,1}$ is now parametrized by

$$\{(r,\theta)\,|\,r \neq 0,\ r,\theta \in \mathbb{R}\} = \{(r,\theta)\,|\,r < 0\} \cup \{(r,\theta)\,|\,r > 0\}.$$

Using this parametrization one has the following result:

5.5.1-5. Theorem. Let t denote a tangent vector, represented by $(\dot{r},\dot{\theta})^T$ in the coordinates (r,θ). Then the Riemannian metric is given with respect to the coordinates (r,θ), by

$$g(t,t) = \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix} \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}.$$

Proof. This is a straightforward computation. Let

$$J = \frac{\partial \begin{pmatrix} a \\ b \end{pmatrix}}{\partial (r,\theta)} \tag{5.5.1-6}$$

denote the Jacobian of the transformation of variables. Then $\begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix} = J \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}$. So one has

$$\begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix}^T \begin{bmatrix} \frac{-b^2}{4a^3} & \frac{b}{4a^2} \\ \frac{b}{4a^2} & \frac{-1}{2a} \end{bmatrix} \begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix} = \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}^T J^T \begin{bmatrix} \frac{-b^2}{4a^3} & \frac{b}{4a^2} \\ \frac{b}{4a^2} & \frac{-1}{2a} \end{bmatrix} J \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}. \tag{5.5.1-7}$$

J can be calculated to be

$$J = \begin{bmatrix} 0 & -2e^{2\theta} \\ \sqrt{2}.e^{\theta} & \sqrt{2}.r.e^{\theta} \end{bmatrix}. \tag{5.5.1-8}$$

Furthermore, substitution of $a = -e^{2\theta}$, $b = \sqrt{2}.re^{\theta}$ gives the following expression for the matrix in the middle of (5.5.1-7):

$$\begin{bmatrix} \frac{-b^2}{4a^3} & \frac{b}{4a^2} \\ \frac{b}{4a^2} & \frac{-1}{2a} \end{bmatrix} = \begin{bmatrix} \frac{r^2}{2}e^{-4\theta} & \frac{\sqrt{2}}{4}re^{-3\theta} \\ \frac{\sqrt{2}}{4}re^{-3\theta} & \frac{e^{-2\theta}}{2} \end{bmatrix}. \tag{5.5.1-9}$$

So one obtains

$$(5.5.1\text{-}10)\quad J^T \begin{bmatrix} \frac{-b^2}{4a^3} & \frac{b}{4a^2} \\ \frac{b}{4a^2} & \frac{-1}{2a} \end{bmatrix} J = \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.$$

Q.E.D.

Now notice that the Riemannian metric tensor of (5.5.1-5) is precisely the one of polar coordinates in the Euclidean plane. However, there are two differences with polar coordinates:

(i) (r,θ) does not denote the same point as $(r,\theta+2\pi)$,
(ii) r can take both positive and negative values here.

This means that the Riemannian geometry in this case is the geometry of two infinite sheeted Riemann surfaces: (r,θ) is a pair of (generalized) polar coordinates, $\theta \in \mathbb{R}$ is the generalized angle, and $|r|$ is the radius; the points with $r > 0$ form one Riemann surface and the points with $r < 0$ the other. The zero system ($r=0$) connects the two.
A formal construction of the Riemann surface is as follows.
Consider the set

$$\overset{\circ}{R} := \{(r\cos\theta, r\sin\theta, |r|\theta) \mid r \neq 0, \theta \in \mathbb{R}\}$$

$$= \{(x,y,\phi) \mid |r| = \sqrt{(x^2+y^2)} \neq 0, (x,y) = \pm(|r| \cos(\tfrac{\phi}{|r|}), |r| \sin(\tfrac{\phi}{|r|}))\} \subseteq \mathbb{R}^3$$

This is a two dimensional surface (a submanifold) in $\mathbb{R}^3$. Let π denote the projection $\pi(x,y,\phi) = (x,y)$. Then locally π is invertible, i.e. for each (x,y,ϕ) with $x^2+y^2 \neq 0$, there exists an open neighbourhood U of (x,y,ϕ) such that $\pi|_{\overset{\circ}{R}\cap U} : \overset{\circ}{R}\cap U \to \mathbb{R}^2$ is injective. Now let the metric on $\overset{\circ}{R}$ be locally the same as the Euclidean metric on $\mathbb{R}^2$, via the projection π. I.e. if $(\dot{x},\dot{y},\dot{\phi})$ denotes a tangent vector t, then the Riemannian metric on $\overset{\circ}{R}$ is given by

(5.5.1-11) $g_{\underset{R}{\circ}}(t,t) = \begin{pmatrix}\dot{x}\\ \dot{y}\end{pmatrix}^T \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\dot{x}\\ \dot{y}\end{pmatrix} = \dot{x}^2 + \dot{y}^2.$

Let $\overset{\circ}{R}$ be endowed with the inner metric d, corresponding to this Riemannian metric. Then $\overset{\circ}{R}$ is a model for the double, infinite sheeted Riemann surface. So we have

5.5.1-12. Theorem. $(\overset{\circ}{R},d)$ is isometric to $M^{m,\ell}_{1,1,1}$ (endowed with the inner metric induced by the Riemannian metric).

Proof. Consider the parametrization $\{(r,\theta)\,|\,r \in \mathbb{R}\backslash\{0\},\ \theta \in \mathbb{R}\}$ of $M^{m,\ell}_{1,1,1}$ as before, and with Riemannian metric tensor as in (5.5.1-5). Consider the mapping $\mu\colon M^{m,\ell}_{1,1,1} \to \overset{\circ}{R}$, $(r,\theta) \longmapsto (r\cos\theta, r\sin\theta, |r|\theta)$. Then it can easily be shown that μ is bijective and a local isometry. Because both the metrics in $\overset{\circ}{R}$ and $M^{m,\ell}_{1,1,1}$ are inner metrics (i.e. infimal pathlength metrics) the metric spaces must be isometric.

Q.E.D.

Next we want to add the zero system to the metric space. First note

5.5.1-13. Proposition. $|r| = |b|(2|a|)^{-\frac{1}{2}}$ is the norm of the system (as defined in 5.2-11).

Proof. The general formula is given in (5.2-11): $\|\Sigma\|^2 = \operatorname{tr} C.M(BB^T).C^T + DD^T$.

In this case D = 0, C = 1, B = b, A = a, and so $M(BB^T) = -\frac{b^2}{2a}$, and so

$$\|\Sigma\| = \sqrt{\left(\frac{b^2}{2|a|}\right)} = |b|(2|a|)^{-\frac{1}{2}}.$$

Q.E.D.

The straight line in Hilbert space between a system Σ and the zero system is given by $[\Sigma] := \{\lambda\Sigma\,|\,\lambda \in \mathbb{R}\}$. If $\Sigma \in M^{m,\ell}_{1,1,1}$ then $[\Sigma] \subseteq M^{m,\ell}_{1,1,1} \cup \{0\}$. It follows that $[\Sigma]$ is a geodesic, in the Hilbert space and therefore a fortiori in $M^{m,\ell}_{1,1,1} \cup \{0\}$, and so one has $d(\Sigma,0) = \|\Sigma\|$ in $M^{m,\ell}_{1,1,1} \cup \{0\}$. In this way the inner metric is defined on the whole set $M^{m,\ell}_{1,1,1} \cup \{0\}$. At the other hand, one can define the distance d of a point $(x,y,\phi) \in \overset{\circ}{R}$ to $(0,0,0) \in \overset{\circ}{R}$ to be $\sqrt{(x^2+y^2)}$,

in accordance with the rule that locally the metric on $\overset{\circ}{R}$ is equal to the Euclidean metric on $\pi(\overset{\circ}{R})$. From the previous theorem and the previous proposition it follows that:

5.5.1-14. Theorem. $M^{m,\ell}_{1,1,1} \cup \{0\}$, endowed with the inner metric described above, and $(\overset{\circ}{R} \cup \{0\},d)$, are isometric.

5.5.1-15. Remarks. (i) Because the geometry of $M^{m,\ell}_{1,1,1}$ is locally Euclidean, the curvature is zero.
(ii) Note that $M^{m,\ell}_{1,1,1}$ and $M^{m,\ell}_{1,1,1} \cup \{0\}$ are non-compact, and $M^{m,\ell}_{1,1,1} \cup \{0\}$ is not even locally compact in 0. In fact the 'sphere' $S_\varepsilon(0) = \{\Sigma \in M^{m,\ell}_{1,1,1} \mid \|\Sigma\| = \varepsilon\}$ consists topologically of two disjoint lines, for each $\varepsilon > 0$.
(iii) The geodesic between any two points in $M^{m,\ell}_{1,1,1} \cup \{0\}$ can be found easily as follows: Compute the 'polar' coordinates (r_o,θ_o) and (r_1,θ_1) of the points. If the difference $|\theta_o-\theta_1| \leq \pi$ and $\text{sign}(r_o) = \text{sign}(r_1)$, then the straight line segment ℓ between $\begin{pmatrix} r_o \cos\theta_o \\ r_o \sin\theta_o \end{pmatrix}$ and $\begin{pmatrix} r_1 \cos\theta_1 \\ r_1 \sin\theta_1 \end{pmatrix}$ corresponds to a unique continuous curve γ on $\overset{\circ}{R}$, starting in $(r_o \cos\theta_o, r_o \sin\theta_o, |r|\theta_o)$, and ending in $(r_1 \cos\theta_1, r_1 \sin\theta_1, |r|\theta_1)$, and such that $\pi(\gamma)$ is equal to the line segment ℓ. In other words γ is the lifting of ℓ. Then γ is the geodesic between the two points, with length $(r_o^2-2r_or_1\cos(\theta_o-\theta_1)+r_1^2)^{\frac{1}{2}}$.
If $|\theta_o-\theta_1| > \pi$ or $\text{sign}(r_o) \neq \text{sign}(r_1)$, then the geodesic consists of the union of the straight line segments from (r_o,θ_o) to the origin and from the origin to (r_1,θ_1). The length of such a geodesic is $|r_o| + |r_1|$.
(iv) It is interesting to compare the distance of two points in $M^{m,\ell}_{1,1,1}$ along the manifold (i.e. the inner metric) with the Hilbert space metric ('the outer metric') induced by the norm: As before let Σ_o have 'polar coordinates' (r_o,θ_o), and Σ_1: (r_1,θ_1). Then it is not difficult to compute that

$$\|\Sigma_o-\Sigma_1\| = \left\{r_o^2 - \frac{2r_or_1}{\cosh(\theta_o-\theta_1)} + r_1^2\right\}^{\frac{1}{2}}. \tag{5.5.1-16}$$

Note that $0 < (\cosh(\theta_o-\theta_1))^{-1} \leq 1$, and that $\cos(x) - (\cosh(x))^{-1} = \overset{\circ}{0}(x^4)$, which shows that $\cos(x)$ and $(\cosh(x))^{-1}$ have a comparable behaviour around zero. For larger values of $|\theta_o-\theta_1|$, the two distance concepts, (the inner and the 'outer' distance)) differ considerably. With r_o and r_1 fixed the limit for

$|\theta_o-\theta_1| \to \infty$ of (5.5.1-16) is $\sqrt{(r_1^2+r_2^2)}$, while the analogous limit of (5.5.1-4) is equal to $|r_1| + |r_2|$ (and it is reached already for $|\theta_o - \theta_1| = \pi$). Of course $|r_1| + |r_2| \geq \sqrt{(r_1^2+r_2^2)}$.

Next let us consider the discrete time case:

$$(5.5.1\text{-}17)\begin{cases} x_{t+1} = ax_t + bu_t, & -1 < a < 1,\ b \neq 0,\ c \neq 0, \\ y_t = cx_t \end{cases}$$

with x_t, y_t, u_t, a, b, c all real scalars. We choose a canonical form, by taking

$$(5.5.1\text{-}18)\quad c = 1.$$

So $M^{m,a}_{1,1,1}$ is parametrized by

$$\{(a,b) \mid -1 < a < 1,\ b \neq 0\} =$$
$$\{(a,b) \mid -1 < a < 1,\ b < 0\} \cup \{(a,b) \mid -1 < a < 1,\ b > 0\}.$$

(This parametrization is parameter identifiable). Just as in the continuous time case, it consists of two components, each of which is connected. Let t denote a tangent vector of $M^{m,a}_{1,1,1}$, represented by $(\dot{a},\dot{b})^T$ in this parametrization. The Riemannian metric tensor g is now given by

$$(5.5.1\text{-}19)\quad g(t,t) = \begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix}^T \begin{bmatrix} \dfrac{(1+a^2)b^2}{(1-a^2)^3} & \dfrac{ab}{(1-a^2)^2} \\ \dfrac{ab}{(1-a^2)^2} & \dfrac{1}{1-a^2} \end{bmatrix} \begin{bmatrix} \dot{a} \\ \dot{b} \end{bmatrix}.$$

This can be calculated by substitution of $(A,B,C) = (a,b,c)$ and $(\dot{A},\dot{B},\dot{C}) = (\dot{a},\dot{b},0)$ in formula (5.2-39). In this case consider the following transformation

$$(5.5.1\text{-}20)\begin{cases} \theta = \operatorname{arctgh}(a) \\ r = b\cosh(\theta) \end{cases}, \text{ or, equivalently} \begin{cases} a = \operatorname{tgh}(\theta) \\ b = \dfrac{r}{\cosh(\theta)} \end{cases}.$$

Remarks:

(i) Recall that $\mathrm{tgh}(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}} \in (-1,1)$.

(ii) Of course $\mathrm{arctgh}(y) = \frac{1}{2}\ell n\left(\frac{1+y}{1-y}\right)$.

(iii) A simple calculation shows that $r = \dfrac{b}{\sqrt{(1-a^2)}}$, and so r is equal to the norm of the system, as defined in (5.2-8) (with D = 0). Using (r,θ) as new coordinates, $M^{m,a}_{1,1,1}$ is parametrized by

$$\{(r,\theta) \mid r \neq 0,\ r,\theta \in \mathbb{R}\} = \{(r,\theta) \mid r > 0\} \cup \{(r,\theta) \mid r < 0\}.$$

It turns out that (r,θ) are 'generalized polar coordinates', just as in the continuous case:

5.5.1-21. Theorem. The Riemannian metric tensor is given with respect to the new coordinates (r,θ), in the discrete time case, by

$$g(t,t) = \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix}^T \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix} \begin{bmatrix} \dot{r} \\ \dot{\theta} \end{bmatrix},$$

where t denotes the tangent vector that is represented by $(\dot{r},\dot{\theta})^T$.

Proof. Just as in the proof of (5.5.1-5) we have to compute the Jacobian

$$J = \frac{\partial \binom{a}{b}}{\partial (r,\theta)}$$

of the transformation. It is equal to

$$(5.5.1\text{-}22) \quad J = \begin{bmatrix} 0 & (\cosh(\theta))^{-2} \\ (\cosh(\theta))^{-1} & r\,\mathrm{tgh}(\theta).(\cosh(\theta))^{-1} \end{bmatrix}.$$

Substitution of $a = \mathrm{tgh}(\theta)$, $b = r(\cosh(\theta))^{-1}$ in the matrix of formula (5.5.1-19) leads to the matrix:

$$(5.5.1\text{-}23) \begin{bmatrix} \cosh(\theta)^4(1+\mathrm{tgh}(\theta)^2)r^2 & \mathrm{tgh}(\theta)\cosh(\theta)^3 r \\ \mathrm{tgh}(\theta)\ \cosh(\theta)^3 r & \cosh(\theta)^2 \end{bmatrix}.$$

Premultiplication with J^T and postmultiplication with J leads to the Riemannian metric tensor of the coordinates (r,θ):

$$(5.5.1\text{-}24)\quad \begin{bmatrix} 1 & 0 \\ 0 & r^2 \end{bmatrix}.$$

Q.E.D.

Now let us compare the continuous time case and the discrete time case. In the continuous time case a transformation to polar coordinates on the Riemann surface is given by

$$(5.5.1\text{-}25)\quad \begin{cases} r_o = b_o(2|a_o|)^{-\frac{1}{2}}, \\ \theta_o = \ell n(|a_o|^{-\frac{1}{2}}), \end{cases} \text{ or, equivalently } \begin{cases} a_o = -e^{-2\theta_o}, \\ b_o = \sqrt{2}r_o e^{-\theta_o}, \end{cases}$$

where the subindex 0 now stands for the continuous time case. In the discrete time case the transformation to polar coordinates on the Riemann surface is given by (see (5.5.1-20)):

$$(5.5.1\text{-}26)\quad \begin{cases} r_1 = b_1 \cosh(\theta_1), \\ \theta_1 = \operatorname{arctgh}(a_1), \end{cases} \text{ or equivalently } \begin{cases} a_1 = \operatorname{tgh}(\theta_1) \\ b_1 = r_1 \cosh(\theta_1)^{-1}. \end{cases}$$

where the subindex 1 now stands for the discrete time case. Consider the mapping $Q: M^{m,a}_{1,1,1} \to M^{m,\ell}_{1,1,1}$ which maps the point represented by (r,θ) in $M^{m,a}_{1,1,1}$ to the point represented by (r,θ) in $M^{m,\ell}_{1,1,1}$. Then Q is an isometry. In terms of (a_o,b_o) and (a_1,b_1), Q^{-1} is given by

$$(5.5.1\text{-}27)\quad \begin{aligned} &(a_1,b_1) = Q^{-1}(a_o,b_o) = \\ &\left(-\operatorname{tgh}(\ell n|a_o|^{\frac{1}{2}})\right), b_o(2|a_o|)^{-\frac{1}{2}}\{\cosh(\ell n(a_o)^{\frac{1}{2}})\}^{-1}) = \left(\frac{a_o+1}{-a_o+1}, \frac{(\sqrt{2})b_o}{1-a_o}\right). \end{aligned}$$

5.5.1-28. Proposition. Q is equal to the mapping P_o, defined in (5.4-1), for the case $m = n = m' = 1$.

Proof. Compute P_o^{-1} in terms of a_o and b_o for this case. This gives precisely the formula for $Q^{-1}(a_o,b_o)$ just found. Q.E.D.

So the isometry Q is identical to the isometry P_o of section 5.4.
Next, consider the group of isometries of the space $M^{m,a}_{1,1,1}$ (mutatis mutandis the same holds for $M^{m,\ell}_{1,1,1}$). Let $M^{m,a}_{1,1,1} = M^+ \cup M^-$, where M^+ is the component of $M^{m,a}_{1,1,1}$ in which (b is positive, or equivalently) r is positive, and M^- is the other component. The group of isometries of $M^+ \cup M^-$ with respect to the Riemannian metric is generated by:

(i) Rotations around the origin; in terms of the polar coordinates such a rotation is a mapping $(r,\theta)\longmapsto(r,\theta+\gamma)$. Because M^+ and M^- are disconnected one can rotate M^+ and M^- differently.

(ii) Reflection of the generalized angle with respect to the zero angle. In terms of polar coordinates $(r,\theta)\longmapsto(r,-\theta)$. This can be done in none, one or both components at the same time.

(iii) Reflection with respect to the origin, in terms of the generalized polar coordinates $(r,\theta)\longmapsto(-r,\theta)$. In this case M^+ is mapped onto M^- and vice versa.

Now let us restrict ourselves to the isometries of M^+ (i.e. the infinite sheeted Riemann surface). They are given by the group $\{K_\gamma\} \cup \{K^-_\gamma\}$, where $K_\gamma(r,\theta) = (r,\theta+\gamma)$, in polar coordinates, and $K^-_\gamma(r,\theta) = (r,-(\theta+\gamma)) = K^-_o \circ K_\gamma$ with $K^-_o(r,\theta) = (r,-\theta)$. This can be compared with the group of isometries $\{F_\beta\} \cup \{F^-_\beta\}$, found in section 5.4, theorem (5.4-22) for the case m=n=m'=1, and restricted to the component M^+.

5.5.1-29. Theorem. $\forall\gamma \in \mathbb{R}$: $K_\gamma = F_{-2\gamma}|_{M^+}$ and $K^-_\gamma = F^-_{-2\gamma}|_{M^+}$.

Proof. The mapping F_β in terms of the polar coordinates (r,θ) can be calculated explicitly using $a = \mathrm{tgh}(\theta)$, $b = r(\cosh(\theta))^{-1}$, etc. The explicit calculation shows $K_\gamma = F_{-2\gamma}|_{M^+}$ and $K^-_o = F^-_o|_{M^+}$.
Q.E.D.

5.5.1-30. Remark. It follows that the 'hyperbolic angle' which occurs in F and ϕ is transformed to an angle in the Riemann surface!

We want to conclude this subsection with making some remarks on the relation of the infinite sheeted Riemann surface M^+ and complex function theory, especially the theory of conformal mapping. As is well-known the infinite sheeted Riemann surface can be considered as the domain of a complex analytic continuation of the logarithm on the positive real line. This analytic function maps the Riemann surface conformally onto $\mathbb{C}$. In terms of polar coordinates it is easy to see what this means: The analytic continuation of the logarithm maps the point on the Riemann surface with generalized polar coordinates (r,θ) to the point $\ell nr+i\theta \in \mathbb{C}$. If we put $x = \ell nr$, $y = \theta$, then the Riemannian metric tensor corresponding to these coordinates is $r^2 I_2$, (with $r = e^x$). So the Riemannian metric tensor is given by a (socalled) scalar matrix. This is typical for conformal mapping, because a map is conformal iff the Riemannian metric in an image point is a positive scalar factor, depending on the image point, times the Riemannian metric in the original point (see e.g. [Ko-N], vol. I, p. 309, Note 11). It may be interesting to note that to determine the Riemannian gradient direction of a function in a point, one needs to know the Riemannian metric tensor up to a positive scalar factor (see also chapter 6). So in this case, using
$(x,y) = (\ell nb_1 - \frac{1}{2}\ell n(1-a_1) - \frac{1}{2}\ell n(1+a_1), \frac{1}{2}\ell n(1+a_1) - \frac{1}{2}\ell n(1-a_1))$ as coordinates, calculation of the Riemannian gradient direction of the function V on M^+ could be done by calculating $(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y})^T$.

In the continuous time case, one could take
$(x,y) = (\ell nb_o + \ell n(|a_o|^{-\frac{1}{2}}), \ell n|a_o|^{-\frac{1}{2}}) = (z+y,y)$, with $z = \ell nb_o$. The Riemannian gradient direction would be given by $(\frac{\partial V}{\partial x}, \frac{\partial V}{\partial y})^T = (\frac{\partial \tilde{V}}{\partial z}, -\frac{\partial \tilde{V}}{\partial z} + \frac{\partial \tilde{V}}{\partial y})^T$, where

$$V = V(x,y),\ \tilde{V} = \tilde{V}(z,y) = V(z+y,y).$$

5.5.2. Some other special cases

(i) The first one is

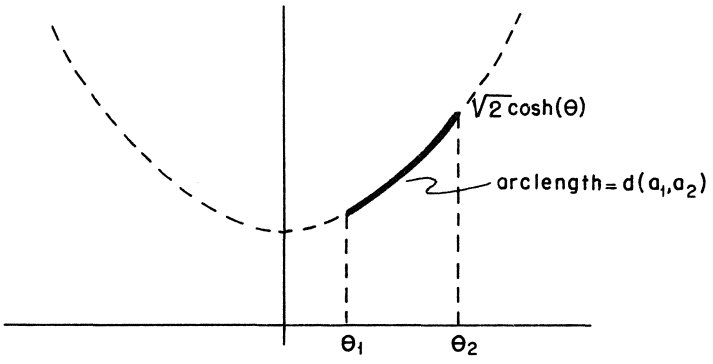
$$(5.5.2\text{-}1)\quad \begin{cases} x_{t+1} = ax_t + u_t, & a \in \mathbb{R}, \\ y_t = x_t \end{cases}$$

i.e. we have put b = 1. Each such system is completely characterized by the value of a. The problem is to find the inner distance d. Using again the transformation $a = \mathrm{tgh}(\theta)$, one finds

$$(5.5.2\text{-}2) \quad d(a_1,a_2) = \int_{a_1}^{a_2} \frac{1}{1-a^2}\sqrt{\left(\frac{1+a}{1-a}\right)}da = \int_{\theta_1}^{\theta_2} \sqrt{\cosh(2\theta)}d\theta =$$

$$= \int_{\theta_1}^{\theta_2} \sqrt{\left(1+f'(\theta)^2\right)}d\theta \text{ where } f(\theta) = \left(\sqrt{2}\right)\cosh(\theta).$$

This means that the distance between $a_1 = \mathrm{tgh}(\theta_1)$ and $a_2 = \mathrm{tgh}(\theta_2)$ is equal to the arclength of the graph of the function $\left(\sqrt{2}\right)\cosh(\theta)$, between θ_1 and θ_2.



(ii) Now consider

$$(5.5.2\text{-}3) \quad \begin{cases} x_{t+1} = ax_t + u_t, & a \in \mathbb{C}, \\ y_t = x_t \end{cases}$$

and use the P-norm given by (5.2-25). In section 5.2 we promised to return to this case. The sequence of Markov parameters is $(1,a,a^2,a^3,\ldots)$, and a tangent vector is of the form $(0,\dot{a},2a\dot{a},3a^2\dot{a},4a^3\dot{a},\ldots)$, which has P-norm equal to $\frac{a^2}{1-|a|^2}$. This is the Riemannian metric tensor of the well-known Poincaré metric. (See e.g. [Ne-P], section 3.11). This implies that we are in fact dealing with the hyperbolic geometry in the open unit disk. The geodesics and the corresponding inner distance function of this geometry are well-known. The points on the unit circle lie at 'infinite distance', in the Poincaré metric, i.e. the Poincaré metric is only well-defined on the open unit disk. This is

remarkable because the P-norm, given by (5.2-25), (and consequently the outer metric) is well-defined also for elements on the unit circle!

5.6. Degeneration of the Riemannian metric tensor

In this section we want to make some remarks about the behaviour of the Riemannian metric tensor at points of McMillan degree < n. Depending on the parametrization, it may or may not degenerate at such a point. For example in the Riemann surface case (i.e. $M^{m,a}_{1,1,1}$) of section 5.5, one could use rectangular coordinates, instead of polar coordinates, for each 'halfplane' $\{(r,\theta)\,|\,|\theta-\theta_o| \leq \pi/4\}$, $r > 0$ (and similarly for $r < 0$). One simply has to take $x = r\cos(\theta)$, $y = r\sin(\theta)$. In (x,y)-coordinates the Riemannian metric tensor is the identity matrix, and will therefore not degenerate for $r = \sqrt{(x^2+y^2)} \to 0$. On the other hand in terms of the parameters (a,b), with $c \equiv 1$, the Riemannian metric tensor has a degenerate limit for $b \to 0$:

$$(5.6\text{-}1)\qquad \lim_{b\to o}\begin{bmatrix} \dfrac{(1+a^2)b^2}{(1-a^2)^3} & \dfrac{ab}{(1-a^2)^2} \\ \dfrac{ab}{(1-a^2)^2} & \dfrac{1}{1-a^2} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \dfrac{1}{1-a^2} \end{bmatrix}$$

and clearly this matrix is positive semi-definite and not positive definite. Now let us define formally what we mean by a degenerate Riemannian metric. (not to be confused with a pseudo-Riemannian metric).

5.6-2. Definition. (i) Let $k = \mathbb{R}$ or $\mathbb{C}$. A degenerate inner product g on a k-vector space V is a mapping $g: V\times V \to k$ with the properties:

(a) $g(\lambda v+w,u) = \lambda g(v,u) + g(w,u)$

and

$g(v,\lambda w+u) = \bar{\lambda} g(v,w)+g(v,u)$,

for all $\lambda \in k$; $v,w,u \in V$.

(b) $g(v,v) \geq 0$.

(ii) A degenerate Riemannian metric g is an assignment for each $\Sigma \in M$, of a degenerate inner product g_Σ to the tangent space T_Σ at M in Σ, such that g_Σ varies smoothly with Σ (to be determinate, let us require that g_Σ is a C^∞ tensor).

5.6-3. Remark. The difference of a degenerate inner product with a usual inner product is, that vectors are allowed which have zero inner product with any

other vector.

To such a degenerate Riemannian metric corresponds a degenerate inner metric, because different points may have zero distance. This problem can be overcome by considering equivalence classes as follows: Two points η_1,η_2 will be called equivalent if they have zero distance according to the degenerate inner metric considered. Then the degenerate inner metric induces a non-degenerate metric on the set of equivalence classes.

Consider now the Riemannian metric g, defined before on $M^{m,a}_{m',n,m}$, (as before completely analogous results hold for the continuous time case). On the other hand, consider the manifold $L^{m,a}_{m',n,m}$ of all matrix triples (A,B,C), A: n×n, B: n×m', C: m×n, corresponding to minimal, asymptotically stable systems. Let η = vec(A,B,C) be the vector consisting of all the components of A, B and C. An element $\dot{\eta}$ of the tangent space $TL^{m,a}_{m',n,m}$ at η corresponds naturally to a tangent vector t of $TM^{m,a}_{m',n,m}$ at $\Sigma(A,B,C)$. Note that t = 0 does not imply $\dot{\eta}$ = 0. Let R = R(η) denote the $(n^2+nm+nm') \times (n^2+nm+nm')$ symmetric positive semi-definite matrix such that the Riemannian length of a tangent vector $t \in TM^{m,a}_{m',n,m}$, represented by a tangent vector $\dot{\eta} \in TL^{m,a}_{m',n,m}$ is given by

$$(5.6\text{-}4) \qquad g_\Sigma(t,t)^{\frac{1}{2}} = (\dot{\eta}^T R \dot{\eta})^{\frac{1}{2}}.$$

Clearly R(η) defines a (degenerate) Riemannian metric tensor on $L^{m,a}_{m',n,m}$. That it is really degenerate for each $n \geq 1$, follows from the following observation.

5.6-5. Theorem. $\forall \eta \in L^{m,a}_{m',n,m}$: rk(R($\eta$)) = n(m+m') (and so its corank is n^2).

Proof. According to corollary (4.6-6), $\beta: L^{m,a}_{m',n,m} \to M^{m,a}_{m',n,m}$ is a (locally trivial) principal fibre bundle. It follows that locally $L^{m,a}_{m',n,m}$ is diffeomorphic to $M^{m,a}_{m',n,m} \times G\ell_n$. From this and the fact that the Riemannian metric is non-degenerate on $M^{m,a}_{m',n,m}$ it follows directly that

$$\text{rk}(R(\eta)) \geq n(m+m'),$$

but also, because each tangent vector that corresponds with a tangent vector of $G\ell_n$ is in the kernel of R(η), $\text{rk}(R(\eta)) \leq (nm+nm'+n^2) - (n^2) = nm + nm'$.

This shows that

(5.6-6) $\quad rk(R(\eta)) = n(m+m')$.

Q.E.D.

The point we want to make in this section is that the rank of $R(\eta)$ will be smaller than $n(m+m')$ at points $\eta = vec(A,B,C)$ of McMillan degree smaller than n. (As noted before this is a property of the parametrization. If, instead of working with the parameters (A,B,C) one works with other parameters, the corresponding metric tensor may not degenerate at points of McMillan degree < n). We will not investigate the statement in general, but give a proof only for a special case, as an example.

5.6-7. Theorem. Consider the set $L^a_{1,n,1}$ of matrix triples (A,b,c), each describing an asymptotically stable s.i.s.o. system. Suppose (A,b,c) is observable, but non-reachable, and has McMillan degree n-1.
Let $\eta = vec(A,b,c)$. Then

(5.6-8) $\quad rank(R(\eta)) < n(m+m') = 2n$.

Proof. Because the system described by (A,b,c) is observable, but has McMillan degree n-1, there exists a basis transformation of the state space, which transforms the matrix triple (A,b,c) into one of the following form (this is the well-known Kalman decomposition, see e.g. [Che], theorem 5.17):

$$(5.6\text{-}9) \qquad \left[\begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}, \begin{pmatrix} b_1 \\ 0 \end{pmatrix}, (c_1, c_2)\right],$$

where A_1 is (n-1) × (n-1), A_2 is 1 × (n-1), A_3 is 1×1, b_1 is (n-1) × 1, c_1 is 1 × (n-1) and c_2 is 1×1; (A_1,b_1,c_1) is a minimal realization of the system. Without loss of generality we may assume that (A,b,c) is of the form (5.6-9). A change in A_2, A_3 or c_2 does not affect the input-output map. Thus all tangent vectors of the form

$$(5.6\text{-}10)\qquad \dot{\eta} = \operatorname{vec}\left[\begin{pmatrix} 0 & \dot{A}_2 \\ 0 & \dot{A}_3 \end{pmatrix},\ 0,\ (0\ \ \dot{c}_2)\right]$$

constitute an (n+1)-dimensional subspace of the kernel of the matrix R. On the other hand consider the set of all tangent vectors of the form

$$(5.6\text{-}11)\qquad \dot{\eta} = \operatorname{vec}(-TA+AT, -Tb, cT),\ T \in \mathbb{R}^{n\times n} \text{ arbitrary.}$$

They constitute a subspace of the kernel of R. This can be seen as follows. Consider the curve $G(t) \in Gl_n$, $|t| \leq \varepsilon$ and $G(0) = I_n$, $\dot{G}(0) = T$. Then the curve

$$(G(t)^{-1}AG(t),\ G(t)^{-1}b,\ cG(t))$$

in $L^a_{1,n,1}$ corresponds for each $t \in [-\varepsilon,\varepsilon]$ to the same system (and the same Hankel matrix). Therefore its tangent vectors are in the kernel of R. At $t = 0$ its tangent vector is of the form (5.6-11). As T was chosen arbitrarily, each vector $\dot{\eta}$ of the form (5.6-11) is in the kernel of R indeed.
Partition T as

$$(5.6\text{-}12)\qquad T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix},$$

T_{11} is (n-1) × (n-1), T_{12} is (n-1) × 1, T_{21} is 1 × (n-1) and T_{22} is 1×1. Clearly T can be written as

$$(5.6\text{-}13)\qquad T = \begin{bmatrix} T_{11} & 0 \\ T_{21} & 0 \end{bmatrix} + \begin{bmatrix} 0 & T_{12} \\ 0 & T_{22} \end{bmatrix},$$

and the space $\left\{\begin{pmatrix} T_{11} & 0 \\ T_{21} & 0 \end{pmatrix}\right\}$ has dimension n(n-1). Consider the corresponding space of tangent vectors

$$(5.6\text{-}14)\qquad \left\{\dot{\eta} = \operatorname{vec}(-TA+AT, -Tb, cT)\,\middle|\, T = \begin{pmatrix} T_{11} & 0 \\ T_{21} & 0 \end{pmatrix}\right\}.$$

If this space too has dimension n(n-1), and at the same time has zero intersection with the (n+1)-dimensional subspace of tangent vectors of the form (5.6-10), then it follows that the dimension of the kernel is larger than or equal to $n(n-1) + n + 1 = n^2 + 1 > n^2$, and the theorem then

follows. The space in (5.6-14) has dimension n(n-1), if the equation

$$(5.6\text{-}15) \quad (-TA+AT,-Tb,cT) = 0 \text{ with } T = \begin{pmatrix} T_{11} & 0 \\ T_{21} & 0 \end{pmatrix}$$

has T=0 as its only solution, while the intersection of the space in (5.6-14) with the one defined by (5.6-10) is zero if the equation

$$(5.6\text{-}16) \quad (-TA+AT,-Tb,cT) = \left[\begin{pmatrix} 0 & \dot{A}_2 \\ 0 & \dot{A}_3 \end{pmatrix}, 0, \begin{pmatrix} 0 & c_2 \end{pmatrix}\right],$$

with $T = \begin{pmatrix} T_{11} & 0 \\ T_{21} & 0 \end{pmatrix}$, has zero as its only solution.

Now (5.6-16) contains (5.6-15), so we only have to solve equation (5.6-16). Part of it can be rewritten as

$$(5.6\text{-}17) \quad \begin{aligned} &(a) - T_{11}A_1 + A_1T_{11} + A_2T_{21} = 0, \\ &(b) - T_{21}A_1 + A_3T_{21} = 0, \\ &(c) - T_{11}b_1 = 0, \\ &(d) - T_{21}b_1 = 0, \\ &(e)\ c_1T_{11} + c_2T_{21} = 0. \end{aligned}$$

Let us show by induction that $T_{11}A_1^k b_1 = 0$ and $T_{21}A_1^k b_1 = 0$, $k = 0,1,2,\ldots$. For $k = 0$ this follows from (c) and (d). Suppose it holds for k. Postmultiply (a) and (b) by $A_1^k b_1$. Then, using the induction hypothesis it follows that $T_{11}A_1^{k+1}b_1 = 0$ and $T_{21}A_1^{k+1}b_1 = 0$. Because (A_1,b_1,c_1) is a minimal realization, it follows that $T_{11} = 0$, $T_{21} = 0$. Q.E.D.

Because of the continuity of R as a function of the parameters (A,B,C), it follows that in a neighbourhood of a system of McMillan degree smaller than n, the matrix R has - apart from the n zero eigenvalues corresponding to the tangent vectors of Gl_n, as described in the proof of theorem (5.6-5) - at least one eigenvalue close to zero. This means that the input-output map (or equivalently, the Markov parameters), will be very insensitive to a change of the parameters in the direction of a corresponding eigenvector. So in this respect the parametrization by (A,B,C) does not behave very well. This holds for all locally continuous canonical forms in which n^2 components of (A,B,C) are fixed to be 0 or 1, and all other components are free to vary in an n(m+m')-dimensional open set. An interesting research problem is to find

(practical) parametrizations which do not have this kind of insensitivity property, i.e. for which the Riemannian metric tensor does not degenerate if a system of McMillan degree smaller than n is approached. In the $M^{m,a}_{1,1,1}$ case the (x,y) parametrization from the beginning of this section satisfies these requirements.

5.7. A Riemannian metric on the state bundle

In this section we will define a Riemannian metric on the whole state bundle $E = M^{m,a}_{1,m',n,m}$ (see section 4.6), which is an extension of the Riemannian metric on $M^{m,a}_{m',n,m}$ as defined before.

We will use the notation $\Sigma(A,B,C;x)$ to denote a system represented by A,B,C, with corresponding state, represented by x. One can identify $\Sigma(A,B,C;x)$ with the orbit of (A,B,C;x) under the action of $G\ell_n(\mathbb{R})$ ('change of basis of the state space'):

$$(5.7\text{-}1) \qquad \Sigma(A,B,C;x) := \{(TAT^{-1},TB,CT^{-1};Tx) \,|\, T \in G\ell_n(\mathbb{R})\}.$$

Consider a system with state $\Sigma(A,B,C;x)$. The corresponding system $\Sigma(A,B,C)$ is determined completely by the sequence of Markov matrices $\{CA^kB\}_{k=o}^{\infty}$, while the system-with-state $\Sigma(A,B,C;x)$ is determined completely by the sequences $\{CA^kB\}_{k=o}^{\infty}$ and $\{CA^kx\}_{k=o}^{\infty}$. As before, we define the norm of $\Sigma(A,B,C)$ to be the ℓ^2-norm of $\{CA^kB\}_o^{\infty}$:

$$(5.7\text{-}2) \qquad \|\Sigma(A,B,C)\|^2 = \sum_{k=o}^{\infty} \|CA^kB\|_F^2.$$

Analogously, we define the norm of x, for given (A,B,C), by the ℓ^2-norm of the sequence $\{CA^kx\}_o^{\infty}$, i.e.:

$$(5.7\text{-}3) \qquad \|x\|^2_{(A,B,C)} := x^T \sum_{k=o}^{\infty} (A^T)^k C^T CA^k x.$$

(Both norms can be computed by solving a Lyapunov equation). Of course, one could formally define a norm on the set of systems-with-states by using the embedding $M^{m,a}_{1,m',n,m} \to M^{m,a}_{m'+1,n,m}$, induced by $(A,B,C,x) \longmapsto (A,[B,x],C)$. This has, however, the disadvantage that the distance between systems would depend on the state. To avoid this, we will use another metric, which will presently be derived.

Consider a smooth curve in the state bundle E, given by

$$t \longmapsto \Sigma(A(t), B(t), C(t); x(t)), \quad t \in [-\varepsilon,\varepsilon], \tag{5.7-4}$$

for some $\varepsilon > 0$. Because each system with state corresponds to a pair of sequences as just described, this curve corresponds to the mapping

$$t \longmapsto [\{C(t)A(t)^k B(t)\}_{k=0}^{\infty}, \{C(t)A(t)^k x(t)\}_{k=0}^{\infty}], \quad t \in [-\varepsilon,\varepsilon]. \tag{5.7-5}$$

Therefore, the <u>tangent vector</u> of the curve at $t = 0$ corresponds to the pair of sequences.

$$\{\frac{\partial}{\partial t}[C(t)A(t)^k B(t)]_{t=0}\}_{k=0}^{\infty}, \ \{\frac{\partial}{\partial t}[C(t)A(t)^k x(t)]_{t=0}\}_{k=0}^{\infty}. \tag{5.7-6}$$

The <u>state space</u> $\mathbb{R}^n$ corresponds at $t = 0$ to the space of sequences

$$X_\Sigma := \{\{CA^k x\}_{k=0}^{\infty} | x \in \mathbb{R}^n\}, \tag{5.7-7}$$

where we write C for C(0), A for A(0) etc, for simplicity of notation. Note that X_Σ is a <u>coordinate free</u> description of the state space at $\Sigma(A,B,C)$. The sequence $\{\frac{\partial}{\partial t}[C(t)A(t)^k x(t)]_{t=0}\}_{k=0}^{\infty}$ can be <u>decomposed</u> into a part that is orthogonal to the 'state space' X_Σ, in the ℓ^2 sense, and a part that is <u>in the direction</u> of X_Σ. With the sequence $\{\frac{\partial}{\partial t}[C(t)A(t)^k B(t)]_{t=0}\}_{k=0}^{\infty}$ we want to measure the infinitesimal change in the <u>system</u>, and with the sequence $\{\frac{\partial}{\partial t}[C(t)A(t)^k x(t)]_{t=0}\}_{k=0}^{\infty}$ we want to measure the infinitesimal change in the <u>state</u> (only). Therefore we are <u>only</u> interested in the part in the orthogonal decomposition of this last sequence, that is <u>in the direction</u> of the space X_Σ. Let P_X denote the orthogonal projection on X_Σ. We define a Riemannian metric on the state bundle E as follows:

5.7-8. <u>Definition</u>. Let the norm squared of the tangent vector $\dot{\Sigma}(A,B,C;x)$ of an arbitrary smooth curve through $\Sigma(A,B,C;x)$ be given by

$$\begin{aligned} \|\dot{\Sigma}(A,B,C;x)\|^2 &:= \|\{\frac{\partial}{\partial t}[C(t)A(t)^k B(t)]_{t=0}\}_{k=0}^{\infty}\|_{\ell^2}^2 + \\ &+ \|P_X(\{\frac{\partial}{\partial t}[C(t)A(t)^k x(t)]_{t=0}\}_{k=0}^{\infty})\|_{\ell^2}^2. \end{aligned} \tag{5.7-9}$$

Here $\|,\|_{\ell^2}$ stands for the ℓ^2-norm, i.e. the square root of the sum of squares of all entries in the sequence, as before. Because of the asymptotic stability these norms are indeed finite.
If instead of the point t = 0 we consider an arbitrary point t of the curve, this formula becomes

$$\|\dot{\Sigma}(t)\|^2 = \|\{\tfrac{\partial}{\partial t}[C(t)A(t)^k B(t)]\}_{k=o}^{\infty}\|_{\ell^2}^2 + \|P_{X(t)}(\{\tfrac{\partial}{\partial t}[C(t)A(t)^k x(t)]\}_{k=o}^{\infty})\|_{\ell^2}^2 \tag{5.7-10}$$

where $P_{X(t)}$ denotes the orthogonal projection in ℓ^2 on the space of sequences

$$X_{\Sigma}(t) = \{\{C(t)A(t)^k x\}_{k=o}^{\infty} | x \in \mathbb{R}^n\}. \tag{5.7-11}$$

The curve $\Sigma(A(t),B(t),C(t);x(t))$ in E is called a parallel displacement of the state, along the curve $\Sigma(A(t),B(t),C(t))$ in $M^{m;a}_{m',n,m,}$, if the last term in (5.7-10) is zero for all $t \in [-\varepsilon,\varepsilon]$. If the state at t = 0 is determined to be x(0) (say) and the (smooth) curve $\Sigma(A(t),B(t),C(t))$ in $M^{m;a}_{m',n,m}$ is determined, then there is a unique parallel displacement $\Sigma(A(t),B(t),C(t);x(t))$. (This is a standard result from differential geometry, see e.g. [Ko-N]); in this case, it can also be shown directly, using formula (5.7-10)). The Riemannian metric (5.7-9) is constructed such that if the curve $\Sigma(A(t),B(t),C(t);x(t))$ is a parallel displacement of the state, then the length of the curve is independent of the state, and equal to the Riemannian length of the curve $\Sigma(A(t),B(t),C(t))$ as defined before! (In this sense the Riemannian metric (5.7-9) is a generalization to E of the Riemannian metric on $M^{m;a}_{m',n,m}$ defined before).
It can easily be shown that the parallel displacement of the state as just described corresponds to a metric connection, which means that the norm of the state, as defined by (5.7-3), remains constant along a parallel displacement. For an application of this Riemannian metric, we refer to [Hnz 85b].

5.8. Riemannian metrics for families of stochastic linear systems

We will consider subsets of the family of all stationary, ergodic, linear, finite dimensional, zero mean Gaussian, full rank stochastic processes $\{y_t\}$, (so $Ey_t y_t^T$ is positive definite). The subsets that we will consider are those for which the dimension of the outputs and the McMillan degree are fixed. (see section 2.4.3 for the definition of the McMillan degree of such a stochastic process).

In chapter 2, especially section 2.4(cf. (2.4.1-1)), it has been shown that such a stochastic system Γ can be represented by

$$(5.8\text{-}1) \qquad \begin{cases} x_{t+1} = Ax_t + Bw_t, \; x_t \in \mathbb{R}^n, \\ y_t = Cx_t + Dw_t, \; y_t \in \mathbb{R}^m, w_t \in \mathbb{R}^{m'}, \end{cases}$$

$\{w_t\}$ Gaussian white noise, $w_t \sim N(0,\bar{\Omega})$, $\bar{\Omega}$ a positive definite matrix, $T(s) = D + C(sI-A)^{-1}B$ has its poles all within the open unit disk $D(0,1)$ (we do not require (A,B,C,D) to be observable and reachable here, until such is explicitly stated). It is important to note that the representation (5.8-1) embraces other representations as special cases:

(i) If $D = I$ and $\sigma(A-BC) \subseteq \overline{D(0,1)}$ and (A,B,C) a minimal realization, then (5.8-1) is an innovations representation (see [An-M], chapter 9, theorem 4.4). This representation is unique ([An-M], ch.9, thm 4.1) among the various representations of the stochastic process. See also chapter 4.

(ii) If $D\bar{\Omega}B^T = 0$ one has the often encountered case of stochastic independence of process noise (Bw) and measurement noise (Dw).

(iii) The stochastic arma model can be rewritten in the form of (5.8-1). See e.g. section 2.3.3, especially equation (2.3.3-23)ff or [An-M], chapter 9, lemma 3.2. Note again that we do not require (A,B,C,D) to be a minimal representation, only that $\sigma(A) \subseteq D(0,1)$.

It follows that if we can give a formula for a Riemannian metric tensor in terms of the parameters $(A,B,C,D,\bar{\Omega})$ of (5.8-1), then the corresponding Riemannian metric tensor for any of these other representations can be directly derived from such a formula. Consider the sequence of covariance matrices $\{\Gamma_k\}$ of the process $\{y_t\}$. We found before (in (2.4.2-2) and

(2.4.2-3)) expressions for Γ_k in terms of $A,B,C,D,\bar{\Omega}$.
Now let

$$(5.8\text{-}2) \qquad G = G(A,B,C,D,\bar{\Omega}) := B\bar{\Omega}D^T + APC^T,$$

where $P = L(B\bar{\Omega}B^T)$ (i.e. P is as in (2.4.1-2) and (2.4.1-3)), and let

$$(5.8\text{-}3) \qquad J = J(A,B,C,D,\bar{\Omega}) := CPC^T + D\bar{\Omega}D^T.$$

Then (see (2.4.2-2) and (2.4.2-3)):

$$(5.8\text{-}4) \qquad \Gamma_o = J, \text{ positive definite,}$$

and

$$\Gamma_k = CA^{k-1}G, \quad k = 1,2,3,\ldots$$

(and the equality $\Gamma_k = \Gamma_{-k}^T$ can be used to find the Γ_k for $k < 0$).
It follows that the sequence $\{\Gamma_k\}_{k=o}^{\infty}$ is the sequence of Markov matrices of the deterministic system (A,G,C,J). It is shown in (2.4.3-25) that the McMillan degree of this system is equal to the McMillan degree of the stochastic system. It follows that the set of stochastic systems, as described above, with fixed McMillan degree (and fixed output dimension) can be represented by a subset of $M_{m,n,m}^{m,a} \times \mathrm{Pos}(m)$, namely the image of the mapping

$$(5.8\text{-}5) \qquad (A,B,C,D,\bar{\Omega}) \longmapsto ([A,G,C)],J) \in M_{m,n,m}^{m,a} \times \mathrm{Pos}(m),$$

with $G = G(A,B,C,D,\bar{\Omega})$ and $J = J(A,B,C,D,\bar{\Omega})$ as above. (Note that under a state space base change that transforms the state x into Sx (S nonsingular) G transforms to SG and J to J.) If we restrict to stochastic systems which have no zero on the unit circle, i.e.

$$(5.8\text{-}6) \qquad \det \sum_{k\in\mathbb{Z}} \Gamma_k e^{ik\phi} \neq 0, \forall \phi \in [0,2\pi),$$

then the corresponding subset forms an open submanifold of $M_{m,n,m}^{m,a} \times \mathrm{Pos}(m)$. (This follows from the analysis, below, of the special case of the innovations representation). Consider the Riemannian metric on

$\overset{\circ}{M}{}^{m,a}_{m,n,m}$ ($= M^{m,a}_{m,n,m} \times \mathbb{R}^{m\times m}$) defined in section 5.2. This induces a Riemannian metric on any submanifold. Especially, this defines a Riemannian metric on our manifold of stochastic systems. In terms of the representation $(A,B,C,D,\bar{\Omega})$ of the stochastic system Γ, the Riemannian metric tensor h_Γ (say) can be obtained from the formula (5.2-39): Let $(\dot{A},\dot{B},\dot{C},\dot{D},\dot{\bar{\Omega}})$ denote a tangent vector at $(A,B,C,D,\bar{\Omega})$.
Then

$$(5.8\text{-}7) \qquad \|(\dot{A},\dot{B},\dot{C},\dot{D},\dot{\bar{\Omega}})\|^2_{h_\Gamma} := \|(\dot{A},\dot{G},\dot{C},\dot{J})\|^2_{g_\Sigma},$$

(see (5.2.3-9)), where Σ is the deterministic system represented by (A,G,C,J) (i.e. the covariance system), and where

$$(5.8\text{-}8) \qquad \dot{G} = \dot{B}\Omega D^T + B\dot{\bar{\Omega}}D^T + B\bar{\Omega}\dot{D}^T + \dot{A}PC^T + A\dot{P}C^T + APC^T,$$

where $\dot{P}$ is the solution of

$$(5.8\text{-}9) \qquad \dot{P} - \dot{A}PA^T - A\dot{P}A^T - APA^T = \dot{B}\bar{\Omega}B + B\bar{\Omega}\dot{B}^T + B\dot{\bar{\Omega}}B^T,$$

i.e.

$$(5.8\text{-}10) \qquad \dot{P} = L(\dot{A}PA^T + APA^T + \dot{B}\bar{\Omega}B^T + B\bar{\Omega}\dot{B}^T + B\dot{\bar{\Omega}}B^T)$$

and

$$(5.8\text{-}11) \qquad \dot{J} = \dot{C}PC^T + CP\dot{C}^T + CPC^T + \dot{D}\bar{\Omega}D^T + D\bar{\Omega}\dot{D}^T + D\dot{\bar{\Omega}}D^T.$$

5.8-12. Remarks: (i) In specific cases the formulas for $\dot{G}$ and $\dot{J}$ simplify considerably.
(ii) For the case n = 1, m = 1, m' = 1 the Riemannian metric tensor is given explicitly in appendix 5A.
(iii) Another possibility, a minor variation of equal mathematical elegance, is to use the mapping

$$(5.8\text{-}13) \qquad (A,B,C,D,\bar{\Omega}) \longmapsto ([A,G,C)],\tfrac{1}{2}J) \in M^{m,a}_{m,n,m} \times \mathrm{Pos}(m)$$

instead of

$$(A,B,C,D,\bar{\Omega}) \longmapsto ([A,G,C)],J) \in M^{m,a}_{m,n,m} \times \mathrm{Pos}(m).$$

Clearly, then we get a Riemannian metric tensor

$$(5.8\text{-}14) \qquad \|(\dot{A},\dot{B},\dot{C},\dot{D},\dot{\bar{\Omega}})\|^2_{k_\Gamma} := \|(\dot{A},\dot{G},\dot{C},\tfrac{1}{2}\dot{J})\|^2_{g_\Sigma},$$

where Σ is now the deterministic system represented by $(A,G,C,\frac{1}{2}J)$. Of course the McMillan degree is not affected by this. The reason that this possibility is also quite attractive is, that k_Γ is, up to a positive factor, the Riemannian metric induced by the ℓ^2-norm on the covariance sequences $\{\Gamma_k\}_{k=-\infty}^{\infty}$, whereas h_Γ is induced by the ℓ^2-norm on the covariance sequences $\{\Gamma_k\}_{k=o}^{\infty}$.

Now let us consider the special case of the innovations representation (compare case (i) below (5.8-1)). We will take $D = I$, (A,B) reachable and (C,A) observable. Then the assumptions made before imply that $\sigma(A) \subseteq D(0,1)$ and $\sigma(A-BC) \subseteq D(0,1)$. Such an innovations representation always exists and is unique up to state space isomorphism (theorem (4.8-6)). (Note that we are now in the situation in which we can apply theorem (4.8-8)).
In this case the mapping (5.8-5) can be replaced by

$$(5.8\text{-}15) \qquad (A,B,C,\bar{\Omega}) \longmapsto ([(A,G,C)],J) \in M^{m,a}_{m,n,m} \times \mathrm{Pos}(m).$$

Because the state space in the innovations representation is determined up to state space isomorphism, one has the following commutative diagram

$$(5.8\text{-}16) \qquad \begin{array}{ccc} (A,B,C,\bar{\Omega}) & \xrightarrow{\phi_1} & ([(A,G,C)],J) \in M^{m,a}_{m,n,m} \times \mathrm{Pos}(m) \\ & \searrow^{\phi_2} \quad \nearrow_{\phi_3} & \\ ([(A,B,C)],\bar{\Omega}) \in M^{m,a,f}_{m,n,m} \times \mathrm{Pos}(m). & & \end{array}$$

This can be seen as follows. In this case the mappings G and J are given by

$$(5.8\text{-}17) \qquad G = G(A,B,C,\bar{\Omega}) = B\bar{\Omega} + APC^T,$$

where

(5.8-18) $P = L(B\bar{\Omega}B^T)$, as before,

and

(5.8-19) $J = J(A,B,C,\bar{\Omega}) = CPC^T + \bar{\Omega}$.

As noted above, if the basis of the state space is changed i.e. $(A,B,C) \longmapsto (SAS^{-1},SB,CS^{-1})$, then

$$G \longmapsto SG \text{ and } J \longmapsto J,$$

so the diagram (5.8-16) is commutative indeed.
Theorem (4.8-8) tells us that ϕ_2 is surjective onto $M^{m,a,f}_{m,n,m}$ and from the uniqueness of the innovations representation, up to state space isomorphism, we know that ϕ_3 is injective.
Note that the dimensions of the manifolds $M^{m,a,f}_{m,n,m} \times Pos(m)$ and $M^{m,a}_{m,n,m} \times Pos(m)$ are the same.

5.8-20. Theorem. The Riemannian metric h_Γ defined above, induces a (non-degenerate) Riemannian metric on $M^{m,a,f}_{m,n,m} \times Pos(m)$, via ϕ_3.

Before we go to the proof, let us make some remarks.

5.8-21. Remarks.

(i) The theorem is equivalent to the assertion that the Jacobian of ϕ_3 is nonsingular everywhere on $M^{m,a,f}_{m,n,m} \times Pos(m)$.

(ii) Together with theorem (4.8-9) this theorem tells us that $M^{m,a,f}_{m,n,m} \times Pos(m)$, together with the induced (non-degenerate) Riemannian metric forms a differentiable manifold isometric to the Riemannian space of stochastic systems defined above.

Proof. We will show that one has a non-degenerate Riemannian metric on $M^{m,a,f}_{m,n,m} \times Pos(m)$, by showing that any tangent vector with zero Riemannian length is zero. In fact what we will show is that a tangent vector $(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}})$ has zero Riemannian length iff $\dot{\bar{\Omega}} = 0$ and $(\dot{A},\dot{B},\dot{C})$ is of the form

$$(\dot{A},\dot{B},\dot{C}) = (-TA+AT,-TB,CT)$$

for some $T \in \mathbb{R}^{n\times n}$ (compare (5.6-11)). If $(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}})$ has zero Riemannian length, then the corresponding $(\dot{A},\dot{G},\dot{C},\dot{J})$ has zero Riemannian length in the sense of deterministic systems, cf. (5.8-15).
Therefore an argument analogous to the one following (5.6-11), combined with theorem (5.6-5), shows that

$$(\dot{A},\dot{G},\dot{C},) = (-TA+AT,-TG,CT)$$

for some $T \in \mathbb{R}^{n\times n}$. Furthermore $\dot{J} = 0$ will hold.
Now it is clear that $(-TA+AT,-TB,CT,0)$ at $(A,B,C,\bar{\Omega})$ corresponds to a zero tangent vector to the manifold $M^{m,a,f}_{m,n,m} \times \mathrm{Pos}(m)$. Therefore one can substract from $(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}})$, without loss of generality this tangent vector. Therefore we only have to prove for each tangent vector of the form $(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}}) = (0,\dot{B},0,\dot{\bar{\Omega}})$, that if it has zero Riemannian length then it is itself zero. Now consider the change in the spectral matrix due to a change $(0,\dot{B},0,\dot{\bar{\Omega}})$; because the Riemannian length of $(0,\dot{B},0,\dot{\bar{\Omega}})$ is zero this must be zero.
So we have

$$\begin{aligned} 0 &= \frac{d}{dt}[I+C(sI-A)^{-1}B_t]\bar{\Omega}_t[I+C(s^{-1}I-A)^{-1}B_t]^T = \\ &= C(sI-A)^{-1}\dot{B}\bar{\Omega}[I+C(s^{-1}I-A)^{-1}B_t]^T + \\ &\quad T(s)\dot{\bar{\Omega}}T(s^{-1})^T + T(s)\bar{\Omega}[C(s^{-1}I-A)^{-1}\dot{B}]^T, \end{aligned} \tag{5.8-22}$$

where $T(s) = I + C(sI-A)^{-1}B$. One has $T(s)^{-1} = I - C(sI-A+BC)^{-1}B$, and A-BC is asymptotically stable. Pre- and postmultiplication with $T(s)^{-1}$ and $T(s^{-1})^{-T}$ gives

$$0 = T(s)^{-1}C(sI-A)^{-1}\dot{B}\bar{\Omega} + \dot{\bar{\Omega}} + \bar{\Omega}\dot{B}^T(s^{-1}I-A^T)^{-1}C^TT(s^{-1})^{-T}. \tag{5.8-23}$$

So, clearly if $\dot{B} = 0$, then $\dot{\bar{\Omega}} = 0$ follows. Because $T(s)^{-1}$ and $(sI-A)^{-1}$ have no poles outside the open unit disk, the expression $T(s)^{-1}C(sI-A)^{-1}\dot{B}\bar{\Omega}$ can be expanded in a power series in s^{-1} around infinity, with a s^{-1} convergence radius $\rho > 1$:

$$(5.8\text{-}24) \qquad T(s)^{-1}C(sI-A)^{-1}\dot{B}\dot{\bar{\Omega}} = \sum_{k=1}^{\infty} R_k s^{-k}, \quad |s^{-1}| < \rho.$$

Substituting this in (5.8-23), one obtains:

$$(5.8\text{-}25) \qquad 0 = \sum_{k=1}^{\infty} R_k s^{-k} + \dot{\bar{\Omega}} + \sum_{k=1}^{\infty} R_k^T s^k, \quad \rho^{-1} < |s| < \rho.$$

It follows that $\dot{\bar{\Omega}} = 0$ and $R_k = 0$, $k = 1,2,3,\ldots$, and therefore

$$(5.8\text{-}26) \qquad T(s)^{-1}C(sI-A)^{-1}\dot{B}\dot{\bar{\Omega}} = 0.$$

By premultiplication with $T(s)$, and postmultiplication with $\bar{\Omega}^{-1}$, one obtains

$$(5.8\text{-}27) \qquad C(sI-A)^{-1}\dot{B} = 0.$$

It follows from the observability of (A,C) that $\dot{B} = 0$.

Q.E.D.

5.8-28. Remark. The choice of an appropriate Riemannian metric may depend on the use of the model. E.g. if the model is used for prediction, then one could argue that the coefficients of the prediction formula should be the starting point for the definition of the Riemannian metric. To be more specific, consider the following innovations representation model

$$(5.8\text{-}29) \qquad \begin{cases} x_{t+1} = Ax_t + Bw_t, & x_t \in \mathbb{R}^n \\ y_t = Cx_t + w_t, & y_t \in \mathbb{R}^m,\ w_t \in \mathbb{R}^m, \end{cases}$$

w_t Gaussian white noise, $w_t \sim N(0,\bar{\Omega})$, $\bar{\Omega} > 0$, (A,B,C) minimal, $\sigma(A) \subseteq D(0,1)$ and $\sigma(A-BC) \subseteq D(0,1)$. The corresponding steady state prediction filter is

$$(5.8\text{-}30) \qquad \begin{cases} \hat{x}_{t+1} = (A-BC)\hat{x}_t + By_t, \\ \hat{y}_{t+1} = C\hat{x}_{t+1}. \end{cases}$$

So the filter is the deterministic system $(A-BC,B,C)$, with $\{y_t\}$ as inputs, and $\{\hat{y}_{t+1}\}$ as outputs. One could now propose to use the Riemannian metric for

deterministic systems, but applied to $(A-BC,B,C)$. This means that one can use formula (5.2-39) with $A-BC$ instead of A, and $\dot{A}-B\dot{C}-BC$ instead of $\dot{A}$.
Note that in this case systems with different $\bar{\Omega}$ matrix, but equal otherwise, are considered as equivalent because they lead to the same prediction filter (5.8-30).

5.9. The Fisher metric and Hellinger distance

A important Riemannian metric for stochastic models is the socalled Fisher metric. It is the metric which has the well-known Fisher information matrix as its metric tensor. See e.g. [Ama 82], [Ama 85]. In the next subsection the Fisher metric will be applied to stochastic linear systems. In this section we want to stress the relationship between the Fisher metric and the socalled Hellinger distance between probability measures. It will turn out that the inner metric corresponding to the Fisher metric is the same as the inner metric induced by the Hellinger distance. This will also lead to a natural generalization of the Fisher metric in cases where the Fisher information matrix is not well defined, or singular. We do not claim originality in the following derivation. In fact most of the steps have been taken before (see e.g.[Daw], [Ama 85], [Kaz] and [Kaz-Pa]). Still we include this, because it shows that the Fisher metric falls into the same over all pattern as the metrics derived before, and because we believe that the importance of this relationship has not been fully realized.
Let us define the Hellinger distance. Consider a sample space Ω, together with a σ-algebra $\mathcal{F}$. Suppose three probability measures μ_1, μ_2 and λ are defined on $(\Omega,\mathcal{F})$ and μ_1 and μ_2 are absolutely continuous with respect to λ, so μ_1 and μ_2 have densities p_1 resp. p_2, say. Let $q_1 = \sqrt{p_1}$ and $q_2 = \sqrt{p_2}$ be the roots of p_1 and p_2. Then

$$\int q_i^2 d\lambda = 1, \quad i = 1,2, \tag{5.9-1}$$

and so

$$q_i \in L^2(\lambda), \quad i = 1,2.$$

One can define the following distance function

$$d(\mu_1,\mu_2) := \|q_1-q_2\|_{L^2(\lambda)} = \{\int (q_1-q_2)^2 d\lambda\}^{\frac{1}{2}}, \tag{5.9-2}$$

i.e. $d(\mu_1,\mu_2)$ is the distance in $L^2(\lambda)$ between q_1 and q_2. Now it is a very useful and important observation that this distance function is independent of the choice of the measure λ, as long as both μ_1 and μ_2 are absolutely continuous with respect to λ. One can always find a probability measure λ such that μ_1 and μ_2 are absolutely continuous with respect to λ; e.g. take $\lambda = \frac{1}{2}\mu_1 + \frac{1}{2}\mu_2$. Therefore the distance function is well defined on the set of all probability measures on $(\Omega,\mathring{F})$. (see [Kaku], who remarks that this idea was communicated to him by Von Neumann; this idea can also be found in the literature of quantum mechanics, see e.g. [Ma]).

5.9-3. Definition. (a) Consider a sample space Ω together with a σ-algebra $\mathring{F}$ of subsets of Ω. The distance d between each pair of probability measures on $(\Omega,\mathring{F})$ is called the Hellinger distance between probability measures on $(\Omega,\mathring{F})$.
(b) The corresponding inner product

$$\langle\mu_1,\mu_2\rangle := \langle q_1,q_2\rangle_{L^2(\lambda)} = \int q_1 q_2 d\lambda \tag{5.9-4}$$

is called the Hellinger affinity between μ_1 and μ_2.

5.9-5. Remarks. (i) In fact one can extend the definition to all nonnegative measures on $(\Omega,\mathring{F})$, without any problem.
(ii) For probability measures the distance can be derived from the affinity and vice versa:

$$\begin{aligned} d(\mu_1,\mu_2)^2 &= \|q_1-q_2\|^2_{L^2(\lambda)} = \langle q_1-q_2, q_1-q_2\rangle_{L^2(\lambda)} \\ &= \langle q_1,q_1\rangle_{L^2(\lambda)} + \langle q_2,q_2\rangle_{L^2(\lambda)} - 2\langle q_1,q_2\rangle_{L^2(\lambda)} = \\ &= 2-2\langle q_1,q_2\rangle_{L^2(\lambda)} = 2(1-\langle q_1,q_2\rangle_{L^2(\lambda)}) \end{aligned} \tag{5.9-6}$$

Note that $\langle q_1,q_2\rangle_{L^2(\lambda)} \in [0,1]$.

See also remark (iv).

(iii) Let $L^2(\lambda)^+$ denote the subset of $L^2(\lambda)$ consisting of nonnegative functions (the nonnegative 'orthant'). Let ψ: $L^2(\lambda)^+ \to$ {non-negative measures $<<\lambda$}, $d\psi(q) := q^2 d\lambda$. Then the set of all non-negative measures is $\bigcup_\lambda \psi(L^2(\lambda)^+)$, where λ runs through all non-negative measures. The Hellinger distance is the distance concept inherited from the $L^2(\lambda)^+$, via ψ.

(iv) Now let $B(\lambda) \subseteq L^2(\lambda)^+$ be the intersection of $L^2(\lambda)^+$ with the unit sphere around the origin in $L^2(\lambda)$. Then the set of probability measures is $\bigcup_\lambda \psi(B(\lambda))$. The Hellinger distance induces an inner metric d_B on the set of all probability measures (on a given $(\Omega,\mathring{F})$ of course): it corresponds exactly to the inner metric induced by the $L^2(\lambda)$ metric on the unit sphere $B(\lambda)$ in $L^2(\lambda)$. Therefore it can easily be computed. If q_1 and q_2 are as before, then it is just the angle between q_1 and q_2, if they are considered as vectors from the origin:

$$(5.9\text{-}7) \qquad d_B(\mu_1,\mu_2) = \arccos\langle q_1,q_2\rangle \in [0,\tfrac{\pi}{2}].$$

(cf. [Rin], p.4., formula (5)).

Because arccos is monotonic it follows that the affinity itself is also a measure of closeness, for probability measures. The geodesic between μ_1 and μ_2, along the 'sphere' of probability measures is $\mu(t)$, $t \in [0,1]$, with $\mu(0) = \mu$, $\mu(1) = \mu_2$, and

$$(5.9\text{-}8) \qquad \mu(t) = \frac{t^2\mu_1+(1-t)^2\mu_2+2t(1-t)\sqrt{(\mu_1\mu_2)}}{t^2+(1-t)^2+\langle\mu_1,\mu_2\rangle 2t(1-t)},$$

where $\sqrt{(\mu_1\mu_2)}$ denotes the geometric mean of μ_1 and μ_2, which has, by definition, density q_1q_2 with respect to λ. Here too it can be shown easily that $\sqrt{(\mu_1\mu_2)}$ is in fact independent of the choice of λ, if only μ_1 and μ_2 are both absolutely continuous with respect to λ. Note that $\int q_1q_2 d\lambda = \langle q_1,q_2\rangle_{L^2(\lambda)} = \langle\mu_1,\mu_2\rangle$, the affinity between μ_1 and μ_2.

Let $q(t)^2 = \frac{d\mu(t)}{d\lambda}$, then $q(t) = \frac{tq_1+(1-t)q_2}{\|tq_1+(1-t)q_2\|_{L^2(\lambda)}}$, from which (5.9-8) follows.

Now let $\{\mu(\theta) | \theta \in \Theta\}$ denote a parametrized set of probability measures, Θ a differentiable manifold. Suppose there exists a measure λ such that $\mu(\theta)$ is absolutely continuous with respect to λ for all $\theta \in \Theta$. Also, suppose that $\mu(\theta)$ varies 'smoothly' with θ, such that the Hellinger metric d on the probability measures, (or, what amounts to the same thing, the metric d_B on the probability measures) induces a Riemannian metric on the tangent space at Θ in θ, as follows. If $\theta = \theta(s)$, $|s| < \varepsilon$, is a smooth curve in Θ, and $\dot{\theta} = \dot{\theta}(0)$ is the tangent vector of this curve at $s = 0$, then the Riemannian metric is given, in local coordinates, which will also be denoted by θ (by (a common) abuse of notation), by:

$$(5.9\text{-}9) \qquad g_{\theta(0)}(\dot{\theta},\dot{\theta}) = \lim_{s \downarrow o} \frac{d(\mu(\theta(s)),\mu(\theta(0)))^2}{s^2} = \dot{\theta}^T \int \frac{\partial q}{\partial \theta} \frac{\partial q}{\partial \theta^T} d\lambda \, \dot{\theta}.$$

Therefore the Riemannian metric tensor is equal to

$$(5.9\text{-}10) \qquad \int \frac{\partial q}{\partial \theta} \frac{\partial q}{\partial \theta^T} d\lambda = \frac{1}{4} \int \frac{\partial \ell np}{\partial \theta} \frac{\partial \ell np}{\partial \theta^T} p d\lambda = \frac{1}{4} E\left[\frac{\partial \ell np}{\partial \theta} \frac{\partial \ell np}{\partial \theta^T} | \theta(0)\right].$$

This is-apart from the factor $\frac{1}{4}$ - the Fisher information matrix, and therefore the Riemannian metric that it describes is called the Fisher metric.

5.9-11. Remark. The Fisher metric can also be viewed as a limiting case of the Kullback-Liebler distance. The difference with the Hellinger distance is that the Kullback-Liebler distance is not symmetric, and therefore it does not satisfy the requirements for a metric, in contradistinction to the Hellinger distance.

5.10. The Fisher metric on spaces of stochastic systems

The Hellinger metric restricted to the space of stochastic systems Γ, as described at the beginning of section 5.8, turns out to be a discrete metric. (This is not so surprising, if one realizes that these systems are identifiable from the sequence of all data $\{y_t\}_{t=-\infty}^{\infty}$). This implies that the corresponding Fisher metric, as defined in section 5.9 is degenerate. If one restricts to T consecutive data $\{y_t\}_{t=1}^{T}$, then the corresponding Fisher information matrix g_T^F (= Fisher metric tensor) is nondegenerate. As it turns out,

(5.10-1) $$g^F := \lim_{T\to\infty} \frac{1}{T} g_T^F$$

exists and is nondegenerate. In the literature (5.10-1) is therefore taken as the definition of the Fisher metric for linear stochastic systems (see e.g. [Kri 77], [Ama 86]).

5.10-2. Theorem. Let $(A(\tau),B(\tau),C(\tau),\bar{\Omega}(\tau))$, $-\varepsilon < \tau < \varepsilon$, denote a smooth curve of stochastic systems in innovations representation (5.8-29) and let $(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}})$ denote the tangent vector at this curve at $\tau = 0$. Then

$$\|(\dot{A},\dot{B},\dot{C},\dot{\bar{\Omega}}\|^2_{g^F} =$$

(5.10-3)
$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\tilde{T}(e^{i\theta})^{-1} \frac{\partial\tilde{T}(e^{i\theta})}{\partial\tau} \frac{\partial\tilde{T}(e^{i\theta})^*}{\partial\tau}\{\tilde{T}(e^{i\theta})^{-1}\}^* d\theta +$$

$$+ 2\left[\mathrm{tr}\left\{(\bar{\Omega})^{-\frac{1}{2}} \frac{\partial(\bar{\Omega})^{\frac{1}{2}}}{\partial\tau}\right\}\right]^2,$$

where

$$\tilde{T}(s) := C(sI-A)^{-1}B.(\bar{\Omega})^{\frac{1}{2}} + (\bar{\Omega})^{\frac{1}{2}} = T(s)(\bar{\Omega})^{\frac{1}{2}}.$$

Proof. cf. [Meh], p. 216-217.

Various other expressions for the Fisher information matrix (and its inverse, which is the asymptotic normalized covariance matrix for maximum likelihood estimators) have been derived, cf. e.g. [Kri 77], [Go-P], section 5.5 etc.

5.10-4. Remark. $\tilde{T}(s)$ is the transfer matrix of the so-called standardized innovations representation. The standardized innovation is defined as

(5.10-5) $$v_t := (\bar{\Omega})^{-\frac{1}{2}} w_t,$$

where $w_t \sim N(0,\bar{\Omega})$ is the innovation at time t. Clearly $v_t \sim N(0,I)$. Let $U(s) = \tilde{T}(\bar{s})^{-*}$, i.e. U(s) is the transfer matrix of the system

(5.10-6)
$$\begin{cases} x_{t+1} = (A-BC)^T x_t + C^T.(\bar{\Omega})^{-\frac{1}{2}} u_t, \\ y_t = -B^T x_t + (\bar{\Omega})^{-\frac{1}{2}} u_t. \end{cases}$$

Note that the zeroes of the original system are the poles of the new system and vice versa.

5.10-7. Theorem. Let N be some open subset of a Euclidean space and consider a local parametrization $(A(\phi),B(\phi),C(\phi),\bar{\Omega}(\phi)),\phi\in N$, of stochastic systems in innovations representation (5.8-30). The Fisher information matrix of these stochastic systems with respect to this parametrization is the same as the Fisher information matrix of the following stochastic systems in innovations representation:

$$\left((A(\phi)-B(\phi)C(\phi)),^{T}C(\phi)\ ,^{T}-B(\phi)^{T},\bar{\Omega}(\phi)^{-1}\ \right),\ \phi\ \epsilon\ N.$$

Proof. Let $(\bar{\bar{\Omega}})^{\frac{1}{2}} := (\bar{\Omega})^{-\frac{1}{2}}$. Going from the original system to the new one, corresponds to going from $\tilde{T}(s)$ to $U(s)$, defined as above, and to going from $(\bar{\Omega})^{\frac{1}{2}}$ to $(\bar{\bar{\Omega}})^{\frac{1}{2}}$.
Fix a value $\phi_o\ \epsilon\ N$ and a direction $\dot{\phi}$. Let $\frac{\partial}{\partial\tau}$ denote the directional derivative in $\phi = \phi_o$ in the direction of $\dot{\phi}$. One has

$$\tilde{T}(e^{i\theta}) = U(e^{-i\theta})^{-*} \tag{5.10-8}$$

and

$$\frac{\partial\tilde{T}(e^{i\theta})}{\partial\tau} = -U(e^{-i\theta})^{-*}\ \frac{\partial U(e^{-i\theta})^{*}}{\partial\tau}U(e^{-i\theta})^{-*}. \tag{5.10-9}$$

Substituting this in (5.10-3) gives

$$\begin{aligned}&\|\dot{A},\dot{B},\dot{C};\dot{\bar{\Omega}}\|_{g_F}^{2} = \\ &\frac{1}{2\pi}\int_{-\pi}^{\pi}\mathrm{tr}\Big[U(e^{-i\theta})^{*}\Big\{-U(e^{-i\theta})^{-*}\ \frac{\partial U(e^{-i\theta})^{*}}{\partial\tau}\ U(e^{-i\theta})^{-*}\Big\}\times \\ &\Big\{-U(e^{-i\theta})^{-1}\ \frac{\partial U(e^{-i\theta})}{\partial t}U(e^{-i\theta})^{-1}\Big\}\Big]d\theta + 2\Big\{\mathrm{tr}\Big[(\bar{\bar{\Omega}})^{\frac{1}{2}}\ \frac{\partial(\bar{\bar{\Omega}})^{-\frac{1}{2}}}{\partial\tau}\Big]\Big\}^{2} =\end{aligned} \tag{5.10-10}$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[U(e^{-i\theta})^{-1}\,\frac{\partial U(e^{-i\theta})}{\partial \tau}\,\frac{\partial U(e^{-i\theta})^{*}}{\partial \tau}\,U(e^{-i\theta})^{-*}\right]d\theta +$$

$$+\ 2\left\{\mathrm{tr}\left[(\bar{\bar{\Omega}})^{-\frac{1}{2}}\,\frac{\partial(\bar{\bar{\Omega}})}{\partial \tau}\right]\right\}^2 =$$

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{tr}\left[U(e^{i\theta})^{-1}\,\frac{\partial U(e^{i\theta})}{\partial \tau}\,\frac{\partial U(e^{i\theta})^{*}}{\partial \tau}\,U(e^{i\theta})^{-*}\right]d\theta + 2\left\{\mathrm{tr}\left[(\bar{\bar{\Omega}})^{-\frac{1}{2}}\,\frac{\partial(\bar{\bar{\Omega}})}{\partial \tau}\right]\right\}^2.$$

The last formula is exactly the right-hand side of (5.10-3) with $\tilde{T}$ replaced by U and $(\bar{\Omega})^{\frac{1}{2}}$ by $(\bar{\bar{\Omega}})^{\frac{1}{2}}$.

Q.E.D.

5.10-11. Corollary. The mapping of stochastic systems given in terms of their innovations representations by

(5.10-12) $\quad (A,B,C;\bar{\Omega}) \longmapsto \left((A-BC)^{T},C^{T},-B^{T};(\bar{\Omega})^{-1}\right)$

is an isometry for the Fisher metric.

5.10-13. Comment. From this theorem it follows that if one approaches the boundary of the manifold of stochastic systems (which consists of systems with zeroes on the unit circle) then the Fisher metric, and the corresponding inner metric, will be the same as when one approaches the unstable systems, using the mapping (5.10-12). If an unstable system is infinitely far, (as is always the case in the metric of section 5.8) then the same holds for the system with zeroes on the unit circle, that corresponds to it via (5.10-12). It turns out, however, that this is not always the case. Whether it is sometimes the case or not is still an open question, which I am trying to figure out. If it turns out to be the case sometimes, i.e. if there are (asymptotically stable) stochastic systems with zeroes on the unit circle that are infinitely far from stochastic systems with no zeroes on the unit circle, then this would be a disadvantage for the Fisher metric on stochastic systems, as defined by (5.10-1). The metric defined in section 5.8 does not have this disadvantage. Why would one call such a property a disadvantage? Because stochastic systems with zeroes on the unit circle are intuitively not exceptional from the point of view of identification of stochastic systems. (They are exceptional from the point of view of prediction, because the corresponding predictor is not asymptotically stable).

The definition of (5.10-1) could also be based on asymptotic maximum likelihood estimator theory, because the inverse of g^F is the asymptotic covariance matrix of the maximum likelihood estimator (see e.g. [Go-P], section 5.0, although they do not mention the word Fisher information matrix explicitly there). The results for the dynamic case are comparable to the case of i.i.d. random variables, as long as the prediction errors of a model (which needs not be the true one) at times far apart are almost uncorrelated. (See also the next chapter.) However, this may no longer hold for systems with zeroes on the unit circle. All this suggests that further research of the behaviour of the Fisher metric in the neighbourhood of the boundary of the manifold is interesting and important.

5.11. Some remarks on identifiability and Riemannian geometry

It is well-known that if the Fisher information matrix is nonsingular at a point in the parameter space, then the parametrization (of the model space) is locally identifiable at that parameter point. (cf. [Ro]). This can be easily understood if the Fisher information matrix is interpreted as metric tensor of the Fisher metric, and the result can be immediately generalized to other Riemannian metrics. Also this geometrical interpretation points the way how these results can be generalized.

In section 3.2.1.3 local (parameter-) identifiability was defined. Here, again, we assume that the model space is system identifiable. The following theorem is standard differential geometry.

5.11-1. Theorem. Consider a Riemannian manifold M,d = dim (M). Let

ϕ_i: $\Theta_i \to M$, $i \in I$ be an indexed family of mappings,

$\Theta_i \subseteq \mathbb{R}^d$, Θ_i open, $\forall i \in I$, such that

$$\bigcup_{i \in I} \mu(\Theta_i) = M.$$

Let furthermore the Riemannian metric tensor with respect to the parameters in Θ_i be well-defined and nonsingular, for all $\theta \in \Theta_i$, and for all $i \in I$. Then the $\{(\Theta_i, \phi_i) \mid i \in I\}$ form a (complete) set of coordinate charts for the manifold. It follows immediately that the corresponding parametrization ϕ (notation as in section 3.2.1.3) is locally identifiable.

One can deduce from this that a parametrization ϕ: $\Theta \to M$, with

$\Theta = \cup \{(\Theta_i, i) | i \in I\}$, Θ_i open, $\Theta_i \subseteq \mathbb{R}^d$, $\forall i \in I$, is <u>locally identifiable at</u> $\theta \in \Theta_i$, $i \in I$, if the Riemannian metric tensor with respect to the parameters in Θ_i is <u>well-defined and nonsingular at θ.</u>

Because the Fisher information matrix is a metric tensor it follows immediately that if for a given parametrization the Fisher information matrix is well-defined and nonsingular at a point, then the parametrization is locally identifiable there. But the same conclusion holds for any other Riemannian metric tensor! Because the Fisher information matrix is rather difficult to compute for stochastic linear dynamical systems, one can, for testing local identifiability, consider to compute the determinant of the Riemannian metric tensor for such systems that we proposed in section 5.8. If this determinant is nonzero then local identifiability follows!.

Appendix 5A. Symbolic computation of the Riemannian metric tensors

Making use of the formulas in theorem (5.2-38) in the deterministic case and of the formulas in section 5.8, especially (5.8-8) ff., for the stochastic case, W Mak has computed symbolically the Riemannian metric tensor (R.m.t.) for a number of cases, using the computer program 'Reduce'. The parametrization has been chosen as follows

$$(5A\text{-}1) \qquad A = \begin{bmatrix} 0 & \cdots & 0 & -a_0 \\ 1 & & \vdots & \vdots \\ 0 & & 0 & -a_{n-2} \\ \vdots & \ddots & & \\ 0 & \cdots & 1 & -a_{n-1} \end{bmatrix}, \quad B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad C = (0,0,\ldots,0,1),$$

with n = 1,2 or 3. Due to the large amount of computer memory space that is used by Reduce, only in these cases (apart from some minor variations) the R.m.t. has been computed, for the corresponding deterministic continuous - time systems. For the deterministic discrete-time systems the R.m.t. is computed for n = 1,2 and for the stochastic discrete-time case the R.m.t. is computed for n = 1 only. (For the cases with more than one input, but with one output ('m.i.s.o') the R.m.t. can easily be derived from the s.i.s.o. case; by dualization one can then also obtain the R.m.t. for the 's.i.m.o.' case). Let a tangent vector be denoted by

$$t = (\dot{a}_0, \dot{a}_1, \ldots, \dot{a}_{n-1}, \dot{b}_1, \dot{b}_2, \ldots, \dot{b}_n) \in \mathbb{R}^{2n},$$

and let the Riemannian metric tensor R be such that the Riemannian length of t is given by $(t^T R t)^{\frac{1}{2}}$. For the various cases R is as follows. (To avoid confusion, note the minus signs in the parameters of the matrix A in (5A-1). This differs from the parametrization used in section 5.5!).

The deterministic continuous-time case

If $n = 1$, $R = \begin{bmatrix} \frac{b_1^2}{4a_o^3} & \frac{-b_1}{4a_o^2} \\ \frac{-b_1}{4a_o^2} & \frac{1}{2a_o} \end{bmatrix}$

If $n = 2$, $R =$

$$\begin{bmatrix} \frac{a_o^2 b_2^2 + a_o b_1^2 + a_1^2 b_1^2}{4a_o^3 a_1^3} & 0 & -\left(\frac{a_o b_2 + a_1 b_1}{4a_o^2 a_1^2}\right) & \frac{b_1}{4a_o a_1^2} \\ 0 & \frac{a_o b_2^2 + b_1^2}{4a_o a_1^3} & \frac{-b_1}{4a_o a_1^2} & \frac{-b_2}{4a_1^2} \\ -\left(\frac{a_o b_2 + a_1 b_1}{4a_o^2 a_1^2}\right) & \frac{-b_1}{4a_o a_1^2} & \frac{1}{2a_o a_1} & 0 \\ \frac{b_1}{4a_o a_1^2} & \frac{-b_2}{4a_1^2} & 0 & \frac{1}{2a_1} \end{bmatrix}$$

If $n = 3$ the expressions become rather big, and will therefore not be given here. We refer to [Mak]; a copy of that paper can be sent upon request.

The deterministic discrete-time case.

If $n = 1$, $R = \begin{bmatrix} \frac{-b_1^2(a_o^2+1)}{(a_o^2-1)^3} & \frac{a_o b_1}{(a_o^2-1)^2} \\ \frac{a_o b_1}{(a_o^2-1)^2} & \frac{-1}{a_o^2-1} \end{bmatrix}$

If $n = 2$, the expressions are again rather big and we refer again to [Mak]. It is remarkable that the parametrization for the corresponding continuous-time systems leads to a much more concise representation of the Riemannian metric!

The stochastic discrete-time case

It should be noted that without loss of generality, one takes here $\bar{\Omega} = 1$ and $D = 0$ (cf. section 5.8). If $n = 1$,

$$R = \begin{bmatrix} \dfrac{-(9a_o^2+1)b_1^4}{(a_o^2-1)^5} & \dfrac{6a_o b_1^3}{(a_o^2-1)^4} \\ \dfrac{6a_o b_1^3}{(a_o^2-1)^4} & \dfrac{-4b_1^2}{(a_o^2-1)^3} \end{bmatrix}.$$

Making use of these R.m.t's, the curvature of the Riemannian metrics has been computed with Reduce in the two-dimensional model space cases (in which the curvature is given by a scalar), i.e. in the cases with $n = 1$. For the spaces $M_{1,1,1}^{m,a}$ and $M_{1,1,1}^{m,\ell}$ of deterministic systems it did not come as a surprise that the curvature turned out to be zero, because this already followed from the fact that the space is isometric to a Riemann surface (cf. section 5.5.1). But it did come as a surprise that in the stochastic discrete-time case, the curvature is also zero. Further analysis must show whether in this case too, the space is isometric to a Riemann surface.

MATHEMATICAL CENTRE TRACTS

1 T. van der Walt. *Fixed and almost fixed points.* 1963.

2 A.R. Bloemena. *Sampling from a graph.* 1964.

3 G. de Leve. *Generalized Markovian decision processes, part I: model and method.* 1964.

4 G. de Leve. *Generalized Markovian decision processes, part II: probabilistic background.* 1964.

5 G. de Leve, H.C. Tijms, P.J. Weeda. *Generalized Markovian decision processes, applications.* 1970.

6 M.A. Maurice. *Compact ordered spaces.* 1964.

7 W.R. van Zwet. *Convex transformations of random variables.* 1964.

8 J.A. Zonneveld. *Automatic numerical integration.* 1964.

9 P.C. Baayen. *Universal morphisms.* 1964.

10 E.M. de Jager. *Applications of distributions in mathematical physics.* 1964.

11 A.B. Paalman-de Miranda. *Topological semigroups.* 1964.

12 J.A.Th.M. van Berckel, H. Brandt Corstius, R.J. Mokken, A. van Wijngaarden. *Formal properties of newspaper Dutch.* 1965.

13 H.A. Lauwerier. *Asymptotic expansions.* 1966, out of print; replaced by MCT 54.

14 H.A. Lauwerier. *Calculus of variations in mathematical physics.* 1966.

15 R. Doornbos. *Slippage tests.* 1966.

16 J.W. de Bakker. *Formal definition of programming languages with an application to the definition of ALGOL 60.* 1967.

17 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 1.* 1968.

18 R.P. van de Riet. *Formula manipulation in ALGOL 60, part 2.* 1968.

19 J. van der Slot. *Some properties related to compactness.* 1968.

20 P.J. van der Houwen. *Finite difference methods for solving partial differential equations.* 1968.

21 E. Wattel. *The compactness operator in set theory and topology.* 1968.

22 T.J. Dekker. *ALGOL 60 procedures in numerical algebra, part 1.* 1968.

23 T.J. Dekker, W. Hoffmann. *ALGOL 60 procedures in numerical algebra, part 2.* 1968.

24 J.W. de Bakker. *Recursive procedures.* 1971.

25 E.R. Paërl. *Representations of the Lorentz group and projective geometry.* 1969.

26 European Meeting 1968. *Selected statistical papers, part I.* 1968.

27 European Meeting 1968. *Selected statistical papers, part II.* 1968.

28 J. Oosterhoff. *Combination of one-sided statistical tests.* 1969.

29 J. Verhoeff. *Error detecting decimal codes.* 1969.

30 H. Brandt Corstius. *Exercises in computational linguistics.* 1970.

31 W. Molenaar. *Approximations to the Poisson, binomial and hypergeometric distribution functions.* 1970.

32 L. de Haan. *On regular variation and its application to the weak convergence of sample extremes.* 1970.

33 F.W. Steutel. *Preservation of infinite divisibility under mixing and related topics.* 1970.

34 I. Juhász, A. Verbeek, N.S. Kroonenberg. *Cardinal functions in topology.* 1971.

35 M.H. van Emden. *An analysis of complexity.* 1971.

36 J. Grasman. *On the birth of boundary layers.* 1971.

37 J.W. de Bakker, G.A. Blaauw, A.J.W. Duijvestijn, E.W. Dijkstra, P.J. van der Houwen, G.A.M. Kamsteeg-Kemper, F.E.J. Kruseman Aretz, W.L. van der Poel, J.P. Schaap-Kruseman, M.V. Wilkes, G. Zoutendijk. *MC-25 Informatica Symposium.* 1971.

38 W.A. Verloren van Themaat. *Automatic analysis of Dutch compound words.* 1972.

39 H. Bavinck. *Jacobi series and approximation.* 1972.

40 H.C. Tijms. *Analysis of (s,S) inventory models.* 1972.

41 A. Verbeek. *Superextensions of topological spaces.* 1972.

42 W. Vervaat. *Success epochs in Bernoulli trials (with applications in number theory).* 1972.

43 F.H. Ruymgaart. *Asymptotic theory of rank tests for independence.* 1973.

44 H. Bart. *Meromorphic operator valued functions.* 1973.

45 A.A. Balkema. *Monotone transformations and limit laws.* 1973.

46 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 1: the language.* 1973.

47 R.P. van de Riet. *ABC ALGOL, a portable language for formula manipulation systems, part 2: the compiler.* 1973.

48 F.E.J. Kruseman Aretz, P.J.W. ten Hagen, H.L. Oudshoorn. *An ALGOL 60 compiler in ALGOL 60, text of the MC-compiler for the EL-X8.* 1973.

49 H. Kok. *Connected orderable spaces.* 1974.

50 A. van Wijngaarden, B.J. Mailloux, J.E.L. Peck, C.H.A. Koster, M. Sintzoff, C.H. Lindsey, L.G.L.T. Meertens, R.G. Fisker (eds.). *Revised report on the algorithmic language ALGOL 68.* 1976.

51 A. Hordijk. *Dynamic programming and Markov potential theory.* 1974.

52 P.C. Baayen (ed.). *Topological structures.* 1974.

53 M.J. Faber. *Metrizability in generalized ordered spaces.* 1974.

54 H.A. Lauwerier. *Asymptotic analysis, part 1.* 1974.

55 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 1: theory of designs, finite geometry and coding theory.* 1974.

56 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry.* 1974.

57 M. Hall, Jr., J.H. van Lint (eds.). *Combinatorics, part 3: combinatorial group theory.* 1974.

58 W. Albers. *Asymptotic expansions and the deficiency concept in statistics.* 1975.

59 J.L. Mijnheer. *Sample path properties of stable processes.* 1975.

60 F. Göbel. *Queueing models involving buffers.* 1975.

63 J.W. de Bakker (ed.). *Foundations of computer science.* 1975.

64 W.J. de Schipper. *Symmetric closed categories.* 1975.

65 J. de Vries. *Topological transformation groups, 1: a categorical approach.* 1975.

66 H.G.J. Pijls. *Logically convex algebras in spectral theory and eigenfunction expansions.* 1976.

68 P.P.N. de Groen. *Singularly perturbed differential operators of second order.* 1976.

69 J.K. Lenstra. *Sequencing by enumerative methods.* 1977.

70 W.P. de Roever, Jr. *Recursive program schemes: semantics and proof theory.* 1976.

71 J.A.E.E. van Nunen. *Contracting Markov decision processes.* 1976.

72 J.K.M. Jansen. *Simple periodic and non-periodic Lamé functions and their applications in the theory of conical waveguides.* 1977.

73 D.M.R. Leivant. *Absoluteness of intuitionistic logic.* 1979.

74 H.J.J. te Riele. *A theoretical and computational study of generalized aliquot sequences.* 1976.

75 A.E. Brouwer. *Treelike spaces and related connected topological spaces.* 1977.

76 M. Rem. *Associons and the closure statement.* 1976.

77 W.C.M. Kallenberg. *Asymptotic optimality of likelihood ratio tests in exponential families.* 1978.

78 E. de Jonge, A.C.M. van Rooij. *Introduction to Riesz spaces.* 1977.

79 M.C.A. van Zuijlen. *Emperical distributions and rank statistics.* 1977.

80 P.W. Hemker. *A numerical study of stiff two-point boundary problems.* 1977.

81 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 1.* 1976.

82 K.R. Apt, J.W. de Bakker (eds.). *Foundations of computer science II, part 2.* 1976.

83 L.S. van Benthem Jutting. *Checking Landau's "Grundlagen" in the AUTOMATH system.* 1979.

84 H.L.L. Busard. *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?), books vii-xii.* 1977.

85 J. van Mill. *Supercompactness and Wallman spaces.* 1977.

86 S.G. van der Meulen, M. Veldhorst. *Torrix I, a programming system for operations on vectors and matrices over arbitrary fields and of variable size.* 1978.

88 A. Schrijver. *Matroids and linking systems.* 1977.

89 J.W. de Roever. *Complex Fourier transformation and analytic functionals with unbounded carriers.* 1978.

90 L.P.J. Groenewegen. *Characterization of optimal strategies in dynamic games.* 1981.

91 J.M. Geysel. *Transcendence in fields of positive characteristic.* 1979.

92 P.J. Weeda. *Finite generalized Markov programming.* 1979.

93 H.C. Tijms, J. Wessels (eds.). *Markov decision theory.* 1977.

94 A. Bijlsma. *Simultaneous approximations in transcendental number theory.* 1978.

95 K.M. van Hee. *Bayesian control of Markov chains.* 1978.

96 P.M.B. Vitányi. *Lindenmayer systems: structure, languages, and growth functions.* 1980.

97 A. Federgruen. *Markovian control problems; functional equations and algorithms.* 1984.

98 R. Geel. *Singular perturbations of hyperbolic type.* 1978.

99 J.K. Lenstra, A.H.G. Rinnooy Kan, P. van Emde Boas (eds.). *Interfaces between computer science and operations research.* 1978.

100 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1.* 1979.

101 P.C. Baayen, D. van Dulst, J. Oosterhoff (eds.). *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2.* 1979.

102 D. van Dulst. *Reflexive and superreflexive Banach spaces.* 1978.

103 K. van Harn. *Classifying infinitely divisible distributions by functional equations.* 1978.

104 J.M. van Wouwe. *Go-spaces and generalizations of metrizability.* 1979.

105 R. Helmers. *Edgeworth expansions for linear combinations of order statistics.* 1982.

106 A. Schrijver (ed.). *Packing and covering in combinatorics.* 1979.

107 C. den Heijer. *The numerical solution of nonlinear operator equations by imbedding methods.* 1979.

108 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 1.* 1979.

109 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science III, part 2.* 1979.

110 J.C. van Vliet. *ALGOL 68 transput, part I: historical review and discussion of the implementation model.* 1979.

111 J.C. van Vliet. *ALGOL 68 transput, part II: an implementation model.* 1979.

112 H.C.P. Berbee. *Random walks with stationary increments and renewal theory.* 1979.

113 T.A.B. Snijders. *Asymptotic optimality theory for testing problems with restricted alternatives.* 1979.

114 A.J.E.M. Janssen. *Application of the Wigner distribution to harmonic analysis of generalized stochastic processes.* 1979.

115 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 1.* 1979.

116 P.C. Baayen, J. van Mill (eds.). *Topological structures II, part 2.* 1979.

117 P.J.M. Kallenberg. *Branching processes with continuous state space.* 1979.

118 P. Groeneboom. *Large deviations and asymptotic efficiencies.* 1980.

119 F.J. Peters. *Sparse matrices and substructures, with a novel implementation of finite element algorithms.* 1980.

120 W.P.M. de Ruyter. *On the asymptotic analysis of large-scale ocean circulation.* 1980.

121 W.H. Haemers. *Eigenvalue techniques in design and graph theory.* 1980.

122 J.C.P. Bus. *Numerical solution of systems of nonlinear equations.* 1980.

123 I. Yuhász. *Cardinal functions in topology - ten years later.* 1980.

124 R.D. Gill. *Censoring and stochastic integrals.* 1980.

125 R. Eising. *2-D systems, an algebraic approach.* 1980.

126 G. van der Hoek. *Reduction methods in nonlinear programming.* 1980.

127 J.W. Klop. *Combinatory reduction systems.* 1980.

128 A.J.J. Talman. *Variable dimension fixed point algorithms and triangulations.* 1980.

129 G. van der Laan. *Simplicial fixed point algorithms.* 1980.

130 P.J.W. ten Hagen, T. Hagen, P. Klint, H. Noot, H.J. Sint, A.H. Veen. *ILP: intermediate language for pictures.* 1980.

131 R.J.R. Back. *Correctness preserving program refinements: proof theory and applications.* 1980.

132 H.M. Mulder. *The interval function of a graph.* 1980.

133 C.A.J. Klaassen. *Statistical performance of location estimators.* 1981.

134 J.C. van Vliet, H. Wupper (eds.). *Proceedings international conference on ALGOL 68.* 1981.

135 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part I.* 1981.

136 J.A.G. Groenendijk, T.M.V. Janssen, M.J.B. Stokhof (eds.). *Formal methods in the study of language, part II.* 1981.

137 J. Telgen. *Redundancy and linear programs.* 1981.

138 H.A. Lauwerier. *Mathematical models of epidemics.* 1981.

139 J. van der Wal. *Stochastic dynamic programming, successive approximations and nearly optimal strategies for Markov decision processes and Markov games.* 1981.

140 J.H. van Geldrop. *A mathematical theory of pure exchange economies without the no-critical-point hypothesis.* 1981.

141 G.E. Welters. *Abel-Jacobi isogenies for certain types of Fano threefolds.* 1981.

142 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 1.* 1981.

143 J.M. Schumacher. *Dynamic feedback in finite- and infinite-dimensional linear systems.* 1981.

144 P. Eijgenraam. *The solution of initial value problems using interval arithmetic; formulation and analysis of an algorithm.* 1981.

145 A.J. Brentjes. *Multi-dimensional continued fraction algorithms.* 1981.

146 C.V.M. van der Mee. *Semigroup and factorization methods in transport theory.* 1981.

147 H.H. Tigelaar. *Identification and informative sample size.* 1982.

148 L.C.M. Kallenberg. *Linear programming and finite Markovian control problems.* 1983.

149 C.B. Huijsmans, M.A. Kaashoek, W.A.J. Luxemburg, W.K. Vietsch (eds.). *From A to Z, proceedings of a symposium in honour of A.C. Zaanen.* 1982.

150 M. Veldhorst. *An analysis of sparse matrix storage schemes.* 1982.

151 R.J.M.M. Does. *Higher order asymptotics for simple linear rank statistics.* 1982.

152 G.F. van der Hoeven. *Projections of lawless sequences.* 1982.

153 J.P.C. Blanc. *Application of the theory of boundary value problems in the analysis of a queueing model with paired services.* 1982.

154 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part I.* 1982.

155 H.W. Lenstra, Jr., R. Tijdeman (eds.). *Computational methods in number theory, part II.* 1982.

156 P.M.G. Apers. *Query processing and data allocation in distributed database systems.* 1983.

157 H.A.W.M. Kneppers. *The covariant classification of two-dimensional smooth commutative formal groups over an algebraically closed field of positive characteristic.* 1983.

158 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 1.* 1983.

159 J.W. de Bakker, J. van Leeuwen (eds.). *Foundations of computer science IV, distributed systems, part 2.* 1983.

160 A. Rezus. *Abstract AUTOMATH.* 1983.

161 G.F. Helminck. *Eisenstein series on the metaplectic group, an algebraic approach.* 1983.

162 J.J. Dik. *Tests for preference.* 1983.

163 H. Schippers. *Multiple grid methods for equations of the second kind with applications in fluid mechanics.* 1983.

164 F.A. van der Duyn Schouten. *Markov decision processes with continuous time parameter.* 1983.

165 P.C.T. van der Hoeven. *On point processes.* 1983.

166 H.B.M. Jonkers. *Abstraction, specification and implementation techniques, with an application to garbage collection.* 1983.

167 W.H.M. Zijm. *Nonnegative matrices in dynamic programming.* 1983.

168 J.H. Evertse. *Upper bounds for the numbers of solutions of diophantine equations.* 1983.

169 H.R. Bennett, D.J. Lutzer (eds.). *Topology and order structures, part 2.* 1983.

CWI TRACTS

1 D.H.J. Epema. *Surfaces with canonical hyperplane sections.* 1984.

2 J.J. Dijkstra. *Fake topological Hilbert spaces and characterizations of dimension in terms of negligibility.* 1984.

3 A.J. van der Schaft. *System theoretic descriptions of physical systems.* 1984.

4 J. Koene. *Minimal cost flow in processing networks, a primal approach.* 1984.

5 B. Hoogenboom. *Intertwining functions on compact Lie groups.* 1984.

6 A.P.W. Böhm. *Dataflow computation.* 1984.

7 A. Blokhuis. *Few-distance sets.* 1984.

8 M.H. van Hoorn. *Algorithms and approximations for queueing systems.* 1984.

9 C.P.J. Koymans. *Models of the lambda calculus.* 1984.

10 C.G. van der Laan, N.M. Temme. *Calculation of special functions: the gamma function, the exponential integrals and error-like functions.* 1984.

11 N.M. van Dijk. *Controlled Markov processes; time-discretization.* 1984.

12 W.H. Hundsdorfer. *The numerical solution of nonlinear stiff initial value problems: an analysis of one step methods.* 1985.

13 D. Grune. *On the design of ALEPH.* 1985.

14 J.G.F. Thiemann. *Analytic spaces and dynamic programming: a measure theoretic approach.* 1985.

15 F.J. van der Linden. *Euclidean rings with two infinite primes.* 1985.

16 R.J.P. Groothuizen. *Mixed elliptic-hyperbolic partial differential operators: a case-study in Fourier integral operators.* 1985.

17 H.M.M. ten Eikelder. *Symmetries for dynamical and Hamiltonian systems.* 1985.

18 A.D.M. Kester. *Some large deviation results in statistics.* 1985.

19 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 1: Philosophy, framework, computer science.* 1986.

20 B.F. Schriever. *Order dependence.* 1986.

21 D.P. van der Vecht. *Inequalities for stopped Brownian motion.* 1986.

22 J.C.S.P. van der Woude. *Topological dynamix.* 1986.

23 A.F. Monna. *Methods, concepts and ideas in mathematics: aspects of an evolution.* 1986.

24 J.C.M. Baeten. *Filters and ultrafilters over definable subsets of admissible ordinals.* 1986.

25 A.W.J. Kolen. *Tree network and planar rectilinear location theory.* 1986.

26 A.H. Veen. *The misconstrued semicolon: Reconciling imperative languages and dataflow machines.* 1986.

27 A.J.M. van Engelen. *Homogeneous zero-dimensional absolute Borel sets.* 1986.

28 T.M.V. Janssen. *Foundations and applications of Montague grammar, part 2: Applications to natural language.* 1986.

29 H.L. Trentelman. *Almost invariant subspaces and high gain feedback.* 1986.

30 A.G. de Kok. *Production-inventory control models: approximations and algorithms.* 1987.

31 E.E.M. van Berkum. *Optimal paired comparison designs for factorial experiments.* 1987.

32 J.H.J. Einmahl. *Multivariate empirical processes.* 1987.

33 O.J. Vrieze. *Stochastic games with finite state and action spaces.* 1987.

34 P.H.M. Kersten. *Infinitesimal symmetries: a computational approach.* 1987.

35 M.L. Eaton. *Lectures on topics in probability inequalities.* 1987.

36 A.H.P. van der Burgh, R.M.M. Mattheij (eds.). *Proceedings of the first international conference on industrial and applied mathematics (ICIAM 87).* 1987.

37 L. Stougie. *Design and analysis of algorithms for stochastic integer programming.* 1987.

38 J.B.G. Frenk. *On Banach algebras, renewal measures and regenerative processes.* 1987.

39 H.J.M. Peters, O.J. Vrieze (eds.). *Surveys in game theory and related topics.* 1987.

40 J.L. Geluk, L. de Haan. *Regular variation, extensions and Tauberian theorems.* 1987.

41 Sape J. Mullender (ed.). *The Amoeba distributed operating system: Selected papers 1984-1987.* 1987.

42 P.R.J. Asveld, A. Nijholt (eds.). *Essays on concepts, formalisms, and tools.* 1987.

43 H.L. Bodlaender. *Distributed computing: structure and complexity.* 1987.

44 A.W. van der Vaart. *Statistical estimation in large parameter spaces.* 1988.

45 S.A. van de Geer. *Regression analysis and empirical processes.* 1988.

46 S.P. Spekreijse. *Multigrid solution of the steady Euler equations.* 1988.

47 J.B. Dijkstra. *Analysis of means in some non-standard situations.* 1988.

48 F.C. Drost. *Asymptotics for generalized chi-square goodness-of-fit tests.* 1988.

49 F.W. Wubs. *Numerical solution of the shallow-water equations.* 1988.

50 F. de Kerf. *Asymptotic analysis of a class of perturbed Korteweg-de Vries initial value problems.* 1988.

51 P.J.M. van Laarhoven. *Theoretical and computational aspects of simulated annealing.* 1988.

52 P.M. van Loon. *Continuous decoupling transformations for linear boundary value problems.* 1988.

53 K.C.P. Machielsen. *Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space.* 1988.

54 L.C.R.J. Willenborg. *Computational aspects of survey data processing.* 1988.

55 G.J. van der Steen. *A program generator for recognition, parsing and transduction with syntactic patterns.* 1988.

56 J.C. Ebergen. *Translating programs into delay-insensitive circuits.* 1989.

57 S.M. Verduyn Lunel. *Exponential type calculus for linear delay equations.* 1989.

58 M.C.M. de Gunst. *A random model for plant cell population growth.* 1989.

59 D. van Dulst. *Characterizations of Banach spaces not containing l^1.* 1989.

60 H.E. de Swart. *Vacillation and predictability properties of low-order atmospheric spectral models.* 1989.

61 P. de Jong. *Central limit theorems for generalized multilinear forms.* 1989.

62 V.J. de Jong. *A specification system for statistical software.* 1989.

63 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part I.* 1989.

64 B. Hanzon. *Identifiability, recursive identification and spaces of linear dynamical systems, part II.* 1989.

65 B.M.M. de Weger. *Algorithms for diophantine equations.* 1989.

66 A. Jung. *Cartesian closed categories of domains.* 1989.

67 J.W. Polderman. *Adaptive control & identification: Conflict of conflux?.* 1989.

68 H.J. Woerdeman. *Matrix and operator extensions.* 1989.

69 B.G. Hansen. *Monotonicity properties of infinitely divisible distributions.* 1989.

70 J.K. Lenstra, H.C. Tijms, A. Volgenant (eds.). *Twenty-five years of operations research in the Netherlands: Papers dedicated to Gijs de Leve.* 1990.

71 P.J.C. Spreij. *Counting process systems. Identification and stochastic realization.* 1990.

72 J.F. Kaashoek. *Modeling one dimensional pattern formation by anti-diffusion.* 1990.

73 A.M.H. Gerards. *Graphs and polyhedra. Binary spaces and cutting planes.* 1990.

74 B. Koren. *Multigrid and defect correction for the steady Navier-Stokes equations. Application to aerodynamics.* 1990.

75 M.W.P. Savelsbergh. *Computer aided routing.* 1990.