TIME-VARYING SYSTEM THEORY AND COMPUTATIONAL MODELING

Realization, Approximation, and Factorization

TIME-VARYING SYSTEM THEORY AND COMPUTATIONAL MODELING

Realization, Approximation, and Factorization

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Delft, op gezag van de Rector Magnificus, prof. drs. P.A. Schenck, in het openbaar te verdedigen ten overstaan van een commissie aangewezen door het College van Dekanen, op 1 juni 1993 te 14.00 uur door

Alle-Jan van der Veen

geboren te Voorschoten
elektrotechnisch ingenieur

Dit proefschrift is goedgekeurd door de promotor,
Prof. dr. ir. P.M. Dewilde

ISBN 90-5326-005-6 / CIP

SUMMARY

Time-varying linear systems are an important generalization of the more familiar time-invariant concept. We study time-varying systems in discrete time as bounded input-output operators acting on the Hilbert space of ℓ_2-sequences. Such operators have matrix representations, and in our notation, causal systems correspond to operators whose matrix representations are upper triangular. We allow the number of inputs and outputs of systems to be time-varying, which yields matrix representations in which the entries are block matrices themselves. The block entries need not have equal dimensions, and we allow that some (or most) of the dimensions are zero. If all but a finite number of dimensions are zero, then the input-output operator reduces to a finite (block) matrix, and the application of the system to an input sequence reduces to a finite matrix-vector multiplication. This gives a connection between the fields of linear algebra and linear system theory which proves to be quite fruitful.

For general linear time-varying systems, we are in particular interested in *state realizations*. Such a realization can be viewed as the *computational model* by which the system computes the output sequence from its input sequence; the states are the intermediate quantities in the computation. The following aspects are treated.

- *Realization theory:* given the input-output operator of a bounded causal system, determine a minimal state realization. An important role is played by a generalization of the Hankel operator, familiar in time-invariant system theory.
- *Optimal model reduction:* given a state realization of a system, determine an 'optimal' approximant of it that has lower state dimensions. The approximation norm that is used is a generalization of the time-invariant Hankel norm as introduced in the theory of Adamjan, Arov, and Krein. The construction of a state realization of the approximant and the derivation of its complexity are among the main results of the thesis.
- *Inner-outer factorization, spectral factorization, and lossless embedding ('unitary ex-*

tension') play an important role in time-invariant system theory. State-space algorithms are derived to compute such factorizations for time-varying systems, which gives Riccati equations with time-varying coefficients. These results are instrumental in, for example, the solution of robust control problems for time-varying systems.

- *Cascade factorization of inner systems,* which leads to numerically stable implementations of contractive systems, using a minimal number of parameters.

Finite (block) matrices can be viewed as special cases of time-varying systems, and with this interpretation, the above issues translate to new insights in the field of *computational linear algebra.* In particular, a state realization of a large matrix which has low state dimensions represents a computational model of low complexity by which a matrix-vector multiplication can be done. When a state realization is known, the connection with system theory provides efficient ways to do *matrix inversion, Cholesky factorization* and *QR factorization.* In addition, the Hankel-norm model reduction theory can be used to derive, for the given matrix, an approximating computational model of lower complexity.

PREFACE

According to Israel Gohberg, in an after-dinner speech delivered at the end of a meeting in Amsterdam (1992), "Considering extensions to time-varying systems gives 'cheap generalizations'." And indeed, a reader versed in time-invariant system theory will find that many of the results in this thesis look at least familiar. Despite these 'trivialities', this book has grown to be rather heavy. The catch lies, of course, in proving the details, and in discovering which facts from time-invariant system theory do generalize, and which facts don't. The facts which don't readily generalize typically make use of frequency-domain techniques: Fourier transforms, poles, winding numbers, and although each of these notions can be extended in some sense, the extensions turned out to be not really applicable. Some other results which require 'non-trivial' new proofs are those which, in the time-invariant context, make explicit use of the eigenvalues or invertibility of certain matrices, whereas, in the generalization, these matrices need not even be square.

A second reason for the bulkiness of this book lies in the fact that it incorporates not a single point of view, but rather provides cross-links between aspects of problems that belong to different fields of science.

- The main part of the thesis is at a system theoretic level, and features mathematical proofs of generalizations to the time-varying context of many important instruments in system theory and control: state realizations, Hankel-norm approximations, Nehari extensions, inner-outer factorizations and spectral factorizations.
- The second part is on computational techniques for large structured matrix problems in linear algebra. The assumed structure is such that state space techniques from the above-mentioned system theory can be used.

These two parts are tightly connected, and the main difference is in the customary terminology. A causal linear system in time-varying system theory is related to a block-upper matrix in linear algebra, one speaks of spectral factorization versus Cholesky factorization,

orthogonal embedding or Nehari problems versus unitary extension problems, inner-outer factorization versus *QR* factorization. The system theoretic level is more general in the sense that it can also take infinite size matrices into account. Such matrices can be specified by a finite amount of data in a number of special cases, such as extensions of finite matrices with borders that are constant, or periodically varying, or (under conditions) even indeterminate. Ordinary linear algebra methods fail to handle such generalizations.

The discussion on these subjects is mixed throughout the thesis. I have tried to keep the treatment of the computational algebra part in simple terms, in order to make potential applications visible. Hence, most chapters start with an introduction in terms of finite matrices, while mathematical details are subsequently treated in a systems theory context.

The scope of this thesis can be viewed as being defined by a disk that touches on the interests of three professors whose interaction with me most influenced my research during the past years: Ed Deprettere (who likes lots of illustrations and 'engineering' explanations), Patrick Dewilde (illustrations are allowed, but never explain a theorem by its proof), and Harry Dym (no illustrations, no physical interpretations of formulas: that way you only prove what is true instead of what you expect to be true). Not surprisingly, my interests turn out to lie somewhere in the center of the disk. I have tried to make the compromise clean. As a result: Ed, don't read the theorems; Patrick, only read the theorems; Harry, I hope that my proofs provide sufficient inspiration for a Real Mathematical Proof.

Acknowledgements

During the past four years, the research reported in this thesis has been sponsored by a number of organizations. It was supported in part by the commission of the EC under the ESPRIT Basic Research Action program 3280 and 6632, called NANA ('novel parallel algorithms for new real-time VLSI architectures'). In addition, it received a grant from Delft University of Technology in the context of 'beleidsruimte'. Travel support was granted twice by the Dutch NWO foundation, and by Shell Nederland B.V.

Part of my research was performed during a visit of three months (in 1990) to the Department of Theoretical Mathematics, Weizmann Institute of Science in Rehovot, Israel. I would like to thank Prof. Harry Dym for his efforts and hospitality, and The Karyn Kupcinet International Summer School program for their support.

No man is an island. It would be impossible for me (and I will not even try) to mention everyone by name who contributed to my research or otherwise helped me during the past years. Hence, to stay general, I would like to thank all inhabitants of the Network Theory section and researchers around the world for being fine colleagues, and, in other ways, my parents and family, roommates and friends, for 'being there' and keeping me from work.

CONTENTS

Chapter 1

INTRODUCTION

In this work, two settings play a major role. The first is the field of linear algebra, and in this setting, we are concerned with the derivation of efficient algorithms to do certain matrix calculations. The second setting is concerned with linear time-varying system theory, which will be treated at a fairly abstract level. The purpose of this section is to provide some clear links between the two settings, by introducing how linear time-varying system theory can be used to solve linear algebra problems.

1.1 COMPUTATIONAL ALGEBRA AND TIME-VARYING MODELING

Concepts

In the intersection of linear algebra and system theory is the field of *computational linear algebra*. In this field, the purpose is to find efficient algorithms for linear algebra problems, such as matrix multiplication, inversion, and approximation. A useful model for matrix computations is provided by the state equations that are used in dynamical system theory. Such a state model is often quite natural: in any algorithm for matrix multiplication or inversion, the global operation is decomposed into a sequence of local operations that each act on a limited number of matrix entries (ultimately two), assisted by intermediate quantities that connect the local operations. These quantities can be called the states of the algorithm, and translate to the state of the dynamical system that is the computational model of the matrix operation. Although many matrix operations can be represented in this way by some linear dynamical system, our interest is in matrices that possess some kind of structure which allows for efficient ("fast") algorithms: algorithms that exploit this structure. Structure in a matrix has its origin in the linear algebra problem, and is, to our purposes, typically caused by the modeling of some (physical) dynamical system. Many problems in signal processing, inverse scattering and least-squares estimation produce structured matrices that can indeed be modeled by a low complexity network.

Besides sparse matrices (matrices with many zero entries), two classical examples of structured matrices are the Toeplitz and Hankel matrices (matrices that are constant along diagonals or anti-diagonals), which translate to linear time-invariant (LTI) systems. The associated computational algorithms are well known. For example, for Toeplitz systems we have

- Schur recursions for LU and Cholesky factorization [1, 2],
- Levinson recursions for the factorization of the inverse [3],
- Gohberg/Semencul recursions for computing the inverse [4],
- Recursions for QR factorization [5].

The above algorithms have computational complexity of order $O(n^2)$ for matrices of size $(n \times n)$, as compared to $O(n^3)$ for algorithms that do not take the Toeplitz structure into account. Generalizations of the Toeplitz structure are obtained by considering matrices which have a *displacement structure* [6, 7, 8, 9]: matrices G for which there are (simple) matrices F_1, F_2 such that

$$G - F_1^* G F_2 \tag{1.1}$$

is of low rank, α say. This type of matrices occurs, *e.g.*, in such stochastic adaptive prediction problems as the covariance matrix of the received stochastic signal; the matrix is called α-stationary. Toeplitz matrices are a special case for which $F_1 = F_2$ are shift matrices Z and $\alpha = 2$. Related examples are block-Toeplitz and Toeplitz-block matrices, and, *e.g.*, the inverse of a Toeplitz matrix, which is itself not Toeplitz yet has a displacement rank of $\alpha = 2$. An overview of inversion and factorization algorithms for such matrices can be found in [10].

In this thesis, we pursue a complementary notion of structure which we will call a state structure. The state structure applies to upper triangular matrices and is seemingly unrelated to the Toeplitz or displacement structure mentioned above. A primary purpose of the computational schemes considered in this thesis is to perform a desired linear transformation T on some vector ('input sequence') u,

$$u = [u_0 \quad u_1 \quad u_2 \quad \cdots \quad u_n]$$

which yields an output vector or sequence $y = uT$. The key idea is that we can associate with this matrix-vector multiplication a computational network that takes u and computes y, and that matrices with a sparse state structure have a computational network of low complexity so that using the network to compute y is more efficient than computing uT directly. To introduce this notion, consider an upper triangular matrix T along with its

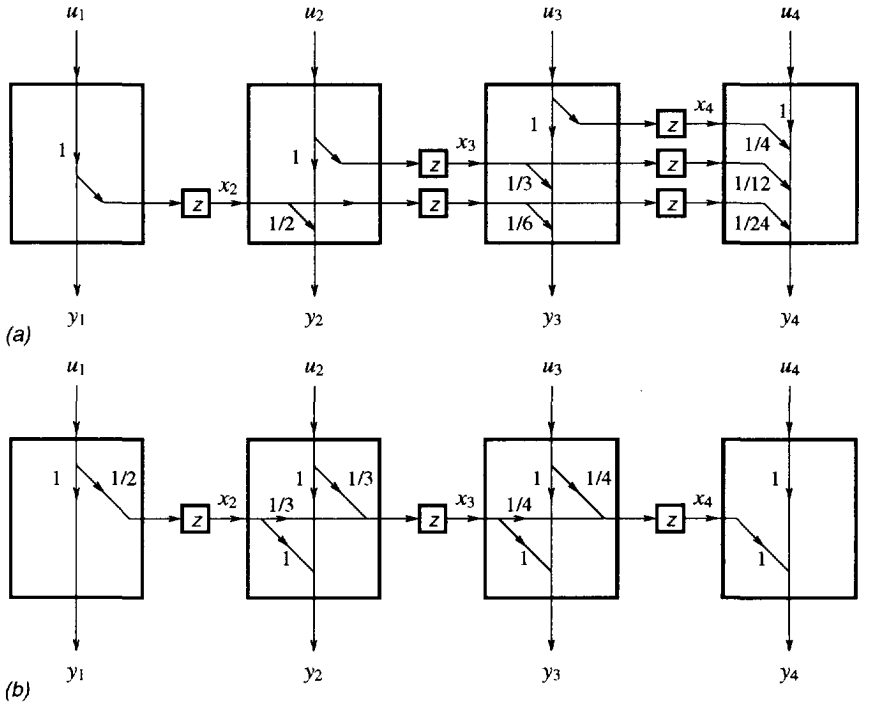


Figure 1.1. Computational networks corresponding to T. (a) Direct (trivial) realization, (b) minimal realization.

inverse,

$$T = \begin{bmatrix} 1 & 1/2 & 1/6 & 1/24 \\ & 1 & 1/3 & 1/12 \\ & & 1 & 1/4 \\ & & & 1 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 1 & -1/2 & & \\ & 1 & -1/3 & \\ & & 1 & -1/4 \\ & & & 1 \end{bmatrix}.$$

The inverse of T is sparse, which is an indication of a sparse state structure. A computational network that models multiplication by T is depicted in figure 1.1(a), and it is readily verified that this network does indeed compute $[y_1\ y_2\ y_3\ y_4] = [u_1\ u_2\ u_3\ u_4]T$ by trying vectors of the form [1 0 0 0] up to [0 0 0 1]. The computations in the network are split into sections, which we will call *stages*, where the k-th stage consumes u_k and produces y_k. At each point k the processor in the stage at that point takes its input data u_k from the input sequence u and computes new output data y_k which is part of the output sequence y generated by the system. The dependence of y_k on u_i ($i < k$) introduces

intermediate quantities x_k called *states*. To execute the computation, the processor will use some remainder of its past history, *i.e.*, the state x_k, which has been computed by the previous stages and which was temporarily stored in registers indicated by the symbol z. The complexity of the computational network is equal to the number of states at each point. A non-trivial computational network to compute $y = uT$ which requires less states is shown in figure 1.1(*b*). The total number of multiplications required in this network that are different from 1 is 5, as compared to 6 in a direct computation using T. Although we have gained only one multiplication here, for a less moderate example, say an $(n \times n)$ upper triangular matrix with $n = 10000$ and $d \ll n$ states at each point, the number of multiplications in the network can be as low as $O(8dn)$, instead of $O(1/2\, n^2)$ for a direct computation using T.

The computations in the network can be summarized by the following recursion, for $k = 1$ to n:

$$y = uT \quad \Leftrightarrow \quad \begin{aligned} x_{k+1} &= x_k A_k + u_k B_k \\ y_k &= x_k C_k + u_k D_k \end{aligned} \tag{1.2}$$

or

$$\begin{bmatrix} x_{k+1} & y_k \end{bmatrix} = \begin{bmatrix} x_k & u_k \end{bmatrix} \mathbf{T}_k, \qquad \mathbf{T}_k = \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix}$$

in which x_k is the state vector at time k (taken to have d_k entries), A_k is a $d_k \times d_{k+1}$ (possibly non-square) matrix, B_k is a $1 \times d_{k+1}$ vector, C_k is a $d_k \times 1$ vector, and D_k is a scalar. More general computational networks have their number of inputs and outputs at each stage not necessarily equal to one, and possibly also varying from stage to stage. In the example, we have a sequence of realization matrices

$$\mathbf{T}_1 = \begin{bmatrix} \cdot & \cdot \\ 1/2 & 1 \end{bmatrix} \qquad \mathbf{T}_2 = \begin{bmatrix} 1/3 & 1 \\ 1/3 & 1 \end{bmatrix} \qquad \mathbf{T}_3 = \begin{bmatrix} 1/4 & 1 \\ 1/4 & 1 \end{bmatrix} \qquad \mathbf{T}_4 = \begin{bmatrix} \cdot & 1 \\ \cdot & 1 \end{bmatrix},$$

where the '$\cdot$' indicates entries that actually have dimension 0 because the corresponding states do not exist. The recursion in equation (1.2) shows that it is a recursion for increasing values of k: the order of computations in the network is strictly from left to right, and we cannot compute y_k unless we know x_k, *i.e.*, unless we have processed $u_1, \cdots, u_{k-1}$. Note that y_k does not depend on $u_{k+1}, \cdots, u_n$. This is a direct consequence of the fact that T has been chosen upper triangular, so that such an ordering of computations is indeed possible.

Time-varying systems

A link with system theory is obtained when T is regarded as the input-output map, alias the *transfer operator*, of a *non-stationary* causal linear system with input u and output $y = uT$. The i-th row of T then corresponds to the impulse response of the system when excited by an impulse at time instant i, that is, the output y caused by an input u with $u_k = \delta_k^i$. The case where T has a Toeplitz structure then corresponds with a time-invariant

system for which the response to an impulse at time $i+1$ is just the same as the response to an impulse at time i, shifted over one position. The computational network is called a state realization of T, and the number of states at each point of the computational network is called the system order of the realization at that point in time. For time-invariant systems, the state realization can be chosen constant in time. Since for time-varying systems the number of state variables need not be constant in time, but can increase and shrink, it is seen that in this respect the time-varying realization theory is much richer, and it will be seen later that a time-varying number of states will enable the accuracy of some approximating computational network of T to be varied in time at will. If the network is regarded as the model of a physical time-varying system rather than a computational network, then the interpretation of a time-varying number of states is that the network contains switches that can switch on or off a certain part of the system and thus can make some states inaccessible for inputs or outputs at certain points in time.

Sparse computational models

If the number of state variables is relatively small, then the computation of the output sequence is efficient in comparison with a straight computation of $y = uT$. One example of an operator with a small number of states is the case where T is an upper triangular band matrix: $T_{ij} = 0$ for $j - i > p$. In this case, the state dimension is equal to or smaller than $p-1$, since only $p-1$ of the previous input values need be remembered at any point in the multiplication. However, the state model can be much more general, *e.g.*, if a banded matrix has an inverse, then this inverse is known to have a sparse state realization (of the same complexity) too, as we had in the example above. Moreover, this inversion can be easily carried out by local computations on the realization of T:[1] if $T^{-1} = S$, then $u = yS$ can be computed via

$$\left\{ \begin{array}{lcl} x_{k+1} & = & x_k A_k + u_k B_k \\ y_k & = & x_k C_k + u_k D_k \end{array} \right. \quad \Leftrightarrow \quad \left\{ \begin{array}{lcll} x_{k+1} & = & x_k(A_k - C_k D_k^{-1} B_k) & + \; y_k D_k^{-1} B_k \\ u_k & = & -x_k C_k D_k^{-1} & + \; y_k D_k^{-1} \end{array} \right.$$

hence S has a computational model given by

$$\mathbf{S}_k = \begin{bmatrix} A_k - C_k D_k^{-1} B_k & -C_k D_k^{-1} \\ D_k^{-1} B_k & D_k^{-1} \end{bmatrix} \tag{1.3}$$

Observe that the model for $S = T^{-1}$ is obtained in a *local* way from the model of T: $\mathbf{S}_k$ depends only on $\mathbf{T}_k$. Sums and products of matrices with sparse state structures have again sparse state structures with number of states at each point not larger than the sum of the number of states of its component systems, and computational networks of these compositions (but not necessarily minimal ones) can be easily derived from those of its

[1] This applies to finite matrices only, for which the inverse of the matrix is automatically upper triangular again. For infinite matrices (operators) and block matrices with non-uniform dimensions, the requirement is that T must be *outer*. See chapter 4.

components. Finally, a matrix T' that is not upper triangular can be split (or factored) into an upper triangular and a strictly lower triangular part, each of which can be separately modeled by a computational network. The computational model of the lower triangular part has a recursion that runs backward:

$$\begin{aligned} x'_{k-1} &= x'_k A'_k + u_k B'_k \\ y_k &= x'_k C'_k + u_k D'_k . \end{aligned}$$

The model of the lower triangular part can be used to determine a model of a unitary upper matrix U which is such that U^*T is upper and has a sparse state structure. Thus, computational methods derived for upper matrices, such as the above inversion formula, can be generalized to matrices of mixed type.

Besides matrix inversion, other matrix operations that can be computed efficiently using sparse computational models are for example the QR factorization (chapter 4) and the Cholesky factorization (chapter 8).

At this point, one might wonder for which class of matrices T there exists a sparse computational network (or state realization) that realizes the same multiplication operator. A general criterion will be derived in chapter 3, along with a recursive algorithm to determine such a network for a given matrix T. The criterion itself is not very complicated, but in order to specify it, we have to introduce an additional concept. For an upper triangular $(n \times n)$ matrix T, define matrices H_i $(1 \leq i \leq n)$, which are mirrored submatrices of T, as

$$H_i = \begin{bmatrix} T_{i-1,i} & T_{i-1,i+1} & \cdots & T_{i-1,n} \\ T_{i-2,i} & T_{i-2,i+1} & & \vdots \\ \vdots & & \ddots & T_{2,n} \\ T_{1,i} & \cdots & T_{1,n-1} & T_{1,n} \end{bmatrix}$$

(see figure 1.2). The H_i are called (time-varying) Hankel matrices, as they have a Hankel structure (constant along anti-diagonals) if T has a Toeplitz structure.[2] In terms of the Hankel matrices, the criterion by which matrices with a sparse state structure can be detected is given by the following theorem, proven in chapter 3.

THEOREM 1.1. *The number of states that are required at stage k in a minimal computational network of an upper triangular matrix T is equal to the rank of its k-th Hankel matrix* H_k.

[2]Warning: in the current context (arbitrary upper triangular matrices) the H_i do not have a Hankel structure and the predicate 'Hankel matrix' could lead to misinterpretations. The motivation for the use of this terminology can be found in system theory, where the H_i are related to an abstract operator H_T which is commonly called the Hankel operator. For time-invariant systems, H_T reduces to an operator with a matrix representation that has indeed a Hankel structure.

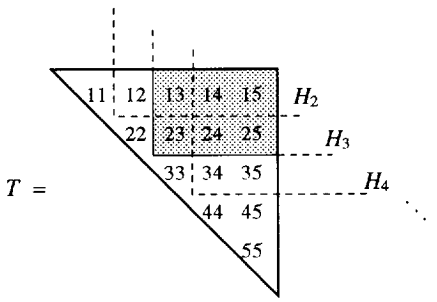


Figure 1.2. Hankel matrices are (mirrored) submatrices of T.

Let's verify this statement for our example. The Hankel matrices are

$$H_1 = [\,\cdot\ \cdot\ \cdot\,], \qquad H_2 = [1/2\ \ 1/6\ \ 1/24\,],$$

$$H_3 = \begin{bmatrix} 1/3 & 1/12 \\ 1/6 & 1/24 \end{bmatrix}, \qquad H_4 = \begin{bmatrix} 1/4 \\ 1/12 \\ 1/24 \end{bmatrix}.$$

Since $\text{rank}(H_1) = 0$, no states x_1 are necessary. One state is required for x_2 and one for x_4, because $\text{rank}(H_2) = \text{rank}(H_4) = 1$. Finally, also only one state is required for x_3, because $\text{rank}(H_3) = 1$. In fact, this is (for this example) the only non-trivial rank condition: if one of the entries in H_3 would have been different, then two states would have been necessary. In general, $\text{rank}(H_i) \leq \min(i-1, n-i+1)$, and for a general upper triangular matrix T without state structure, a computational model indeed requires at most $\min(i-1, n-i+1)$ states for x_i. The statement is also readily verified for matrices with a band structure: if the band width of the matrix is equal to d, then the rank of each Hankel matrix is at most equal to d. As we have seen previously, the inverse of such a band matrix (if it exists) has again a low state structure, *i.e.*, the rank of the Hankel matrices of the inverse is again at most equal to d. For $d = 1$, such matrices have the form (after scaling of each row so that the main diagonal entries are equal to 1)

$$T = \begin{bmatrix} 1 & -a_1 & & \\ & 1 & -a_2 & \\ & & 1 & -a_3 \\ & & & 1 \end{bmatrix} \qquad T^{-1} = \begin{bmatrix} 1 & a_1 & a_1a_2 & a_1a_2a_3 \\ & 1 & a_2 & a_2a_3 \\ & & 1 & a_3 \\ & & & 1 \end{bmatrix}$$

and it is seen that H_3 of T^{-1} is indeed of rank 1.

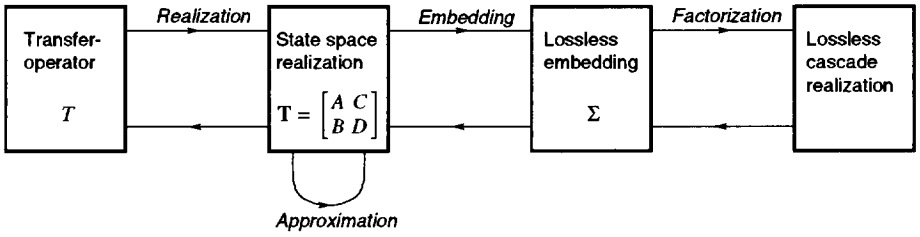


Figure 1.3. Objectives of computational modeling for matrix multiplication.

1.2 OBJECTIVES OF COMPUTATIONAL LINEAR ALGEBRA

Operations

With the preceding section as background material, we are now in a position to identify some of the objectives of computational modeling, as covered by the thesis. We assume most of the time that the given operators or matrices are upper triangular. Applications which involve other types of matrices are viable if they provide some transformation by which upper triangular matrices are obtained. For example, if a matrix can be decomposed into a (block) upper and a (block) lower part, and if each of these parts has a sparse computational network, then these parts can be treated separately (in matrix multiplications), or the matrix can be factored into a product of a lower and an upper triangular matrix, and the factors can be treated independently (in inversion problems). In addition, we assume that the concept of a sparse state structure is *meaningful* for the problem, in other words that in the application, a typical matrix has a sequence of Hankel matrices that all have low rank (relative to the size of the matrix), or that an approximation of that matrix by one whose Hankel matrices have low rank would indeed yield a useful approximation of the underlying (physical) problem that is described by the original matrix.

Much of the thesis is characterized by the objective to determine for a given matrix a computational model $\{\mathbf{T}_k\}_1^n$ of minimal complexity, by which multiplications of vectors by T are effectively carried out, but in a *computationally efficient* and *numerically stable* manner. This objective is divided into four subproblems, connected schematically as in figure 1.3: (1) realization of a given matrix T by a computational model, (2) embedding of this realization in a larger model that consists entirely of unitary (lossless) stages, (3) factorization of the stages of the embedding into a cascade of elementary (degree-1) lossless sections. It could very well be that the matrix that was originally given has a computational model of a very high order. Then intermediate in the above sequence of steps is (4) the approximation of a given realization of T by one of lower complexity. These steps are reasoned below.

Realization. The first step is, given an upper triangular matrix T, to determine any minimal computational network $\mathbf{T} = \{A_k, B_k, C_k, D_k\}_1^n$ that models T. This problem is known as the *realization problem.* If the Hankel matrices of T have low rank, then $\mathbf{T}$ is a computationally efficient realization of the operation 'multiplication by T'.

Orthogonal embedding. From $\mathbf{T}$, all other minimal realizations of T can be derived. Not all of these have the same numerical stability. This is because the computational network has introduced a recursive aspect to the multiplication: states are used to extract information from the input vector u, and a single state x_k gives a contribution both to the current output y_k and to the sequence x_{k+1}, x_{k+2} etc. In particular, a perturbation in x_k (or u_k) also carries over to this sequence. Suppose that T is bounded in norm by some number, say $\|T\| \leq 1$,[3] so that we can measure perturbation errors relative to 1. Then a realization of T is said to be error insensitive if $\|\mathbf{T}_k\| \leq 1$, too. In this case, an error in $[x_k \quad u_k]$ is not magnified by $\mathbf{T}_k$, and the resulting error in $[x_{k+1} \quad y_k]$ is smaller than the original perturbation. Hence the question is: is it possible to obtain a realization for which $\|\mathbf{T}_k\| \leq 1$ if T is such that $\|T\| \leq 1$? The answer is yes, and an algorithm to obtain such a realization is given by the solution of the *orthogonal embedding problem.* This problem is the following: for a given matrix T with $\|T\| \leq 1$, determine a computational model $\{\mathbf{\Sigma}_k\}_1^n$ such that (1) each $\mathbf{\Sigma}_k$ is a unitary matrix, and (2) T is the transfer operator of a subsystem of the transfer operator Σ that corresponds to $\{\mathbf{\Sigma}_k\}$. The latter requirement means that T is the transfer operator from a subset of the inputs of Σ to a subset of its outputs: Σ can be partitioned conformably as

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \qquad T = \Sigma_{11}.$$

The fact that T corresponds to a subsystem of Σ implies that a certain submatrix of $\mathbf{\Sigma}_k$ is a realization $\mathbf{T}_k$ of T, and hence from the unitarity of $\mathbf{\Sigma}_k$ we have that $\|\mathbf{T}_k\| \leq 1$. From the construction of the solution to the embedding problem, it will follow that we can ensure that this realization is minimal, too.

Cascade factorization. Assuming that we have obtained such a realization $\mathbf{\Sigma}_k$, the next question is whether it is possible to break down the operation 'multiplication by $\mathbf{\Sigma}_k$' on vectors $[x_k \quad u_k]$ into a number of elementary operations, each in turn acting on two entries of this vector. Because $\mathbf{\Sigma}_k$ is unitary, we can use elementary unitary operations (acting on scalars) of the form

$$[a_1 \quad b_1] \begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} = [a_2 \quad b_2], \qquad cc^* + ss^* = 1,$$

[3] $\|T\|$ is the operator norm (matrix 2-norm) of T: $\|T\| = \sup_{\|u\|_2 \leq 1} \|uT\|_2$.

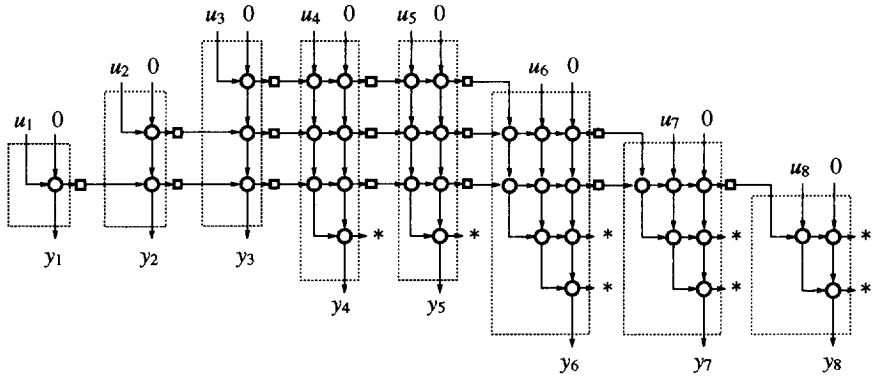


Figure 1.4. Cascade realization of a contractive 8×8 matrix T, with a maximum of 3 states at each point.

i.e., elementary rotations. The use of such elementary operations will ensure that $\mathbf{\Sigma}_k$ is internally numerically stable, too. In order to make the number of elementary rotations minimal, the realization $\mathbf{\Sigma}$ is transformed to an equivalent realization $\mathbf{\Sigma}'$, which realizes the same transfer operator Σ, is still unitary and which still contains a realization $\mathbf{T}'$ for T. A factorization of $\mathbf{\Sigma}'_k$ into elementary rotations is known as a *cascade realization* of Σ. A possible minimal computational model for T that corresponds to such a cascade realization is drawn in figure 1.4. In this figure, each circle indicates an elementary rotation. The precise form of the realization depends on whether the state dimension is constant, shrinks or grows. The realization can be divided into elementary *sections*, where each section describes how a single state entry of x_k is mapped to an entry of the 'next state' vector x_{k+1}.

Besides the numerical property mentioned above, the cascade realization in figure 1.4 has a number of other interesting properties. Firstly, it provides a realization of T with a minimal number of parameters, and in each stage, the number of operations to compute the next state and output is linear in the number of states at that point, rather than quadratic as would be the case for a general (non-factored) realization. Another property is that the network is *pipelinable*, which is interesting if the operation 'multiplication by T' is to be carried out on a collection of vectors u on a parallel computer or on a hardware implementation of the computational network. The property is a consequence of the fact that the signal flow in the network is strictly uni-directional: from top left to bottom right, so that computations on a new vector u (a new u_k and a new x_k) can commence in the top-left part of the network, while computations on the previous u are still being carried out in the bottom-right part.

Approximation. In the previous items, we have assumed that the matrix T has indeed a computational model of an order that is low enough to favor a computational network over an ordinary matrix multiplication. However, if the rank of the Hankel matrices of T (the system order) is not low, then it often makes sense to approximate T by a new upper triangular matrix T_a that has lower complexity. For example, it could happen that the given matrix T is not of low complexity because numerical inaccuracies of the entries of T have increased the rank of the Hankel matrices of T, since the rank of a matrix is a very sensitive (ill-conditioned) parameter. But even if the given matrix T is known to be exact, an approximation by a reduced-order model could be appropriate, for example for design purposes in engineering, to capture the essential behavior of the model. With such a reduced-complexity model, the designer can more easily detect that certain features are not desired and can possibly predict the effects of certain changes in the design; an overly detailed model would rather mask these features.

While it is fairly well known in linear algebra how to obtain a (low-rank) approximant to a matrix in a certain norm (*e.g.*, by use of the singular value decomposition (SVD)), such approximations are not necessarily appropriate for our purposes, because the approximant should be upper triangular again, and have a lower system order. Because the minimal system order at each point is given by the rank of the Hankel matrix at that point, a possible approximation scheme is to approximate each Hankel operator by one that is of lower rank (this could be done using the SVD). The approximation error could then very well be defined in terms of the individual Hankel matrix approximations as the supremum over these approximations. Because the Hankel matrices have many entries in common, it is not immediately clear whether such an approximation scheme is feasible: replacing one Hankel matrix by one of lower rank in a certain norm might make it impossible for the next Hankel matrix to find an optimal (in that norm) approximant such that the part that it has in common with the previous Hankel matrix will be approximated by the same matrix. In other words: each individual local optimization might prevent a global optimum. The severity of this dilemma is mitigated by a proper choice of the error criterion. In fact, it is remarkable that this dilemma has a neat solution, and that this solution can be obtained in a closed form. The error for which a solution is obtained is measured in *Hankel norm:* it is the supremum over the spectral norm (the matrix 2-norm) of each individual Hankel matrix,

$$\|T\|_H = \sup_i \|H_i\|,$$

and a generalization of the Hankel norm for time-invariant systems. In terms of the Hankel norm, the following theorem holds true and generalizes the model reduction techniques based on the Adamjan-Arov-Krein paper [11] to time-varying systems:

THEOREM 1.2. ([12]) *Let T be a strictly upper triangular matrix and let $\Gamma = \mathrm{diag}(\gamma_i)$ be a diagonal Hermitian matrix which parametrizes the acceptable approximation tolerance ($\gamma_i > 0$). Let H_k be the Hankel matrix of $\Gamma^{-1}T$ at stage k, and suppose that, for each k, none*

of the singular values of H_k are equal to 1. Then there exists a strictly upper triangular matrix T_a whose system order at stage k is equal to the number of singular values of H_k that are larger than 1, such that

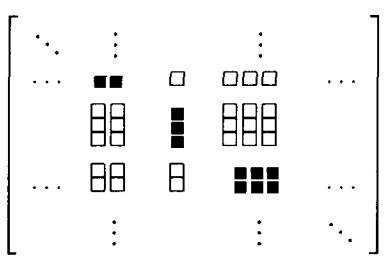
$$\| \Gamma^{-1}(T - T_a) \|_H \leq 1 .$$

In fact, there is an algorithm that determines a state model for T_a directly from a model of T. Γ can be used to influence the local approximation error. For a uniform approximation, $\Gamma = \gamma I$, and hence $\|T - T_a\|_H \leq \gamma$: the approximant is γ-close to T in Hankel norm, which implies in particular that the approximation error in each row or column of T is less than γ. If one of the γ_i is made larger than γ, then the error at the i-th row of T can become larger also, which might result in an approximant T_a that has fewer states. Hence Γ can be chosen to yield an approximant that is accurate at certain points but less tight at others, and whose complexity is minimal.

The realization problem is treated in chapter 3, the embedding problem is the subject of chapter 7, while the cascade factorization algorithm appears in chapter 9. The Hankel norm approximation problem is solved in chapter 6. As applications, the *QR* factorization is treated in chapter 4, and the Cholesky factorization in chapter 8.

Operands

In the preceding section, the types of operations (realization, embedding, factorization, approximation) that are considered in this thesis were introduced. I now outline below the types of operands to which these operations are applied. In principle, we work with bounded linear operators on Hilbert spaces of (vector) sequences. From an engineering point of view, such operators can be regarded as infinite-size matrices. The entries in turn can be block matrices. In general, they could even be operators, but we do not here consider that case. There is no need for the block entries all to have the same size: the only requirement is that all entries on a row of the operator have an equal number of rows, and all entries on a column of the operator have an equal number of columns, to ensure that all vector-matrix products are well defined. Consequently, the upper triangular matrices can have an "appearance" that is not upper triangular. For example, consider

$T =$ 

where in this case each box represents a complex number. The main diagonal is distinguished here by filled boxes. We call such a general matrix a *tableau*.

We say that such an operator describes the input-output behavior of a linear time-varying system. The system is time invariant if the matrix representation of the operator is (block) Toeplitz: constant along diagonals. In general, we allow the upper triangular part to have an arbitrary structure, or even no structure at all. Special cases are periodically varying systems, which give block-Toeplitz operators, and systems that are time-invariant outside a finite interval in time, which give oprators that are constant at the borders. A sequence on which the operator can be applied (the input of the system) is represented by a row vector whose entries are again finite-size vectors conforming to the block entries of the operator. This corresponds to a system with block inputs and block outputs. If the size of the block entries is not constant, then the system has a time-varying number of inputs and outputs, which corresponds physically to a system with switches that are used to switch on or off certain inputs and outputs at certain times. It is possible to model finite matrices this way, as was shown in the introduction. For finite matrices, there are no inputs and outputs before and after a certain interval in time.

A causal system corresponds to an operator whose matrix representation is upper triangular. We are interested in such systems because causality implies a computational direction: usually we can start calculations at the top-left end of the tableau and work towards the bottom-right end. Causality also introduces the notion of state. The number of states is allowed to be time varying: think, for example, of switches that switch on or off a certain part of the system. The concept of a time-varying number of states allows the incorporation of a finer level of detail at certain intervals in time.

1.3 ABOUT THIS THESIS

The thesis contains an account of time-varying systems theory for Hilbert space operators. A special feature is that a new notation is used that allows for expressions which are mostly index free. The thesis is to a large extent self-contained, in the sense that all the system theoretic results required are (re-)derived instead of 'borrowed' from literature. Algorithms are usually described up to the state-space level, and only in some places at the elementary (2×2) level of local computations. The algorithms are intended to show how global matrix computations can be reduced to a state-space level, but they are not really optimized for efficiency, as this would typically require knowledge of a particular application and would not help to make the overall picture clear.

We usually consider Hilbert space operators ('infinite matrices') instead of finite matrices although, in the end, our main interest is in such matrices. Our original reason for considering infinite matrices was that, when applying the basic operation of (row or column) shifts to such objects, the result is (1) again an object in the same space, and (2) the shift is invertible. At first, the impression was that this would be impossible for finite matrices (see [13]). However, as was discovered later, with operators whose rows and columns are permitted to have varying dimensions, it is possible to construct a theory

on finite matrices in which the shift operator is still invertible. Such a theory is easier to describe because then, for example, the concern of the boundedness of operators is not a main issue. At present, the reason to consider operators rather than finite matrices is that there is some interest in special cases such as systems which do not change outside a finite time-invariant interval, and systems which periodically change. These cases are usually highlighted separately at the end of the chapters.

Most of the discussion in the thesis is at a system theoretic level, and not in terms of computational modeling (as done in this chapter). The results obtained can be listed under the following categories.

- *Realization theory* for time-varying systems (chapter 3): a Ho-Kalman-like algorithm is derived which can be used to obtain a minimal realization of a given transfer operator. Issues such as stability, controllability, realization equivalence etc. are defined.
- The realization theory can be specialized to apply to *inner systems* (chapter 4), and this provides the necessary background to derive coprime factorizations and inner-outer factorizations in state-space terms. The latter can be used, *e.g.*, to determine the inverse of a causal system, which is not trivial because the inverse is not necessarily causal.
- The realization theory can also be specialized to apply to *J-unitary systems* (chapter 5), providing results that are useful in the solution of certain constrained interpolation problems.
- One such interpolation problem occurs in the solution of the *Hankel-norm approximation problem* (chapter 6). We show what the number of states of an approximant will be, given a certain error tolerance, how a state model of an approximant can be computed, and how all other Hankel-norm approximants can be obtained. As a special case, a state realization for the solution of the *Nehari problem* is derived.
- The *orthogonal embedding problem* is solved in chapter 7. This problem is in fact a *spectral factorization problem,* and such problems are investigated in their own right in chapter 8. In both cases, solutions are described by a Riccati recursive equation with time-varying coefficients, and some properties of this recursion are investigated. In particular, it is shown that if there is a positive semi-definite solution, then this solution is unique and yields outer factors. It is also proven that (under certain conditions) the recursion converges to the exact solution, even if it is started from an approximate initial value.
- Finally, *cascade factorizations* for inner operators are derived (chapter 9).

The definitions and results presented in chapters 2 and 3 are used throughout the thesis. The discussion in chapter 6 is strongly dependent on that of chapter 5, but the other chapters can be read quite independently from each other.

Bibliography

[1] I. Schur, "Uber Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I," *J. Reine Angew. Math.*, vol. 147, pp. 205–232, 1917. Eng. Transl. *Operator Theory: Adv. Appl.*, vol. 18, pp. 31-59, Birkhäuser Verlag, 1986.

[2] T. Kailath, "A theorem of I. Schur and its impact on modern signal processing," in *Operator Theory: Advances and Applications*, vol. 18, pp. 9–30, Basel: Birkhäuser Verlag, 1986.

[3] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.

[4] I. Gohberg and A. Semencul, "On the inversion of finite Toeplitz matrices and their continuous analogs," *Mat. Issled.*, vol. 2, pp. 201–233, 1972.

[5] J. Chun, T. Kailath, and H. Lev-Ari, "Fast parallel algorithms for *QR* and triangular factorizations," *SIAM J. Sci. Stat. Comp.*, vol. 8, no. 6, pp. 899–913, 1987.

[6] T. Kailath, S.Y. Kung, and M. Morf, "Displacement ranks of matrices and linear equations," *J. Math. Anal. Appl.*, vol. 68, no. 2, pp. 395–407, 1979.

[7] H. Lev-Ari and T. Kailath, "Lattice filter parametrization and modeling of non-stationary processes," *IEEE Trans. Informat. Th.*, vol. 30, pp. 2–16, Jan. 1984.

[8] H. Lev-Ari and T. Kailath, "Triangular factorizations of structured Hermitian matrices," in *Operator Theory: Advances and Applications*, vol. 18, pp. 301–324, Birkhäuser Verlag, 1986.

[9] H. Lev-Ari and T. Kailath, "Lossless arrays and fast algorithms for structured matrices," in *Algorithms and Parallel VLSI Architectures* (Ed. F. Deprettere and A.J. van der Veen, eds.), vol. A, pp. 97–112, Elsevier, 1991.

[10] J. Chun, *Fast Array Algorithms for Structured Matrices*. PhD thesis, Stanford Univ., Stanford, CA, 1989.

[11] V.M. Adamjan, D.Z. Arov, and M.G. Krein, "Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem," *Math. USSR Sbornik*, vol. 15, no. 1, pp. 31–73, 1971. (transl. of *Iz. Akad. Nauk Armjan. SSR Ser. Mat. 6 (1971)).*

[12] P.M. Dewilde and A.J. van der Veen, "On the Hankel-norm approximation of upper-triangular operators and matrices," *to appear in Integral Equations and Operator Theory*, 1993.

[13] H. Nelis, *Sparse Approximations of Inverse Matrices*. PhD thesis, Delft Univ. Techn., The Netherlands, 1989.

Chapter 2

SIGNALS AND SYSTEM DEFINITIONS

In this chapter, we introduce a precise notation by which signals and transfer operators can be compactly described. In this notation, operators are decomposed into diagonals and shift operators. The notation was originally introduced by Alpay and Dewilde in [1] (and subsequently in Alpay, Dewilde and Dym [2]), who developed a generalization of the z-transform for upper non-commutative operators, called the W-transform, and investigated the interpolating properties of lossless time-varying (or non-stationary) systems represented by these operators. The notation has been refined a number of times, to allow for time-varying state spaces [3] and time-varying input and output spaces [4]. The basic mathematical properties were proven in [2] and additional properties later in Dewilde and Dym [5].

There are a number of other approaches to describe time-varying systems. Starting in the 1950s [6] (or even earlier), time-varying network theory and extensions of important system theoretic notions to the time-varying case have been discussed by many authors. While most of the early work is on continuous-time linear systems and differential equations with time-varying coefficients (see, *e.g.*, [7] for a 1960 survey), discrete-time systems have gradually come into favor. There are some more recent approaches which are important, running in parallel with the time-varying state-space realization theory discussed in chapter 3. These are presented in the monograph by Feintuch and Saeks [8], in which a Hilbert resolution space setting is taken, and in recent work by Kamen, Poolla and Khargonekar [9, 10, 11], where time-varying systems are put into an algebraic framework of polynomial noncommutative rings. In the latter approach, a different kind of generalized z-transform is introduced. However, many of these results, in particular on controllability, detectability, stabilizability etc., have been discussed by many authors without using these specialized mathematical means, but rather by simply time indexing the state-space matrices $\{A, B, C, D\}$ and deriving expressions (iterations) in terms of these matrices. There is usually a one-to-one correspondence between these expressions and their equivalent in our notation.

The original formulation in [2] of the notation that we use here favored a representation of linear maps as vector-matrix multiplications (as in 'uT') over the more common matrix-vector multiplications (as in 'Tu').[1] As a consequence of this choice, the order of matrices in familiar (state-space) expressions would appear to be reversed. Because of the close relation of matrices with linear operators, we also apply operators on sequences that sit at the left of the operator instead of at the right. Especially for projection operators, this could lead to confusion, and therefore we make in this case an exception and write for example $\mathbf{P}(uT)$ instead of $uT\mathbf{P}$. This compromise makes the notation used here compatible in most respects with the notation in [2, 5], while at the same time it retains the possibility of physically correct interpretations (in terms of 'signals', 'signal spaces' and 'systems'). In [2, 5], operators T that are applied to physical sequences u as in uT are analyzed via Tf, in which f does not have an interpretation as a signal. This duality is avoided here.

2.1 HILBERT SPACE DEFINITIONS AND PROPERTIES

Hilbert space plays an important role throughout, and it thus seems proper to start our treatise with a brief review of those definitions and results that are relevant to later chapters. The material in this section is basic and can be found in textbooks such as Akhiezer-Glazman [12] (which we follow here), Halmos [13], and Fuhrmann [14, chap. 2]. The main focus is on the properties of subspaces of Hilbert space. Hilbert space theory in this section is called 'abstract': it is axiomatic in character, and indeed a wide variety of linear systems satisfy the axioms. Starting in section 2.2, the abstract theory is specialized by considering only Hilbert spaces over sequences, although the objects are more general than ordinary sequences in the sense that they are taken *non-uniform*: the entries of a sequence are (finite) vectors of possibly different dimensions. This imposes a certain structure on the Hilbert space that is not found in standard textbooks.

Linear manifold

In this section, we consider *complex vector spaces* whose elements ('vectors') are not further specified (they could, for example, indeed be vectors in the usual n-dimensional Euclidean space $\mathbf{C}^n$, or more in general, be infinite-dimensional vectors). Besides a set of elements, for a complex vector space $\mathcal{H}$ two operations are defined: the addition of two elements of $\mathcal{H}$ and the multiplication of an element of $\mathcal{H}$ by a complex number, and $\mathcal{H}$ should contain a unique null element for addition. Elements $f_1, f_2, \cdots, f_n$ in $\mathcal{H}$ are called *linearly independent* if (for complex numbers α_i)

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots \alpha_n f_n = 0 \quad \Leftrightarrow \quad \alpha_1, \cdots, \alpha_n = 0.$$

[1] The original motivation was that, by doing so, the formula for the reproducing kernel of a Hilbert space subspace becomes a format that is familiar from functional analysis.

$\mathcal{H}$ is finite dimensional (say n-dimensional) if at most a finite number of n elements are linearly independent. Such spaces are studied in linear algebra and yield a specialization of the Hilbert space theory to follow below. A set $\mathcal{M}$ of elements of a complex vector space $\mathcal{H}$ is called a *linear manifold* if

$$f \in \mathcal{M},\ g \in \mathcal{M} \quad \Rightarrow \quad \alpha f + \beta g \in \mathcal{M}$$

for all scalars α, β. A set $\mathcal{M}$ is called the *direct sum* of a finite number of linear manifolds $\mathcal{M}_k \subset \mathcal{H}$,

$$\mathcal{M} = \mathcal{M}_1 \dot{+} \cdots \dot{+} \mathcal{M}_n, \tag{2.1}$$

if for every $g \in \mathcal{M}$ there is one and only one expression in the form of a sum

$$g = g_1 + g_2 + \cdots + g_n$$

where $g_k \in \mathcal{M}_k$, and if any sum of this form is in $\mathcal{M}$. $\mathcal{M}$ is a linear manifold itself. A set of n linear manifolds $\{\mathcal{M}_k\}_1^n$ is called linearly independent if

$$f_1 + f_2 + \cdots + f_n = 0 \quad (f_i \in \mathcal{M}_i) \quad \Rightarrow \quad f_1, \cdots, f_n = 0.$$

Linear independence is both a necessary and a sufficient condition to be able to construct the direct sum in (2.1).

Inner product

A complex vector space $\mathcal{H}$ is an *inner product space* if a functional $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \to \mathbf{C}$ is defined such that, for every $f, g \in \mathcal{H}$ and $\alpha_1, \alpha_2 \in \mathbf{C}$,

$$\begin{array}{lrcl} (i) & (g,f) & = & \overline{(f,g)} \\ (ii) & (\alpha_1 f_1 + \alpha_2 f_2, g) & = & \alpha_1 (f_1, g) + \alpha_2 (f_2, g) \\ (iii) & (f,f) \geq 0; & & (f,f) = 0 \quad \Leftrightarrow \quad f = 0. \end{array}$$

The overbar denotes complex conjugation. The *norm* of $f \in \mathcal{H}$, induced by the inner product, is defined by

$$\|f\|_2 = (f,f)^{1/2}.$$

Some properties that follow from the definitions (i)–(iii) are

$$\begin{array}{lcll} \|\alpha f\|_2 & = & |\alpha| \cdot \|f\|_2 & (\alpha \in \mathbf{C}) \\ |(f,g)| & \leq & \|f\|_2 \cdot \|g\|_2 & \text{(Schwarz's inequality)} \\ \|f+g\|_2 & \leq & \|f\|_2 + \|g\|_2 & \text{(triangle inequality)}. \end{array}$$

Orthogonality

Two vectors f, g are said to be *orthogonal*, $f \perp g$, if $(f, g) = 0$. Given a set $\mathcal{M}$, we write $f \perp \mathcal{M}$ if $f \perp m$ (all $m \in \mathcal{M}$). A set of vectors $\{f_i\}$ is an *orthogonal set* if $(f_i, f_j) = 0$ $(i \neq j)$. A vector f is *normalized* if $\|f\|_2 = 1$. An *orthonormal set* is an orthogonal set of normalized vectors.

Metric space

A *metric space* is a set $\mathcal{H}$ for which a *distance* $d(f, g)$ is defined, which satisfies

$$\begin{array}{lllll} (i) & d(f,g) & = & d(g,f) > 0 & \text{when } f \neq g \\ (ii) & d(f,f) & = & 0 & \\ (iii) & d(f,g) & \leq & d(f,h) + d(g,h) & \text{(triangle ineq.)} \end{array}$$

Hence, an inner product space is a metric space where $d(f, g) = \|f - g\|_2$.

A sequence of elements f_n in $\mathcal{H}$ has a point $f \in \mathcal{H}$ as its limit: $f_n \to f$, if

$$\lim_{n \to \infty} d(f_n, f) = 0. \tag{2.2}$$

We say that $\{f_n\}$ converges to f in norm, and call this *strong* or *norm convergence*. From (*iii*) it follows that (2.2) implies

$$\lim_{m,n \to \infty} d(f_n, f_m) = 0. \tag{2.3}$$

A sequence $\{f_n\}$ that satisfies (2.3) is called a *Cauchy sequence*. It is not true for every metric space $\mathcal{H}$ that a Cauchy sequence $\{f_n\}$ converges to an element of the set: (2.3) does not imply (2.2). If it does, then $\mathcal{H}$ is called *complete*.

A limit point of a set $\mathcal{M} \subset \mathcal{H}$ is any point $f \in \mathcal{H}$ such that any ε-neighborhood $\{g : d(f, g) < \varepsilon\}$ $(\varepsilon > 0)$ of f contains infinitely many points of $\mathcal{M}$. A set that contains all its limit points is said to be *closed*. The process of adding to $\mathcal{M}$ all its limit points is called *closure*, the set yielded is denoted by $\overline{\mathcal{M}}$: the closure of $\mathcal{M}$. A set is *dense* in another set if the closure of the first set yields the second set. As an example, the set of rational numbers is dense in $\mathbb{R}$.

If in a metric space there is a countable set whose closure coincides with the whole space, then the space is said to be *separable*. In this case, the countable set is *everywhere dense*.

Hilbert space

A *Hilbert space* is an inner product space that is complete, relative to the metric induced by the inner product. The prime example of a Hilbert space is the space ℓ_2 of sequences $f = [\cdots \; f_0 \; f_1 \; f_2 \; \cdots] = [f_i]_{-\infty}^{\infty}$ of complex numbers f_i such that $\|f\|_2 < \infty$. The inner product in this space is defined by[2]

$$(f, g) = \sum_{-\infty}^{\infty} f_i \overline{g_i}.$$

[2]The meaning of the infinite sum is defined via a limit process of sums over finite sets, in case these sums converge. See Halmos [13, §7].

This space is separable: a countable set is for example the set of all vectors with a finite number of non-zero rational components f_i. The space ℓ_2 is complete, and it is infinite dimensional since the unit vectors

$$\begin{array}{rcl} & \vdots & \\ e_0 & = & [\cdots 0 \quad 1 \quad 0 \quad 0 \quad \cdots] \\ e_1 & = & [\cdots 0 \quad 0 \quad 1 \quad 0 \quad \cdots] \\ e_2 & = & [\cdots 0 \quad 0 \quad 0 \quad 1 \quad \cdots] \\ & \vdots & \end{array} \tag{2.4}$$

are linearly independent.

A *closed* linear manifold in a Hilbert space $\mathcal{H}$ is called a *subspace*. A subspace is itself a Hilbert space. An example of a subspace is, given some vector $y \in \mathcal{H}$, the set $\{x \in \mathcal{H} : (x, y) = 0\}$. (The main issue in proving that this set is a subspace is the proof that it is closed; this goes via the fact that $x_n \to x \Rightarrow (x_n, y) \to (x, y)$. See [12].) More in general, given a set $\mathcal{M} \subset \mathcal{H}$, define

$$\mathcal{M}^{\perp} = \{x \in \mathcal{H} : (x, y) = 0, \forall y \in \mathcal{M}\}.$$

Again, $\mathcal{M}^{\perp}$ is a subspace. If $\mathcal{M}$ is a subspace, then $\mathcal{M}^{\perp}$ is called the *orthogonal complement* of $\mathcal{M}$. For a subspace $\mathcal{M}$ and vector $f \in \mathcal{H}$, there exists a *unique* vector $f_1 \in \mathcal{M}$ such that $\|f - f_1\|_2 < \|f - g\|_2$ for all $g \in \mathcal{M}$ $(g \neq f_1)$. This vector f_1 is called the *component* of f in $\mathcal{M}$, or the *orthogonal projection* of f onto the subspace $\mathcal{M}$. The vector $f_2 = f - f_1$ is readily shown to be orthogonal to $\mathcal{M}$, *i.e.*, $f_2 \in \mathcal{M}^{\perp}$. With respect to $\mathcal{H}$, we have obtained the decomposition

$$\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^{\perp}, \tag{2.5}$$

where '$\oplus$' denotes the direct sum ($\dotplus$) of orthogonal spaces. The orthogonal complement $\mathcal{M}^{\perp}$ is likewise written as

$$\mathcal{M}^{\perp} = \mathcal{H} \ominus \mathcal{M}.$$

Projection onto a finite-dimensional subspace

Let $\{e_i\}_1^n$ be a set of n orthonormal vectors in a Hilbert space $\mathcal{H}$, and let $\mathcal{M}$ be the finite-dimensional subspace spanned by linear combinations of the $\{e_i\}$:

$$\mathcal{M} = \{m : m = \alpha_1 e_1 + \alpha_2 e_2 + \cdots + \alpha_n e_n, \quad \text{all } \alpha_i \in \mathbb{C}\}.$$

Because the $\{e_i\}$ are linearly independent, any $m \in \mathcal{M}$ can be written as a unique linear combination of the $\{e_i\}$. It immediately follows that $(m, e_i) = \alpha_i$, so that

$$m = \sum_1^n (m, e_i) e_i$$

(where $(m, e_i)e_i$ can be regarded as the projection of m onto e_i), and because the $\{e_i\}$ are orthonormal,

$$\| m \|_2^2 = \sum_1^n |(m, e_i)|^2 .$$

Let $f \in \mathcal{H}$, then we have seen that there is a unique decomposition $f = f_1 + f_2$, with $f_1 \in \mathcal{M}$, $f_2 \in \mathcal{M}^\perp$. Since $(f_2, e_i) = 0$, we have $(f, e_i) = (f_1, e_i)$ and hence

$$f = \sum_1^n (f, e_i)e_i + f_2 \qquad (f_2 \in \mathcal{M}^\perp).$$

Hence the projection of f onto $\mathcal{M}$ is obtained explicitly as $\sum_1^n (f, e_i)e_i$. The projection formula can be extended to infinite dimensional subspaces which are spanned by a countable sequence of orthonormal elements $\{e_i\}_1^\infty$.

Basis

A sequence $\{\phi_i\}_1^\infty$ of vectors of a Hilbert space $\mathcal{H}$ is called a *basis* of this space if every vector $f \in \mathcal{H}$ can be expanded in a unique way in a series

$$f = \sum_1^\infty \alpha_i \phi_i = \lim_{n \to \infty} \sum_1^n \alpha_i \phi_i$$

which converges in the norm of $\mathcal{H}$. A Hilbert space can have a basis if and only if it is separable. Such a basis satisfies the following properties: [12]

1. the sequence of vectors is *complete*:[3] a set of vectors is complete if there is no non-zero vector in $\mathcal{H}$ which is orthogonal to every vector in the set;

2. the sequence of vectors is *closed*: its linear envelope $\sum \alpha_i \phi_i$ is dense in $\mathcal{H}$. A set of orthonormal vectors $\{e_i\}$ is closed if and only if for an arbitrary vector $h \in \mathcal{H}$,

$$\| h \|_2^2 = \sum_1^\infty |(h, e_i)|^2 \tag{2.6}$$

(Parseval's equation). This equation can be generalized to non-orthonormal sequences.

An infinite sequence of vectors is complete in a Hilbert space $\mathcal{H}$ if and only if it is closed in $\mathcal{H}$ [12]. In a separable Hilbert space, any complete sequence of orthonormal

[3] In Fuhrmann [14], the definitions of a complete and a closed set of vectors appear the other way round comparing to the definitions in Akhiezer-Glazman [12]. For Hilbert spaces, they are actually equivalent [12].

vectors $\{e_i\}$ forms a basis. In addition, the cardinalities of two orthonormal bases of a separable Hilbert space are equal: they are at most countably infinite, and if there is a finite orthonormal basis $\{e_i\}_1^n$, then any other orthonormal basis has also n elements. The *dimension* of $\mathcal{H}$ is defined as the number of elements in any complete orthonormal basis. Any subspace of a separable Hilbert space is again separable; the dimension of a subspace is defined in the same way. The dimension of a linear manifold $\mathcal{L}$ is defined to be the dimension of its closure $\overline{\mathcal{L}}$.

If two Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ have the same dimension, then they are *isomorphic* in the sense that a one-to-one correspondence between the elements of $\mathcal{H}$ and $\mathcal{H}'$ can be set up, such that, if $f, g \in \mathcal{H}$ and $f', g' \in \mathcal{H}'$ correspond to f, g, then

1. $\alpha f' + \beta g'$ corresponds to $\alpha f + \beta g$;
2. $(f', g')_{\mathcal{H}'} = (f, g)_{\mathcal{H}}$.

In fact, the isometry is defined by the transformation of a complete orthonormal basis in $\mathcal{H}$ into such a basis in $\mathcal{H}'$.

Non-orthogonal basis; Gram matrix

At this point, we include a somewhat less succinct account of the subject of non-orthogonal bases, and in particular on the role of the Gram matrix of such bases.

Let $\{f_1, \cdots, f_n\}$ be a set of n vectors in a Hilbert space $\mathcal{H}$. Consider the matrix $\Lambda_n = [(f_j, f_i)]_{i,j=1}^n$ of inner products of the f_i, *i.e.*,

$$\Lambda_n = \begin{bmatrix} (f_1,f_1) & (f_2,f_1) & \cdots & (f_n,f_1) \\ (f_1,f_2) & (f_2,f_2) & & (f_n,f_2) \\ \vdots & & \ddots & \vdots \\ (f_1,f_n) & (f_2,f_n) & \cdots & (f_n,f_n) \end{bmatrix}.$$

(For an orthonormal set, $\Lambda_n = I$.) The set of vectors is linearly independent if and only if Λ_n is non-singular (*i.e.*, invertible). This can readily be shown from the definition of linear independence: let $f = f_1\alpha_1 + f_2\alpha_2 + \cdots + f_n\alpha_n$ be a vector in the linear manifold generated by the f_i, and suppose that not all α_i are equal to zero. By definition, the set of vectors is linearly independent if $f = 0 \Rightarrow \alpha_i = 0$ $(i = 1, \cdots, n)$. Because $f = 0 \Rightarrow (f, f_i) = 0$ $(i = 1, \cdots, n)$, we obtain upon substituting the definition of f the set of linear equations

$$\begin{cases} (f_1,f_1)\alpha_1 + (f_2,f_1)\alpha_2 + \cdots + (f_n,f_1)\alpha_n = 0 \\ \cdots \\ (f_1,f_n)\alpha_1 + (f_2,f_n)\alpha_2 + \cdots + (f_n,f_n)\alpha_n = 0 \end{cases}$$

and hence $\alpha_i = 0$ $(i = 1, \cdots, n)$ follows if and only if Λ_n is invertible.

Λ_n is called the Gram matrix of the set of vectors. Gram matrices play an important role in the analysis of non-orthogonal bases, as is illustrated by the following. Let $\{f_k\}_1^\infty$ be a complete system of vectors in a Hilbert space $\mathcal{H}$, and let Λ_n be the sequence of Gram matrices $\Lambda_n = [(f_j, f_i)]_{i,j=1}^n$. If

$$\begin{aligned} \lim_{n\to\infty} \|\Lambda_n\| &< \infty \\ \lim_{n\to\infty} \|\Lambda_n^{-1}\| &< \infty \end{aligned}$$

(where $\|\cdot\|$ denotes the matrix 2-norm), then $\{f_k\}_1^\infty$ is a basis in $\mathcal{H}$ [12]. Such a basis is called a Riesz basis. It is said to be equivalent to an orthonormal basis because there is a boundedly invertible transformation (based on Λ) of $\{f_k\}$ to an orthonormal basis.

It can be proven [12] that a sequence of linear independent vectors $\{f_k\}$ is closed if and only if, for an arbitrary element $f \in \mathcal{H}$,

$$\lim_{n\to\infty} \sum_{i,j=1}^{n} (\Lambda_n^{-1})_{ij} (f_i, f)(f, f_j) = \|f\|_2^2 .$$

This is a generalization of Parseval's equation (2.6). Since an infinite sequence of vectors is complete in $\mathcal{H}$ if and only if it is closed in $\mathcal{H}$, a sequence of linearly independent vectors is a basis in $\mathcal{H}$ if and only if it satisfies the generalized Parseval's equation.

The precise way in which the Gram matrix enters into the above is perhaps more clearly seen in the following application. Let $\{f_k\}_1^\infty$ be a non-orthogonal basis in $\mathcal{H}$, and let $\{q_k\}_1^\infty$ be an orthonormal basis. Then the $\{f_i\}$ can be expressed in terms of the $\{q_i\}$ as

$$f_j = \sum_k q_k R_{kj}, \qquad \text{where} \quad R_{ij} = (f_j, q_i). \tag{2.7}$$

Define $R = [R_{ij}]_{i,j=1}^\infty$. The Gram matrix $\Lambda = [(f_j, f_i)]$ can be written in terms of R by the above expansion of f_j as

$$\Lambda_{ij} = \sum_k (q_k, f_i)(f_j, q_k) = \sum_k (R^*)_{ik} R_{kj}$$

so that $\Lambda = R^*R$. Suppose that both R and R^{-1} are bounded. Then Λ and Λ^{-1} are bounded, so that $\{f_k\}$ is a Riesz basis, and the expression $\sum_k R_{ik}(R^{-1})_{kj} = \delta_{ij}$ shows, with (2.7), that each q_j can be written in terms of the $\{f_i\}$:

$$q_j = \sum_i f_i (R^{-1})_{ij} .$$

Hence $\{f_i\}$ can be orthonormalized by R^{-1}, where R is a boundedly invertible factor of Λ. Any $f \in \mathcal{H}$ can be written as

$$\begin{aligned} f &= \sum_j q_j (f, q_j) \\ &= \sum_j \sum_i f_i (R^{-1})_{ij} (f, q_j) \\ &= \sum_i f_i \sum_j (R^{-1})_{ij} (f, q_j) \\ &= \sum_i f_i \sum_j (R^{-1})_{ij} (f, \sum_k f_k (R^{-1})_{kj}) \\ &= \sum_i f_i \sum_j \sum_k (R^{-1})_{ij} [(R^{-1})_{kj}]^* (f, f_k) \\ &= \sum_i f_i \sum_k \left[\sum_j (R^{-1})_{ij} (R^{-*})_{jk} \right] (f, f_k) \\ &= \sum_i f_i \sum_k (R^* R)^{-1}{}_{ik} (f, f_k) \\ &= \sum_i f_i \sum_k (\Lambda^{-1})_{ik} (f, f_k). \end{aligned}$$

(The assumptions on R ensure that the partial sums are bounded, so that the order of summations can indeed be switched.) Hence any $f \in \mathcal{H}$ can be written as $f = \sum_i f_i c_i$, with coefficients $c_i = \sum_k (\Lambda^{-1})_{ik} (f, f_k)$. The generalized Parseval's equation is directly recovered from these expressions, since $\|f\|_2^2 = (f, f) = \sum (f_i, f) c_i$. Further, note that the same derivation holds if the $\{f_i\}$ and $\{q_i\}$ only span a subspace $\mathcal{M}$ in $\mathcal{H}$, so that the projection of $f \in \mathcal{H}$ onto $\mathcal{M}$ can be written as

$$P_{\mathcal{M}} f = \sum_i f_i \sum_k (\Lambda^{-1})_{ik} (f, f_k). \tag{2.8}$$

Bounded linear operators

Let $\mathcal{H}_1$ and $\mathcal{H}_2$ be Hilbert spaces, and let D denote a set in $\mathcal{H}_1$. A function (mapping) T which associates with each element $f \in D$ some element $g = Tf$ in $\mathcal{H}_2$ is called an operator. $D = D(T)$ is called the domain of T, while $\text{ran}(T) = \{Tf : f \in D\}$ is its range. T is linear if D is a linear manifold and $T(\alpha f + \beta g) = \alpha Tf + \beta Tg$ for all $f, g \in D$ and all complex numbers α, β. The *norm* of a linear operator T is

$$\| T \| = \sup_{f \in D, \| f \|_2 < 1} \| Tf \|_2,$$

and T is bounded if $\| T \| < \infty$. A bounded linear operator is continuous: for every $f_0 \in D$,

$$\lim_{f \to f_0} Tf = Tf_0 \qquad (f \in D).$$

If S is another bounded linear operator such that the product ST is defined, then $\| ST \| \leq \| S \| \cdot \| T \|$.

A linear operator T is finite dimensional if it is bounded and if ran (T) is a finite-dimensional subspace of $\mathcal{H}$. Let $\{h_k\}$ be a basis in ran(T), then the operator can be expressed as

$$Tf = \sum_1^n (f, g_k) h_k$$

where $\{g_k\}$ is a finite system of vectors, not depending on f.

Let $T: \mathcal{H}_1 \to \mathcal{H}_2$ be a bounded linear operator defined on the whole of $\mathcal{H}_1$. The adjoint of T is the operator $T^* : \mathcal{H}_2 \to \mathcal{H}_1$, such that

$$(Tf, g) = (f, T^* g)$$

for all $f, g \in \mathcal{H}_1$. T^* exists and is unique, $(T^*)^* = T$, $(ST)^* = T^* S^*$, and if T^{-1} exists then $(T^{-1})^* = (T^*)^{-1}$. T is called self-adjoint if $T = T^*$; a self-adjoint operator is called positive if $(Tf, f) \geq 0$ for all $f \in \mathcal{H}_1$.

Let $\{e_k\}_1^\infty$ be an orthonormal basis in $\mathcal{H}$. The trace of an operator T is defined as

$$\text{trace}(T) := \sum_1^\infty (Te_k, e_k),$$

whenever this series converges absolutely. In this case, the sum does not depend on the basis chosen. Operators with a finite trace are the so-called *nuclear operators* [12].

The null-space or kernel of a bounded linear operator $T: \mathcal{H}_1 \to \mathcal{H}_2$ is the linear manifold

$$\ker(T) = \{f \in \mathcal{H}_1 : Tf = 0\}.$$

This linear manifold is actually closed, hence $\ker(T)$ is a subspace. On the other hand, the range of T is a linear manifold which is not necessarily closed; it is closed if and only if the range of its adjoint is closed. $\mathcal{H}_1$ and $\mathcal{H}_2$ satisfy an orthogonal decomposition as

$$\begin{aligned} \mathcal{H}_1 &= \ker(T) \oplus \overline{\text{ran}(T^*)} \\ \mathcal{H}_2 &= \ker(T^*) \oplus \overline{\text{ran}(T)}. \end{aligned} \tag{2.9}$$

T is said to be *injective* (one-to-one) if $Tf = Tg \Rightarrow f = g$, which reduces for linear operators to the condition $Tf = 0 \Rightarrow f = 0$, *i.e.*, T is injective if and only if $\ker(T) = 0$. Hence if the range of T^* is dense in $\mathcal{H}_1$, then T is one-to-one. T is *surjective* (onto) if its range is all of $\mathcal{H}_2$. T with domain restricted to $\ker(T)^\perp$ maps one-to-one to the closure of its range, but is not necessarily surjective. If T is both injective and surjective, then (by the closed graph theorem [15]) it is boundedly invertible.

An operator P is a projection if it satisfies $P^2 = P$. It is called an orthogonal projection if, in addition, $P^* = P$. If $\mathcal{M}$ is a subspace in $\mathcal{H}$, then $\mathcal{H} = \mathcal{M} \oplus \mathcal{M}^\perp$; the orthogonal projector $P_{\mathcal{M}}$ whose range is $\mathcal{M}$ is unique.

The following theorem gives necessary and sufficient conditions for the range of an operator to be closed (*cf.* [13, §21], [16]):

THEOREM 2.1. *Let T be a bounded operator on a Hilbert space.*

$$\text{ran}(T^*) \text{ is closed} \quad \Leftrightarrow \quad \exists\, \varepsilon > 0 : \quad \|Tx\| \geq \varepsilon \|x\| \quad \text{for all } x \in \overline{\text{ran}(T^*)}. \tag{2.10}$$

We know already that $\text{ran}(T^*)$ is closed if and only if $\text{ran}(T)$ is closed.

A linear manifold (subspace) $\mathcal{M}$ is called an invariant manifold (subspace) under an operator T if $\mathcal{M}T \subset \mathcal{M}$. $\mathcal{M}$ is invariant under T if and only if $P_{\mathcal{M}} T P_{\mathcal{M}} = T P_{\mathcal{M}}$.

An operator U is called an isometry if it satisfies $U^*U = I$, a co-isometry if $UU^* = I$, and unitary if it satisfies both. If U is unitary, then it is invertible, and $U^{-1} = U^*$. Two Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are isometrically isomorphic if there exists an invertible transformation U such that

$$(Uf, Ug)_2 = (f, g)_1 \qquad (\text{for all } f, g \in \mathcal{H}_1).$$

In this case, U is unitary.

2.2 NON-UNIFORM SIGNALS AND TRANSFER OPERATORS

In the previous section, we encountered the space ℓ_2 as an example of a separable Hilbert space. In this section, we generalize this space to sequences whose entries are not elements of the same space. The 'non-uniform sequences' yielded are still vectors in a separable Hilbert space, and since two separable Hilbert spaces of the same dimension are isomorphic, they are in fact isomorphic to ordinary ℓ_2-sequences. While this is true (and allows standard results from operator theory to apply in the context of non-uniform sequences), the additional structure admits an interpretation of these sequences in terms of signals that play a role in time-varying system theory. The necessary notation is defined in this section.

Signals and index sequences

A signal is by definition any quantity that varies with time, space and/or possibly other independent quantities. We consider only one-dimensional signals, which are functions of one index: 'time'. Also, we consider only discrete-time signals: our signal is an infinite sequence (written as a row vector)

$$u = [\, \cdots \quad u_{-1} \quad \boxed{u_0} \quad u_1 \quad u_2 \quad \cdots \,] \tag{2.11}$$

of components u_i, the value of the signal at time instant i. The square surrounding u_0 identifies this entry as the entry on the zeroth position. If the u_i are scalar, then u is a

one-channel signal. A more general situation is obtained by taking the u_i to be (row) vectors themselves, which makes u a multi-channel signal. It is not necessary that all u_i have equal dimensions: we allow for a time-varying number of channels, or equivalently, for non-uniform sequences. In order to describe such objects properly, we introduce the notion of *index sequences*.

Let $\{N_i \in \mathbb{N}, i \in \mathbb{Z}\}$ be an indexed collection of finite natural numbers.[4]

$$N = [N_i]_{-\infty}^{\infty} = [\cdots \quad N_{-1} \quad \boxed{N_0} \quad N_1 \quad N_2 \quad \cdots] \in \mathbb{N}^{\mathbb{Z}}$$

is called an index sequence. Using N, signals (2.11) live in the space of non-uniform sequences which is the Cartesian product of the $\mathcal{N}_i$:

$$\mathcal{N} = \cdots \times \mathcal{N}_{-1} \times \boxed{\mathcal{N}_0} \times \mathcal{N}_1 \times \mathcal{N}_2 \times \cdots \in \mathbb{C}^N,$$

where $\mathcal{N}_i \in \mathbb{C}^{N_i}$ so that N_i is the dimension of $\mathcal{N}_i$. A signal in such a space can be viewed as an infinite sequence that has been partitioned into an infinite number of finite dimensional components. Some of these components may have zero dimension to reflect the fact that no input signal is present at that point in time: we define $\mathbb{C}^0 = \emptyset$.[5] Thus, finite dimensional vectors are also incorporated in the space of non-uniform sequences, by putting $N_i = 0$ for i outside a finite interval. We adopt the notation '$\cdot n$' for an index set with all its components equal to n, so that $\mathbb{C}^{\cdot n}$ is a (uniform) sequence of n-dimensional vectors. If $\mathcal{N} = \mathbb{C}^N$, then to retrieve N from $\mathcal{N}$ we write

$$N = \#(\mathcal{N}).$$

It is sometimes convenient to have named operators to construct a sequence from its entries. Following [2], we define for a given space sequence $\mathcal{N}$,

$$\pi_k : a \in \mathcal{N}_k \quad \rightarrow \quad u = a\pi_k \in \mathcal{N} \text{ where } \begin{cases} u_i = a, & (i = k), \\ u_i = 0, & (i \neq k), \end{cases} \tag{2.12}$$

with adjoint

$$\pi_k^* : u \in \mathcal{N} \quad \rightarrow \quad u_k = u\pi_k^* \in \mathcal{N}_k.$$

The operator π_k constructs a sequence out of an element of $\mathcal{N}_k$, by embedding it into a sequence which is otherwise zero: it corresponds to the operator $\pi_k = [\cdots \; 0 \; I_{\mathcal{N}_k} \; 0 \; \cdots]$. Its adjoint π_k^* retrieves the k-th (block) entry of the sequence. We often implicitly use the fact that $\pi_k \pi_k^* = 1$ and $\sum_k \pi_k^* \pi_k = I_{\mathcal{N}}$ (the identity operator on $\mathcal{N}$).

[4] 0 is included in $\mathbb{N}$.

[5] Actually, $\mathbb{C}^0$ contains one element, the 'zero-vector' of size 1×0. We will not make a distinction between the two.

The inner product of two non-uniform sequences f, g in $\mathcal{N}$ is defined in terms of the usual inner product of (row)-vectors in $\mathcal{N}_i$ as

$$(f, g) = \sum_i (f_i, g_i)$$

where $(f_i, g_i) = f_i g_i^*$ is defined to be 0 if $N_i = 0$.[6] The norm of a non-uniform sequence is the standard 2-norm (vector norm) defined on this inner product:

$$u = [u_i]_{-\infty}^{\infty} : \quad \| u \|_2^2 = (u, u) = \sum_{i=-\infty}^{\infty} \| u_i \|_2^2$$

so that $\| u \|_2^2$ represents the energy of the signal. The space of non-uniform sequences in $\mathcal{N}$ with finite 2-norm is denoted by $\ell_2^{\mathcal{N}}$:

$$\ell_2^{\mathcal{N}} = \{u \in \mathcal{N} : \| u \|_2 < \infty\}.$$

This space can be viewed as an ordinary separable Hilbert space of sequences on which a certain regrouping (of scalars into finite dimensional vectors) has been superimposed. Consequently, properties of Hilbert spaces carry over to the present context when this grouping is suppressed.

To illustrate some of the above, let $N = [\cdots \ 0 \ \boxed{1} \ 3 \ 2 \ 0 \ \cdots]$. An element of the non-uniform sequence $\mathcal{N} = \mathbb{C}^N$ is, *e.g.*, the vector $u = [\boxed{6}, [3\ 2\ 1], [4\ 2]]$ (suppressing entries with zero dimensions). The norm of u is then given by $\| u \|_2^2 = 6^2 + (3^2 + 2^2 + 1^2) + (4^2 + 2^2)$.

Operator spaces

Let $\mathcal{M}$ and $\mathcal{N}$ be spaces of sequences corresponding to index sequences M, N. When we consider sequences in these spaces as signals, then a system that maps ingoing signals in $\mathcal{M}$ into outgoing signals in $\mathcal{N}$ is described by an operator from $\mathcal{M}$ to $\mathcal{N}$:

$$T : \mathcal{M} \to \mathcal{N}, \quad y = (uT).$$

From this point of view, we will call such an operator on signals the input-output map or *transfer operator* of the system. A transfer operator is linear if for all scalars α_1, α_2 and $u_1, u_2 \in \mathcal{M}$,

$$(\alpha_1 u_1 + \alpha_2 u_2)T = \alpha_1 (u_1 T) + \alpha_2 (u_2 T)$$

in which case T is a linear operator. Linear transfer operators correspond to linear operators $T : \mathcal{M} \to \mathcal{N}$ if we write the operands at the left of the operator (as we have done here).

[6] More in general, we define the product of an $n \times 0$ matrix with a $0 \times m$ matrix to be the zero matrix of dimensions $n \times m$.

We denote by $\mathcal{X}(\mathcal{M}, \mathcal{N})$ the space of *bounded* linear operators $\ell_2^{\mathcal{M}} \to \ell_2^{\mathcal{N}}$: an operator T is in $\mathcal{X}(\mathcal{M}, \mathcal{N})$ if and only if for each $u \in \ell_2^{\mathcal{M}}$, the result $y = uT$ is in $\ell_2^{\mathcal{N}}$, in which case the induced operator norm of T,

$$\|T\| = \sup_{\|u\|_2 \le 1} \|uT\|_2,$$

is bounded. It is well known that a bounded operator defined everywhere on a separable Hilbert space $\mathcal{H}_1$ and mapping into a separable Hilbert space $\mathcal{H}_2$ admits a matrix representation which uniquely determines the operator [12]. The matrix representation hinges on the choice of orthonormal bases in $\mathcal{H}_1$ and $\mathcal{H}_2$ for which one takes, typically, the standard basis $\{e_k\}$ of equation (2.4). The same is true in the present context [2]: $T \in \mathcal{X}(\mathcal{M}, \mathcal{N})$ has a matrix representation $[T_{ij}]_{i,j=-\infty}^{\infty}$ such that

$$y = uT \quad \Leftrightarrow \quad \forall j: \; y_j = \sum_i u_i T_{ij}. \tag{2.13}$$

As is usual, we identify T with its matrix representation and write

$$T = [T_{ij}]_{i,j=-\infty}^{\infty} = \begin{bmatrix} \ddots & & \vdots & & \iddots \\ & T_{-1,-1} & T_{-1,0} & T_{-1,1} & \\ \cdots & T_{0,-1} & \boxed{T_{00}} & T_{01} & \cdots \\ & T_{1,-1} & T_{10} & T_{11} & \\ \iddots & & \vdots & & \ddots \end{bmatrix} \tag{2.14}$$

(where the square identifies the 00-entry) such that it fits the usual vector-matrix multiplication rules. The block entry T_{ij} is an $M_i \times N_j$ matrix and is given by $T_{ij} = \pi_i T \pi_j^*$. This is seen by substitution into equation (2.13), which yields the following steps to compute y from u: (*i*) constructing a sequence from each u_i, (*ii*) applying T to these sequences, (*iii*) selecting the j-th entry from each of the resulting sequences, and (*iv*) summing these entries over i. With regard to (2.14), the operator $T_i = \pi_i T$ can be called the i-th (block) row of T, while $T\pi_j^*$ is the j-th column of T.

T is called a Toeplitz operator if $T_{ij} = T_{i-j}$ for all i, j, *i.e.*, if T is constant along the diagonals of its matrix representation. Such operators correspond to time-invariant systems, as will be discussed in chapter 3.

In $\mathcal{X}(\mathcal{M}, \mathcal{N})$, we define the space of bounded upper operators

$$\mathcal{U}(\mathcal{M}, \mathcal{N}) = \{T \in \mathcal{X}(\mathcal{M}, \mathcal{N}) : T_{ij} = 0 \quad (i > j)\},$$

the space of bounded lower operators

$$\mathcal{L}(\mathcal{M}, \mathcal{N}) = \{T \in \mathcal{X}(\mathcal{M}, \mathcal{N}) : T_{ij} = 0 \quad (i < j)\},$$

and the space of bounded diagonal operators

$$\mathcal{D} = \mathcal{U} \cap \mathcal{L}.$$

As a matter of notational convenience, we often just write $\mathcal{X}, \mathcal{U}, \mathcal{L}, \mathcal{D}$ when the underlying spaces are clear from the context or are of no particular importance. For $A \in \mathcal{D}$, "A_i" serves as shorthand for the entry A_{ii}, and we write

$$A = \mathrm{diag}[\cdots\ A_{-1}\ \boxed{A_0}\ A_1\ \cdots] = \mathrm{diag}[A_i].$$

$\mathcal{U}$, $\mathcal{L}$ and $\mathcal{D}$ satisfy the following elementary properties [2]:

$$\begin{array}{lll} \mathcal{U}\cdot\mathcal{U} \subset \mathcal{U} & & \mathcal{L}^* = \mathcal{U} \\ \mathcal{L}\cdot\mathcal{L} \subset \mathcal{L} & & \mathcal{U}^* = \mathcal{L} \\ \mathcal{D}\cdot\mathcal{D} \subset \mathcal{D}. & & \end{array} \tag{2.15}$$

The operators in $\mathcal{U}$ form the space of bounded causal operators. They are called causal because the output to an input that starts at a certain time i, say (for $i = 0$),

$$u = [\cdots\ 0\ \boxed{u_0}\ u_1\ u_2\ \cdots]$$

yields an output that is zero before time i:

$$y = uT = [\cdots\ 0,\ \boxed{u_0T_{00}},\ u_1T_{11} + u_0T_{01},\ u_2T_{22} + u_1T_{12} + u_0T_{02},\ \cdots]$$

If $D \in \mathcal{D}$ and invertible, then $D^{-1} \in \mathcal{D}$, and $(D^{-1})_i = (D_i)^{-1}$ [2]. However, unlike the situation for finite-size matrices on uniform sequences, the spaces $\mathcal{U}$ and $\mathcal{L}$ are not closed under inversion: if an upper operator $T \in \mathcal{U}$ is boundedly invertible, then the inverse is not necessarily upper. A simple example of this is given by the pair of Toeplitz operators

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & & \\ & \boxed{1} & -2 & 0 & \\ & & 1 & -2 & \ddots \\ & \mathbf{0} & & 1 & \ddots \\ & & & & \ddots \end{bmatrix}, \qquad T^{-1} = \begin{bmatrix} \ddots & & & & \\ \ddots & \boxed{0} & & & \mathbf{0} \\ & -1/2 & 0 & & \\ & -1/4 & -1/2 & 0 & \\ \cdots & -1/8 & -1/4 & -1/2 & 0 \\ & \vdots & & & \ddots\ \ddots \end{bmatrix}$$

But also for finite-size matrices based on *non-uniform* space sequences, the same can occur:

$$T = \begin{array}{c} \\ \mathbb{C}^2 \\ \mathbb{C} \\ \varnothing \end{array} \overset{\begin{array}{ccc} \mathbb{C} & \mathbb{C} & \mathbb{C} \end{array}}{\begin{bmatrix} \boxed{1} & 0 & 0 \\ \boxed{1/2} & 2 & 0 \\ 0 & \underline{1/4} & 1 \\ \cdot & \cdot & \underline{\cdot} \end{bmatrix}}, \qquad T^{-1} = \begin{array}{c} \\ \mathbb{C} \\ \mathbb{C} \\ \mathbb{C} \end{array} \overset{\begin{array}{cccc} & \mathbb{C}^2 & & \mathbb{C} \quad \varnothing \end{array}}{\begin{bmatrix} \boxed{1 \quad 0} & 0 & \cdot \\ -1/4 \quad 1/2 & \underline{0} & \cdot \\ 1/16 \quad -1/8 & 1 & \underline{\cdot} \end{bmatrix}} \tag{2.16}$$

(the underscore identifies the position of the 0-th diagonal). When viewed as matrices without considering their structure, T^{-1} is of course just the matrix inverse of T. Mixed cases (the inverse has a lower and an upper part) can also occur, and these inverses are not trivially computed, as they require a 'dichotomy': a splitting of spaces into a part that determines the upper part and a part that gives the lower part. The topic will be investigated in chapter 4. An important special case of upper operators with upper inverses is the following. An operator of the form $(I-X)$, where X is a bounded operator, has an inverse that is given by the series expansion (Neumann expansion)

$$(I-X)^{-1} = I + X + X^2 + \cdots \tag{2.17}$$

only if the series converges in norm. It is known in operator theory that the series converges strongly if and only if the geometric series $1 + \|X\| + \|X^2\| + \cdots$ converges, which it is known to do if the spectral radius $r(X)$ of X is smaller than 1:[7]

$$r(X) := \lim_{n\to\infty} \|X^n\|^{1/n} < 1.$$

The above is consolidated in the following proposition.

PROPOSITION 2.2. *If $X \in \mathcal{U}$ and $r(X) < 1$, then $(I-X)^{-1}$ is given by (2.17) and is also in $\mathcal{U}$.*

We will use two properties related to the spectral radius of an operator in later chapters. The first is that $r(X) \le \|X\|$, because $\|X^n\|^{1/n} \le (\|X\|^n)^{1/n}$. The second property makes this more precise: the sequence $\|X^n\|^{1/n}$ monotonically decreases when n goes to infinity:

$$\|X^{n+1}\|^{\frac{1}{n+1}} = \|(X^n)^{\frac{n+1}{n}}\|^{\frac{1}{n+1}} \le \|X^n\|^{\frac{n+1}{n}\frac{1}{n+1}} = \|X^n\|^{1/n}. \tag{2.18}$$

Hilbert-Schmidt operators

Based on the space $\mathcal{X}(\mathcal{M},\mathcal{N})$ of $[\ell_2^{\mathcal{M}} \to \ell_2^{\mathcal{N}}]$ operators bounded in operator norm, we define the Hilbert-Schmidt space $\mathcal{X}_2(\mathcal{M},\mathcal{N})$, consisting of elements of $\mathcal{X}$ that are also bounded in Hilbert-Schmidt norm. The *Hilbert-Schmidt norm* is defined as

$$\|A\|_{HS}^2 = \sum_{i,j} \|A_{ij}\|_2^2 \qquad (A \in \mathcal{X}(\mathcal{M},\mathcal{N})),$$

where $\|A_{ij}\|_2^2$ is, in turn, equal to the sum of the entries of A_{ij} squared. For finite matrices, the Hilbert-Schmidt norm is usually called the Frobenius norm. $\mathcal{X}_2(\mathcal{M},\mathcal{N})$ is thus given by

$$\mathcal{X}_2(\mathcal{M},\mathcal{N}) = \{A \in \mathcal{X}(\mathcal{M},\mathcal{N}) : \|A\|_{HS}^2 < \infty\}.$$

[7]For readers not familiar with the concept of spectral radius, we mention that for a finite matrix X, $r(X)$ is equal to the largest eigenvalue of X. In the context of operators, however, the spectrum is more complicated. See [12].

On $\mathcal{X}_2(\mathcal{M}, \mathcal{N})$, the Hilbert-Schmidt inner product

$$\langle A, B\rangle_{HS} = \text{trace}(AB^*)$$

can be defined, and the Hilbert Schmidt norm satisfies $\| A \|_{HS}^2 = \langle A, A\rangle_{HS} = \text{trace}(AA^*)$. $\mathcal{X}_2$ is a Hilbert space for the Hilbert-Schmidt inner product (it becomes an ordinary Hilbert space of sequences if the entries A_{ij} are written as one sequence). Related spaces are the spaces of upper, lower and diagonal Hilbert-Schmidt spaces, given by, respectively,

$$\begin{aligned} \mathcal{U}_2 &= \mathcal{X}_2 \cap \mathcal{U} \\ \mathcal{L}_2 &= \mathcal{X}_2 \cap \mathcal{L} \\ \mathcal{D}_2 &= \mathcal{X}_2 \cap \mathcal{D}. \end{aligned} \tag{2.19}$$

We define $\mathbf{P}_{\mathcal{H}}$ as the orthogonal projection operator of $\mathcal{X}_2$ onto some subspace $\mathcal{H}$ of $\mathcal{X}_2$. The following projections are designated by their own symbol:

$$\begin{aligned} &\mathbf{P} \ : \text{ the projection operator of } \mathcal{X}_2 \text{ onto } \mathcal{U}_2, \text{ and} \\ &\mathbf{P}_0 \ : \text{ the projection operator of } \mathcal{X}_2 \text{ onto } \mathcal{D}_2. \end{aligned} \tag{2.20}$$

These projections are bounded operators on Hilbert-Schmidt spaces in the induced Hilbert-Schmidt operator norm. They are in general not bounded operators on $\mathcal{X}$ (which is one of the reasons for introducing Hilbert-Schmidt spaces). This situation generalizes what already happens with Toeplitz operators. An example of this, taken from [8], is the following. Consider the semi-infinite Toeplitz matrix

$$T = \begin{bmatrix} \ddots & \vdots & & & & & \vdots & \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ & 0 & \boxed{0} & -1 & -1/2 & -1/3 & -1/4 & \cdots \\ & 0 & 1 & 0 & -1 & -1/2 & -1/3 & \\ & 0 & 1/2 & 1 & 0 & -1 & -1/2 & \\ & 0 & 1/3 & 1/2 & 1 & 0 & -1 & \\ \cdots & 0 & 1/4 & 1/3 & 1/2 & 1 & 0 & \cdots \\ & \vdots & \vdots & & & & \vdots & \ddots \end{bmatrix}$$

The only candidate matrix representation for the upper part $\mathbf{P}(T)$ of T is

$$\mathbf{P}(T) = \begin{bmatrix} \ddots & \vdots & & & & & \vdots & \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ & 0 & \boxed{0} & -1 & -1/2 & -1/3 & -1/4 & \cdots \\ & 0 & 0 & 0 & -1 & -1/2 & -1/3 & \\ & 0 & 0 & 0 & 0 & -1 & -1/2 & \\ & 0 & 0 & 0 & 0 & 0 & -1 & \\ \cdots & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ & \vdots & \vdots & & & & \vdots & \ddots \end{bmatrix}$$

It is proven in [8] using Fourier theory that T is bounded: $\|T\| = \pi$, while $\mathbf{P}(T)$ is unbounded. However, the projection onto the central diagonal, $\mathbf{P}_0$, is also bounded on $\mathcal{X}$: each diagonal of a bounded operator, taken by itself, is again a bounded operator with its norm not exceeding the norm of the original operator.

We use the following properties of $\mathbf{P}_0$:

$$\begin{aligned} \mathbf{P}_0(D_1 X D_2) &= D_1\,\mathbf{P}_0(X)\,D_2 \qquad (D_{1,2} \in \mathcal{D},\ X \in \mathcal{X}) \\ [\mathbf{P}_0(X)]^* &= \mathbf{P}_0(X^*) \end{aligned}$$

Operators in $\mathcal{X}_2$ satisfy the 'two-sided ideal' properties: if $A \in \mathcal{X}_2$, $B \in \mathcal{X}$ such that the product AB is defined, then $AB \in \mathcal{X}_2$; a similar result holds for BA if this product is defined. A consequence of this property is that the Hilbert-Schmidt spaces $\mathcal{X}_2$ can be considered to be input or output spaces for transfer operators in $\mathcal{X}$, as a generalization of ℓ_2 sequences. This will be the topic of section 2.3.

Shift operators

For some index sequence $N = [\cdots\ N_{-1}\ \boxed{N_0}\ N_1\ \cdots]$, the k-th right-shifted sequence is denoted $N^{(k)} = [\cdots\ N_{-k-1}\ \boxed{N_{-k}}\ N_{-k+1}\ \cdots]$. The corresponding right-shifted space sequence is denoted $\mathcal{N}^{(k)} = \mathbf{C}^{N^{(k)}}$. The right (or causal) bilateral shift operator $Z = Z_{\mathcal{N}}$ on sequences $u \in \mathcal{N}$ is defined by $(uZ)_i = u_{i-1}$, *i.e.*,

$$[\cdots\ \boxed{u_0}\ u_1\ u_2\ \cdots]\,Z = [\cdots\ \boxed{u_{-1}}\ u_0\ u_1\ \cdots].$$

$Z_{\mathcal{N}}$ is an operator $\ell_2^{\mathcal{N}} \to \ell_2^{\mathcal{N}^{(1)}}$. It is readily checked from its definition that

$$Z_{ij} = \pi_i Z \pi_j^* = \begin{cases} I, & \text{if } j = i+1, \\ 0, & \text{otherwise,} \end{cases}$$

so that $Z \in \mathcal{U}$ and Z has a matrix representation

$$Z = \begin{bmatrix} \ddots & \ddots & & & \\ & 0 & I_{N_{-1}\times N_{-1}} & & \mathbf{0} \\ & & \boxed{0} & I_{N_0\times N_0} & \\ & & & 0 & I_{N_1\times N_1} \\ & \mathbf{0} & & & 0 & \ddots \\ & & & & & \ddots \end{bmatrix}.$$

Z is unitary on $\ell_2^{\mathcal{N}}$: $ZZ^* = I$, $Z^*Z = I$, so that $Z^{-1} = Z^*$. The operator $Z^{[k]}$ denotes the k-times repeated application of Z:

$$Z^{[k]} = Z_{\mathcal{N}} Z_{\mathcal{N}^{(1)}} \cdots Z_{\mathcal{N}^{(k-1)}}.$$

Note that formally Z^k is not well defined because the dimensions in the multiplications do not match. Nonetheless, as a relaxation of notation we will, in future sections, usually suppress dimension information in formulas and just write Z^k instead of $Z^{[k]}$.

Since $Z \in \mathcal{U}$, the properties in equation (2.15) specialize to [2]

$$\begin{array}{lcll} Z\mathcal{U} & \subset & \mathcal{U} & \\ \mathcal{U}Z & \subset & \mathcal{U} & \mathcal{L} \cap Z\mathcal{U} = 0 \\ Z^{-1}\mathcal{L} & \subset & \mathcal{L} & \mathcal{L}Z^{-1} \cap \mathcal{U} = 0 \\ \mathcal{L}Z^{-1} & \subset & \mathcal{L} & \end{array}$$

The same type of properties hold for $\mathcal{U}_2$ and $\mathcal{L}_2$.

It is a fundamental fact (and proven in [2]) that $\mathcal{U}_2 \perp \mathcal{L}_2 Z^{-1}$ and $\mathcal{U}_2 \perp Z^{-1}\mathcal{L}_2$, and that $\mathcal{X}_2$ admits an orthogonal decomposition

$$\mathcal{X}_2 = \mathcal{L}_2 Z^{-1} \oplus \mathcal{U}_2 = \mathcal{L}_2 Z^{-1} \oplus \mathcal{D}_2 \oplus \mathcal{U}_2 Z.$$

Diagonal shifts

One of the aspects of the operators in $\mathcal{X}$ is that they do not commute with the shift operator: let $T \in \mathcal{X}(\mathcal{M}, \mathcal{N})$, then if we define $T^{(1)}$ by

$$Z_{\mathcal{M}} T^{(1)} = T Z_{\mathcal{N}},$$

that is, $T^{(1)} = Z^* T Z$, then $T^{(1)}$ is the operator T shifted one position in the southeast direction: $(T^{(1)})_{i,j} = T_{i-1,j-1}$. If T does commute with the shift operator, $T^{(1)} = T$, then $T_{i,j} = T_{i-1,j-1}$ and T is a Toeplitz operator. More in general, the k-th diagonal shift of $T \in \mathcal{X}(\mathcal{M}, \mathcal{N})$ in the southeast direction along the diagonals of T is defined by

$$T^{(k)} = \left[Z^{[k]}\right]^* T Z^{[k]},$$

which is in $\mathcal{X}(\mathcal{M}^{(k)}, \mathcal{N}^{(k)})$. This is equivalent to saying that $(T^{(k)})_{ij} = T_{i-k,j-k}$. The diagonal shift takes each of the spaces $\mathcal{L}$, $\mathcal{U}$ and $\mathcal{D}$ into themselves (albeit with shifted index sequences); it is readily verified that if $S, T \in \mathcal{X}$ such that the product ST is well defined, then

$$(ST)^{(k)} = S^{(k)} T^{(k)}, \qquad T^{(k+m)} = (T^{(k)})^{(m)}.$$

We will in future chapters often run across products $(AZ)^n$, where $A \in \mathcal{X}(\mathcal{N}, \mathcal{N}^{(-1)})$. These are evaluated as

$$\begin{array}{rcl} (AZ)^n & = & (AZ)(AZ)\cdots(AZ) \\ & = & Z^{[n]} A^{(n)} A^{(n-1)} \cdots A^{(1)} \\ & =: & Z^{[n]} A^{\{n\}} \end{array}$$

where $A^{\{n\}}$ is defined as

$$\begin{array}{rcl} A^{\{0\}} & = & I \\ A^{\{n\}} & = & A^{(n)} A^{\{n-1\}} = A^{(n)} A^{(n-1)} \cdots A^{(1)}. \end{array} \tag{2.21}$$

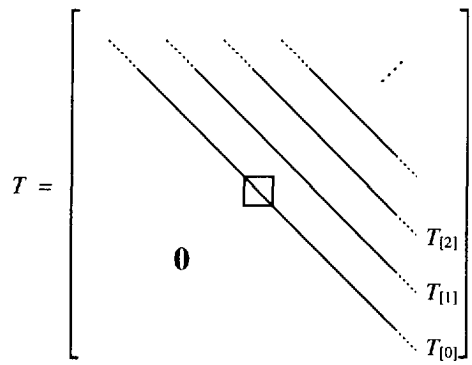


Figure 2.1. Diagonal decomposition of an operator $T \in \mathcal{U}$.

Diagonal representation

For $T \in \mathcal{X}(\mathcal{M}, \mathcal{N})$, let $T_{[k]} \in \mathcal{D}(\mathcal{M}^{(k)}, \mathcal{N})$ denote the k-th subdiagonal above the central (0-th) diagonal of T:

$$T_{[k]} = \mathbf{P}_0(Z^{-k}T), \qquad (T_{[k]})_i = T_{i-k,i}.$$

Based on a recursive use of the property $\mathcal{U} = \mathcal{D} + Z\mathcal{U}$, it is proven in [2] that, for $T \in \mathcal{U}$,

$$T - \sum_{k=0}^{n} Z^{[k]}T_{[k]} \in Z^{[n+1]}\mathcal{U}$$

so that T formally has a decomposition into a sum of shifted diagonals (see figure 2.1). Although the collection $\{T_{[k]}\}_0^\infty$ uniquely specifies T, the sum need not converge to T for $n \to \infty$ in a uniform sense [2]. However, for operators in $\mathcal{U}_2$ the sum does converge in the Hilbert-Schmidt norm, which provides another reason for the use of Hilbert-Schmidt spaces:

$$U \in \mathcal{X}_2: \qquad U = \sum_{-\infty}^{\infty} Z^{[k]}U_{[k]}, \qquad U_{[k]} = \mathbf{P}_0(Z^{[-k]}U).$$

For an operator $T \in \mathcal{U}$, we can give meaning to the formal expression

$$T = \sum_{k=0}^{\infty} Z^{[k]}T_{[k]} \tag{2.22}$$

by premultiplying T with an element U of $\mathcal{X}_2$, and analyzing the diagonals of $Y = UT$ in terms of those of U and T. In analogy with the definition of a matrix representation of

an operator [12, §29], we say that an operator T admits a diagonal representation $\{T_{[k]}\}_0^\infty$ written formally as (2.22), if Y in the multiplication $Y = UT$, with $U, Y \in \mathcal{X}_2$,

$$\begin{array}{rclcrcl} U & = & \sum_i Z^{[i]} U_{[i]} & \text{where} & U_{[i]} & = & \mathbf{P}_0(Z^{[-i]}U), \\ Y & = & \sum_i Z^{[i]} Y_{[i]} & \text{where} & Y_{[i]} & = & \mathbf{P}_0(Z^{[-i]}Y) \end{array} \tag{2.23}$$

is given by

$$Y_{[i]} = \sum_{k=0}^{\infty} U^{(k)}_{[i-k]} T_{[k]} \tag{2.24}$$

where the latter sum is to converge in the Hilbert-Schmidt metric. (Note the diagonal shift in $U^{(k)}_{[i-k]}$.) The formula for $Y_{[i]}$ can be regarded as a convolution product of diagonals.

Every bounded linear operator $T \in \mathcal{U}$ defined throughout $\mathcal{X}_2$ admits such a diagonal representation: it is straightforward to show that the summation for $Y_{[i]}$ is convergent. The definition (2.24) allows the replacement of the summations that follow from the formal series (2.22),

$$\begin{array}{rcl} Y = UT & = & \sum_{k=0}^{\infty} UZ^{[k]} T_{[k]} \\ & = & \sum_{k=0}^{\infty} \sum_i Z^{[i]} U_{[i]} Z^{[k]} T_{[k]} \\ & = & \sum_{k=0}^{\infty} \sum_i Z^{[i+k]} U^{(k)}_{[i]} T_{[k]} \\ & = & \sum_{k=0}^{\infty} \sum_i Z^{[i]} U^{(k)}_{[i-k]} T_{[k]}, \end{array}$$

by $Y = \sum_i \sum_{k=0}^{\infty} Z^{[i]} U^{(k)}_{[i-k]} T_{[k]} = \sum_i Z^{[i]} Y_{[i]}$, where $Y_{[i]}$ is as given in (2.24).

2.3 THE DIAGONAL ALGEBRA OF $\mathcal{X}_2$

In the previous section, we introduced the Hilbert-Schmidt space $\mathcal{X}_2$, which is a Hilbert space, and we remarked that operators in this space satisfy the 'ideal' property

$$U \in \mathcal{X}_2, \; T \in \mathcal{X} \qquad \Rightarrow \qquad Y = UT \in \mathcal{X}_2.$$

In addition, if $U \in \mathcal{X}_2$, then the sum of the squared entries of its matrix representation is finite. In particular, if $U \in \mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$, then the i-th row $U_i = \pi_i U$ is in $\ell_2^{\mathcal{M}}$, which is the same space for all i (recall that $\mathbf{C}^{\bullet 1}$ is the space of uniform sequences with scalar entries). Hence, we can view U as a stack of independent sequences $\{U_i\}$ in $\ell_2^{\mathcal{M}}$, or, mathematically, $\mathcal{X}_2^{\mathcal{M}}$ can be viewed as the Hilbert space of Cartesian products '$\cdots \times \ell_2^{\mathcal{M}} \times \ell_2^{\mathcal{M}} \times \cdots$' of $\ell_2^{\mathcal{M}}$-sequences. The rows of $U \in \mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$ act as independent inputs sequences to T:

$$Y = UT \quad \Leftrightarrow \quad \forall i: \; \pi_i Y = \pi_i UT = (\pi_i U)T \quad \Leftrightarrow \quad \forall i: \; Y_i = U_i T, \tag{2.25}$$

so that applying T to U is equivalent to applying T to all rows U_i independently.

In future chapters, we will often study systems by considering the application of a collection of signals to a transfer operator T. From the properties mentioned above, it follows that we can regard an element of $\mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$ as a generalized input operator, consisting of an infinite collection of $\ell_2^{\mathcal{M}}$ sequences. Applying T directly to such generalized inputs will then lead to a significant simplification of notation, as less indices are required. For example, in chapter 3 we will study the effect of applying inputs U_i that stop at a certain point in time, say $(U_i)_k = 0$ $(k > i)$. This set of input signals can be applied all at once by taking $U = [(U_i)_k] \in \mathcal{L}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$ as a generalized input operator.

Mathematically, the above concept of a generalized input sequence translates to viewing an element of $\mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$ as a (row) sequence of *diagonals*. Based on this idea, a non-commutative algebra can be set up in which diagonals play the role of scalars and the Hilbert space of ℓ_2-sequences becomes a Hilbert space *module* of sequences of diagonals (*cf.* [17]). In the same way, the scalar Hilbert space inner product translates to a diagonal inner product in the Hilbert space module. This idea of such a diagonal algebra originates to Alpay, Dewilde and Dym [2]. We omit the (rather standard) proof that an *algebra* is obtained, and confine ourselves to proving the properties that we actually need.

Diagonal inner product

Let $\mathcal{M}$ be a space of (non-uniform) sequences. Define

$$\mathcal{X}_2^{\mathcal{M}} = \mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M}).$$

An operator $U \in \mathcal{X}_2^{\mathcal{M}}$ consists of rows $U_i = \pi_i U \in \ell_2^{\mathcal{M}}$ such that $U = \sum_i \pi_i^* U_i$. $\mathcal{X}_2^{\mathcal{M}}$ is the direct orthogonal sum of its subspaces $\pi_i^* \pi_i \mathcal{X}_2^{\mathcal{M}} = \cdots 0 \times 0 \times \ell_2^{\mathcal{M}} \times 0 \times \cdots$, which are isomorphic to $\ell_2^{\mathcal{M}}$, and the rows of $\mathcal{X}_2^{\mathcal{M}}$ act as independent input sequences to T. Consequently, the norm of an operator on ℓ_2 is equal to

$$\|T\| = \sup_{U \in \mathcal{X}_2^{\mathcal{M}}} \frac{\|UT\|_{HS}}{\|U\|_{HS}}$$

In the space $\mathcal{X}_2^{\mathcal{M}}$, we define the *diagonal inner product* as [2][8]

$$\{A, B\} := \mathbf{P}_0(AB^*) \qquad (A, B \in \mathcal{X}_2^{\mathcal{M}}). \tag{2.26}$$

Some properties are that $\{A, B\} \in \mathcal{D}_2(\mathcal{M}, \mathcal{M})$, and that $\langle A, B\rangle_{HS} = \text{trace}\{A, B\}$. The i-th entry of $\{A, B\}$ on the diagonal is equal to the ordinary inner product of ℓ_2-sequences (A_i, B_i):

$$\{A, B\} = \text{diag}[\,(A_i, B_i)\,]_{-\infty}^{\infty},$$

[8]The diagonal inner product does not evaluate to a scalar and hence it is not an inner product in the usual Hilbert space theory, but rather in a Hilbert space module sense.

where $A_i = \pi_i A$ and $B_i = \pi_i B$ are the i-th rows of A and B, respectively. In particular, we have that

$$A = 0 \quad \Leftrightarrow \quad \langle A, A\rangle_{HS} = 0 \quad \Leftrightarrow \quad \{A, A\} = 0, \tag{2.27}$$

$$\langle DA, B\rangle_{HS} = 0 \quad (\text{all } D \in \mathcal{D}) \quad \Leftrightarrow \quad \{A, B\} = 0. \tag{2.28}$$

The observation that the diagonal inner product does not render a single number but rather a diagonal of inner products is useful in the determination of projections onto subspaces, discussed later in this section.

Positive and contractive operators

A Hermitian operator A in $\mathcal{X}(\mathcal{M}, \mathcal{M})$ is *positive*, $A \geq 0$, if for all $u \in \ell_2^{\mathcal{M}}$,

$$(uA, u) \geq 0.$$

We say that A is *strictly positive*,[9] notation $A \gg 0$, if there is an $\varepsilon > 0$ such that, for all u in $\ell_2^{\mathcal{M}}$,

$$(uA, u) \geq \varepsilon(u, u).$$

It is known that a positive operator $A \in \mathcal{X}$ is strictly positive if and only if A is boundedly invertible in $\mathcal{X}$. The above definitions can be formulated in terms of the diagonal inner product, as follows.

PROPOSITION 2.3. *Let $A \in \mathcal{X}(\mathcal{M}, \mathcal{M})$ be a bounded Hermitian operator.*

$$\begin{aligned} A \geq 0 \quad &\Leftrightarrow \quad \text{for all } U \in \mathcal{X}_2^{\mathcal{M}} : \quad \{UA, U\} \geq 0, \\ A \gg 0 \quad &\Leftrightarrow \quad \exists\, \varepsilon > 0 : \quad \text{for all } U \in \mathcal{X}_2^{\mathcal{M}} : \{UA, U\} \geq \varepsilon\{U, U\}. \end{aligned}$$

PROOF This is an immediate consequence of the fact that a Hermitian diagonal operator is positive if and only if all its diagonal entries are positive, and the fact that the diagonal inner product is a diagonal of ordinary inner products: $\{UA, U\} = \operatorname{diag}[(U_iA, U_i)]_{-\infty}^{\infty}$, where $U_i = \pi_i U$ is the i-th row of U. Thus

$$\begin{aligned} \{UA, U\} \geq 0 \quad &\Leftrightarrow \quad (U_iA, U_i) \geq 0 \qquad (\text{all } i) \\ &\Leftrightarrow \quad \langle UA, U\rangle_{HS} = \textstyle\sum_i (U_iA, U_i) \geq 0. \end{aligned}$$

The same reasoning applies to the second part of the proposition. □

Let T be an operator in $\mathcal{X}(\mathcal{M}, \mathcal{N})$. T is said to be contractive if $y = uT \Rightarrow \|y\| \leq \|u\|$, that is, if $(uT, uT) \leq (u, u)$ for all $u \in \ell_2^{\mathcal{M}}$. T is strictly contractive if there is $\varepsilon > 0$ such

[9] More precisely, uniformly strictly positive.

that $(uT, uT) \leq (1-\varepsilon)(u,u)$ for all $u \in \ell_2^{\mathcal{M}}$. Hence T is contractive, respectively strictly contractive, if

$$I - TT^* \geq 0, \qquad \text{resp.} \qquad I - TT^* \gg 0.$$

In the latter case, $I-TT^*$ is boundedly invertible. Because of the identity $I+T^*(I-TT^*)^{-1}T = (I-T^*T)^{-1}$ it is clear that $I-TT^* \gg 0$ implies that $I-T^*T \gg 0$ also.

Left D-invariant subspaces

Consider the space $\mathcal{X}_2^{\mathcal{M}}$. Since it is a Hilbert space in the Hilbert-Schmidt inner product, we can talk about subspaces of $\mathcal{X}_2^{\mathcal{M}}$: closed linear manifolds of a subset of elements of $\mathcal{X}_2^{\mathcal{M}}$. Here, we are interested in subspaces that satisfy the additional property of *left D invariance*: a subspace (or linear manifold) $\mathcal{H}$ in $\mathcal{X}_2^{\mathcal{M}}$ is said to be left D invariant if $F \in \mathcal{H} \Rightarrow DF \in \mathcal{H}$ for any diagonal $D \in \mathcal{D}(\mathbf{C}^{\cdot 1}, \mathbf{C}^{\cdot 1})$, *i.e.*,

$$\mathcal{D}\mathcal{H} \subset \mathcal{H}.$$

A left D-invariant subspace has the property that it naturally falls apart into independent *slices*: in the same sense as we wrote $\mathcal{X}_2^{\mathcal{M}} = \cdots \ell_2^{\mathcal{M}} \times \ell_2^{\mathcal{M}} \times \cdots$ earlier, we can write

$$\mathcal{H} = \cdots \times \mathcal{H}_0 \times \mathcal{H}_1 \times \cdots \tag{2.29}$$

where each $\mathcal{H}_i = \pi_i \mathcal{H}$ is an ordinary subspace in $\ell_2^{\mathcal{M}}$. This is because if $F \in \mathcal{H}$, then $DF \in \mathcal{H}$. By taking D equal to $[D_i = 1, D_k = 0\ (k \neq i)]$, that is, $D = \pi_i^* \pi_i \in \mathcal{D}$, it follows that $\pi_i^* \pi_i F = \pi_i^* F_i \in \mathcal{H}$, so that $F \in \mathcal{H} \Rightarrow F_i \in \mathcal{H}_i$. Because the $\mathcal{H}_i$ are uncoupled, it is seen that the D-invariance property inhibits subspaces $\mathcal{H}$ to have a 'vertical' structure, in which one row F_i gives conditions on other rows of F. A closely related alternative to the description (2.29) is provided by the following lemma:

LEMMA 2.4. *Let $\mathcal{H}$ be a left D-invariant subspace, and let $\mathcal{H}_i = \pi_i \mathcal{H}$. The spaces $\pi_i^* \mathcal{H}_i$ are subspaces of $\mathcal{H}$ which are pairwise orthogonal and together span $\mathcal{H}$:*

$$\mathcal{H} = \cdots \oplus \pi_0^* \mathcal{H}_0 \oplus \pi_1^* \mathcal{H}_1 \oplus \cdots.$$

PROOF Note that an element of $\pi_i^* \mathcal{H}_i$ has all its rows equal to zero, except possibly for the i-th row. $\pi_i^* \mathcal{H}_i$ is a subspace of $\mathcal{H}$ because $\pi_i^* \mathcal{H}_i = \pi_i^* \pi_i \mathcal{H} = D\mathcal{H} \subset \mathcal{H}$, where $D = \pi_i^* \pi_i \in \mathcal{D}$. $\pi_i^* \mathcal{H}_i$ is orthogonal to $\pi_j^* \mathcal{H}_j$ if $i \neq j$ because, for $F_i \in \mathcal{H}_i, F_j \in \mathcal{H}_j$, we have that $\langle \pi_i^* F_i, \pi_j^* F_j \rangle_{HS} = \text{trace } \pi_i^* (F_i, F_j) \pi_j$ and trace $\pi_i^* \pi_j = 0$ $(i \neq j)$. The collection $\{\pi_i^* \mathcal{H}_i\}$ span $\mathcal{H}$ because $\sum_i \pi_i^* \pi_i = I$. □

Let $\mathcal{H}$ be a left D invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$. Each of its slices $\mathcal{H}_i$ is a subspace in the Hilbert space $\ell_2^{\mathcal{M}}$. Let N_i be the dimension of the subspace $\mathcal{H}_i$. If each of these dimensions is finite then we say that $\mathcal{H}$ is of *locally finite dimension*. Note that the dimension of $\mathcal{H}$

is equal to the sum of all N_i, and hence $\mathcal{H}$ can still be an infinite dimensional subspace in $\mathcal{X}_2^{\mathcal{M}}$. The index sequence $N = [N_i]_{-\infty}^{\infty}$ is called the (left) dimension sequence of the left D-invariant subspace $\mathcal{H}$, and we write

$$N = \text{s-dim}(\mathcal{H}).$$

The orthogonal complement of a subspace $\mathcal{H}$ in $\mathcal{X}_2^{\mathcal{M}}$ is

$$\mathcal{H}^{\perp} = \{F \in \mathcal{X}_2 : \langle F, G\rangle_{HS} = 0, \text{ all } G \in \mathcal{H}\}.$$

Since $\mathcal{X}_2^{\mathcal{M}}$ is a Hilbert space, $\mathcal{H}^{\perp}$ is a subspace, and $\mathcal{H} \oplus \mathcal{H}^{\perp} = \mathcal{X}_2^{\mathcal{M}}$.

PROPOSITION 2.5. *If $\mathcal{H}$ is a left D invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$, then $\mathcal{H}^{\perp}$ is left D invariant too, and*

$$\mathcal{H}^{\perp} = \{F \in \mathcal{X}_2^{\mathcal{M}} : \{F, G\} = 0, \text{ all } G \in \mathcal{H}\}.$$

PROOF A straightforward proof uses (2.28) twice. Let $F \in \mathcal{H}^{\perp}$, $G \in \mathcal{H}$, then the D-invariance property of $\mathcal{H}$ implies

$$\langle F, DG\rangle_{HS} = 0 \quad (\text{all } D \in \mathcal{D}) \quad \Leftrightarrow \quad \{F, G\} = 0 \quad \Leftrightarrow \quad \langle DF, G\rangle_{HS} = 0 \quad (\text{all } D \in \mathcal{D}),$$

so that $DF \in \mathcal{H}^{\perp}$. □

Consequently, $\mathcal{H}^{\perp}$ also falls apart into subspaces $(\mathcal{H}^{\perp})_i$, and it is easy to show that $(\mathcal{H}^{\perp})_i = (\mathcal{H}_i)^{\perp}$, so that the orthogonal complement of a left D-invariant subspace $\mathcal{H}$ consists of the complement of its slices $\mathcal{H}_i$.

We list some more properties of D-invariant subspaces. If $\mathcal{A} \subset \mathcal{X}_2$ is a left D-invariant subspace, then so is $\mathcal{A}^{\perp}$. If $\mathcal{A}$ and $\mathcal{B}$ are left D-invariant subspaces, then so are $\mathbf{P}_{\mathcal{A}}(\mathcal{B})$ and $\mathbf{P}_{\mathcal{A}^{\perp}}(\mathcal{B})$, the projections of $\mathcal{B}$ onto $\mathcal{A}$ and $\mathcal{A}^{\perp}$, respectively. If $\mathcal{A}$ or $\mathcal{B}$ is locally finite, then so is $\mathbf{P}_{\mathcal{A}}(\mathcal{B})$.

If two linearly independent subspaces $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{X}_2^{\mathcal{M}}$ are locally finite, then so is their direct sum $\mathcal{A} + \mathcal{B}$. If $\mathcal{A}$ is a left D-invariant subspace and $B \in \mathcal{X}$ is a bounded linear operator, then $\overline{\mathcal{A}B}$ is also a left D-invariant subspace, with

$$\text{s-dim}(\overline{\mathcal{A}B}) \leq \text{s-dim}(\mathcal{A}). \tag{2.30}$$

Bases of locally finite subspaces

Let $\mathcal{H}$ be a left D-invariant subspace of $\mathcal{X}_2^{\mathcal{M}}$. Since $\mathcal{X}_2^{\mathcal{M}}$ is separable in the Hilbert-Schmidt metric, $\mathcal{H}$ has an orthonormal basis. We have seen that $\mathcal{H}$ falls apart into slices $\mathcal{H}_i = \pi_i\mathcal{H}$, which are subspaces in $\ell_2^{\mathcal{M}}$. If each of these subspaces has finite dimension (N_i, say), then we have called $\mathcal{H}$ locally finite. We consider basis representations for such subspaces in this section.

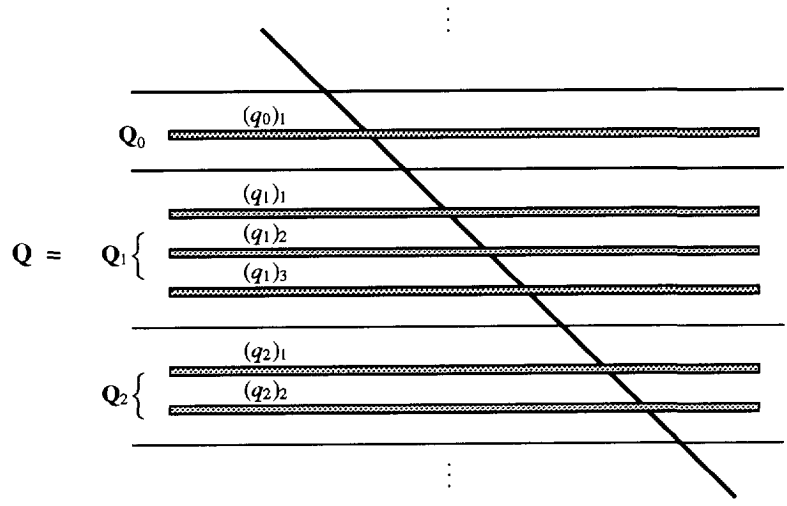


Figure 2.2. Basis representation $\mathbf{Q}$ of some subspace in $\mathcal{X}_2$.

Let $\mathcal{H}_i$ have an orthonormal basis $\{(q_i)_1, \cdots, (q_i)_{N_i}\}$, with each $(q_i)_j \in \ell_2^{\mathcal{M}}$. Because of lemma 2.4, an orthonormal basis of $\mathcal{H}$ is the set $\{\pi_i^*(q_i)_j\}$ $(j = 1, \cdots, N_i,\ i = -\infty, \cdots, \infty)$. It is notationally convenient to collect the set of $(q_i)_j$ into one operator $\mathbf{Q}$. This is done in two steps.

- Stack $\{(q_i)_j\}_{j=1..N_i}$ into one operator $\mathbf{Q}_i \in [\mathbb{C}^{N_i} \to \ell_2^{\mathcal{M}}]$. Note that $\Lambda_i = \mathbf{Q}_i\mathbf{Q}_i^*$ is the Gram matrix of the basis of $\mathcal{H}_i$. In the current situation, the basis is orthonormal and $\Lambda_i = I$. $\mathcal{H}_i$ is generated by $\mathbf{Q}_i$ in the sense that $\mathcal{H}_i = \mathbb{C}^{N_i}\mathbf{Q}_i$: it consists of all linear combinations of the $(q_i)_j$.

- Stack the $\mathbf{Q}_i$ into one operator

$$\mathbf{Q} = \sum_i \pi_i^* \mathbf{Q}_i \tag{2.31}$$

  with rows $\pi_i\mathbf{Q} = \mathbf{Q}_i$. See figure 2.2.

We call $\mathbf{Q}$ an (orthonormal) basis representation of the given basis of $\mathcal{H}$. A number of properties of such a basis operator are listed below.

PROPOSITION 2.6. *Let $\mathcal{H}$ be a locally finite D-invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$, with* s-dim$(\mathcal{H}) = N$, *and let $\mathbf{Q}$ be an orthonormal basis representation for $\mathcal{H}$. Let $\mathcal{N} = \mathbb{C}^N$. Then any $F \in \mathcal{H}$*

can be uniquely written as

$$F = D_F \mathbf{Q},$$

for a certain $D_F \in \mathcal{D}_2^{\mathcal{N}}$. *In particular,* $\mathbf{Q}$ *is bounded on* $\mathcal{D}_2^{\mathcal{N}}$ *and generates* $\mathcal{H}$ *via*

$$\mathcal{H} = \mathcal{D}_2^{\mathcal{N}} \mathbf{Q}.$$

PROOF Let us start from the orthonormal basis $\{(q_i)_1, \cdots, (q_i)_{N_i}\}$ of each $\mathcal{H}_i$. Because $\{\pi_i^*(q_i)_j\}$ $(j = 1, \cdots, N_i,\ i = -\infty, \cdots, \infty)$ is a basis of $\mathcal{H}$, any $F \in \mathcal{H}$ can be written as the linear combination of the basis sequences

$$F = \sum_{i,j} (\alpha_i)_j \cdot \pi_i^*(q_i)_j, \tag{2.32}$$

where the coefficients $(\alpha_i)_j$ are uniquely determined by F and $\sum_{ij} |(\alpha_i)_j|^2 = \| F \|_{HS}^2 < \infty$. Using $\mathbf{Q}_i$, equation (2.32) becomes

$$F = \sum_i \alpha_i \cdot \pi_i^* \mathbf{Q}_i, \tag{2.33}$$

where $\alpha_i = [(\alpha_i)_1, \cdots, (\alpha_i)_{N_i}] \in \mathbb{C}^{1 \times N_i}$ satisfies $\sum_i \| \alpha_i \|_2^2 < \infty$. In terms of $\mathbf{Q}$, equation (2.33) in turn becomes

$$F = D_F \mathbf{Q}, \qquad D_F = \mathrm{diag}[\alpha_i]_{-\infty}^{\infty} \in \mathcal{D}_2^{\mathcal{N}}, \tag{2.34}$$

so that $\mathcal{H} = \mathcal{D}_2^{\mathcal{N}} \mathbf{Q}$. The expression $\mathcal{H} = \mathcal{D}_2^{\mathcal{N}} \mathbf{Q}$ shows that $\mathbf{Q}$ is bounded as a $[\mathcal{D}_2 \to \mathcal{X}_2]$ operator. □

$\mathbf{Q}$ can be viewed as an operator mapping $\mathcal{X}_2^{\mathcal{N}}$ to $\mathcal{X}_2^{\mathcal{M}}$, but then it need not be a bounded operator. A simple example of an unbounded $\mathbf{Q}$ is obtained by taking $\mathbf{Q}_i = [\cdots\ 0\ \boxed{1}\ 0\ \cdots]$ (all i), and recalling that an ℓ_2-sequence need not be summable. Although it is usually enough to consider $\mathbf{Q}$ with domain restricted to $\mathcal{D}_2$, sometimes we need properties which seem to involve a more general domain, and we derive these properties below. (A reader not interested in these details can continue with proposition 2.6.)

To start, note that along with $\mathbf{Q}$, operators $D\mathbf{Q}$ and $\mathbf{Q}X$ $(D \in \mathcal{D}, X \in \mathcal{X})$ are also bounded $[\mathcal{D}_2 \to \mathcal{X}_2]$ operators since $D\mathcal{D}_2 \in \mathcal{D}_2$, $\mathcal{X}_2 X \in \mathcal{X}_2$. The domain of definition of $\mathbf{Q}$ can be extended: for example, the operator $Z\mathbf{Q}$ is also a bounded $[\mathcal{D}_2 \to \mathcal{X}_2]$ operator, and is consistently defined via

$$D(Z\mathbf{Q}) = Z(D^{(1)}\mathbf{Q}). \tag{2.35}$$

$\mathbf{Q}$ is also bounded on all finite sums of terms of the type $DZ^{[k]}\mathbf{Q}$. The result is that $\mathbf{Q}$ is densely defined on $\mathbf{X}_2^{\mathcal{N}}$.

We have defined, in equation (2.20), the operator $\mathbf{P}_0$ on $\mathcal{X}_2$ as the projection onto $\mathcal{D}_2$. We have already extended $\mathbf{P}_0$ to operators in $\mathcal{X}$: $\mathbf{P}_0(X) = \mathrm{diag}[X_{ii}] \in \mathcal{D}$, where $X_{ii} = \pi_i X \pi_i^*$

is bounded for each i. $\mathbf{P}_0$ can also be extended to unbounded operators that are bounded as $[\mathcal{D}_2 \to \mathcal{X}_2]$ operators: because $\pi_i^*\pi_i \in \mathcal{D}_2$ and hence $\pi_i^*\pi_i\mathbf{Q} \in \mathcal{X}_2$, $\mathbf{Q}_{ii} = \pi_i\mathbf{Q}\pi_i^* = \pi_i(\pi_i^*\pi_i\mathbf{Q})\pi_i^*$ is bounded for each i. Thus $\mathbf{P}_0(\mathbf{Q}) = \mathrm{diag}(\mathbf{Q}_{ii})$ is well defined and bounded: $\mathbf{P}_0(\mathbf{Q}) \in \mathcal{D}$. The extension satisfies the usual 'homogeneity' rule for $\mathbf{P}_0$: if $D_{1,2} \in \mathcal{D}$, then $\mathbf{P}_0(D_1\mathbf{Q}D_2) = D_1\mathbf{P}_0(\mathbf{Q})D_2$.

PROPOSITION 2.7. *Let $\mathbf{Q}$ be an operator densely defined on $\mathcal{X}_2^{\mathcal{N}}$ which is bounded as a $[\mathcal{D}_2^{\mathcal{N}} \to \mathcal{X}_2^{\mathcal{M}}]$ operator. $\mathbf{Q}$ has a (not necessarily bounded) adjoint $\mathbf{Q}^*$ acting from $\mathcal{X}_2^{\mathcal{M}}$ to $\mathcal{X}_2^{\mathcal{N}}$.*

The operator $\mathbf{P}_0(\cdot\,\mathbf{Q}^*)$ is a bounded operator in $\mathcal{X}_2$, and it is the adjoint of the operator $\mathbf{Q}$ restricted to $\mathcal{D}_2$.

PROOF Because $\mathbf{Q}$ is densely defined on $\mathcal{X}_2^{\mathcal{N}}$, it has, according to [12, §44], an adjoint operator $\mathbf{Q}^*$ acting from $\mathcal{X}_2^{\mathcal{M}}$ to $\mathcal{X}_2^{\mathcal{N}}$.

Let $\mathrm{dom}(\mathbf{Q})$ be the domain of $\mathbf{Q}$ in $\mathcal{X}_2^{\mathcal{N}}$. The domain of $\mathbf{Q}^*$ consists of all elements $G \in \mathcal{X}_2^{\mathcal{M}}$ for which there is a $G' \in \mathcal{X}_2^{\mathcal{N}}$ such that

$$\langle F\mathbf{Q}, G\rangle_{HS} = \langle F, G'\rangle_{HS}, \qquad G' = G\mathbf{Q}^* \tag{2.36}$$

for every $F \in \mathrm{dom}(\mathbf{Q})$. The existence of $\mathbf{Q}^*$ implies: if $F \in \mathrm{dom}(\mathbf{Q})$ then for all $G \in \mathrm{dom}(\mathbf{Q}^*)$ there exists a $G' = G\mathbf{Q}^* \in \mathcal{X}_2^{\mathcal{N}}$ such that (2.36) holds, and G' is unique. Restricting F to $\mathcal{D}_2 \subset \mathrm{dom}(\mathbf{Q})$ in which it is a bounded operator, it follows for all $F \in \mathcal{D}_2^{\mathcal{N}}$ that

$$\langle F, G'\rangle_{HS} = \langle F, \mathbf{P}_0(G')\rangle_{HS}$$

and $\langle F\mathbf{Q}, G\rangle_{HS} = \langle F, \mathbf{P}_0(G\mathbf{Q}^*)\rangle_{HS}$. Hence $\mathbf{P}_0(\cdot\,\mathbf{Q}^*)$ is the adjoint operator of [$\mathbf{Q}$ restricted to $\mathcal{D}_2^{\mathcal{N}}$]. Since the latter operator is bounded, its adjoint is a bounded $[\mathcal{X}_2^{\mathcal{M}} \to \mathcal{D}_2^{\mathcal{N}}]$ operator. □

As a corollary, $\mathbf{P}_0(\cdot\,\mathbf{Q}\mathbf{Q}^*)$ is a bounded $[\mathcal{D}_2^{\mathcal{N}} \to \mathcal{D}_2^{\mathcal{N}}]$ operator, hence

$$\Lambda_{\mathbf{Q}} := \mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*) \in \mathcal{D}(\mathcal{N}, \mathcal{N})$$

is well defined by the extension of the domain of $\mathbf{P}_0$ discussed earlier. The operator $\Lambda_{\mathbf{Q}}$ is the Gram operator of the basis $\{(\pi_i^* q_i)_j\}$ of $\mathcal{H}$. It is a diagonal operator whose entries $\Lambda_i = \mathbf{Q}_i\mathbf{Q}_i^*$ contain the Gram matrices of the bases of the subspaces $\mathcal{H}_i$ of $\mathcal{H}$. Because these bases have been chosen orthonormal, $\Lambda_{\mathbf{Q}} = I$.

Finally, using the definition (2.35), the adjoint of $\cdot\,(Z\mathbf{Q})$ restricted to $\mathcal{D}_2$ is formally equal to $\mathbf{P}_0(Z^{-1}\cdot\,\mathbf{Q}^*)^{(-1)}$: let $D \in \mathcal{D}_2$, $X \in \mathcal{X}_2$, then

$$\begin{aligned} \{DZ\mathbf{Q}, X\} &= \{ZD^{(1)}\mathbf{Q}, X\} \\ &= \{D^{(1)}\mathbf{Q}, Z^{-1}X\}^{(-1)} \end{aligned} \tag{2.37}$$

$$= \{D^{(1)}, \mathbf{P}_0(Z^{-1}X\mathbf{Q}^*)\}^{(-1)}$$
$$= \{D, \mathbf{P}_0(Z^{-1}X\mathbf{Q}^*)^{(-1)}\}.$$

The computing rules on unbounded basis operators introduced so far are sufficient for our purposes. The importance of such basis representations is illustrated by the following proposition.

PROPOSITION 2.8. *Let $\mathcal{H}$ be a locally finite D-invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$, and let $\mathbf{Q}$ be a basis representation of $\mathcal{H}$. Let $F \in \mathcal{H}$, then*

$$F = \mathbf{P}_0(F\mathbf{Q}^*)\,\mathbf{Q}.$$

PROOF Let $N = \text{s-dim}\,\mathcal{H}$ and $\mathcal{N} = \mathbb{C}^N$. According to (2.34), any element F of $\mathcal{H}$ has a representation $F = D_F\mathbf{Q}$ in terms of $\mathbf{Q}$, where $D_F \in \mathcal{D}_2^{\mathcal{N}}$. The diagonal of coefficients D_F is obtained as

$$D_F = \mathbf{P}_0(F\mathbf{Q}^*).$$

Since $F \in \mathcal{X}_2^{\mathcal{M}}$, we have indeed that $D_F \in \mathcal{D}_2^{\mathcal{N}}$. □

Non-orthogonal bases of locally finite subspaces

The preceding discussion can be generalized to non-orthonormal bases. Again, let $\mathcal{H}$ be a locally finite left D-invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$, falling apart into subspaces $\mathcal{H}_i = \pi_i\mathcal{H}$ with finite dimensions N_i. For each i, let $\{(f_i)_1, \cdots, (f_i)_{N_i}\}$ be a Riesz basis of $\mathcal{H}_i$, *i.e.*, a complete system of vectors whose Gram matrix $\Lambda_i = [((f_i)_j, (f_i)_k)]_{j,k=1}^{N_i}$ is bounded and boundedly invertible. The total collection $\{\pi_i^*(f_i)_j\}$ ($j = 1, \cdots, N_i$, all i) is a Riesz basis of $\mathcal{H}$ if the corresponding Gram operator Λ is bounded and boundedly invertible. The latter condition is equivalent to demanding that Λ be strictly positive: $\Lambda \gg 0$. We call such a basis a *strong basis*. For a strong basis, we can in the same way as for an orthonormal basis construct operators $\mathbf{F}_i$ and stack these into an operator $\mathbf{F}$. We obtain the same results: $\mathbf{F}$ generates $\mathcal{H}$ via

$$\mathcal{H} = \mathcal{D}_2^{\mathcal{N}}\mathbf{F},$$

it is in general an unbounded operator, densely defined on $\mathcal{X}_2^{\mathcal{N}}$, but it is bounded as a $[\mathcal{D}_2^{\mathcal{N}} \to \mathcal{X}_2^{\mathcal{M}}]$ operator, and its adjoint $\mathbf{F}^*$ exists in $\mathcal{X}_2$, but is also unbounded in general. The operator $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$: $\mathcal{X}_2^{\mathcal{M}} \to \mathcal{D}_2^{\mathcal{N}}$ is well defined and bounded, and is the adjoint of $\mathbf{F}$ with domain restricted to $\mathcal{D}_2^{\mathcal{N}}$. Consequently, the operator $\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*)$ is in $\mathcal{D}(\mathcal{N}, \mathcal{N})$, and in fact equal to the Gram operator Λ of the chosen basis:

$$\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) = \text{diag}[\Lambda_i].$$

If $\mathbf{Q}$ is an orthonormal basis representation of $\mathcal{H}$, then $\mathbf{F}$ can be expressed in terms of $\mathbf{Q}$:

$$\mathbf{F} = R\mathbf{Q}, \qquad R \in \mathcal{D}(\mathcal{N}, \mathcal{N}),$$

where R is given explicitly as $R = \mathbf{P}_0(\mathbf{F}\mathbf{Q}^*)$.

If $\mathbf{F}$ is a given strong basis representation, then it can be orthonormalized by factoring $\Lambda_\mathbf{F}$ into invertible factors R as

$$\Lambda_\mathbf{F} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) =: RR^* .$$

Since $\Lambda_\mathbf{F} \gg 0$, this is always possible. The orthonormal basis representation $\mathbf{Q}$ is given by $\mathbf{Q} = R^{-1}\mathbf{F}$; indeed

$$\Lambda_\mathbf{Q} = \mathbf{P}_0(R^{-1}\mathbf{F}\mathbf{F}^*R^{-*}) = R^{-1}\mathbf{P}_0(\mathbf{F}\mathbf{F}^*)R^{-*} = I.$$

Orthogonal projection onto subspaces

Having investigated left D-invariant subspaces and their basis representations, we turn our attention to the projection onto subspaces. The following proposition is used.

PROPOSITION 2.9. *Let $\mathcal{H}$ be a locally finite left D-invariant subspace in $\mathcal{X}_2^\mathcal{M}$, and let $\mathbf{Q}$ be an orthonormal basis representation of $\mathcal{H}$, then (for $X \in \mathcal{X}_2^\mathcal{M}$),*

$$X \perp \mathcal{H} \qquad \Leftrightarrow \qquad \mathbf{P}_0(X\mathbf{Q}^*) = 0.$$

PROOF Any Y in $\mathcal{H}$ can be written as $Y = D\mathbf{Q}$, for some $D \in \mathcal{D}_2$. Then $X \perp Y \Leftrightarrow \{X, Y\} = \mathbf{P}_0(XY^*) = 0$, and $\mathbf{P}_0(XY^*) = \mathbf{P}_0(X\mathbf{Q}^*D^*) = \mathbf{P}_0(X\mathbf{Q}^*)D^*$. Letting Y range over $\mathcal{H}$, this expression is $= 0$ for all D in $\mathcal{D}_2$, and it follows that $\mathbf{P}_0(X\mathbf{Q}^*) = 0$. □

Let $\mathcal{H}$ be a subspace in $\mathcal{X}_2^\mathcal{M}$. Then $\mathcal{X}_2^\mathcal{M} = \mathcal{H} \oplus \mathcal{H}^\perp$, so that every $X \in \mathcal{X}_2^\mathcal{M}$ can be written (uniquely) as $X = X_1 + X_2$, where $X_1 \in \mathcal{H}$ and $X_2 \in \mathcal{H}^\perp$. The operator of (orthogonal) projection onto $\mathcal{H}$ is defined as $\mathbf{P}_\mathcal{H}(X) = X_1$.

THEOREM 2.10. *Let $\mathcal{H}$ be a locally finite left D-invariant subspace in $\mathcal{X}_2^\mathcal{M}$, and let $\mathbf{Q}$ be an orthonormal basis representation of $\mathcal{H}$. The orthogonal projection of any $X \in \mathcal{X}_2^\mathcal{M}$ onto $\mathcal{H}$ is given by*

$$\mathbf{P}_\mathcal{H}(X) = \mathbf{P}_0(X\mathbf{Q}^*)\mathbf{Q}. \tag{2.38}$$

PROOF Let $X = X_1 + X_2$, where $X_1 = \mathbf{P}_\mathcal{H}(X) \in \mathcal{H}$ and $X_2 \in \mathcal{H}^\perp$. Then

$$\begin{aligned} \mathbf{P}_0(X\mathbf{Q}^*)\mathbf{Q} &= \mathbf{P}_0((X_1 + X_2)\mathbf{Q}^*)\mathbf{Q} \\ &= \mathbf{P}_0(X_1\mathbf{Q}^*)\mathbf{Q} + \mathbf{P}_0(X_2\mathbf{Q}^*)\mathbf{Q} \\ &= \mathbf{P}_0(X_1\mathbf{Q}^*)\mathbf{Q} && \text{[prop. 2.9]} \\ &= X_1 && \text{[prop. 2.8]} \end{aligned}$$

Hence $\mathbf{P}_\mathcal{H}(X) = \mathbf{P}(X\mathbf{Q}^*)\mathbf{Q}$. □

COROLLARY 2.11. *Let $\mathcal{H}$ be a locally finite left D-invariant subspace in $\mathcal{X}_2^{\mathcal{M}}$, and let $\mathbf{F}$ be a strong basis representation of $\mathcal{H}$. The orthogonal projection of $X \in \mathcal{X}_2^{\mathcal{M}}$ onto $\mathcal{H}$ is given by*

$$\mathbf{P}_{\mathcal{H}}(X) = \mathbf{P}_0(X\mathbf{F}^*)\,\Lambda_{\mathbf{F}}^{-1}\,\mathbf{F}. \tag{2.39}$$

PROOF If $\mathbf{F}$ is a strong basis representation generating $\mathcal{H}$, then $\mathbf{F} = R\mathbf{Q}$, where $\mathbf{Q}$ is an orthonormal basis representation and $R \in \mathcal{D}$ is any boundedly invertible factor of $\Lambda_{\mathbf{F}} = RR^*$. Inserting $\mathbf{Q} = R^{-1}\mathbf{F}$ in (2.38), the result is obtained. □

Equation (2.39) generalizes the projection formula (2.8) to the present diagonal algebra context. As in the classical context, an operator $\mathbf{P}$ defined everywhere on $\mathcal{X}_2$ is an orthogonal projector if and only if it is idempotent and Hermitian: $\mathbf{P}\,\mathbf{P} = \mathbf{P}$, $\mathbf{P}^* = \mathbf{P}$. These properties are readily verified for the definition in (2.38):

$\mathbf{P}_{\mathcal{H}}$ is idempotent since

$$\begin{aligned}\mathbf{P}_{\mathcal{H}}(\mathbf{P}_{\mathcal{H}}(X)) &= \mathbf{P}_0\left(\mathbf{P}_0(X\mathbf{Q}^*)\cdot\mathbf{Q}\mathbf{Q}^*\right)\cdot\mathbf{Q} = \\ &\mathbf{P}_0(X\mathbf{Q}^*)\,\mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*)\cdot\mathbf{Q} = \mathbf{P}_0(X\mathbf{Q}^*)\cdot\mathbf{Q} = \mathbf{P}_{\mathcal{H}}(X).\end{aligned}$$

$\mathbf{P}_{\mathcal{H}}$ is Hermitian if $\{\mathbf{P}_{\mathcal{H}}(A), B\} = \{A, \mathbf{P}_{\mathcal{H}}(B)\}$ for all $A, B \in \mathcal{X}_2$. Expanding the first term yields

$$\{\mathbf{P}_{\mathcal{H}}(A), B\} = \mathbf{P}_0\left(\mathbf{P}_0(A\mathbf{Q}^*)\cdot\mathbf{Q}B^*\right) = \mathbf{P}_0(A\mathbf{Q}^*)\,\mathbf{P}_0(\mathbf{Q}B^*).$$

The second term is equal to

$$\begin{aligned}\{A, \mathbf{P}_{\mathcal{H}}(B)\} &= \mathbf{P}_0(A\left[\mathbf{P}_0(B\mathbf{Q}^*)\cdot\mathbf{Q}\right]^*) \\ &= \mathbf{P}_0\left(A\mathbf{Q}^*\,\mathbf{P}_0(\mathbf{Q}B^*)\right) \\ &= \mathbf{P}_0(A\mathbf{Q}^*)\,\mathbf{P}_0(\mathbf{Q}B^*).\end{aligned}$$

Hence $\mathbf{P}_{\mathcal{H}}$ is Hermitian.

Matrix representations of operators in $\mathcal{X}_2 \to \mathcal{X}_2$

In standard Hilbert space theory, it is known that any bounded operator T mapping a separable Hilbert space $\mathcal{H}_1$ into a separable Hilbert space $\mathcal{H}_2$ admits a matrix representation $[T_{ij}]$. The entries T_{ij} follow by selecting orthonormal bases in $\mathcal{H}_1$ and $\mathcal{H}_2$. For example, if u and y are (scalar) ℓ_2-sequences, and $\{e_k\}$ is the standard orthonormal basis in ℓ_2, then T_{ij} is defined by the inner product $T_{ij} = (e_iT, e_j)$. This definition is such that $y = uT$ can be written as the (strongly converging) summation

$$y_j = \sum_i u_i T_{ij}$$

where $u_i = (u, e_i)$ and $y_j = (y, e_j)$. Indeed,

$$y_j = (y, e_j) = (uT, e_j) = \sum_i u_i(e_iT, e_j) = \sum_i u_i T_{ij}.$$

A similar representation holds for non-uniform sequences. We now consider the case where T is a bounded operator mapping $\mathcal{X}_2$ into $\mathcal{X}_2$. Elements of $\mathcal{X}_2$ have matrix representations, rather than vector representations: they have entries which are doubly indexed. Consequently, a general operator T from $\mathcal{X}_2$ to $\mathcal{X}_2$ does not have a matrix representation $[T_{ij}]$ but rather a *tensor* representation $[T_{fgij}]$ with four indices, corresponding to

$$Y = UT \qquad \Leftrightarrow \qquad Y_{ij} = \sum_{f,g} U_{fg} T_{fgij}$$

Because $\{(f, g, i, j)\}$ is a four-dimensional lattice, $[T_{fgij}]$ is a four-dimensional object which cannot, in general, be represented by a two-dimensional matrix. T has a matrix representation as an $[\ell_2 \to \ell_2]$ operator only in the special case where T is associative in Z, in the sense that $Z^k(UT) = (Z^k U)T$. This is equivalent to the existence of an ℓ_2 operator T' (which we identify with T), such that

$$(\pi_k^* U_k)T = \pi_k^*(U_k T') \tag{2.40}$$

(where $U_k = \pi_k U$ is the k-th row of U). This is in fact the rule used in equation (2.25), and

$$\begin{aligned} Y = UT &= (\textstyle\sum_k \pi_k^* U_k)T \\ &= \textstyle\sum_k (\pi_k^* U_k)T \\ &= \textstyle\sum_k \pi_k^*(U_k T') \qquad \text{[by (2.40)]} \\ &= \textstyle\sum_k \pi_k^* Y_k . \end{aligned}$$

If (2.40) holds, we can use the usual ℓ_2 matrix representation of T' to represent T. However, a general operator in $[\mathcal{X}_2 \to \mathcal{X}_2]$ need not satisfy (2.40), and a typical example is provided by the operator $\mathbf{P}$ defined in (2.20): the projection onto the upper Hilbert-Schmidt space $\mathcal{U}_2$, which is not associative with Z. Before deriving a tensor representation for such operators, we remark that, throughout the thesis, all operators are left D invariant, in the sense that

$$D(UT) = (DU)T, \qquad \text{all } D \in \mathcal{D} .$$

Consequently, in the expression $Y = UT$, the entry Y_{ij} is only dependent on the entries of the i-th row U_i, rather than all entries $\{U_{fg}\}$, so that it suffices to have a representation of T in only three indices. To reduce this number to two, we sometimes use a representation of elements in $\mathcal{X}_2$ as ℓ_2-sequences of *diagonals*. The result is a representation of T as a matrix whose entries are diagonals.

Thus, for $U = \sum Z^{[i]} U_{[i]} \in \mathcal{X}_2$, define the 'vector representation' $\tilde{U}$, which we call its *diagonal expansion*, as

$$\begin{aligned} U &= \cdots + Z^{-1}U_{[-1]} + U_{[0]} + ZU_{[1]} + Z^2 U_{[2]} + \cdots \\ &= \cdots + U_{[-1]}^{(1)} Z^{-1} + U_{[0]} + U_{[1]}^{(-1)} Z + U_{[2]}^{(-2)} Z^2 + \cdots \\ \tilde{U} &= \begin{bmatrix} \cdots & U_{[-1]}^{(1)} & \boxed{U_{[0]}} & U_{[1]}^{(-1)} & U_{[2]}^{(-2)} & \cdots \end{bmatrix} . \end{aligned} \tag{2.41}$$

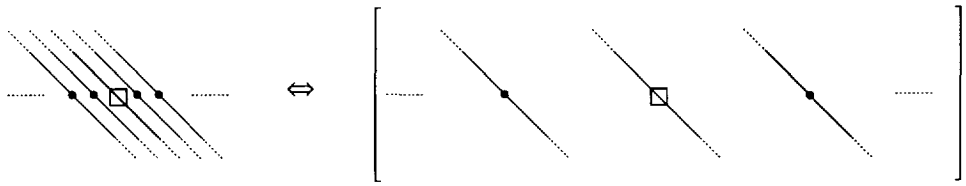


Figure 2.3. Diagonal expansion of an operator in $\mathcal{X}_2$ into a sequence of its diagonals.

See figure 2.3. $\tilde{U}$ is an element of the space $\ell_2(\mathcal{D})$ of square summable sequences whose entries are in $\mathcal{D}$. With the inner product $\langle \tilde{U}, \tilde{Y} \rangle = \text{trace}\left(\sum U_{[k]}^{(-k)} Y_{[k]}^{(-k)*}\right) = \text{trace}(\tilde{U}\tilde{Y}^*)$, this space is isometrically isomorphic to $\mathcal{X}_2$, and with regard to diagonal inner products, we have that

$$\mathbf{P}_0(UY^*) = \tilde{U}\tilde{Y}^* . \tag{2.42}$$

The correspondence between $U \in \mathcal{X}_2$ and $\tilde{U}$ can be viewed as a generalization of the scalar z-transform, which associates to an ℓ_2-sequence a polynomial in z. The generalization is obtained by replacing z by Z and scalars by diagonals. Other generalizations (*i.e.*, other isomorphisms) along these lines are possible, but the definition (2.41) is such that $(DU)^{\sim} = D\tilde{U}$ (where $D \in \mathcal{D}$): it keeps entries U_{ij} that are on the same (i-th) row of U also on the same 'row' of $\tilde{U}$. Here, the i-th row of $\tilde{U}$ is defined to be the sequence obtained by selecting the i-th element of every diagonal in the sequence of $\tilde{U}$. We keep the notation $\cdot^{(n)}$ for the operator on $\ell_2(\mathcal{D})$ that is isomorphic to the same operator in $\mathcal{X}_2$, so that $(U^{(n)})^{\sim} = (\tilde{U})^{(n)}$ is the diagonal shift of the entries in the sequence.

Operators $T \in [\mathcal{X}_2 \to \mathcal{X}_2]$ have corresponding operators $\tilde{T} \in [\ell_2(\mathcal{D}) \to \ell_2(\mathcal{D})]$, where the correspondence is provided by the chosen isomorphism between $\mathcal{X}_2$ and $\ell_2(\mathcal{D})$. If T is a left D-invariant operator, then this property carries over to $\tilde{T}$, because

$$D\tilde{Y} = (DY)^{\sim} = (DU)^{\sim}\tilde{T} = (D\tilde{U})\tilde{T}.$$

This shows that $\tilde{T}$ maps rows of $\tilde{U}$ into rows of $\tilde{Y}$. Consequently, $\tilde{T}$ can be represented by a matrix whose entries are *diagonals:*

$$\tilde{T} = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & \tilde{T}_{-1,-1} & \tilde{T}_{-1,0} & \tilde{T}_{-1,1} & \ddots \\ \ddots & \tilde{T}_{0,-1} & \boxed{\tilde{T}_{0,0}} & \tilde{T}_{0,1} & \ddots \\ \ddots & \tilde{T}_{1,-1} & \tilde{T}_{1,0} & \tilde{T}_{1,1} & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{2.43}$$

where $\tilde{T}_{ij} \in \mathcal{D}(\mathcal{M}, \mathcal{N})$ if $T \in [\mathcal{X}_2^{\mathcal{M}} \to \mathcal{X}_2^{\mathcal{N}}]$. We call $\tilde{T}$ the diagonal expansion of the operator T. $\tilde{T}$ has the same operator norm as T, in the metric induced by the diagonal expansion.

A third representation of T is obtained by constructing a sequence of ℓ_2 operators $\tilde{T}_k$ from $\tilde{T}$. These operators are such that entry (i,j) of the matrix representation of $\tilde{T}_k$ is equal to the k-th entry along the diagonal of $\tilde{T}_{ij}$. We call the $\tilde{T}_k$ the *snapshots* of the operator T, and k has the interpretation of 'point in time' in this representation. The connection between $\tilde{T}$ and $\{\tilde{T}_k\}$ amounts to a simple reordering of a matrix of diagonals into a sequence ('diagonal') of matrices. $\tilde{T}_k$ maps the k-th row $\tilde{U}_k$ of $\tilde{U}$ to the k-th row $\tilde{Y}_k$ of $\tilde{Y}$.

The operators $\{\tilde{T}_k\}$ can be obtained from T directly, without reverting to $\tilde{T}$. With $Y = UT$ and U_k denoting the k-th row of U as before, let the ℓ_2 operator T_k be equal to the mapping of U_k to Y_k, *i.e.*, $Y_k = U_k T_k$. [Note that, because of the left D invariance of T, there is no transfer of U_k to Y_i for $i \neq k$.] T_k satisfies the equation

$$(\pi_k^* U_k)T = \pi_k^*(U_k T_k) \tag{2.44}$$

which is seen to be a generalization of (2.40): the T_k are not all equal to the same T'. The ℓ_2-sequences U_k and $\tilde{U}_k$ are closely related: $U_k = \tilde{U}_k Z^k$, so that U_k is just a shifted version of $\tilde{U}_k$. It follows that $\tilde{T}_k = Z^k T_k Z^{-k} = T_k^{(-k)}$, so that $\tilde{T}_k$ is equal to T_k, modulo a diagonal upward shift over k positions.

To illustrate the above definitions, consider the operator $\mathbf{P}$, the projection of $\mathcal{X}_2$ onto $\mathcal{U}_2$. The diagonal expansion of $\mathbf{P}$ is

$$\tilde{\mathbf{P}} = \begin{bmatrix} \ddots & \ddots & & & \ddots \\ \ddots & 0 & 0 & & \\ & 0 & \boxed{I} & 0 & \\ & & 0 & I & \ddots \\ \ddots & & & \ddots & \ddots \end{bmatrix}$$

and it maps sequences $\tilde{U} = \begin{bmatrix} \cdots & U_{-1}^{(1)} & \boxed{U_{[0]}} & U_{[1]}^{(-1)} & U_{[2]}^{(-2)} & \cdots \end{bmatrix}$ to sequences $\tilde{U}\tilde{\mathbf{P}} = \begin{bmatrix} \cdots & 0 & 0 & \boxed{U_{[0]}} & U_{[1]}^{(-1)} & U_{[2]}^{(-2)} & \cdots \end{bmatrix}$. The snapshots $\tilde{\mathbf{P}}_k$ and $\mathbf{P}_k$ are

$$\tilde{\mathbf{P}}_k = \begin{bmatrix} \ddots & \ddots & & & \ddots \\ \ddots & 0 & 0 & & \\ & 0 & \boxed{1} & 0 & \\ & & 0 & 1 & \ddots \\ \ddots & & & \ddots & \ddots \end{bmatrix} \qquad \mathbf{P}_k = \begin{bmatrix} \ddots & \ddots & & & \ddots \\ \ddots & 0 & 0 & & \\ & 0 & \underline{1} & 0 & \\ & & 0 & 1 & \ddots \\ \ddots & & & \ddots & \ddots \end{bmatrix}$$

where the underlined entry of $\mathbf{P}_k$ is at the (k, k)-th position.

Bibliography

[1] D. Alpay and P. Dewilde, "Time-varying signal approximation and estimation," in *Signal Processing, Scattering and Operator Theory, and Numerical Methods* (M.A. Kaashoek, J.H. van Schuppen, and A.C.M. Ran, eds.), vol. III of *Proc. Int. Symp. MTNS-89*, pp. 1–22, Birkhäuser Verlag, 1990.

[2] D. Alpay, P. Dewilde, and H. Dym, "Lossless Inverse Scattering and reproducing kernels for upper triangular operators," in *Extension and Interpolation of Linear Operators and Matrix Functions* (I. Gohberg, ed.), vol. 47 of *Operator Theory, Advances and Applications*, pp. 61–135, Birkhäuser Verlag, 1990.

[3] A.J. van der Veen and P.M. Dewilde, "Time-varying system theory for computational networks," in *Algorithms and Parallel VLSI Architectures, II* (P. Quinton and Y. Robert, eds.), pp. 103–127, Elsevier, 1991.

[4] P.M. Dewilde and A.J. van der Veen, "On the Hankel-norm approximation of upper-triangular operators and matrices," *to appear in Integral Equations and Operator Theory*, 1993.

[5] P. Dewilde and H. Dym, "Interpolation for upper triangular operators," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 153–260, Birkhäuser Verlag, 1992.

[6] L.A. Zadeh, "Frequency analysis of variable networks," *Proc. IRE*, vol. 38, pp. 291–299, Mar. 1950.

[7] L.A. Zadeh, "Time-varying networks, I," *Proc. IRE*, vol. 49, pp. 1488–1503, Oct. 1961.

[8] A. Feintuch and R. Saeks, *System Theory: A Hilbert Space Approach*. Academic Press, 1982.

[9] E.W. Kamen, P.P. Khargonekar, and K.R. Poolla, "A transfer-function approach to linear time-varying discrete-time systems," *SIAM J. Control and Optimization*, vol. 23, pp. 550–565, July 1985.

[10] P.P. Khargonekar and K. Poolla, "On polynomial matrix fraction representations for linear time-varying discrete-time systems," *Lin. Alg. Appl.*, vol. 80, pp. 1–37, 1986.

[11] K. Poolla and P. Khargonekar, "Stabilizability and stable-proper factorizations for linear time-varying systems," *SIAM J. Control and Optimization*, vol. 25, pp. 723–736, May 1987.

[12] N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space*, vol. I and II. Pitman Publishing Ltd, London, 1981.

[13] P.R. Halmos, *Introduction to Hilbert Space*. Chelsea Publ. Comp., NY, 1951.

[14] P.A. Fuhrmann, *Linear Systems and Operators in Hilbert Space*. McGraw-Hill, 1981.

[15] N. Dunford and J.T. Schwartz, *Linear Operators*, vol. 1, 2. New York: Interscience, 1963.

[16] R.G. Douglas, "On majorization, factorization and range inclusion of operators on Hilbert space," *Proc. Amer. Math. Soc.*, vol. 17, pp. 413–415, 1966.

[17] R.P. Gilbert and G.N. Hile, "Hilbert function modules with reproducing kernels," *Non-linear Analysis, Methods and Applications*, vol. 1, no. 2, pp. 135–150, 1977.

Chapter 3

REALIZATION THEORY

With the notation and preliminary results given in chapter 2, the ground has been prepared to solve the *realization problem*: the problem to determine a (state-space) model that matches a given transfer operator (input-output mapping). In addition to a treatment of this problem, a number of important issues in realization theory pass in review: controllability, Lyapunov equations and state similarity. We consider only systems for which the number of states is finite at any point in time. The resulting state-space theory has aspects of both LTI finite- and infinite-dimensional systems theory.

This chapter has two levels. We start with an introduction of time-varying state-space realizations and some elementary properties, and derive a realization algorithm for finite matrices without actually using the diagonal notations (section 3.2). With this preparation, a mathematically more rigorous approach is taken, fully exploiting chapter 2, to derive reminiscent realization algorithms for more general bounded upper operators (sections 3.3 and 3.4). These are used to prove a Kronecker-type theorem which claims the existence of minimal realizations for certain upper operators, and relates the minimal number of states that are needed at each point in time to the rank of a Hankel operator. The chapter finishes with some examples and computational issues.[1]

3.1 REALIZATIONS OF A TRANSFER OPERATOR

Transfer operator

In chapter 2 we introduced the spaces of bounded non-uniform sequences as signal spaces. Let $\ell_2^{\mathcal{M}}$ and $\ell_2^{\mathcal{N}}$ be two such spaces. The input-output behavior of a linear time-varying system is described by its transfer operator: an operator T mapping signals from $\ell_2^{\mathcal{M}}$ to $\ell_2^{\mathcal{N}}$. Here, $\mathcal{M}$ is the input space of the system, and $\mathcal{N}$ is the output space. In general, the

[1] The material in this chapter was reported in part in [1].

number of inputs and outputs of the system can be time-varying, too: even if we start with the assumption that a system has a constant number of inputs and outputs, factorizations of T in future chapters introduce in a natural way new systems that have time-varying signal spaces. Hence $\mathcal{M}$ and $\mathcal{N}$ need not be 'uniform' space sequences, but can be of the full generality as introduced in chapter 2.

All transfer operators which we consider are assumed to be bounded operators acting on Hilbert spaces: systems that map signals of bounded energy to other signals of bounded energy. Other spaces, such as ℓ_∞, could have been considered as signal spaces [2], but ℓ_2 is mathematically more attractive: many facts in operator theory are simplest for Hilbert spaces, and some facts, such as the existence of an adjoint operator, are only true for such spaces. One could restrict the attention further and consider only compact sequences: signals which are non-zero only on a finite number of instances in time. The argument for doing so is that most of the mathematical complications which are still present in the Hilbert space context disappear, and since such sequences are dense in ℓ_2, the resulting system theory (save for the mathematical details) is closely related to the Hilbert space realization theory. This is the approach taken in the parallel development of a time-varying system theory by Gohberg, Kaashoek and Lerer in [3].

The Hilbert space setting is generalized by considering the space $\mathcal{X}_2^{\mathcal{M}} = \mathcal{X}_2(\mathbf{C}^{\bullet 1}, \mathcal{M})$ as signal space, where $\mathcal{M}$ is some space sequence, and viewing a signal in this space as a collection of $\ell_2^{\mathcal{M}}$-signals whose total energy is bounded. Considering such a collection has notational advantages in the analysis of time-varying systems, as exemplified in the following paragraph. As a relaxation of notation, we will often write just $\mathcal{X}_2$ instead of $\mathcal{X}_2^{\mathcal{M}}$ when the precise form of $\mathcal{M}$ is of no particular interest. We will also often use the subspaces $\mathcal{U}_2$ and $\mathcal{L}_2$ of $\mathcal{X}_2$ as signal spaces: the spaces of upper and lower Hilbert-Schmidt operators as defined in (2.19).

A transfer operator T is said to be causal if it is an upper operator: $T \in \mathcal{U}$. T is causal if and only if

$$U \in \mathcal{U}_2 \qquad \Rightarrow \qquad Y = UT \in \mathcal{U}_2 . \tag{3.1}$$

An expression of causality in terms of ℓ_2-sequences is more elaborate, and we have to define the space of sequences $\ell_2(a, b)$ of sequences whose support lies in the interval $[a, b]$. The transfer operator T is causal if it maps a sequence that is zero before time i into one that is also zero before time i: $u \in \ell_2(i, \infty) \Rightarrow y = uT \in \ell_2(i, \infty)$, for any i. This collection of signals is contained in $\mathcal{U}_2$ in a natural way: take $U = \pi_i^* u$ (*cf.* equation (2.25)). We call $\mathcal{U}_2$ the space of 'future' input or output signals, with respect to the central diagonal. The subspace $\mathcal{L}_2 Z^{-1}$ is complementary to $\mathcal{U}_2$ with respect to $\mathcal{X}_2$ and corresponds to signals lying in the past.

The rows of T can be viewed as the impulse responses of the system. Indeed, in the single-input single-output case, if the unit impulse at time i, ($u \in \ell_2, u_i = 1, u_j = 0\ (j \neq i)$), is the input signal, then $y = uT = [\cdots\ 0\ \ T_{ii}\ \ T_{i,i+1}\ \ T_{i,i+2}\ \cdots]$ is the resulting output, which

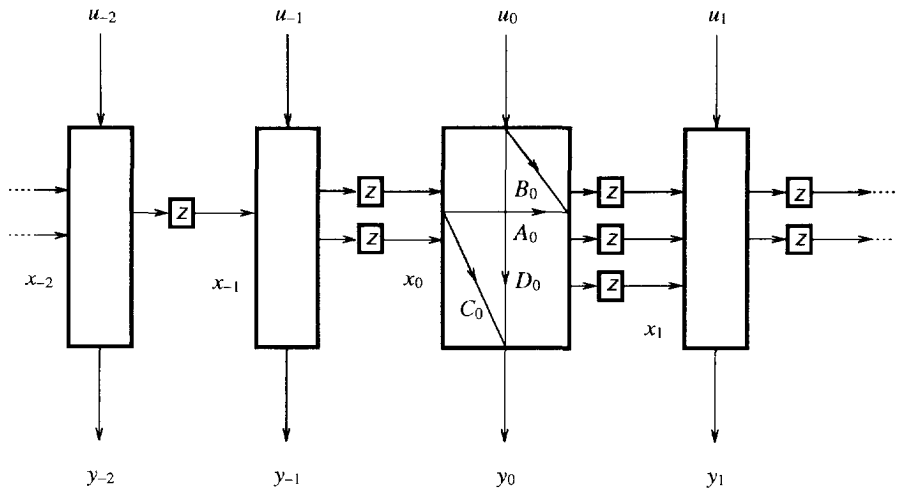


Figure 3.1. Time-varying state realization

is precisely the i-th row of T. An obvious extension holds for general sequences.

Realizations

We are interested in systems that admit a dynamical realization in the form of the state recursion

$$\begin{aligned} x_{k+1} &= x_k A_k + u_k B_k \\ y_k &= x_k C_k + u_k D_k \end{aligned} \qquad \mathbf{T}_k = \begin{bmatrix} A_k & C_k \\ B_k & D_k \end{bmatrix} \tag{3.2}$$

in which we require the matrices $\{A_k, B_k, C_k, D_k\}$ to be uniformly bounded and to have finite (but not necessarily fixed) dimensions. The realization is causal automatically: $y_k = 0$ $(k \le i)$ if $u_k = 0$ $(k \le i)$ with the assumption that $x_{-\infty} = 0$. See figure 3.1.

Realizations of the type (3.2) can be rewritten by assembling the matrices $\{A_k\}$, $\{B_k\}$ etc. into diagonal operators on spaces of sequences of appropriate dimensions:

$$A = \begin{bmatrix} \ddots & & \mathbf{0} \\ & A_k & \\ \mathbf{0} & & \ddots \end{bmatrix} \qquad C = \begin{bmatrix} \ddots & & \mathbf{0} \\ & C_k & \\ \mathbf{0} & & \ddots \end{bmatrix} \tag{3.3}$$

$$B = \begin{bmatrix} \ddots & & \mathbf{0} \\ & B_k & \\ \mathbf{0} & & \ddots \end{bmatrix} \qquad D = \begin{bmatrix} \ddots & & \mathbf{0} \\ & D_k & \\ \mathbf{0} & & \ddots \end{bmatrix}$$

Let $\ell_2^{\mathcal{M}}$ be the space of input sequences, $\ell_2^{\mathcal{N}}$ the space of output sequences, and let us define $\mathcal{B} = \cdots \times \mathcal{B}_0 \times \mathcal{B}_1 \times \cdots$ as the sequence of spaces to which the state belongs. Then

$$\begin{array}{rcccccl} u & = & [\cdots & \boxed{u_0} & u_1 & u_2 \cdots] & \in \ \ell_2^{\mathcal{M}} \\ y & = & [\cdots & \boxed{y_0} & y_1 & y_2 \cdots] & \in \ \ell_2^{\mathcal{N}} \\ x & = & [\cdots & \boxed{x_0} & x_1 & x_2 \cdots] & \in \ \mathcal{B} \\ xZ^{-1} & = & [\cdots & \boxed{x_1} & x_2 & x_3 \cdots] & \in \ \mathcal{B}^{(-1)}. \end{array}$$

A discrete-time causal time-varying linear realization $\mathbf{T}$ consists of the set of four maps

$$\begin{array}{lllllll} A & \in & \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)}), & & C & \in & \mathcal{D}(\mathcal{B}, \mathcal{N}), \\ B & \in & \mathcal{D}(\mathcal{M}, \mathcal{B}^{(-1)}), & & D & \in & \mathcal{D}(\mathcal{M}, \mathcal{N}), \end{array} \tag{3.4}$$

which together represent the dynamical state equations

$$\begin{array}{rcl} xZ^{-1} & = & xA + uB \\ y & = & xC + uD \end{array} \qquad \mathbf{T} = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \tag{3.5}$$

This definition constitutes the same set of time-varying state equations as in (3.2), but now written in an index-free form. The state equations (3.2) are recovered by taking the k-th entry of each sequence and the corresponding k-th entry along the diagonal of each realization matrix. A difference between the equations (3.2) and (3.5) is that the former equations suggest a *recursion* which can be carried out to obtain the next state x_{k+1} and current output y_k from the current state x_k and input u_k, whereas the equations (3.5) are implicit conditions which some sequences u, x and y have to satisfy. In this case, the solutions are more general, as we will see shortly.

If $(I - AZ)$ is boundedly invertible, then x can be eliminated from equations (3.5):

$$\begin{array}{rcl} x & = & uBZ(I - AZ)^{-1} \\ y & = & u\left[D + BZ(I - AZ)^{-1}C\right] \end{array}$$

Hence the transfer operator of which $\mathbf{T}$ is a realization is given by $T = D + BZ(I - AZ)^{-1}C$. It also follows that $BZ(I - AZ)^{-1}$ is a bounded operator, so that x is a bounded signal: $x \in \ell_2^{\mathcal{B}}$. Note the similarity of this expression for the transfer operator T and the familiar expression of the transfer *function* $T(z) = d + bz(1 - az)^{-1}c$ for time-invariant systems with a time-invariant realization $\{a, b, c, d\}$. The difference is that the transfer function is not a function of Z. Formal replacement of Z by a diagonal operator $V \in \mathcal{D}$ leads to the W-transform [4], which is the generalization of the time-invariant z-transform to the present context. This transform is briefly discussed in section 5.3.

However, note that although $(I - AZ) \in \mathcal{U}$, the inverse $(I - AZ)^{-1}$, even if it is bounded, is not necessarily upper, as the example in section 2.2 showed. This means that the transfer operator described by equations (3.5) is not necessarily causal, contrary to the

causal recursion (3.2) from which it was derived. The situation is as follows. According to proposition 2.2, $(I-AZ)$ has an inverse which is upper and given by

$$(I-AZ)^{-1} = I + AZ + AZAZ + \cdots$$

if and only if the spectral radius $\ell_A := r(AZ) < 1$. We say that the realization is strictly stable if $\ell_A < 1$. If this is the case, then $xZ^{-1} = xA + uB$ implies

$$\begin{aligned} x &= uBZ + xAZ \\ &= uBZ + uBZ(AZ) + x(AZ)^2 \\ &= uBZ + uBZ(AZ) + uBZ(AZ)^2 + \cdots \end{aligned} \tag{3.6}$$

which is now convergent for any $u \in \ell_2$ and equal to $x = uBZ(I-AZ)^{-1}$. Hence, the solution of the realization equations (3.5) for a given u is the same as the solution generated by the recursion (3.2), and

$$\begin{aligned} y &= uD + uBZC + uBZAZC + uBZ(AZ)^2C + \cdots \\ &= uD + uZB^{(1)}C + uZ^{[2]}B^{(2)}A^{(1)}C + uZ^{[3]}B^{(3)}A^{\{2\}}C + \cdots . \end{aligned} \tag{3.7}$$

If $\ell_A = 1$, then (3.6) may or may not converge to a sequence x with bounded entries, depending on u and B. We limit our attention to the causal interpretation of (3.5), that is to solutions x and y of inputs u for which (3.6) converges pointwise to a sequence x, which, however, need not necessarily be in ℓ_2. We say that the realization is *bounded* if the entries x_k of x are bounded for all $u \in \ell_2$. This condition is satisfied at any rate if $\ell_A < 1$; for $\ell_A \geq 1$, it depends also on B. The analysis of ℓ_A to characterize strictly stable ($\ell_A < 1$), marginally stable ($\ell_A = 1$) and unstable ($\ell_A > 1$) systems replaces the notion in LTI systems theory of poles (eigenvalues of A) that lie in, on, or outside the unit disc.

The above summarizes down to the following definition. $\mathbf{T}$ is said to be a realization of a transfer operator $T \in \mathcal{U}$ if its diagonals $T_{[k]}$ satisfy (*cf.* equation (3.7)

$$T_{[k]} = \begin{cases} 0, & k < 0, \\ D, & k = 0, \\ B^{(k)}A^{\{k-1\}}C, & k > 0. \end{cases} \tag{3.8}$$

This is equivalent with requiring that the entries T_{ij} of T are given by

$$T_{ij} = \begin{cases} 0, & i > j \\ D_i, & i = j \\ B_iA_{i+1}\cdots A_{j-1}C_j, & i < j \end{cases}$$

and in both cases we obtain

$$T = \begin{bmatrix} \ddots & \vdots & & & \vdots & \\ & D_{-1} & B_{-1}C_0 & B_{-1}A_0C_1 & B_{-1}A_0A_1C_2 & \cdots \\ & & \boxed{D_0} & B_0C_1 & B_0A_1C_2 & \\ & & & D_1 & B_1C_2 & \\ & \mathbf{0} & & & D_2 & \cdots \\ & & & & & \ddots \end{bmatrix} \tag{3.9}$$

If an upper operator has a state realization with state-space sequence $\mathcal{B}$ where each $\mathcal{B}_k$ has finite dimension, then we say that it is a *locally finite* state operator. This is a generalization of the concept of rational operators to time-varying systems. The *order of the realization* is the index sequence $\#(\mathcal{B})$ of the state space $\mathcal{B}$.

The realization (3.5) can be generalized further, by considering generalized inputs U in $\mathcal{X}_2^{\mathcal{M}}$ and outputs Y in $\mathcal{X}_2^{\mathcal{M}}$:

$$\begin{aligned} XZ^{-1} &= XA + UB \\ Y &= XC + UD \end{aligned} \qquad \mathbf{T} = \begin{bmatrix} A & C \\ B & D \end{bmatrix} \tag{3.10}$$

If $\ell_A < 1$, then again $X = UBZ(I - AZ)^{-1}$, so that $X \in \mathcal{X}_2^{\mathcal{B}}$. Realization (3.5) is recovered by selecting rows of U, Y and X. The recursions corresponding to this realization are generalizations of (3.2), and are obtained by selecting the k-th diagonal of U, Y, X in (3.10):

$$\begin{aligned} X_{[k+1]}^{(-1)} &= X_{[k]}A + U_{[k]}B \\ Y_{[k]} &= X_{[k]}C + U_{[k]}D \end{aligned} \tag{3.11}$$

Note the diagonal shift in $X_{[k+1]}^{(-1)}$. The same remarks on the relation between this recursive realization and the equations (3.10) as made on the ℓ_2-realizations are in order here. The realization is bounded if all $X_{[k]}$ are in $\mathcal{D}_2^{\mathcal{B}}$ for any $U \in \mathcal{X}_2^{\mathcal{M}}$.

State transformations

Two realizations $\mathbf{T}$ and $\mathbf{T}'$ are called equivalent if they realize the same transfer operator T, that is, if and only if

$$\begin{aligned} D &= D' \\ B^{(k)}A^{\{k-1\}}C &= B'^{(k)}A'^{\{k-1\}}C' \qquad (\text{all } k \geq 0). \end{aligned} \tag{3.12}$$

If $\{A, B, C, D\}$ is a bounded realization of a system with transfer operator T, then an equivalent bounded realization can be found by applying a state transformation R: $x = x'R$

on the state sequence x of the system, where $R \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ is a bounded and boundedly invertible diagonal operator. The transition operator $\mathbf{T}$ is then transformed to

$$\mathbf{T}' = \begin{bmatrix} R & \\ & I \end{bmatrix} \begin{bmatrix} A & C \\ B & D \end{bmatrix} \begin{bmatrix} [R^{(-1)}]^{-1} & \\ & I \end{bmatrix}. \tag{3.13}$$

(Note the diagonal shift in $[R^{(-1)}]^{-1}$. We write, for shorthand, $R^{-(-1)} = [R^{(-1)}]^{-1}$.) This is readily derived by starting with the given realization

$$\begin{bmatrix} xZ^{-1} &= xA + uB \\ y &= xC + uD \end{bmatrix}$$

and inserting $x = x'R$. Then

$$\begin{bmatrix} x'RZ^{-1} &= x'RA + uB \\ y &= x'RC + uD \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} x'Z^{-1}R^{(-1)} &= x'RA + uB \\ y &= x'RC + uD \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} x'Z^{-1} &= x'RAR^{-(-1)} + uBR^{-(-1)} \\ y &= x'RC + uD \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} x'Z^{-1} &= x'A' + uB' \\ y &= x'C' + uD' \end{bmatrix}$$

$\mathbf{T}$ and $\mathbf{T}'$ indeed realize the same transfer operator T: we have already $D = D'$, and

$$\begin{aligned} B'^{(k)}A'^{\{k-1\}}C' &= B^{(k)}R^{-(k-1)} \cdot R^{(k-1)}A^{\{k-1\}}R^{-(k-2)} \cdot R^{(k-2)}A^{\{k-2\}}R^{-(k-3)} \cdots R^{(1)}A^{(1)}R^{-1} \cdot RC \\ &= B^{(k)}A^{\{k-1\}}C. \end{aligned}$$

Stability and strict stability are properties that are preserved under the transformation:

$$\begin{aligned} \ell_{RAR^{-(-1)}} &= \lim_{n\to\infty} \| (RAR^{-(-1)}Z)^n \|^{1/n} \\ &= \lim_{n\to\infty} \| (RAZR^{-1})^n \|^{1/n} \\ &= \lim_{n\to\infty} \| R(AZ)^n R^{-1} \|^{1/n} \\ &\leq \lim_{n\to\infty} \| R \|^{1/n} \cdot \| (AZ)^n \|^{1/n} \cdot \| R^{-1} \|^{1/n} \\ &= \ell_A \end{aligned} \tag{3.14}$$

since $\| R \|^{1/n} \to 1$ and $\| R^{-1} \|^{1/n} \to 1$. Because $\ell_A \leq \ell_{RAR^{-(-1)}}$ can be proven in the same way, it follows that $\ell_A = \ell_{RAR^{-(-1)}}$.

Subclasses of time-varying systems

Although the main line of this thesis considers (locally finite) bounded upper operators in general, it is often instructive to examine the behavior of certain subclasses of systems.

One of the main reasons is a practical one: it takes an infinite amount of data to describe a general time-varying system, so that computations on this data are also infinite. Here, computations are typically recursions such as in (3.2), another prototype example being

$$\Lambda_{k+1} = A_k^* \Lambda_k A_k + B_k^* B_k, \qquad k = \cdots, -1, 0, 1, \cdots \tag{3.15}$$

(a Lyapunov recursion). Hence, computations are not independent from each other: they depend, at the k-th stage, on data of the previous stages, and we will encounter later also some cases where computations depend on data of later stages.

Finite-size matrices can be embedded in the above definitions in many ways. For example, if the input space sequence $\mathcal{M} = \cdots \times \mathcal{M}_{-1} \times \boxed{\mathcal{M}_0} \times \mathcal{M}_1 \times \cdots$ has $\mathcal{M}_i = \emptyset$ for i outside a finite interval, $[1, n]$ say, and if also the output space sequence $\mathcal{N}$ has $\mathcal{N}_i = \emptyset$ for i outside $[1, n]$, then $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ is an upper triangular $n \times n$ (block) matrix. In this case, the state space $\mathcal{B}$ can be chosen to have zero dimensions outside the index interval $[2, n]$, so that we start and end with zero states. Doing so yields the computational networks in the form described in chapter 1. Finite matrices are an important subclass of the bounded operators, because (*i*) initial conditions are known precisely: we start with zero states, and (*ii*) computations are typically finite, so that boundedness and convergence are not issues. For example, $\ell_A = 0$ because we start and end with zero states, so that $A^{\{i\}} = [\,\cdot\,]$ for $i \geq n$. The Lyapunov recursion (3.15) can be solved for $k > 1$ by starting with initial value $\Lambda_1 = [\,\cdot\,]$.

A second important subclass of time-varying systems are systems which are time-invariant outside a finite time interval, again say $[1, n]$. This class properly contains the finite matrix case. Computations on these systems can typically be split into computations on the time-invariant part, for which classical solutions can be obtained, and computations on a time-varying part, which will typically involve recursions starting from initial values provided by the time-invariant part. Boundedness is usually reduced to a time-invariant issue. For example, ℓ_A is equal to $\max(r(A_{-\infty}), r(A_{\infty}))$ which is governed by the behavior of the time-invariant parts. Turning to our Lyapunov recursion: an initial value for $k = 1$ is obtained from $k = 0$. Since $\Lambda_1 = \Lambda_0$ because of the time-invariance before $k = 1$, the recursion becomes a Lyapunov *equation* for $k \leq 1$,

$$\Lambda_0 = A_0^* \Lambda_0 A_0 + B_0^* B_0 .$$

This equation can be solved analytically using an eigenvalue decomposition (Schur decomposition) of A_0.

A third subclass is the class of periodically varying systems. If a system has a period n, then it can be viewed as a time-invariant system T with block entries $T_{ij} = T_{i-j}$ of size

$n \times n$: T is a block Toeplitz operator. For the Lyapunov recursion, this yields

$$A' = A_1 A_2 \cdots A_n, \qquad B' = \begin{bmatrix} B_1 A_2 A_3 \cdots A_n \\ B_2 A_3 \cdots A_n \\ \vdots \\ B_n \end{bmatrix}$$

where A' and B' are finite matrices (not diagonals). Again, classical time-invariant solutions can be computed for this block system, providing exact initial conditions for recursions within the block. The Lyapunov recursion becomes a finite matrix equation:

$$\Lambda' = A'^* \Lambda' A' + B'^* B' .$$

It follows that $\Lambda_1 = \Lambda'$. From (3.15), $\Lambda_2, \cdots, \Lambda_n$ can be determined.

Finally, the largest class which we consider in any detail is the class of strictly stable systems: systems which have a realization for which $\ell_A < 1$. Recursions on such systems are typically convergent, that is, independent of the precise initial value at $k = -\infty$. This means that it is possible to limit attention to a finite time-interval, and to obtain arbitrarily accurate initial values for this interval by performing a finite recursion on data outside the interval, starting with initial values set to 0. For the Lyapunov recursion example, Λ_1 can be determined as

$$\begin{aligned} \Lambda_1 &= A_0^* \Lambda_0 A_0 + B_0^* B_0 \\ &= A_0^* A_{-1}^* \Lambda_{-1} A_{-1} A_0 + B_0^* B_0 + A_0^* B_{-1}^* B_{-1} A_0 \\ &= A_0^* \cdots A_{-n}^* \Lambda_{-n} A_{-n} \cdots A_0 + \\ &\quad + \left\{ B_0^* B_0 + A_0^* B_{-1}^* B_{-1} A_0 + \textstyle\sum_{i=2}^{n} A_0^* \cdots A_{-i+1}^* B_{-i}^* B_{-i} A_{-i+1} \cdots A_0 \right\} \end{aligned}$$

If the system is strictly stable, then $\|A_{-n} \cdots A_0\|$ can be made arbitrarily small by choosing n large enough. Neglecting for this n the first term gives an approximation for Λ_1. The same approximation would have been obtained by choosing $\Lambda_{-n} = 0$, and computing Λ_1 via the recursion (3.15).

3.2 REALIZATIONS FOR FINITE MATRICES

The purpose of this section is to present some results on the realization theory for time-varying systems in a leisurely manner, as an introduction of some important concepts and as a preparation for a more formal and detailed discussion in section 3.4. Throughout, an important role is played by an operator related to the given transfer operator, mapping "past" inputs to "future" outputs: the Hankel operator. The relevance of this operator to the realization theory of time-invariant systems has been known since the early 1960s and resulted in Ho and Kalman's canonical realization algorithm in 1966 [5]. It was

stressed that the fundamental properties that enable one to derive a realization are not the linearity or time invariance of the system, but rather its causality and the existence of a factorization of the Hankel operator[2] into a surjective and an injective part [6]. Thus, the problem of realization was moved into the algebraic context of module theory and consisted basically of the identification of invariant factors (or invariant subspaces) of the Hankel operator. The algorithm derived by Ho does not require knowledge of these invariant factors but uses the underlying structure to derive properties of a Hankel matrix which is a representation of the Hankel operator. These properties are (1) the factoring into full rank factors $\mathcal{C}$ and $\mathcal{O}$, where the rank is equal to the degree of the system, and (2) shift-invariance. The exploitation of these properties provides explicit formulas for the realization matrices. The description of the algorithm has been simplified throughout the years by a number of authors, to the point where a three-line algorithm suffices: construct a Hankel matrix, determine a factorization into minimal-rank factors (*i.e.*, determine bases for its column space and row space), and construct the realization matrices from these factors.

In the 1970s, a new tool carried over from linear algebra into the world of system theory: the singular value decomposition. With this tool, a numerically robust way became available to compute the factorization of the Hankel matrix. The SVD was incorporated into the realization algorithm by Moore in 1978 (see [7, 8]) in the context of continuous-time systems for the purpose of *balancing* the realization. There are closely related papers by Zeiger and McEwen [9] and by Pernebo and Silverman [10]. It was realized at that time that a balanced realization can be approximated very straightforwardly, and the resulting combination (reported by Kung in 1978 [11] for discrete-time systems) gave rise to a class of robust *identification* algorithms, called Principal Component identification techniques.

This description paved the way for modern subspace-based identification algorithms, where the purpose is to determine a realization from measured input/output data in the presence of noise. Although the Hankel matrix is not known in these algorithms, the key operation is again to identify bases for the column space and row space of the matrix nonetheless. Such algorithms rely on the linearity of the system: by taking linear combinations of the available data, inputs "in the past" are constructed, along with their outputs "in the future". An overview of these identification algorithms appears in [12, 13]. Related applications are direction-of-arrival estimation in antenna array signal processing, which concerns the estimation of the angles of arrival of a number of narrowband plane waves impinging upon an antenna array. The shift-invariance property is provided by structural properties of the antenna array, which should typically consist of (at least) two subarrays, where the second array is equal to a spatially translated copy of the first array.

In this chapter, we follow related strategies. We first derive, in this section, a realization algorithm for finite upper triangular matrices. The algorithm is based on the properties of a sequence of "Hankel matrices" and generalizes Ho's algorithm to time-varying systems.

[2] called the "restricted input/output map" in [6]

A more formal framework results if a Hankel operator is defined which represents the sequence of Hankel matrices. We first explain the connection between both descriptions. Subsequently, section 3.4 covers the abstract realization theory, based on an analysis of this Hankel operator.

Realization algorithm for upper triangular matrices

To avoid a discussion on boundedness or convergence at this point, let us assume that we are given a finite upper triangular matrix T, as a special case of a bounded operator. Assume that we know some time-varying realization $\{A, B, C, D\}$ of T, where A, B, C, D are diagonals with entries A_i, B_i, C_i, D_i, as in equation (3.3), and the realization equations are given by (3.5) or equivalently, by (3.2):

$$\begin{aligned} x_{k+1} &= x_k A_k + u_k B_k \\ y_k &= x_k C_k + u_k D_k \end{aligned}$$

The objective is to find those properties of this realization that would enable us to derive it directly from T if it had been unknown, *i.e.*, to find a realization scheme.

The key idea leading to such a realization scheme is the following observation. Denote a certain time instant as 'current time', say $k = 0$. Apply an input sequence $u \in \ell_2$ to the system which has values only up to $k = -1$ and which is equal to 0 from $k = 0$ on. Such an input is said to be in 'the past', with respect to time $k = 0$. The corresponding output sequence $y = uT$ is taken into account only from time $k = 0$ on, *i.e.*, we record only the 'future' part of y. See figure 3.2(*a*). The following two observations form the cornerstone of realization theory. Let $y_{f(i)}$ denote the half-sided sequence $y_{f(i)} = [\, y_i \;\; y_{i+1} \;\; \cdots] \in \ell_2^+$, and likewise define $u_{p(i)} = [\, u_{i-1} \;\; u_{i-2} \;\; \cdots] \in \ell_2^-$. First, note that the future output sequence is dependent only on x_0:

$$y_{f(0)} = [\, y_0 \;\; y_1 \;\; \cdots] = x_0 [\, C_0 \;\; A_0 C_1 \;\; A_0 A_1 C_2 \;\; \cdots].$$

Hence upon applying all possible inputs that stop at $k = -1$, the corresponding possible outputs $y_{f(0)}$ are seen to be limited by the finite dimension of x_0 to be in a two-dimensional subspace in ℓ_2^+. This subspace is called the *output state space* at time $k = 0$. The key feature of this subspace is, for realizations with finite-size state matrices $\{A_k, B_k, C_k, D_k\}$ (which are the only ones we consider), that its dimension is not infinite but finite (and for our purposes typically low): its dimension is at most equal to the number of states of the realization at time $k = 0$. Of course, if we select another point in time as current time, then, mutatis mutandis, the same is true.

A second observation is almost trivial. If we stop the input at $k = -1$, but now only record the output from $k = 1$ on, then we reach a subset of the subspace $\{y_{f(1)}\}$. This subset is again a subspace, now of the form

$$x_0 A_0 [C_1 \;\; A_1 C_2 \;\; A_1 A_2 C_3 \;\; \cdots]. \tag{3.16}$$

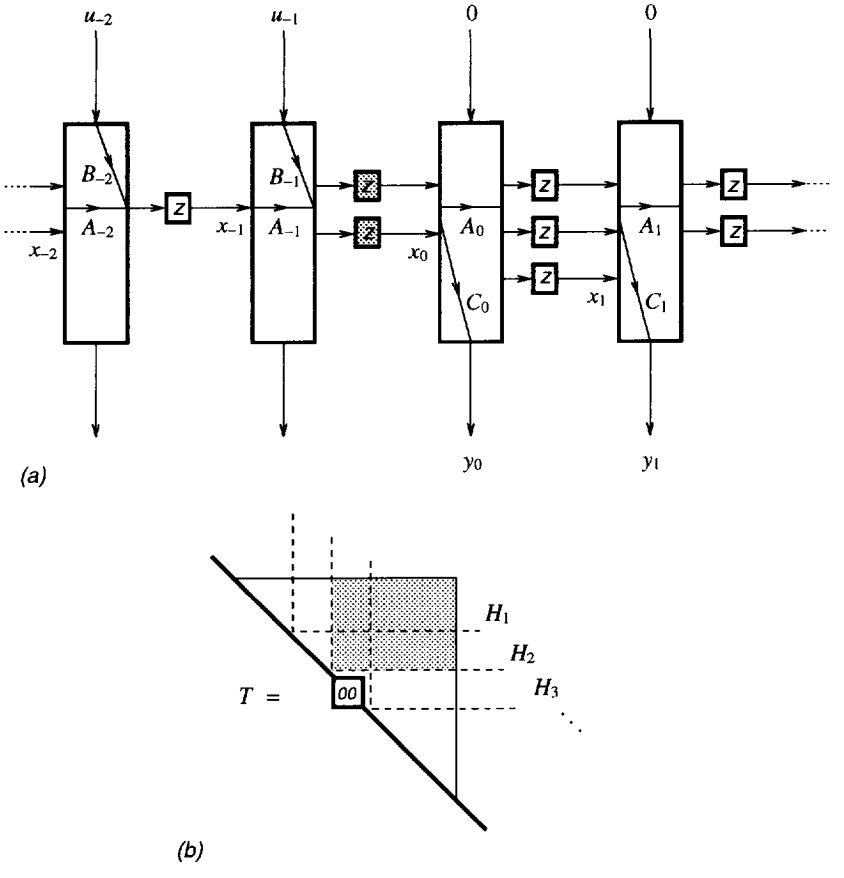


Figure 3.2. (*a*) Principle of the identification of a time-varying state-space model. In this diagram, the 'current time' is 0. All possible inputs up till time $k = -1$ ('the past') are applied, and the corresponding output sequences are recorded from time $k = 0$ on ('the future'). Thus, only part of T is used: H_0, a Hankel operator at instant $k = 0$ (*b*). This should be done in turn for all k.

A refinement of this observation leads to the mathematical concept of *shift invariance*. The appearance of A_0 in this expression enables us to identify it.

Write $u_{p(0)} = [u_{-1} \quad u_{-2} \quad u_{-3} \quad \cdots]$. Then from the relation $y = uT$ follows $y_{f(0)} = u_{p(0)}H_0$, where

$$H_0 = \begin{bmatrix} T_{-1,0} & T_{-1,1} & T_{-1,2} & \cdots \\ T_{-2,0} & T_{-2,1} & & \\ T_{-3,0} & & \ddots & \\ \vdots & & & \end{bmatrix}. \tag{3.17}$$

H_0 is a mirrored submatrix of T (see figure 3.2(*b*)): it is a submatrix because the part below row (-1) of T is multiplied by u_0, u_1, etc., which are all equal to zero, and because the part corresponding to the left of the 0-th column of T is not included in $y_{f(0)}$. The mirroring is caused by the definition of $u_{p(0)}$ as a reversed sequence of the non-zero part of u. The reversal in this definition is not essential but is traditionally included because this keeps the infinite sides of sequences and matrices like H_0 at the right and bottom. Repeating the same exercise for the signal pairs $u_{p(0)}$, $y_{f(0)}$, it follows that H_0 generalizes to a sequence of operators H_k, with similar definitions. We call the H_k the (time-varying) Hankel operators at time k. This is in analogy with the time-invariant case, where T has a Toeplitz structure so that the H_k are all the same and do indeed possess a Hankel structure (constant along anti-diagonals). Although we have lost the traditional anti-diagonal Hankel structure, we retain two important properties: the rank property and a shift-invariance property. The rank property was formulated in chapter 1 as follows:

THEOREM 3.1. *The number of states that are needed at stage k in a minimal computational network of an upper triangular matrix T is equal to the rank of its k-th Hankel matrix* H_k.

This is a Kronecker-type theorem. We are now ready to prove this theorem.

PROOF The rank property is the following. Suppose that $\{A_k, B_k, C_k, D_k\}$ is a realization for T so that the entries T_{ij} are given in terms of the $\{A_k B_k, C_k, D_k\}$ as in equation (3.9). Then a typical Hankel matrix has the following factorization:

$$\begin{aligned} H_0 &= \begin{bmatrix} B_1C_2 & B_1A_2C_3 & B_1A_2A_3C_4 & \cdots \\ B_0A_1C_2 & B_0A_1A_2C_3 & & \\ B_{-1}A_0A_1C_2 & & \ddots & \\ \vdots & & & \end{bmatrix} \\ &= \begin{bmatrix} B_{-1} \\ B_{-2}A_{-1} \\ B_{-3}A_{-2}A_{-1} \\ \vdots \end{bmatrix} [C_0 \quad A_0C_1 \quad A_0A_1C_2 \quad \cdots] =: \mathcal{C}_0\mathcal{O}_0. \end{aligned} \tag{3.18}$$

Hence, the multiplication $y_{f(0)} = u_{p(0)}H_0$ is split into two stages using an intermediate quantity x_0 which is precisely the state at time $k = 0$:

$$\begin{aligned} x_0 &= u_{p(0)}\mathcal{C}_0 \\ y_{f(0)} &= x_0\mathcal{O}_0 \end{aligned}$$

This factorization is very typical of any state realization. From the decomposition $H_k = \mathcal{C}_k\mathcal{O}_k$ it is directly inferred that if A_k is of size $(d_k \times d_{k+1})$, then $\mathcal{C}_0$ and $\mathcal{O}_0^*$ have d_k columns so that $\text{rank}(H_k)$ is at most equal to d_k. If the decomposition is *minimal*, that is, if $\mathcal{C}$ and $\mathcal{O}$ are full-rank factors ($\mathcal{C}_k^*\mathcal{C}_k > 0$, $\mathcal{O}_k\mathcal{O}_k^* > 0$), then $\text{rank}(H_k) = d_k$. It remains to prove the *existence* of a realization $\{A_k, B_k, C_k, D_k\}$ for which $d_k = \text{rank}(H_k)$: if it does, then clearly this must be a minimal realization. To find such a minimal realization, take any minimal factorization $H_k = \mathcal{C}_k\mathcal{O}_k$ into full rank factors $\mathcal{C}_k$ and $\mathcal{O}_k$. We must show that there are matrices $\{A_k, B_k, C_k, D_k\}$ such that

$$\mathcal{C}_k = \begin{bmatrix} B_{k-1} \\ B_{k-2}A_{k-1} \\ \vdots \end{bmatrix} \qquad \mathcal{O}_k = [C_k \quad A_kC_{k+1} \quad A_kA_{k+1}C_{k+2} \quad \cdots] . \tag{3.19}$$

To this end, we use the fact that H_k satisfies a restricted shift-invariance property: for example, with $H_0^{\leftarrow}$ denoting H_0 without its first column, we have

$$H_0^{\leftarrow} = \begin{bmatrix} B_{-1} \\ B_{-2}A_{-1} \\ B_{-3}A_{-2}A_{-1} \\ \vdots \end{bmatrix} \cdot A_0 \cdot [C_1 \quad A_1C_2 \quad A_1A_2C_3 \quad \cdots] .$$

In general, $H_k^{\leftarrow} = \mathcal{C}_kA_k\mathcal{O}_{k+1}$, and in much the same way, $H_k^{\downarrow} = \mathcal{C}_{k-1}A_{k-1}\mathcal{O}_k$, where $H_k^{\downarrow}$ is H_k deprived from its first row. The shift-invariance properties carry over to $\mathcal{C}_k$ and $\mathcal{O}_k$, *e.g.*, $\mathcal{O}_k = A_k\mathcal{O}_{k+1}$. This is the property hinted at in equation (3.16).

Let be given the sequence of H_k, and for each H_k any full rank decomposition $H_k = \mathcal{C}_k\mathcal{O}_k$. From the shift-invariance property $\mathcal{O}_k = A_k\mathcal{O}_{k+1}$ we can determine A_k, as follows. Because $\mathcal{O}_{k+1}$ has full rank, $\mathcal{O}_{k+1}\mathcal{O}_{k+1}^* > 0$ so that $(\mathcal{O}_{k+1}\mathcal{O}_{k+1}^*)$ is invertible, and hence $\mathcal{O}_{k+1}$ has a right inverse $\mathcal{O}_{k+1}^{\dagger} = \mathcal{O}_{k+1}^*(\mathcal{O}_{k+1}\mathcal{O}_{k+1}^*)^{-1}$ such that $\mathcal{O}_{k+1}\mathcal{O}_{k+1}^{\dagger} = I$. It follows that $A_k = \mathcal{O}_k\mathcal{O}_{k+1}^{\dagger}$. The other state matrices can be determined from the factors $\mathcal{O}_k$ and $\mathcal{C}_k$ as well: C_k follows as the first column of the chosen $\mathcal{O}_k$, while B_k is the first row of $\mathcal{C}_{k+1}$. It remains to verify that $\mathcal{C}_k$ and $\mathcal{O}_k$ are indeed generated by this realization. This is rendered straightforward by a recursive use of the shift-invariance properties. □

The construction in the above proof produces a realization algorithm (algorithm 3.1) which we formulate for finite $(n \times n)$ upper triangular matrices T. In this algorithm, a Matlab

In: T (an upper triangular $n \times n$ matrix)
Out: $\{\mathbf{T}_k\}$ (a minimal realization, in output normal form)

$\mathcal{O}_{n+1} = [\,\cdot\,], \mathcal{C}_{n+1} = [\,\cdot\,]$
for $k = n, \cdots, 1$
$H_k =: U_k \Sigma_k V_k^*$
$d_k = \text{rank}(\Sigma_k)$
$\mathcal{C}_k = (U_k \Sigma_k)(:, 1{:}d_k)$
$\mathcal{O}_k = V_k^*(1{:}d_k, :)$
$A_k = \mathcal{O}_k [\,0 \quad \mathcal{O}_{k+1}\,]^*$
$C_k = \mathcal{O}_k(:, 1)$
$B_k = \mathcal{C}_{k+1}(1, :)$
$D_k = T(k, k)$
end

Algorithm 3.1. The realization algorithm.

notation is used: $A(:, 1{:}p)$ denotes the first p columns of A, and $A(1{:}p, :)$ the first p rows. The key part of the algorithm is to obtain a basis $\mathcal{O}_k$ for the row space of each Hankel matrix H_k of T. The singular value decomposition (SVD) [14] is a robust tool for doing this. It is a decomposition of H_k into factors U_k, Σ_k, V_k, where U_k and V_k are unitary matrices whose columns contain the left and right singular vectors of H_k, and Σ_k is a diagonal matrix with positive entries (the singular values of H_k) on the diagonal. The integer d_k is set equal to the number of non-zero singular values of H_k, and $V_k^*(1{:}d_k, :)$ contains the corresponding singular vectors. The rows of $V^*(1{:}d_k, :)$ span the row space of H_k. Note that it is natural that $d_1 = 0$ and $d_{n+1} = 0$, so that the realization starts and ends with zero number of states. The rest of the realization algorithm is straightforward in view of the shift-invariance property. It is in fact very reminiscent of the Principal Component identification method in system theory [11]. For later use, we remark that this algorithm has the property $A_k A_k^* + C_k C_k^* = I$: the realization is in 'output normal form'. We show in the following section how the 'canonical observer realization' theorem yields such realizations: in fact, it uses the same algorithm but for operators rather than matrices.

The above is only an algorithmic outline. Because H_{k+1} has a large overlap with H_k, an efficient SVD updating algorithm can be devised that takes this structure into account. Other decompositions from linear algebra that identify subspaces can be used instead. In theory a QR factorization of the H_k should work, although this is not advisable in practise because QR is not rank revealing: the addition of a small amount of noise on the entries of T will make all Hankel matrices have full rank, thus producing a realization of high

order. Decompositions that can be used instead of QR are rank revealing QR [15, 16, 17], and the URV decomposition [18], which is equivalent to SVD but computationally less demanding.

Note that, based on the singular values of H_k, a reduced order model can be obtained by taking a smaller basis for $\mathcal{O}_k$, a technique that is known in the time-invariant context as balanced model reduction. Although widely used for time-invariant systems, this is in fact a "heuristic" model reduction theory, as the modeling error norm is not known. A precise approximation theory results if the tolerance on the error is given in terms of the *Hankel norm*, which is the subject of chapter 6.

Numerical example

As an example of the realization theorem and algorithm 3.1, let the transfer matrix be given by

$$T = \left[\begin{array}{ccc|ccc} 1 & .800 & .200 & .050 & .013 & .003 \\ 0 & .900 & .600 & .240 & .096 & .038 \\ 0 & 0 & .800 & .500 & .250 & .125 \\ \hline 0 & 0 & 0 & .700 & .400 & .240 \\ 0 & 0 & 0 & 0 & .600 & .300 \\ 0 & 0 & 0 & 0 & 0 & .500 \end{array}\right] \tag{3.20}$$

The position of the Hankel matrix H_4 is indicated (recall that this submatrix must be mirrored to obtain H_4). The SVDs of the Hankel matrices are

$$H_1 = [\,\cdot\,]$$

$$H_2 = \begin{bmatrix} .800 & .200 & .050 & .013 & .003 \end{bmatrix} = 1 \cdot 0.826 \cdot \begin{bmatrix} .968 & .242 & .061 & .015 & .004 \end{bmatrix}$$

$$H_3 = \begin{bmatrix} .600 & .240 & .096 & .038 \\ .200 & .050 & .013 & .003 \end{bmatrix}$$

$$= \begin{bmatrix} .955 & .298 \\ .298 & -.955 \end{bmatrix} \begin{bmatrix} .685 & 0 \\ 0 & .033 \end{bmatrix} \begin{bmatrix} .922 & .356 & .139 & .055 \\ -.374 & .729 & .511 & .259 \end{bmatrix}$$

etcetera. In the above, columns and rows that correspond to zero singular values have been omitted. The non-zero singular values of the Hankel operators of T are

H_1	H_2	H_3	H_4	H_5	H_6
	.826	.685	.631	.553	.406
		.033	.029	.023	
			.001		

Hence T has a state-space realization which grows from zero states ($i = 1$) to a maximum of 3 states ($i = 4$), and then shrinks back to 0 states ($i > 6$). The realization algorithm (algorithm 3.1) yields as time-varying state realization for T

$$\mathbf{T}_1 = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline .826 & 1.000 \end{array}\right] \qquad \mathbf{T}_2 = \left[\begin{array}{cc|c} .247 & -.038 & .968 \\ \hline .654 & .010 & .900 \end{array}\right]$$

$$\mathbf{T}_3 = \left[\begin{array}{ccc|c} .384 & -.038 & -.000 & .922 \\ .913 & .158 & -.037 & -.374 \\ \hline .573 & .012 & .000 & .800 \end{array}\right] \qquad \mathbf{T}_4 = \left[\begin{array}{cc|c} .470 & -.030 & .882 \\ .861 & .237 & -.450 \\ -.196 & .971 & .138 \\ \hline .466 & .012 & .700 \end{array}\right]$$

$$\mathbf{T}_5 = \left[\begin{array}{c|c} .493 & .870 \\ .870 & -.493 \\ \hline .300 & .600 \end{array}\right] \qquad \mathbf{T}_6 = \left[\begin{array}{c|c} \cdot & 1.000 \\ \hline \cdot & .500 \end{array}\right]$$

As is seen from the table of singular values, H_4 is close to a singular matrix, and hence one expects that T can be approximated by a matrix close to it such that only two states are needed. That this is indeed possible will be shown in chapter 6.

3.3 THE HANKEL OPERATOR

Hankel operator definitions

A more formal approach to the derivation of a realization of a time-varying transfer operator T is based on the properties of a restriction of the domain and range of T to become an operator mapping inputs in $\mathcal{L}_2\mathcal{Z}^{-1}$ (representing inputs in 'the past') to the part of the corresponding outputs in $\mathcal{U}_2$ (the part in the 'future'). In later chapters, other restrictions of T to operators between subspaces of $\mathcal{X}_2$ will play important roles as well. Using the projection operators defined in chapter 2, define the past of a signal $U \in \mathcal{X}_2$ as $U_p = \mathbf{P}_{\mathcal{L}_2\mathcal{Z}^{-1}}(U)$, and its future part as $U_f = \mathbf{P}(U)$, so that $U = U_p + U_f$. The same definitions apply to the past and future part of an output Y. With these definitions, the action of T on an input $U \in \mathcal{X}_2$ can be broken down into three operators,

$$Y = UT \quad \Leftrightarrow \quad \begin{cases} Y_p &= U_p K_T \\ Y_f &= U_p H_T + U_f E_T \end{cases} \tag{3.21}$$

where

$$\begin{array}{lll} H_T: & \mathcal{L}_2\mathcal{Z}^{-1} \to \mathcal{U}_2, & U_p H_T = \mathbf{P}(U_p T) \\ K_T: & \mathcal{L}_2\mathcal{Z}^{-1} \to \mathcal{L}_2\mathcal{Z}^{-1}, & U_p K_T = \mathbf{P}_{\mathcal{L}_2\mathcal{Z}^{-1}}(U_p T) \\ E_T: & \mathcal{U}_2 \to \mathcal{U}_2, & U_f E_T = \mathbf{P}(U_f T) = U_f T. \end{array} \tag{3.22}$$

Note that due to causality there is no transfer from U_f to Y_p. H_T is called the Hankel operator of T: it is the map of inputs in $\mathcal{L}_2\mathcal{Z}^{-1}$ to the part in $\mathcal{U}_2$ of the corresponding outputs. See figure 3.3(b).

The first property to note on these operators is that they are left D invariant: if $Y_f = U_p H_T = \mathbf{P}(U_p T)$, then $DY_f = D\mathbf{P}(U_p T) = \mathbf{P}(DU_p T) = (DU_p)H_T$. Consequently, as discussed in section 2.3, these tensor operators have matrix representations in the form of *snapshots* and *diagonal expansions*. With the operator π_k as defined in (2.12), snapshots H_k are obtained according to equation (2.44),

$$(\pi_k^* U_k)H_T = \pi_k^*(U_k H_k) \qquad (U \in \mathcal{L}_2\mathcal{Z}^{-1}) \tag{3.23}$$

where $U_k = \pi_k U$ is the k-th row of U, so that H_k is obtained as the operation of H_T on an input restricted to the ℓ_2-sequence U_k. Since $U \in \mathcal{L}_2\mathcal{Z}^{-1}$, U_k is a sequence which has zero entries from its k-th entry on. Likewise, $Y \in \mathcal{U}_2$ has rows Y_k which have zero entries before time k. Hence, in the computation of $Y_k = U_k H_k = \pi_k[(\pi_k^* U_k)H_T]$, only the quadrant from the right of and strictly above entry (k, k) of the matrix representation of T is used. Consequently, H_k has a one-sided infinite matrix representation, and as it is customary to have the infinite sides at the right and bottom of the matrix representation, we write it as a mirrored submatrix of T, which yields H_k as used in the previous section:

$$H_k = \begin{bmatrix} T_{k-1,k} & T_{k-1,k+1} & T_{k-1,k+2} & \cdots \\ T_{k-2,k} & T_{k-2,k+1} & & \\ T_{k-3,k} & & \ddots & \\ \vdots & & & \end{bmatrix}. \tag{3.24}$$

H_k in equation (3.24) maps sequences $u_{p(k)} = [\, U_{k,k-1} \;\; U_{k,k-2} \;\; \cdots]$, which are isomorphic to U_k, to sequences $y_{f(k)} = [\, Y_{k,k} \;\; Y_{k,k+1} \;\; Y_{k,k+2} \;\; \cdots]$ which is isomorphic to Y_k with the zero entries before entry $Y_{k,k}$ omitted. H_k according to this definition is not precisely the same but isomorphic to the definition in (3.23), where the isomorphism consists of the removal of zero rows and columns and a mirroring. We use the definition of H_k in (3.24) from now on.

A second representation of H_T is obtained in the form of diagonal expansions, as introduced in chapter 2. Inputs and outputs are represented as vectors whose entries are diagonals. This defines Hilbert spaces which are isomorphic to $\mathcal{L}_2\mathcal{Z}^{-1}$ and $\mathcal{U}_2$ and are such that H_T can be represented by a matrix of diagonal entries. Because we act on $\mathcal{L}_2\mathcal{Z}^{-1}$ and $\mathcal{U}_2$, rather than $\mathcal{X}_2$, the definition of the diagonal expansion of a signal is done slightly more specifically than in chapter 2. For $Y \in \mathcal{U}_2$ we define the diagonal expansion of Y as a one-sided sequence of diagonals $\tilde{Y}$,

$$\begin{aligned} Y &= Y_{[0]} + ZY_{[1]} + Z^2 Y_{[2]} + \cdots = Y_{[0]} + Y_{[1]}^{(-1)} Z + Y_{[2]}^{(-2)} Z^2 + \cdots \\ \tilde{Y} &= \begin{bmatrix} Y_{[0]} & Y_{[1]}^{(-1)} & Y_{[2]}^{(-2)} & \cdots \end{bmatrix}. \end{aligned} \tag{3.25}$$

$\tilde{Y}$ is an element of the space $\ell_2^+(\mathcal{D})$ of one-sided square-summable sequences whose entries are diagonals. It is isomorphic to $\mathcal{U}_2$. Analogously, for $U \in \mathcal{L}_2\mathcal{Z}^{-1}$, the diagonal expansion

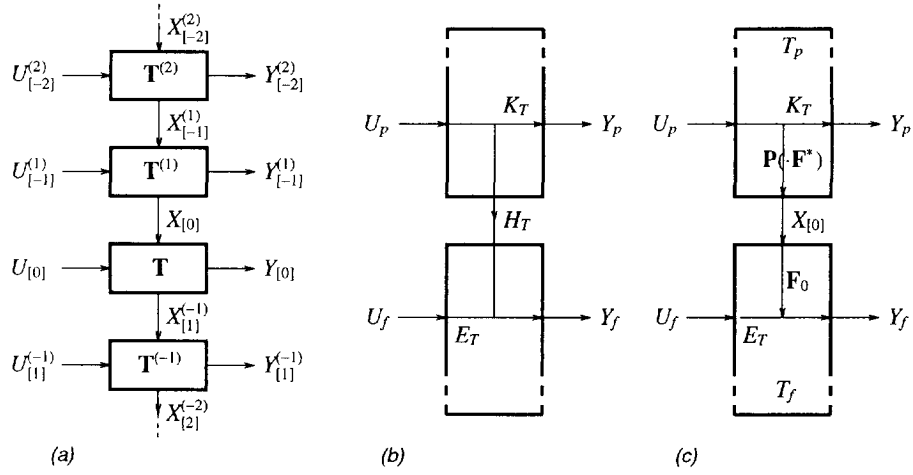


Figure 3.3. (*a*) realization **T**, (*b*) splitting into past and future signals, (*c*) representation by T_p and T_f.

of U is also designated by $\tilde{U} \in \ell_2^-(\mathcal{D})$, now defined by

$$\begin{aligned} U &= Z^{-1}U_{[-1]} + Z^{-2}U_{[-2]} + \cdots = U^{(+1)}_{[-1]}Z^{-1} + U^{(+2)}_{[-2]}Z^{-2} + \cdots \\ \tilde{U} &= \begin{bmatrix} U^{(+1)}_{[-1]} & U^{(+2)}_{[-2]} & \cdots \end{bmatrix}. \end{aligned} \tag{3.26}$$

Take $U \in \mathcal{L}_2 Z^{-1}$. Corresponding to the operator H_T, the operator $\tilde{H}_T$ acts on diagonal expansions $\tilde{U}$. The definition

$$\tilde{H}_T = \begin{bmatrix} T_{[1]} & T^{(-1)}_{[2]} & T^{(-2)}_{[3]} & \cdots \\ T_{[2]} & T^{(-1)}_{[3]} & & \\ T_{[3]} & & \ddots & \\ \vdots & & & \end{bmatrix} \tag{3.27}$$

is such that

$$Y_f = UH_T \in \mathcal{U}_2 \qquad \Leftrightarrow \qquad \tilde{Y}_f = \tilde{U}\tilde{H}_T.$$

The connection of $\tilde{H}_T$ with the sequence of snapshots H_k is obtained by selecting the k-th entry of each diagonal in $\tilde{H}_T$ and constructing a matrix from it: this yields precisely H_k. The same expansions can be done for K_T and E_T; for example, $\tilde{K}_T : \ell_2^-(\mathcal{D}) \to \ell_2^-(\mathcal{D})$ and

$K_i : \ell_2^- \to \ell_2^-$ are given by

$$\tilde{K}_T = \begin{bmatrix} T_{[0]}^{(+1)} & & \mathbf{0} & \\ T_{[1]}^{(+1)} & T_{[0]}^{(+2)} & & \\ T_{[2]}^{(+1)} & T_{[1]}^{(+2)} & T_{[0]}^{(+3)} & \\ \vdots & & \vdots & \ddots \end{bmatrix}, \qquad K_i = \begin{bmatrix} K_{i-1,i-1} & & \mathbf{0} & \\ K_{i-2,i-1} & K_{i-2,i-2} & & \\ K_{i-3,i-1} & K_{i-3,i-2} & K_{i-3,i-3} & \\ \vdots & & \vdots & \ddots \end{bmatrix}.$$

Factorization of H_T

If a realization $\{A, B, C, D\}$ of T is given, then we know from (3.18) that the H_k admit a factorization into $H_k = \mathcal{C}_k \mathcal{O}_k$. An obvious question that emerges is whether H_T admits such a factorization. The answer should be affirmative, of course, in view of the close connection between H_T and its snapshots H_k. While a mathematical factorization into operators is certainly possible, we relegate that to the following paragraph and first check the factorization property on the diagonal expansion $\tilde{H}_T$ of H_T. With a state realization $\{A, B, C, D\}$ the diagonals $T_{[i]}$ of T are given by equation (3.8). $\tilde{H}_T$ in terms of A, B, C, D then follows as the matrix of diagonal operators

$$\tilde{H}_T = \begin{bmatrix} B^{(1)}C & B^{(1)}AC^{(-1)} & B^{(1)}AA^{(-1)}C^{(-2)} & \cdots \\ B^{(2)}A^{(1)}C & B^{(2)}A^{(1)}AC^{(-1)} & B^{(2)}A^{(1)}AA^{(-1)}C^{(-2)} & \\ B^{(3)}A^{(2)}A^{(1)}C & B^{(3)}A^{(2)}A^{(1)}AC^{(-1)} & \ddots & \\ \vdots & & & \end{bmatrix}.$$

We can distinguish operators (column and row sequences with entries that are diagonal operators)

$$\mathcal{C} := \begin{bmatrix} B^{(1)} \\ B^{(2)}A^{(1)} \\ B^{(3)}A^{(2)}A^{(1)} \\ \vdots \end{bmatrix} \qquad \mathcal{O} := \begin{bmatrix} C & AC^{(-1)} & AA^{(-1)}C^{(-2)} & \cdots \end{bmatrix}. \tag{3.28}$$

$\mathcal{C}$ is called the controllability operator, while $\mathcal{O}$ is called the observability operator. These are generalizations of the corresponding concepts in time-invariant systems theory.[3] It is straightforward to verify that if $\{A, B, C, D\}$ is a realization of T, then $\tilde{H}_T$ admits a decomposition $\tilde{H}_T = \mathcal{C}\mathcal{O}$. The operators $\mathcal{C}_k$ and $\mathcal{O}_k$ of (3.18) are obtained by selecting the k-th entry of each diagonal in $\mathcal{C}$ and $\mathcal{O}$.

If $U_f = 0$, then $Y_f = U_p H_T$. The following construction shows that the existence of a realization implies that H_T can be factored into two operators. According to the state

[3] $\mathcal{C}$ is often called the reachability operator in recent literature.

equations in (3.11), $X_{[0]}$ is equal to

$$\begin{aligned} X_{[0]} &= U^{(1)}_{[-1]}B^{(1)} + X^{(1)}_{[-1]}A^{(1)} \\ &= U^{(1)}_{[-1]}B^{(1)} + U^{(2)}_{[-2]}B^{(2)}A^{(1)} + X^{(2)}_{[-2]}A^{(2)}A^{(1)} \\ &= \begin{bmatrix} U^{(1)}_{[-1]} & U^{(2)}_{[-2]} & \cdots \end{bmatrix} \begin{bmatrix} B^{(1)} \\ B^{(2)}A^{(1)} \\ B^{(3)}A^{(2)}A^{(1)} \\ \vdots \end{bmatrix} \\ &= \tilde{U}_p \mathcal{C} \end{aligned} \tag{3.29}$$

Hence $X_{[0]} = \tilde{U}_p \mathcal{C}$. We can continue the above derivation via

$$\begin{aligned} X_{[0]} &= \mathbf{P}_0 \left([\, U^{(1)}_{[-1]}Z^{-1} + U^{(2)}_{[-2]}Z^{-2} + \cdots][\, ZB^{(1)} + Z^2B^{(2)}A^{(1)} + Z^3B^{(3)}A^{(2)}A^{(1)} + \cdots] \right) \\ &= \mathbf{P}_0 \left([Z^{-1}U_{[-1]} + Z^{-2}U_{[-2]} + \cdots][\, BZ + BZAZ + BZAZAZ + \cdots] \right) \\ &= \mathbf{P}_0 (U_p [\, BZ + BZAZ + BZ(AZ)^2 + \cdots]) \\ &= \mathbf{P}_0 (U_p \mathbf{F}^*) \end{aligned} \tag{3.30}$$

where

$$\mathbf{F}^* := BZ + BZAZ + BZ(AZ)^2 + \cdots . \tag{3.31}$$

(We will keep this definition for $\mathbf{F}$ throughout the rest of the section.) The summation in (3.31) need not converge to a bounded operator, but if the realization is bounded ($X_{[k]} \in \mathcal{D}_2$ for all $U \in \mathcal{U}_2$), then $\mathbf{P}_0(\cdot\, \mathbf{F}^*)$ is a bounded $[\mathcal{U}_2 \to \mathcal{D}_2^{\mathcal{B}}]$ operator. From now on, we assume that the realization satisfies this condition. [The realization theory in section 3.4 establishes that bounded operators T which are locally finite always admit bounded realizations, so that it is enough to consider such realizations.] The expression $X_{[0]} = \tilde{U}_p \mathcal{C}$ shows that the diagonal expansion of $\mathbf{P}_0(\cdot\, \mathbf{F}^*)$ is equal to the controllability operator $\mathcal{C}$, that is

$$\left[\mathbf{P}_0(\cdot\, \mathbf{F}^*)\right]^{\sim} = \mathcal{C}. \tag{3.32}$$

If $\ell_A < 1$, then the summation in the definition of $\mathbf{F}^*$ can be summarized to $\mathbf{F}^* = BZ(I - AZ)^{-1}$, which is now a bounded operator: $F \in \mathcal{L}Z^{-1}$, and hence $X_{[0]}$ is given by

$$X_{[0]} = \mathbf{P}_0(U\, BZ(I - AZ)^{-1})$$

This expression could have been obtained directly from the closed-form expression for X which we obtained earlier, $X = UBZ(I - AZ)^{-1}$, by computing $X_{[0]} = \mathbf{P}_0(X)$.

If $U_f = 0$ then, for $k \geq 0$, $Y_{[k]} = X_{[k]}C = X^{(k)}_{[0]}A^{(k)}A^{(k-1)} \cdots A^{(1)}C$, so that

$$Y_f = Y_{[0]} + ZY_{[1]} + Z^2Y_{[2]} + \cdots$$

$$
\begin{aligned}
&= X_{[0]}(C + ZA^{(1)}C + Z^2A^{(2)}A^{(1)}C + \cdots) \\
&= X_{[0]}(I + AZ + AZAZ + \cdots)C \\
&= X_{[0]}\mathbf{F}_0
\end{aligned} \tag{3.33}
$$

where

$$
\mathbf{F}_0 := (I + AZ + (AZ)^2 + \cdots)C. \tag{3.34}
$$

Again, this operator need not be bounded in $\mathcal{X}$, but it is bounded as a $[\mathcal{D}_2^{\mathcal{B}} \to \mathcal{X}_2]$ operator. If $\ell_A < 1$, then $\mathbf{F}_0 = (I - AZ)^{-1}C$, and

$$
Y_f = X_{[0]}(I - AZ)^{-1}C
$$

A reformulation of (3.33) leads to

$$
\begin{aligned}
\tilde{Y}_f &= [\,Y_{[0]} \quad Y_{[1]}^{(-1)} \quad Y_{[2]}^{(-2)} \quad \cdots\,] \\
&= X_{[0]}\left[C \quad AC^{(-1)} \quad AA^{(-1)}C^{(-2)} \quad \cdots\right] \\
&= X_{[0]}\mathcal{O}.
\end{aligned}
$$

This shows that the diagonal expansion of $\mathbf{F}_0$ is equal to the observability operator: $\tilde{\mathbf{F}}_0 = \mathcal{O}$.

Combination of (3.29) and (3.33) yields the observation that the map $H_T : U_p \mapsto Y_f$ can be split into a map $\mathbf{P}_0(\cdot\,\mathbf{F}^*) : U_p \mapsto X_{[0]}$ followed by the map $\mathbf{F}_0 : X_{[0]} \mapsto Y_f$. Hence H_T has a factorization:

THEOREM 3.2. *Let $T \in \mathcal{U}$ have a bounded locally finite realization $\{A, B, C, D\}$. Let $\mathbf{F}$ and $\mathbf{F}_0$ be as given in (3.31) and (3.34). Then H_T has a factorization into*

$$
H_T = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0. \tag{3.35}
$$

This factorization is totally equivalent to $\tilde{H}_T = \mathcal{C}\mathcal{O}$ and $H_k = \mathcal{C}_k\mathcal{O}_k$. In view of (3.21) and the factorization of the Hankel operator, the computation of $Y = UT$ can be split into two operations,

$$
\left\{ \begin{array}{l} \begin{bmatrix} X_{[0]} & Y_p \end{bmatrix} = U_pT_p \\ Y_f = [\,X_{[0]} \quad U_f\,]T_f \end{array} \right. \qquad \text{where} \qquad \left\{ \begin{array}{lll} T_p &=& \begin{bmatrix} \mathbf{P}_0(\cdot\,\mathbf{F}^*) & K_T \end{bmatrix} \\ T_f &=& \begin{bmatrix} \mathbf{F}_0 \\ E_T \end{bmatrix} \end{array} \right. \tag{3.36}
$$

where T_p is a 'past' operator and T_f is a 'future' operator. See figure 3.3(*c*). The expression of H_T in equation (3.35) recalls the projection formula (2.38) in chapter 2, where $\mathbf{F}$ and $\mathbf{F}_0$ play the role of basis representations. The precise relations are investigated in section 3.4, but with the above observation in mind, we define at the present point the 'Gram' operators or Gramians corresponding to $\mathbf{F}$ and $\mathbf{F}_0$ as

$$
\begin{aligned}
\Lambda_{\mathbf{F}} &= \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) \quad \in \mathcal{D}(\mathcal{B}, \mathcal{B}) \\
\Lambda_{\mathbf{F}_0} &= \mathbf{P}_0(\mathbf{F}_0\mathbf{F}_0^*) \quad \in \mathcal{D}(\mathcal{B}, \mathcal{B}).
\end{aligned}
$$

Controllability and observability

An important aspect of the factorization $H_T = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0$ is the investigation of the (local) minimality of this factorization, since this implies the minimality of the dimension sequence $\#(\mathcal{B})$ of $X_{[0]}$ and thus the minimality of the realization. Let $\{A, B, C, D\}$ be a bounded locally finite realization of T where $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$.

DEFINITION 3.3. *Let $\mathbf{F}^*$ be given by (3.31). A realization is said to be controllable if the range of $\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}}$ is dense in $\mathcal{D}_2^{\mathcal{B}}$ and uniformly controllable if its range is all of $\mathcal{D}_2^{\mathcal{B}}$, that is, if $\mathbf{P}_0(\mathcal{L}_2\mathcal{Z}^{-1}\,\mathbf{F}^*) = \mathcal{D}_2^{\mathcal{B}}$ (the operator is 'onto').*

If $\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}}$ is regarded as an operator from $\mathcal{L}_2\mathcal{Z}^{-1} \to \mathcal{D}_2^{\mathcal{B}}$, then its adjoint operator is $\cdot\,\mathbf{F}$ with domain $\mathcal{D}_2^{\mathcal{B}}$. In view of (2.9), we obtain the decomposition $\mathcal{D}_2^{\mathcal{B}} = \overline{\mathrm{ran}}\big[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}}\big] \oplus \ker\big[\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}\big]$. The realization is controllable if and only if $\ker\big[\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}\big] = 0$, *i.e.*, if $D\mathbf{F} = 0$ $(D \in \mathcal{D}_2^{\mathcal{B}}) \Rightarrow D = 0$, so that the operator $\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}$ is 'one-to-one'. In terms of diagonal inner products, we know that $D\mathbf{F} = 0 \Leftrightarrow \{D\mathbf{F}, D\mathbf{F}\} = 0$, and $\{D\mathbf{F}, D\mathbf{F}\} = \mathbf{P}_0(D\mathbf{F}\mathbf{F}^*D) = D\mathbf{P}_0(\mathbf{F}\mathbf{F}^*)D^*$. This implies that the realization is controllable if and only if the Gram operator $\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) > 0$. The realization is uniformly controllable if $\Lambda_{\mathbf{F}}$ is uniformly positive. Reverting to diagonal expansions, application of (2.42) to (3.32) gives $\mathbf{P}_0(\mathbf{F}\mathbf{F}^*) = \mathcal{C}^*\mathcal{C}$, so that the realization is controllable if $\mathcal{C}^*\mathcal{C} > 0$ and uniformly controllable if $\mathcal{C}^*\mathcal{C}$ is uniformly positive, that is, if $\mathcal{C}^*\mathcal{C} \gg 0$. In summary:

PROPOSITION 3.4. *A realization is controllable if and only if $\Lambda_{\mathbf{F}} := \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) = \mathcal{C}^*\mathcal{C} > 0$, and uniformly controllable if and only if $\Lambda_{\mathbf{F}} = \mathcal{C}^*\mathcal{C} \gg 0$.*

Observability is defined in much the same way.

DEFINITION 3.5. *Let $\mathbf{F}_0^*$ be given by (3.34). A realization is observable if $\mathbf{P}_0(\mathcal{U}_2\mathbf{F}_0^*)$ is dense in $\mathcal{D}_2^{\mathcal{B}}$ (the operator $\cdot\,\mathbf{F}_0\big|_{\mathcal{D}_2}$ is one-to-one), and uniformly observable if $\mathbf{P}_0(\mathcal{U}_2\mathbf{F}_0^*) = \mathcal{D}_2^{\mathcal{B}}$.*

DEFINITION 3.6. *A realization which is both controllable and observable is said to be minimal.*

PROPOSITION 3.7. *A realization is observable if and only if $\Lambda_{\mathbf{F}_0} := \mathbf{P}_0(\mathbf{F}_0\mathbf{F}_0^*) = \mathcal{O}\mathcal{O}^* > 0$, and uniformly observable if and only if $\Lambda_{\mathbf{F}_0} = \mathcal{O}\mathcal{O}^* \gg 0$.*

The map $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$: $\overline{\mathrm{ran}}\big[\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}\big] \mapsto \overline{\mathrm{ran}}\big[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big]$ is one-to-one and hence has an algebraic inverse, although it is not necessarily bounded. If a realization is controllable, then

$\overline{\text{ran}}\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right] = \mathcal{D}_2^{\mathcal{B}}$, so that for any $X_{[0]} \in \mathcal{D}_2^{\mathcal{B}}$ there is an input $U \in \mathcal{L}_2\mathcal{Z}^{-1}$ such that $X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*)$. However, this U need not be bounded. If a realization is *uniformly* controllable, then only a bounded input is needed to reach any state in $\mathcal{D}_2^{\mathcal{B}}$:

LEMMA 3.8. *A realization is uniformly controllable if and only if there exists $M < \infty$ such that*

$$\forall X_{[0]} \in \mathcal{D}_2^{\mathcal{B}} \quad \exists U \in \mathcal{L}_2\mathcal{Z}^{-1} \;:\; \begin{cases} (1) & X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*) \\ (2) & \|U\|_{HS} \leq M\|X_{[0]}\|_{HS} \end{cases}$$

PROOF We apply theorem 2.1 to the operator $\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}}$ and its adjoint $\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}$:

$$\begin{aligned} &\text{ran}\,\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}} \text{ is closed} \\ &\Leftrightarrow \quad \exists\,\varepsilon > 0 :\; \forall\, U \in \overline{\text{ran}}\left[\cdot\,\mathbf{F}\big|_{\mathcal{D}_2}\right] :\; \|\mathbf{P}_0(U\mathbf{F}^*)\|_{HS} \geq \varepsilon\|U\|_{HS}. \end{aligned} \tag{3.37}$$

In the remainder of the proof, regard $\cdot\,\mathbf{F} = \cdot\,\mathbf{F}\big|_{\mathcal{D}_2}$, with adjoint $\mathbf{P}_0(\cdot\,\mathbf{F}^*) = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2\mathcal{Z}^{-1}}$.

($\Rightarrow$) If the realization is uniformly controllable, then ran $\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right] = \mathcal{D}_2^{\mathcal{B}}$ is closed. A first implication is that for any $X_{[0]} \in \mathcal{D}_2^{\mathcal{B}}$, there is an $U \in \overline{\text{ran}}\,[\cdot\,\mathbf{F}]$ such that $X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*)$. Taking $M = 1/\varepsilon$ in (3.37), it follows that for any $U \in \overline{\text{ran}}\,[\cdot\,\mathbf{F}]$ and $X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*)$, it holds that $\|U\| \leq M\|X_{[0]}\|$.

($\Leftarrow$) Suppose, conversely, that there is $M < \infty$ such that for all $X_{[0]} \in \mathcal{D}_2^{\mathcal{B}}$ we can find a $U \in \mathcal{L}_2\mathcal{Z}^{-1}$ such that (1) $X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*)$ and (2) $\|U\|_{HS} \leq M\|X_{[0]}\|_{HS}$. It then follows from (1) that $\overline{\text{ran}}\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right] = \mathcal{D}_2^{\mathcal{B}}$, and it remains to show that ran $\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right]$ is closed. The map $\mathbf{P}_0(\cdot\,\mathbf{F}^*) :\; \overline{\text{ran}}\,[\cdot\,\mathbf{F}] \mapsto \overline{\text{ran}}\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right]$ is one-to-one, so that (2) holds for all $U \in \overline{\text{ran}}\,[\cdot\,\mathbf{F}]$ and $X_{[0]} = \mathbf{P}_0(U\mathbf{F}^*)$. Applying (3.37) with $\varepsilon = 1/M$ shows that ran $\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right]$ is closed, so that the realization is uniformly controllable. □

Lyapunov equations

Equation (3.31) yields the expression

$$\mathbf{F}^* = BZ + \mathbf{F}^*AZ.$$

However, care must be taken in the use of this expression, since $\mathbf{F}^*$ is not necessarily a bounded operator on $\mathcal{X}_2$. However, $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$ is a bounded operator on $\mathcal{X}_2$, which leads to the following proposition.

PROPOSITION 3.9. *Let $\{A, B, C, D\}$ be a bounded realization, and let the operator $\mathbf{F}^*$ be given by equation (3.31), with Gramian $\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*)$. Then $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$ is a bounded operator on $\mathcal{X}_2$, and satisfies*

$$\mathbf{P}_0(\cdot\,\mathbf{F}^*) = \mathbf{P}_0(\cdot\,[BZ + \mathbf{F}^*AZ]) \tag{3.38}$$

The adjoint of $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$ *satisfies (on* $\mathcal{D}_2^{\mathcal{B}}$*, and by extension everywhere on* $\mathcal{X}_2$ *where it is bounded)*

$$\mathbf{F} = Z^*B^* + Z^*A^*\mathbf{F} \tag{3.39}$$

$\Lambda_{\mathbf{F}}$ *satisfies the equation*

$$\Lambda_{\mathbf{F}}^{(-1)} = B^*B + A^*\Lambda_{\mathbf{F}}A\,. \tag{3.40}$$

PROOF It was argued before that $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$ is a bounded operator on $\mathcal{X}_2$. The indicated equation follows directly from rewriting equation (3.31) for $\mathbf{P}_0(\cdot\,\mathbf{F}^*)$. The expression for the adjoint is then obtained along the lines of the proof of proposition 2.7. Consequently, the Gram operator $\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*)$ satisfies

$$\begin{aligned}
\Lambda_{\mathbf{F}}^{(-1)} &= \mathbf{P}_0(Z^*\mathbf{F}\mathbf{F}^*Z) \\
&= \mathbf{P}_0(Z[Z^*B^* + Z^*A^*\mathbf{F}]\,[BZ + \mathbf{F}^*AZ]Z^*) \\
&= \mathbf{P}_0(B^*B) + \mathbf{P}_0(A^*\mathbf{F}\mathbf{F}^*A) + \mathbf{P}_0(B^*\mathbf{F}^*A) + \mathbf{P}_0(A^*\mathbf{F}B) \\
&= B^*B + A^*\mathbf{P}_0(\mathbf{F}\mathbf{F}^*)A + 0 + 0\,.
\end{aligned}$$

□

Equations of the type

$$A^*MA + B^*B = M^{(-1)}, \qquad M \in \mathcal{D}(\mathcal{B}, \mathcal{B}). \tag{3.41}$$

are known as Lyapunov or Lyapunov-Stein equations. As discussed at the end of section 3.1, this equation in diagonals can be viewed as a recursive relation $M_{k+1} = A_k^*M_kA_k + B_k^*B_k$ of the entries of the diagonals. The recursion is obtained by selecting the k-th entry of each diagonal. If $\ell_A < 1$, then, as is easy to verify by substitution, the equation has a solution given by

$$M = \sum_{k=0}^{\infty} (A^{\{k\}})^*(B^*B)^{(k+1)}A^{\{k\}}\,,$$

where $A^{\{k\}} = A^{(k)}\cdots A^{(1)}$ was defined in equation (2.21). This summation is precisely equal to the summation that results in the computation of $\Lambda_{\mathbf{F}} = \mathcal{C}^*\mathcal{C}$, and hence $M = \Lambda_{\mathbf{F}}$ and the summation converges. Moreover, this solution is unique: if Λ is another solution, then

$$\begin{aligned}
(M-\Lambda)^{(-1)} &= A^*(M-\Lambda)A \\
\Rightarrow \quad M-\Lambda &= (A^{\{k\}})^*(M-\Lambda)^{(k+1)}A^{\{k\}}
\end{aligned}$$

and $\ell_A < 1$ implies $A^{\{k\}} \to 0$ so that $\Lambda = M$. If $\ell_A = 1$, then the Lyapunov equation does not have a unique solution. For example, if $A = I$ and $B = 0$, then the resulting equation is $M^{(-1)} = M$ so that any $M = \alpha I$ with $\alpha \in \mathbb{C}$ will do, whereas $\Lambda_{\mathbf{F}} = 0$ in this example.

In the same way, we obtain the dual to proposition 3.9.

PROPOSITION 3.10. *Let $\{A, B, C, D\}$ be a bounded realization, and let the operator $\mathbf{F}_0$ be given by equation (3.34), with Gramian $\Lambda_{\mathbf{F}_0} = \mathbf{P}_0(\mathbf{F}_0\mathbf{F}_0^*)$. Then*

$$\begin{array}{lrcl} \text{On } \mathcal{D}_2^{\mathcal{B}}: & \mathbf{F}_0 & = & C + AZ\mathbf{F}_0 \\ \text{On } \mathcal{X}_2: & \mathbf{P}_0(\cdot\,\mathbf{F}_0^*) & = & \mathbf{P}_0(\cdot\,[C^* + \mathbf{P}_0(\cdot\,\mathbf{F}_0^*)Z^*A^*]) \end{array} \tag{3.42}$$

The expression on $\mathcal{D}_2^{\mathcal{B}}$ is also bounded on dense domains in $\mathcal{X}_2^{\mathcal{B}}$. $\Lambda_{\mathbf{F}_0}$ satisfies the dual Lyapunov equation

$$\Lambda_{\mathbf{F}_0} = CC^* + A\Lambda_{\mathbf{F}_0}^{(-1)}A^*. \tag{3.43}$$

Again, if $\ell_A < 1$, then the solution to the equation $Q = CC^* + AQ^{(-1)}A^*$ is unique and equal to $\Lambda_{\mathbf{F}_0}$.

Lyapunov equations arise in the *normalization* of a given realization. Suppose that we are given a bounded minimal realization $\{A, B, C, D\}$ of some locally finite operator $T \in \mathcal{U}$. The objective is to find a similar realization $\{A', B', C', D\}$ which is in input normal form, *i.e.*, for which $\Lambda_{\mathbf{F}'} = I$. In view of (3.40), such a realization satisfies $A'^*A' + B'^*B' = I$. Let $\mathbf{F}$ and $\mathbf{F}_0$ be the controllability and observability operators of $\mathbf{T}$ as in (3.31) and (3.34), and define $\mathbf{F}'$ and $\mathbf{F}_0'$ likewise for $\mathbf{T}'$. If R is a state transformation that brings $\mathbf{T}$ into $\mathbf{T}'$ according to (3.13), then $\mathbf{F} = R^*\mathbf{F}'$ and $R\mathbf{F}_0 = \mathbf{F}_0'$, and the corresponding Gram operators satisfy

$$\begin{array}{rcl} \Lambda_{\mathbf{F}} & = & R^*\Lambda_{\mathbf{F}'}R \\ \Lambda_{\mathbf{F}_0'} & = & R\Lambda_{\mathbf{F}_0}R^* \end{array} \tag{3.44}$$

The first equation gives

$$\Lambda_{\mathbf{F}} = R^*R$$

so that the required state transformation R is given by a factor of $\Lambda_{\mathbf{F}}$. R is boundedly invertible if and only if $\Lambda_{\mathbf{F}}$ is uniformly positive, that is, if the given realization is uniformly controllable. If $\ell_A < 1$, then R is obtained by solving the Lyapunov equation (3.41) for M, followed by solving the factorization $M = R^*R$. Another way to arrive at the Lyapunov equation directly is by inserting the relations $A' = RAR^{(-1)}$ and $B' = BR^{(-1)}$ into the normalization condition $A'^*A' + B'^*B' = I$, and putting $M = R^*R$. Likewise, a realization in output normal form (for which $\Lambda_{\mathbf{F}_0'} = I$ so that $A'A'^* + C'C'^* = I$) is obtained by factoring $\Lambda_{\mathbf{F}_0} = R^{-1}R^{-*}$, and it is seen that the given realization must be uniformly observable. Again, if $\ell_A < 1$, then R can be obtained by solving the Lyapunov equation $Q = CC^* + AQ^{(-1)}A^*$ for Q after which R is obtained as a factor of Q^{-1}. The Lyapunov equation is directly obtained by inserting the relations $A' = RAR^{(-1)}$ and $C' = RC$ into the condition $A'A'^* + C'C'^* = I$.

Nerode state-space definitions

Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a bounded causal transfer operator of a linear time-varying system mapping signals in $\mathcal{X}_2^{\mathcal{M}}$ to signals in $\mathcal{X}_2^{\mathcal{N}}$. $H_T = \mathbf{P}(\cdot\,T)\big|_{\mathcal{L}_2 Z^{-1}}$ is the operator T with domain

and range restricted to two complementary subspaces, $\mathcal{L}_2Z^{-1}$ and $\mathcal{U}_2$, respectively. Since T is bounded and the projection operator $\mathbf{P}$ is contractive on $\mathcal{X}_2$, H_T obviously is also bounded, so that its adjoint $H_T : \mathcal{U}_2 \to \mathcal{L}_2Z^{-1}$ exists and is bounded. In the study of the Hankel operator, we study the effect of inputs in $\mathcal{L}_2Z^{-1}$ onto outputs in $\mathcal{U}_2$, *i.e.*, we study the range and kernel of H_T and H_T^*. Note that neither $\mathrm{ran}(H_T)$ nor $\mathrm{ran}(H_T^*)$ have to be closed.

Define

$$\begin{array}{rcl} \mathcal{K} & = & \ker(H_T) = \{U \in \mathcal{L}_2Z^{-1} : \mathbf{P}(UT) = 0\} \\ \mathcal{H} & = & \mathrm{ran}(H_T^*) = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T^*). \end{array} \tag{3.45}$$

$\mathcal{K}$ is called the *input null space*. It is a left D-invariant subspace in $\mathcal{L}_2Z^{-1}$. $\mathcal{K}$ defines equivalence classes in $\mathcal{L}_2Z^{-1}$: we say that an input $U_1 \in \mathcal{L}_2Z^{-1}$ is Nerode equivalent to $U_2 \in \mathcal{L}_2Z^{-1}$ if they have the same future outputs: $\mathbf{P}(U_1T) = \mathbf{P}(U_2T)$. Consequently, $\mathbf{P}[(U_1 - U_2)T] = 0$, hence U_1 is Nerode equivalent to U_2 if $U_1 - U_2 \in \mathcal{K}$.

$\mathcal{H}$ is called the (natural) *input state space*. It is a left D-invariant linear manifold in $\mathcal{L}_2Z^{-1}$. Since $\ker(H_T) \oplus \overline{\mathrm{ran}(H_T^*)} = \mathcal{L}_2Z^{-1}$ (*cf.* equation (2.9)), the space $\overline{\mathcal{H}}$ is the complement of $\mathcal{K}$ in $\mathcal{L}_2Z^{-1}$:

$$\overline{\mathcal{H}} \oplus \mathcal{K} = \mathcal{L}_2Z^{-1}. \tag{3.46}$$

In the same way, define the (natural) *output state space* $\mathcal{H}_0$ to be the the range of H_T, and the *output null space* $\mathcal{K}_0$ to be the kernel of T^*:

$$\begin{array}{rcl} \mathcal{H}_0 & = & \mathrm{ran}(H_T) = \mathbf{P}(\mathcal{L}_2Z^{-1}T) \\ \mathcal{K}_0 & = & \ker(H_T^*) = \{Y \in \mathcal{U}_2 : \mathbf{P}(YT^*) = 0\}. \end{array} \tag{3.47}$$

$\mathcal{H}_0$ is the left D-invariant manifold containing the projections in $\mathcal{U}_2$ of all outputs of the system that can be generated from inputs in $\mathcal{L}_2Z^{-1}$. $\mathcal{K}_0$ is its complement in $\mathcal{U}_2$:

$$\overline{\mathcal{H}}_0 \oplus \mathcal{K}_0 = \mathcal{U}_2. \tag{3.48}$$

The null and state spaces satisfy the following relations:

$$\begin{array}{rcl} \mathbf{P}(\mathcal{K}T) & = & 0 \\ \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{K}_0T^*) & = & 0 \\ \mathcal{H}_0 = \overline{\mathcal{H}}H_T & = & \mathbf{P}(\overline{\mathcal{H}}T) \\ \mathcal{H} = \overline{\mathcal{H}}_0H_T^* & = & \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\overline{\mathcal{H}}_0T^*). \end{array} \tag{3.49}$$

(The last two equations follow from inserting (3.46) and (3.48) into the definitions of $\mathcal{H}$ and $\mathcal{H}_0$, and using the first two equations.) These relations ensure that $\mathcal{H}$ and $\mathcal{H}_0$ have the same dimension sequences:

LEMMA 3.11. *If $\overline{\mathcal{H}}$ and $\overline{\mathcal{H}}_0$ are locally finite subspaces, then*

$$\text{s-dim}(\overline{\mathcal{H}}) = \text{s-dim}(\overline{\mathcal{H}}_0).$$

PROOF Apply the property mentioned in equation (2.30) to (3.49): $\overline{\mathcal{H}}_0 = \overline{\overline{\mathcal{H}} H_T}$ and $\overline{\mathcal{H}} = \overline{\overline{\mathcal{H}}_0 H_T^*}$. This yields $\text{s-dim}(\overline{\mathcal{H}}_0) \leq \text{s-dim}(\overline{\mathcal{H}})$ and $\text{s-dim}(\overline{\mathcal{H}}) \leq \text{s-dim}(\overline{\mathcal{H}}_0)$. □

PROPOSITION 3.12. *Let $\{A, B, C, D\}$ be a bounded locally finite realization of T, $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$, and let $\mathbf{F}$ and $\mathbf{F}_0$ be the associated controllability and observability operators (equations (3.31), (3.34)). Then $\mathcal{H}_0 \subset \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$ and $\mathcal{H} \subset \mathcal{D}_2^{\mathcal{B}} \mathbf{F}$.*

If the realization is controllable, then $\mathcal{K}_0 = \ker \mathbf{P}_0(\cdot\, \mathbf{F}_0^*)\big|_{\mathcal{U}_2}$, $\overline{\mathcal{H}}_0 = \overline{\mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0}$.

If the realization is uniformly controllable, then $\mathcal{H}_0 = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$.

If the realization is observable, then $\mathcal{K} = \ker \mathbf{P}_0(\cdot\, \mathbf{F}^*)\big|_{\mathcal{L}_2 Z^{-1}}$, $\overline{\mathcal{H}} = \overline{\mathcal{D}_2^{\mathcal{B}} \mathbf{F}}$.

If the realization is uniformly observable, then $\mathcal{H} = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}$.

PROOF H_T has the factorization $H_T = \mathbf{P}_0(\cdot\, \mathbf{F}^*)\mathbf{F}_0$, so that $\mathcal{H}_0 = \text{ran}\, H_T \subset \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$. If the realization is uniformly controllable, then $\mathbf{P}_0(\mathcal{L}_2 Z^{-1}\, \mathbf{F}^*) = \mathcal{D}_2^{\mathcal{B}}$, so that, indeed, $\mathcal{H}_0 = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$. Also, $\mathcal{K}_0 = \ker H_T^* = \ker \mathbf{P}_0(\cdot\, \mathbf{F}_0^*)\mathbf{F}\big|_{\mathcal{U}_2}$. If the realization is controllable, then $\mathbf{F}$ is one-to-one and $\mathcal{K}_0 = \ker \mathbf{P}_0(\cdot\, \mathbf{F}_0^*)\big|_{\mathcal{U}_2}$, with complement $\overline{\mathcal{H}} = \overline{\text{ran}}(\cdot\, \mathbf{F}_0)\big|_{\mathcal{D}_2} = \overline{\mathcal{D}_2^{\mathcal{B}} \mathbf{F}}$. The remaining statements are proven in the same manner. □

PROPOSITION 3.13. *If a realization of T is both uniformly controllable and uniformly observable, then $\mathcal{H}_0 = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$ and $\mathcal{H} = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}$ are closed subspaces, i.e., the ranges of H_T and H_T^* are closed subspaces.*

Conversely, let $\mathcal{H}_0$ and $\mathcal{H}$ be closed subspaces. If the realization is controllable, then it is uniformly controllable. Likewise, if the realization is observable, then it is uniformly observable.

PROOF The first part of the lemma follows immediately from lemma 3.12: since the realization is uniformly controllable, $\text{ran}\, H_T = \mathbf{P}_0(\mathcal{L}_2 Z^{-1} \mathbf{F})\, \mathbf{F}_0 = \mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$. Because the realization is uniformly observable, $\mathcal{D}_2^{\mathcal{B}} \mathbf{F}_0$ is a closed subspace, and hence $\text{ran}\, H_T$ is a closed subspace.

To prove the second part, we again apply theorem 2.1, in the form

$$\begin{aligned} &\text{ran}\, \mathbf{P}_0(\cdot\, \mathbf{F}^*)\big|_{\mathcal{L}_2 Z^{-1}} \text{ is closed} \\ &\quad \Leftrightarrow \quad \exists\, \varepsilon > 0 : \ \forall\, D \in \overline{\text{ran}}\, \mathbf{P}_0(\cdot\, \mathbf{F}^*)\big|_{\mathcal{L}_2 Z^{-1}} : \quad \| D\mathbf{F} \|_{HS} \geq \varepsilon \| D \|_{HS}. \end{aligned}$$

Let the realization be controllable, then $\mathcal{H}$ closed implies with proposition 3.12 that $\mathcal{D}_2^{\mathcal{B}}\mathbf{F}$ (the range of $\mathbf{F}$ restricted to $\mathcal{D}_2^{\mathcal{B}}$) is closed. Then the range of the adjoint of $\mathbf{F}$ is also closed, that is, ran $\mathbf{P}_0(\cdot\,\mathbf{F}^*)\big|_{\mathcal{L}_2 Z^{-1}}$ is closed, and hence by definition 3.3 it is equal to $\mathcal{D}_2^{\mathcal{B}}$. Hence $\|D\mathbf{F}\|_{HS} \geq \varepsilon \|D\|_{HS}$ for all $D \in \mathcal{D}_2^{\mathcal{B}}$. But $\|D\mathbf{F}\|_{HS}^2 = \text{trace}\, D\Lambda_{\mathbf{F}}D^*$, so this inequality implies $\Lambda_{\mathbf{F}} \gg 0$: $\mathbf{F}$ is uniformly controllable. □

Proposition 3.12 has a direct corollary, which is part of a Kronecker theorem for time-varying systems. The second part appears as theorem 3.28 in the next section.

COROLLARY 3.14. (KRONECKER'S THM, I) *Let $T \in \mathcal{U}$ be a locally finite transfer operator which has a bounded realization with state-space sequence $\mathcal{B}$. If the realization is minimal, then* s-dim $\mathcal{H}$ = s-dim $\mathcal{H}_0$ *is equal to the sequence of dimensions of $\mathcal{B}$.*

This corollary is also true at the local level: if the realization is minimal and the k-th slice $\pi_k\mathcal{H}_0 = (\mathcal{H}_0)_k$ of $\mathcal{H}_0$ has a dimension d_k, then d_k is equal to the number of states of the realization at point k. It is also true that $(\mathcal{H}_0)_k$ is isomorphic to the range of H_k, the k-th snapshot of the Hankel operator, where the isomorphism consists of the conversion of ℓ_2-sequences to ℓ_2^+-sequences, and that d_k is equal to the rank of the Hankel operator at point k, *i.e.*, the rank of h_k.

It remains to prove the converse of the corollary, *i.e.*, to show that if s-dim $\mathcal{H}$ = s-dim $\mathcal{H}_0$ = $[\cdots\ d_0\ d_1\ d_2\ \cdots]$ is a uniformly bounded sequence of dimensions, where $d_k = \text{rank}\, H_k$, then there exist realizations of T with d_k equal to the system order at point k. We call the sequence the *minimal system order* of T. The actual construction of such minimal realizations is the subject of the following section, where the converse of corollary 3.14 appears as theorem 3.28.

Numerical example

To illustrate some of the above with a numerical example, consider again the transfer matrix T given in equation (3.20). The range of the Hankel operator H_T is given locally by the row spaces of the Hankel matrices $\{H_k\}$. These are given in turn by the V-matrices of the SVDs of the $\{H_k\}$ that have already been computed in the previous section. Hence, for example,

$$\mathcal{O}_1 = (\tilde{\mathbf{F}}_0)_1 = [\,\cdot\,]$$

$$\mathcal{O}_2 = (\tilde{\mathbf{F}}_0)_2 = \begin{bmatrix} .968 & .242 & .061 & .015 & .004 \end{bmatrix}$$

$$\mathcal{O}_3 = (\tilde{\mathbf{F}}_0)_3 = \begin{bmatrix} .922 & .356 & .139 & .055 \\ -.374 & .729 & .511 & .259 \end{bmatrix}$$

etcetera. The operator $\mathbf{F}_0$ as used in the present section is obtained by stacking these matrices into one upper operator. This gives

$$\mathbf{F} = \left[\begin{array}{cccccccccc} & & \vdots & & & & & \vdots & & \\ \hline \cdots & \cdot & \boxed{\cdot} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ \cdots & \cdot & 0 & .968 & .242 & .061 & .015 & .004 & \cdot & \cdots \\ \hline \cdots & \cdot & 0 & 0 & .922 & .356 & .139 & .055 & \cdot & \cdots \\ \cdots & \cdot & 0 & 0 & -.374 & .729 & .511 & .259 & \cdot & \cdots \\ \hline & & \vdots & & & & & \vdots & & \end{array}\right] \tag{3.50}$$

Using the realization of T derived in the previous section, it is readily verified that, indeed, $\mathbf{F}$ satisfies equation (3.34). A straightforward way to do this is to check that, for each k, $\mathcal{O}_k = [C_k \;\; A_k C_{k+1} \;\; A_k A_{k+1} C_{k+2} \;\; \cdots]$.

3.4 ABSTRACT REALIZATION THEORY

We continue with the analysis of H_T and its characteristic subspaces, $\mathcal{H}$ and $\mathcal{H}_0$. We show how a shift-invariance property on these spaces, along with the choice of a basis in either one of them, produces minimal realizations which are either in 'input normal form' (or in 'canonical controller form') or in 'output normal form' (canonical observer form). In all cases, bounded realizations with $\ell_A \leq 1$ are obtained.

Shift-invariance properties

In chapter 2, we defined Z to be the bilateral right shift operator in ℓ_2, and hence by extension the bilateral shift in $\mathcal{X}_2$. We now define the unilateral or restricted shift operator $\mathcal{Z}$ with domain and range restricted to $\mathcal{L}_2 Z^{-1}$ as $\mathcal{Z}U = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(ZU) = ZU - \mathbf{P}_0(ZU)$ for $U \in \mathcal{L}_2 Z^{-1}$. The restricted shift operator $\mathcal{S} = Z^{-1}$ on $\mathcal{U}_2$ is defined as $\mathcal{S}Y = \mathbf{P}(Z^{-1}Y)$.

The null and state spaces satisfy the following shift-invariance properties, which are of crucial importance for the derivation of state realizations.

LEMMA 3.15. *Let $\mathcal{H}$, $\mathcal{K}$ and $\mathcal{H}_0$, $\mathcal{K}_0$ be as defined in equations (3.45), (3.47). Then*

$$\begin{array}{rcl} Z^{-1}\mathcal{K} & \subset & \mathcal{K} \\ \mathcal{Z}\mathcal{H} & \subset & \mathcal{H}; \qquad \mathcal{Z}\overline{\mathcal{H}} \subset \overline{\mathcal{H}} \end{array}$$

$$\begin{array}{rcl} Z\mathcal{K}_0 & \subset & \mathcal{K}_0 \\ \mathcal{S}\mathcal{H}_0 & \subset & \mathcal{H}_0; \qquad \mathcal{S}\overline{\mathcal{H}}_0 \subset \overline{\mathcal{H}}_0 \end{array}$$

PROOF

($Z^{-1}\mathcal{K} \subset \mathcal{K}$): If $U \in \mathcal{K}$, then $\mathbf{P}(UT) = 0$, hence $UT \in \mathcal{L}_2Z^{-1}$ and thus $Z^{-1}UT \in \mathcal{L}_2Z^{-1}$, too. But this means that $\mathbf{P}(Z^{-1}UT) = 0$ so that $Z^{-1}U \in \mathcal{K}$.

($Z\mathcal{H} \subset \mathcal{H}$):

$$\begin{aligned} Z\mathcal{H} &= \mathbf{P}_{\mathcal{L}_2Z^{-1}}\left[Z\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T^*)\right] \\ &= \mathbf{P}_{\mathcal{L}_2Z^{-1}}\left[Z\mathcal{U}_2T^*\right] \\ &\subset \mathbf{P}_{\mathcal{L}_2Z^{-1}}\left[\mathcal{U}_2T^*\right] = \mathcal{H}. \end{aligned}$$

($Z\overline{\mathcal{H}} \subset \overline{\mathcal{H}}$): In general, let $U, X \in \mathcal{L}_2Z^{-1}$. Then

$$\begin{aligned} \{ZU, X\} &= \{\mathbf{P}_{\mathcal{L}_2Z^{-1}}(ZU), X\} \\ &= \{ZU, X\} \\ &= \{U^{(-1)}, Z^*X^{(-1)}\} = \{U, Z^*X\}^{(-1)}. \end{aligned}$$

(Use has been made of the fact that $ZD = D^{(-1)}Z$.) In particular, if $U \in \overline{\mathcal{H}}$, $X \in \mathcal{K}$, then the shift-invariance of $\mathcal{K}$ implies $Z^*X \in \mathcal{K}$, hence $U \perp Z^*X$, *i.e.*, $\{U, Z^*X\} = 0$. But this implies that $ZU \perp X$. Since X can be any element of $\mathcal{K}$, $ZU \perp \mathcal{K}$, hence $Z\overline{\mathcal{H}} \subset \overline{\mathcal{H}}$.

The remaining three properties (and proofs) are dual to the above. □

Canonical controller operator realizations

Let T be a given bounded linear causal time-varying system transfer operator in $\mathcal{U}(\mathcal{M}, \mathcal{N})$, and assume that its shift-invariant input/output state and null spaces, $\mathcal{H}, \mathcal{H}_0, \mathcal{K}$ and $\mathcal{K}_0$, are known. $\mathcal{H}$ is such that $\mathbf{P}(\mathcal{L}_2Z^{-1}T) = \mathbf{P}(\overline{\mathcal{H}}T)$, hence the effect of any input in the past ($\mathcal{L}_2Z^{-1}$) on the future output in $\mathcal{U}_2$ is equivalently described by a (unique) representative element $\mathbf{X}$ of $\overline{\mathcal{H}}$, called the state. The point is that $\overline{\mathcal{H}}$, as a subspace of $\mathcal{L}_2Z^{-1}$, is assumed to be a much smaller dimensional space than $\mathcal{L}_2Z^{-1}$, so that the state indeed "summarizes" the past input. A refinement of these observations yields the construction of an operator state-space model (since $\mathbf{X}$ is an operator in $\mathcal{L}_2Z^{-1}$), in a way that is already familiar from a number of other contexts (see *e.g.*, [6, 19, 20, 21, 22]). By choosing a basis in $\overline{\mathcal{H}}$, the desired result, a minimal state-space realization involving only diagonal operators, is obtained. The realization that is obtained by choosing the state in $\overline{\mathcal{H}}$ is called the canonical controller state realization, in analogy with the canonical controller realizations described in Kailath [23]. Alternatively, we can choose the state operator in the output state space, which we call the canonical observer realization. We present both solutions.

For a given input U in $\mathcal{X}_2$ and instant k, we have defined the past input with respect to time instant 0 to be $U_p = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(U)$. More in general, let the past input $U_{p(k)}$ (with respect to instant k) be denoted by $U_{p(k)} = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(Z^{-k}U)$. To obtain a canonical controller operator state realization, define the state $\mathbf{X}_k \in \overline{\mathcal{H}}$ at instant k to be the projection of the

past input onto $\overline{\mathcal{H}}$:

$$\mathbf{X}_k = \mathbf{P}_{\mathcal{H}}(U_{p(k)}) = \mathbf{P}_{\mathcal{H}}(Z^{-k}U) \in \mathcal{L}_2 Z^{-1}. \tag{3.51}$$

We use, in the following theorem, diagonal representations of $U, Y \in \mathcal{X}_2$ as in equation (2.23), *viz.*

$$\begin{aligned} U &= \sum Z^k U_{[k]}, & U_{[k]} &= \mathbf{P}_0(Z^{-k}U), \\ Y &= \sum Z^k Y_{[k]}, & Y_{[k]} &= \mathbf{P}_0(Z^{-k}Y). \end{aligned}$$

THEOREM 3.16. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a given transfer operator with input state space $\mathcal{H}$. With $U \in \mathcal{X}_2^{\mathcal{M}}$, $Y \in \mathcal{X}_2^{\mathcal{N}}$, let $Y_{[k]} \in \mathcal{D}_2^{\mathcal{N}}$ and $\mathbf{X}_k \in \overline{\mathcal{H}}$ be given by the operator state equations*

$$\begin{bmatrix} \mathbf{X}_{k+1} &= \mathbf{X}_k\mathbf{A} + U_{[k]}\mathbf{B} \\ Y_{[k]} &= \mathbf{X}_k\mathbf{C} + U_{[k]}\mathbf{D} \end{bmatrix} \tag{3.52}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are bounded operators satisfying

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{P}_{\mathcal{H}}(Z^{-1}\cdot) & \mathbf{P}_0(\cdot\, T) \\ \mathbf{P}_{\mathcal{H}}(Z^{-1}\cdot) & \mathbf{P}_0(\cdot\, T) \end{bmatrix}, \qquad \begin{aligned} \mathbf{A}&: \overline{\mathcal{H}} \to \overline{\mathcal{H}} & \mathbf{C}&: \overline{\mathcal{H}} \to \mathcal{D}_2^{\mathcal{N}} \\ \mathbf{B}&: \mathcal{D}_2^{\mathcal{M}} \to \overline{\mathcal{H}} & \mathbf{D}&: \mathcal{D}_2^{\mathcal{M}} \to \mathcal{D}_2^{\mathcal{N}} \end{aligned} \tag{3.53}$$

Then $Y = UT$ and $\mathbf{X}_k = \mathbf{P}_{\mathcal{H}}(U_{p(k)})$.

PROOF We first show that defining Y by $Y = UT$ and $\mathbf{X}_k$ by (3.51) implies the realization equations (3.52). Since $U_{p(k)} \in \mathcal{L}_2 Z^{-1} = \overline{\mathcal{H}} \oplus \mathcal{K}$, and $\mathbf{P}_0(\mathcal{K}T) = 0$ by definition of $\mathcal{K}$, we have

$$\begin{aligned} \mathbf{P}_0(U_{p(k)}T) &= \mathbf{P}_0\left[\mathbf{P}_{\mathcal{H}}(U_{p(k)})T + \mathbf{P}_{\mathcal{K}}(U_{p(k)})T\right] \\ &= \mathbf{P}_0(\mathbf{X}_k T). \end{aligned}$$

Hence

1. $Y = UT \quad \Leftrightarrow \quad$
$$\begin{aligned} Y_{[k]} &= \mathbf{P}_0(Z^{-k}Y) \\ &= \mathbf{P}_0(Z^{-k}UT) \\ &= \mathbf{P}_0\left[\mathbf{P}_{\mathcal{L}_2 Z^{-1}}(Z^{-k}U)T\right] + \mathbf{P}_0\left[\mathbf{P}_0(Z^{-k}U)T\right] \\ &= \mathbf{P}_0(U_{p(k)}T) + \mathbf{P}_0(U_{[k]}T) \\ &= \mathbf{P}_0(\mathbf{X}_k T) + \mathbf{P}_0(U_{[k]}T) \\ &=: \mathbf{X}_k\mathbf{C} + U_{[k]}\mathbf{D}. \end{aligned}$$

2. $$\begin{aligned} \mathbf{X}_{k+1} &= \mathbf{P}_{\mathcal{H}}(U_{p(k+1)}) \\ &= \mathbf{P}_{\mathcal{H}}(Z^{-k-1}U) \\ &= \mathbf{P}_{\mathcal{H}}(Z^{-1}U_{p(k)} + Z^{-1}U_{[k]}) \\ &= \mathbf{P}_{\mathcal{H}}\left[Z^{-1}\mathbf{P}_{\mathcal{H}}(U_{p(k)}) + Z^{-1}\mathbf{P}_{\mathcal{K}}(U_{p(k)})\right] + \mathbf{P}_{\mathcal{H}}(Z^{-1}U_{[k]}) \\ &= \mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{X}_k) + \mathbf{P}_{\mathcal{H}}(Z^{-1}U_{[k]}) \\ &=: \mathbf{X}_k\mathbf{A} + U_{[k]}\mathbf{B}. \end{aligned}$$

where in making the last step the fact is used that $\mathcal{K}$ is shift-invariant ($Z^{-1}\mathcal{K} \subset \mathcal{K}$) and that $\overline{\mathcal{H}} \perp \mathcal{K}$.

To prove that the equations yield $Y = UT$ and $\mathbf{X}_k$ as specified, we first show

$$\cdot\, \mathbf{A}^n = \mathbf{P}_{\mathcal{H}}(Z^{-n}\cdot). \tag{3.54}$$

Indeed, let $\mathbf{G}_n = \mathbf{G}\mathbf{A}^n$, for some $\mathbf{G} \in \overline{\mathcal{H}}$. Then

$$\begin{aligned}
\mathbf{G}_1 = \mathbf{G}\mathbf{A} &= \mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{G}),\\
\mathbf{G}_2 = \mathbf{G}\mathbf{A}^2 &= (\mathbf{G}\mathbf{A})\mathbf{A}\\
&= \mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{G}_1)\\
&= \mathbf{P}_{\mathcal{H}}\left[Z^{-1}\mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{G})\right]\\
&= \mathbf{P}_{\mathcal{H}}\left[Z^{-1}\mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{G}) + Z^{-1}\mathbf{P}_{\mathcal{K}}(Z^{-1}\mathbf{G})\right]\\
&= \mathbf{P}_{\mathcal{H}}(Z^{-2}\mathbf{G})
\end{aligned}$$

where we have used the fact that $\mathcal{K}$ is Z^{-1} invariant so that $Z^{-1}\mathcal{K}$ is orthogonal to $\overline{\mathcal{H}}$. The result on $\mathbf{A}^n$ follows by repeating these arguments.

With this result, $\mathbf{X}_k$ given by (3.52) is obtained as

$$\begin{aligned}
\mathbf{X}_k &= U_{[k-1]}\mathbf{B} + U_{[k-2]}\mathbf{B}\mathbf{A} + U_{[k-3]}\mathbf{B}\mathbf{A}^2 + \cdots\\
&= \mathbf{P}_{\mathcal{H}}(Z^{-1}U_{[k-1]} + Z^{-2}U_{[k-2]} + \cdots)\\
&= \mathbf{P}_{\mathcal{H}}(U_{p(k)}),
\end{aligned}$$

and hence satisfies (3.51). With this $\mathbf{X}_k$, $Y_{[k]} = \mathbf{P}_0(Z^{-k}Y)$ follows from $Y_{[k]} = \mathbf{X}_k\mathbf{C} + U_{[k]}\mathbf{D}$ by reversing the derivations of the equalities in the first step. □

$\mathbf{A}$ is the restricted shift operator in $\overline{\mathcal{H}}$. It is clear that $\|\mathbf{A}\| \leq 1$, and that if there exists an $\hat{\mathbf{X}} \in \overline{\mathcal{H}}$ such that $Z^{-1}\hat{\mathbf{X}} \in \overline{\mathcal{H}}$, then $\|\mathbf{A}\| = 1$. Let $r(\mathbf{A})$ denote the spectral radius of $\mathbf{A}$:

$$r(\mathbf{A}) = \lim_{n\to\infty} \|\mathbf{A}^n\|^{1/n}.$$

Since $\|\mathbf{A}\| \leq 1$ we have that $r(\mathbf{A}) \leq 1$ also.

Canonical controller realization

Although the above state-space description in terms of operators is the core of any state realization, it is in its present form not yet very useful for our purposes. If we assume the state space to be of locally finite dimension, then by choosing an orthonormal basis representation $\mathbf{Q}$ in $\overline{\mathcal{H}}$, it is possible to "precompute" the effect of the operators $\mathbf{A}$, $\mathbf{B}$ and T on $\mathbf{Q}$, and arrive at a state-space description with diagonal operators A, B, C, D only. This is demonstrated in the following theorem, where the crucial step is the definition of A via $\mathbf{Q}\mathbf{A} = A^{(1)}\mathbf{Q}$. Some care must be taken if $\mathbf{Q}$ is an unbounded operator on $\mathcal{X}_2$. It can be shown that this happens only if $r(\mathbf{A}) = 1$, and that $r(\mathbf{A}) = 1$ coincides with $\ell_A = 1$, where $\ell_A = r(ZA)$ is the spectral radius of the operator ZA. Nonetheless, $\mathbf{Q}$ is bounded as a $[\mathcal{D}_2^{\mathcal{B}} \to \mathcal{X}_2]$ operator, and this property is sufficient to prove the theorem.

THEOREM 3.17. (CANONICAL CONTROLLER REALIZATION) *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a given transfer operator with input state space $\mathcal{H}$ of locally finite dimensions. Let $d = \text{s-dim}(\mathcal{H})$, and put $\mathcal{B} = \mathbb{C}^d$. Let $\mathbf{Q}$ be an orthonormal basis representation of $\overline{\mathcal{H}}$: $\overline{\mathcal{H}} = \mathcal{D}_2^{\mathcal{B}}\mathbf{Q}$, $\Lambda_{\mathbf{Q}} = I$. Take $U \in \mathcal{X}_2^{\mathcal{M}}$, $Y \in \mathcal{X}_2^{\mathcal{N}}$ with diagonals $U_{[k]} = \mathbf{P}_0(Z^{-k}U)$ and $Y_{[k]} = \mathbf{P}_0(Z^{-k}Y)$.*

Then

$$Y = UT \quad \Leftrightarrow \quad \begin{bmatrix} X_{[k+1]}^{(-1)} &= X_{[k]}A + U_{[k]}B \\ Y_{[k]} &= X_{[k]}C + U_{[k]}D \end{bmatrix}, \tag{3.55}$$

where

$$\begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)^{(-1)} & \mathbf{P}_0(\mathbf{Q}T) \\ \mathbf{P}_0(Z^{-1}\mathbf{Q}^*)^{(-1)} & \mathbf{P}_0(T) \end{bmatrix} \qquad \begin{matrix} A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)}) & C \in \mathcal{D}(\mathcal{B}, \mathcal{N}) \\ B \in \mathcal{D}(\mathcal{M}, \mathcal{B}^{(-1)}) & D \in \mathcal{D}(\mathcal{M}, \mathcal{N}). \end{matrix}$$

PROOF For a given $\mathbf{X}_k$ in $\overline{\mathcal{H}}$, proposition 2.8 showed that $\mathbf{X}_k$ can be written in terms of the basis $\mathbf{Q}$ of $\overline{\mathcal{H}}$ as $\mathbf{X}_k = X_{[k]}\mathbf{Q}$, where $X_{[k]} = \mathbf{P}_0(\mathbf{X}_k\mathbf{Q}^*) \in \mathcal{D}_2^{\mathcal{B}}$. Starting with the operator realization in theorem 3.16 for a certain k and $\mathbf{X}_k$, write the new state $\mathbf{X}_{k+1} \in \overline{\mathcal{H}}$ as $\mathbf{X}_{k+1} = X_{[k+1]}\mathbf{Q}$. Then

$$\begin{aligned} \mathbf{X}_{k+1} = X_{[k+1]}\mathbf{Q} &= \mathbf{X}_k\mathbf{A} &&+ U_{[k]}\mathbf{B} \\ &= \mathbf{P}_{\mathcal{H}}(Z^{-1}\mathbf{X}_k) &&+ \mathbf{P}_{\mathcal{H}}(Z^{-1}U_{[k]}) \\ &= \mathbf{P}_0(Z^{-1}\mathbf{X}_k\mathbf{Q}^*)\mathbf{Q} &&+ \mathbf{P}_0(Z^{-1}U_{[k]}\mathbf{Q}^*)\mathbf{Q} \qquad \text{[thm. 2.10]} \\ &= \mathbf{P}_0(Z^{-1}X_{[k]}\mathbf{Q}\mathbf{Q}^*)\mathbf{Q} &&+ \mathbf{P}_0(Z^{-1}U_{[k]}\mathbf{Q}^*)\mathbf{Q} \\ &= X_{[k]}^{(1)}\mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)\mathbf{Q} &&+ U_{[k]}^{(1)}\mathbf{P}_0(Z^{-1}\mathbf{Q}^*)\mathbf{Q}, \end{aligned}$$

that is,

$$X_{[k+1]} = X_{[k]}^{(1)}\,\mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*) + U_{[k]}^{(1)}\,\mathbf{P}_0(Z^{-1}\mathbf{Q}^*).$$

Putting $A^{(1)} = \mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)$ and $B^{(1)} = \mathbf{P}_0(Z^{-1}\mathbf{Q}^*)$, *i.e.*, $A = \mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)^{(-1)}$ and $B = \mathbf{P}_0(Z^{-1}\mathbf{Q}^*)^{(-1)}$, gives the first part of the result. In the same way, $C = \mathbf{P}_0(\mathbf{Q}T)$ is derived via

$$\begin{aligned} \mathbf{X}_k\mathbf{C} = \mathbf{P}_0(\mathbf{X}_kT) &= \mathbf{P}_0(X_{[k]}\mathbf{Q}T) \\ &= X_{[k]}\,\mathbf{P}_0(\mathbf{Q}T). \end{aligned}$$

□

The theorem shows that knowledge of an orthonormal basis of the input state space is sufficient to construct A and B operators, and that knowledge of the response of the system to this basis gives the corresponding C operator. The realization corresponds to a factorization of H_T as

$$\cdot H_T = \mathbf{P}_0(\cdot\,\mathbf{Q}^*)\mathbf{P}(\mathbf{Q}T)$$

because $\mathbf{X}_k = \mathbf{P}_{\mathcal{H}}(U_{p(k)}) = X_{[k]}\mathbf{Q}$ with $X_{[k]} = \mathbf{P}_0(U_{p(k)}\mathbf{Q}^*)$, and $Y_{f(k)} = \mathbf{P}(\mathbf{X}_kT) = X_{[k]}\mathbf{P}(\mathbf{Q}T)$. The realization is uniformly controllable: $\Lambda_{\mathbf{Q}} = I$. Comparing with the factorization of

H_T obtained in section 3.3, it is seen that the operator $\mathbf{F}_0 = \mathbf{P}(\mathbf{Q}T)$ is the observability operator. $\mathbf{F}_0$ is one-to-one, because

$$\begin{aligned} D\mathbf{F}_0 = 0 \quad &\Leftrightarrow \quad \mathbf{P}(D\mathbf{Q}T) = 0 \\ &\Leftrightarrow \quad D\mathbf{Q} \in \mathcal{K} \\ &\Rightarrow \quad D = 0. \end{aligned}$$

Hence the canonical controller realization is observable and minimal. It is not necessarily uniformly observable: if it is, then the range of H_T is a closed subspace (proposition 3.13), which need not be the case.

Before deriving realizations of T for more general bases in $\mathcal{H}$, and for bases in $\mathcal{H}_0$, we first investigate the canonical controller realization in more detail.

The above definition of A is connected to the definition of $\mathbf{A}$ via the chosen basis as

$$D\mathbf{Q}\mathbf{A} = D^{(1)}A^{(1)}\mathbf{Q} \qquad (\text{any } D \in \mathcal{D}_2^{\mathcal{B}}).$$

From $\mathbf{A}^n = \mathbf{P}_{\mathcal{H}}(Z^{-n}\cdot)$, viz. equation (3.54), it follows by recursive application of the above expression that

$$D\mathbf{Q}\mathbf{A}^n = D^{(n)}A^{\{n\}}\mathbf{Q} \qquad (\text{any } D \in \mathcal{D}_2^{\mathcal{B}}),$$

where $A^{\{n\}} = A^{(n)} \cdots A^{(1)}$. Application of the projection formula of theorem 2.10, $\mathbf{P}_{\mathcal{H}}(\cdot) = \mathbf{P}_0(\cdot\,\mathbf{Q}^*)\mathbf{Q}$, results in

$$\begin{aligned} D\mathbf{Q}\mathbf{A}^n &= \mathbf{P}_{\mathcal{H}}(Z^{-n}D\mathbf{Q}) = \mathbf{P}_0(Z^{-n}D\mathbf{Q}\mathbf{Q}^*)\mathbf{Q} \\ &= D^{(n)}\mathbf{P}_0(Z^{-n}\mathbf{Q}\mathbf{Q}^*)\mathbf{Q} \\ &= D^{(n)}A^{\{n\}}\mathbf{Q} \end{aligned}$$

which yields, because $\mathbf{Q}$ is a strong basis representation and the relation is valid for any $D \in \mathcal{D}_2^{\mathcal{B}}$, the expression

$$A^{\{n\}} = \mathbf{P}_0(Z^{-n}\mathbf{Q}\mathbf{Q}^*) \qquad (n \geq 0). \tag{3.56}$$

Hence $\mathbf{Q}$ and A are closely connected. In particular, it will be possible to relate the stability properties of A to the boundedness of $\mathbf{Q}$, as is shown in the following lemma.

LEMMA 3.18. *Let $\mathbf{Q}$ be an orthonormal basis representation of $\mathcal{H}$, and let the operator $\mathbf{A}$ and A be as given in theorems 3.16 and 3.17. Then*

1. $\ell_A = r(\mathbf{A}) \leq \|\mathbf{A}\| = \|A\| \leq 1$,
2. *$\mathbf{Q}$ is bounded on $\mathcal{X}_2 \Leftrightarrow \ell_A < 1$.*

PROOF

1. Because the set of operators $\{X\mathbf{Q}\}$ spans $\overline{\mathcal{H}}$ for $X \in \mathcal{D}_2^{\mathcal{B}}$, and since $\mathbf{Q}$ is an orthonormal basis representation, $\mathbf{X} = X\mathbf{Q}$ satisfies $\|\mathbf{X}\|_{HS} = \|X\|_{HS} = \|X^{(n)}\|_{HS}$. Hence

$$\begin{aligned} \|(AZ)^n\| = \|A^{\{n\}}\| &= \sup_{\substack{X \in \mathcal{D}_2^{\mathcal{B}} \\ \|X\|_{HS} \le 1}} \|X^{(n)}A^{\{n\}}\|_{HS} \\ &= \sup_{\substack{X \in \mathcal{D}_2^{\mathcal{B}} \\ \|X\|_{HS} \le 1}} \|X^{(n)}A^{\{n\}}\mathbf{Q}\|_{HS} \\ &= \sup_{\substack{\mathbf{X} \in \overline{\mathcal{H}} \\ \|\mathbf{X}\|_{HS} \le 1}} \|\mathbf{X}\mathbf{A}^n\|_{HS} \\ &= \|\mathbf{A}^n\|. \end{aligned} \tag{3.57}$$

Consequently, $\ell_A = r(\mathbf{A})$. The relation $r(\mathbf{A}) \le \|\mathbf{A}\|$ is immediate, and (3.57) specializes to $\|\mathbf{A}\| = \|A\|$. Finally, $\mathbf{A}^n = \mathbf{P}_{\mathcal{H}}(Z^{-n}\cdot)$ implies $\|\mathbf{A}\| \le 1$, as remarked before.

2. Because we know already that $\ell_A \le 1$, the proof that $\mathbf{Q}$ is bounded if and only if $\ell_A < 1$ can consist of the two steps,

   (a) $\ell_A = 1 \Rightarrow$ the operator $[I + AZ + (AZ)^2 + \cdots]$ is unbounded on $\mathcal{D}_2^{\mathcal{B}}$,

   (b) $\mathbf{Q}$ bounded $\Rightarrow$ the operator $[I + AZ + (AZ)^2 + \cdots]$ is bounded on $\mathcal{D}_2^{\mathcal{B}}$.

   Proof of step 2a. By definition, $\ell_A = r(AZ) = \lim_{n\to\infty} \|(AZ)^n\|^{1/n}$. Because $\|AZ\| \le 1$, and $\|(AZ)^n\|^{1/n}$ monotonically decreases when n goes to infinity (*viz.* equation (2.18)),

$$\begin{aligned} \ell_A = 1 \quad &\Rightarrow \quad \|(AZ)^n\|^{1/n} = 1 \quad (\text{for all } n) \\ &\Rightarrow \quad \|(AZ)^n\| = 1 \quad (\text{for all } n) \\ &\Rightarrow \quad \sup_{\|D\|_{HS} \le 1} \|D(AZ)^n\|_{HS} = 1 \quad (\text{for all } n). \end{aligned} \tag{3.58}$$

Because $\|AZ\| \le 1$ implies $\|D(AZ)^{n-1}\|_{HS} \ge \|D(AZ)^n\|_{HS}$ for any $D \in \mathcal{D}_2$, we have from (3.58) that

$$\sup_D \sum_{k=0}^{n} \|D(AZ)^k\|_{HS}^2 \ge \sup_D n\|D(AZ)^n\|_{HS}^2 = n. \tag{3.59}$$

But since, for any n,

$$\sup_D \|D\left[I + AZ + (AZ)^2 + \cdots\right]\|_{HS}^2 \ge \sup_D \sum_{k=0}^{n} \|D(AZ)^k\|_{HS}^2,$$

it follows from (3.59) by taking the limit for $n \to \infty$, that the left-hand side of the above expression is equal to infinity. This proves that $[I + AZ + (AZ)^2 + \cdots]$ is unbounded on $\mathcal{D}_2$.

Proof of step 2b. If $\mathbf{Q}$ is a bounded operator, then for any choice of $D \in \mathcal{D}_2$, the operator $\mathbf{P}(D\mathbf{Q}\mathbf{Q}^*)$ is bounded and in $\mathcal{U}_2$. But, using (3.56), $\mathbf{P}(D\mathbf{Q}\mathbf{Q}^*)$ can be evaluated as

$$\begin{aligned}\mathbf{P}(D\mathbf{Q}\mathbf{Q}^*) &= \textstyle\sum_0^\infty Z^n \mathbf{P}_0(Z^{-n}D\mathbf{Q}\mathbf{Q}^*)\\ &= \textstyle\sum_0^\infty Z^n D^{(n)} \mathbf{P}_0(Z^{-n}\mathbf{Q}\mathbf{Q}^*)\\ &= \textstyle\sum_0^\infty Z^n D^{(n)} A^{\{n\}}\\ &= D \textstyle\sum_0^\infty (AZ)^n\\ &= D\left[I + AZ + (AZ)^2 + \cdots\right].\end{aligned}$$

Hence $\mathbf{Q}$ bounded implies that $[I + AZ + (AZ)^2 + \cdots]$ is bounded on $\mathcal{D}_2$. □

The following lemma also summarizes some material from section 3.3.

LEMMA 3.19. *The realization in theorem 3.17 has the following properties:*

1.
$$A^{\{n\}} = \mathbf{P}_0(Z^{-n}\mathbf{Q}\mathbf{Q}^*), \qquad (n \geq 0). \tag{3.60}$$

2.
$$\begin{aligned}&\text{On } \mathcal{X}_2: && \mathbf{P}_0(Z^{-1}\cdot\mathbf{Q}^*)^{(-1)} &&= \mathbf{P}_0(\cdot\,[\mathbf{Q}^*A + B])\\ &\text{On } \mathcal{D}_2^{\mathcal{B}}: && \cdot Z\mathbf{Q} &&= \cdot A^*\mathbf{Q} + \cdot B^*\\ &\text{On } \mathcal{X}_2: && \mathbf{P}_0(\cdot\,T) &&= \mathbf{P}_0(\cdot\,[D + \mathbf{Q}^*C])\\ &\text{On } \mathcal{X}_2: && \cdot T^* &&= \cdot D^* + \cdot C^*\mathbf{Q}.\end{aligned} \tag{3.61}$$

The property on $\mathcal{D}_2^{\mathcal{B}}$ can be extended to (dense domains in) $\mathcal{X}_2^{\mathcal{B}}$.

3. $\mathbf{Q}^*$ *has diagonals* $(\mathbf{Q}^*)_{[i]} = \mathbf{P}_0(Z^{-i}\mathbf{Q}^*)$ *given by*

$$(\mathbf{Q}^*)_{[i]} = \begin{cases} 0, & i \leq 0,\\ B^{(i)}A^{\{i-1\}}, & i > 0.\end{cases} \tag{3.62}$$

so that $\mathbf{Q}^*$ satisfies the relation $\mathbf{Q}^*Z^{-1} = \mathbf{Q}^*A + B$. The operator $\mathbf{P}_0(\cdot\,\mathbf{Q}^*)$ is bounded on $\mathcal{X}_2$ and has a diagonal representation, in the sense of (2.22), given by

$$\mathbf{P}_0(\cdot\,\mathbf{Q}^*) = \mathbf{P}_0(\cdot\,[BZ + BZAZ + BZ(AZ)^2 + \cdots]),$$

4. $A^*A + B^*B = I$.
5. *If $\ell_A < 1$ then the above equations simplify to $\mathbf{Q}^* = BZ(I - AZ)^{-1}$, so that $\mathbf{Q}^* \in Z\mathcal{U}$ is itself bounded on $\mathcal{X}_2$, and $\mathbf{Q} \in \mathcal{L}Z^{-1}$ is bounded on $\mathcal{X}_2$, too.*

PROOF

1. This is just equation (3.56).

2. Substituting $X_{[k]} = \mathbf{P}_0(Z^{-k}U\mathbf{Q}^*)$ and $X_{[k+1]} = \mathbf{P}_0(Z^{-k-1}U\mathbf{Q}^*)$ into the relation $X_{[k+1]}^{(-1)} = X_{[k]}A + U_{[k]}B$ yields $\mathbf{P}_0(Z^{-k-1}U\mathbf{Q}^*)^{(-1)} = \mathbf{P}_0(Z^{-k}U\mathbf{Q}^*)A + \mathbf{P}_0(Z^{-k}U)B$ for all $U \in \mathcal{X}_2$, and hence
$$\mathbf{P}_0(Z^{-1} \cdot \mathbf{Q}^*)^{(-1)} = \mathbf{P}_0(\cdot\,[\mathbf{Q}^*A + B]). \tag{3.63}$$
By taking adjoints (*cf.* equation (2.37)), it is seen that $\mathbf{Q}$ satisfies the relation $Z\mathbf{Q} = A^*\mathbf{Q} + B^*$ on $\mathcal{D}_2^{\mathcal{B}}$, but by extension everywhere on $\mathcal{X}_2^{\mathcal{B}}$ where it is bounded. From the factorization $\cdot H_T = \mathbf{P}(\cdot\, T)\big|_{\mathcal{L}_2Z^{-1}} = \mathbf{P}_0(\cdot\,\mathbf{Q}^*)\mathbf{F}_0$, where $\mathbf{P}_0(\mathbf{F}_0) = C$, and $\mathbf{P}_0(T) = D$, the remaining two equations follow.

3. The above result implies $\mathbf{P}_0(Z^{-k-1}\mathbf{Q}^*)^{(-1)} = \mathbf{P}_0(Z^{-k}\mathbf{Q}^*)A + \mathbf{P}_0(Z^{-k}B)$ which is equivalent to
$$(\mathbf{Q}^*)_{[k+1]}^{(-1)} = (\mathbf{Q}^*)_{[k]}A + \mathbf{P}_0(Z^{-k}B) \tag{3.64}$$
where of course $\mathbf{P}_0(Z^{-k}B) = B$ $(k = 0)$; $\mathbf{P}_0(Z^{-k}B) = 0$ $(k \neq 0)$. Since $\mathbf{Q}$ is a basis representation for a subspace in $\mathcal{L}_2Z^{-1}$, it follows directly that $(\mathbf{Q}^*)_{[k]} = 0$, $(k \leq 0)$. Evaluating (3.64) for $k = 0, 1, \cdots$ leads to

$$\begin{array}{lllllll}
(\mathbf{Q}^*)_{[1]}^{(-1)} & = & B & \Rightarrow & (\mathbf{Q}^*)_{[1]} & = & B^{(1)} \\
(\mathbf{Q}^*)_{[2]}^{(-1)} & = & (\mathbf{Q}^*)_{[1]}A & \Rightarrow & (\mathbf{Q}^*)_{[2]} & = & B^{(2)}A^{(1)} \\
(\mathbf{Q}^*)_{[k]}^{(-1)} & = & (\mathbf{Q}^*)_{[k-1]}A\ (k \geq 1) & \Rightarrow & (\mathbf{Q}^*)_{[k]} & = & B^{(k)}A^{(k-1)} \cdots A^{(1)} = B^{(k)}A^{\{k-1\}}.
\end{array}$$

4. Combining the first two expressions in equation (3.63) and using the orthonormality of the basis, $\Lambda_{\mathbf{Q}} = I$, gives
$$\begin{aligned}
I = \Lambda_{\mathbf{Q}}^{(-1)} &= \mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*)^{(-1)} \\
&= \mathbf{P}_0(\,[A^*\mathbf{Q} + B^*]\,[\mathbf{Q}^*A + B]\,) \\
&= A^*\,\mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*)A + B^*B \\
&= A^*\Lambda_{\mathbf{Q}}A + B^*B \\
&= A^*A + B^*B.
\end{aligned}$$

5. From (3.62), it is seen that we can write
$$\begin{aligned}
\mathbf{Q}^* &= \textstyle\sum_1^\infty Z^{[i]}(\mathbf{Q}^*)_{[i]} \\
&= \textstyle\sum_1^\infty Z^{[i]}B^{(i)}A^{\{i-1\}} \\
&= BZ + BZAZ + BZ(AZ)^2 + \cdots
\end{aligned}$$
if this sum converges. If $\ell_A < 1$, then the sum $I + AZ + (AZ)^2 + \cdots$ converges and is equal to $(I - AZ)^{-1}$, so that in this case $\mathbf{Q}^* = BZ(I - AZ)^{-1}$. At the same time, this shows that $\mathbf{Q}^*$ and hence $\mathbf{Q}$ are bounded operators. □

Property 3 is the same as equation (3.31). It has been derived here based on the definitions of A and B in terms of $\mathbf{Q}$.

Related realizations can be derived if a different, possibly non-orthogonal basis in $\overline{\mathcal{H}}$ is chosen. For canonical results, we have to require that this alternative basis is a strong basis. The realization that is obtained in this case is linked to the realization based on $\mathbf{Q}$ via an invertible state transformation.

THEOREM 3.20. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a given transfer operator with input state space $\mathcal{H}$ of locally finite dimensions. Let $d = \text{s-dim}(\mathcal{H})$, and put $\mathcal{B} = \mathbb{C}^d$.*

If $\mathbf{F}$ is a strong basis representation of $\mathcal{H}$, such that $\Lambda_{\mathbf{F}} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) \gg 0$ and $\Lambda_{\mathbf{F}} < \infty$, then T has a state realization

$$\begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} \Lambda_{\mathbf{F}}^{-1}\mathbf{P}_0(Z^{-1}\mathbf{F}\mathbf{F}^*)^{(-1)} & \Lambda_{\mathbf{F}}^{-1}\mathbf{P}_0(\mathbf{F}T) \\ \mathbf{P}_0(Z^{-1}\mathbf{F}^*)^{(-1)} & \mathbf{P}_0(T) \end{bmatrix}$$

$$\begin{array}{ll} A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)}) & C \in \mathcal{D}(\mathcal{B}, \mathcal{N}) \\ B \in \mathcal{D}(\mathcal{M}, \mathcal{B}^{(-1)}) & D \in \mathcal{D}(\mathcal{M}, \mathcal{N}). \end{array}$$

PROOF The realization follows from theorem 3.16 in the same way as the realization in theorem 3.17 was derived, but now with the projector onto $\overline{\mathcal{H}}$ written in terms of $\mathbf{F}$: $\mathbf{P}_{\mathcal{H}}(\cdot) = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\Lambda_{\mathbf{F}}^{-1}\mathbf{F}$ (*viz.* equation (2.39)), and the choice of $X_{[k]} = \mathbf{P}_0(\mathbf{X}_k\mathbf{F}^*)$ so that $\mathbf{X}_k = X_{[k]}\Lambda_{\mathbf{F}}^{-1}\mathbf{F}$. (Rest of proof omitted.) □

When $\mathbf{F}$ is written in terms of an orthonormal basis representation $\mathbf{Q}$ of $\overline{\mathcal{H}}$,

$$\begin{aligned} \mathbf{F} &= R^*\mathbf{Q} \\ \Lambda_{\mathbf{F}} &= \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) = R^*R \end{aligned}$$

(where $R \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ is a boundedly invertible factor of $\Lambda_{\mathbf{F}}$), then the above realization based on $\mathbf{F}$ can be 'normalized' to obtain the realization based on $\mathbf{Q}$ via a state transformation $X \to X'R$, where X' is a state in the realization based on $\mathbf{Q}$. That this is the case is readily verified by the derivation

$$\begin{aligned} \mathbf{X}_k = X'_{[k]}\mathbf{Q} &= X_{[k]}R^{-1}\mathbf{Q} \\ &= X_{[k]}R^{-1}R^{-*}\mathbf{F} \\ &= X_{[k]}\Lambda_{\mathbf{F}}^{-1}\mathbf{F} \\ &= \mathbf{P}_0(\mathbf{X}_k\mathbf{F}^*)\Lambda_{\mathbf{F}}^{-1}\mathbf{F}. \end{aligned}$$

It can also be verified on the realizations. Let $\{A', B', C', D\}$ be the realization based on $\mathbf{Q}$ in theorem 3.17, then *e.g.*,

$$\begin{aligned} A' = RAR^{-(-1)} &= R\,\Lambda_{\mathbf{F}}^{-1}\,\mathbf{P}_0(Z^{-1}\mathbf{F}\mathbf{F}^*)^{(-1)}R^{-(-1)} \\ &= R\,(R^{-1}R^{-*})\,\mathbf{P}_0(Z^{-1}\mathbf{F}\mathbf{F}^*)^{(-1)}R^{-(-1)} \\ &= \mathbf{P}_0(Z^{-1}R^{-*}\mathbf{F}\mathbf{F}^*R^{-1})^{(-1)} \\ &= \mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)^{(-1)}. \end{aligned}$$

The realization based on the basis representation $\mathbf{F}$ of $\mathcal{H}$ provides a factorization of H_T into

$$H_T = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}T).$$

The realization is uniformly controllable by construction: the controllability Gramian is $\Lambda_{\mathbf{F}} \gg 0$. Writing $\mathbf{F}_0 = \Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}T)$, the fact that $\mathbf{F}_0$ is one-to-one on $\mathcal{D}_2^{\mathcal{B}}$ is proven in the same way as done for the realization based on $\mathbf{Q}$, and hence the realization is observable and minimal.

Using the state transformation by R, it is straightforward to derive the following equivalent to lemma 3.19.

LEMMA 3.21. *The realization in theorem 3.17 has the following properties:*

1. $A^{\{n\}} = \Lambda_{\mathbf{F}}^{-(n)}\mathbf{P}_0(Z^{-n}\mathbf{F}\mathbf{F}^*), \qquad (n \geq 0).$
2. $\ell_A \leq 1.$
3. $$\begin{array}{lrcl} \text{On } \mathcal{X}_2: & \mathbf{P}_0(Z^{-1}\cdot\mathbf{F}^*)^{(-1)} & = & \mathbf{P}_0(\cdot\,[\mathbf{F}^*A + B]), \\ \text{on } \mathcal{D}_2^{\mathcal{B}}: & \cdot Z\mathbf{F} & = & \cdot A^*\mathbf{F} + \cdot B^*, \\ \text{on } \mathcal{X}_2: & \mathbf{P}_0(\cdot\,T) & = & \mathbf{P}_0(\cdot\,[D + \mathbf{F}^*C]), \\ \text{on } \mathcal{X}_2: & \cdot T^* & = & \cdot D^* + \cdot C^*\mathbf{F}. \end{array}$$

   The property on $\mathcal{D}_2$ can be extended to (dense domains in) $\mathcal{X}_2$.
4. $\mathbf{F}^*$ *has diagonals given by*

$$(\mathbf{F}^*)_{[i]} = \begin{cases} 0 & i \leq 0, \\ B^{(i)}A^{\{i-1\}} & i > 0. \end{cases}$$

   so that $\mathbf{F}$ *satisfies the relation* $\mathbf{F}^*Z^{-1} = \mathbf{F}^*A + B$.
5. $A^*\Lambda_{\mathbf{F}}A + B^*B = \Lambda_{\mathbf{F}}^{(-1)}.$
6. *If* $\ell_A < 1$ *then* $\mathbf{F}^*$ *satisfies* $\mathbf{F}^* = BZ(I - AZ)^{-1}$, *so that* $\mathbf{F}^* \in Z\mathcal{U}$ *is a bounded operator, and* $\mathbf{F} \in \mathcal{L}Z^{-1}$ *is also bounded.*

Canonical observer realizations

In the previous section, we defined the state $\mathbf{X}_k$ at point k to be the projection of the 'past input' $U_{p(k)} = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(Z^{-k}U)$ onto the input state space $\overline{\mathcal{H}}$. Selecting an orthonormal basis or another strong basis produced a canonical realization which we have called the canonical controller realizations because they are defined at the input side of the transformation by the system T, that is, the side at which the state is controlled. It is possible to derive realizations based on the definition of state at the output side of the system, which will give canonical realizations in the observer form (in which the state is observed at the

output). To this end, define the state $\mathbf{X}_k$ to be the projection of the past input, after transformation by T, onto the output state space $\overline{\mathcal{H}}_0$:

$$\mathbf{X}_k = \mathbf{P}(U_{p(k)}T) \quad \in \mathcal{H}_0 . \tag{3.65}$$

THEOREM 3.22. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a given transfer operator with output state space $\mathcal{H}_0$. With $U \in \mathcal{X}_2^{\mathcal{M}}$, $Y \in \mathcal{X}_2^{\mathcal{N}}$ having diagonals $U_{[k]} = \mathbf{P}_0(Z^{-k}U)$, $Y_k = \mathbf{P}_0(Z^{-k}Y)$, let $Y_{[k]} \in \mathcal{D}_2^{\mathcal{N}}$ and $\mathbf{X}_k \in \mathcal{U}^N$ be given by the operator state equations*

$$\begin{bmatrix} \mathbf{X}_{k+1} &= \mathbf{X}_k\mathbf{A} + U_{[k]}\mathbf{B} \\ Y_{[k]} &= \mathbf{X}_k\mathbf{C} + U_{[k]}\mathbf{D} \end{bmatrix} \tag{3.66}$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are bounded operators satisfying

$$\begin{bmatrix} \mathbf{A} & \mathbf{C} \\ \mathbf{B} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{P}(Z^{-1}\cdot) & \mathbf{P}_0(\cdot) \\ \mathbf{P}(Z^{-1}\cdot T) & \mathbf{P}_0(\cdot T) \end{bmatrix} \qquad \begin{aligned} \mathbf{A}&: \ \overline{\mathcal{H}}_0 \to \overline{\mathcal{H}}_0 \qquad \mathbf{C}: \ \overline{\mathcal{H}}_0 \to \mathcal{D}_2^{\mathcal{N}} \\ \mathbf{B}&: \ \mathcal{D}_2^{\mathcal{M}} \to \overline{\mathcal{H}}_0 \qquad \mathbf{D}: \ \mathcal{D}_2^{\mathcal{M}} \to \mathcal{D}_2^{\mathcal{N}} \end{aligned}$$

Then $Y = UT$, i.e., $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ forms an operator realization of T.

PROOF First assume that $Y = UT$ and $\mathbf{X}_k$ is given by (3.65). Then the realization equations (3.66) are implied:

1. $$\begin{aligned} \mathbf{X}_{k+1} &= \mathbf{P}(U_{p(k+1)}T) \\ &= \mathbf{P}(\mathbf{P}_{\mathcal{L}_2Z^{-1}}(Z^{-k-1}U)T) \\ &= \mathbf{P}([Z^{-1}\mathbf{P}_{\mathcal{L}_2Z^{-1}}(Z^{-k}U) + Z^{-1}U_{[k]}]T) \\ &= \mathbf{P}(Z^{-1}U_{p(k)}T \quad + \quad Z^{-1}U_{[k]}T) \\ &= \mathbf{P}(Z^{-1}U_{p(k)}T) \quad + \quad \mathbf{P}(Z^{-1}U_{[k]}T) \\ &= \mathbf{P}\left(Z^{-1}\mathbf{P}(U_{p(k)}T)\right) \quad + \quad \mathbf{P}(Z^{-1}U_{[k]}T) \\ &= \mathbf{P}(Z^{-1}\mathbf{X}_k) \quad + \quad \mathbf{P}(Z^{-1}U_{[k]}T) \\ &=: \mathbf{X}_k\mathbf{A} \quad + \quad U_{[k]}\mathbf{B}. \end{aligned}$$

   Note: if $\mathbf{X}_k \in \overline{\mathcal{H}}_0$, then $\mathbf{X}_k\mathbf{A} = \mathbf{P}(Z^{-1}\mathbf{X}_k) \in \overline{\mathcal{H}}_0$ because of the shift-invariance property of $\overline{\mathcal{H}}_0$ (lemma 3.15).

2. $$\begin{aligned} Y_{[k]} &= \mathbf{P}_0(Z^{-k}Y) \\ &= \mathbf{P}_0(Z^{-k}UT) \\ &= \mathbf{P}_0(U_{p(k)}T) + \mathbf{P}_0(U_{[k]}T) \\ &= \mathbf{P}_0(\mathbf{X}_k) + U_{[k]}\mathbf{P}_0(T) \\ &=: \mathbf{X}_k\mathbf{C} + U_{[k]}\mathbf{D}. \end{aligned}$$

Let $Y_{[k]}$ be as defined in the state equations (3.66). To prove that $\{\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}\}$ forms indeed an operator realization of T, we first show

$$\begin{aligned} \mathbf{A} &= \mathbf{P}(Z^{-1}\cdot) \\ \mathbf{A}^2 &= \mathbf{P}(Z^{-1}\mathbf{P}(Z^{-1}\cdot)) = \mathbf{P}(Z^{-2}\cdot) \\ \Rightarrow \quad \mathbf{A}^n &= \mathbf{P}(Z^{-n}\cdot). \end{aligned}$$

so that $\mathbf{B}\mathbf{A}^{n-1} = \mathbf{P}\left(Z^{-n+1}\mathbf{P}(Z^{-1}\cdot T)\right) = \mathbf{P}(Z^{-n}\cdot T)$. Hence $\mathbf{X}_k$ given by the state equations is

$$\begin{aligned}\mathbf{X}_k &= \textstyle\sum_{n=1}^{\infty} U_{[k-n]}\mathbf{B}\mathbf{A}^{n-1} \\ &= \mathbf{P}\left(\textstyle\sum_{n=1}^{\infty} Z^{-n}U_{[k-n]}T\right) \\ &= \mathbf{P}\left(U_{p(k)}T\right),\end{aligned}$$

according to the definition (3.65). With this $\mathbf{X}_k$, $Y_{[k]}$ given by the state equations is seen to be equal to $\mathbf{P}_0(Z^{-k}Y)$ by reversing the line of equalities in the second item above. □

THEOREM 3.23. (CANONICAL OBSERVER REALIZATION) *Let $T \in \mathcal{U}(\mathcal{M},\mathcal{N})$ be a given transfer operator with output state space $\mathcal{H}_0$ of locally finite dimensions. Let d = s-dim$(\mathcal{H}_0)$, and put $\mathcal{B} = \mathbb{C}^d$. Let $\mathbf{G}$ be an orthonormal basis representation of $\overline{\mathcal{H}}_0$: $\overline{\mathcal{H}}_0 = \mathcal{D}_2^{\mathcal{B}}\mathbf{G}$, $\Lambda_{\mathbf{G}} = I$. Take $U \in \mathcal{X}_2^{\mathcal{M}}$, $Y \in \mathcal{X}_2^{\mathcal{N}}$ with diagonals $U_{[k]} = \mathbf{P}_0(Z^{-k}U)$ and $Y_{[k]} = \mathbf{P}_0(Z^{-k}Y)$.*

Then

$$Y = UT \quad \Leftrightarrow \quad \begin{bmatrix} X_{[k+1]}^{(-1)} &= X_{[k]}A + U_{[k]}B \\ Y_{[k]} &= X_{[k]}C + U_{[k]}D, \end{bmatrix} \tag{3.67}$$

where

$$\begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0(Z^{-1}\mathbf{G}\mathbf{G}^*)^{(-1)} & \mathbf{P}_0(\mathbf{G}) \\ \mathbf{P}_0(Z^{-1}T\mathbf{G}^*)^{(-1)} & \mathbf{P}_0(T) \end{bmatrix} \qquad \begin{matrix} A \in \mathcal{D}(\mathcal{B},\mathcal{B}^{(-1)}) & C \in \mathcal{D}(\mathcal{B},\mathcal{N}) \\ B \in \mathcal{D}(\mathcal{M},\mathcal{B}^{(-1)}) & D \in \mathcal{D}(\mathcal{M},\mathcal{N}). \end{matrix}$$

PROOF The proof follows closely that of theorem 3.17. For a given $\mathbf{X}_k$ in $\overline{\mathcal{H}}_0$, put $\mathbf{X}_k = X_{[k]}\mathbf{G}$, for $X_{[k]} \in \mathcal{D}_2^{\mathcal{B}}$. Then

$$\begin{aligned}\mathbf{X}_{k+1} = X_{[k+1]}\mathbf{G} &= \mathbf{P}(Z^{-1}\mathbf{X}_k) &&+ \mathbf{P}(Z^{-1}U_{[k]}T) \\ &= \mathbf{P}_{\mathcal{H}_0}(Z^{-1}\mathbf{X}_k) &&+ \mathbf{P}_{\mathcal{H}_0}(Z^{-1}U_{[k]}T) \\ &= \mathbf{P}_{\mathcal{H}_0}(Z^{-1}X_{[k]}\mathbf{G}) &&+ \mathbf{P}_{\mathcal{H}_0}(Z^{-1}U_{[k]}T) \\ &= \mathbf{P}_0(Z^{-1}X_{[k]}\mathbf{G}\mathbf{G}^*)\mathbf{G} &&+ \mathbf{P}_0(Z^{-1}U_{[k]}T\mathbf{G}^*)\mathbf{G} \\ &= X_{[k]}^{(1)}\mathbf{P}_0(Z^{-1}\mathbf{G}\mathbf{G}^*)\mathbf{G} &&+ U_{[k]}^{(1)}\mathbf{P}_0(Z^{-1}T\mathbf{G}^*)\mathbf{G}.\end{aligned}$$

Hence $A = \mathbf{P}_0(Z^{-1}\mathbf{G}\mathbf{G}^*)^{(-1)}$ and $B = \mathbf{P}_0(Z^{-1}T\mathbf{G}^*)^{(-1)}$. In the same way,

$$\begin{aligned}\mathbf{P}_0(\mathbf{X}_k) &= \mathbf{P}_0(X_{[k]}\mathbf{G}) \\ &= X_{[k]}\mathbf{P}_0(\mathbf{G}),\end{aligned}$$

hence $C = \mathbf{P}_0(\mathbf{G})$. □

The factorization of H_T that corresponds to this realization is given by

$$H_T = \mathbf{P}_0(\cdot\, T\mathbf{G}^*)\mathbf{G}$$

because $Y_{f(k)} = \mathbf{P}(U_{p(k)}T) = \mathbf{X}_k = \mathbf{P}_{\mathcal{H}_0}(\mathbf{X}_k) = \mathbf{P}_0(\mathbf{X}_k\mathbf{G}^*)\mathbf{G}$, and hence

$$\begin{aligned} & Y_{f(k)} = U_{p(k)}H_T \\ \Leftrightarrow \quad & \begin{cases} X_{[k]} &= \mathbf{P}_0(\mathbf{X}_k\mathbf{G}^*) = \mathbf{P}_0(\mathbf{P}(U_{p(k)}T)\mathbf{G}^*) \\ &= \mathbf{P}_0(U_{p(k)}T\mathbf{G}^*) \\ Y_{f(k)} &= X_{[k]}\mathbf{G}. \end{cases} \end{aligned}$$

The observability Gramian is equal to $\Lambda_{\mathbf{G}} = I$. The controllability operator $\mathbf{F}^*$ is given by $\mathbf{F}^* = \mathbf{P}_{\mathcal{Z}l}(T\mathbf{G}^*)$, or $\mathbf{F} = \mathbf{P}_{\mathcal{L}\mathcal{Z}^{-1}}(\mathbf{G}T^*)$. Its kernel $\ker \mathbf{F} = 0$ because, for any $D \in \mathcal{D}_2^{\mathcal{B}}$,

$$\begin{aligned} D\mathbf{F} = 0 \quad & \Leftrightarrow \quad \mathbf{P}_{\mathcal{L}_2\mathcal{Z}^{-1}}(D\mathbf{G}T^*) = 0 \\ & \Leftrightarrow \quad D\mathbf{G} \in \mathcal{K}_0 \\ & \Rightarrow \quad D = 0. \end{aligned}$$

Hence, the realization is controllable (but not necessarily uniformly).

LEMMA 3.24. *The realization in theorem 3.23 has the following properties:*

1. $\|A\| \leq 1$, $\quad \ell_A \leq 1$.
2. $\mathbf{G}$ *has diagonals* $\mathbf{G}_{[i]} = \mathbf{P}_0(Z^{-i}\mathbf{G})$ *given by*

$$\mathbf{G}_{[i]} = \begin{cases} 0 & i < 0, \\ A^{\{i\}}C & i \geq 0. \end{cases} \tag{3.68}$$

$\mathbf{G}$ *satisfies the relations*

$$\begin{aligned} \text{on } \mathcal{D}_2: \quad & \cdot\mathbf{G} &&= \cdot C + \cdot AZ\mathbf{G}, \\ \text{on } \mathcal{X}_2: \quad & \mathbf{P}_0(\cdot\,\mathbf{G}^*) &&= \mathbf{P}_0(Z^{-1}\cdot\mathbf{G}^*)^{(-1)}A^* + \mathbf{P}_0(\cdot\,C^*), \\ \text{on } \mathcal{X}_2: \quad & \mathbf{P}_0(\cdot\,T^*) &&= \mathbf{P}_0(\cdot\,D^*) + \mathbf{P}_0(Z^{-1}\cdot\mathbf{G}^*)^{(-1)}B^*, \\ \text{on } \mathcal{X}_2: \quad & \cdot T &&= \cdot D + \cdot BZ\mathbf{G}. \end{aligned}$$

The property on $\mathcal{D}_2$ *can be extended to (dense domains in)* $\mathcal{X}_2$. *In the sense of (2.22),* $\mathbf{G}$ *has (on* $\mathcal{D}_2$*) a diagonal representation*

$$\mathbf{G} = C + AZC + (AZ)^2C + \cdots.$$

3. $AA^* + CC^* = I$.
4. *If* $\ell_A < 1$ *then* $\mathbf{G} = (I - AZ)^{-1}C$, *so that* $\mathbf{G} \in \mathcal{U}$ *is a bounded operator.*

PROOF

1. $\|A\| \leq 1$ is proven as in lemma 3.19.

2. According to lemma 3.15, $\mathbf{P}(Z^{-1}\overline{\mathcal{H}}_0) \subset \overline{\mathcal{H}}_0$. Hence, for any $D \in \mathcal{D}_2$,

$$\begin{aligned} \mathbf{P}(Z^{-1}D\mathbf{G}) &= \mathbf{P}_{\mathcal{H}_0}(Z^{-1}D\mathbf{G}) \\ &= \mathbf{P}_0(Z^{-1}D\mathbf{G}\mathbf{G}^*)\,\mathbf{G} \\ &= D^{(1)}\mathbf{P}_0(Z^{-1}\mathbf{G}\mathbf{G}^*)\,\mathbf{G} \\ &= D^{(1)}A^{(1)}\mathbf{G}. \end{aligned} \tag{3.69}$$

Likewise,

$$\begin{aligned} \mathbf{P}(Z^{-n}D\mathbf{G}) &= \mathbf{P}(Z^{-n+1}\mathbf{P}(Z^{-1}D\mathbf{G})) \\ &= \mathbf{P}(Z^{-n+1}D^{(1)}A^{(1)}\mathbf{G}) \\ &= \cdots = D^{(n)}A^{\{n\}}\mathbf{G}. \end{aligned}$$

Since we know already that $\mathbf{P}_0(\mathbf{G}) = C$, it follows that

$$\begin{aligned} \mathbf{G}_{[n]} = \mathbf{P}_0(Z^{-n}\mathbf{G}) &= A^{\{n\}}\mathbf{P}_0(\mathbf{G}) \\ &= A^{\{n\}}C. \end{aligned}$$

From (3.69) it is also inferred that

$$\begin{aligned} D\mathbf{G} &= \mathbf{P}_0(D\mathbf{G}) + Z\mathbf{P}(Z^{-1}D\mathbf{G}) \\ &= D\mathbf{P}_0(\mathbf{G}) + ZD^{(1)}A^{(1)}\mathbf{G} \\ &= D\,[C + AZ\mathbf{G}]\,, \end{aligned}$$

so that $\mathbf{G} = C + AZ\mathbf{G}$. Taking the adjoint of the above expression (*cf.* equation (2.37)) leads to

$$\mathbf{P}_0(\cdot\,\mathbf{G}^*) = \mathbf{P}_0(Z^{-1}\cdot\,\mathbf{G}^*)^{(-1)}A^* + \mathbf{P}_0(\cdot\,C^*).$$

The remaining two equations are proven in the same way as in lemma 3.19.

3. Inserting $\mathbf{G} = C + AZ\mathbf{G}$ into the expression for its adjoint leads to

$$\begin{aligned} I = \Lambda_{\mathbf{G}} &= \mathbf{P}_0(\mathbf{G}\mathbf{G}^*) \\ &= \mathbf{P}_0(Z^{-1}AZ\mathbf{G}\mathbf{G}^*)^{(-1)}A^* + \mathbf{P}_0(CC^*) \\ &= A\mathbf{P}_0(\mathbf{G}\mathbf{G}^*)^{(-1)}A^* + CC^* \\ &= A\Lambda_{\mathbf{G}}^{(-1)}A^* + CC^* \\ &= AA^* + CC^*. \end{aligned}$$

4. If $\ell_A < 1$, then $(I - AZ)$ is invertible in $\mathcal{U}$, so that

$$\mathbf{G} = C + AZ\mathbf{G} \quad \Leftrightarrow \quad (I - AZ)\mathbf{G} = C \quad \Leftrightarrow \quad \mathbf{G} = (I - AZ)^{-1}C.$$

□

THEOREM 3.25. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a given transfer operator with output state space $\mathcal{H}_0$ of locally finite dimensions. Let $d = \text{s-dim}(\mathcal{H}_0)$, and put $\mathcal{B} = \mathbb{C}^d$.*

If $\mathbf{F}_0$ is a strong basis representation of $\overline{\mathcal{H}}_0$, such that $\Lambda_{\mathbf{F}_0} = \mathbf{P}_0(\mathbf{F}_0\mathbf{F}_0^*) \gg 0$ and $\Lambda_{\mathbf{F}_0} < \infty$, then T has a state realization

$$\begin{bmatrix} A & C \\ B & D \end{bmatrix} = \begin{bmatrix} \mathbf{P}_0(Z^{-1}\mathbf{F}_0\mathbf{F}_0^*)^{(-1)}\Lambda_{\mathbf{F}_0}^{-(-1)} & \mathbf{P}_0(\mathbf{F}_0) \\ \mathbf{P}_0(Z^{-1}T\mathbf{F}_0^*)^{(-1)}\Lambda_{\mathbf{F}_0}^{-(-1)} & \mathbf{P}_0(T) \end{bmatrix}$$

$$\begin{aligned} A &\in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)}) & C &\in \mathcal{D}(\mathcal{B}, \mathcal{N}) \\ B &\in \mathcal{D}(\mathcal{M}, \mathcal{B}^{(-1)}) & D &\in \mathcal{D}(\mathcal{M}, \mathcal{N}). \end{aligned}$$

PROOF The proof follows from theorem 3.23 and follows the lines of the proof of theorem 3.20, with state transformation $X = X'R$, and an orthogonal basis $\mathbf{G}$ such that $\mathbf{F}_0 = R\mathbf{G}$. □

The factorization of H_T corresponding to this realization is

$$H_T = \mathbf{P}_0(\cdot\, T\mathbf{F}_0^*)\Lambda_{\mathbf{F}_0}^{-1}\mathbf{F}_0.$$

The realization is uniformly observable by construction: the observability Gramian is $\Lambda_{\mathbf{F}_0} \gg 0$. The controllability operator is given by $\mathbf{F}^* = \mathbf{P}_{Z\mathcal{U}}(T\mathbf{F}_0^*)\Lambda_{\mathbf{F}_0}^{-1}$; the fact that $\mathbf{F}$ is one-to-one on $\mathcal{D}_2^{\mathcal{B}}$ is proven in the same way as before, and hence the realization is controllable and minimal.

LEMMA 3.26. *The realization in theorem 3.25 has the following properties:*

1. $\ell_A \leq 1$.
2. $\mathbf{F}_0$ *has diagonals* $(\mathbf{F}_0)_{[i]} = \mathbf{P}_0(Z^{-i}\mathbf{F}_0)$ *given by*

$$\mathbf{F}_{0[i]} = \begin{cases} 0 & i < 0, \\ A^{\{i\}}C & i \geq 0, \end{cases} \tag{3.70}$$

so that

$$\begin{aligned} &\text{on } \mathcal{D}_2^{\mathcal{B}}: && \cdot\,\mathbf{F}_0 &&= \cdot\,C + \cdot\,AZ\mathbf{F}_0, \\ &\text{on } \mathcal{X}_2: && \mathbf{P}_0(\cdot\,\mathbf{F}_0^*) &&= \mathbf{P}_0(Z^{-1}\cdot\,\mathbf{F}_0^*)^{(-1)}A^* + \mathbf{P}_0(\cdot\,C^*), \\ &\text{on } \mathcal{X}_2: && \mathbf{P}_0(\cdot\,T^*) &&= \mathbf{P}_0(\cdot\,D^*) + \mathbf{P}_0(Z^{-1}\cdot\,\mathbf{F}_0^*)^{(-1)}B^*, \\ &\text{on } \mathcal{X}_2: && \cdot\,T &&= \cdot\,D + \cdot\,BZ\mathbf{F}_0. \end{aligned}$$

The property on $\mathcal{D}_2^{\mathcal{B}}$ can be extended to (dense domains in) $\mathcal{X}_2^{\mathcal{B}}$.

3. $A\Lambda_{\mathbf{F}_0}^{(-1)}A^* + CC^* = \Lambda_{\mathbf{F}_0}$.
4. *If* $\ell_A < 1$, *then* $\mathbf{F}_0 = (I - AZ)^{-1}C$.

Connection between controller and observer realizations

Because the canonical controller and observer realizations both provide a factorization of the Hankel operator H_T, there is a connection between the two representations.

THEOREM 3.27. *Given a bounded system transfer operator $T \in \mathcal{U}$ with finite dimensional state spaces $\overline{\mathcal{H}}$ and $\overline{\mathcal{H}}_0$, let $\mathbf{F}$ be the representation of a strong basis in $\overline{\mathcal{H}}$. Put*

$$\mathbf{F}_0 = \Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}T)$$

and suppose that $\mathbf{F}_0$ represents a strong basis ($\Lambda_{\mathbf{F}_0} \gg 0$). Then the canonical realization based on $\mathbf{F}$ (theorem 3.20) is identical to the canonical realization based on $\mathbf{F}_0$ (theorem 3.25).

PROOF Let X be the state of the canonical realization based on $\mathbf{F}$, and X' be that of $\mathbf{F}_0$. We will prove that, when $\mathbf{F}_0 = \Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}T)$, these states are the same. The proof hinges on the fact that $\mathbf{P}(U_{p(k)}T) = \mathbf{P}(\,\mathbf{P}_{\mathcal{H}}(U_{p(k)})\,T\,)$ by definition of $\mathcal{H}$. Let

$$\begin{bmatrix} \mathbf{X}_k &=& \mathbf{P}_{\mathcal{H}}(U_{p(k)}) \\ \mathbf{X}'_k &=& \mathbf{P}(U_{p(k)}T) \end{bmatrix}, \qquad \begin{bmatrix} \mathbf{X}_k &=& X_{[k]}\Lambda_{\mathbf{F}}^{-1}\mathbf{F} \\ \mathbf{X}'_k &=& X'_{[k]}\mathbf{F}_0 \end{bmatrix}$$

(according to the definitions (3.51), (3.65) leading to theorems 3.20 and 3.25). Then

$$\begin{aligned} \mathbf{X}'_k &= \mathbf{P}(U_{p(k)}T) \\ &= \mathbf{P}(\,\mathbf{P}_{\mathcal{H}}(U_{p(k)})\,T\,) \\ &= \mathbf{P}(\mathbf{X}_k T) \\ &= \mathbf{P}(X_{[k]}\Lambda_{\mathbf{F}}^{-1}\mathbf{F}T) \\ &= X_{[k]}\,\Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}T) \\ &= X_{[k]}\mathbf{F}_0\,. \end{aligned}$$

If $\mathbf{F}_0$ is a strong basis representation, then $X'_{[k]}\mathbf{F}_0 = X_{[k]}\mathbf{F}_0$ implies $X'_{[k]} = X_{[k]}$. □

The above theorems, along with proposition 3.13, prove the converse of corollary 3.14:

THEOREM 3.28. (KRONECKER'S THM, II) *Let $T \in \mathcal{U}$ be a locally finite operator. If the range of the Hankel operator H_T has s-dim$\mathcal{H}_0 = d$, then there exist bounded minimal realizations $\{A, B, C, D\}$ for T with $\ell_A \le 1$ and $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$, where $\mathcal{B} = \mathbf{C}^d$.*

It is possible to choose the minimal realization to be either uniformly controllable or uniformly observable, but both can occur for the same realization if and only if the range of H_T is closed.

A generic way to choose the basis representations $\mathbf{Q}$ and $\mathbf{G}$ is obtained via a singular value decomposition of H_T. Let $T \in \mathcal{U}$ be locally finite. Then there exists $\mathbf{Q}$, $\mathbf{G}$, $\hat{\Sigma}$ such

that

$$\cdot H_T = \mathbf{P}_0(\cdot\,\mathbf{Q}^*)\hat{\Sigma}\mathbf{G} \qquad \text{with} \qquad \begin{array}{rcll} \mathcal{D}_2^{\mathcal{B}}\mathbf{Q} & = & \overline{\mathcal{H}}, & \Lambda_{\mathbf{Q}} = I \\ \mathcal{D}_2^{\mathcal{B}}\mathbf{G} & = & \overline{\mathcal{H}}_0, & \Lambda_{\mathbf{G}} = I \\ \hat{\Sigma} & \in & \mathcal{D}(\mathcal{B},\mathcal{B}), & \hat{\Sigma}^* = \hat{\Sigma}. \end{array} \tag{3.71}$$

This factorization of H_T is obtained by computing the singular value decomposition of its snapshots H_k (as in section 3.2), putting the singular vectors whose span is the range of H_k^* and H_k into $\mathbf{Q}_k$ and $\mathbf{G}_k$, and putting the non-zero singular values into $\hat{\Sigma}_k$. Then $\mathbf{Q}$, $\mathbf{G}$ are obtained by stacking the $\mathbf{Q}_i$ and $\mathbf{G}_i$ (like was done in equations (2.31) and (3.50)), and setting $\hat{\Sigma} = \mathrm{diag}[\hat{\Sigma}_k]_{-\infty}^{\infty}$. Since $\|H_k\| = \|\hat{\Sigma}_k\|$, also $\|H_T\| = \|\hat{\Sigma}\|$. Based on the above factorization of H_T, the factorizations corresponding to the canonical realizations as derived in this section are

$$\begin{array}{rcll} H_T & = & \left[\mathbf{P}_0(\cdot\,\mathbf{Q}^*)\right]\,[\hat{\Sigma}\mathbf{G}] = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\,\mathbf{F}_0, & (\mathbf{F} = \mathbf{Q},\ \mathbf{F}_0 = \hat{\Sigma}\mathbf{G}) \\ & = & \left[\mathbf{P}_0(\cdot\,\mathbf{Q}^*)\hat{\Sigma}\right]\,\mathbf{G} = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)\,\mathbf{F}_0', & (\mathbf{F}' = \hat{\Sigma}\mathbf{Q},\ \mathbf{F}_0' = \mathbf{G}). \end{array}$$

The factorization of H_T on the first line corresponds to a canonical controller realization on $\mathbf{Q}$ for which $\Lambda_{\mathbf{F}_0} = \hat{\Sigma}^2$, while the second factorization corresponds to a canonical observer realization based on $\mathbf{G}$ and has $\Lambda_{\mathbf{F}'} = \hat{\Sigma}^2$. The actual construction of the realization based on $\mathbf{G}$, according to theorem 3.23, can be done along the lines of algorithm 3.1 in section 3.2.

Anomalies

In the previous sections, some anomalies were noted:

1. The basis representations $\mathbf{Q}$, $\mathbf{G}$ of $\mathcal{H}$ and $\mathcal{H}_0$ can be unbounded, which occurs if and only if $\ell_A = 1$.
2. H_T, H_T^* can have ranges $\mathcal{H}_0$, $\mathcal{H}$ which are not closed.

We show that these phenomena are unconnected, and that the second item is governed by the singular values $\hat{\Sigma}$ of the Hankel operator.

PROPOSITION 3.29. *Let $T \in \mathcal{U}$ be a locally finite operator, and let its Hankel operator H_T have a singular value decomposition given by (3.71). $\mathcal{H}$ and $\mathcal{H}_0$ are closed subspaces if and only if $\hat{\Sigma}$ is boundedly invertible, and a realization of T which is uniformly controllable and uniformly observable exists if and only if this condition holds.*

PROOF Consider the SVD-based factorization of H_T in terms of (3.71). A realization based on $\mathbf{Q}$ is uniformly controllable, and because $\mathbf{F}_0 = \hat{\Sigma}\mathbf{G}$, the observability Gramian is $\Lambda_{\mathbf{F}_0} = \hat{\Sigma}^2$. Hence the realization is observable. It is uniformly observable, $\hat{\Sigma}^2 \gg 0$, if and only if $\hat{\Sigma}^{-1}$ is bounded. According to proposition 3.13, this occurs if and only if $\mathcal{H}$ and $\mathcal{H}_0$

are both closed subspaces. Proposition 3.13 already implied that any other realization can be both uniformly controllable and uniformly observable if and only if these subspaces are closed. □

If $\mathbf{Q}$ is bounded and $\hat{\Sigma}^{-1}$ is bounded, then $\mathbf{G}$ and $\mathbf{F}_0$ are bounded, because $H_T = \mathbf{P}_0(\cdot\,\mathbf{Q}^*)\hat{\Sigma}\mathbf{G}$ is bounded, and also $\mathbf{P}_0(\cdot\,\mathbf{Q}^*)$ is bounded. If $\mathbf{Q}$ is unbounded and $\hat{\Sigma}^{-1}$ is bounded, then $\mathbf{G}$ and $\mathbf{F}_0$ are unbounded.

It is however not true that $\mathbf{Q}$ and $\mathbf{G}$ bounded implies that $\hat{\Sigma}^{-1}$ is bounded. A counterexample is provided by taking

$$T = \begin{bmatrix} 0 & 1/2 & & \mathbf{0} & \\ & \boxed{0} & 1/4 & & \\ & & 0 & 1/8 & \\ \mathbf{0} & & & \ddots & \ddots \end{bmatrix}$$

$\mathbf{Q}$, $\mathbf{G}$ and $\hat{\Sigma}$ are given by

$$\mathbf{Q} = \begin{bmatrix} 1 & \boxed{0} & \mathbf{0} & \\ & 1 & 0 & \\ \mathbf{0} & & 1 & 0 \\ & & & \ddots & \ddots \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \boxed{1} & \mathbf{0} \\ & 1 \\ \mathbf{0} & & 1 \\ & & & \ddots \end{bmatrix}, \quad \hat{\Sigma} = \begin{bmatrix} \boxed{1/2} & \mathbf{0} & \\ & 1/4 & \\ \mathbf{0} & & 1/8 \\ & & & \ddots \end{bmatrix}$$

$\mathbf{Q}$ and $\mathbf{G}$ are bounded, but $\hat{\Sigma}^{-1}$ is unbounded. A realization based on $\mathbf{Q}$ yields $A_k = 0$, $B_k = 1$, $C_k = 2^{-k-1}$, $D_k = 0$ $(k \geq 0)$. Indeed, the realization is not uniformly observable.

It is also not true that $\hat{\Sigma}^{-1}$ bounded implies that $\mathbf{Q}$, $\mathbf{G}$ are bounded. A counterexample is obtained by considering inner operators (operators T which are both unitary and upper). As shown in chapter 4, such operators have Hankel matrices H_k that are isometries, so that $\hat{\Sigma} = I_{\mathcal{B}}$. We also show in that chapter that a unitary realization $\mathbf{T} = \{A, B, C, D\}$ realizes a unitary operator T. It is, however, possible to construct a sequence of unitary matrices $\mathbf{T}_k$ such that $\ell_A = 1$, a trivial example being

$$\mathbf{T}_k = \begin{bmatrix} c_k & s_k \\ -s_k^* & c_k^* \end{bmatrix}, \qquad c_k^* c_k + s_k^* s_k = 1,$$

where $c_k \to 1$ for $k \to \infty$. With $\ell_A = 1$, $\mathbf{Q}$ and $\mathbf{G}$ are unbounded.

It has thus been shown that there is no connection between the properties $\ell_A < 1$ ($\mathbf{Q}$ and $\mathbf{G}$ bounded) and the fact that $\mathcal{H}$ and $\mathcal{H}_0$ are closed subspaces ($\hat{\Sigma}$ boundedly invertible).

As a pathological example in which some of the above-mentioned aspects occur, consider the operator

$$T = \begin{bmatrix} \boxed{0} & 1 & 1/2 & 1/4 & 1/8 & 1/16 & \cdots \\ & 0 & 1/2 & 1/4 & 1/8 & 1/16 & \cdots \\ & & 0 & 1/4 & 1/8 & 1/16 & \cdots \\ & & & 0 & \vdots & & \ddots \end{bmatrix}$$

T is a bounded operator: it is equal to a diagonal scaling of the bounded LTI system $z(1 - 1/2\,z)^{-1}$. One possible (SVD-based) factorization of its Hankel operators H_k is (for $k > 0$)

$$H_k = \sigma_k \cdot \frac{1}{\sqrt{k}} \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \end{bmatrix} \Big\} k \cdot \frac{1}{p} \begin{bmatrix} 1 & 1/2 & 1/4 & \cdots \end{bmatrix}$$

where $\sigma_k = \frac{p\sqrt{k}}{2^{k-1}}$ and p is equal to the norm of the vector $[\,1 \quad 1/2 \quad 1/4 \quad \cdots\,]$. Each Hankel matrix H_k has only one singular value unequal to 0, and $\sigma_k \to 0$ if $k \to \infty$, hence $\hat{\Sigma}$ is not boundedly invertible. $\mathbf{Q}$ and $\mathbf{G}$ follow from the above decomposition as

$$\mathbf{Q} = \begin{bmatrix} \boxed{\cdot} & & & & \\ 1 & 0 & & & \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & & \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \\ \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & \frac{1}{\sqrt{4}} & 0 \\ & \vdots & & & \ddots \end{bmatrix} \qquad \mathbf{G} = \begin{bmatrix} \boxed{\cdot} & \frac{1}{p} & \frac{1}{2p} & \frac{1}{4p} & \frac{1}{8p} & \cdots \\ & & \frac{1}{p} & \frac{1}{2p} & \frac{1}{4p} & \cdots \\ & & & \frac{1}{p} & \frac{1}{2p} & \cdots \\ & & & & \ddots & \end{bmatrix}$$

$\mathbf{G}$ is bounded, but $\mathbf{Q}$ is unbounded, which can be seen, *e.g.*, from the fact that the norms of its columns are unbounded. A realization based on $\mathbf{G}$ has

$$A_k = \frac{1}{2}, \qquad C_k = \frac{1}{p}, \qquad (k > 0), \tag{3.72}$$

so that $\ell_A = 1/2$, but $B_k = \frac{p}{2^{k-1}} \to 0$ $(k \to \infty)$ and the realization is not uniformly controllable. A realization based on $\mathbf{Q}$ is

$$\begin{aligned} A_k &= \frac{1}{\sqrt{k}} \frac{k}{\sqrt{k+1}} = \frac{\sqrt{k}}{\sqrt{k+1}} \to 1 \\ B_k &= \frac{1}{\sqrt{k+1}} \to 0 \end{aligned} \tag{3.73}$$

and indeed $\ell_A = 1$, which was to be expected as $\mathbf{Q}$ is unbounded.

3.5 STATE-SPACE ISOMORPHISMS

We now go into some details of the concept of state similarity, and the equivalence of two state models that realize the same transfer operator. This subject has been studied for finite dimensional time-invariant systems in, among others, [6]. For infinite dimensional systems, the situation is more complex and a number of extra conditions have to be introduced concerning the boundedness and closed-rangeness of the controllability and observability operators. This theory was set up by Helton [20] and Fuhrmann [24, 19], and explained in detail in [25]. The results are readily applicable to our context, where instead of a time-invariant infinite number of states, we have an infinite sequence of a finite number of states. The relevant extra concepts with regard to controllability and observability have already been introduced in section 3.3. We follow the treatment in Fuhrmann [25].

Let $\mathbf{T} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$, $\mathbf{T}' = \begin{bmatrix} A' & C' \\ B' & D' \end{bmatrix}$, A map $R \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ is said to *intertwine* realizations $(\mathbf{T}, \mathbf{T}')$ if $D = D'$ and

$$\begin{cases} B &= B'R^{(-1)} \\ RA &= A'R^{(-1)} \\ RC &= C' \end{cases} \tag{3.74}$$

(Note that the property is not symmetrical.) $\mathbf{T}$ is a *transformation* of $\mathbf{T}'$ if there exists also an operator $R' \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ such that

$$\begin{cases} B' &= BR'^{(-1)} \\ R'A' &= AR'^{(-1)} \\ R'C' &= C \end{cases} \tag{3.75}$$

$\mathbf{T}$ and $\mathbf{T}'$ are *similar* if there exists a boundedly invertible R such that R satisfies (3.74) and $R' = R^{-1}$ satisfies (3.75), in which case $\mathbf{T}$ and $\mathbf{T}'$ are related by the state transformation formula (3.13). $\mathbf{T}$ and $\mathbf{T}'$ are *quasi-similar* if R is algebraically invertible but the inverse is not necessarily bounded, that is, if R is one-to-one with dense range.

Let $\mathbf{F}$ and $\mathbf{F}_0$ be the controllability and observability operators of $\mathbf{T}$, and let $\mathbf{F}'$ and $\mathbf{F}'_0$ be likewise for $\mathbf{T}'$. If R intertwines $(\mathbf{T}, \mathbf{T}')$, then $\mathbf{F} = R^*\mathbf{F}'$ and $R\mathbf{F}_0 = \mathbf{F}'_0$, and the corresponding Gramians satisfy

$$\begin{aligned} \Lambda_{\mathbf{F}} &= R^*\Lambda_{\mathbf{F}'}R \\ \Lambda_{\mathbf{F}'_0} &= R\Lambda_{\mathbf{F}_0}R^*. \end{aligned}$$

LEMMA 3.30. *If R intertwines $(\mathbf{T}, \mathbf{T}')$, then $\mathbf{T}$ and $\mathbf{T}'$ realize the same operator T.*

PROOF Immediate, in view of (3.12): for $k > 0$, the diagonals $T_{[k]}$ of T are

$$\begin{aligned} T_{[k]} &= B'^{(k)} A'^{\{k-1\}} C' \\ &= B'^{(k)} A'^{(k-1)} \cdots A'^{(1)} RC \\ &= B'^{(k)} R^{(k)} A^{(k-1)} \cdots A^{(1)} C \\ &= B^{(k)} A^{\{k-1\}} C. \end{aligned}$$

□

The converse of this lemma is not true, unless further assumptions are made on the realizations $\mathbf{T}$ and $\mathbf{T}'$, as done in theorem 3.34 below.

LEMMA 3.31. *If* $\mathbf{T}$ *and* $\mathbf{T}'$ *are both controllable, then if these realizations are quasi-similar, they are similar.*

PROOF If $\mathbf{T}$ and $\mathbf{T}'$ are quasi-similar, then there is an R such that (3.74) holds, and an R' such that (3.75) holds. The operator $R'R$ intertwines $\mathbf{T}$ with itself, that is

$$\begin{aligned} B &= BR'^{(-1)} R^{(-1)}, \\ R'RA &= AR'^{(-1)} R^{(-1)}, \\ R'RC &= C. \end{aligned}$$

so that $B^{(k)} A^{\{k-1\}} R'R = B^{(k)} A^{\{k-1\}}$. Consequently, if $\mathbf{F}$ defined in equation (3.31) is the basis of the input state space corresponding to the chosen realization, then $\mathbf{F}^* R'R = \mathbf{F}^*$, or $\mathbf{F}^*(I - R'R) = 0$, so that $(I - R'R)^* \mathbf{F} = 0$. But since the realization is controllable, $\ker \mathbf{F} = 0$, hence $R'R = I$. Similarly, follows that $\mathbf{T}'$ intertwines with itself, which yields $RR' = I$. Hence R is boundedly invertible. □

LEMMA 3.32. *Let be given realizations* $\mathbf{T}$ *and* $\mathbf{T}'$. *If* $\mathbf{T}'$ *is controllable, then if* R *intertwines* $(\mathbf{T}, \mathbf{T}')$, *it is unique. The same holds if* $\mathbf{T}$ *is observable.*

PROOF If R intertwines $(\mathbf{T}, \mathbf{T}')$, then $\mathbf{F} = R^* \mathbf{F}'$. If R_1 also intertwines, then $(R - R_1)^* \mathbf{F}' = 0$. Since $\mathbf{T}'$ is controllable, $\ker \mathbf{F}' = 0$, and $(R - R_1)^* \mathbf{F}' = 0$ implies $R = R_1$. □

THEOREM 3.33. *Let be given realizations* $\mathbf{T}$ *and* $\mathbf{T}'$, *and an operator* R *intertwining* $(\mathbf{T}, \mathbf{T}')$.

- *If* R *has dense range and* $\mathbf{T}'$ *is controllable, then* $\mathbf{T}$ *is controllable. Conversely, if* $\mathbf{T}$ *is controllable, then* R *has dense range.*
- *If* R *is surjective and* $\mathbf{T}'$ *is uniformly controllable, then* $\mathbf{T}$ *is uniformly controllable. Conversely, if* $\mathbf{T}$ *is uniformly controllable, then* R *is surjective.*

- *If R is one-to-one, then $\mathbf{T}$ observable implies that $\mathbf{T}'$ is observable. If $\mathbf{T}'$ is observable, then R is one-to-one.*
- *If R^* is surjective, then $\mathbf{T}$ uniformly controllable implies that $\mathbf{T}'$ is uniformly controllable. If $\mathbf{T}'$ is uniformly observable, then R^* is surjective.*

PROOF If R intertwines $(\mathbf{T}, \mathbf{T}')$, then $\mathbf{F} = R^*\mathbf{F}'$. R has dense range if and only if $\ker R^* = 0$. $\mathbf{T}$ is controllable if and only if $\ker \mathbf{F} = 0$. If $\ker \mathbf{F}' = 0$ and $\ker R^* = 0$ then $\ker \mathbf{F} = 0$ so that $\mathbf{T}$ is controllable. Conversely, $\mathbf{F} = R^*\mathbf{F}' \Rightarrow \ker \mathbf{F} \supset \ker R^*$, so that if $\mathbf{T}$ is controllable, then $\ker R^* = 0$. This proves the first item.

If R is surjective, then $\operatorname{ran} R = \mathcal{D}_2^{\mathcal{B}}$. $\mathbf{T}$ is uniformly controllable if and only if $\operatorname{ran}\left[\mathbf{P}_0(\cdot\,\mathbf{F}^*)\right] = \mathcal{D}_2^{\mathcal{B}}$. Since $\mathbf{F}^* = \mathbf{F}'^*R$, this shows that if $\mathbf{T}'$ is uniformly controllable and R is surjective, then $\mathbf{T}$ is also uniformly controllable. It also proves the converse statement.

The remaining two items are dual to the above. □

THEOREM 3.34. *Let $\mathbf{T}$ and $\mathbf{T}'$ realize the same transfer operator. (1) If $\mathbf{T}$ is controllable and observable, and $\mathbf{T}'$ is observable and uniformly controllable, then there exists an $R \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ which is one-to-one and has dense range, and intertwines $(\mathbf{T}, \mathbf{T}')$. (2) If in addition $\mathbf{T}$ is uniformly controllable, then R is boundedly invertible and the two realizations are similar.*

PROOF If $\mathbf{T}$ and $\mathbf{T}'$ realize the same transfer operator T, then $H_T = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0 = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)\mathbf{F}_0'$. We first show that there exists an R which is one-to-one and has dense range (hence is algebraically invertible) such that $\mathbf{P}_0(\cdot\,\mathbf{F}^*) = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)R$ and $R\mathbf{F}_0 = \mathbf{F}_0'$.

Indeed,

$$\begin{aligned}\mathbf{F}'H_T &= \mathbf{P}_0(\mathbf{F}'\mathbf{F}'^*)\mathbf{F}_0' = \Lambda_{\mathbf{F}'}\mathbf{F}_0' \\ &= \mathbf{P}_0(\mathbf{F}'\mathbf{F}^*)\mathbf{F}_0\,.\end{aligned}$$

Hence $\mathbf{F}_0' = R\mathbf{F}_0$ with $R = \Lambda_{\mathbf{F}'}^{-1}\mathbf{P}_0(\mathbf{F}'\mathbf{F}^*)$. Because $\mathbf{T}'$ is observable, $\ker R \subset \ker \mathbf{F}_0' = 0$ and R is one-to-one. Inserting the expression for $\mathbf{F}_0'$ into that of H_T yields $H_T = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0 = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)R\mathbf{F}_0$. Because $\mathbf{T}$ is observable, $\mathbf{P}_0(\cdot\,\mathbf{F}^*) = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)R$ and $\mathbf{F} = R^*\mathbf{F}'$. Because $\mathbf{F}$ is controllable, $\ker \mathbf{F} = 0$, and hence $\ker R^* = 0$ so that R has dense range.

It remains to show that R intertwines $(\mathbf{T}, \mathbf{T}')$. Firstly, $R\mathbf{F}_0 = \mathbf{F}_0'$ implies $RC = C'$. Likewise, $\mathbf{P}_0(\cdot\,\mathbf{F}^*) = \mathbf{P}_0(\cdot\,\mathbf{F}'^*)R$ implies $B = B'R^{(-1)}$. Finally, again using this relation and $\mathbf{F}^* = BZ + \mathbf{F}^*AZ$ (*cf.* equation (3.39)) gives

$$\begin{aligned}A'R^{(-1)} &= \Lambda_{\mathbf{F}'}^{-1}\mathbf{P}_0(Z^{-1}\mathbf{F}'\mathbf{F}'^*)^{(-1)}R^{(-1)} \\ &= \Lambda_{\mathbf{F}'}^{-1}\mathbf{P}_0(Z^{-1}\mathbf{F}'\mathbf{F}^*)^{(-1)} \\ &= \Lambda_{\mathbf{F}'}^{-1}\mathbf{P}_0(Z^{-1}\mathbf{F}'[BZ + \mathbf{F}^*AZ])^{(-1)} \\ &= \Lambda_{\mathbf{F}'}^{-1}\mathbf{P}_0(Z^{-1}\mathbf{F}'\mathbf{F}^*AZ)^{(-1)}\end{aligned}$$

$$
\begin{aligned}
&= \Lambda_{\mathbf{F}'}^{-1} \mathbf{P}_0(\mathbf{F}'\mathbf{F}^*)A \\
&= RA .
\end{aligned}
$$

If $\mathbf{T}$ is also uniformly controllable, then $\mathcal{D}_2^{\mathcal{B}} = \operatorname{ran} \mathbf{P}_0(\cdot\, \mathbf{F}^*) = \operatorname{ran} \mathbf{P}_0(\cdot\, \mathbf{F}'^*)R$. Because $\operatorname{ran} \mathbf{P}_0(\cdot\, \mathbf{F}'^*) = \mathcal{D}_2^{\mathcal{B}}$ also, it follows that R is of full range. Because R is also one-to-one, it is boundedly invertible and the two realizations are similar. □

A dual result holds if $\mathbf{T}$ is uniformly observable. A consequence of this theorem is that all uniformly controllable minimal realizations are similar, and all uniformly observable minimal realizations are similar. Theorem 3.33, in addition, yields that a uniformly controllable realization is similar to a uniformly observable realization only if H_T has closed range, so that the realizations are both uniformly controllable and uniformly observable.

An example of the latter situation is obtained by taking canonical SVD-based realizations, as before:

$$
\begin{aligned}
H_T &= \left[\mathbf{P}_0(\cdot\, \mathbf{Q}^*)\hat{\Sigma}\right] \mathbf{G} &&= \mathbf{P}_0(\cdot\, \mathbf{F}'^*)\, \mathbf{F}_0', && (\mathbf{F}' = \hat{\Sigma}\mathbf{Q},\ \mathbf{F}_0' = \mathbf{G}) \\
&= \left[\mathbf{P}_0(\cdot\, \mathbf{Q}^*)\right] \left[\hat{\Sigma}\mathbf{G}\right] &&= \mathbf{P}_0(\cdot\, \mathbf{F}^*)\, \mathbf{F}_0, && (\mathbf{F} = \mathbf{Q},\ \mathbf{F}_0 = \hat{\Sigma}\mathbf{G}).
\end{aligned}
$$

The intertwining operator between the realizations is given by $R = \hat{\Sigma}$. Hence R is a state transformation only if $\hat{\Sigma}$ is boundedly invertible, that is, if H_T has closed range. In this case, both realizations are uniformly controllable and uniformly observable: the first realization has $\Lambda_{\mathbf{F}}' = \hat{\Sigma}^2$, $\Lambda_{\mathbf{F}_0'} = I$, while the second realization has $\Lambda_{\mathbf{F}} = I$, $\Lambda_{\mathbf{F}_0} = \hat{\Sigma}^2$.

Finally, we remark that if two realizations $\mathbf{T}$ and $\mathbf{T}'$ are similar with similarity transformation R, then $A' = RAR^{(-1)}$. As noted in equation (3.14), $\ell_A = \ell_{A'}$, so that the two realizations have the same stability properties. However, if the two realizations are not similar (the intertwining map R is not boundedly invertible), then it is not necessarily true that $\ell_A = \ell_{A'}$. An example of this was given in the previous section, equations (3.72) and (3.73). This effect is comparable to what occurs in infinite-dimensional LTI system theory, where it was noted that for infinite-dimensional systems, minimal realizations of the same system do not necessarily have A-operators with the same spectrum [25].

Finally, we remark that, even in cases where R is not boundedly invertible, it is still possible that $RAR^{(-1)}$ and $BR^{(-1)}$ are bounded. For example, the canonical controller and observer realizations are connected to each other via a state transformation $R = \hat{\Sigma}$, and both are bounded realizations even if $\hat{\Sigma}$ is not boundedly invertible. For time-invariant infinite-dimensional discrete-time systems, it was proven by Young [22] that also a state transformation by $\hat{\Sigma}^{1/2}$ (which will lead to *balanced realizations*: realizations with equal and diagonal controllability and observability Gramians) give bounded realizations. We expect the same to be true in the time-varying context.

3.6 DISCUSSION

Historical notes

The concept of state originated as an abstraction of computer memory in automaton theory [26]. It entered system theory in the late 1950s when the connection with first-order differential equations became clear. During the 1960s, much effort was put into the construction of state models for continuous-time LTI and LTV systems specified by their impulse response ('weighting pattern') $H(T, \tau)$, such that $y(t) = \int H(t, \tau)u(\tau)d\tau$. Among the initial results was the proof that realizability is equivalent to the separability of the impulse response matrix into $H(t, \tau) = \Psi(t)\Theta(\tau)$. However, the effective construction of this factorization was difficult, and even not always possible, and the direct realizations that were produced were not always asymptotically stable [27]. For LTI systems, state-space realization synthesis began with the work of Kalman and his co-workers [28, 5, 6], Gilbert [29] and Youla [30]. The use of the Hankel matrix, which does not require a separable form of the impulse response matrix, resulted in the Ho-Kalman algorithm [5], which was independently obtained by Youla and Tissi [31]. For continuous-time time-varying systems with a constant system order, a realization theory was developed by Silverman and Meadows [32, 33, 34]. Controllability and stability issues were treated also in [35]. Kamen extended Kalman's algebraic module theory to incorporate a continuous-time pure delay operator [36, 37], and considered the realization by state-space models of systems $Ay(t) = Bu(t)$, where A and B are matrix polynomials in the differential operator p and unit delay operator d. For time-varying systems, these results could be extended by using a non-commutative ring of polynomials [38].

Discrete-time realization theory for LTV systems started its development in the 1970s with the work of Weiss [39] and Evans [40]. The concepts of controllability, observability and minimality were defined (see also [41]), but the realization theory was limited to state dimensions of constant rank. An algebraic approach was followed by Kamen, Khargonekar, and Poolla [42, 43, 44], who defined time-varying systems via modules of non-commutative rings of polynomials acting on signals in $\ell_\infty(\mathbb{Z})$. Many definitions and results in [43] can be translated directly into the diagonal algebra considered in this thesis: instead of Z, two operators z and σ are used, where σ is a time-shift operator on sequences, and z is an algebraic symbol. The description of objects using z and σ is equivalent to our description of diagonals and polynomials (in Z) of diagonals. The aspect of varying state dimensions was generalized only more recently in [1], and in parallel in a realization theory for lower triangular block matrices by Gohberg, Kaashoek and Lerer [3], in which again operators on $\ell_\infty(\mathbb{Z})$ were considered. As discussed by Murray in [2], operators on L_∞ in their generality do not have an adjoint operator. As a remedy, he defined a 'crossed product' algebra as a subalgebra of operators on L_∞ such that an involution (*) can be defined, and proved some aspects of state realizations in that algebra.

In a parallel evolution, mathematicians and 'fundamental' engineers considered state-

space theory for operators on a Hilbert space. Besides the mathematical elegance, Hilbert space theory seemed necessary to incorporate infinite dimensional systems in a state space theory. Such systems arise in a natural way in the time-continuous context of systems which contain 'pure delays', *e.g.*, networks with lossless transmission lines. Scattering theory for such networks was developed by Phillips and Lax [45], but without using state-space theory. Connections between the fields of Hilbert space operator theory (in particular the work of Sz.-Nagy and Foias [46] and network synthesis were made by Livsic in 1965 in Russia and with other viewpoints by Dewilde [47], Helton [48, 20, 49] and Fuhrmann [50, 51, 52, 25] in the U.S.A. These efforts put the algebraic realization theory of Kalman into the Hardy space context of shift-invariant subspaces à la Helson [53], Beurling-Lax representations of such subspaces by inner functions [54, 55], and coprime factorizations. More recently, additional results on this type of realization theory (the existence of balanced realizations for infinite-dimensional discrete-time systems) have been obtained by Young [22]. These ideas and results on infinite-dimensional realization theory of operators in Hilbert space are fundamental to the time-varying realization theory as treated in this chapter, and to a number of results in the chapters to come.

Finally, one different but related approach to the time-varying realizations of operators in Hilbert space is the work of Feintuch and Saeks [21]. Their theory is based on a Hilbert space resolution of the identity in terms of a nested series of projectors that endow the abstract Hilbert space with a time structure. The projectors are projectors of sequences onto the past, with respect to each point k in time. With the projectors, one can define various types of causality, and the theory provides operators with a state structure via a factorization of the Hankel operator, which is also defined in terms of the projections. Many of the issues mentioned in the present chapter are also discussed in the book [21], but in a different language.

Computational issues

We mention some issues related to theorem 3.1 and the corresponding realization algorithm, which are of some importance for a practical implementation of the algorithm.

Let T be a given upper triangular matrix, and consider its sequence of Hankel matrices $\{H_k\}$, where H_k has rank d_k. If for each H_k a submatrix $\hat{H}_k$ is known such that $\text{rank}(\hat{H}_k) = d_k$ also, then it is possible to determine a realization of T based on factorizations of the $\hat{H}_k$ rather than factorizations of H_k. This generalization of the time-invariant analog [56] is useful since it can yield considerable computational savings if the $\hat{H}_k$ are small in comparison with H_k. A remaining practical problem is how to obtain the $\hat{H}_k$ in an efficient way, because, unlike the time-invariant case, T need not be diagonally dominant even if its Hankel matrices have low rank, so that the $\hat{H}_k$ can still be matrices of large size. A trivial example of the latter is provided by taking T to be an $n \times n$ matrix consisting of zeros, except for the $(1, n)$-entry.

In this section, we use the matrix $\pi_r := [I_r \ 0 \ 0 \ \cdots]$ to select the first r rows of a matrix

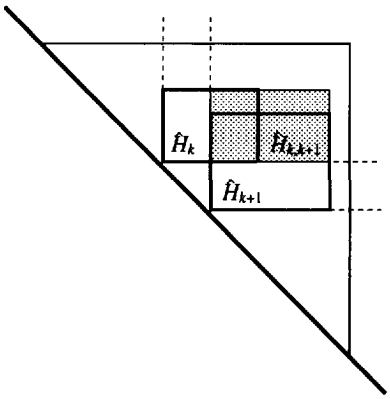


Figure 3.4. Relation between $\hat{H}_k$ and $\hat{H}_{k+1}$.

at its right. We use, as before, the notation $H_k^{\leftarrow}$ to denote H_k with its first column deleted, and let $\dagger$ denote the generalized (left or right) inverse of a matrix. The following result (and proof) can be found in [3].

THEOREM 3.35. *Let T be an upper triangular matrix with Hankel matrices H_k having rank d_k. For each k, suppose that the numbers $r(k)$ and $c(k)$ are such that the submatrices $\hat{H}_k = \pi_{r(k)} H_k \pi_{c(k)}^*$ has rank d_k. Let $\hat{H}_k = \hat{\mathcal{C}}_k \hat{\mathcal{O}}_k$ be a factorization of $\hat{H}_k$ into minimal rank factors. Then a realization of T is given by*

$$\begin{aligned} \hat{A}_k &= \hat{\mathcal{C}}_k^{\dagger} \hat{H}_{k,k+1} \hat{\mathcal{O}}_k^{\dagger}, & \hat{C}_k &= \hat{\mathcal{O}}_k \pi_1^*, \\ \hat{B}_k &= \pi_1 \hat{\mathcal{C}}_k, & \hat{D}_k &= T_{k,k}, \end{aligned}$$

where $\hat{H}_{k,k+1} = \pi_{r(k)} H_k^{\leftarrow} \pi_{c(k+1)}$.

PROOF A diagram of the relations between $\hat{H}_k$, $\hat{H}_{k+1}$ and $\hat{H}_{k,k+1}$ is provided in figure 3.4. The proof consists of two parts. We first verify that H_k has a minimal factorization into rank d_k factors $\mathcal{C}_k$ and $\mathcal{O}_k$ such that

$$\hat{\mathcal{C}}_k = \pi_{r(k)} \mathcal{C}_k, \qquad \hat{\mathcal{O}}_k = \mathcal{O}_k \pi_{c(k)}^*. \tag{3.76}$$

Indeed, let $H_k = \tilde{\mathcal{C}}_k \tilde{\mathcal{O}}_k$ be a minimal factorization, then $\hat{H}_k = \pi_{r(k)} H_k \pi_{c(k)}^* = (\pi_{r(k)} \tilde{\mathcal{C}}_k)(\tilde{\mathcal{O}}_k \pi_{c(k)}^*)$. Because $\operatorname{rank}(\hat{H}_k) = d_k$ also, it follows that $\pi_{r(k)} \tilde{\mathcal{C}}_k$ and $\tilde{\mathcal{O}}_k \pi_{c(k)}^*$ are full rank factors of $\hat{H}_k$, so that these are related to the given factorization $\hat{H}_k = \hat{\mathcal{C}}_k \hat{\mathcal{O}}_k$ as $\hat{\mathcal{C}}_k = (\pi_{r(k)} \tilde{\mathcal{C}}_k) R_k$ and $\hat{\mathcal{O}}_k = R_k^{-1} (\tilde{\mathcal{O}}_k \pi_{c(k)}^*)$, where R_k is an invertible state transformation. Putting $\mathcal{C}_k = \tilde{\mathcal{C}}_k R_k$ and $\mathcal{O}_k = R_k^{-1} \tilde{\mathcal{O}}_k$ gives (3.76).

The second step is to verify that $\{\hat{A}_k, \hat{B}_k, \hat{C}_k, \hat{D}_k\}$ is a realization of T. This is done by proving that this realization is precisely equal to the realization based on the full-size factors $\mathcal{C}_k$ and $\mathcal{O}_k$. The main issue is in proving that $A_k = \mathcal{C}_k^{\dagger} H_k^{\leftarrow} \mathcal{O}_k^{\dagger}$ is equal to $\hat{A}_k$. Expressions for these generalized inverses are

$$\begin{aligned} \mathcal{C}_k^{\dagger} &= \hat{\mathcal{C}}_k^{\dagger} \pi_{r(k)} \\ \mathcal{O}_k^{\dagger} &= \pi_{c(k)}^{*} \hat{\mathcal{O}}_k^{\dagger} \end{aligned}$$

because $\mathcal{C}_k^{\dagger}\mathcal{C}_k = \hat{\mathcal{C}}_k^{\dagger}\pi_{r(k)}\mathcal{C}_k = \hat{\mathcal{C}}_k^{\dagger}\hat{\mathcal{C}}_k = I_{d_k}$, and likewise for $\mathcal{O}_k^{\dagger}$. Hence $A_k = \mathcal{C}_k^{\dagger} H_k^{\leftarrow} \mathcal{O}_k^{\dagger} = \hat{\mathcal{C}}_k^{\dagger}\pi_{r(k)} H_k^{\leftarrow} \pi_{c(k)}^{*}\hat{\mathcal{O}}_k^{\dagger} = \hat{\mathcal{C}}_k^{\dagger}\hat{H}_{k,k+1}\hat{\mathcal{O}}_k^{\dagger} = \hat{A}_k$. With less effort, it follows that $B_k = \pi_1 \mathcal{C}_k = \pi_1 \pi_{r(k)}\mathcal{C}_k = \pi_1 \hat{\mathcal{C}}_k = \hat{B}_k$, and likewise $C_k = \hat{C}_k$. □

The content of the theorem is such that we can work with finite matrices without loss of accuracy, as long as we are sure that the finite Hankel matrix has a rank equal to the actual system order at that point. Hence, a remaining problem is to obtain the submatrices $\{\hat{H}_k\}$ as required in the theorem. If, for some k, the submatrix $\hat{H}_k$ of rank d_k is known, then it is possible to determine the next submatrix $\hat{H}_{k+1}$ iteratively in a number of ways. One solution is as follows. For ease of description, the single-input single-output case is considered, so that the rank of H_{k+1} differs at most 1 from the rank of H_k.

1. Remove the first column of $\hat{H}_k$, and add a new row on top. This is a skeleton for $\hat{H}_{k+1}$, but its rank might be too low (by 1 in the scalar case).
2. Add columns until the rank of the matrix increases by 1. This might not occur, in which case we try to remove as many columns from the right side without decreasing the rank.
3. Finally, try to remove as many of the bottom rows without decreasing the rank of the matrix.

An important step in such an updating scheme is the determination of the rank of a matrix that differs from the previous one by a column or a row. Since, in the end, we also need a factorization of the matrix, it makes sense to combine the determination of the rank and the factorization of the new matrix. The combination is provided, for example, by an SVD updating scheme for computing the SVD of $A = [A_1 \;\; b]$ from an SVD of A_1 (updating), or for computing the SVD of A_1 if the SVD of A is known (deflation). Algorithms to do this are known (see *e.g.*, [57]); they require $\mathcal{O}(n^3)$ operations for matrices of size $n \times p$. Since the SVD is rather expensive to update, it might be useful to consider other types of decompositions. In [18], Stewart introduces the *URV* decomposition of a matrix, along with updating schemes. The *URV* decomposition of a matrix A is not unique; it is a decomposition of A into

$$A = U \begin{bmatrix} R & F \\ 0 & G \end{bmatrix} V^{*}$$

where R and G are upper triangular, R is invertible, and F and G are small in norm. In principle, F and G are to be equal to 0, but the *URV* decomposition is designed to be 'rank revealing', which makes it robust in the presence of numerical errors. The notions of 'small' and 'invertible' must be made more precise. With $\sigma_1 \geq \cdots \geq \sigma_k > \sigma_{k+1} \geq \cdots \geq \sigma_p$, it is required that σ_k is large compared to σ_{k+1}, the smallest singular value of R is approximately equal to σ_k, and $\| F \|_F^2 + \| G \|_F^2 \simeq \sigma_{k+1}^2 + \cdots + \sigma_p^2$. The user must furnish a tolerance to distinguish 'small' from 'large' singular values, *i.e.*, specify the maximal condition number of R. The *URV* decomposition is cheaper to update than the SVD, since R need not be made upper triangular. The algorithm described in [18] is of order $\mathcal{O}(n^2)$. A crucial step is the estimation of the condition number of an augmented matrix. An overview of numerical algorithms to carry out this estimation is described in [58]. Such algorithms also typically require $\mathcal{O}(n^2)$ flops.

Bibliography

[1] A.J. van der Veen and P.M. Dewilde, "Time-varying system theory for computational networks," in *Algorithms and Parallel VLSI Architectures, II* (P. Quinton and Y. Robert, eds.), pp. 103–127, Elsevier, 1991.

[2] J. Murray, "Time-varying systems and crossed products," *Math. Systems Theory*, vol. 17, pp. 217–241, 1984.

[3] I. Gohberg, M.A. Kaashoek, and L. Lerer, "Minimality and realization of discrete time-varying systems," in *Time Variant Systems and Interpolation* (I. Gohberg, ed.), vol. OT 56, pp. 261–296, Birkhäuser Verlag, 1992.

[4] D. Alpay, P. Dewilde, and H. Dym, "Lossless Inverse Scattering and reproducing kernels for upper triangular operators," in *Extension and Interpolation of Linear Operators and Matrix Functions* (I. Gohberg, ed.), vol. 47 of *Operator Theory, Advances and Applications*, pp. 61–135, Birkhäuser Verlag, 1990.

[5] B.L. Ho and R.E. Kalman, "Effective construction of linear, state-variable models from input/output functions," *Regelungstechnik*, vol. 14, pp. 545–548, 1966.

[6] R.E. Kalman, P.L. Falb, and M.A. Arbib, *Topics in Mathematical System Theory*. Int. Series in Pure and Applied Math., McGraw-Hill, 1970.

[7] B.C. Moore, "Singular value analysis of linear systems," in *Proc. IEEE Conf. Dec. Control*, pp. 66–73, 1979.

[8] B.C. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," *IEEE Trans. Automat. Control*, vol. 26, pp. 17–32, Feb. 1981.

[9] H.P. Zeiger and A.J. McEwen, "Approximate linear realizations of given dimension via Ho's algorithm," *IEEE Trans. Automat. Control*, vol. 19, p. 153, Apr. 1974.

[10] L. Pernebo and L.M. Silverman, "Balanced systems and model reduction," in *Proc. IEEE Conf. Dec. Control*, pp. 865–867, 1979.

[11] S.Y. Kung, "A new identification and model reduction algorithm via singular value decomposition," in *Twelfth Asilomar Conf. on Circuits, Systems and Comp.*, (Asilomar, CA.), pp. 705–714, Nov. 1978.

[12] A.J. van der Veen, E.F. Deprettere, and A.L. Swindlehurst, "SVD-based estimation of low rank system parameters," in *Algorithms and Parallel VLSI Architectures* (E.F. Deprettere and A.J. van der Veen, eds.), vol. A, pp. 203–228, Elsevier, 1991.

[13] A.J. van der Veen, E.F. Deprettere, and A.L. Swindlehurst, "Subspace based signal analysis using singular value decomposition," *acc. for Proc. IEEE*, Feb. 1993.

[14] G. Golub and C.F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1984.

[15] L.V. Foster, "Rank and null space calculations using matrix decomposition without column interchanges," *Lin. Alg. Appl.*, vol. 74, pp. 47–71, 1986.

[16] T.F. Chan, "Rank-revealing *QR*-factorizations," *Lin. Alg. Appl.*, vol. 88/89, pp. 67–82, 1987.

[17] C.H. Bischof and G.M. Schroff, "On updating signal subspaces," *IEEE Trans. Signal Processing*, vol. 40, pp. 96–105, Jan. 1992.

[18] G.W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, pp. 1535–1541, June 1992.

[19] P.A. Fuhrmann, "Exact controllability and observability and realization theory in Hilbert space," *J. Math. Anal. Appl.*, vol. 53, pp. 377–392, Feb. 1976.

[20] J.W. Helton, "Discrete time systems, operator models, and scattering theory," *J. Functional Anal.*, vol. 16, pp. 15–38, May 1974.

[21] A. Feintuch and R. Saeks, *System Theory: A Hilbert Space Approach*. Academic Press, 1982.

[22] N.J. Young, "Balanced realizations in infinite dimensions," in *Operator Theory and Systems* (H. Bart, I. Gohberg, and M.A. Kaashoek, eds.), vol. OT-19, pp. 449–471, Birkhäuser Verlag, 1986.

[23] T. Kailath, *Linear Systems*. Prentice Hall, Englewood Cliffs, NJ, 1980.

[24] P.A. Fuhrmann, "Realization theory in Hilbert space for a class of transfer functions," *J. Functional Anal.*, vol. 18, pp. 338–349, Apr. 1975.

[25] P.A. Fuhrmann, *Linear Systems and Operators in Hilbert Space*. McGraw-Hill, 1981.

[26] A. Nerode, "Linear automaton transformations," *Proc. American Mathematical Society*, vol. 9, pp. 541–544, 1958.

[27] E.W. Kamen, "New results in realization theory for linear time-varying analytic systems," *IEEE Trans. Automat. Control*, vol. 24, pp. 866–878, Dec. 1979.

[28] R.E. Kalman, "Mathematical description of linear dynamical systems," *SIAM J. Control*, vol. 1, pp. 152–192, 1963.

[29] E.G. Gilbert, "Controllability and observability in multivariable control systems," *SIAM J. Control*, vol. 1, pp. 128–151, 1963.

[30] D.C. Youla, "The synthesis of linear dynamical systems from prescribed weighting patterns," *SIAM J. Applied Math.*, vol. 14, pp. 527–549, May 1966.

[31] D.C. Youla and P. Tissi, "*n*-Port synthesis via reactance extraction–part I," *IEEE Int. Conf. Rec.*, vol. 14, no. 7, pp. 183–205, 1966.

[32] L.M. Silverman and H.E. Meadows, "Equivalence and synthesis of time-variable linear systems," in *Proc. 4-th Allerton Conf. Circuit and Systems Theory*, pp. 776–784, 1966.

[33] L.M. Silverman and B.D.O. Anderson, "Controllability, observability and stability of linear systems," *SIAM J. Control*, vol. 6, no. 1, pp. 121–130, 1968.

[34] L.M. Silverman and H.E Meadows, "Equivalent realizations of linear systems," *SIAM J. Applied Math.*, vol. 17, pp. 393–408, 1969.

[35] B.D.O. Anderson and J.B. Moore, "New results in linear system stability," *SIAM J. Control*, vol. 7, pp. 398–414, Aug. 1969.

[36] E.W. Kamen, "On an algebraic theory of systems defined by convolution operators," *Math. Systems Theory*, vol. 9, no. 1, pp. 57–74, 1975.

[37] E.W. Kamen, "Module structure of infinite-dimensional systems with applications to controllability," *SIAM J. Control and Optimization*, vol. 14, pp. 389–408, May 1976.

[38] E.W. Kamen, "Representation and realization of operational differential equations with time-varying coefficients," *Journal of the Franklin Inst.*, vol. 301, pp. 559–571, June 1976.

[39] L. Weiss, "Controllability, realization, and stability of discrete-time systems," *SIAM J. Control and Optimization*, vol. 10, pp. 230–251, 1972.

[40] D.S. Evans, "Finite-dimensional realization of discrete-time weighting patterns," *SIAM J. Applied Math.*, vol. 22, pp. 45–67, 1972.

[41] B.D.O. Anderson and J.B. Moore, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM J. Control and Optimization*, vol. 19, pp. 20–32, Jan. 1981. Comments in *IEEE Trans. Automat. Control*, vol. 37, no. 3, 1992, pp. 409-410.

[42] E.W. Kamen and K.M. Hafez, "Algebraic theory of linear time-varying systems," *SIAM J. Control and Optimization*, vol. 17, pp. 500–510, July 1979.

[43] E.W. Kamen, P.P. Khargonekar, and K.R. Poolla, "A transfer-function approach to linear time-varying discrete-time systems," *SIAM J. Control and Optimization*, vol. 23, pp. 550–565, July 1985.

[44] K. Poolla and P. Khargonekar, "Stabilizability and stable-proper factorizations for linear time-varying systems," *SIAM J. Control and Optimization*, vol. 25, pp. 723–736, May 1987.

[45] P.D. Lax and R.S. Phillips, *Scattering Theory*. New York: Academic Press, 1967.

[46] B. Sz.-Nagy and C. Foias, *Harmonic Analysis of Operators on Hilbert Space*. Amsterdam: North-Holland, 1970.

[47] P. Dewilde, "Input-output description of roomy systems," *SIAM J. Control and Optimization*, vol. 14, pp. 712–736, July 1976.

[48] J.W. Helton, "The characteristic functions of operator theory and electrical network realization," *Indiana Univ. Math. J.*, vol. 22, no. 5, pp. 403–414, 1972.

[49] J.W. Helton, "Systems with infinite-dimensional state space: The Hilbert space approach," *Proceedings of the IEEE*, vol. 64, pp. 145–160, Jan. 1976.

[50] P.A. Fuhrmann, "On realizations of linear systems and applications to some questions of stability," *Math. Systems Theory*, vol. 8, pp. 132–141, 1974.

[51] P.A. Fuhrmann, "Realization theory in Hilbert space for a class of transfer functions," *J. Functional Anal.*, vol. 18, pp. 338–349, Apr. 1975.

[52] P.A. Fuhrmann, "Exact controllability and observability and realization theory in Hilbert space," *J. Math. Anal. Appl.*, vol. 53, no. 2, pp. 377–392, 1976.

[53] H. Helson, *Lectures on Invariant Subspaces*. New York: Academic Press, 1964.

[54] A. Beurling, "On two problems concerning linear transformations in Hilbert space," *Acta Math.*, vol. 81, pp. 239–255, 1949.

[55] P.D. Lax, "Translation invariant subspaces," *Acta Math.*, vol. 101, pp. 163–178, 1959.

[56] R.E. Kalman, "Irreducible realizations and the degree of a rational matrix," *SIAM J. Applied Math.*, vol. 13, pp. 520–545, 1965.

[57] J.R. Bunch and C.P. Nielsen, "Updating the singular value decomposition," *Numerische Mathematik*, vol. 31, pp. 111–129, 1978.

[58] N.J. Higham, "A survey of condition number estimation for triangular matrices," *SIAM Review*, vol. 29, pp. 575–596, Dec. 1987.

Chapter 4

INNER OPERATORS

4.1 REALIZATION OF INNER OPERATORS

An operator $V \in \mathcal{X}$ is an isometry if $VV^* = I$, a co-isometry if $V^*V = I$, and unitary if both $VV^* = I$ and $V^*V = I$, or $V^{-1} = V^*$. Equivalently, an operator is an isometry if its domain and range are closed subspaces in $\mathcal{X}_2$ and if inner products are conserved: for $F, G \in \mathcal{X}_2$, $\langle FV, GV\rangle_{HS} = \langle F, G\rangle_{HS}$, or $\{FV, GV\} = \{F, G\}$. An operator is inner if it is both unitary and upper. Systems described by isometric or inner operators are interesting for a number of reasons. For example, isometric systems satisfy an energy conservation property: let $U, Y \in \mathcal{X}_2$,

$$\begin{array}{llllll} \text{if} & VV^* = I & \text{then} & Y = UV & \Rightarrow & \|Y\|_{HS} = \|U\|_{HS} \\ \text{if} & V^*V = I & \text{then} & Y = UV^* & \Rightarrow & \|Y\|_{HS} = \|U\|_{HS}. \end{array}$$

Another elementary property is that they leave orthogonality intact:

$$\begin{array}{llll} \text{if} & VV^* = I & \text{then} & X \perp Y \Leftrightarrow XV \perp YV \\ \text{if} & V^*V = I & \text{then} & X \perp Y \Leftrightarrow XV^* \perp YV^*. \end{array}$$

If V is an isometry, then it maps closed sets into closed sets, since distances between elements of the set are preserved.

For finite matrices (operators in $\mathcal{U}(\mathcal{M}, \mathcal{N})$ with index sequences that vanish outside a finite interval), a non-trivial inner matrix is possible only if the dimensions of $\mathcal{M}$ and $\mathcal{N}$ are varying. This is because a scalar upper triangular and unitary matrix is diagonal.

Let $\mathbf{V} = \left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right]$ be a realization operator. The realization is called unitary if $\mathbf{V}\mathbf{V}^* = I$ and $\mathbf{V}^*\mathbf{V} = I$.

We first show that if an operator is inner and locally finite, then it has a realization that is unitary. Conversely, if a realization is unitary and has $\ell_A < 1$, then the corresponding transfer operator is inner. With these results, we look at certain standard factorizations of

transfer operators T. The first factorization that we consider is what we call the *external* (or *inner-coprime*) factorization: factorizations of the type

$$T = \Delta^* V$$

where V is inner and $\Delta \in \mathcal{U}$. In this factorization, V captures the output state space of T: $\mathcal{H}_0(V) = \overline{\mathcal{H}_0(T)}$. Because of this property, inner operators play an important role in the derivation of reduced-order models discussed in chapter 6. The factorization can be derived in two ways: via a constructive proof using realizations, and in a very straightforward way via a generalization of the classical Beurling-Lax theorem to the present context. This theorem also provides a second factorization: the *inner-outer* factorization

$$T = T_0 V$$

where, again, V is inner, and $T_0 \in \mathcal{U}$ is outer: it satisfies the range condition $\overline{\mathcal{U}_2 T_0} = \mathcal{U}_2$.

For *time-invariant* single-input single-output systems, the inner-outer factorization is a factorization of an analytical (causal) transfer function $T(z)$ into the product of an inner and an outer system: $T(z) = V(z)T_0(z)$. The inner factor $V(z)$ is analytical (*i.e.*, has its poles outside the unit disc) and has modulus 1 on the unit circle, whereas the outer factor $T_0(z)$ is analytical and may have zeros outside the unit disc only.[1] Such functions are called *minimum phase* in engineering. For example, (with $|\alpha|, |\beta| < 1$)

$$z\frac{z-\alpha^*}{1-\beta z} = z\frac{z-\alpha^*}{1-\alpha z}\cdot\frac{1-\alpha z}{1-\beta z}.$$

The resulting outer factor is such that its inverse is again a stable system, provided there are no zeros on the unit circle. For multi-input multi-output systems, the definition of the outer factor is more abstract (see *e.g.*, Halmos [1]) and takes the form of a range condition: $T_0(z)$ is outer if $T_0(z)H_m^2 = H_n^2$, where H_m^2 is the Hardy space of analytical m-dimensional vector-valued functions. A generalization of this definition applies in the time-varying context.

The existence of inner-outer factorizations in any context is more or less fundamental to analytical Hilbert spaces. There are abstract mathematical formulations of it which also apply to the time-varying setting (for example, [2, 3]), but a computational scheme acting on state-space realizations was still lacking up to now. One of the aspects of time-varying systems is that the state dimension can vary, and therefore, the number of 'zeros' in the inner and outer factors can vary, too. The theory in this chapter handles such variations automatically.[2]

An application of the inner-outer factorization is the computation of inverse systems: if T is a causal and invertible system, then its inverse is not necessarily causal: the inversion

[1] Here, we use the mathematical convention to write the z-transform as a series in z rather than z^{-1}.

[2] The material in this chapter has been submitted for publication in [4].

might have introduced an anti-causal part. This effect is known as a dichotomy; it is in general not a trivial task to determine the causal and anti-causal parts of T^{-1}. With the inner-outer factorization, however, the inverse of the outer factor is again causal, whereas the inverse of the inner factor is fully anti-causal, and determines which part of the inverse outer factor is made anti-causal. This application of the inner-outer factorization plays a crucial role, *e.g.*, in the computation of optimal feedback controllers [5]. This application is briefly discussed in chapter 10.

State-space properties

For an operator $T \in \mathcal{U}$, we defined the input/output state and null spaces in chapter 3 in terms of the ranges and kernels of the Hankel operator H_T and its adjoint (equations (3.45), (3.47)):

$$\begin{array}{lcll} \mathcal{K}(T) & = & \ker(H_T) & = \{U \in \mathcal{L}_2Z^{-1} : \mathbf{P}(UT) = 0\} \\ \mathcal{H}(T) & = & \mathrm{ran}(H_T^*) & = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T^*) \\ \mathcal{H}_0(T) & = & \mathrm{ran}(H_T) & = \mathbf{P}(\mathcal{L}_2Z^{-1}T) \\ \mathcal{K}_0(T) & = & \ker(H_T^*) & = \{Y \in \mathcal{U}_2 : \mathbf{P}(YT^*) = 0\}. \end{array}$$

These subspaces provide decompositions of $\mathcal{L}_2Z^{-1}$ and $\mathcal{U}_2$ as

$$\begin{array}{lcll} \overline{\mathcal{H}}(T) & \oplus & \mathcal{K}(T) & = \mathcal{L}_2Z^{-1} \\ \overline{\mathcal{H}}_0(T) & \oplus & \mathcal{K}_0(T) & = \mathcal{U}_2. \end{array}$$

For inner operators V, the null spaces take on a more specific structure.

PROPOSITION 4.1. *Let $V \in \mathcal{U}$ be an inner operator. Then*

$$\begin{array}{lcllcl} \mathcal{K}(V) & = & \mathcal{L}_2Z^{-1}V^*, & \mathcal{H}(V) & = & \mathcal{L}_2Z^{-1} \ominus \mathcal{L}_2Z^{-1}V^* \\ \mathcal{K}_0(V) & = & \mathcal{U}_2V, & \mathcal{H}_0(V) & = & \mathcal{U}_2 \ominus \mathcal{U}_2V. \end{array}$$

$\mathcal{H}$ and $\mathcal{H}_0$ are closed subspaces.

PROOF In general, for a bounded operator, $\mathcal{X}_2V \subset \mathcal{X}_2$. From $VV^* = I$, it follows that $\mathcal{X}_2 = \mathcal{X}_2V^* = (\mathcal{L}_2Z^{-1} \oplus \mathcal{U}_2)V^*$. Because $V^*V = I$, $\mathcal{L}_2Z^{-1}V^* \perp \mathcal{U}_2V^*$, so that

$$\mathcal{X}_2 = \mathcal{L}_2Z^{-1}V^* \oplus \mathcal{U}_2V^*.$$

Both $\mathcal{L}_2Z^{-1}V^*$ and $\mathcal{U}_2V$ are closed subspaces, and because $V \in \mathcal{U}$, $\mathcal{L}_2Z^{-1}V^* \subset \mathcal{L}_2Z^{-1}$. Projecting onto $\mathcal{L}_2Z^{-1}$ yields that

$$\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2V^*) = \mathcal{L}_2Z^{-1} \ominus \mathcal{L}_2Z^{-1}V^*$$

is a closed subspace, so that $\mathcal{H}$ is closed. Hence

$$\mathcal{L}_2Z^{-1} = \mathcal{L}_2Z^{-1}V^* \oplus \mathcal{H}, \tag{4.1}$$

so that $\mathcal{K} = \mathcal{L}_2 Z^{-1} V^*$. The result on $\mathcal{K}_0$ follows in the same way by dual arguments. □

COROLLARY 4.2. *Let $V \in \mathcal{U}$ be an inner operator. Then*

$$\begin{aligned} \mathcal{H}_0 &= \mathcal{H}V \\ \mathcal{H} &= \mathcal{H}_0 V^*. \end{aligned}$$

PROOF From (4.1), it follows immediately that

$$\mathcal{L}_2 Z^{-1} \oplus \mathcal{H}V = \mathcal{L}_2 Z^{-1} V.$$

Hence $\mathcal{H}V \subset \mathcal{U}_2$, and $\mathcal{H}V = \mathbf{P}(\mathcal{L}_2 Z^{-1} V) = \mathcal{H}_0$. Dually, we have that $\mathcal{H} = \mathcal{H}_0 V^*$. □

For general operators T, we had already that $\mathcal{H}_0 = \mathbf{P}(\mathcal{H}T)$. The fact that, for inner operators, $\mathcal{H}_0 = \mathcal{H}V$ shows that the Hankel operator of V, H_V, satisfies $\cdot H_V = \cdot V$ on $\mathcal{H}$. Since $\cdot H_V = 0$ on $\mathcal{K}$, it is seen that H_V is an isometry. Consequently, its non-zero singular values are all equal to 1: $\hat{\mathbf{\Sigma}}$ in the SVD-based factorization $H_V = \mathbf{P}_0(\cdot\, \mathbf{Q}^*)\hat{\mathbf{\Sigma}}\mathbf{G}$ (equation (3.71)) is $\hat{\mathbf{\Sigma}} = I$.

COROLLARY 4.3. *Let $V \in \mathcal{U}$ be a locally finite inner operator. If $\mathbf{Q}$ is an orthonormal basis representation of the input state space $\mathcal{H}$ of V, then $\mathbf{G} = \mathbf{Q}V$ is an orthonormal basis representation of its output state space $\mathcal{H}_0$, and the canonical controller realization based on $\mathbf{Q}$ (theorem 3.17) and the canonical observer realization based on $\mathbf{G}$ (theorem 3.23) are equal.*

PROOF According to theorem 3.27, a basis of $\mathcal{H}_0$ is obtained as $\mathbf{F}_0 = \mathbf{P}(\mathbf{Q}V)$. Because $\mathcal{H}_0 = \mathcal{H}V$, it follows that $\mathbf{F}_0 = \mathbf{P}(\mathbf{Q}V) = \mathbf{Q}V = \mathbf{G}$. $\mathbf{G}$ is an orthonormal basis of $\mathcal{H}_0$, because $\Lambda_{\mathbf{G}} = \mathbf{P}_0(\mathbf{Q}VV^*\mathbf{Q}^*) = \Lambda_{\mathbf{Q}} = I$. The canonical realizations are obtained from theorems 3.17 and 3.23, respectively, as

$$\mathbf{V} = \begin{bmatrix} \mathbf{P}_0(Z^{-1}\mathbf{Q}\mathbf{Q}^*)^{(-1)} & \mathbf{P}_0(\mathbf{Q}V) \\ \mathbf{P}_0(Z^{-1}\mathbf{Q}^*)^{(-1)} & \mathbf{P}_0(V) \end{bmatrix} \quad \text{and} \quad \mathbf{V}' = \begin{bmatrix} \mathbf{P}_0(Z^{-1}\mathbf{G}\mathbf{G}^*)^{(-1)} & \mathbf{P}_0(\mathbf{G}) \\ \mathbf{P}_0(Z^{-1}V\mathbf{G}^*)^{(-1)} & \mathbf{P}_0(V) \end{bmatrix}. \tag{4.2}$$

The fact that both realizations are equal follows directly by inserting $\mathbf{G} = \mathbf{Q}V$. □

If V is not inner, but rather an isometry, then the properties listed in proposition 4.1 and corollary 4.2 hold only partially:

PROPOSITION 4.4. *Let $V \in \mathcal{U}$. Then*

$$VV^* = I \quad \Rightarrow \quad \begin{cases} \mathcal{K}_0 = \mathcal{U}_2 V \oplus \ker(\cdot V^*|_{\mathcal{U}_2}), \\ \ker(\cdot V^*|_{\mathcal{U}_2}) = 0 \quad \Rightarrow \quad V \text{ is inner} \\ \mathcal{H} = \overline{\mathcal{H}_0 V^*} \\ \overline{\mathcal{U}_2 V^*} = \mathcal{U}_2 \oplus \overline{\mathcal{H}} \end{cases}$$

$$V^*V = I \quad \Rightarrow \quad \begin{cases} \mathcal{K} = \mathcal{L}_2 Z^{-1} V^* \oplus \ker(\cdot V|_{\mathcal{L}_2 Z^{-1}}). \\ \ker(\cdot V|_{\mathcal{L}_2 Z^{-1}}) = 0 \quad \Rightarrow \quad V \text{ is inner} \\ \mathcal{H}_0 = \overline{\mathcal{H}} V \\ \overline{\mathcal{L}_2 Z^{-1} V} = \mathcal{L}_2 Z^{-1} \oplus \overline{\mathcal{H}_0} \end{cases}$$

PROOF Let $VV^* = I$. Because V is an isometry, the subspace $\mathcal{X}_2 V = \mathrm{ran}(V)$ is closed. Because $\mathcal{X}_2 V = \mathcal{L}_2 Z^{-1} V \oplus \mathcal{U}_2 V$, both $\mathcal{U}_2 V$ and $\mathcal{L}_2 Z^{-1} V$ are closed subspaces.

$\mathcal{U}_2 V \subset \mathcal{K}_0$, because $\mathbf{P}_{\mathcal{L}_2 Z^{-1}}([\mathcal{U}_2 V]\, V^*) = 0$. The remaining subspace $\mathcal{K}_0 \ominus \mathcal{U}_2 V$ consists of elements

$$\begin{aligned} \mathcal{K}_0 \ominus \mathcal{U}_2 V &= \{X \in \mathcal{U}_2 : \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(XV^*) = 0 \ \wedge \ \mathbf{P}(XV^*) = 0\} \\ &= \{X \in \mathcal{U}_2 : XV^* = 0\} \\ &= \ker(\cdot V^*|_{\mathcal{U}_2}). \end{aligned}$$

Hence $\mathcal{K}_0 = \mathcal{U}_2 V \oplus \ker(\cdot V^*|_{\mathcal{U}_2})$.

To prove $\mathcal{H} = \overline{\mathcal{H}_0 V^*}$, take $U \in \mathcal{L}_2 Z^{-1}$. Then $UV = U_1 + Y$, where $U_1 \in \mathcal{L}_2 Z^{-1}$ and $Y = \mathbf{P}(UV) \in \mathcal{H}_0 \subset \mathcal{U}_2$. All of $\mathcal{H}_0$ can be reached by Y if U ranges over $\mathcal{L}_2 Z^{-1}$. Multiplication by V^* gives $U = U_1 V^* + YV^*$, and since $V^* \in \mathcal{L}$, it follows that $YV^* \in \mathcal{L}_2 Z^{-1}$, and this is true for all $Y \in \mathcal{H}_0$. Hence $\mathcal{H}_0 V^* \subset \mathcal{L}_2 Z^{-1}$ and also

$$\overline{\mathcal{H}_0 V^*} \subset \mathcal{L}_2 Z^{-1}.$$

Since $\mathcal{H} = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 V^*) = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\overline{\mathcal{H}_0 V^*})$, we obtain $\mathcal{H} = \overline{\mathcal{H}_0 V^*}$.

The expressions for $\mathcal{H}_0$ and $\mathcal{K}_0$ combined give

$$\mathcal{U}_2 = \overline{\mathcal{H}_0} \oplus \mathcal{U}_2 V \oplus \ker(\cdot V^*|_{\mathcal{U}_2})$$

hence $\mathcal{U}_2 V^* = \overline{\mathcal{H}_0 V^*} + \mathcal{U}_2$. Because $\overline{\mathcal{H}_0 V^*} = \mathcal{H} \in \mathcal{L}_2 Z^{-1}$, the two components are actually orthogonal.

If $\ker(\cdot V^*|_{\mathcal{U}_2}) = 0$, then

$$X \in \mathcal{U}_2, \quad XV^* = 0 \quad \Rightarrow \quad X = 0.$$

This implies

$$X \in Z^n \mathcal{U}_2, \quad XV^* = 0 \quad \Rightarrow \quad X = 0 \qquad (\text{all } n \in \mathbf{Z}),$$

since $(Z^n X)V^* = 0 \Leftrightarrow XV^* = 0$. Letting $n \to \infty$ yields $\ker(\cdot V^*) = 0$, so that V has a left inverse, which must be equal to the right inverse V^*. Hence $V^*V = I$ and V is inner.

Dual results hold in case $V^*V = I$. □

Unitary realizations

The purpose of this section is twofold. We show that (*i*) if a locally finite operator V is inner, then it has a unitary realization $\mathbf{V}$ (which is obtained by a canonical realization based on $\mathbf{Q}$ or $\mathbf{G}$); and conversely, (*ii*) if $\mathbf{V}$ is a unitary realization with $\ell_A < 1$, then the corresponding operator V is inner.

THEOREM 4.5. *Let $V \in \mathcal{U}$ be a locally finite inner operator. Let $\mathbf{Q}$ be an orthonormal basis representation for $\mathcal{H}(V)$.*

The canonical controller realization $\mathbf{V}$ based on $\mathbf{Q}$ is unitary ($\mathbf{V}\mathbf{V}^* = I, \mathbf{V}^*\mathbf{V} = I$), and identical to the canonical observer realization based on $\mathbf{G} = \mathbf{Q}V$.

PROOF Let $\mathbf{V}$ be given by the canonical realization (4.2). This realization satisfies the properties (lemma 3.19):

$$Z\mathbf{Q} = A^*\mathbf{Q} + B^*, \tag{4.3}$$

$$A^*A + B^*B = I. \tag{4.4}$$

$$\begin{aligned} \mathbf{P}_0(\cdot V) &= \mathbf{P}_0(\cdot\,[D + \mathbf{Q}^*C]) \\ V^* &= D^* + C^*\mathbf{Q}. \end{aligned} \tag{4.5}$$

To verify that $\mathbf{V}^*\mathbf{V} = I$, we have to show the following three relations:

$$\begin{aligned} A^*A + B^*B &= I \\ C^*C + D^*D &= I \\ A^*C + B^*D &= 0 \end{aligned}$$

$A^*A + B^*B = I$ is equal to equation (4.4). To prove $C^*C + D^*D = I$, use equation (4.5) and the fact that $\mathbf{Q}$ is strictly lower:

$$\begin{aligned} \mathbf{P}_0(V^*V) = I \quad \Rightarrow \quad & \mathbf{P}_0([D^* + C^*\mathbf{Q}][D + \mathbf{Q}^*C]) \\ & = D^*D + C^*\mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*)C \\ & = D^*D + C^*C = I. \end{aligned}$$

To prove $A^*C + B^*D = 0$, use $\mathbf{P}_0(ZQV) = \mathbf{P}_0(ZG) = 0$ and equations (4.3), (4.5):

$$\begin{aligned} \mathbf{P}_0(ZQV) = 0 \quad \Rightarrow \quad & \mathbf{P}_0([B^* + A^*\mathbf{Q}][D + \mathbf{Q}^*C]) \\ & = B^*D + A^*\mathbf{P}_0(\mathbf{Q}\mathbf{Q}^*)C \\ & = B^*D + A^*C = 0. \end{aligned}$$

Hence $\mathbf{V}^*\mathbf{V} = I$. Dually, we have for the realization $\mathbf{V}'$ in equation (4.2) the properties listed in lemma 3.24:

$$\mathbf{G} = C + AZ\mathbf{G}, \tag{4.6}$$

$$AA^* + CC^* = I.$$

$$\begin{aligned} \mathbf{P}_0(\cdot V^*) &= \mathbf{P}_0(\cdot D^*) + \mathbf{P}_0(Z^{-1} \cdot \mathbf{G}^*)^{(-1)} B^* \\ V &= D + BZ\mathbf{G}. \end{aligned} \tag{4.7}$$

We have to show $\mathbf{V}'\mathbf{V}'^* = I$, *i.e.*,

$$\begin{aligned} AA^* + CC^* &= I \\ BB^* + DD^* &= I \\ AB^* + CD^* &= 0. \end{aligned}$$

The first equation is again immediate. To prove $BB^* + DD^* = I$, use (4.7) and $VV^* = I$:

$$\begin{aligned} \mathbf{P}_0(VV^*) = I \quad \Rightarrow \quad & \mathbf{P}_0([D + BZ\mathbf{G}]D^*) + \mathbf{P}_0(Z^{-1}[D + BZ\mathbf{G}]\mathbf{G}^*)^{(-1)}B^* \\ & = DD^* + B\mathbf{P}_0(\mathbf{G}\mathbf{G}^*)^{(-1)}B^* \\ & = DD^* + BB^* = I. \end{aligned}$$

Finally, to prove $AB^* + CD^* = 0$, use $\mathbf{P}_0(\mathbf{G}V^*) = \mathbf{P}_0(\mathbf{Q}) = 0$, and equations (4.6), (4.7):

$$\begin{aligned} \mathbf{P}_0(\mathbf{G}V^*) = 0 \quad \Rightarrow \quad & \mathbf{P}_0([C + AZ\mathbf{G}]D^*) + \mathbf{P}_0(Z^{-1}[C + AZ\mathbf{G}]G^*)^{(-1)}B^* \\ & = CD^* + A\mathbf{P}_0(\mathbf{G}\mathbf{G}^*)^{(-1)}B^* \\ & = CD^* + AB^* = 0. \end{aligned}$$

Hence $\mathbf{V}'\mathbf{V}'^* = I$. Since $\mathbf{V} = \mathbf{V}'$ if $\mathbf{G} = \mathbf{Q}V$ (corollary 4.3), it follows that $\mathbf{V}$ is inner. □

The converse of this theorem is in general true only if, in addition, $\ell_A < 1$. If $\ell_A = 1$, then additional assumptions on the controllability and observability of the realization must be made.

THEOREM 4.6. *Let* $\mathbf{V} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$ *be a state realization of a bounded transfer operator* V. *Let* $\Lambda_{\mathbf{F}}$ *and* $\Lambda_{\mathbf{F}_0}$ *be the controllability and the observability Gramians of the given realization. If* $\ell_A < 1$, *then*

$$\begin{aligned} \mathbf{V}^*\mathbf{V} = I \quad &\Rightarrow \quad V^*V = I, \quad \Lambda_{\mathbf{F}} = I, \\ \mathbf{V}\mathbf{V}^* = I \quad &\Rightarrow \quad VV^* = I, \quad \Lambda_{\mathbf{F}_0} = I. \end{aligned} \tag{4.8}$$

If $\ell_A \leq 1$, *then*

$$\begin{aligned} \mathbf{V}^*\mathbf{V} = I, \quad \Lambda_{\mathbf{F}} = I \quad &\Rightarrow \quad V^*V = I, \\ \mathbf{V}\mathbf{V}^* = I, \quad \Lambda_{\mathbf{F}_0} = I \quad &\Rightarrow \quad VV^* = I. \end{aligned}$$

PROOF If $\ell_A < 1$, then $\mathbf{V}^*\mathbf{V} = I$ implies *a.o.* $A^*A + B^*B = I$. This expression can be compared with the Lyapunov equation for $\mathbf{F}$: $A^*\Lambda_{\mathbf{F}}A + B^*B = \Lambda_{\mathbf{F}}^{(-1)}$. Since $\ell_A < 1$, the equation has a unique solution, which must be $\Lambda_{\mathbf{F}} = I$. A dual result holds for $\Lambda_{\mathbf{F}_0}$ in case $\mathbf{V}\mathbf{V}^* = I$.

Assume $\mathbf{V}^*\mathbf{V} = I$ and $\Lambda_{\mathbf{F}} = I$. Since it is an orthonormal basis, we write $\mathbf{Q}$ for $\mathbf{F}$ from now on. Equations (4.5) hold:

$$\begin{aligned} \mathbf{P}_0(\cdot\, V) &= \mathbf{P}_0(\cdot\, [D + \mathbf{Q}^* C]) \\ V^* &= D^* + C^* \mathbf{Q}. \end{aligned}$$

To show $V^*V = I$, we show that $\mathbf{P}_0(Z^{-n}V^*V)$ is $= I$ for $n = 0$, and $= 0$ otherwise. For $n = 0$:

$$\begin{aligned} \mathbf{P}_0(V^*V) &= \mathbf{P}_0(\,[D^* + C^*\mathbf{Q}]\,[D + \mathbf{Q}^*C]\,) \\ &= \mathbf{P}_0(D^*D) + \mathbf{P}_0(D^*\mathbf{Q}^*C) + \mathbf{P}_0(C^*\mathbf{Q}D) + \mathbf{P}_0(C^*\mathbf{Q}\mathbf{Q}^*C) \\ &= D^*D + C^*C = I. \end{aligned}$$

For $n > 0$,

$$\begin{aligned} \mathbf{P}_0(Z^{-n}V^*V) &= \mathbf{P}_0(Z^{-n}[D^* + C^*\mathbf{Q}]\,[D + \mathbf{Q}^*C]\,) \\ &= \mathbf{P}_0(Z^{-n}D^*D) + \mathbf{P}_0(Z^{-n}D^*\mathbf{Q}^*C) + \mathbf{P}_0(Z^{-n}C^*\mathbf{Q}D) + \mathbf{P}_0(Z^{-n}C^*\mathbf{Q}\mathbf{Q}^*C) \end{aligned}$$

Using equations (3.60) and (3.62), *viz.*

$$\begin{aligned} \mathbf{P}_0(Z^{-n}\mathbf{Q}\mathbf{Q}^*) &= A^{\{n\}} && (n \geq 0) \\ \mathbf{P}_0(Z^{-n}\mathbf{Q}^*) &= B^{(n)}A^{\{n-1\}} && (n > 0), \end{aligned}$$

gives

$$\begin{aligned} \mathbf{P}_0(Z^{-n}V^*V) &= 0 + 0 + D^{*(n)}B^{(n)}A^{\{n-1\}}C + C^{*(n)}A^{\{n\}}C \\ &= [D^*B + C^*A]^{(n)}A^{\{n-1\}}C \\ &= 0. \end{aligned}$$

Taking adjoints shows that $\mathbf{P}_0(Z^{-n}V^*V) = 0$ for $n < 0$, too. Hence $V^*V = I$.
The fact $[\mathbf{V}\mathbf{V}^* = I,\ \Lambda_{\mathbf{F}_0} = I] \Rightarrow VV^* = I$ can be shown in a dual way. □

If $\ell_A < 1$, then more elementary means suffice to verify the theorem: one can evaluate $I - V^*V$ and $I - VV^*$. The proof goes as follows.

$$\begin{aligned} I - V^*V &= I - [D + BZ(I - AZ)^{-1}C]\,[D + BZ(I - AZ)^{-1}C] \\ &= I - D^*D - C^*(I - Z^*A^*)^{-1}Z^*B^*D - D^*BZ(I - AZ)^{-1}C \\ &\quad - C^*(I - Z^*A^*)^{-1}Z^*B^*BZ(I - AZ)^{-1}C \\ &= I - D^*D + C^*(I - Z^*A^*)^{-1}Z^*A^*C + C^*AZ(I - AZ)^{-1}C + \\ &\quad - C^*(I - Z^*A^*)^{-1}Z^*\{I - A^*A\}Z(I - AZ)^{-1}C \\ &= C^*C + C^*(I - Z^*A^*)^{-1}\{Z^*A^* + AZ - I - Z^*A^*AZ\}(I - AZ)^{-1}C \end{aligned}$$

since $B^*D = -A^*C$, $B^*B = I - A^*A$ and $I - D^*D = C^*C$, and hence

$$\begin{aligned} I - V^*V &= C^*(I - Z^*A^*)^{-1}\{(I - Z^*A^*)(I - AZ) + \\ &\quad + Z^*A^* + AZ - I - Z^*A^*AZ\}(I - AZ)^{-1}C \\ &= 0. \end{aligned}$$

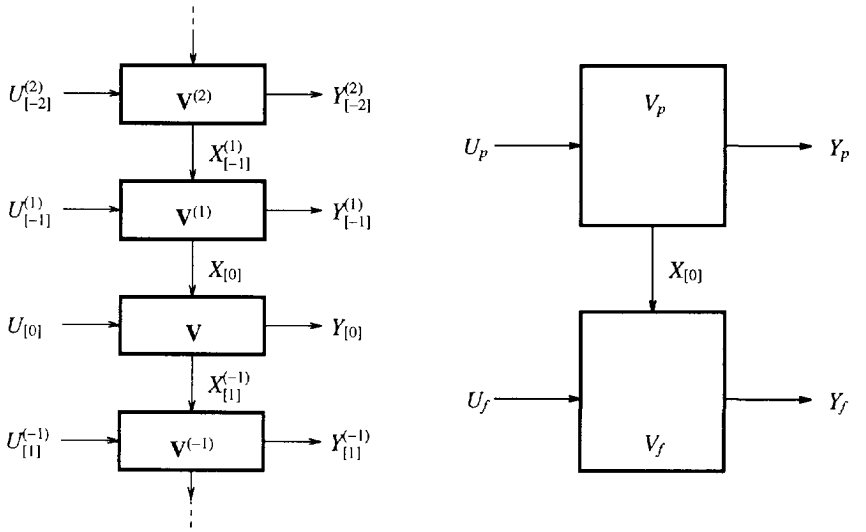


Figure 4.1. A lossless realization.

$I - VV^* = 0$ is verified by an analogous procedure.

Theorems 4.5 and 4.6 have an interpretation in terms of conservation of energy, and these interpretations lead to alternative proofs. Let $\mathbf{V}$ be a realization for some bounded operator, such that $\mathbf{V}\mathbf{V}^* = I$. With $[X^{(-1)}_{[k+1]} \;\; Y_{[k]}] = [X_{[k]} \;\; U_{[k]}]\mathbf{V}$, this property ensures that, for each k,

$$\| [X^{(-1)}_{[k+1]} \;\; Y_{[k]}] \|^2_{HS} = \| [X_{[k]} \;\; U_{[k]}] \|^2_{HS} \tag{4.9}$$

Summing this equation over all k yields

$$\| Y \|^2_{HS} + \| X \|^2_{HS} = \| U \|^2_{HS} + \| X \|^2_{HS}.$$

If $\ell_A < 1$, then $X \in \mathcal{X}_2$ so that $\| X \|^2_{HS} < \infty$, and it follows that $\| Y \|_{HS} = \| U \|_{HS}$, so that $VV^* = I$. In the case where $\ell_A = 1$, $\| X \|^2_{HS}$ can be unbounded: energy can remain in the state $X_{[k]}$ for $k \to \infty$, so that the system is not lossless. If the realization has observability Gramian equal to I, this can in fact not occur, but observability cannot be determined from $AA^* + CC^* = I$ if $\ell_A = 1$.

An alternative proof of theorem 4.5 is as follows. Let $VV^* = I$ and let $\mathbf{V}$ be a realization based on an orthonormal basis representation $\mathbf{G}$. As in equation (3.36), see figure 4.1, the computation of $Y = UV$ can be split into a past and a future part, according to the decomposition $U = U_p + U_f \in \mathcal{L}_2 Z^{-1} \oplus \mathcal{U}_2$, and a similar decomposition of Y. The state $X_{[0]}$ is given by $X_{[0]} = \mathbf{P}_0(U_p V \mathbf{G}^*)$. We show that the energy conservation relation between

the past and future parts can be expressed as

$$\begin{array}{rcl} \| U_p \|_{HS}^2 & = & \| Y_p \|_{HS}^2 + \| X_{[0]} \|_{HS}^2 \\ \| Y_f \|_{HS}^2 & = & \| U_f \|_{HS}^2 + \| X_{[0]} \|_{HS}^2 \end{array} \tag{4.10}$$

Indeed, the first equation can be proven by taking $U = U_p$ ($U_f = 0$). Then $Y = Y_p + Y_f = Y_p + X_{[0]}\mathbf{G}$, so that $\| Y \|^2 = \| Y_p \|^2 + \| Y_f \|^2 = \| Y_p \|^2 + \| X_{[0]} \|^2$. Since $VV^* = I$, $\| Y \|^2 = \| U_p \|^2$, which proves the first equation in (4.10). The result also holds when $U_f \neq 0$, because the realization is causal. The second equation is derived in much the same way: $Y_f = \mathbf{P}(U_p V) + U_f V$. Because $\mathbf{P}(U_p V) = X_{[0]}\mathbf{G} \in \mathcal{H}_0$ and $U_f V \in \mathcal{K}_0$, these two components are orthogonal and $\| Y_f \|^2 = \| X_{[0]} \|^2 + \| U_f \|^2$.

Equation (4.10) is obtained for a decomposition into past and future parts with respect to the 0-th diagonal. Generalizing for the k-th diagonal, equation (4.9) can be obtained by taking differences between the relations at point k and at point $k+1$. Since the realization is minimal, $X_{[k]}$ can take on all possible values: $[X_{[k]} \quad U_{[k]}]$ is full range. Then (4.9) implies $\mathbf{V}\mathbf{V}^* = I$.

Example

As an example, let $V \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be given by

$$V = \begin{bmatrix} \boxed{d_0} & b_0 & 0 & 0 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ 0 & 0 & \underline{d_2} & b_2 & 0 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdots \\ 0 & 0 & 0 & 0 & \underline{d_4} & b_4 & \cdots \\ & & & & & & \ddots \end{bmatrix} \qquad \begin{array}{rcl} \#\mathcal{M} & = & [1\ 0\ 1\ 0\ 1\ 0\ \cdots] \\ \#\mathcal{N} & = & [1\ 1\ 1\ 1\ 1\ 1\ \cdots] \\ \#\mathcal{B} & = & [0\ 1\ 0\ 1\ 0\ 1\ \cdots], \end{array}$$

where $d_i^2 + b_i^2 = 1$ (the underlined entries form the main diagonal). V is an isometry: $VV^* = I$. It has an isometric realization, $\mathbf{V}\mathbf{V}^* = I$, given by

$$\mathbf{V}_i = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline b_i & d_i \end{array}\right] \quad \text{(even } i\text{)}, \qquad \mathbf{V}_i = \left[\begin{array}{c|c} \cdot & 1 \\ \hline \cdot & \cdot \end{array}\right] \quad \text{(odd } i\text{)}.$$

See figure 4.2. Let $b_i \to 0$, for $i \to \infty$. Then the output state space $\mathcal{H}_0(V) = \mathbf{P}(\mathcal{L}_2 Z^{-1} V)$ is not a closed subspace: it is the range of the Hankel operator H_V of V, with snapshots

$$(H_V)_i = 0 \quad \text{(even } i\text{)}, \qquad (H_V)_i = \begin{bmatrix} b_{i-1} & 0 & \cdots \\ 0 & 0 & \\ \vdots & & \ddots \end{bmatrix} \quad \text{(odd } i\text{)}.$$

The row range of $(H_V)_i$ determines i-th slice of $\mathcal{H}_0(V)$. For odd i, the Hankel matrix has rank 1, but the range of the whole collection is not closed because $b_i \to 0$ but never becomes equal to 0.

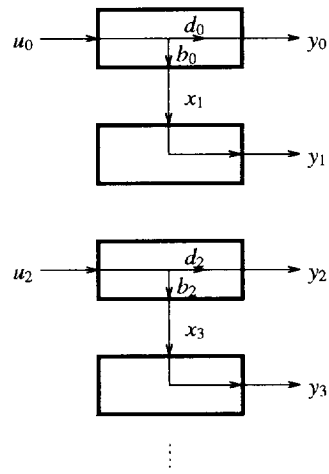


Figure 4.2. A simple isometric system.

V can be extended to an inner operator V, by adding extra inputs. This is straightforwardly done by completing each realization matrix $\mathbf{V}_i$ to a unitary matrix $\mathbf{W}_i$, which yields

$$\mathbf{W}_i = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline b_i & d_i \\ -d_i & b_i \end{array}\right] \quad (\text{even } i), \qquad \mathbf{W}_i = \left[\begin{array}{c|c} \cdot & 1 \\ \hline \cdot & \cdot \end{array}\right] \quad (\text{odd } i).$$

$$W = \begin{bmatrix} \boxed{\begin{matrix} d_0 \\ b_0 \end{matrix}} & \begin{matrix} b_0 \\ -d_0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \cdots \\ \begin{matrix} \cdot \\ 0 \end{matrix} & \begin{matrix} \cdot \\ 0 \end{matrix} & \begin{matrix} \cdot \\ \underline{d_2} \end{matrix} & \begin{matrix} \cdot \\ b_2 \end{matrix} & \begin{matrix} \cdot \\ 0 \end{matrix} & \cdots \\ \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \end{matrix} & \begin{matrix} \underline{b_2} \\ \cdot \end{matrix} & \begin{matrix} -d_2 \\ \cdot \end{matrix} & \begin{matrix} 0 \\ \cdot \end{matrix} & \cdots \\ & & & & \ddots & \end{bmatrix}$$

$$\begin{aligned} \#\mathcal{M}_W &= [2\ 0\ 2\ 0\ 2\ 0\ \cdots] \\ \#\mathcal{N}_W &= [1\ 1\ 1\ 1\ 1\ 1\ \cdots] \\ \#\mathcal{B}_W &= [0\ 1\ 0\ 1\ 0\ 1\ \cdots], \end{aligned}$$

W satisfies $WW^* = I_{\mathcal{M}_W}$ and $W^*W = I_{\mathcal{N}_W}$. Its output state space is closed, and it is the closure of the output state space of V: $\mathcal{H}_0(W) = \overline{\mathcal{H}_0(V)}$. Indeed, the snapshots of the Hankel operator of W are given by

$$(H_W)_i = 0 \quad (\text{even } i), \qquad (H_W)_i = \begin{bmatrix} b_{i-1} & 0 & \cdots \\ d_{i-1} & 0 & \\ 0 & 0 & \\ \vdots & & \ddots \end{bmatrix} \quad (\text{odd } i),$$

and each odd Hankel operator snapshot has one nonzero singular value, equal to 1.

4.2 INNER FACTORIZATIONS

External factorization

Let $T \in \mathcal{U}$ be some transfer operator. We call an *external factorization* a factorization of the form

$$T = \Delta^* V,$$

where $\Delta = VT^* \in \mathcal{U}$ and $V \in \mathcal{U}$ is an inner operator. The factorization is such that V is the inner system (of smallest possible degree) such that $\Delta = VT^*$ is upper. We show that if T has a locally finite state space and has a uniformly observable realization for which $\ell_A < 1$, then such a factorization exists and can be readily computed from this realization. V has the same output state space as T. However, if $\ell_A = 1$, then if V is constructed in the same way, it can happen that it is isometric rather than inner. An isometric V is not acceptable: although $\Delta = VT^*$ is upper, $T \neq \Delta^* V$. The construction of an inner V that meets both conditions gives systems with have a larger state space than T, and presumably V is not locally finite.

The external factorization is a 'poor man's' coprime factorization (or stable-proper factorization). For example, a left coprime factorization of $T \in \mathcal{X}$ is a factorization

$$T = M^{-1}N, \qquad M, N \in \mathcal{U},$$

such that $[M \;\; N]$ has a right inverse in $\mathcal{U}$. Without the latter constraint, the connection for $T \in \mathcal{U}$ is trivial: take $M = I$, $N = T$. For $T \in \mathcal{L}$, let $T_1 = T^* \in \mathcal{U}$, then an external factorization gives

$$T_1 = \Delta^* U \quad \Rightarrow \quad T = U^* \Delta$$

so that $T = M^{-1}N$ with $M = U$, $N = \Delta$. The omission of the constraint that $[M \;\; N]$ has a right inverse in $\mathcal{U}$ makes the external factorization easier to compute. We show later in this section that the factors Δ and U can be such that they are *inner coprime*, *i.e.*, such that they do not contain common inner factors. In this case, the external factorization can be called an inner-coprime factorization [6]. Time-varying coprime factorizations have been reported in [7, 8, 9].

To obtain a better understanding of the external (inner-coprime) factorization, consider the scalar time-invariant case. Let

$$T = \frac{z - \alpha^*}{1 - \beta z}, \qquad |\alpha|, |\beta| < 1.$$

Then T has an inner-coprime factorization as

$$T = \Delta^* V = \frac{z - \alpha^*}{z - \beta^*} \cdot \frac{z - \beta^*}{1 - \beta z}, \qquad \Delta = \frac{1 - \alpha z}{1 - \beta z}, \qquad V = \frac{z - \beta^*}{1 - \beta z}.$$

Hence the poles of T are collected in the inner factor V. These poles also appear as poles of Δ, unless they are matched by complementary zeros of T.

The following simple observation is crucial in the computation of the inner factor.

PROPOSITION 4.7. *Let be given operators $T \in \mathcal{U}$ and $V \in \mathcal{U}$. Then $\Delta := VT^*$ is upper if and only if $\mathcal{U}_2 V \subset \mathcal{K}_0(T)$.*

PROOF $\Delta \in \mathcal{U} \Leftrightarrow \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 \Delta) = 0$. Substitution of $\Delta = VT^*$ yields

$$\begin{aligned} \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 \Delta) &= \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 V T^*) \\ &\subset \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{K}_0(T) T^*) \\ &= 0. \end{aligned}$$

□

$\mathcal{K}_0(T)$ is the largest subspace in $\mathcal{U}_2$ which is mapped by T^* to upper. Hence the system V of lowest complexity such that $\Delta = VT^* \in \mathcal{U}$ is obtained if $\mathcal{U}_2 V = \mathcal{K}_0(T)$. If V is inner, then from proposition 4.1, we have that $\mathcal{K}_0(V) = \mathcal{U}_2 V$, which provides the following additional result.

COROLLARY 4.8. *If V is inner, then $\Delta = VT^*$ is upper if and only if $\overline{\mathcal{H}}_0(T) \subset \mathcal{H}_0(V)$.*

The next step to prove the existence of the external factorization is to construct an operator V such that $\mathcal{U}_2 V \subset \mathcal{K}_0(T)$, or, assuming V can be inner and in view of proposition 4.1, $\mathcal{H}_0(V) = \overline{\mathcal{H}}_0(T)$. This can be done in a state-space context, by acting in a local way on a realization of T. Let T be a locally finite operator in $\mathcal{U}$. We start from a realization of T in output normal form, such that

$$AA^* + CC^* = I, \tag{4.11}$$

which means that at each point k in time the equation $A_k A_k^* + C_k C_k^* = I$ is satisfied. Such a realization is obtained from a canonical observer realization (*viz.* lemma 3.24). We assume that $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$, with state-space sequence $\mathcal{B}$, so that $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$. For each time instant k, augment the state transition matrices $[A_k \quad C_k]$ of T with as many extra rows as needed to yield a unitary (hence square) matrix $\mathbf{V}_k$:

$$\mathbf{V}_k = \begin{matrix} \\ \mathcal{B}_k \\ (\mathcal{M}_V)_k \end{matrix} \begin{matrix} \begin{matrix} \mathcal{B}_{k+1} & \mathcal{N}_k \end{matrix} \\ \begin{bmatrix} A_k & C_k \\ (B_V)_k & (D_V)_k \end{bmatrix} \end{matrix}. \tag{4.12}$$

The added rows introduce a space $(\mathcal{M}_V)_k$ with dimensions satisfying $\#\mathcal{B}_k + \#(\mathcal{M}_V)_k = \#\mathcal{B}_{k+1} + \#\mathcal{N}_k$. From $A_k A_k^* + C_k C_k^* = I$ it follows that $\#\mathcal{B}_{k+1} + \#\mathcal{N}_k \geq \#\mathcal{B}_k$, hence $\#(\mathcal{M}_V)_k \geq 0$. Assemble the individual matrices $\{A_k, (B_V)_k, C_k, (D_V)_k\}$ in diagonal operators $\{A, B_V, C, D_V\}$,

and define V by taking the corresponding operator $\mathbf{V}$ as a state-space realization for V. When we assume that this realization of V is minimal (in this case: uniformly controllable), then V is inner, and because $\mathbf{T}$ and $\mathbf{V}$ have the same (A, C)-matrices, $\mathcal{H}_0(V) = \overline{\mathcal{H}}_0(T)$, as required to make $\Delta \in \mathcal{U}$.

Although the construction is the same whether $\ell_A < 1$ or $\ell_A = 1$, the proof that it yields an external factorization is less elementary (and only conditionally true) for the case $\ell_A = 1$, so these cases are treated separately.

THEOREM 4.9. *Let T be a locally finite operator in $\mathcal{U}$. If T has a realization that is uniformly observable and has $\ell_A < 1$, then there exists an inner operator V with a unitary realization $\mathbf{V}$ such that*

$$T = \Delta^* V$$

where $\Delta = VT^* \in \mathcal{U}$.

PROOF Under the present conditions on T, it has a minimal realization $\mathbf{T}$ which is in output normal form and has $\ell_A < 1$. Then the above construction gives a unitary realization $\mathbf{V}$. Since this realization has $\ell_A < 1$, theorem 4.6 ensures that $\mathbf{V}$ is a minimal realization and that the corresponding operator V is inner, by construction such that $\mathcal{H}_0(V) = \overline{\mathcal{H}}_0(T)$. Application of corollary 4.8 shows that $\Delta := VT^*$ is upper. Because V is inner, this implies that $T = \Delta^* U$. □

Instead of applying corollary 4.8, the fact that $\Delta = VT^*$ is upper can also be verified by a direct computation of Δ, in which we make use of the relations $AA^* + CC^* = I$, $B_V A^* + D_V C^* = 0$:

$$\begin{aligned}
\Delta = VT^* &= \left[D_V + B_V Z(I-AZ)^{-1}C\right]\left[D^* + C^*(I-Z^*A^*)^{-1}Z^*B^*\right] \\
&= \left[D_V + B_V Z(I-AZ)^{-1}C\right] D^* + \underline{D_V C^*}(I-Z^*A^*)^{-1}Z^*B^* + \\
&\quad + B_V Z(I-AZ)^{-1}\underline{CC^*}(I-Z^*A^*)^{-1}Z^*B^* \\
&= \left[D_V + B_V Z(I-AZ)^{-1}C\right] D^* - B_V A^*(I-Z^*A^*)^{-1}Z^*B^* + \\
&\quad + B_V Z(I-AZ)^{-1}(I-AA^*)(I-Z^*A^*)^{-1}Z^*B^*.
\end{aligned}$$

Now, we make use of the relation

$$\begin{aligned}
&Z(I-AZ)^{-1}(I-AA^*)(I-Z^*A^*)^{-1}Z^* \\
&\quad = (I-ZA)^{-1}Z(I-AA^*)(Z-A^*)^{-1} \\
&\quad = (I-ZA)^{-1} + A^*(Z-A^*)^{-1} \\
&= (I-ZA)^{-1} + A^*(I-Z^*A^*)^{-1}Z^*
\end{aligned}$$

which is easily verified by pre- and postmultiplying with $(I-ZA)$ and $(Z-A^*)$, respectively. Plugging this relation into the expression for Δ, it is seen that the anti-causal parts of the expression cancel, and we obtain

$$\begin{aligned}
\Delta &= \left[D_V + B_V Z(I-AZ)^{-1}C\right] D^* + B_V(I-ZA)^{-1}B^* \\
&= D_V D^* + B_V B^* + B_V Z(I-AZ)^{-1}(AB^* + CD^*).
\end{aligned}$$

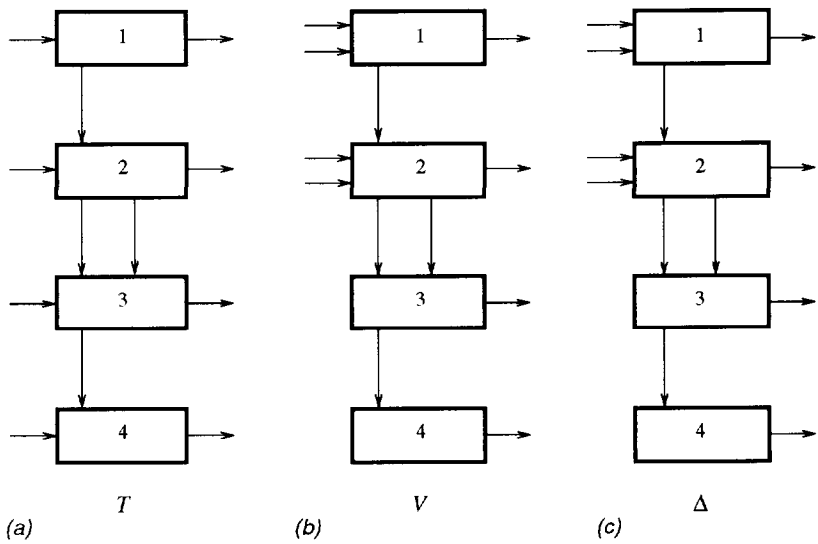


Figure 4.3. External factorization: (*a*) The structure of a state realization for an example T, (*b*) the structure of the corresponding inner factor V and (*c*) of Δ such that $T = \Delta^* V$.

Hence Δ has a realization

$$\mathbf{\Delta} = \begin{bmatrix} A & AB^* + CD^* \\ B_V & D_V D^* + B_V B^* \end{bmatrix}. \tag{4.13}$$

This realization is not necessarily minimal: if, for example, T is itself inner, then $B = B_V$ and $D = D_V$, so that $C_\Delta = 0$.

Because the A_k are not necessarily square matrices, the dimension of the state space may vary in time. A consequence of this is that the number of inputs of V varies in time for an inner V having minimal state dimension. The varying number of inputs of V are of course matched by a varying number of outputs of Δ^*. Figure 4.3 illustrates this point. (It is possible to extend V such that its state dimension becomes constant.)

If $\ell_A = 1$, then the proof becomes more complicated: the realization $\mathbf{V}$ need not be uniformly controllable, so that it is not minimal. Consequently, the corresponding operator V need not be inner, but it is in any case an isometry. It is possible to show that $\Delta := VT^*$ is again upper. However, if V is not inner, then $T \neq \Delta^* V$. Although it is possible to extend V in a minimal way to an inner system, *i.e.*, to construct an inner system W such that $\mathcal{H}_0(W) = \mathcal{H}_0(V)$, this does not really help because now $\Delta := WT^*$ is no longer upper. The remainder of this section is devoted to a derivation of these observations.

PROPOSITION 4.10. *Let $T \in \mathcal{U}$ be locally finite. Let* $\mathbf{T}$ *be a minimal realization of T in output normal form. Let* $\mathbf{V}$ *be as constructed in equation (4.12). Then V is an isometry ($VV^* = I$) and is such that $\Delta = VT^*$ is upper.*

If $\mathbf{V}$ *is a minimal realization, then V is inner and $T = \Delta^* V$.*

PROOF T has a minimal realization $\mathbf{T}$ in output normal form ($\Lambda_{\mathbf{G}} = I$), so that $\ell_A \leq 1$. Construct $\mathbf{V}$ as before. It is a unitary realization, and because it shares A and C with T, it also has the same observability operator $\mathbf{G}$. Hence theorem 4.6 ensures that the corresponding operator V is an isometry: $VV^* = I$.

From $\mathbf{V}\mathbf{V}^* = I$, we can prove $\Delta = VT^*$ is upper, by showing that $\mathbf{P}_0(Z^{-n}\Delta^*) = 0$ $(n > 0)$. To show this, use equation (4.7) and

$$\mathbf{P}_0(Z^{-n}\mathbf{G}\mathbf{G}^*) = A^{\{n\}}, \quad (n > 0); \qquad \mathbf{P}_0(Z^{-n}\mathbf{G}) = A^{\{n\}}C, \quad (n \geq 0).$$

(lemma 3.24). Then, indeed, for $n > 0$,

$$\begin{aligned}
\mathbf{P}_0(Z^{-n}\Delta^*) &= \mathbf{P}_0(Z^{-n}TV^*) \\
&= \mathbf{P}_0(Z^{-n}[D + BZ\mathbf{G}]D_V^*) + \mathbf{P}_0(Z^{-n-1}[D + BZ\mathbf{G}^*])^{(-1)}B_V^* \\
&= 0 + B^{(n)}\mathbf{P}_0(Z^{-n+1}\mathbf{G})D_V^* + 0 + B^{(n)}\mathbf{P}_0(Z^{-n}\mathbf{G}\mathbf{G}^*)^{(-1)}B_V^* \\
&= B^{(n)}A^{\{n-1\}}CD_V^* + B^{(n)}[A^{\{n\}}]^{(-1)}B_V^* \\
&= B^{(n)}A^{\{n-1\}}[CD_V^* + AB_V^*] \\
&= 0.
\end{aligned}$$

(By doing the same for $\mathbf{P}_0(Z^{-n}\Delta)$, we see that $\mathbf{\Delta}$ in (4.13) is still a model for Δ.)

For the case $\ell_A = 1$, the construction yields an operator V which is isometric and whose realization is uniformly observable (as inherited from $\mathbf{T}$). However, note that we have not shown that the realization is uniformly controllable, or even controllable at all: the realization might be too large. According to proposition 4.4, $\mathcal{K}_0(V) = \mathcal{U}_2 V \oplus \ker(\cdot\, V^*|_{\mathcal{U}_2})$, so that

$$\begin{aligned}
\mathcal{U}_2 &= \overline{\mathcal{H}}_0(T) \oplus \mathcal{K}_0(T) \\
&= \mathcal{D}_2\mathbf{G} \oplus \mathcal{K}_0(T) \\
&= \overline{\mathcal{H}}_0(V) \oplus \ker(\cdot\, V^*|_{\mathcal{U}_2}) \oplus \mathcal{U}_2 V \\
&= \overline{\mathcal{H}}_0(V) \oplus \mathcal{K}_0(V)
\end{aligned} \tag{4.14}$$

If $\mathbf{V}$ is minimal, then $\overline{\mathcal{H}}_0(V) = \mathcal{D}_2\mathbf{G}$, *i.e.*, $\ker(\cdot\, V^*|_{\mathcal{U}_2}) = 0$, so that V is inner (proposition 4.4). In this case, $\mathcal{H}_0(V)$ is closed and $\mathcal{H}_0(V) = \overline{\mathcal{H}}_0(T)$, so that $T = \Delta^* V$ with Δ upper.

□

The term $\ker(\cdot\, V^*|_{\mathcal{U}_2})$ represents the fact that V is an isometry, rather than unitary. It consists of elements Y in $\mathcal{U}_2$ for which $YV^* = 0$, so that there is no input U such that $UV = Y$. In the following section, we show how V can be extended with an isometry U ($UU^* = I$), defined by $\mathcal{U}_2 U = \ker(\cdot\, V^*|_{\mathcal{U}_2})$, so that $UV^* = 0$ and $V^*V + U^*U = I$. This can

be used to show that $T \neq \Delta^* V$ if V is not inner. Indeed, because $\mathcal{U}_2 U = \mathcal{K}_0(V) \ominus \mathcal{K}_0(T)$ (by equation (4.14)),

$$\begin{aligned} T = \Delta^* V \quad &\Leftrightarrow \quad T^* = V^* V T^* \\ &\Leftrightarrow \quad (I - V^* V) T^* = 0 \\ &\Leftrightarrow \quad U^* U T^* = 0 \\ &\Leftrightarrow \quad U T^* = 0 \\ &\Rightarrow \quad \mathcal{U}_2 U \subset \ker(\cdot\, T^*|_{\mathcal{U}_2}). \end{aligned}$$

But $\ker(\cdot\, T^*|_{\mathcal{U}_2}) \subset \mathcal{K}_0(T)$, while $\mathcal{U}_2 U = \mathcal{K}_0(V) \ominus \mathcal{K}_0(T)$. Hence $T \neq \Delta^* V$, and an isometric system is not suitable for an external factorization. $W = \begin{bmatrix} V \\ U \end{bmatrix}$ is inner, and is a minimal unitary extension of V in the sense that $\mathcal{K}_0(W) = \mathcal{U}_2 W = \mathcal{U}_2 V \oplus \mathcal{U}_2 U = \mathcal{K}_0(V)$, and $\mathcal{H}_0(W) = \overline{\mathcal{H}}_0(V)$. Because $\mathcal{H}_0(W) \subset \overline{\mathcal{H}}_0(T)$ (equation (4.14)), WT^* is not upper and W is also not suitable for an external factorization.

One remaining issue on the external factorization is to explain why (and when) it can be called inner coprime. Two upper operators T_1 and T_2 are called (left inner) coprime if they do not have a common left inner factor [6], *i.e.*, if

$$\begin{aligned} T_1 &= W T_1' \\ T_2 &= W T_2' \end{aligned}$$

(where $T_{1,2}' \in \mathcal{U}$ and W is inner) implies $W \in \mathcal{D}$. With this definition of inner coprimeness, it is possible to show that Δ and V in the factorization $T = \Delta^* V$ are inner coprime if $\mathcal{K}_0(T) = \mathcal{U}_2 V = \mathcal{K}_0(V)$. Indeed, suppose that they have a common left inner factor W, then $T = \Delta_1^* V_1$, where

$$\begin{aligned} \Delta_1 &= W^* \Delta \in \mathcal{U} \\ V_1 &= W^* V \in \mathcal{U}. \end{aligned}$$

On the one hand, $\mathcal{U}_2 V = \mathcal{U}_2 W V_1 \subset \mathcal{U}_2 V_1$. On the other hand, $\Delta_1 \in \mathcal{U} \Rightarrow \mathcal{U}_2 \Delta_1 = \mathcal{U}_2 [V_1 T^*] = [\mathcal{U}_2 V_1] T^* \subset \mathcal{U}_2$, hence $\mathcal{U}_2 V_1 \subset \mathcal{U}_2 V$, since $\mathcal{U}_2 V = \mathcal{K}_0(T)$ is the largest subspace in $\mathcal{U}_2$ that is mapped by T^* to $\mathcal{U}_2$. Combining both observations gives $\mathcal{U}_2 V_1 = \mathcal{U}_2 V$, so that V_1 is equal to V, up to a left diagonal unitary factor.

Beurling-Lax theorem

The existence of the external factorization was shown to depend on the construction of an inner operator V such that $\mathcal{U}_2 V$ is equal to some specified subspace $\mathcal{K}_0(T)$, the output null space of the system T. There is, however, a more fundamental result, which claims that any subspace $\mathcal{K}_0$ that is left D-invariant and Z-invariant (*i.e.*, such that $Z\mathcal{K}_0 \subset \mathcal{K}_0$) is of the form $\mathcal{U}_2 V$, for some isometric operator V. Such a theorem is known in the Hardy space setting as a Beurling-Lax theorem [10, 11, 1]. It not only provides the external factorization, but other factorizations as well, such as the inner-outer factorization.

In the theorem below, it follows that the input space sequence of $V \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ satisfying $\mathcal{K}_0 = \mathcal{U}_2^{\mathcal{M}} V$ is of locally finite dimension only if $\mathcal{K}_0 \ominus Z\mathcal{K}_0$ is a locally finite subspace.

Although $\mathcal{M}$ will be locally finite in the application to inner-outer factorization, we will prove theorem 4.11 for the more general situation. This calls for an extension of some of the definitions in chapter 2, to include operators with matrix representations whose entries are again operators. The extensions are straightforward (see [12]).

THEOREM 4.11. (BEURLINGLAX,I) *All DZ-invariant subspaces $\mathcal{K}_0$ in $\mathcal{U}_2^{\mathcal{N}}$ have the form $\mathcal{K}_0 = \mathcal{U}_2^{\mathcal{M}} V$, where $V \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ is an isometry ($VV^* = I$).*

PROOF Let $\mathcal{R}_0 = \mathcal{K}_0 \ominus Z\mathcal{K}_0$. This is a D-invariant subspace in $\mathcal{U}_2^{\mathcal{N}}$. We can assume it is non-empty, for else $\mathcal{K}_0 = Z\mathcal{K}_0 = Z^n\mathcal{K}_0$ for all $n \geq 0$, and since $X \in \mathcal{U}_2 \Rightarrow \lim_{n\to\infty} \mathbf{P}(Z^{-n}X) = 0$, this implies that $\mathcal{K}_0 = 0$, and there is nothing to prove. Likewise, define $\mathcal{R}_n = Z^n\mathcal{K}_0 \ominus Z^{n+1}\mathcal{K}_0$. Then $\mathcal{R}_n = Z^n\mathcal{R}_0$, and $\mathcal{K}_0 = \mathcal{R}_0 \oplus \mathcal{R}_1 \oplus \mathcal{R}_2 \oplus \cdots$.

Suppose s-dim $\mathcal{R}_0 = M$, and define the sequence of Hilbert spaces $\mathcal{M}$ to have entries $\mathcal{M}_k = \mathbb{C}^{M_k}$ ($\mathcal{M}_k = \ell_2$ if $M_k = \infty$).[3] Then there exist isometries $V_k : \mathcal{M}_k \to (\mathcal{R}_0)_k$ such that $(\mathcal{R}_0)_k = \mathcal{M}_k V_k$. Let V be the operator whose k-th block-rows is equal to V_k. Stacking the V_k into one operator V, we obtain an orthonormal basis representation of $\mathcal{R}_0$, as in chapter 2, such that

$$\mathcal{R}_0 = \mathcal{D}_2^{\mathcal{M}} V, \qquad \mathbf{P}_0(VV^*) = I.$$

Then $\mathcal{R}_n = \mathcal{D}_2 Z^n V$. Because $\mathcal{R}_i \perp \mathcal{R}_j$ ($i \neq j$), it follows that $D_1 Z^n V \perp D_2 V$ ($n \geq 1$) for all $D_{1,2} \in \mathcal{D}_2$, *i.e.*,

$$\begin{aligned} \mathbf{P}_0(Z^n VV^*) &= 0 \\ \mathbf{P}_0(VV^* Z^{-n}) &= 0 \end{aligned}$$

so that $VV^* = I$: V is an isometry. The orthogonal collection $\{\mathcal{D}_2 Z^n V\} \in \mathcal{K}_0$ ($n \geq 0$), and together spans the space $\mathcal{U}_2 V$. Hence $\mathcal{K}_0 = \{\mathcal{D}_2 Z^n V\} = \mathcal{U}_2 V$.

If V_1 is another orthonormal basis representation of $\mathcal{R}_0$, then $V = V_1 R$, with $R \in \mathcal{D}(\mathcal{N}, \mathcal{N})$ and R unitary. This is proven by

$$\begin{aligned} \mathcal{U}_2 V_1 = \mathcal{U}_2 V \quad &\Rightarrow \quad \begin{cases} \mathcal{U}_2 V_1 V^* = \mathcal{U}_2 \\ \mathcal{U}_2 V V_1^* = \mathcal{U}_2 \end{cases} \\ &\Rightarrow \quad \begin{cases} V_1 V^* \in \mathcal{U} \\ V V_1^* \in \mathcal{U} \end{cases} \\ &\Rightarrow \quad V_1 V^* \in \mathcal{D}. \end{aligned}$$

□

[3] Let N be the index sequence corresponding to $\mathcal{N}$, with entries N_i. It follows that the dimension sequence M has entries $M_i < N_i + N_{i+1} + \cdots$. Although M_i can be infinite, an orthonormal basis for $(\mathcal{R}_0)_i = \pi_i \mathcal{R}_0$ is still *countable*, and the construction of an orthonormal basis representation of $\mathcal{R}_0$ can be done as explained in the proof of the theorem.

The above proof is in the style of the proof given in Helson [1, §VI.3] for the time-invariant Hardy space setting. This proof was in turn based on Beurling's work [10] for the scalar (SISO) case and Lax [11] for the extension to vector valued functions.

A remaining issue is to give conditions under which V is actually unitary. For time-invariant systems, this condition is that $\mathcal{K}_0$ is "full range" [1]. Systems T for which $\mathcal{K}_0(T)$ is full range were called "roomy" in [6]. Systems of finite degree are roomy: if $\mathcal{H}_0(T)$ is finite dimensional, then its complement $\mathcal{K}_0(T)$ is automatically full range. For time-varying systems, only less definite results can be obtained.

If V is inner, then $\mathcal{K}_0(V) = \mathcal{U}_2 V = \mathcal{K}_0$. If V is an isometry but is not inner, then

$$\mathcal{U}_2 = \overline{\mathcal{H}}_0(V) \oplus \mathcal{K}_0' \oplus \mathcal{K}_0,$$

where $\mathcal{K}_0(V) = \mathcal{K}_0 \oplus \mathcal{K}_0'$, $\mathcal{K}_0 = \mathcal{U}_2 V$ and $\mathcal{K}_0' = \{X \in \mathcal{U}_2 : XV^* = 0\}$ (proposition 4.4). Let us call $\mathcal{K}_0$ a "full range subspace" if V is inner, *i.e.*, if $\mathcal{K}_0' = 0$ (proposition 4.4). We want to determine conditions on $\mathcal{K}_0$ for which this happens. To this end, $\mathcal{K}_0'$ is constructed from $\mathcal{K}_0$ without use of V.

In this section, let $\mathcal{K}_n = \mathbf{P}(Z^{-n}\mathcal{K}_0)$. Define $\mathcal{H}_0 = \mathcal{U}_2 \ominus \mathcal{K}_0$, and, for $n > 0$, $\mathcal{H}_n = \mathbf{P}(Z^{-n}\mathcal{H}_0)$. We show that $\mathcal{K}_0$ is a "full range subspace" if $\bigcup_0^\infty \mathcal{K}_n = \mathcal{U}_2$.

LEMMA 4.12. *With the above definitions, for $n \geq 0$,*

$$\begin{aligned} \mathcal{K}_n &\subset \mathcal{K}_{n+1} \\ \mathcal{K}_0' &= \mathcal{U}_2 \ominus \textstyle\bigcup_0^\infty \mathcal{K}_n. \end{aligned}$$

PROOF Because $Z\mathcal{K}_0 \subset \mathcal{K}_0$, it follows that $\mathcal{K}_0 = \mathbf{P}(Z^{-1}Z\mathcal{K}_0) \subset \mathbf{P}(Z^{-1}\mathcal{K}_0) = \mathcal{K}_1$. Repeating the argument gives $\mathcal{K}_n \subset \mathcal{K}_{n+1}$. Let $X \in \mathcal{U}_2$. Then, because $\mathcal{K}_0 = \mathcal{U}_2 V$,

$$\begin{aligned} X \in \mathcal{K}_0' &\Leftrightarrow XV^* = 0 \\ &\Leftrightarrow \mathbf{P}_0(XV^*Z^n) = 0 \qquad (\text{all } n \in \mathbf{Z}) \\ &\Leftrightarrow X \perp \mathbf{P}(Z^{-n}\mathcal{K}_0) \qquad (\text{all } n \geq 0) \\ &\Leftrightarrow X \perp \textstyle\bigcup_0^\infty \mathbf{P}(Z^{-n}\mathcal{K}_0) \quad (\text{all } n \geq 0). \end{aligned}$$

□

Hence $\mathcal{K}_0$ is full range if and only if $\bigcup_0^\infty \mathcal{K}_n = \mathcal{U}_2$. This property can also be given in terms of $\mathcal{H}_n$:

LEMMA 4.13. *With the above definitions, for $n \geq 0$,*

$$\begin{aligned} \mathcal{H}_n &= \mathcal{U}_2 \ominus \mathcal{K}_n, \\ \mathcal{H}_{n+1} &\subset \mathcal{H}_n, \\ \mathcal{K}_0' &= \textstyle\bigcap_0^\infty \mathcal{H}_n. \end{aligned}$$

PROOF

$$\begin{aligned} X \in \mathcal{U}_2 \ominus \mathcal{K}_n &\Leftrightarrow X \in \mathcal{U}_2, \quad X \perp \mathcal{K}_n \\ &\Leftrightarrow X \in \mathcal{U}_2, \quad Z^n X \perp \mathcal{K}_0 \\ &\Leftrightarrow X \in \mathcal{U}_2, \quad Z^n X \in \mathcal{H}_0 \\ &\Leftrightarrow X \in \mathcal{U}_2, \quad X \in Z^{-n}\mathcal{H}_0 \\ &\Leftrightarrow X \in \mathbf{P}(Z^{-n}\mathcal{H}_0) \quad = \quad \mathcal{H}_n. \end{aligned}$$

Hence $\mathcal{H}_n = \mathcal{U}_2 \ominus \mathcal{K}_n$. The remaining issues are a corollary of lemma 4.12. □

LEMMA 4.14. *$\mathcal{K}_0'$ is a doubly shift-invariant subspace: $Z\mathcal{K}_0' \subset \mathcal{K}_0'$, $\mathbf{P}(Z^{-1}\mathcal{K}_0') = \mathcal{K}_0'$.*

PROOF $Z\mathcal{K}_0' \subset \mathcal{K}_0'$, because

$$\begin{aligned} Z\mathcal{K}_0' &= \{ZX : X \in \mathcal{K}_0'\} \\ &= \{ZX : X \in \mathcal{U}_2 \wedge XV^* = 0\} \\ &= \{ZX : X \in \mathcal{U}_2 \wedge ZXV^* = 0\} \\ &= \{Y \in Z\mathcal{U}_2 : YV^* = 0\} \\ &\subset \mathcal{K}_0'. \end{aligned}$$

On the other hand, $\mathbf{P}(Z^{-1}\mathcal{K}_0') \subset \mathcal{K}_0'$, because $\mathbf{P}(Z^{-1}\mathcal{H}_n) = \mathcal{H}_{n+1} \subset \mathcal{H}_n$, and $\mathcal{H}_n = \bigcap_{k=0}^n \mathcal{H}_n$. Hence $\mathbf{P}(Z^{-1} \bigcap_{k=0}^n \mathcal{H}_n) = \mathcal{H}_{n+1} \subset \bigcap_{k=0}^n \mathcal{H}_n$. Letting $n \to \infty$ yields $\mathbf{P}(Z^{-1}\mathcal{K}_0') \subset \mathcal{K}_0'$. □

There are connections between the fact that a subspace is doubly shift-invariant and the occurrence of $\ell_A = 1$ in realizations which have this subspace in their observability space $\mathcal{D}_2\mathbf{F}_0$. For example, if $\mathcal{K}_0 = \mathcal{U}_2 V$ where V is not inner but an isometry, and if V has a realization which is unitary and uniformly observable, then the realization cannot be uniformly controllable (for else V is inner), so that we must have $\ell_A = 1$ (theorem 4.6). In this case, $\mathcal{K}_0(V) = \mathcal{U}_2 V \oplus \mathcal{K}_0'$, and $\overline{\mathcal{H}}_0(V) = \mathcal{D}_2\mathbf{F}_0 \ominus \mathcal{K}_0'$.

The above lemmas are summarized in the following theorem.

THEOREM 4.15. (BEURLINGLAX,II) *In theorem 4.11, V is inner if and only if $\bigcap_0^\infty \mathcal{H}_n = 0$, where $\mathcal{H}_0 = \mathcal{U}_2 \ominus \mathcal{K}_0$ and $\mathcal{H}_n = \mathbf{P}(Z^{-n}\mathcal{H}_0)$ $(n \geq 0)$.*

The fact that $\mathcal{K}_0'$ is shift-invariant (lemma 4.14) ensures, according to theorem 4.11, the existence of an isometry U such that $\mathcal{K}_0' = \mathcal{U}_2 U$:

COROLLARY 4.16. *If $V \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ is an isometry, then there exists an isometry $U \in \mathcal{U}(\mathcal{M}_U, \mathcal{N})$ such that* $\ker(\cdot\, V^*|_{\mathcal{U}_2^{\mathcal{N}}}) = \mathcal{U}_2^{\mathcal{M}_U} U$. *The operator*

$$W = \begin{bmatrix} U \\ V \end{bmatrix}$$

is inner, with $\mathcal{H}_0(W) = \overline{\mathcal{H}}_0(V)$.

PROOF If V is an isometry, then (proposition 4.4)

$$\mathcal{U}_2^{\mathcal{N}} = \overline{\mathcal{H}}_0(V) \oplus \ker(\cdot\, V^*|_{\mathcal{U}_2^{\mathcal{N}}}) \oplus \mathcal{U}_2^{\mathcal{M}} V, \tag{4.15}$$

where $\mathcal{K}_0' := \ker(\cdot\, V^*|_{\mathcal{U}_2})$ is shift-invariant, so that according to theorem 4.11 there exists an isometry $U \in \mathcal{U}(\mathcal{M}_U, \mathcal{N})$ such that $\mathcal{K}_0' = \mathcal{U}_2^{\mathcal{M}_U} U$. In view of proposition 4.4, W is inner if $WW^* = I$ and $\ker(\cdot\, W^*|_{\mathcal{U}_2}) = 0$. $WW^* = I$ requires $UV^* = 0$, which is true because $\mathcal{U}_2 V \perp \mathcal{U}_2 U$. Hence $\mathcal{U}_2 W = \mathcal{U}_2 U \oplus \mathcal{U}_2 V$, and since $\overline{\mathcal{H}}_0(W) \supset \overline{\mathcal{H}}_0(V)$, we must have (from equation (4.15)) that $\overline{\mathcal{H}}_0(W) = \overline{\mathcal{H}}_0(V)$ and $\ker(\cdot\, W^*|_{\mathcal{U}_2}) = 0$. Hence W is inner, and $\mathcal{H}_0(W)$ is closed. □

Inner-outer factorizations

An operator $T_0 \in \mathcal{U}$ is said to be (left) *outer* if

$$\overline{\mathcal{U}_2 T_0} = \mathcal{U}_2 . \tag{4.16}$$

Other definitions are possible;[4] the above definition is such that $\overline{\text{ran}}(\cdot\, T_0) = \overline{\mathcal{X}_2 T_0} = \mathcal{X}_2$, so that $\ker(\cdot\, T_0^*) = 0$ and T_0 has an algebraic left inverse (which can be unbounded if $\mathcal{X}_2 T_0$ is not closed).

A factorization of an operator T into $T = T_0 V$, where T_0 is outer and V is inner (or an isometry) is called an outer-inner factorization. This factorization can be obtained from theorem 4.11 by taking a different definition of $\mathcal{K}_0$ than was the case in the external factorization (where we took $\mathcal{K}_0$ equal to the output null space $\mathcal{K}_0(T)$). Note that the closure in (4.16) is necessary: for example, the system $T = I - Z$ has inner factor $V = I$ and of necessity an outer factor $T_0 = I - Z$. T_0 is not boundedly invertible, and $\mathcal{U}_2 T_0$ is only dense in $\mathcal{U}_2$. This happens when $\mathcal{U}_2 T$ is not a closed subspace. The time-invariant equivalent of this example is $T(z) = 1 - z$, which has a zero on the unit disc. Again, $V(z) = 1$, and $T_0(z) = T(z)$ is not boundedly invertible. Also note that if T is not an invertible operator, then it is not possible to obtain an inner factor: only an isometric operator can be obtained, since we have chosen T_0 to be invertible.

[4] See *e.g.*, Arveson [2], who, translated to our notation, requires that $\mathcal{U}_2 T_0$ is dense in $\mathbf{P}(\mathcal{X}_2 T_0)$ and that the projection operator onto the range of T_0 is diagonal.

THEOREM 4.17. (INNER-OUTER FACTORIZATION) *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$. Then T has a factorization*

$$T = T_0 V$$

where $V \in \mathcal{U}(\mathcal{M}_V, \mathcal{N})$ is an isometry ($VV^* = I$), $T_0 \in \mathcal{U}(\mathcal{M}, \mathcal{M}_V)$ is outer, and $\mathcal{M}_V \subset \mathcal{M}$ (entrywise). V is inner if and only if $\ker(\cdot T^*) = 0$.

PROOF Define $\mathcal{K}_0 = \overline{\mathcal{U}_2 T}$. Then $\mathcal{K}_0$ is a D-invariant subspace which is shift-invariant: $Z\mathcal{K}_0 \subset \mathcal{K}_0$. According to theorem 4.11, there is a space sequence $\mathcal{M}_V$ and an isometric operator $V \in \mathcal{U}(\mathcal{M}_V, \mathcal{N})$ such that $\overline{\mathcal{U}_2^{\mathcal{M}} T} = \mathcal{U}_2^{\mathcal{M}_V} V$. By construction, $\overline{\mathcal{U}_2^{\mathcal{M}} T} = \mathcal{D}_2^{\mathcal{M}_V} V \oplus \overline{Z\mathcal{U}_2^{\mathcal{M}} T}$ with $\mathcal{M}_V$ of minimal dimensions. Because also $\overline{\mathcal{U}_2^{\mathcal{M}} T} = \overline{[\mathcal{D}_2^{\mathcal{M}} \oplus Z\mathcal{U}_2^{\mathcal{M}}]T}$, but $\mathcal{D}_2^{\mathcal{M}} T$ is not necessarily orthogonal to $Z\mathcal{U}^{\mathcal{M}} T$, it follows that $\mathcal{M}_V \subset \mathcal{M}$. In particular, the entries of $\mathcal{M}_V$ are finite vector spaces.

Define $T_0 = TV^*$. Then $\overline{\mathcal{U}_2 T_0} = \overline{\mathcal{U}_2 TV^*} = \overline{\overline{\mathcal{U}_2 T} V^*} = \overline{\mathcal{U}_2 VV^*} = \mathcal{U}_2$, so that T_0 is outer. It remains to prove that $T = T_0 V$, *i.e.*, $T = TV^* V$. This is immediate if V is inner. If V is not inner, then corollary 4.16 ensures the existence of an isometry U such that

$$\mathcal{U}_2 = \overline{\mathcal{H}}_0(V) \oplus \mathcal{U}_2 U \oplus \mathcal{U}_2 V,$$

where $\mathcal{K}_0' := \mathcal{U}_2 U = \ker(\cdot V^*|_{\mathcal{U}_2})$, and $W = \begin{bmatrix} U \\ V \end{bmatrix}$ is inner and such that $\mathcal{H}_0(W) = \overline{\mathcal{H}}_0(V)$. Then $U^* U + V^* V = I$, $VU^* = 0$, and

$$\begin{aligned} T = TV^* V \quad &\Leftrightarrow \quad T(I - V^* V) = 0 \\ &\Leftrightarrow \quad TU^* U = 0. \end{aligned}$$

But $\mathcal{U}_2 TU^* \subset \mathcal{U}_2 VU^* = 0$, which implies $TU^* = 0$. Hence $T = T_0 V$.

In terms of the definitions of the previous section, we have $\mathcal{K}_0 = \overline{\mathcal{U}_2 T}$, and

$$\begin{aligned} \mathcal{H}_0 &= \mathcal{K}_0 \ominus \overline{\mathcal{U}_2 T} = \{X \in \mathcal{U}_2 : \mathbf{P}(XT^*) = 0\} \\ \mathcal{H}_n &= \mathbf{P}(Z^{-n}\mathcal{H}_0) \\ &= \{\mathbf{P}(Z^{-n}X) : X \in \mathcal{U}_2 \wedge \mathbf{P}(XT^*) = 0\} \\ &= \{\mathbf{P}(Z^{-n}X) : X \in \mathcal{U}_2 \wedge \mathbf{P}\left(Z^n (Z^{-n}X) T^*\right) = 0\} \\ &= \{\mathbf{P}(Z^{-n}X) : X \in \mathcal{U}_2 \wedge \mathbf{P}\left(Z^n \mathbf{P}(Z^{-n}X) T^*\right) = 0\} \\ &= \{X \in \mathcal{U}_2 : \mathbf{P}(Z^n XT^*) = 0\}. \end{aligned}$$

Hence $\mathcal{K}_0' = \cap \mathcal{H}_n = \{X \in \mathcal{U}_2 : XT^* = 0\}$. $\mathcal{K}_0'$ is empty if and only if $\ker(\cdot T^*|_{\mathcal{U}_2}) = 0$, that is, if and only if $\ker(\cdot T^*) = 0$. The rest follows from theorem 4.15. □

A more general result was obtained by Arveson [2], who showed the existence of the inner-outer factorization in the general context of nest algebras which also applies to our model of time-varying systems.

If $\overline{\text{ran}}\, T$ is not all of $\mathcal{N}_2$, then V is not full range either: V is not inner. One can in this case also define a factorization based on the extension of V to an inner operator W, where W is defined as in the above proof. Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$, $V \in \mathcal{U}(\mathcal{M}_V, \mathcal{N})$, and $W \in \mathcal{U}(\mathcal{M}_W, \mathcal{N})$. Define $T_0 = TW^* \in \mathcal{X}(\mathcal{M}, \mathcal{M}_W)$. Then $T = T_0 W$, and

$$\begin{aligned} \overline{\mathcal{U}_2^{\mathcal{M}} T_0} &= \overline{\mathcal{U}_2 T} W^* = \mathcal{U}_2 V W^* = \mathcal{U}_2 [0 \quad I] \\ &= \mathcal{U}_2^{\mathcal{M}_V} \subset \mathcal{U}_2^{\mathcal{M}_W}, \end{aligned}$$

so that T_0 is upper but not precisely outer:[5] it reaches only a subset of $\mathcal{U}_2^{\mathcal{M}_W}$. This is the best we can hope for, in view of the fact that T is not full range.

The inner-outer factorization is based on the identification of a subspace $\mathcal{K}_0 = \overline{\mathcal{U}_2 T}$ as $\mathcal{K}_0 = \mathcal{U}_2 V$. The complement in $\mathcal{U}_2$ of this space is $\overline{\mathcal{H}}_0(V) \oplus \mathcal{K}_0'$ and is characterized by the elements $X \in \mathcal{U}_2$ satisfying $\mathbf{P}_0(\mathcal{U}_2 T X^*) = 0$, that is, $XT^* \perp \mathcal{U}_2$. Hence

$$\overline{\mathcal{H}}_0(V) \oplus \mathcal{K}_0' = \{X \in \mathcal{U}_2 : XT^* \in \mathcal{L}_2 Z^{-1}\} = \{X \in \mathcal{U}_2 : \mathbf{P}(XT^*) = 0\}. \tag{4.17}$$

In this expression, $\mathcal{K}_0' = \mathcal{U}_2 U = \ker(\cdot\, V^*|_{\mathcal{U}_2})$ according to its definition. We now show that also $\mathcal{K}_0' = \ker(\cdot\, T^*|_{\mathcal{U}_2}) = \{X \in \mathcal{U}_2 : XT^* = 0\}$. Indeed, if $X \in \mathcal{K}_0'$, then $X = X_1 U$ for some $X_1 \in \mathcal{U}_2$, and because $UT^* = 0$, it follows that $XT^* = 0$. Conversely, if $XT^* = 0$, then $XV^* T_0^* = 0$, and because $\ker(\cdot\, T_0^*) = 0$, it follows that $XV^* = 0$ so that $X \in \mathcal{K}_0'$. Hence $\mathcal{K}_0' = \ker(\cdot\, T^*|_{\mathcal{U}_2})$.

In equation (4.17), $\overline{\mathcal{H}}_0(V) \oplus \mathcal{K}_0'$ is the subspace in $\mathcal{U}_2$ on which $\cdot\, T^* = \cdot\, H_T^*$. Note that if T is itself inner, then $\mathcal{K}_0' = 0$ and $\cdot\, T^* = \cdot\, H_T^*$ on all of $\mathcal{H}_0(T)$, so in this case the result is $V = T$ and $T_0 = I$, save for unitary diagonal factors.

Computation of the inner-outer factorization $T = VT_0$

In this section, we work with a dual factorization of T: $T = VT_0$ (for different V and T_0), where T_0 is 'right outer': $\overline{\mathcal{L}_2 Z^{-1} T_0^*} = \mathcal{L}_2 Z^{-1}$ (or $\overline{T_0 \mathcal{U}_2} = \mathcal{U}_2$), and where the left inner (isometric) factor V satisfies $V^* V = I$ and is obtained by identifying the subspace $\mathcal{K}(V) = \mathcal{L}_2 Z^{-1} V^*$ with $\overline{\mathcal{L}_2 Z^{-1} T^*}$. For this factorization,

$$\overline{\mathcal{H}}(V) \oplus \mathcal{K}' = \{U \in \mathcal{L}_2 Z^{-1} : UT \in \mathcal{U}_2\}, \qquad \mathcal{K}' = \ker(\cdot\, T|_{\mathcal{L}_2 Z^{-1}}).$$

We have defined in chapter 3 the decomposition of T, restricted to $\mathcal{L}_2 Z^{-1}$, as

$$\cdot\, T|_{\mathcal{L}_2 Z^{-1}} = \cdot\, K_T + \cdot\, H_T, \qquad \cdot\, K_T = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\cdot\, T).$$

It is thus seen that $\overline{\mathcal{H}}(V)$ is the *largest* subspace in $\mathcal{L}_2 Z^{-1}$ for which $\overline{\mathcal{H}}(V) K_T = 0$ and which is orthogonal to $\mathcal{K}'$. This property provides a way to compute the inner-outer factorization.

[5] T_0 is outer according to Arveson's definition [2].

Let $\mathbf{Q}$ be an orthonormal basis representation of $\overline{\mathcal{H}}(V)$: $\overline{\mathcal{H}}(V) = \mathcal{D}_2\mathbf{Q}$, and let $\mathbf{F}_0$ be a basis representation of $\overline{\mathcal{H}}_0(T)$, or more generally, a subspace in $\mathcal{U}_2$ containing $\overline{\mathcal{H}}_0(T)$. The fact that $\overline{\mathcal{H}}(V)K_T = 0$ translates to the condition $\mathbf{Q}T \in \mathcal{U}$. Because $\overline{\mathcal{H}}(V)T \subset \overline{\mathcal{H}}_0(T)$, we must have that $\mathbf{Q}T = Y\mathbf{F}_0$ for some bounded diagonal operator Y, which plays an instrumental role in the derivation of a state realization for V. It remains to implement the condition $\overline{\mathcal{H}}(V) \perp \mathcal{K}'$. Suppose that $\mathbf{Q}$ has a component in $\mathcal{K}'$, so that $D\mathbf{Q} \in \mathcal{K}'$, for some $D \in \mathcal{D}_2$. Then, since $\mathcal{K}' = \ker(\cdot\, T|_{\mathcal{L}_2 Z^{-1}})$,

$$D\mathbf{Q} \in \mathcal{K}' \quad \Leftrightarrow \quad D\mathbf{Q}T = DY\mathbf{F}_0 = 0 \quad \Leftrightarrow \quad D \in \ker(\cdot\, Y)$$

Hence $\overline{\mathcal{H}}(V) = \mathcal{D}_2\mathbf{Q}$ can be described as the largest subspace $\mathcal{D}_2\mathbf{Q}$ for which $\mathbf{Q}T = Y\mathbf{F}_0$ with $\ker(\cdot\, Y) = 0$.

If $\mathcal{B}$ is the state space sequence of T, and $\mathcal{B}_V$ is the state space sequence of V, then $Y \in \mathcal{D}(\mathcal{B}_V, \mathcal{B})$. The condition $\ker(\cdot\, Y) = 0$ implies that $\mathcal{B}_V \subset \mathcal{B}$ (pointwise), so that the state dimension of V is at each point in time less than or equal to the state dimension of T at that point.

PROPOSITION 4.18. *Let $T \in \mathcal{U}$ be a locally finite transfer operator, let $\mathbf{T} = \{A, B, C, D\}$ be an observable realization of T, and assume $\ell_A < 1$. Let V be a left inner (isometric) factor of T so that $T_0 = V^*T$ is right outer. Then the pair (A_V, B_V) that corresponds to an orthonormal basis representation $\mathbf{Q}$ of $\overline{\mathcal{H}}(V)$ satisfies*

$$\begin{array}{lrcl} (i) & A_V^* Y A + B_V^* B & = & Y^{(-1)} \\ (ii) & A_V^* Y C + B_V^* D & = & 0 \\ (iii) & A_V^* A_V + B_V^* B_V & = & I \\ (iv) & \ker(\cdot\, Y) & = & 0. \end{array}$$

and conversely, all solutions (A_V, B_V) of these equations give basis representations of $\overline{\mathcal{H}}(V)$.

PROOF Let $\mathbf{F}_0 = (I - AZ)^{-1}C$. We use in this proof the relations

$$\begin{array}{rcl} T & = & D + BZ\mathbf{F}_0 \\ \mathbf{F}_0 & = & C + AZ\mathbf{F}_0 \\ Z\mathbf{Q} & = & A_V^*\mathbf{Q} + B_V^*. \end{array}$$

(*cf.* lemmas 3.19, 3.26). Observability implies that $\mathcal{H}_0(T) \subset \mathcal{D}_2\mathbf{F}_0$, and hence $\mathbf{P}(\mathbf{Q}T) = Y\mathbf{F}_0$ for some bounded $Y \in \mathcal{D}$, and we will show that Y is given by a solution to equation (*i*). Indeed, let Y be defined by $\mathbf{P}(\mathbf{Q}T) = Y\mathbf{F}_0$. Then $\mathbf{P}(Z^{-1}Y\mathbf{F}_0) = Y^{(1)}\mathbf{P}(Z^{-1}\mathbf{F}_0) = (YA)^{(1)}\mathbf{F}_0$. On the other hand,

$$\begin{array}{rcl} A_V^{*(1)}\mathbf{P}(Z^{-1}\mathbf{Q}T) & = & \mathbf{P}(Z^{-1}[A_V^*\mathbf{Q}]T) \\ & = & \mathbf{P}(Z^{-1}[Z\mathbf{Q} - B_V^*]T) \\ & = & \mathbf{P}(\mathbf{Q}T) - B_V^{*(1)}B^{(1)}\mathbf{F}_0 \\ & = & Y\mathbf{F}_0 - B_V^{*(1)}B^{(1)}\mathbf{F}_0. \end{array}$$

Hence, because observability means that $\cdot \mathbf{F}_0\big|_{\mathcal{D}_2}$ is one-to-one (definition 3.5),

$$\begin{aligned} & \mathbf{P}(Z^{-1}Y\mathbf{F}_0) = \mathbf{P}(Z^{-1}\mathbf{Q}T) \\ \Leftrightarrow \quad & (A_V^*YA)^{(1)}\mathbf{F}_0 + (B_V^*B)^{(1)}\mathbf{F}_0 = Y\mathbf{F}_0 \\ \Leftrightarrow \quad & A_V^*YA + B_V^*B = Y^{(-1)}. \end{aligned}$$

Conversely, since $\ell_A < 1$ implies that any solution Y of (i) must be unique, it follows that this solution will satisfy $\mathbf{P}(\mathbf{Q}T) = Y\mathbf{F}_0$.

Let Y be given by $\mathbf{P}(\mathbf{Q}T) = Y\mathbf{F}_0$. To derive the equivalence of (ii) with the condition $\mathbf{Q}T = \mathcal{U}$, we use the fact that $\mathbf{Q}T \in \mathcal{U} \Leftrightarrow \mathbf{P}_0(Z^n\mathbf{Q}T) = 0$ for all $n > 0$.

$$\begin{aligned} n = 1: \qquad \mathbf{P}_0(Z\mathbf{Q}T) &= \mathbf{P}_0([A_V^*\mathbf{Q} + B_V^*]T) \\ &= A_V^*\mathbf{P}_0(\mathbf{Q}T) + B_V^*D \\ &= A_V^*YC + B_V^*D \end{aligned}$$

Hence $\mathbf{P}_0(Z\mathbf{Q}T) = 0 \Leftrightarrow A_V^*YC + B_V^*D = 0$. For $n > 1$, assume $\mathbf{P}_0(Z^{n-1}\mathbf{Q}T) = 0$. Then

$$\begin{aligned} \mathbf{P}_0(Z^n\mathbf{Q}T) &= \mathbf{P}_0(Z^{n-1}[Z\mathbf{Q}T]) \\ &= \mathbf{P}_0(Z^{n-1}[A_V^*\mathbf{Q}]T) + \mathbf{P}_0(Z^{n-1}B_V^*T) \\ &= A_V^{*(n-1)}\mathbf{P}_0(Z^{n-1}\mathbf{Q}T) + B_V^{*(n-1)}\mathbf{P}_0(Z^{n-1}T) \\ &= 0 + 0. \end{aligned}$$

Hence (ii) is both necessary and sufficient for the condition $\mathbf{Q}T \in \mathcal{U}$ to be satisfied. The fact that we took $\mathbf{Q}$ to be an orthonormal basis representation implies condition (iii), and condition (iv) has already been derived. □

It is possible to construct solutions (A_V, B_V) for the four equations in proposition 4.18, and from these solutions a realization $\mathbf{V}$ for the inner (isometric) factor V of T follows. Taking the k-th entry of each diagonal in (i)–(iv) gives the recursive equations

$$\begin{cases} (i) & A_{V,k}^* Y_k A_k + B_{V,k}^* B_k = Y_{k+1} \\ (ii) & A_{V,k}^* Y_k C_k + B_{V,k}^* D_k = 0 \\ (iii) & A_{V,k}^* A_{V,k} + B_{V,k}^* B_{V,k} = I \\ (iv) & Y_{k+1} \text{ full row-rank.} \end{cases}$$

A_V and B_V can be computed from these equations starting at some point in time, once an initial value for Y is known (this is discussed below). The recursion for Y_{k+1} is convergent

because $\ell_A < 1$. At each point in time, the computation requires four steps:

$$
\begin{aligned}
&(a) \quad \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix} = \begin{bmatrix} Y_k C_k \\ D_k \end{bmatrix}^{\perp} && \text{[for } (ii)\text{]} \\
&(b) \quad Y'_{k+1} = [A'^{*}_{V,k} \quad B'^{*}_{V,k}] \begin{bmatrix} Y_k A_k \\ B_k \end{bmatrix} && \text{[for } (i)\text{]} \\
&(c) \quad \begin{bmatrix} Y_{k+1} \\ 0 \end{bmatrix} = \begin{bmatrix} Q_{1,k} \\ Q_{2,k} \end{bmatrix} Y'_{k+1} && \text{[QR-factorization of } Y'_{k+1} \text{ for } (iv)\text{]} \\
&(d) \quad \begin{bmatrix} A_{V,k} \\ B_{V,k} \end{bmatrix} = \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix} Q^{*}_{1,k},
\end{aligned}
$$

where $[\,\cdot\,]^{\perp}$ denotes the linear algebra operation of taking a minimal orthonormal basis of the full orthogonal complement of the column space of its argument (the basis vectors form the columns of the result). Steps (a) and (b) determine Y'_{k+1}, which can be too large: its kernel is not necessarily empty. In step (c), a unitary matrix $Q = \left[\begin{smallmatrix} Q_{1,k} \\ Q_{2,k} \end{smallmatrix}\right]$ is computed such that the kernel of Y_{k+1} is determined as the span of the rows of $Q_{2,k}$, and subsequently removed, which yields Y_{k+1} and $A_{V,k}, B_{V,k}$.

With A_V and B_V known, we can proceed in two directions. It was noted in the previous section that it will not always be possible to obtain an inner factor V: if $\ker(\,\cdot\,T|_{\mathcal{L}_2 Z^{-1}}) \neq 0$, then V will be isometric. V can be extended to an inner operator $W = [U \quad V]$, where U is the isometry satisfying $\mathcal{L}_2 Z^{-1} U^{*} = \ker(\,\cdot\,T|_{\mathcal{L}_2 Z^{-1}})$. The resulting W is too large in the sense that $U^{*}T = 0$, but since $\overline{\mathcal{H}}(W) = \overline{\mathcal{H}}(V)$, a realization $\mathbf{W}$ is readily obtained from A_V, B_V:

$$
\mathbf{W} = \begin{bmatrix} A_V & C_W \\ B_V & D_W \end{bmatrix}
$$

where C_W and D_W are obtained by the condition that $\mathbf{W}$ is unitary: $\mathbf{W}\mathbf{W}^{*} = I$, $\mathbf{W}^{*}\mathbf{W} = I$ (*cf.* theorem 4.6). This leads to the condition

$$
(v) \qquad \begin{bmatrix} C_{W,k} \\ D_{W,k} \end{bmatrix} = \begin{bmatrix} A_{V,k} \\ B_{V,k} \end{bmatrix}^{\perp}
$$

A realization for U is obtained from the condition $U^{*}T = 0$, where $U^{*}T$ evaluates as

$$
\begin{aligned}
U^{*}T &= [D_U^{*} + C_U^{*}\mathbf{Q}]T \\
&= D_U^{*}T + C_U^{*}\mathbf{Q}T \\
&= D_U^{*}[D + BZ\mathbf{F}_0] + C_U^{*}Y\mathbf{F}_0 \\
&= [D_U^{*}D + C_U^{*}YC] + [D_U^{*}B + C_U^{*}YA]Z\mathbf{F}_0
\end{aligned}
\tag{4.18}
$$

Hence

$$
U^{*}T = 0 \quad \Leftrightarrow \quad \begin{cases} C_U^{*}YA + D_U^{*}B = 0 \\ C_U^{*}YC + D_U^{*}D = 0 \end{cases}
$$

and in view of the above steps (*a*)–(*d*), it follows that

$$\begin{bmatrix} C_{U,k} \\ D_{U,k} \end{bmatrix} = \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix} Q^*_{2,k}.$$

Finally, C_V and D_V can be obtained as the complement of $\begin{bmatrix} A_V & C_U \\ B_V & D_U \end{bmatrix}$, *i.e.*,

$$(e) \quad \begin{bmatrix} C_{V,k} \\ D_{V,k} \end{bmatrix} = \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix}^{\perp}$$

With $\mathbf{V}$ known, a realization for the outer factor T_0 is obtained by evaluating $T_0 = V^* T$ in terms of state-space quantities. This yields, much as in equation (4.18)

$$T_0 = [C_V^* Y C + D_V^* D] + [C_V^* Y A + D_V^* B]\,\mathbf{F}_0$$

Hence a realization of T_0 is given by

$$\mathbf{T}_0 = \begin{bmatrix} A & C \\ C_V^* Y A + D_V^* B & C_V^* Y C + D_V^* D \end{bmatrix}. \tag{4.19}$$

An algorithm to compute $\mathbf{V}$ and $\mathbf{T}_0$ from a realization of T for finite $n \times n$ (block) matrices is given as algorithm 4.1. The body of the algorithm consists of the steps explained above. One issue that remains to be discussed concerns the initialization of Y. In the algorithm for finite matrices, we can take $Y_1 = [\,\cdot\,]$ because the input space $\mathcal{M}$ for T (and hence V) has empty dimensions before time instant 1, so that a minimal realization for V has zero states before time instant 1. For the more general class of systems which are time invariant before, say, point 1 in time, an initial value for Y is determined in the following way. Y_1 now has to satisfy an equation rather than a recursion:

$$Y_1 = Y_0 = A_{V,0}^* Y_0 A_0 + B_{V,0}^* B_0,$$

where, as before,

$$\begin{aligned} A_{V,0}^* A_{V,0} + B_{V,0}^* B_{V,0} &= I, \\ A_{V,0}^* Y_0 C_0 + B_{V,0}^* D_0 &= 0. \end{aligned}$$

We show that the solution of these equations is the same as the classical solution of the inner-outer factorization, and is determined by the zeros of the time-invariant part of T that are in the unit disc. For convenience of notation, define $y = Y_0$, $a = A_0$, $b = B_0$, $c = C_0$, $d = D_0$, $\alpha = A_{V,0}$, $\beta = B_{V,0}$. We also assume that d (and hence T) is invertible, and that its zeros are distinct. Then

$$\begin{array}{rcl} y &=& \alpha^* y a + \beta^* b \\ 0 &=& \alpha^* y c + \beta^* d \\ I &=& \alpha^* \alpha + \beta^* \beta \end{array} \quad \Leftrightarrow \quad \begin{array}{rcl} \beta^* &=& -\alpha^* y c d^{-1} \\ y &=& \alpha^* y (a - c d^{-1} b) \\ I &=& \alpha^* \alpha + \beta^* \beta. \end{array} \tag{4.20}$$

In: $\{\mathbf{T}_k\}$ (an observable realization of T)
Out: $\{\mathbf{V}_k\}, \{(\mathbf{T}_0)_k\}$ (realizations of the isometric and outer factors)

$Y_1 = [\,\cdot\,]$

for $k = 1, \cdots, n$

$$
\begin{aligned}
(a) \quad & \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix} = \begin{bmatrix} Y_k C_k \\ D_k \end{bmatrix}^{\perp} \\
(b) \quad & Y'_{k+1} = \begin{bmatrix} A'^{*}_{V,k} & B'^{*}_{V,k} \end{bmatrix} \begin{bmatrix} Y_k A_k \\ B_k \end{bmatrix} \\
(c) \quad & \begin{bmatrix} Y_{k+1} \\ 0 \end{bmatrix} = \begin{bmatrix} Q_{1,k} \\ Q_{2,k} \end{bmatrix} Y'_{k+1} \qquad \text{[QR-factorization of } Y'_{k+1}\text{]} \\
(d) \quad & \begin{bmatrix} A_{V,k} \\ B_{V,k} \end{bmatrix} = \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix} Q^{*}_{1,k} \\
(e) \quad & \begin{bmatrix} C_{V,k} \\ D_{V,k} \end{bmatrix} = \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix}^{\perp} \\
& \mathbf{V}_k = \begin{bmatrix} A_{V,k} & C_{V,k} \\ B_{V,k} & D_{V,k} \end{bmatrix} \\
& (\mathbf{T}_0)_k = \begin{bmatrix} A_k & C_k \\ C^{*}_{V,k} Y_k A_k + D^{*}_{V,k} B_k & C^{*}_{V,k} Y_k C_k + D^{*}_{V,k} D_k \end{bmatrix}
\end{aligned}
$$

end

Algorithm 4.1. Inner-outer factorization algorithm for $T = VT_0$ (finite matrix case).

Bring in eigenvalue decompositions of α and $(a - cd^{-1}b)$:

$$\alpha = r\phi r^{-1}; \qquad a - cd^{-1}b = s\psi s^{-1}.$$

Then

$$(r^{*}ys) = \phi^{*}(r^{*}ys)\psi.$$

Because both ϕ and ψ are diagonal matrices, the above expression shows that $(r^{*}ys)$ must be a rectangular diagonal matrix (or a permutation thereof), and hence the diagonal entries of ϕ are equal to a subset of the diagonal entries of ψ^{-*}. In view of the requirement $\alpha^{*}\alpha = I - \beta^{*}\beta$, ϕ can contain only the entries of ψ^{-*} that are smaller than 1. Because V must be of the highest possible system order and y must have full row rank, ϕ is precisely equal to those entries.

It remains to note that the entries of $\psi^{-1} = \mathrm{eig}(a - cd^{-1}b)^{-1}$ are equal to the zeros of T. This

is because $T^{-1} = d^{-1} + d^{-1}bz\left[I-(a-cd^{-1}b)z\right]^{-1}cd^{-1}$ has poles equal to $\mathrm{eig}(a-cd^{-1}b)^{-1}$. With the poles of the inner system thus determined, it is a straightforward matter (involving a Lyapunov equation) to compute α, β, and y from (4.20). In particular, one can choose the non-zero values of $y' := r^*ys$ in any way. Then substitution in the equations (4.20) leads to the Lyapunov equation

$$\phi^* r^* r\phi + y'(s^{-1}cd^{-1})(s^{-1}cd^{-1})^* y'^* = r^* r.$$

This gives r, and then y follows as $y = r^{-*}y's^{-1}$.

Closed-form expression for the outer factor realization

In the time-invariant setting, it is well known that the outer factor T_0 of T can be written in closed form in terms of the original state matrices $\{A, B, C, D\}$ of T and only one unknown intermediate quantity, which is the solution of a Riccati equation with $\{A, B, C, D\}$ as parameters. One way to obtain the Riccati equation is by performing a spectral factorization of the squared relation $T^*T = T_0^*T_0$. Riccati equations can be solved recursively; efficient solution methods for the recursive version are the *square-root algorithms*, in which extra intermediate quantities are introduced to avoid the computation of inverses and square roots. The algorithm to compute the realization for T_0 given in (4.19) can be viewed as such a square-root algorithm: besides Y, it contains the intermediate quantities A_Y and B_Y. We show in this section how the corresponding Riccati recursion can be derived.

THEOREM 4.19. *Let $T \in \mathcal{U}$ be a locally finite transfer operator, let $\mathbf{T} = \{A, B, C, D\}$ be an observable realization of T, and assume $\ell_A < 1$. Then a realization of the outer factor T_0 of T so that $T_0 = V^*T$ is given by*

$$\mathbf{T}_0 = \begin{bmatrix} I & \\ & R^* \end{bmatrix} \begin{bmatrix} A & C \\ C^*MA + D^*B & C^*MC + D^*D \end{bmatrix}$$

where $M \geq 0$ is the solution of maximal rank of the recursive Riccati equation

$$M^{(-1)} = A^*MA + B^*B - \left[A^*MC + B^*D\right](D^*D + C^*MC)^\dagger \left[D^*B + C^*MA\right] \tag{4.21}$$

and R is a minimal full range factor ($\ker(\cdot R^*) = 0$) of

$$RR^* = (D^*D + C^*MC)^\dagger,$$

provided the pseudo-inverse is bounded (see proposition 4.20 below).

PROOF Let $\mathbf{T}_0$ be given by equation (4.19), so that C_Y and D_Y are given, according to steps (a) and (e), as

$$\begin{bmatrix} C_Y \\ D_Y \end{bmatrix} = \begin{bmatrix} YC \\ D \end{bmatrix}^{\perp\perp} = \begin{bmatrix} YC \\ D \end{bmatrix} R. \tag{4.22}$$

$R \in \mathcal{D}(\mathcal{N}, \mathcal{N}_V)$ is a diagonal whose 'tall' matrix entries R_k make the columns of $\begin{bmatrix} Y_k C_k \\ D_k \end{bmatrix}$ isometric, removing columns that are linearly dependent:

$$R^* (D^*D + C^*MC) R = I_{\mathcal{N}_V}, \qquad \text{where} \quad M := Y^*Y.$$

Let $X = D^*D + C^*MC$, then $R^*XR = I$ implies $RR^* = X^\dagger$, where $(\cdot)^\dagger$ denotes the operator pseudo-inverse [13]. According to step (c), $(Y^*Y)^{(-1)} = (Y'^*Y')^{(-1)}$, so that we obtain from step (b)

$$\begin{aligned}
(Y^*Y)^{(-1)} &= [A^*Y^* \quad B^*] \begin{bmatrix} A'_V \\ B'_V \end{bmatrix} [A'^*_V \quad B'^*_V] \begin{bmatrix} YA \\ B \end{bmatrix} \\
&= [A^*Y^* \quad B^*] \left(I - \begin{bmatrix} YC \\ D \end{bmatrix} RR^* [C^*Y^* \quad D^*] \right) \begin{bmatrix} YA \\ B \end{bmatrix} \\
&= A^*Y^*(I - YCRR^*C^*Y^*)YA + B^*(I - DRR^*D^*)B - \\
&\quad - A^*Y^*(YCRR^*D^*)B - B^*(DRR^*C^*Y^*)YA,
\end{aligned}$$

and with $M = Y^*Y$ this yields

$$M^{(-1)} = A^*MA + B^*B - \left[A^*MC + B^*D\right] RR^* \left[D^*B + C^*MA\right].$$

This equation has more solutions M. As $Y \in \mathcal{D}(\mathcal{B}_V, \mathcal{B})$ has $\mathcal{B}_V$ of maximal possible dimensions such that $\ker(\cdot\, Y) = 0$, the solution M of the Riccati equation must be positive and of maximal rank to yield an outer factor T_0. (Note that if D^*D is invertible, then $M = 0$ is always a solution, and yields $T_0 = T$.) □

The resulting Riccati equation bears a close resemblance to the Riccati equation that will be obtained later in the solution of the time-varying lossless embedding problem (chapter 7). The connection between the two problems is that both problems can be described as a spectral factorization problem. This connection is discussed in chapter 8.

By taking the k-th entry of each diagonal in equation (4.21), we obtain the recursion

$$\begin{aligned}
M_{k+1} &= A_k^* M_k A_k + B_k^* B_k - \\
&\quad - \left[A_k^* M_k C_k + B_k^* D_k\right] (D_k^* D_k + C_k^* M_k C_k)^\dagger \left[D_k^* B_k + C_k^* M_k A_k\right]
\end{aligned} \tag{4.23}$$

Initial conditions for the recursion can be obtained in special cases. For example, when T starts with zero states at some point k_0 in time, then $M_{k_0} = [\,\cdot\,]$. If T is time invariant before k_0, then M_{k_0} is given by a time-invariant Riccati equation. Again, the solution requires eigenvalue decompositions, and must satisfy the side conditions that $M_{k_0} \geq 0$ and has maximal rank. Riccati equations are studied in more detail in chapters 7 and 8.

In the above proof, we required the boundedness of the pseudo-inverse of $(D^*D + C^*MC)^\dagger$ in case this operator is not uniformly positive (this is no issue when D^*D is uniformly positive). For historical reasons, the issue of boundedness is investigated in closer detail

for a related Riccati equation which occurs in the solution of the embedding problem (chapter 7), where we discuss that pseudo-inverses can become unbounded if the range of their operand is not closed. We will show that if $\mathrm{ran}(\cdot T)$ is closed, then the pseudo-inverse is also bounded. This condition is a generalization of the time-invariant "no zeros of T are on the unit circle". If $\mathrm{ran}(\cdot T)$ is not closed, then $\mathcal{L}_2\mathcal{Z}^{-1}T_0$ is dense in $\mathcal{L}_2\mathcal{Z}^{-1}$, but not closed. In this case, T_0 has a one-sided inverse which is unbounded. Similar issues played a role in the embedding problem [14], where it could be shown that, even when R became unbounded, the product $R^*(D^*B + C^*MA)$ and $R^*(C^*MC + D^*D)$ would remain bounded because of range conditions that are automatically satisfied. The same happens here, although the analysis is much simplified by the fact that we know already from the constriction in algorithm 4.1 that there exists a realization $\mathbf{T}_0$ which is bounded.

PROPOSITION 4.20. *In theorem 4.19, $(D^*D + C^*MC)^\dagger$ is bounded if $\mathrm{ran}(\cdot T)$ is closed.*

Whether the range is closed or not, M is bounded, as are the products $R^*(D^*D + C^*MC)$ and $R^*(C^*MA + D^*B)$.

PROOF If $\mathrm{ran}(\cdot T_0)$ is closed, then T_0 has a one-sided inverse which is again upper. It follows that in this case $\mathrm{ran}(\cdot D_{T_0})$ is closed, so that $D_{T_0}^* D_{T_0} = XRR^*X = XX^\dagger X = X = D^*D + C^*MC$ has closed range and a bounded pseudo-inverse. Because $T_0 = V^*T$, $\mathrm{ran}(\cdot T_0)$ can be closed only if $\mathrm{ran}(\cdot T)$ is closed. If $\mathrm{ran}(\cdot T)$ is closed, then $(\mathcal{X}_2 V)T_0$ is closed. But from $V^*V = I$ it follows that $\mathcal{X}_2 V = \mathcal{X}_2$, so that in this case $\mathrm{ran}(\cdot T_0)$ is closed, too.

Because the realization of T is observable, it was argued in proposition 4.18 that Y (and hence M) is bounded. From the first equality in (4.22) we see that $\left[\begin{smallmatrix} C_V \\ D_V \end{smallmatrix}\right]$ is obtained by taking an orthonormal basis in the closure of the range of $\left[\begin{smallmatrix} YC \\ D \end{smallmatrix}\right]$. R is unbounded if the latter range is not closed. Nonetheless, $\left[\begin{smallmatrix} C_V \\ D_V \end{smallmatrix}\right]$ is well-defined and isometric, and $D_{T_0}^* = R(D^*D + C^*MC) = \left[\begin{smallmatrix} C_V \\ D_V \end{smallmatrix}\right]^* \left[\begin{smallmatrix} YC \\ D \end{smallmatrix}\right]$ is bounded. In the same way, it is shown that $C_{T_0} = R^*(C^*MA + D^*B) = \left[\begin{smallmatrix} C_V \\ D_V \end{smallmatrix}\right]^* \left[\begin{smallmatrix} YA \\ B \end{smallmatrix}\right]$ is bounded. □

"Square-root" solution of the Riccati equation

The computation of the outer factor can be done along the lines of theorem 4.19, which essentially boils down to recursively computing M_k in equation (4.23). However, as is well known, M_{k+1} in the Riccati recursion can be computed more efficiently using square-root algorithms (see *e.g.*, Morf [15] for a list of pre-1975 references). In such algorithms, the square-root Y of M is computed, rather than M itself. The square-root algorithm that corresponds to the above equations is very related to algorithm 4.1, but is written in a more structured way. The algorithm acts on data known at the k-th step: the state matrices

A_k, B_k, C_k, D_k, and the matrix Y_k. This data is collected in a matrix $\mathbf{T}'_k$:

$$\mathbf{T}'_k = \left[\begin{array}{c|c} Y_k & \\ \hline & I \end{array}\right] \left[\begin{array}{c|c} A_k & C_k \\ \hline B_k & D_k \end{array}\right]$$

The algorithm consists in computing a unitary matrix $\mathbf{W}_k$ such that $\mathbf{W}_k^* \mathbf{T}'_k$ has zero entries (of maximal dimensions) in the indicated positions:

$$\mathbf{W}_k^* \mathbf{T}'_k = \left[\begin{array}{c|c} Y_{k+1} & 0 \\ 0 & 0 \\ \hline N_k & R_k^{-1} \end{array}\right] := \mathbf{T}''_k, \tag{4.24}$$

where the non-zero block entries in $\mathbf{T}''_k$ are defined as Y_{k+1}, N_k and R_k^{-1}. Given Y_k, $\mathbf{W}_k$ can be obtained by a simple QR-factorization. Y_{k+1} is used in the recursion for the next step. The connection with the Riccati equation is given by the following lemma.

LEMMA 4.21. *In the above square-root algorithm, $M_k = Y_k^* Y_k$ and R_k satisfy the relations in theorem 4.19, and $\{\mathbf{W}_k\}$ is the realization of the inner factor W which embeds V.*

PROOF Since $\mathbf{W}_k$ is unitary, $(\mathbf{T}'_k)^* \mathbf{T}'_k = (\mathbf{T}''_k)^* \mathbf{T}''_k$. Writing out the corresponding equalities gives

$$\begin{cases} C_k^* M_k C_k + D_k^* D_k &= R_k^{-*} R_k^{-1} \\ A_k^* M_k C_k + B_k^* D_k &= N_k^* R_k^{-1} \\ A_k^* M_k A_k + B_k^* B_k &= M_{k+1} + N_k^* N_k \end{cases}$$

$$\Rightarrow \begin{cases} C_k^* M_k C_k + D_k^* D_k = (R_k R_k^*)^{-1} \\ A_k^* M_k A_k + B_k^* B_k - (A_k^* M_k C_k + B_k^* D_k) R_k R_k^* (D_k^* B_k + C_k^* M_k A_k) = M_{k+1}. \end{cases}$$

Hence M_k and R_k satisfy the equations in theorem 4.19.

It remains to show that $\mathbf{W}_k$, when taken equal to the realization of the inner factor W as computed in algorithm 4.1, indeed satisfies equation (4.24). In the notation of that algorithm, let

$$\mathbf{W} = \left[\begin{array}{c|cc} A_V & C_U & C_V \\ \hline B_V & D_U & D_V \end{array}\right] = \left[\begin{array}{cc} \begin{bmatrix} A'_V \\ B'_V \end{bmatrix} Q^* & \begin{matrix} C_V \\ D_V \end{matrix} \end{array}\right]$$

where $Q = \left[\begin{smallmatrix} Q_1 \\ Q_2 \end{smallmatrix}\right]$. Then

$$\mathbf{W}_k^* \left[\begin{array}{c|c} Y_k A_k & Y_k C_k \\ \hline B_k & D_k \end{array}\right] = \left[\begin{array}{c|c} Q_k^* \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix}^* \begin{bmatrix} Y_k A_k \\ B_k \end{bmatrix} & Q_k^* \begin{bmatrix} A'_{V,k} \\ B'_{V,k} \end{bmatrix}^* \begin{bmatrix} Y_k C_k \\ D_k \end{bmatrix} \\ \hline \begin{bmatrix} C_{V,k} \\ D_{V,k} \end{bmatrix}^* \begin{bmatrix} Y_k A_k \\ B_k \end{bmatrix} & \begin{bmatrix} C_{V,k} \\ D_{V,k} \end{bmatrix}^* \begin{bmatrix} Y_k C_k \\ D_k \end{bmatrix} \end{array}\right]$$

$$= \left[\begin{array}{c|c} \begin{bmatrix} Y_{k+1} \\ 0 \end{bmatrix} & 0 \\ \hline * & R_k^{-1} \end{array}\right].$$

□

Inner-outer factorization examples

We finish this chapter with some examples of the inner-outer factorization algorithm on a time-invariant system and on a number of finite (4×4) matrices. In the finite matrix case, interesting things can occur only when T is singular or when the dimensions of T are not uniform.

1. Consider the time-invariant system

$$T = \frac{z - \alpha^*}{1 - \beta z} = \frac{1 - \alpha z}{1 - \beta z} \cdot \frac{z - \alpha^*}{1 - \alpha z}.$$

A state-space realization of T is

$$\mathbf{T} = \begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} \beta & 1 \\ 1 - \alpha^* \beta & -\alpha^* \end{bmatrix}.$$

Its zeros are $(a - cd^{-1}b)^{-1} = (\beta - \alpha^{-*}(1 - \alpha^* \beta))^{-1} = \alpha^*$. Hence α is indeed a solution of equation (4.20). Substitution leads to $\beta = (1 - \alpha^* \alpha)^{1/2}$ and $y = \beta^*$.

2. Using algorithm 4.1 on

$$T = \begin{bmatrix} \underline{0} & 1 & 4 & 6 \\ 0 & \underline{0} & 2 & 5 \\ 0 & 0 & \underline{0} & 3 \\ 0 & 0 & 0 & \underline{0} \end{bmatrix}$$

(the underlined entries form the 0-th diagonal) yields an almost trivial isometric factor V or inner factor W:

$$V = \begin{bmatrix} \cdot & 1 & 0 & 0 \\ \cdot & \underline{0} & 1 & 0 \\ \cdot & 0 & \underline{0} & 1 \\ \cdot & 0 & 0 & \underline{0} \end{bmatrix} \qquad W = \begin{bmatrix} \cdot & 1 & 0 & 0 & 0 \\ \cdot & \underline{0} & 1 & 0 & 0 \\ \cdot & 0 & \underline{0} & 1 & 0 \\ \cdot & 0 & 0 & \underline{0} & \underline{1} \end{bmatrix}$$

$$\begin{aligned} \#\mathcal{M}_W &= [1\ 1\ 1\ 1] \\ \#\mathcal{N}_W &= [0\ 1\ 1\ 2] \\ \#\mathcal{B}_W &= [0\ 1\ 1\ 1] \end{aligned}$$

It is seen that V is not inner, because T is singular. W is the inner extension of V. The only effect of W is a redefinition of time intervals: W acts as a shift operator. $T_0 = W^* T$ is

$$W^* T = \begin{bmatrix} \cdot & \cdot & \cdot & \cdot \\ 0 & \underline{1} & 4 & 6 \\ 0 & 0 & \underline{2} & 5 \\ 0 & 0 & 0 & \underline{3} \\ 0 & 0 & 0 & \underline{0} \end{bmatrix}$$

$$\begin{aligned} \#\mathcal{M}_{T_0} &= [0\ 1\ 1\ 2] \\ \#\mathcal{N}_{T_0} &= [1\ 1\ 1\ 1] \end{aligned}$$

The multiplication by W^* has shifted the rows of T downwards. This is possible: the result T_0 is still upper. V^*T is equal to W^*T with its last row removed.

3. Take

$$T = \begin{bmatrix} \underline{0} & 1 & 4 & 6 \\ 0 & \underline{1} & 2 & 5 \\ 0 & 0 & \underline{1} & 3 \\ 0 & 0 & 0 & \underline{1} \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M} &= [1\ 1\ 1\ 1] \\ \#\mathcal{N} &= [1\ 1\ 1\ 1] \\ \#\mathcal{B} &= [0\ 1\ 2\ 1] \end{aligned}$$

Hence T is again singular, but now a simple shift will not suffice. The algorithm computes W as

$$W = \begin{bmatrix} \vdots & -0.707 & 0.577 & 0.367 & 0.180 \\ \cdot & \underline{-0.707} & -0.577 & -0.367 & -0.180 \\ \cdot & 0 & \underline{0.577} & -0.733 & -0.359 \\ \cdot & 0 & 0 & \underline{-0.440} & \underline{0.898} \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M}_W &= [1\ 1\ 1\ 1] \\ \#\mathcal{N}_W &= [0\ 1\ 1\ 2] \\ \#\mathcal{B}_W &= [0\ 1\ 1\ 1] \end{aligned}$$

$$T_0 = W^*T = \begin{bmatrix} \vdots & \cdot & \cdot & \cdot \\ 0 & \underline{-1.414} & -4.243 & -7.778 \\ 0 & 0 & \underline{1.732} & 2.309 \\ 0 & 0 & 0 & \underline{-2.273} \\ 0 & 0 & 0 & \underline{0} \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M}_{T_0} &= [0\ 1\ 1\ 2] \\ \#\mathcal{N}_{T_0} &= [1\ 1\ 1\ 1] \end{aligned}$$

V is equal to W with its last column removed, so that $T_0 = V^*T$ is equal to the above T_0 with its last row removed.

4. In the previous examples, we considered only systems T with a constant number of inputs and outputs (equal to 1), for which $V \neq I$ only if T is singular. However, a non-identical V can also occur if the number of inputs and outputs of T varies in time. Thus consider

$$T = \begin{bmatrix} \underline{1.000} & 0.500 & 0.250 & 0.125 \\ \underline{1.000} & 0.300 & 0.100 & 0.027 \\ 0 & \underline{1.000} & 0.500 & 0.250 \\ 0 & 0 & \underline{1.000} & 0.300 \\ \cdot & \cdot & \cdot & \vdots \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M} &= [2\ 1\ 1\ 0] \\ \#\mathcal{N} &= [1\ 1\ 1\ 1] \\ \#\mathcal{B} &= [0\ 1\ 2\ 1] \end{aligned}$$

$$V = \begin{bmatrix} \underline{-0.707} & 0.099 & 0.025 & -0.699 \\ \underline{-0.707} & -0.099 & -0.025 & 0.699 \\ 0 & \underline{0.990} & -0.005 & 0.139 \\ 0 & 0 & \underline{0.999} & 0.035 \\ \cdot & \cdot & \cdot & \vdots \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M}_V &= [2\ 1\ 1\ 0] \\ \#\mathcal{N}_V &= [1\ 1\ 1\ 1] \\ \#\mathcal{B}_V &= [0\ 1\ 1\ 1] \end{aligned}$$

In this case, V is itself inner. The outer factor T_0 follows as

$$T_0 = V^*T = \begin{bmatrix} \underline{-1.414} & -0.565 & -0.247 & -0.107 \\ 0 & \underline{1.010} & 0.509 & 0.257 \\ 0 & 0 & \underline{1.001} & 0.301 \\ 0 & 0 & 0 & \underline{-0.023} \end{bmatrix} \qquad \begin{aligned} \#\mathcal{M}_{T_0} &= [1\ 1\ 1\ 1] \\ \#\mathcal{N}_{T_0} &= [1\ 1\ 1\ 1] \end{aligned}$$

An interesting observation from these examples is that the inner-outer factorization of finite matrices T is equal to the QR factorization of T when it is considered as an ordinary matrix without block entries. In combination with the external factorization, this observation can be used to efficiently compute the QR factorization of a general block matrix (mixed upper-lower) if both its upper and its lower parts have state realizations of low dimensions. Let X be such a matrix, then first compute U such that $T = UX$ is upper (U follows from an external factorization of $\mathbf{P}(X^*) =: \Delta^* U$), and subsequently compute the inner-outer factorization of T as $T = VT_0$. Then the QR factorization of X follows as $X = (U^*V)T_0$. Note that, if the square-root algorithm is used, then in this scheme the global QR factorization of X is replaced by local QR factorizations of state-space matrices.

As mentioned in the introduction, another application of the inner-outer factorization is in the inversion of matrices. Equation 2.16 showed that the inverse of a block-upper matrix need not necessarily be upper again. The same is obvious for a matrix of mixed causality, say again X. Assume that X is invertible and that we know state realizations of its upper and lower part of low order. To compute the state realization of the lower and upper parts of X^{-1}, first compute the QR factorization $X = U^*VT_0$ as above. Here, all factors have known state realizations. Then $X^{-1} = T_0^{-1}V^*U$, where V^* and U have known state realizations. A state realization of T_0^{-1} is straightforward to compute, and was given in equation (1.3).

Bibliography

[1] H. Helson, *Lectures on Invariant Subspaces.* New York: Academic Press, 1964.

[2] W. Arveson, "Interpolation problems in nest algebras," *J. Functional Anal.*, vol. 20, pp. 208–233, 1975.

[3] M. Rosenblum and J. Rovnyak, *Hardy Classes and Operator Theory.* Oxford Univ. Press, 1985.

[4] A.J. van der Veen, "Computation of the inner-outer factorization for time-varying systems," in *Challenges of a Generalized System Theory* (M. Verhaegen et al., ed.), Essays of the Royal Dutch Academy of Sciences, (Amsterdam, The Netherlands), 1993.

[5] A. Feintuch and B.A. Francis, "Uniformly optimal control of linear feedback systems," *Automatica*, vol. 21, no. 5, pp. 563–574, 1985.

[6] P. Dewilde, "Input-output description of roomy systems," *SIAM J. Control and Optimization*, vol. 14, pp. 712–736, July 1976.

[7] K. Poolla and P. Khargonekar, "Stabilizability and stable-proper factorizations for linear time-varying systems," *SIAM J. Control and Optimization*, vol. 25, pp. 723–736, May 1987.

[8] W.N. Dale and M.C. Smith, "Existence of coprime factorizations for time-varying systems—an operator-theoretic approach," in *Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing I (Proc. Int. Symp. MTNS-91)* (H. Kimura and S. Kodama, eds.), pp. 177–182, MITA Press, Japan, 1992.

[9] R. Ravi, A.M. Pascoal, and P.P. Khargonekar, "Normalized coprime factorizations for linear time-varying systems," *Systems and Control Letters*, vol. 18, pp. 455–465, 1992.

[10] A. Beurling, "On two problems concerning linear transformations in Hilbert space," *Acta Math.*, vol. 81, pp. 239–255, 1949.

[11] P.D. Lax, "Translation invariant subspaces," *Acta Math.*, vol. 101, pp. 163–178, 1959.

[12] P. Dewilde and H. Dym, "Interpolation for upper triangular operators," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 153–260, Birkhäuser Verlag, 1992.

[13] F.J. Beutler and W.L Root, "The operator pseudo-inverse in control and systems identification," in *Generalized Inverses and Applications* (M. Zuhair Nashed, ed.), pp. 397–494, Academic Press, 1976.

[14] A.J. van der Veen and P.M. Dewilde, "Embedding of time-varying contractive systems in lossless realizations," *subm. Math. Control Signals Systems*, July 1992.

[15] M. Morf and T. Kailath, "Square-root algorithms for Least-Squares Estimation," *IEEE Trans. Automat. Control*, vol. 20, no. 4, pp. 487–497, 1975.

Chapter 5

J-UNITARY OPERATORS

[1] In the previous sections, we studied systems by their transfer operators T, by means of certain characteristic subspaces of T. One other way to study operators is by their *graphs*: let $T \in \mathcal{X}$, and $y = uT$, then the graph of T is the subspace $\mathcal{H} \subset \mathcal{X}_2 \times \mathcal{X}_2$ spanned by

$$[u \quad y] = u[I \quad T] \qquad (\text{all } u \in \mathcal{X}_2). \tag{5.1}$$

In this respect, T is called the *angle* operator for $\mathcal{H}$. A given subspace in $\mathcal{X}_2 \times \mathcal{X}_2$ does not always have an angle operator; if it does, then the subspace is generally called *admissible*. The advantage of using graphs to represent operators is that cascade connections of inner operator two-ports lead to the application of linear matrix operators to subspaces, rather than complicated linear fractional transforms. (This is explained in section 5.1.)

In (5.1), if we call $\| u \|_{HS}^2$ the ingoing energy of the system T, and $\| y \|_{HS}^2$ the outgoing energy, then we say that the system is contractive if $\| u \|_{HS}^2 - \| y \|_{HS}^2$ is positive for all possible u, isometric if it is always equal to zero, and expansive if it is always negative. For a given graph $\mathcal{H}$, we can check the type of the corresponding angle operator by evaluating the sign of $\| x_1 \|_{HS}^2 - \| x_2 \|_{HS}^2$, for all elements $x = [x_1 \quad x_2]$ in $\mathcal{H}$. To this end, it is useful to define an *indefinite* metric for such subspaces, based on an indefinite diagonal inner product $[\cdot, \cdot]$:

$$x = [x_1 \quad x_2] : \qquad \begin{aligned} [x, x] &= \{x_1, x_1\} - \{x_2, x_2\} \\ &= \{xJ, x\}, \qquad \text{where } J = \begin{bmatrix} I & \\ & -I \end{bmatrix}. \end{aligned}$$

Here, $\{\cdot, \cdot\}$ is the usual diagonal inner product as defined in equation (2.26). J is called a signature operator, and elements of subspaces $\mathcal{H}$ are said to be J-positive, J-neutral or J-negative if their 'J-norm' (which is a diagonal) is entrywise positive, zero or negative,

[1] Starting from this section, operators in $\mathcal{X}_2$ are rendered in lower case symbols, to avoid confusion with operators in $\mathcal{X}$.

respectively. Subspaces $\mathcal{H}$ can be classified in the same way, but clearly, mixed (indefinite) cases can also occur. Operators which leave the J-norm of all vectors invariant are J-isometric, or J-unitary if they are also of full range (invertible). Such operators play an important role in this and the next chapter.

We prove, in section 5.2, a theorem of the following form.

THEOREM 5.1. *Let $\mathcal{H}$ be a locally finite left DZ-invariant subspace in $\mathcal{L}_2\mathcal{Z}^{-1}$, with a bounded basis representation $\mathbf{F}$ whose J-Gramian $\Lambda_{\mathbf{F}}^J = \mathbf{P}_0(\mathbf{F}J\mathbf{F}^*)$ is boundedly invertible. Then there exists a bounded J-unitary operator $\Theta \in \mathcal{U}$ such that $\mathcal{H} = \mathcal{H}(\Theta)$, the input state space of Θ.*

Theorem 5.1 plays the role of a generalized Beurling-Lax theorem. It represents a shift-invariant subspace by a J-unitary operator. Similar generalizations (for the time-invariant setting) have been considered by Ball and Helton [1, 2, 3], and earlier in other ways by De Branges in a reproducing kernel Hilbert space theory ([4]; *cf.* [5], theorem 4.1). They were introduced for the purpose of factorization, interpolation and approximation. There is a J-inner-coprime factorization theorem (corollary 5.20), and there are applications such as J-spectral factorization of indefinite Hermitian operators and generalized Wiener-Hopf factorizations. However, the main application of the representation theorem is the description of the solutions of certain interpolation problems, connected with the names of Hermite-Fejér-Carathéodory, Nevanlinna-Pick, Schur-Takagi, Nehari, as well as others. In the generic case, one searches for contractive functions whose Fourier coefficients or z-transforms have prescribed values at prescribed points. The generalization of scalar functions to matrix valued functions exposed the connection of the state space of certain J-unitary operators with the "interpolation subspace", or reproducing kernel Hilbert space as specified by the interpolation data. Such generalizations were studied by many mathematicians, such as Potapov, Sz.-Nagy, Foias, Adamjan, Arov and Krein, Sarason, Rosenblum-Rovnyak, and others, and have been accumulated into two monographs, by Dym [5] and Ball-Gohberg-Rodman [6]. The references in these books provide a more detailed historical perspective. The first book emphasizes connections with functional analysis, while the second is based on state space techniques from systems theory. The past decade saw renewed interest in such constrained interpolation problems, fostered by engineering applications such as stochastic prediction and estimation [7, 8, 9], robust ($\mathcal{H}_\infty$) control [10], optimal filtering, sensitivity minimization and optimal (Hankel-norm) model approximation. The connection of interpolation theory with approximation problem again goes via the representation theory for shift-invariant subspaces. While for interpolation problems, the subspace typically has to be definite (positive or negative) for solutions to exist, it is indefinite in approximation applications: in fact, approximations are obtained by removing the positive or negative part of the subspace.

This chapter gives preliminary and elementary results on J-unitary operators and their

realizations, and gives a time-varying version of the representation theorem. Section 5.3 illustrates this theorem by introducing the time-varying Nevanlinna-Pick problem, and solving a more general ("fundamental") interpolation problem in which the interpolation subspace is uniformly positive. The more general case of indefinite subspaces is deferred to chapter 6, and this chapter provides in fact the reason for including any material on J-unitary operators.

5.1 REALIZATION OF J-UNITARY OPERATORS

Signatures

Let be given a (possibly non-uniform) sequence of spaces $\mathcal{M}$, and consider a partitioning of $\mathcal{M}$ into two space sequences $\mathcal{M}_+$ and $\mathcal{M}_-$ of lower s-dimension:

$$\mathcal{M} = \mathcal{M}_+ \times \mathcal{M}_-, \qquad \#\mathcal{M} = \#\mathcal{M}_+ + \#\mathcal{M}_-,$$

where '#' indicates the sequence of dimensions of its argument. Such a partitioning generates a partitioning of the Hilbert space $\ell_2^{\mathcal{M}}$ into two components $\ell_2^{\mathcal{M}_+}$ and $\ell_2^{\mathcal{M}_-}$,

$$\ell_2^{\mathcal{M}} = \ell_2^{\mathcal{M}_+} \times \ell_2^{\mathcal{M}_-} = [\ell_2^{\mathcal{M}_+} \quad 0] \oplus [0 \quad \ell_2^{\mathcal{M}_-}],$$

and also

$$\mathcal{X}_2^{\mathcal{M}} = \mathcal{X}_2^{\mathcal{M}_+} \times \mathcal{X}_2^{\mathcal{M}_-} = [\mathcal{X}_2^{\mathcal{M}_+} \quad 0] \oplus [0 \quad \mathcal{X}_2^{\mathcal{M}_-}].$$

Let

$$J_{\mathcal{M}} = \begin{bmatrix} I_{\mathcal{M}_+} & \\ & -I_{\mathcal{M}_-} \end{bmatrix}$$

be a signature matrix corresponding to the above partitioning of $\mathcal{M}$. Likewise, consider a second space sequence $\mathcal{N} = \mathcal{N}_+ \times \mathcal{N}_-$, with signature matrix $J_{\mathcal{N}}$. An operator $\Theta : \ell_2^{\mathcal{M}} \to \ell_2^{\mathcal{N}}$ has a partitioning conformably to the partitioning of $\mathcal{M}$, $\mathcal{N}$, as

$$\Theta = \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \tag{5.2}$$

where

$$\begin{aligned} \Theta_{11} &\in \mathcal{X}(\mathcal{M}_+, \mathcal{N}_+), \qquad \Theta_{12} \in \mathcal{X}(\mathcal{M}_+, \mathcal{N}_-), \\ \Theta_{21} &\in \mathcal{X}(\mathcal{M}_-, \mathcal{N}_+), \qquad \Theta_{22} \in \mathcal{X}(\mathcal{M}_-, \mathcal{N}_-). \end{aligned}$$

Such an operator Θ is said to be a $(J_{\mathcal{M}}, J_{\mathcal{N}})$-isometry if $\Theta J_{\mathcal{N}} \Theta^* = J_{\mathcal{M}}$, a $(J_{\mathcal{M}}, J_{\mathcal{N}})$-co-isometry if $\Theta^* J_{\mathcal{M}} \Theta = J_{\mathcal{N}}$, and $(J_{\mathcal{M}}, J_{\mathcal{N}})$-unitary if both $\Theta J_{\mathcal{N}} \Theta^* = J_{\mathcal{M}}$ and $\Theta^* J_{\mathcal{M}} \Theta = J_{\mathcal{N}}$. If Θ is $(J_{\mathcal{M}}, J_{\mathcal{N}})$-unitary, then Θ^{-1} is bounded, and $\Theta^{-1} = J_{\mathcal{N}} \Theta^* J_{\mathcal{M}}$.

If Θ is J-unitary, then it is unitary with respect to an indefinite diagonal inner product

$$[x, y] = \{xJ, y\} = \mathbf{P}_0(xJy^*),$$

that is, with respect to the corresponding indefinite Hilbert-Schmidt inner product trace $[x, y]$. Actually, there are two inner products involved, one corresponding to $\mathcal{M}$, the other to $\mathcal{N}$. This inner product is called indefinite because $[x, x]$ is a diagonal with entries that can be positive, zero, and negative, depending on x. Note that $[x, x] = 0$ can occur even if $x \neq 0$, for example if $x_+ = x_-$. One can show that an operator Θ is J-isometric if

$$[x\Theta, x\Theta] = [x, x] \tag{5.3}$$

for all $x \in \mathcal{X}_2$, and J-unitary if moreover it is invertible. Alternatively, J-unitarity can be defined this way (as is done in [11]) and the properties $\Theta^* J\Theta = J$, $\Theta J\Theta^* = J$ be derived from it.

The Hilbert-Schmidt space $\mathcal{X}_2$, endowed with the indefinite J-inner product, is known as a Krein space, which is a special case of a space with an indefinite metric. The classical operator theory for such spaces goes back to a paper of Dirac on quantum field theory in 1942 and in a more mathematical context to the work of Pontrjagin, Krein, Yokhvidov, Bognar and many others. Standard textbooks are [12, 13, 14], which also provide connections with many of the early papers. We use only the J-unitary operators in this space. In particular, the input and output state spaces $\mathcal{H}(\Theta)$ and $\mathcal{H}_0(\Theta)$ will play an important role.

J-unitary and unitary operators

Associated to a J-unitary operator Θ is a unitary operator Σ which defines the same relations between signal quantities but regards different quantities as inputs and outputs. This very useful property is stated in the following theorem.

THEOREM 5.2. *Let $\Theta \in \mathcal{X}(\mathcal{M}, \mathcal{N})$ be a $(J_{\mathcal{M}}, J_{\mathcal{N}})$-unitary operator with partitioning (5.2). Then*

1. *Θ_{22}^{-1} exists and is bounded,*
2. *$\|\Theta_{22}^{-1}\| \leq 1$, $\|\Theta_{22}^{-1}\Theta_{21}\| < 1$, $\|\Theta_{12}\Theta_{22}^{-1}\| < 1$.*
3. *There exists an operator $\Sigma \in \mathcal{X}(\mathcal{M}_+ \times \mathcal{N}_-, \mathcal{N}_+ \times \mathcal{M}_-)$ such that*

$$[a_1 \quad b_1]\Theta = [a_2 \quad b_2] \quad \Leftrightarrow \quad [a_1 \quad b_2]\Sigma = [a_2 \quad b_1]$$

for $a_1 \in \mathcal{X}_2^{\mathcal{M}_+}$, $a_2 \in \mathcal{X}_2^{\mathcal{N}_+}$, $b_1 \in \mathcal{X}_2^{\mathcal{M}_-}$, $b_2 \in \mathcal{X}_2^{\mathcal{N}_-}$. Σ is unitary, and given by

$$\Sigma = \begin{bmatrix} \Theta_{11} - \Theta_{12}\Theta_{22}^{-1}\Theta_{21} & -\Theta_{12}\Theta_{22}^{-1} \\ \Theta_{22}^{-1}\Theta_{21} & \Theta_{22}^{-1} \end{bmatrix}. \tag{5.4}$$

PROOF The proofs are elementary and well known; see *e.g.*, [15, lemma 5.2], [11].

1. $\Theta J\Theta^* = J$ and $\Theta^* J\Theta = J$ give the relations

$$\Theta_{22}\Theta_{22}^* = I + \Theta_{21}\Theta_{21}^*, \qquad \Theta_{22}^*\Theta_{22} = I + \Theta_{12}^*\Theta_{12}.$$

Hence Θ_{22} and Θ_{22}^* both have closed range and empty kernel, so that Θ_{22} is boundedly invertible, and both $\Theta_{22}^{-1}\Theta_{22}^{-*} \gg 0$ and $\Theta_{22}^{-*}\Theta_{22}^{-1} \gg 0$.

2. Applying Θ_{22}^{-1} and Θ_{22}^{-*} to the above two expressions yields

$$I = \Theta_{22}^{-1}\Theta_{22}^{-*} + (\Theta_{22}^{-1}\Theta_{21})(\Theta_{22}^{-1}\Theta_{21})^*, \qquad I = \Theta_{22}^{-*}\Theta_{22}^{-1} + (\Theta_{12}\Theta_{22}^{-1})^*(\Theta_{12}\Theta_{22}^{-1}).$$

Hence $\Theta_{22}^{-1}\Theta_{22}^{-*} \leq I$ and $\Theta_{22}^{-*}\Theta_{22}^{-1} \leq I$, *i.e.*, $\|\Theta_{22}^{-1}\| \leq 1$. Because $\Theta_{22}^{-1}\Theta_{22}^{-*} \gg 0$ and $\Theta_{22}^{-*}\Theta_{22}^{-1} \gg 0$ it follows that $\| \Theta_{22}^{-1}\Theta_{21} \| < 1$ and $\| \Theta_{12}\Theta_{22}^{-1} \| < 1$.

3. Writing out the expression $[a_1 \quad b_1]\Theta = [a_2 \quad b_2]$ in full gives

$$\begin{cases} a_1\Theta_{11} + b_1\Theta_{21} = a_2 \\ a_1\Theta_{12} + b_1\Theta_{22} = b_2 \end{cases} \Leftrightarrow \begin{cases} a_1(\Theta_{11} - \Theta_{12}\Theta_{22}^{-1}\Theta_{21}) + b_2\Theta_{22}^{-1}\Theta_{21} = a_2 \\ -a_1\Theta_{12}\Theta_{22}^{-1} + b_2\Theta_{22}^{-1} = b_1 \end{cases}$$

as Θ_{22} is invertible. The second set of equations is $[a_1 \quad b_2]\Sigma = [a_2 \quad b_1]$. The fact that Σ is unitary can be verified by computing $\Sigma^*\Sigma$ and $\Sigma\Sigma^*$ in terms of its block entries. □

An alternative proof of the fact that Σ is unitary uses equation (5.3). As Θ is J-unitary, we have for all a_1, b_1:

$$\begin{array}{rrcl} & [[a_1 \quad b_1]\Theta, [a_1 \quad b_1]\Theta] & = & [[a_1 \quad b_1], [a_1 \quad b_1]] \\ \Leftrightarrow & [[a_2 \quad b_2], [a_2 \quad b_2]] & = & [[a_1 \quad b_1], [a_1 \quad b_1]] \\ \Leftrightarrow & \mathbf{P}_0(a_2a_2^*) - \mathbf{P}_0(b_2b_2^*) & = & \mathbf{P}_0(a_1a_1^*) - \mathbf{P}_0(b_1b_1^*) \\ \Leftrightarrow & \mathbf{P}_0(a_2a_2^*) + \mathbf{P}_0(b_1b_1^*) & = & \mathbf{P}_0(a_1a_1^*) + \mathbf{P}_0(b_2b_2^*) \\ \Leftrightarrow & \{[a_2 \quad b_1], [a_2 \quad b_1]\} & = & \{[a_1 \quad b_2], [a_1 \quad b_2]\} \\ \Leftrightarrow & \{[a_1 \quad b_2]\Sigma, [a_1 \quad b_2]\Sigma\} & = & \{[a_1 \quad b_2], [a_1 \quad b_2]\} \end{array}$$

for all a_1, b_2, which shows that Σ is an isometry. The derivation can be repeated for Θ^{-1}, which gives that Σ is a co-isometry.

Σ is known as the scattering operator corresponding to Θ. The relation between Θ and Σ is drawn in figure 5.1. Note that $\Theta \to \Sigma$ exists for any J-unitary operator Θ, but the converse $\Sigma \to \Theta$ gives a bounded operator Θ only if Σ_{22} is invertible, in which case $\Sigma_{22} = \Theta_{22}^{-1}$. Hence considering only bounded operators Θ is in some sense 'special', and indeed gives rise to certain special properties of its state space, as will be discussed later.

Note that, although Θ_{22}^{-1} exists as a bounded operator, it is not necessarily true that $\Theta_{22}^{-1} \in \mathcal{U}$, even if Θ_{22} is itself upper. Θ is called J-inner if the corresponding scattering operator Σ is inner, that is, unitary and *upper*. (An alternative definition avoids the use of Σ and uses projections instead; see [11].) If $\Theta \in \mathcal{U}$, then Θ is J-inner if and only if Θ_{22}^{-1} is upper.

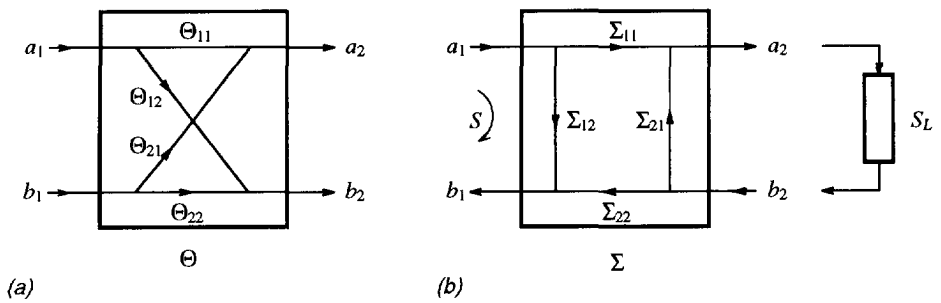


Figure 5.1. The connection between Θ and the corresponding scattering operator Σ.

Linear fractional transformations

An important application of J-unitary operators is in the calculation of the transfer operator of a Σ-section, terminated in a 'load' S_L, as in figure 5.1. If the spectral radius of $\Sigma_{21}S_L$ is smaller than 1 (which in view of theorem 5.2 will always be the case if S_L is contractive: $\| S_L \| \leq 1$), then S can be determined as

$$S = \Sigma_{12} + \Sigma_{11}S_L(I - \Sigma_{21}S_L)^{-1}\Sigma_{22}. \tag{5.5}$$

However, if instead of a single Σ operator, a cascade of such operators is placed between the ports $[a_1 \quad b_1]$ and S_L, the computation of S in terms of S_L and the Σ's becomes more complicated (the resulting formula is known as a Redheffer product). But S can also be determined in terms of Θ, which is interesting because a cascade of Θ-sections is again a Θ operator, and is equal to the product of the individual transfer matrices of the Θ-sections: Θ is a 'chain scattering operator'. It remains to specify S in terms of Θ and S_L. S_L gives a relation between a_2 and b_2: $b_2 = a_2S_L$. Let S be the transfer of a_1 to b_1: $b_1 = a_1S$, then S_L and S are angle operators for the graphs $[a_2 \quad b_2]$ and $[a_1 \quad b_1]$. They satisfy

$$b_2 = a_2S_L \quad \Leftrightarrow \quad [a_2 \quad b_2]\begin{bmatrix} S_L \\ -I \end{bmatrix} = 0,$$

$$b_1 = a_1S \quad \Leftrightarrow \quad [a_1 \quad b_1]\begin{bmatrix} S \\ -I \end{bmatrix} = 0.$$

Combining with the relation $[a_1 \quad b_1]\Theta = [a_2 \quad b_2]$ gives a relation between S and S_L as

$$\begin{bmatrix} S \\ -I \end{bmatrix}\Phi = \Theta\begin{bmatrix} S_L \\ -I \end{bmatrix} \tag{5.6}$$

for some operator Φ, given by $\Phi = \Theta_{22} - \Theta_{21}S_L$. If Φ is invertible, then S is given by the ratio

$$S := T_\Theta[S_L] = (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1}.$$

The expression is usually written as $S = T_\Theta[S_L]$. It remains to note that $\Phi = \Theta_{22}(I - \Theta_{22}^{-1}\Theta_{21}S_L) = \Theta_{22}^{-1}(I - \Sigma_{21}S_L)$ so that a sufficient condition for Φ to be invertible is, again, that S_L is contractive. For later use, the above remarks are collected in the following lemma.

LEMMA 5.3. *Let Θ be a J-unitary operator. Then*

$$S = T_\Theta[S_L] \quad \Leftrightarrow \quad \begin{bmatrix} S \\ -I \end{bmatrix} \Phi = \Theta \begin{bmatrix} S_L \\ -I \end{bmatrix},$$

where $\Phi = \Theta_{22}(I - \Theta_{22}^{-1}\Theta_{21}S_L)$ is invertible if S_L is contractive.

COROLLARY 5.4. *Let Θ be a J-unitary operator, and let S_L be contractive. Then $S = T_\Theta[S_L]$ is strictly contractive if and only if S_L is so.*

PROOF This is clear from $\Phi(I - S^*S)\Phi = I - S_L^*S_L$ (see also [16], lemma 4.1). □

Subspaces

With the connection $\Sigma \leftrightarrow \Theta$ as motivation, we investigate in some more detail the implications of the use of the indefinite (diagonal) inner product $[x, y] = \{xJ, y\}$. As mentioned before, for any $x \in \mathcal{X}_2$, $[x, x]$ is diagonal with entries that can have either sign, or be equal to 0 for $x \neq 0$. Also note that $\|[x, x]\| \leq \{x, x\}$, so that $x \in \mathcal{X}_2$ is also bounded in J-norm. With regard to subspaces in $\mathcal{X}_2$, we call a left D-invariant subspace $\mathcal{H} \subset \mathcal{X}_2$

- J-positive if $x \in \mathcal{H} \Rightarrow [x, x] \geq 0$,
- J-negative if $x \in \mathcal{H} \Rightarrow [x, x] \leq 0$,
- J-neutral if $x \in \mathcal{H} \Rightarrow [x, x] = 0$,
- J-uniformly positive if $\exists\, \varepsilon > 0 : x \in \mathcal{H} \Rightarrow [x, x] \geq \varepsilon\{x, x\}$.
- J-uniformly negative if $\exists\, \varepsilon > 0 : x \in \mathcal{H} \Rightarrow -[x, x] \geq \varepsilon\{x, x\}$.

The J-orthogonal complement of a D-invariant subspace $\mathcal{H}$ is $\mathcal{H}^{[\perp]}$, defined as

$$\mathcal{H}^{[\perp]} = \{x \in \mathcal{X}_2 : [x, y] = 0, \text{ all } y \in \mathcal{H}\} = \mathcal{H}^\perp J.$$

This subspace is also D-invariant.

On J-uniformly definite subspaces, the J-inner product is equivalent to the usual inner product: $\varepsilon\{x, x\} \leq [x, x] \leq \{x, x\}$, which ensures that important properties (such as completeness and closedness) carry over: a J-uniformly definite subspace is a Hilbert space subspace. We are, however, interested in more general cases than only uniformly definite

subspaces, namely cases where a subspace $\mathcal{H}$ can be split into $\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$, where $\mathcal{H}_+$ and $\mathcal{H}_-$ are uniformly positive and negative subspaces, and '$\boxplus$' denotes the J-orthogonal direct sum:

$$\mathcal{H} = \mathcal{A} \boxplus \mathcal{B} \quad \Leftrightarrow \quad \mathcal{H} = \mathcal{A} \dotplus \mathcal{B}, \quad \mathcal{A} \,[\perp]\, \mathcal{B}.$$

This indefinite direct sum is the analog of $\oplus$, but in using $\boxplus$, a number of properties that are a matter of course in Hilbert spaces no longer hold. For example, for a subspace in the usual (definite) inner product, we always have that $\mathcal{H} \cap \mathcal{H}^\perp = 0$ and $\mathcal{H} \oplus \mathcal{H}^\perp = \mathcal{X}_2$. With an indefinite metric, the analogous equations are in general not true. The intersection of $\mathcal{H}$ and $\mathcal{H}^{[\perp]}$ need not be empty: for example, if $\mathcal{H}$ is a neutral subspace, then $\mathcal{H} \subset \mathcal{H}^{[\perp]}$. With neutral subspaces, one can also show that a subspace and its J-complement do not necessarily span the whole space. Consequently, the algebraic sum $\mathcal{H}+\mathcal{H}^{[\perp]}$ need no longer be a direct sum: if one of the subspaces contains a neutral element, the decomposition is not unique.

To describe the situation, we require a number of additional definitions. A full discussion (of the case of 'classical' Hilbert spaces) can be found in [14]; here, we only go to the level of detail that is required in the following sections. Throughout, we consider only subspaces that are D-invariant.

A subspace $\mathcal{H}$ is said to be *projectively complete* if $\mathcal{H} + \mathcal{H}^{[\perp]} = \mathcal{X}_2$. In this case, each $x \in \mathcal{X}_2$ has at least one decomposition into $x = x_0 + x_1 \in \mathcal{H} + \mathcal{H}^{[\perp]}$. A vector $x \in \mathcal{H}$ is called the *J-orthogonal projection* of a vector $y \in \mathcal{X}_2$ if (*i*) $x \in \mathcal{H}$ and (*ii*) $y - x \,[\perp]\, \mathcal{H}$.

Let $\mathcal{H}_0 = \mathcal{H} \cap \mathcal{H}^{[\perp]}$. $\mathcal{H}$ is called a *non-degenerate* subspace if $\mathcal{H}_0 = 0$. It is straightforward to show that $[\mathcal{H} + \mathcal{H}^{[\perp]}]^{[\perp]} = \mathcal{H}^{[\perp]} \cap \mathcal{H}^{[\perp][\perp]} = \mathcal{H}^{[\perp]} \cap \mathcal{H} = \mathcal{H}_0$, so that

$$\mathcal{X}_2 = (\mathcal{H} + \mathcal{H}^{[\perp]}) \oplus \mathcal{H}_0 J. \tag{5.7}$$

It follows that $\mathcal{H}$ can be projectively complete only if it is non-degenerate: $\mathcal{H} \cap \mathcal{H}^{[\perp]} = 0$. In that case, decompositions are unique, so that *if $\mathcal{H}$ is projectively complete, then* $\mathcal{X}_2 = \mathcal{H} \boxplus \mathcal{H}^{[\perp]}$.

If $\mathcal{H}$ is a locally finite D-invariant subspace, then it has some strong basis representation $\mathbf{F}$ such that $\mathcal{H} = \mathcal{D}_2^{\mathcal{B}}\mathbf{F}$ (*cf.* prop. 2.6). Here, $\mathcal{B}$ is the non-uniform space sequence whose dimension $\#\mathcal{B}$ is the sequence of dimensions of the subspace $\mathcal{H}$. In analogy with the definition of the Gram operator $\Lambda_{\mathbf{F}} = \{\mathbf{F}, \mathbf{F}\}$ in chapter 2, we define the J-Gram operator of this basis as the diagonal operator

$$\Lambda_{\mathbf{F}}^J = [\mathbf{F}, \mathbf{F}] = \mathbf{P}_0(\mathbf{F}J\mathbf{F}^*) \in \mathcal{D}(\mathcal{B}, \mathcal{B}). \tag{5.8}$$

$\mathbf{F}$ is J-orthonormal if $\Lambda_{\mathbf{F}}^J = J_{\mathcal{B}}$, where $J_{\mathcal{B}}$ is some signature operator on $\mathcal{B}$. We call $\mathcal{H}$ *regular* if the J-Gram operator of any strong basis is boundedly invertible. Since strong bases are related by invertible diagonal transformations R: $\mathbf{F}' = R\mathbf{F}$, the invertibility properties of the Gram operators of all these bases are the same, so that regularity is a

property of the subspace. Note that $\Lambda_{\mathbf{F}} \gg 0$ does not imply that $\Lambda_{\mathbf{F}}^J$ is boundedly invertible. But it is readily verified that the reverse implication is true: $\Lambda_{\mathbf{F}}^J$ boundedly invertible $\Rightarrow$ $\Lambda_{\mathbf{F}} \gg 0$, so that in particular a J-orthonormal basis is a strong basis.

If $\Lambda_{\mathbf{F}}^J$ is boundedly invertible, then it has a factorization into $\Lambda_{\mathbf{F}}^J = RJ_{\mathcal{B}}R^*$, where R and $J_{\mathcal{B}}$ are diagonals in $\mathcal{D}(\mathcal{B}, \mathcal{B})$, and $J_{\mathcal{B}}$ is the signature matrix of $\Lambda_{\mathbf{F}}^J$: it is a diagonal of matrices

$$(J_{\mathcal{B}})_k = \begin{bmatrix} I & \\ & -I \end{bmatrix}$$

and defines a partitioning of $\mathcal{B}$ into $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_-$. $J_{\mathcal{B}}$ is again independent of the choice of basis in $\mathcal{H}$. We call $J_{\mathcal{B}}$ the *inertia signature matrix* of the subspace $\mathcal{H}$, and the sequences $\#(\mathcal{B}_+)$ and $\#(\mathcal{B}_-)$ corresponding to the number of positive and negative entries of $J_{\mathcal{B}}$ at each point is called the inertia of $\mathcal{H}$. More general subspaces can also have a zero-inertia, corresponding to singularities of $\Lambda_{\mathbf{F}}^J$, but if $\mathcal{H}$ is regular, then it has no zero-inertia. (The zero-inertia is only well defined if the range of $\Lambda_{\mathbf{F}}^J$ is closed, or equivalently, if any of its eigenvalues is either equal to zero or uniformly bounded away from zero.)

The following theorem is proved in [14, thm. 1.7.16] for classical Krein spaces. The proof is easier for locally finite subspaces, and as these are the only ones that we will consider later, we restrict the theorem to this case.

THEOREM 5.5. *Let $\mathcal{H}$ be a locally finite left $\mathcal{D}$-invariant subspace in $\mathcal{X}_2$. The following are equivalent:*

1. *$\mathcal{H}$ is projectively complete; $\mathcal{H} \boxplus \mathcal{H}^{[\perp]} = \mathcal{X}_2$,*
2. *$\mathcal{H}$ is regular,*
3. *$\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$, where $\mathcal{H}_+$ and $\mathcal{H}_-$ are uniformly positive (resp. negative) subspaces,*
4. *Any element in $\mathcal{X}_2$ has at least one J-orthogonal projection onto $\mathcal{H}$.*

PROOF (1) $\Rightarrow$ (2). Let $\mathcal{H} = \mathcal{D}_2^{\mathcal{B}}\mathbf{F}$, where $\mathbf{F}$ is a strong basis representation with Gram operator $\Lambda_{\mathbf{F}} = \{\mathbf{F}, \mathbf{F}\} = \mathbf{P}_0(\mathbf{F}\mathbf{F}^*) \gg 0$. Assume $\mathcal{H} \boxplus \mathcal{H}^{[\perp]} = \mathcal{X}_2$. Then there exists a projection operator $\mathbf{P}_{\mathcal{H}}^J$ onto $\mathcal{H}$. Let $x \in \mathcal{H}$, $y = xJ$, then $y = y_0 + y_1$ where $y_0 = \mathbf{P}_{\mathcal{H}}^J(y) \in \mathcal{H}$ and $y_1 \in \mathcal{H}^{[\perp]}$. We have

$$\{x, x\} = [x, y] = [x, y_0] = \{xJ, y_0\} = \{\mathbf{P}_{\mathcal{H}}(xJ), y_0\}.$$

The projection operator $\mathbf{P}_{\mathcal{H}}(\cdot)$ follows from theorem 2.10 as $\mathbf{P}_{\mathcal{H}}(\cdot) = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\Lambda_{\mathbf{F}}^{-1}\mathbf{F}$. Writing $x = D\mathbf{F}$, $y_0 = D_0\mathbf{F}$, where $D, D_0 \in \mathcal{D}$, we obtain $\mathbf{P}_{\mathcal{H}}(xJ) = D\mathbf{P}_0(\mathbf{F}J\mathbf{F}^*)\Lambda_{\mathbf{F}}^{-1}\mathbf{F} = D\Lambda_{\mathbf{F}}^J\Lambda_{\mathbf{F}}^{-1}\mathbf{F}$, and

$$D\Lambda_{\mathbf{F}}D^* = \{D\Lambda_{\mathbf{F}}^J\Lambda_{\mathbf{F}}^{-1}\mathbf{F}, D_0\mathbf{F}\} = D\Lambda_{\mathbf{F}}^J D_0^*.$$

As this is true for all $D \in \mathcal{D}$ and $\Lambda_{\mathbf{F}} \gg 0$, and $D\Lambda_{\mathbf{F}}^J D_0^* = D_0\Lambda_{\mathbf{F}}^J D^*$, it follows that $\Lambda_{\mathbf{F}}^J$ must be boundedly invertible.

(2) $\Rightarrow$ (3). Let $\Lambda_{\mathbf{F}}^J = \mathbf{P}_0(\mathbf{F}J\mathbf{F}^*)$ be the J-Gramian of $\mathbf{F}$, which by the present assumptions is boundedly invertible. Let $J_{\mathcal{B}}$ be the signature matrix of $\Lambda_{\mathbf{F}}^J$: $\Lambda_{\mathbf{F}}^J = RJ_{\mathcal{B}}R^*$. $J_{\mathcal{B}}$ generates a decomposition of the state space sequence $\mathcal{B}$ into $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_-$. Let $\mathbf{Q} = R^{-1}\mathbf{F}$, then $\Lambda_{\mathbf{Q}}^J = J_{\mathcal{B}}$ so that $\mathbf{Q}$ is a J-orthonormal basis representation of $\mathcal{H}$. $\mathbf{Q}$ splits into a basis representation $\mathbf{Q}_+$ corresponding to the positive entries of $J_{\mathcal{B}}$, and a basis $\mathbf{Q}_-$ that is J-orthogonal to it and that corresponds to the negative entries. Hence $\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$, where $\mathcal{H}_+ = \mathcal{D}_2^{\mathcal{B}_+}\mathbf{Q}_+$ is a uniformly positive subspace and $\mathcal{H}_- = \mathcal{D}_2^{\mathcal{B}_-}\mathbf{Q}_-$ is J-uniformly negative.

(3) $\Rightarrow$ (4). Let $\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$, and let $\mathbf{F}_+$, $\mathbf{F}_-$ be strong basis representations of $\mathcal{H}_+$ and $\mathcal{H}_-$. Because the subspaces $\mathcal{H}_+, \mathcal{H}_-$ are uniformly definite, $\Lambda_{\mathbf{F}_+}^J$ and $\Lambda_{\mathbf{F}_-}^J$ are boundedly invertible, and J-orthonormal projectors onto $\mathcal{H}_+$ and $\mathcal{H}_-$ are

$$\begin{aligned} \mathbf{P}_{\mathcal{H}_+}^J &= \mathbf{P}_0(\cdot\, J\mathbf{F}_+^*)(\Lambda_{\mathbf{F}_+}^J)^{-1}\mathbf{F}_+, \\ \mathbf{P}_{\mathcal{H}_-}^J &= \mathbf{P}_0(\cdot\, J\mathbf{F}_-^*)(\Lambda_{\mathbf{F}_-}^J)^{-1}\mathbf{F}_-. \end{aligned}$$

Hence $y \in \mathcal{X}_2$ has well-defined projections y_+ onto $\mathcal{H}_+$ and y_- onto $\mathcal{H}_-$. Then $\mathbf{P}_{\mathcal{H}}^J(y) = y_+ + y_-$.

(4) $\Rightarrow$ (1). Since any $y \in \mathcal{X}_2$ has a J-orthogonal projection y_0 onto $\mathcal{H}$: $y = y_0 + y_1$ (where $y_0 \in \mathcal{H}, y_1 \in \mathcal{H}^{[\perp]}$), we must have that $\mathcal{H} + \mathcal{H}^{[\perp]} = \mathcal{X}_2$, that is, $\mathcal{H}$ is projectively complete. □

COROLLARY 5.6. *Let $\mathcal{H}$ be a locally finite regular left D-invariant subspace in $\mathcal{X}_2$ with dimension sequence $\mathcal{B}$, and let $J_{\mathcal{B}}$ be the inertia signature matrix of $\mathcal{H}$. Then $\mathcal{H}$ has a canonical decomposition $\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$ into uniformly definite subspaces, where* s-dim $\mathcal{H}_+ = \#\mathcal{B}_+ = \#_+(J_{\mathcal{B}})$ *and* s-dim $\mathcal{H}_- = \#\mathcal{B}_- = \#_-(J_{\mathcal{B}})$.

State space properties of J-unitary operators

We show that a bounded, block-upper J-unitary operator Θ has input and output state spaces $\mathcal{H}(\Theta)$ and $\mathcal{H}_0(\Theta)$ that are closed, regular subspaces, so that *e.g.*, $\mathcal{H}(\Theta) = \mathcal{H}_+ \boxplus \mathcal{H}_-$ where $\mathcal{H}_+$ and $\mathcal{H}_-$ are uniformly definite.

PROPOSITION 5.7. *Let $\Theta \in \mathcal{X}(\mathcal{M}, \mathcal{N})$ be a $(J_{\mathcal{M}}, J_{\mathcal{N}})$-unitary operator. Then*

$$\begin{aligned} \mathcal{K}(\Theta) &= \mathcal{L}_2\mathcal{Z}^{-1}\,\Theta^* J_{\mathcal{M}}, & \overline{\mathcal{H}}(\Theta) &= \mathcal{L}_2\mathcal{Z}^{-1} \ominus \mathcal{L}_2\mathcal{Z}^{-1}\,\Theta^* J_{\mathcal{M}} \\ \mathcal{K}_0(\Theta) &= \mathcal{U}_2\Theta J_{\mathcal{N}}, & \overline{\mathcal{H}}_0(\Theta) &= \mathcal{U}_2 \ominus \mathcal{U}_2\Theta J_{\mathcal{N}}. \end{aligned}$$

PROOF Let $x \in \mathcal{K}(\Theta) = \{x \in \mathcal{L}_2\mathcal{Z}^{-1} : \mathbf{P}(x\Theta) = 0\}$. Since $\Theta^{-1} = J\Theta^* J \in \mathcal{L}$, it follows that $x \in \mathcal{K}(\Theta)$ if and only if $x \in \mathcal{L}_2\mathcal{Z}^{-1}\Theta^{-1}$. Hence $\mathcal{K}(\Theta) = \mathcal{L}_2\mathcal{Z}^{-1}\Theta^{-1} = \mathcal{L}_2\mathcal{Z}^{-1}J\Theta^* J = \mathcal{L}_2\mathcal{Z}^{-1}\Theta^* J$, and $\overline{\mathcal{H}}(\Theta) = \mathcal{L}_2\mathcal{Z}^{-1} \ominus \mathcal{L}_2\mathcal{Z}^{-1}\,\Theta^* J$.

The results on $\mathcal{K}_0$ and $\mathcal{H}_0$ follow in a dual way. □

PROPOSITION 5.8. *In proposition 5.7, $\mathcal{H}$ and $\mathcal{H}_0$ are closed subspaces, and*

$$\begin{aligned}\mathcal{H}_0(\Theta) &= \mathcal{H}(\Theta)J_{\mathcal{M}}\Theta \\ \mathcal{H}(\Theta) &= \mathcal{H}_0(\Theta)J_{\mathcal{N}}\Theta^*.\end{aligned}$$

PROOF Let $x \in \overline{\mathcal{H}}(\Theta)$. Then

$$\{xJ\Theta, \mathcal{L}_2\mathcal{Z}^{-1}\} = \{x, \mathcal{L}_2\mathcal{Z}^{-1}\Theta^*J\} = \{x, \mathcal{K}(\Theta)\} = 0.$$

Hence $(xJ)\Theta \in \mathcal{U}_2$ so that $\overline{\mathcal{H}}J\Theta \subset \mathcal{H}_0$ and $\overline{\mathcal{H}} \subset \mathcal{H}_0J\Theta^*$. In a dual way, it follows that $\overline{\mathcal{H}}_0J\Theta^* \subset \mathcal{H}$. Combining these two expressions gives

$$\overline{\mathcal{H}}_0J\Theta^* \subset \mathcal{H} \subset \overline{\mathcal{H}} \subset \mathcal{H}_0J\Theta^*,$$

and because $J\Theta^*$ is boundedly invertible, $\mathcal{H}_0(\Theta)$ and $\mathcal{H}(\Theta)$ are closed. In addition, it follows that $\mathcal{H}J\Theta = \mathcal{H}_0$ and $\mathcal{H}_0J\Theta^* = \mathcal{H}$. □

PROPOSITION 5.9. *In proposition 5.7, $\mathcal{H}(\Theta)$ and $\mathcal{H}_0(\Theta)$ are regular;*

$$\begin{aligned}\mathcal{L}_2\mathcal{Z}^{-1} &= \mathcal{H} \boxplus \mathcal{L}_2\mathcal{Z}^{-1}\Theta^* \\ \mathcal{U}_2 &= \mathcal{H}_0 \boxplus \mathcal{U}_2\Theta.\end{aligned}$$

PROOF $\mathcal{H}_0^{[\perp]} = \mathcal{H}_0^{\perp}J = \mathcal{K}_0J = \mathcal{U}_2\Theta$. To prove that $\mathcal{U}_2 = \mathcal{H}_0 \boxplus \mathcal{H}_0^{[\perp]}$, we show that every $y \in \mathcal{U}_2$ has a J-orthogonal projection onto $\mathcal{H}_0$. Let $y \in \mathcal{U}_2$, and define $y\Theta^{-1} = u_1 + y_1$, with $u_1 \in \mathcal{L}_2\mathcal{Z}^{-1}$ and $y_1 \in \mathcal{U}_2$. Further, define $u_2 = u_1J \in \mathcal{L}_2\mathcal{Z}^{-1}$. Then $u_2 \in \mathcal{H}$ because $u_2 = \mathbf{P}_{\mathcal{L}_2\mathcal{Z}^{-1}}(y\Theta^{-1})J = \mathbf{P}_{\mathcal{L}_2\mathcal{Z}^{-1}}\left((yJ)\Theta^*\right)$. It follows that $y = u_2J\Theta + y_1\Theta$, where $u_2J\Theta \in \mathcal{H}_0$ (because of proposition 5.8) and $y_1\Theta \in \mathcal{H}_0^{[\perp]} = \mathcal{U}_2\Theta$. Hence every $y \in \mathcal{U}_2$ has a J-projection onto $\mathcal{H}_0$, so that according to theorem 5.5 $\mathcal{H}_0$ is regular. A dual proof holds for $\mathcal{H}$. □

COROLLARY 5.10. *Let $\Theta \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a locally finite J-unitary operator. If $\mathbf{F}$ is a $J_{\mathcal{B}}$-orthonormal basis representation of $\mathcal{H}(\Theta)$, then $\mathbf{F}_0 = J_{\mathcal{B}}\mathbf{F}J_{\mathcal{M}}\Theta$ is a J-orthonormal basis representation of $\mathcal{H}_0(\Theta)$, and in this case the canonical controller realization based on $\mathbf{F}$ (theorem 3.20) and canonical observer realization based on $\mathbf{F}_0$ (theorem 3.25) are equal.*

PROOF Because $\mathcal{H}(\Theta)$ is regular, there is a J-orthonormal basis representation $\mathbf{F}$ of $\mathcal{H}$: $\Lambda_{\mathbf{F}}^J = \mathbf{P}_0(\mathbf{F}J\mathbf{F}^*) = J_{\mathcal{B}}$. This $\mathbf{F}$ defines a factorization of the Hankel operator of Θ as $H_\Theta = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0$ where $\mathbf{F}_0 = \Lambda_{\mathbf{F}}^{-1}\mathbf{P}(\mathbf{F}\Theta)$ is a basis of the output state space $\mathcal{H}_0$ of Θ. (theorem 3.27). On the other hand, the relation $\mathcal{H}_0 = \mathcal{H}J\Theta$ ensures that $\mathbf{F}_1$, defined as $\mathbf{F}_1 = \mathbf{F}J_{\mathcal{M}}\Theta$, is upper and also a J-orthonormal basis representation of $\mathcal{H}_0$. The

connection between $\mathbf{F}_1$ and $\mathbf{F}_0$ is $\mathbf{F}_1 = \mathbf{F}J_{\mathcal{M}}\Theta = \mathbf{F}J_{\mathcal{M}}H_\Theta = \mathbf{P}_0(\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*)\,\mathbf{F}_0 = J_{\mathcal{B}}\mathbf{F}_0$, so that $\mathbf{F}_0 = J_{\mathcal{B}}\mathbf{F}_1 = J_{\mathcal{B}}\mathbf{F}J_{\mathcal{M}}\Theta$. It is readily verified that $\mathbf{F}_0$ is also J-orthonormal. Theorem 3.27 claimed that the canonical observer realization based on this $\mathbf{F}_0$ is equal to the canonical controller realization of $\mathbf{F}$. □

J-Unitary realizations

A realization matrix $\boldsymbol{\Theta} \in \mathcal{U}(\mathcal{B}\times\mathcal{M}, \mathcal{B}^{(-1)}\times\mathcal{N})$ with signature matrices $\mathbf{J}_1$, $\mathbf{J}_2$,

$$\boldsymbol{\Theta} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}, \qquad \mathbf{J}_1 = \begin{bmatrix} J_{\mathcal{B}} & \\ & J_{\mathcal{M}} \end{bmatrix}, \quad \mathbf{J}_2 = \begin{bmatrix} J_{\mathcal{B}}^{(-1)} & \\ & J_{\mathcal{N}} \end{bmatrix} \tag{5.9}$$

is said to be J-unitary if

$$\boldsymbol{\Theta}^*\mathbf{J}_1\boldsymbol{\Theta} = \mathbf{J}_2, \qquad \boldsymbol{\Theta}\mathbf{J}_2\boldsymbol{\Theta}^* = \mathbf{J}_1.$$

We call $J_{\mathcal{B}}$ the state signature matrix of $\boldsymbol{\Theta}$. With '#' indicating the sequence of dimensions of a space sequence, we have that, for each time instant k, the total number of positive entries of the signatures at the left-hand side of each equation is equal to the total positive signature at the right-hand side, and similarly for the total negative signature (the *inertia theorem*):

$$\begin{aligned} \#\mathcal{B}_+ + \#\mathcal{M}_+ &= \#\mathcal{B}_+^{(-1)} + \#\mathcal{N}_+ \\ \#\mathcal{B}_- + \#\mathcal{M}_- &= \#\mathcal{B}_-^{(-1)} + \#\mathcal{N}_-. \end{aligned} \tag{5.10}$$

As for inner systems, J-unitary systems and J-unitary realizations go together. Proofs of this are similar to the unitary case (theorems 4.5 and 4.6).

THEOREM 5.11. *Let $\Theta \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a locally finite $(J_{\mathcal{M}}, J_{\mathcal{N}})$-unitary operator. Let $J_{\mathcal{B}}$ be the inertia signature matrix of $\mathcal{H}(\Theta)$, and let $\mathbf{F}$ be a J-orthonormal basis representation for $\mathcal{H}(\Theta)$.*

Then the canonical controller realization $\boldsymbol{\Theta}$ based on $\mathbf{F}$ is J-unitary, and identical to the canonical observer realization based on $\mathbf{F}_0 = J_{\mathcal{B}}\mathbf{F}J_{\mathcal{M}}\Theta$.

PROOF Let $\boldsymbol{\Theta}$ be given by the canonical controller realization (theorem 3.20). This realization satisfies the properties (lemma 3.21):

$$\begin{cases} Z\mathbf{F} &= A^*\mathbf{F} + B^*, \\ \mathbf{P}_0(Z^{-1}\cdot\mathbf{F}^*)^{(-1)} &= \mathbf{P}_0(\cdot\,[\mathbf{F}^*A + B]) \\ \mathbf{P}_0(\cdot\,\Theta) &= \mathbf{P}_0(\cdot\,[D + \mathbf{F}^*C]), \\ \Theta^* &= D^* + C^*\mathbf{F}. \end{cases} \tag{5.11}$$

To verify that $\boldsymbol{\Theta}^*\mathbf{J}_1\boldsymbol{\Theta} = \mathbf{J}_2$, we have to show that

$$\begin{cases} \Theta^*J_{\mathcal{M}}\Theta &= J_{\mathcal{N}} \\ \mathbf{P}_0(\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*) &= J_{\mathcal{B}} \end{cases} \quad \Rightarrow \quad \begin{cases} A^*J_{\mathcal{B}}A + B^*J_{\mathcal{M}}B &= J_{\mathcal{B}}^{(-1)} \\ C^*J_{\mathcal{B}}C + D^*J_{\mathcal{M}}D &= J_{\mathcal{N}} \\ A^*J_{\mathcal{B}}C + B^*J_{\mathcal{M}}D &= 0. \end{cases}$$

Indeed,

$$\begin{aligned}
\mathbf{P}_0(\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*) = J_{\mathcal{B}} \quad \Rightarrow \quad J_{\mathcal{B}}^{(-1)} &= \mathbf{P}_0(Z^{-1}(Z\mathbf{F})J_{\mathcal{M}}\mathbf{F}^*)^{(-1)} \\
&= \mathbf{P}_0([A^*\mathbf{F}+B^*]J_{\mathcal{M}}[\mathbf{F}^*A+B]) \\
&= A^*J_{\mathcal{B}}A + B^*J_{\mathcal{M}}B,
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_0(\Theta^*J_{\mathcal{M}}\Theta) = J_{\mathcal{N}} \quad \Rightarrow \quad & \mathbf{P}_0([D^*+C^*\mathbf{F}]J_{\mathcal{M}}[D+\mathbf{F}^*C]) \\
& = D^*J_{\mathcal{M}}D + C^*\mathbf{P}_0(\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*)C \\
& = D^*J_{\mathcal{M}}D + C^*J_{\mathcal{B}}C \quad = J_{\mathcal{N}}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{P}_0(Z\mathbf{F}J\Theta) = 0 \quad \Rightarrow \quad & \mathbf{P}_0([B^*+A^*\mathbf{F}]J_{\mathcal{M}}[D+\mathbf{F}^*C]) \\
& = B^*J_{\mathcal{M}}D + A^*\mathbf{P}_0(\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*)C \\
& = B^*J_{\mathcal{M}}D + A^*J_{\mathcal{B}}C \quad = 0.
\end{aligned}$$

Hence $\mathbf{\Theta}^*\mathbf{J}_1\mathbf{\Theta} = \mathbf{J}_2$. The relation $\mathbf{\Theta}\mathbf{J}_2\mathbf{\Theta}^* = \mathbf{J}_1$ can be derived in the same (dual) way as above. The equality of both realizations has been proven in corollary 5.10. □

The converse of this theorem is in general true only if, in addition, $\ell_A < 1$. If $\ell_A = 1$, then additional assumptions on the controllability and observability of the realization must be made.

THEOREM 5.12. *Let* $\mathbf{\Theta} = \begin{bmatrix} A & C \\ B & D \end{bmatrix}$ *be a state realization of a bounded transfer operator* Θ. *Let* $\Lambda_{\mathbf{F}}^J$ *and* $\Lambda_{\mathbf{F}_0}^J$ *be the controllability and the observability J-Gramians of the given realization. If* $\ell_A < 1$, *then*

$$\begin{aligned}
&\mathbf{\Theta}^*\mathbf{J}_1\mathbf{\Theta} = \mathbf{J}_2 \quad \Rightarrow \quad \Theta^*J_{\mathcal{M}}\Theta = J_{\mathcal{N}}, \quad \Lambda_{\mathbf{F}}^J = J_{\mathcal{B}}, \\
&\mathbf{\Theta}\mathbf{J}_2\mathbf{\Theta}^* = \mathbf{J}_1 \quad \Rightarrow \quad \Theta J_{\mathcal{N}}\Theta^* = \mathcal{M}, \quad \Lambda_{\mathbf{F}_0}^J = J_{\mathcal{B}}.
\end{aligned} \tag{5.12}$$

If $\ell_A \leq 1$, *then*

$$\begin{aligned}
&\mathbf{\Theta}^*\mathbf{J}_1\mathbf{\Theta} = \mathbf{J}_2, \quad \Lambda_{\mathbf{F}}^J = J_{\mathcal{B}} \quad \Rightarrow \quad \Theta^*J_{\mathcal{M}}\Theta = J_{\mathcal{N}}, \\
&\mathbf{\Theta}\mathbf{J}_2\mathbf{\Theta}^* = \mathbf{J}_1, \quad \Lambda_{\mathbf{F}_0}^J = J_{\mathcal{B}} \quad \Rightarrow \quad \Theta J_{\mathcal{N}}\Theta^* = J_{\mathcal{M}}.
\end{aligned}$$

PROOF If $\ell_A < 1$, then $\mathbf{\Theta}^*\mathbf{J}_1\mathbf{\Theta} = \mathbf{J}_2$ implies $A^*J_{\mathcal{B}}A + B^*J_{\mathcal{M}}B = J_{\mathcal{B}}^{(-1)}$. The J-Lyapunov equation for the realization based on $\mathbf{F}$ is determined from proposition 3.9 to be $A^*\Lambda_{\mathbf{F}}^J A + B^*JB = (\Lambda_{\mathbf{F}}^J)^{(-1)}$. Since $\ell_A < 1$, the equation has a unique solution, so that $\Lambda_{\mathbf{F}}^J = J_{\mathcal{B}}$, and $\mathbf{F}$ is J-orthonormal. A dual result holds for $\Lambda_{\mathbf{F}_0}^J$ in case $\mathbf{\Theta}\mathbf{J}_2\mathbf{\Theta}^* = \mathbf{J}_1$.

Assume $\mathbf{\Theta}^*\mathbf{J}_1\mathbf{\Theta} = \mathbf{J}_2$ and $\Lambda_{\mathbf{F}}^J = J_{\mathcal{B}}$. We use the equations (*viz.* lemma 3.21)

$$\begin{aligned}
\mathbf{P}_0(\cdot\,\Theta) &= \mathbf{P}_0(\cdot\,[D+\mathbf{F}^*C]) \\
\Theta^* &= D^* + C^*\mathbf{F}.
\end{aligned}$$

$$\begin{aligned} \mathbf{P}_0(Z^{-n}\mathbf{F}J\mathbf{F}^*) &= J_{\mathcal{B}}^{(n)}A^{\{n\}} \qquad (n \geq 0) \\ \mathbf{P}_0(Z^{-n}\mathbf{F}^*) &= B^{(n)}A^{\{n-1\}} \qquad (n > 0). \end{aligned}$$

To show $\Theta^* J_{\mathcal{M}} \Theta = J_{\mathcal{N}}$, we show that $\mathbf{P}_0(Z^{-n}\Theta^* J\Theta)$ is equal to J for $n = 0$, and equal to zero otherwise. For $n = 0$:

$$\begin{aligned} \mathbf{P}_0(\Theta^* J_{\mathcal{M}}\Theta) &= \mathbf{P}_0(\,[D^* + C^*\mathbf{F}]\,J_{\mathcal{M}}\,[D + \mathbf{F}^*C]\,) \\ &= \mathbf{P}_0(D^*J_{\mathcal{M}}D) + \mathbf{P}_0(D^*J_{\mathcal{M}}\mathbf{F}^*C) + \mathbf{P}_0(C^*\mathbf{F}J_{\mathcal{M}}D) + \mathbf{P}_0(C^*\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*C) \\ &= D^*J_{\mathcal{M}}D + C^*J_{\mathcal{B}}C = J_{\mathcal{N}}. \end{aligned}$$

For $n > 0$,

$$\begin{aligned} \mathbf{P}_0(Z^{-n}\Theta^* J_{\mathcal{M}}\Theta) &= \mathbf{P}_0(\,Z^{-n}[D^* + C^*\mathbf{F}]\,J_{\mathcal{M}}\,[D + \mathbf{F}^*C]\,) \\ &= \mathbf{P}_0(Z^{-n}D^*J_{\mathcal{M}}D) + \mathbf{P}_0(Z^{-n}D^*J_{\mathcal{M}}\mathbf{F}^*C) + \\ &\quad + \mathbf{P}_0(Z^{-n}C^*\mathbf{F}J_{\mathcal{M}}D) + \mathbf{P}_0(Z^{-n}C^*\mathbf{F}J_{\mathcal{M}}\mathbf{F}^*C) \\ &= 0 + 0 + D^{*(n)}J_{\mathcal{M}}^{(n)}B^{(n)}A^{\{n-1\}}C + C^{*(n)}J_{\mathcal{B}}^{(n)}A^{\{n\}}C \\ &= [D^*J_{\mathcal{M}}B + C^*J_{\mathcal{B}}A]^{(n)}A^{\{n-1\}}C \\ &= 0. \end{aligned}$$

Taking adjoints shows that $\mathbf{P}_0(Z^{-n}\Theta^* J\Theta) = 0$ for $n < 0$, too. Hence $\Theta^* J\Theta = J$.
The fact $[\boldsymbol{\Theta}\mathbf{J}_2\boldsymbol{\Theta}^* = \mathbf{J}_2,\ \Lambda^J_{\mathbf{F}_0} = J_{\mathcal{B}}] \Rightarrow \Theta J\Theta^* = J$ can be shown in a dual way. □

Again, if $\ell_A < 1$, then a more elementary computation suffices to verify the theorem: evaluating $J_{\mathcal{N}} - \Theta^* J_{\mathcal{M}}\Theta$ and $J_{\mathcal{M}} - \Theta J_{\mathcal{N}}\Theta^*$ yields, for the former expression,

$$\begin{aligned} J_{\mathcal{N}} - \Theta^* J_{\mathcal{M}}\Theta &= J_{\mathcal{N}} - [D + BZ(I - AZ)^{-1}C]^* J_{\mathcal{M}} [D + BZ(I - AZ)^{-1}C] \\ &= C^*J_{\mathcal{B}}C + C^*(I - Z^*A^*)^{-1}Z^*A^*J_{\mathcal{B}}C + C^*J_{\mathcal{B}}AZ(I - AZ)^{-1}C + \\ &\quad - C^*(I - Z^*A^*)^{-1}Z^*\left\{J_{\mathcal{B}}^{(-1)} - A^*J_{\mathcal{B}}A\right\}Z(I - AZ)^{-1}C \end{aligned}$$

since $B^*J_{\mathcal{M}}D = -A^*J_{\mathcal{B}}C$, $B^*J_{\mathcal{M}}B = J_{\mathcal{B}}^{(-1)} - A^*J_{\mathcal{B}}A$ and $J_{\mathcal{N}} - D^*J_{\mathcal{M}}D = C^*J_{\mathcal{B}}C$, and hence

$$\begin{aligned} J_{\mathcal{N}} - \Theta^* J_{\mathcal{M}}\Theta &= C^*(I - Z^*A^*)^{-1}\left\{(I - Z^*A^*)J_{\mathcal{B}}(I - AZ) + Z^*A^*J_{\mathcal{B}}(I - AZ) + \right. \\ &\quad \left. + (I - Z^*A^*)J_{\mathcal{B}}AZ - J_{\mathcal{B}} + Z^*A^*J_{\mathcal{B}}AZ\right\}(I - AZ)^{-1}C \\ &= 0. \end{aligned}$$

The second equality follows by an analogous procedure.

Unitary state representation for Σ

Let $\boldsymbol{\Theta}$ be a J-unitary realization of a J-unitary operator $\Theta \in \mathcal{U}$, with state signature matrix $J_{\mathcal{B}}$. We have seen (in proposition 5.9) that the input and output state spaces $\mathcal{H}(\Theta)$ and $\mathcal{H}_0(\Theta)$ are regular: $\mathcal{H} = \mathcal{H}_+ \boxplus \mathcal{H}_-$, and that this partitioning corresponds to the partitioning of the state space sequence $\mathcal{B}$ into $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_-$ conformably to $J_{\mathcal{B}}$. Because the bases chosen for the state spaces are J-orthonormal ($\boldsymbol{\Theta}$ is J-unitary), the basis representation

$\mathbf{F}$ is partitioned into two J-orthonormal bases $\mathbf{F}_+$ and $\mathbf{F}_-$, such that $\mathcal{H}_+ = \mathcal{D}_2^{\mathcal{B}_+}\mathbf{F}_+$ and $\mathcal{H}_- = \mathcal{D}_2^{\mathcal{B}_-}\mathbf{F}_-$. Hence a state $x \in \mathcal{X}_2^{\mathcal{B}}$ is partitioned into $x = [x_+ \quad x_-] \in \mathcal{X}_2^{\mathcal{B}_+} \times \mathcal{X}_2^{\mathcal{B}_-}$, where x_+ and x_- are the parts of the state that correspond to the positive and negative subspaces in the state space $\mathcal{H}$: $x_+\mathbf{F}_+ \in \mathcal{H}_+$ and $x_-\mathbf{F}_- \in \mathcal{H}_-$. The decomposition of the state defines a partitioning of $\boldsymbol{\Theta}$ according to the equation

$$[x_+ \quad x_- \quad a_1 \quad b_1]\boldsymbol{\Theta} = [x_+Z^{-1} \quad x_-Z^{-1} \quad a_2 \quad b_2]. \tag{5.13}$$

into

$$\boldsymbol{\Theta} = \begin{array}{c} \\ x_+ \\ x_- \\ a_1 \\ b_1 \end{array} \begin{array}{c} \begin{array}{cccc} x_+Z^{-1} & x_-Z^{-1} & a_2 & b_2 \end{array} \\ \left[\begin{array}{cc|cc} A_{11} & A_{12} & C_{11} & C_{12} \\ A_{21} & A_{22} & C_{21} & C_{22} \\ \hline B_{11} & B_{12} & D_{11} & D_{12} \\ B_{21} & B_{22} & D_{21} & D_{22} \end{array}\right] \end{array}. \tag{5.14}$$

We have shown, in theorem 5.2, that associated to Θ is a unitary operator Σ such that

$$[a_1 \quad b_2]\Sigma = [a_2 \quad b_1] \quad \Leftrightarrow \quad [a_1 \quad b_1]\Theta = [a_2 \quad b_2].$$

The question now is whether the given realization of Θ gives rise to a realization of Σ.

A reordering of rows and columns in (5.14) with respect to their signatures converts $\boldsymbol{\Theta}$ into a genuine square-block J-unitary operator, *i.e.*, each matrix

$$\left[\begin{array}{cc|cc} A_{11} & C_{11} & A_{12} & C_{12} \\ B_{11} & D_{11} & B_{12} & D_{12} \\ \hline A_{21} & C_{21} & A_{22} & C_{22} \\ B_{21} & D_{21} & B_{22} & D_{22} \end{array}\right]_k$$

is a square and J-unitary matrix with respect to the signature

$$\begin{bmatrix} I_{(\mathcal{B}_+)_k \times (\mathcal{M}_+)_k} & \\ & -I_{(\mathcal{B}_-)_k \times (\mathcal{M}_-)_k} \end{bmatrix} = \begin{bmatrix} I_{(\mathcal{B}_+)_{k+1} \times (\mathcal{N}_+)_k} & \\ & -I_{(\mathcal{B}_-)_{k+1} \times (\mathcal{N}_-)_k} \end{bmatrix}.$$

In particular, each submatrix

$$\begin{bmatrix} A_{22} & C_{22} \\ B_{22} & D_{22} \end{bmatrix}_k$$

of $\boldsymbol{\Theta}_k$ is square and invertible, and because $\boldsymbol{\Theta}$ is J-unitary, the block-diagonal operator constructed from these submatrices is boundedly invertible as well. It follows that the

following block-diagonal operators are well defined (*cf.* equation (5.4)):

$$\begin{aligned}
\begin{bmatrix} F_{11} & H_{11} \\ G_{11} & K_{11} \end{bmatrix} &= \begin{bmatrix} A_{11} & C_{11} \\ B_{11} & D_{11} \end{bmatrix} - \begin{bmatrix} A_{12} & C_{12} \\ B_{12} & D_{12} \end{bmatrix} \begin{bmatrix} A_{22} & C_{22} \\ B_{22} & D_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21} & C_{21} \\ B_{21} & D_{21} \end{bmatrix} \\
\begin{bmatrix} F_{12} & H_{12} \\ G_{12} & K_{12} \end{bmatrix} &= -\begin{bmatrix} A_{12} & C_{12} \\ B_{12} & D_{12} \end{bmatrix} \begin{bmatrix} A_{22} & C_{22} \\ B_{22} & D_{22} \end{bmatrix}^{-1} \\
\begin{bmatrix} F_{21} & H_{21} \\ G_{21} & K_{21} \end{bmatrix} &= \begin{bmatrix} A_{22} & C_{22} \\ B_{22} & D_{22} \end{bmatrix}^{-1} \begin{bmatrix} A_{21} & C_{21} \\ B_{21} & D_{21} \end{bmatrix} \\
\begin{bmatrix} F_{22} & H_{22} \\ G_{22} & K_{22} \end{bmatrix} &= \begin{bmatrix} A_{22} & C_{22} \\ B_{22} & D_{22} \end{bmatrix}^{-1}
\end{aligned} \tag{5.15}$$

and we obtain the relation

$$[x_+ \quad x_- Z^{-1} \quad a_1 \quad b_2]\,\boldsymbol{\Sigma} = [x_+ Z^{-1} \quad x_- \quad a_2 \quad b_1] \tag{5.16}$$

where

$$\boldsymbol{\Sigma} = \begin{array}{c} \\ x_+ \\ x_- Z^{-1} \\ a_1 \\ b_2 \end{array} \begin{array}{c} \begin{array}{cc} x_+ Z^{-1} & x_- \end{array} \\ \left[\begin{array}{cc|cc} F_{11} & F_{12} & H_{11} & H_{12} \\ F_{21} & F_{22} & H_{21} & H_{22} \\ \hline G_{11} & G_{12} & K_{11} & K_{12} \\ G_{21} & G_{22} & K_{21} & K_{22} \end{array}\right] \end{array} \begin{array}{c} \\ \\ \\ \\ \\ \end{array} \begin{array}{cc} a_2 & b_1 \end{array} . \tag{5.17}$$

See figure 5.2. An important point which can be readily derived from the J-unitarity of $\boldsymbol{\Theta}$ is fact that $\boldsymbol{\Sigma}$ is unitary:

$$\boldsymbol{\Sigma}\boldsymbol{\Sigma}^* = I; \quad \boldsymbol{\Sigma}^*\boldsymbol{\Sigma} = I.$$

Because in (5.16) state quantities with and without Z^{-1} appear in the same argument at the left- and right-hand sides, $\boldsymbol{\Sigma}$ is a kind of generalized or implicit realization for the transfer operator Σ, but is not computable in this form. Σ can be obtained from $\boldsymbol{\Sigma}$ by elimination of x_- and x_+. $\boldsymbol{\Sigma}$ can be interpreted as a realization having an "upward" state x_- and a downward state x_+, as depicted in figure 5.2. Recall that although Σ is unitary, it need not be upper. It can be contemplated that the upward state x_- is instrumental in generating the lower triangular (anti-causal) part of Σ. The precise details will be investigated later (proposition 5.16), but a preliminary result is straightforward to derive.

PROPOSITION 5.13. *Let $\boldsymbol{\Theta}$ be a $(\mathbf{J}_1, \mathbf{J}_2)$-unitary realization for a J-unitary operator Θ. If $J_B = I$, then $\Theta_{22}^{-1} \in \mathcal{U}$, that is, Θ is J-inner and the corresponding unitary operator Σ is upper and hence inner.*

PROOF If $J_B = I$, then the dimension of x_- is zero, so that the implicit state relations $\boldsymbol{\Sigma}$ for Σ in (5.16) are reduced to ordinary state equations $[x_+ Z^{-1} \quad a_2 \quad b_1] = [x_+ \quad a_1 \quad b_2]\boldsymbol{\Sigma}$, which define an upper (causal) operator Σ. □

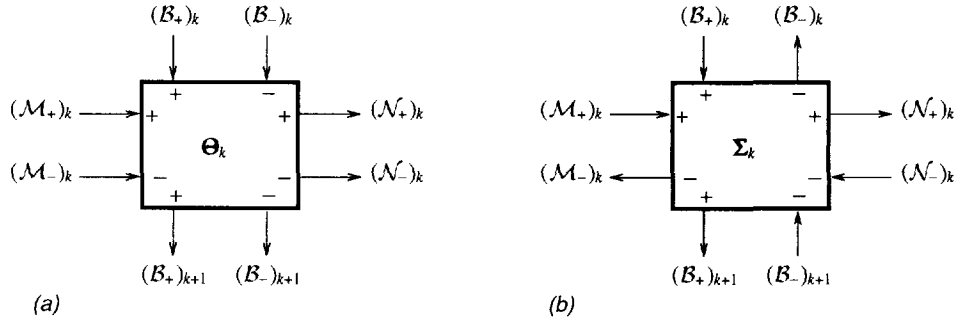


Figure 5.2. (*a*) The spaces connected with a realization for a *J*-unitary block-upper operator Θ which transfers $\ell_2^{\mathcal{M}_+} \times \ell_2^{\mathcal{M}_-}$ to $\ell_2^{\mathcal{N}_+} \times \ell_2^{\mathcal{N}_-}$. The state transition operator is marked as $\boldsymbol{\Theta}$. (*b*) The corresponding scattering situation.

Past and future scattering operators

In section 3.3, we defined for a signal $u \in \mathcal{X}_2$ the decomposition $u = u_p + u_f$, where $u_p = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(u) \in \mathcal{L}_2 Z^{-1}$ is the 'past' part of the signal (with reference to its 0-th diagonal), and $u_f = \mathbf{P}(u) \in \mathcal{U}_2$ is its 'future' part. We also showed how a causal operator T with state realization $\mathbf{T}$ could be split into a past operator T_p which maps u_p to $[x_{[0]} \;\; y_p]$ and a future operator T_f which maps $[x_{[0]} \;\; u_f]$ to y_f. In the present context, let the signals a_1, b_1, a_2, b_2 and the state sequences x_+, x_- be in $\mathcal{X}_2$ and be related by $\boldsymbol{\Theta}$ as in (5.13). With the partitioning of the signals a_1, etc., into a past and a future part, Θ can be split into operators $(\cdot)\Theta_p : Z^{-1}\mathcal{L}_2^{\mathcal{M}} \to [\mathcal{D}_2^{\mathcal{B}} \;\; Z^{-1}\mathcal{L}_2^{\mathcal{N}}]$ and $(\cdot)\Theta_f : [\mathcal{D}_2^{\mathcal{B}} \;\; \mathcal{U}_2^{\mathcal{M}}] \to \mathcal{U}_2^{\mathcal{N}}$ via

$$\begin{aligned} [a_{1p} \;\; b_{1p}]\,\Theta_p &= [x_{+[0]} \;\; x_{-[0]} \;\; a_{2p} \;\; b_{2p}] \\ [x_{+[0]} \;\; x_{-[0]} \;\; a_{1f} \;\; b_{1f}]\,\Theta_f &= [a_{2f} \;\; b_{2f}]. \end{aligned} \tag{5.18}$$

Θ_p and Θ_f can be determined once basis representations for the input and output state spaces of Θ have been chosen. The following procedure is as in section 3.3. The splitting of signals into past and future parts associates to Θ an 'expanded' operator $\hat{\Theta}$, defined such that $(u_p + u_f)\Theta = (y_p + y_f) \Leftrightarrow [u_p \;\; u_f]\hat{\Theta} = [y_p \;\; y_f]$:

$$\hat{\Theta} = \begin{bmatrix} K_\Theta & H_\Theta \\ 0 & E_\Theta \end{bmatrix} \qquad \text{where} \qquad \begin{cases} K_\Theta &= \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\cdot\,\Theta)\big|_{\mathcal{L}_2 Z^{-1}} \\ H_\Theta &= \mathbf{P}(\cdot\,\Theta)\big|_{\mathcal{L}_2 Z^{-1}} \\ E_\Theta &= \mathbf{P}(\cdot\,\Theta)\big|_{\mathcal{U}_2} \end{cases} \tag{5.19}$$

Let $\mathbf{F}$ be a *J*-orthonormal basis in $\mathcal{H}(\Theta)$, and let $\mathbf{F}_0 = J\mathbf{F}J\Theta$ be the corresponding *J*-orthonormal basis in $\mathcal{H}_0(\Theta)$. Then Θ_p and Θ_f are given by (*cf.* equation (3.36))

$$\Theta_p = [\mathbf{P}_0(\cdot\,\mathbf{F}^*) \;\; K_\Theta], \qquad \Theta_f = \begin{bmatrix} \mathbf{F}_0 \\ E_\Theta \end{bmatrix}. \tag{5.20}$$

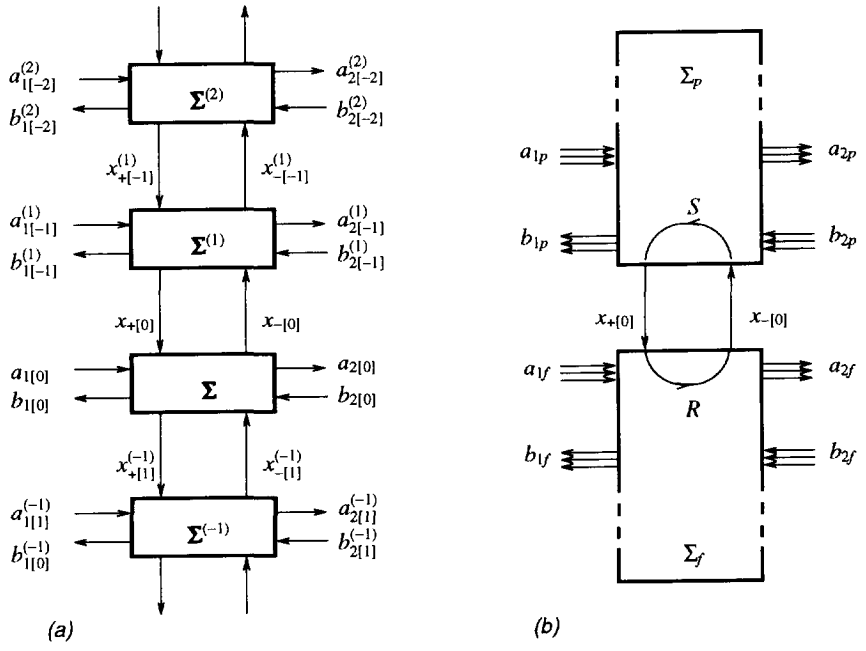


Figure 5.3. (*a*) The state transition scheme for $\boldsymbol{\Sigma}$, (*b*) The decomposition of $\boldsymbol{\Sigma}$ into a past operator $\boldsymbol{\Sigma}_p$ and a future operator $\boldsymbol{\Sigma}_f$ linked by the state $[x_{+[0]} \quad x_{-[0]}]$. This summarizes the figure on the left for all time.

We first show that Θ_p and Θ_f are J-unitary operators. Then, as a consequence, we can define operators $(\cdot)\Sigma_p$, $(\cdot)\Sigma_f$:

$$\begin{aligned} [x_{-[0]} \quad a_{1p} \quad b_{2p}]\,\Sigma_p &= [x_{+[0]} \quad a_{2p} \quad b_{1p}] \\ [x_{+[0]} \quad a_{1f} \quad b_{2f}]\,\Sigma_f &= [x_{-[0]} \quad a_{2f} \quad b_{1f}] \end{aligned}$$

which are the (non-causal) scattering operators corresponding to Θ_p and Θ_f, respectively (see figure 5.3(*b*)). The J-unitarity of Θ_p and Θ_f, and hence the existence and unitarity of Σ_p and Σ_f, is asserted in the following proposition.

PROPOSITION 5.14. *Let $\Theta \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a locally finite J-unitary operator, and let $\boldsymbol{\Theta}$ be a J-unitary realization for Θ. Then Θ_p and Θ_f are J-unitary operators, and Σ_p, Σ_f are well-defined unitary operators.*

PROOF Let $\hat{\Theta}$ be given as in equation (5.19). Since $\hat{\Theta}$ is 'the same' operator as Θ, it is J-unitary as well, which gives the relations

$$\left\{\begin{array}{lcl} E_\Theta J E_\Theta^* & = & J, \\ H_\Theta J E_\Theta^* & = & 0, \\ H_\Theta^* J H_\Theta + E_\Theta^* J E_\Theta & = & J, \end{array}\right. \qquad \left\{\begin{array}{lcl} K_\Theta^* J K_\Theta & = & J, \\ H_\Theta^* J K_\Theta & = & 0, \\ H_\Theta J H_\Theta^* + K_\Theta J K_\Theta^* & = & J. \end{array}\right. \tag{5.21}$$

Let $\mathbf{F}$ be a J-orthonormal basis in $\mathcal{H}(\Theta)$, and let $\mathbf{F}_0 = J\mathbf{F}J\Theta$ be the corresponding J-orthonormal basis in $\mathcal{H}_0(\Theta)$. Note that $\mathbf{F}_0$ is also given by $\mathbf{F}_0 = J\mathbf{F}JH_\Theta$, so that $\mathbf{F}_0 J E_\Theta^* = J\mathbf{F}JH_\Theta J E_\Theta^* = 0$. With the chosen basis, the Hankel operator has a factorization as $H_\Theta = \mathbf{P}_0(\cdot\,\mathbf{F}^*)\mathbf{F}_0$ and $H_\Theta^* = \mathbf{P}_0(\cdot\,\mathbf{F}_0^*)\mathbf{F}$, so that

$$H_\Theta^* J H_\Theta = \mathbf{P}_0(\cdot\,\mathbf{F}_0^*)\mathbf{P}_0(\mathbf{F}J\mathbf{F}^*)\mathbf{F}_0 = \mathbf{P}_0(\cdot\,\mathbf{F}_0^*)J\mathbf{F}_0. \tag{5.22}$$

Θ_f as in (5.20) has adjoint $\Theta_f^* = [\mathbf{P}_0(\cdot\,\mathbf{F}_0^*)\;\; E_0^*]$, so that (with (5.21))

$$\begin{aligned} \Theta_f J \Theta_f^* &= \begin{bmatrix} \mathbf{F}_0 \\ E_\Theta \end{bmatrix} J \begin{bmatrix} \mathbf{P}_0(\cdot\,\mathbf{F}_0^*) & E_\Theta^* \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{P}_0(\mathbf{F}_0 J \mathbf{F}_0^*) & \mathbf{F}_0 J E_\Theta^* \\ \mathbf{P}_0(E_\Theta J \mathbf{F}_0) & E_\Theta J E_\Theta^* \end{bmatrix} \\ &= \begin{bmatrix} J_\mathcal{B} & 0 \\ 0 & J_\mathcal{M} \end{bmatrix} \end{aligned}$$

and with (5.22), also

$$\Theta_f^* J \Theta_f = [\mathbf{P}_0(\cdot\,\mathbf{F}_0^*) \;\; E_\Theta^*] J \begin{bmatrix} \mathbf{F}_0 \\ E_\Theta \end{bmatrix} = \mathbf{P}_0(\cdot\,\mathbf{F}_0^*)J\mathbf{F}_0 + E_\Theta^* J E_\Theta = J.$$

Hence Θ_f is J-unitary. The J-unitarity of Θ_p follows in a dual way. □

State dimension of $\Sigma_{22} = \Theta_{22}^{-1}$

We have seen in proposition 5.13 that if the state signature sequence $J_\mathcal{B} = I$, then $\Sigma_{22} = \Theta_{22}^{-1}$ is upper. In chapter 6, an important role is played by J-unitary operators with non-trivial state signature, so that Θ_{22}^{-1} is not upper. In particular, we are interested in the dimension of the state space sequence $\mathcal{H}(\Theta_{22}^{-*})$ of Θ_{22}^{-1}, determined by the lower (anti-causal) part of Θ_{22}^{-1}. To this end, we use in this section a "conjugate-Hankel" operator, defined as

$$H' := H'_{\Theta_{22}^{-1}} = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\cdot\,\Theta_{22}^{-1})\big|_{\mathcal{U}_2} \tag{5.23}$$

The definition is such that $\mathcal{H}(\Theta_{22}^{-*}) = \mathrm{ran}(H')$.

Let the signals a_1, b_1, a_2, b_2 and the state sequences x_+, x_- be in $\mathcal{X}_2$ and be related by $\boldsymbol{\Theta}$ as in (5.13). As proven in proposition 5.14, Θ can be split into operators $(\cdot)\Theta_p$ and $(\cdot)\Theta_f$ via

$$\begin{aligned} [a_{1p} \;\; b_{1p}]\,\Theta_p &= [x_{+[0]} \;\; x_{-[0]} \;\; a_{2p} \;\; b_{2p}] \\ [x_{+[0]} \;\; x_{-[0]} \;\; a_{1f} \;\; b_{1f}]\,\Theta_f &= [a_{2f} \;\; b_{2f}], \end{aligned} \tag{5.24}$$

and the related scattering operators Σ_p and Σ_f are well defined by

$$\begin{array}{rcl} [x_{-[0]} \quad a_{1p} \quad b_{2p}]\,\Sigma_p &=& [x_{+[0]} \quad a_{2p} \quad b_{1p}] \\ [x_{+[0]} \quad a_{1f} \quad b_{2f}]\,\Sigma_f &=& [x_{-[0]} \quad a_{2f} \quad b_{1f}] \end{array} \tag{5.25}$$

and constitute the same relations as in (5.24).

Because $\Theta_{22}^{-1} = \Sigma_{22}$, the conjugate-Hankel operator H' defined in (5.23) is a restriction of the partial map $\Sigma_{22} : b_2 \mapsto b_1$. Indeed, $H' : b_{2f} \mapsto b_{1p}$ is such that b_{2p} and b_{1p} satisfy the input-output relations defined by Σ under the conditions $a_1 = 0$ and $b_{2p} = 0$ (see also figure 5.3(*b*)). H', as a Hankel operator, can be factored as $H' = \sigma\tau$, where the operators

$$\begin{array}{lcl} \sigma: & b_{2f} & \mapsto \quad x_{-[0]} \\ \tau: & x_{-[0]} & \mapsto \quad b_{1p} \end{array}$$

can be derived from Σ_f and Σ_p by elimination of $x_{+[0]}$, again taking $a_1 = 0$ and $b_{2p} = 0$. We show, in proposition 5.15, that the operator σ is 'onto' while τ is 'one-to-one', so that the factorization of H' into these operators is minimal. It is even *uniformly* minimal: the state $x_{-[0]}$ is uniformly controllable by b_{2f} (*i.e.*, the range of σ spans $\mathcal{D}_2$), and $x_{-[0]}$ as input of τ is uniformly observable. It follows, in proposition 5.16, that the dimension of $x_{-[0]}$ at each point in time determines the local dimension of the subspace $\mathcal{H}(\Theta_{22}^{-*})$ at that point.

PROPOSITION 5.15. *Let $\Theta \in \mathcal{U}$ be a locally finite J-unitary operator, with J-unitary realization $\boldsymbol{\Theta}$ such that $\ell_A < 1$. Let $x_+, x_-, a_1, b_1, a_2, b_2$ satisfy (5.24) and (5.25).*

1. *If $a_{1p} = 0$ and $b_{2p} = 0$, then the map $\tau : x_{-[0]} \mapsto b_{1p}$ is one-to-one and boundedly invertible on its range, i.e.,*

$$\exists\, \varepsilon > 0 : \quad \|b_{1p}\| \geq \varepsilon \|x_{-[0]}\|. \tag{5.26}$$

2. *The map $\sigma : b_{2f} \mapsto x_{-[0]}$ is onto, and moreover, there exists $M < \infty$ such that for any $x_{-[0]}$ there is a b_{2f} in its pre-image such that*

$$\|b_{2f}\| \leq M \|x_{-[0]}\|.$$

PROOF

1. *The map $\tau : x_{-[0]} \mapsto b_{1p}$ is one-to-one.* Put $a_{1p} = 0$ and $b_{2p} = 0$. Equation (5.25) gives $[x_{-[0]} \;\; 0 \;\; 0]\Sigma_p = [x_{+[0]} \;\; a_{2p} \;\; b_{1p}]$, that is, we have for some $x_{+[0]}$ and a_{2p}

$$[0 \quad b_{1p}]\,\Theta_p = [x_{+[0]} \quad x_{-[0]} \quad a_{2p} \quad 0]. \tag{5.27}$$

Since Θ_p is bounded, $\|b_{1p}\| < 1 \;\Rightarrow\; \|x_{-[0]}\| < M$ and hence, with $\varepsilon = 1/M$: $\|x_{-[0]}\| \geq 1 \;\Rightarrow\; \|b_{1p}\| \geq \varepsilon$. It follows that $x_{-[0]} \mapsto b_{1p}$ is one-to-one as claimed, and that (5.26) holds.

2. *The map* σ: $b_{2f} \mapsto x_{-[0]}$ *is onto.* Let be given any $x_{-[0]}$. We have to show that there is a b_{2f} that via Σ_f can generate this state. First, with $a_{1p} = b_{2p} = 0$, Σ_p associates a unique b_{1p} and $x_{+[0]}$ to $x_{-[0]}$. Put also $a_{1f} = b_{1f} = 0$, then Θ generates a corresponding b_{2f} as $b_{2f} = b_1\Theta_{22}$. Because Σ_f is well defined, application of Σ_f to $[x_{+[0]}\ 0\ b_{2f}]$ gives again a state $x'_{-[0]}$; but this must be equal to $x_{-[0]}$ because they both generate the same b_{1p} and the map $x_{-[0]} \mapsto b_{1p}$ is one-to-one. Hence this b_{2f} generates the given state $x_{-[0]}$. In addition, we have from $\|b_{1p}\| \leq \|x_{-[0]}\|$ and $\|\Theta\| \leq M < \infty$ that

$$\begin{aligned} \|b_{2f}\| &\leq \|\Theta_{22}\|\,\|b_{1p}\| \\ &\leq M\|x_{-[0]}\|. \end{aligned}$$

This means that the state $x_{-[0]}$ is uniformly controllable by b_{2f} as well. □

Proposition 5.15 is instrumental in proving that the sequence of the number of states x_- of the anti-causal part of Θ_{22}^{-1} is equal to the sequence of ranks of the Hankel operator H'.

PROPOSITION 5.16. *Let* $\Theta \in \mathcal{U}$ *be a locally finite J-unitary operator, with state signature operator* $J_{\mathcal{B}}$. *The s-dimension of* $\mathcal{H}(\Theta_{22}^{-*})$ *is equal to* $\#_-(J_{\mathcal{B}}) = \#(\mathcal{B}_-)$, *i.e., the sequence of the number of negative entries in* $J_{\mathcal{B}}$.

PROOF

$$\begin{aligned} \mathcal{H}(\Theta_{22}^{-*}) &= \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2\Theta_{22}^{-1}) \\ &= \{\mathbf{P}_{\mathcal{L}_2Z^{-1}}(b_{2f}\Theta_{22}^{-1}) :\ b_2 \in \mathcal{U}_2\}. \end{aligned}$$

Put $a_1 = 0$ and $b_{2p} = 0$ so that $b_{1p} = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(b_{2f}\Theta_{22}^{-1})$. The space $\mathcal{H}(\Theta_{22}^{-*}) = \{b_{1p} :\ b_2 \in \mathcal{U}_2\}$ is generated by the map $H' :\ b_{2f} \mapsto b_{1p}$. But this map can be split into $\sigma :\ b_{2f} \mapsto x_{-[0]}$ and $\tau :\ x_{-[0]} \mapsto b_{1p}$. Because $[x_{-[0]}\ 0\ 0]\Sigma_p = [x_{+[0]}\ a_{2p}\ b_{1p}]$, the signal $x_{-[0]}$ determines b_{1p} completely. In proposition 5.15 we showed that $x_{-[0]} \mapsto b_{1p}$ is one-to-one and that $b_{2f} \mapsto x_{-[0]}$ is onto. Hence, the state $x_{-[0]}$ is both uniformly observable in b_{1p} and uniformly controllable by b_{2f}, *i.e.*, its state dimension sequence for the map $b_{2f} \mapsto b_{1p}$ is minimal at each point in time. Since the number of state variables in $x_{-[0]}$ is given by $\#_-(J_{\mathcal{B}}) = \#(\mathcal{B}_-)$, it follows that

$$\text{s-dim}\ \mathcal{H}(\Theta_{22}^{-*}) = \#(\mathcal{B}_-).$$

□

5.2 J-INNER COPRIME FACTORIZATION

An extension of the external (inner-coprime) factorization in chapter 4 is the case where an operator $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ and a signature $J_{\mathcal{M}}$ is given, and a factorization of the form

$$T^*\Theta = \Delta \tag{5.28}$$

(where $\Theta \in \mathcal{U}$ is J-unitary and Δ is upper) is desired. We take the above definition as dual to the formulation of the external factorization in section 4.2, where we had $T = \Delta^* V$, or $VT^* = \Delta$; the definition in (5.28) connects more easily to the interpolation theory in later sections. The factorization results are also an extension of the theory in section 4.2, although we only consider the case where $\ell_A < 1$.

A dual form of proposition 4.7 holds:

PROPOSITION 5.17. *Let be given operators $T \in \mathcal{U}$ and $\Theta \in \mathcal{U}$. Then $\Delta = T^*\Theta$ is upper if and only if $\mathcal{L}_2 Z^{-1}\Theta^* \subset \mathcal{K}(T)$. If Θ is J-unitary, then $\mathcal{L}_2 Z^{-1}\Theta^* J = \mathcal{K}(\Theta)$, and Δ is upper if and only if Θ satisfies*

$$\overline{\mathcal{H}}(JT) \subset \mathcal{H}(\Theta).$$

The construction of such a Θ is comparable to the construction for inner operators. Let $\{A, B, C, D\}$ be a realization for T. Assume that the realization is uniformly controllable and that $\ell_A < 1$, then $\mathbf{F} = \left(BZ(I - AZ)^{-1}\right)^*$ is a strong basis representation such that $\overline{\mathcal{H}}(T) \subset \mathcal{D}_2\mathbf{F}$. An operator Θ such that $\Delta \in \mathcal{U}$ is obtained by taking $\mathcal{H}(\Theta) = \mathcal{D}_2\mathbf{F}J$, and a J-orthonormal realization of Θ is obtained by making $\mathbf{F}J$ J-orthonormal, which is possible if and only if $\Lambda_{\mathbf{F}}^J = \mathbf{P}_0(\mathbf{F}J\mathbf{F}^*)$ is boundedly invertible. Let $J_{\mathcal{B}}$ be the signature of $\Lambda_{\mathbf{F}}^J$, then $\Lambda_{\mathbf{F}}^J = R^* J_{\mathcal{B}} R$ for some invertible state transformation R, and hence A_Θ and B_Θ of a J-unitary realization are given by

$$\begin{bmatrix} A_\Theta \\ B_\Theta \end{bmatrix} = \begin{bmatrix} R & \\ & I \end{bmatrix} \begin{bmatrix} A \\ JB \end{bmatrix} R^{-(-1)} \tag{5.29}$$

It remains to complete this realization such that

$$\boldsymbol{\Theta} = \begin{bmatrix} A_\Theta & C_\Theta \\ B_\Theta & D_\Theta \end{bmatrix}$$

is $(\mathbf{J}_1, \mathbf{J}_2)$-unitary. This step is less obvious than for inner systems, so we first prove an additional lemma on this before stating the main theorem.

LEMMA 5.18. *Let be given finite matrices α, β, and signature matrices j_1, j_2, j_3 such that*

$$\alpha^* j_1 \alpha + \beta^* j_2 \beta = j_3 .$$

Then there exist matrices γ, δ and a signature matrix j_4 such that $\theta = \left[\begin{smallmatrix} \alpha & \gamma \\ \beta & \delta \end{smallmatrix}\right]$ is a J-unitary matrix, in the sense

$$\theta^* \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix} \theta = \begin{bmatrix} j_3 & \\ & j_4 \end{bmatrix}, \qquad \theta \begin{bmatrix} j_3 & \\ & j_4 \end{bmatrix} \theta^* = \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix}.$$

PROOF Let α be an $(m_\alpha \times n_\alpha)$-dimensional matrix, and $\beta : (m_\beta \times n_\alpha)$. It is clear that if an extension exists, then j_4 is specified by the inertia relations:

$$\begin{aligned} \#_+(j_4) &= \#_+(j_1) + \#_+(j_2) - \#_+(j_3) \\ \#_-(j_4) &= \#_-(j_1) + \#_-(j_2) - \#_-(j_3) \end{aligned}$$

Since the first block column of θ is already J-isometric,

$$[\alpha^* \quad \beta^*] \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = j_3 ,$$

it remains to show that this column can be completed to a J-unitary matrix. Because j_3 is non-singular, the n_α columns of $\left[\begin{smallmatrix} \alpha \\ \beta \end{smallmatrix}\right]$ are linearly independent. Choose a matrix $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ with $m_\alpha + m_\beta - n_\alpha$ independent columns such that

$$[\alpha^* \quad \beta^*] \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix} = 0 \quad \Leftrightarrow \quad \begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} j_1 \alpha \\ j_2 \beta \end{bmatrix}^{\perp} \tag{5.30}$$

We claim that the square matrix $\left[\begin{smallmatrix} \alpha & c \\ \beta & d \end{smallmatrix}\right]$ is invertible. To prove this, it is enough to show that its null space is zero. Suppose that

$$\begin{bmatrix} \alpha & c \\ \beta & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

then

$$[\alpha^* \quad \beta^*] \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix} \begin{bmatrix} \alpha & c \\ \beta & d \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} j_3 x_1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

Hence $x_1 = 0$ and $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right] x_2 = 0$. But the columns of $\left[\begin{smallmatrix} c \\ d \end{smallmatrix}\right]$ are linearly independent, so that $x_2 = 0$. Thus

$$\begin{bmatrix} \alpha^* & \beta^* \\ c & d^* \end{bmatrix} \begin{bmatrix} j_1 & \\ & j_2 \end{bmatrix} \begin{bmatrix} \alpha & c \\ \beta & d \end{bmatrix} = \begin{bmatrix} j_3 & \\ & N \end{bmatrix}$$

where N is a square invertible matrix. By the usual inertia argument, the signature of N is equal to j_4, and hence N has a factorization $N = R^* j_4 R$, where R is invertible. Thus putting

$$\begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix} R^{-1} , \qquad \theta = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$$

ensures that θ is j-unitary as required. □

THEOREM 5.19. *Let be given a subspace $\mathcal{H} = \mathcal{D}_2^{\mathcal{B}} \mathbf{F} J$ in $Z^{-1}\mathcal{L}_2^{\mathcal{M}}$, specified by a bounded basis representation $\mathbf{F} = \left(BZ(I - AZ)^{-1}\right)^*$ ($\ell_A < 1$) which is such that $\Lambda_{\mathbf{F}}^J$ is boundedly invertible. Then there exists a bounded J-unitary operator $\Theta \in \mathcal{U}(\mathcal{M}, \mathcal{N}_\Theta)$ such that*

$\mathcal{H} = \mathcal{H}(\Theta)$. Θ *is unique up to a right diagonal J-unitary factor.* $\mathcal{N}_\Theta$ *is the space sequence with dimension sequence given by*

$$\begin{aligned} \#(\mathcal{N}_\Theta)_+ &= \#\mathcal{B}_+ - \#\mathcal{B}_+^{(-1)} + \#\mathcal{M}_+ \geq 0 \\ \#(\mathcal{N}_\Theta)_- &= \#\mathcal{B}_- - \#\mathcal{B}_-^{(-1)} + \#\mathcal{M}_- \geq 0. \end{aligned} \tag{5.31}$$

PROOF Since $\Lambda_{\mathbf{F}}^J$ is boundedly invertible, there is a signature operator $J_{\mathcal{B}}$ and a boundedly invertible operator $R \in \mathcal{D}$ such that $\Lambda_{\mathbf{F}}^J = R^* J_{\mathcal{B}} R$. The signature $J_{\mathcal{B}}$ implies a space sequence decomposition $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_-$, and since $\Lambda_{\mathbf{F}}^J$ satisfies the J-Lyapunov equation

$$A^* \Lambda_{\mathbf{F}}^J A + B^* J_{\mathcal{M}} B = (\Lambda_{\mathbf{F}}^J)^{(-1)}, \tag{5.32}$$

A_Θ, B_Θ, given by

$$\begin{bmatrix} A_\Theta \\ B_\Theta \end{bmatrix} = \begin{bmatrix} R & \\ & I \end{bmatrix} \begin{bmatrix} A \\ JB \end{bmatrix} R^{-(-1)} \tag{5.33}$$

form a J-isometric block column with diagonal entries. We proceed with the construction of a realization $\boldsymbol{\Theta}$ of the form

$$\boldsymbol{\Theta} = \begin{bmatrix} A_\Theta & C_\Theta \\ B_\Theta & D_\Theta \end{bmatrix} = \begin{bmatrix} R & \\ & I \end{bmatrix} \begin{bmatrix} A & C' \\ JB & D_\Theta \end{bmatrix} \begin{bmatrix} R^{-(-1)} & \\ & I \end{bmatrix} \tag{5.34}$$

which is a square matrix at each point k, and where C_Θ (or C') and D_Θ are yet to be determined. $\boldsymbol{\Theta}$ is to satisfy $\boldsymbol{\Theta}^* \mathbf{J}_1 \boldsymbol{\Theta} = \mathbf{J}_2$, $\boldsymbol{\Theta} \mathbf{J}_2 \boldsymbol{\Theta}^* = \mathbf{J}_1$, for

$$\mathbf{J}_1 = \begin{bmatrix} J_{\mathcal{B}} & \\ & J_{\mathcal{M}} \end{bmatrix}, \qquad \mathbf{J}_2 := \begin{bmatrix} J_{\mathcal{B}}^{(-1)} & \\ & J_{\mathcal{N}_\Theta} \end{bmatrix} \tag{5.35}$$

where $J_{\mathcal{N}_\Theta}$ is still to be determined, and with it the dimensionality of the output space sequence $\mathcal{N}_\Theta$. However, since all other signatures are known at this point, these follow from the inertia property (equation (5.10)) as the space sequence with dimensions given by (5.31). To obtain $\boldsymbol{\Theta}$, it remains to show that $\begin{bmatrix} A_\Theta \\ B_\Theta \end{bmatrix}$ can be completed to form $\boldsymbol{\Theta}$ in (5.34), in such a way that the whole operator is J-unitary. This completion can be achieved for each point k at the local level, and exists as was shown in lemma 5.18. Since $\boldsymbol{\Theta}$ is J-unitary and $\ell_A < 1$, theorem 5.12 implies that the corresponding operator Θ is J-unitary. Finally, $\mathcal{H}(\Theta) = \mathcal{H}$ by construction.

The uniqueness of Θ, up to a left diagonal J-unitary factor, is proven in the same way as for inner operators in the Beurling-Lax theorem (theorem 4.11). Indeed, let Θ_1 be another J-unitary operator such that $\mathcal{H} = \mathcal{H}(\Theta_1)$, then $\mathcal{K} = \mathcal{L}_2 Z^{-1} \ominus \mathcal{H} = \mathcal{L}_2 Z^{-1} \Theta^* = \mathcal{L}_2 Z^{-1} \Theta_1^*$, so that

$$\begin{aligned} \mathcal{L}_2 Z^{-1} \Theta^* J \Theta_1 &= \mathcal{L}_2 Z^{-1} \\ \mathcal{L}_2 Z^{-1} \Theta_1^* J \Theta &= \mathcal{L}_2 Z^{-1} \end{aligned}$$

which implies $\Theta^* J \Theta_1 \in \mathcal{D}$, say $\Theta^* J \Theta_1 = JD$, where $D \in \mathcal{D}$. Then $\Theta_1 = \Theta D$, and D must be J-unitary. □

COROLLARY 5.20. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a locally finite operator with uniformly controllable realization $\{A, B, C, D\}$ such that $\ell_A < 1$, and let be given a signature matrix $J_{\mathcal{M}}$. If the solution Λ^J of the J-Lyapunov equation*

$$A^*\Lambda^J A + B^* J_{\mathcal{M}} B = (\Lambda^J)^{(-1)} \tag{5.36}$$

is such that Λ^J is boundedly invertible, then there exists a bounded J-unitary operator $\Theta \in \mathcal{U}(\mathcal{M}, \mathcal{N}_\Theta)$ such that

$$T^*\Theta = \Delta \in \mathcal{U}.$$

$\mathcal{N}_\Theta$ and its signature are determined by equation (5.31).

PROOF The conditions imply that the subspace $\mathcal{H} = \mathcal{P}_2 \mathbf{F} J = \mathcal{P}_2 \left(BZ(I-AZ)^{-1}\right)^* J$ has $\Lambda^J_{\mathbf{F}} = \Lambda^J$ boundedly invertible. Hence theorem 5.19 asserts that there is a J-unitary operator Θ such that $\mathcal{H}(\Theta) = \mathcal{H}$. Note that a necessary condition for Λ^J to be invertible is that the given realization be uniformly controllable, so that

$$\overline{\mathcal{H}}(JT) = \overline{\mathcal{H}}(T)J \subset \mathcal{P}_2 \mathbf{F} J = \mathcal{H} = \mathcal{H}(\Theta).$$

This in turn implies that $\Delta = T^*\Theta$ is upper. □

For later use, we evaluate $\Delta = T^*\Theta$. Instead of C_Θ, we use $C' = R^{-1}C_\Theta$ (see equation (5.34)), as A_Δ will become equal to the original A in this case. We also apply the relation $J_{\mathcal{B}} \mathbf{F} J_{\mathcal{M}} \Theta = \mathbf{F}_0$, which in case $\ell_A < 1$ reads

$$(I - Z^* A^*_\Theta)^{-1} Z^* B^*_\Theta J \Theta = J_{\mathcal{B}} (I - A_\Theta Z)^{-1} C_\Theta .$$

Thus

$$\begin{aligned}
\Delta &= [D^* + C^*(I - Z^*A^*)^{-1}Z^*B^*]\Theta \\
&= D^*[D_\Theta + B_\Theta Z(I - A_\Theta Z)^{-1}C_\Theta] + C^*R(I - Z^*A^*_\Theta)^{-1}Z^*B^*_\Theta J\Theta \\
&= D^*D_\Theta + D^*B_\Theta Z(I - A_\Theta Z)^{-1}C_\Theta + C^*RJ_{\mathcal{B}}(I - A_\Theta Z)^{-1}C_\Theta \\
&= D^*D_\Theta + D^*JBZ(I - AZ)^{-1}C' + C^*\Lambda^J(I - AZ)^{-1}C' .
\end{aligned}$$

Consequently,

$$\Delta = T^*\Theta = \left\{D^*D_\Theta + C^*\Lambda^J C'\right\} + \left\{D^*JB + C^*\Lambda^J A\right\} Z(I - AZ)^{-1}C', \tag{5.37}$$

where Λ^J is given by (5.36) and C' by (5.34).

5.3 DEFINITE INTERPOLATION

As an application of the subspace representation theorem (theorem 5.19), or the J-inner coprime factorization theorem, we consider the solution of a class of constrained interpolation problems, known as the Nevanlinna-Pick interpolation problem. Starting in 1989,

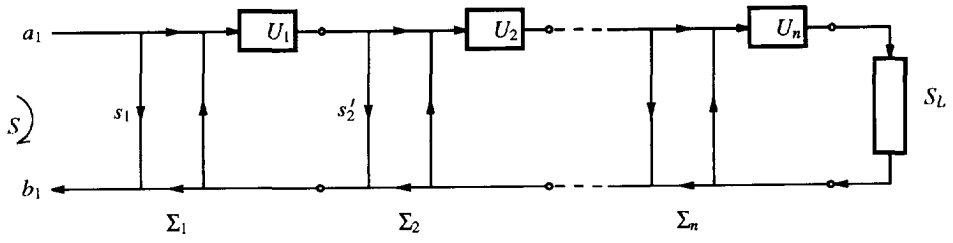


Figure 5.4. Time-invariant interpolation of a function $S(z)$. The cascade as a whole has a structure as in figure 5.1(b).

constrained interpolation problems of increasing generality for time-varying systems have been considered by Alpay, Dewilde and Dym [17, 15, 18, 16]. The topic was adopted by Ball, Gohberg and Kaashoek [19, 11], who previously worked (with Woerdeman) on the related problem of time-varying Nehari extensions [20, 21, 22, 23]; see section 6.5. (A third school is formed by the group of Kailath, working on interpolation problems for systems with a displacement rank, and approaching this problem using Schur and Levinson recursions. See *e.g.*, [24, 25, 26] and references therein.)

Nevanlinna-Pick interpolation

In its simplest form in the time-invariant setting, the Nevanlinna-Pick problem is the following [5]:

> Given n points $\{\omega_i\}$ and $\{s_i\}$ in the open unit disc $\mathbf{D}$, does an analytic and contractive function $S(z)$ exist such that $S(\omega_i) = s_i$ $(i = 1, \cdots, n)$. If so, describe all possible S.

For a single pair of points (ω_1, s_1), the answer to the existence question is of course affirmative: take $S(z) = s_1$. All other interpolating $S(z)$ have the form $S(z) = s_1 + (z-\omega_1)S'(z)$ with $S'(z)$ analytic in $\mathbf{D}$, but this form is too general: the condition that $S(z)$ must also be contractive still has to be implemented. Splitting the problems into two steps, one looks for a function $S''(z)$ which is contractive and satisfies the simpler interpolation condition that $S''(\omega_1) = 0$, and a way to transform this interpolation value to the required value s_1. The former function is given by $S''(z) = U_1(z)S_1(z)$, where $S_1(z)$ is any contractive and analytic function, and $U_1(z)$ is a (Blaschke) function which is lossless and has a zero at ω_1. The transformation of the interpolated value 0 to s_1 is effected by a linear fractional transform (which maps circles into circles and contractive functions into contractive functions), and the result is that all contractive interpolating $S(z)$ are given by

$$S(z) = \frac{s_1 - U_1(z)S_1(z)}{1 - s_1^* U_1(z)S_1(z)}, \qquad U_1(z) = \frac{z - \omega_1}{1 - \omega_1^* z}, \quad \text{(all contractive } S_1\text{, analytic in } \mathbf{D}\text{)}.$$

If there are more interpolation points, then the freedom in $S_1(z)$ can be used to meet each of them without disturbing the first interpolation condition. To this end, the remaining interpolation values $\{s_2, \cdots, s_n\}$ must be translated into conditions on $S_1(z)$ (the interpolation points $\{\omega_1, \cdots, \omega_n\}$ stay the same). Thus, all interpolation points can be treated (order) recursively. The recursion breaks down if at any point in the recursion, one of the transformed interpolation points becomes larger than one in modulus, and in this case the interpolation problem has no solution. It remains to note that the linear fractional transforms correspond to elementary Θ (or Σ)-sections, so that $S(z)$ is a cascade of such sections (see figure 5.4, which has the general structure of 5.1(b)).

In the time-varying setting, a number of things change. To start, the time-invariant z-transform is no longer applicable, and one has to define an extended notion of such transform, such that it has properties that are useful for describing interpolation problems. Such an extension was done by Alpay, Dewilde and Dym in [15], who called the new transform the W-transform. For operators T in $\mathcal{U}(\mathcal{M}, \mathcal{N})$, the W-transform of T at point $V \in \mathcal{D}$ is denoted by $\hat{T}(V)$, and is defined by

$$T = \sum_{i=0}^{\infty} Z^{[i]} T_{[i]} \qquad \rightarrow \qquad \hat{T}(V) = \sum_{i=0}^{\infty} V^{[i]} T_{[i]}$$

whenever the sum converges, which it does at least for $\ell_V < 1$. In the above expression, $V^{[n]} := VV^{(1)} \cdots V^{(n-1)}$, so that the definition makes sense only for operators $V \in \mathcal{D}(\mathcal{M}, \mathcal{M}^{(1)})$, where $\mathcal{M}$ is the input space sequence of T. The definition reduces to the z-transform for Toeplitz operators T and V. We do not use many of the properties of the W-transform, but an important one is [15]

$$\hat{T}(V) = 0 \qquad \Leftrightarrow \qquad T = (Z - V)T' \quad (T' \in \mathcal{U}). \tag{5.38}$$

Hence, there is a notion of the 'left' zeros of T in terms of the W-transform. A second property, useful to prove the above property and also to make a transition to a more familiar context, is

$$\hat{T}(V) = \mathbf{P}_0((I - VZ^*)^{-1}T) \tag{5.39}$$

This is straightforward to derive: for $A, B \in \mathcal{U}$, we have

$$\begin{aligned} \mathbf{P}_0(A^*B) &= \mathbf{P}_0\left(\textstyle\sum (A_{[n]})^* Z^{-n} B\right) \\ &= \textstyle\sum (A_{[n]})^* \mathbf{P}_0(Z^{-n}B) \\ &= \textstyle\sum (A_{[n]})^* B_{[n]} \end{aligned}$$

Taking $A = (I - ZV^*)^{-1}$, one has $A_{[n]} = V^{*(n-1)} \cdots V^{*(1)} V^*$, or $(A_{[n]})^* = V^{[n]}$.

The Nevanlinna-Pick interpolation problem in the time-varying setting can be translated from the above description as

Given a set of n 'points' $\{v_i\}$ in $\mathcal{D}(\mathcal{M}, \mathcal{M}^{(1)})$, each with $\ell_{v_i} < 1$, and n 'values' $\{s_i\}$, all in $\mathcal{D}(\mathcal{M}, \mathcal{N})$, does a contractive operator $S \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ exist such that $\hat{S}(v_i) = s_i$ $(i = 1, \cdots, n)$. If so, describe all solutions.

Alternatively, using equation (5.39), the interpolation condition can be translated into

$$\begin{aligned} \hat{S}(v_i) = s_i \;&\Leftrightarrow\; \mathbf{P}_0\left((I - v_i Z^*)^{-1} S\right) &&= s_i \\ &\Leftrightarrow\; \mathbf{P}_0\left(S^*(I - Z v_i^*)^{-1}\right) &&= s_i^* \\ &\Leftrightarrow\; \mathbf{P}\left(S^*(I - Z v_i^*)^{-1} E\right) &&= s_i^*(I - Z v_i^*)^{-1} E, \qquad (\text{all } E \in \mathcal{D}_2), \end{aligned} \tag{5.40}$$

where the last expression follows from the preceding by a shift-invariance property: write $A = (I - Z v_i^*)^{-1}$, then the structure of A is such that $\mathbf{P}(AZ^{-k}) = AD_k$ (for $D_k = \mathbf{P}_0(AZ^{-k}) \in \mathcal{D}$), and hence, the diagonals of the last expression in (5.40) are given by $\mathbf{P}_0(S^*AZ^{-k}) = \mathbf{P}_0(S^*\mathbf{P}(AZ^{-k})) = \mathbf{P}_0(S^*A)D_k = s_i^* D_k = s_i^* \mathbf{P}_0(AZ^{-k})$. Formally, postmultiplication by $E \in \mathcal{D}_2$ is required to obtain an expression in $\mathcal{X}_2$ rather than $\mathcal{X}$, so that the projection operator $\mathbf{P}$ can safely be used.

It is also useful to collect the n data points into single diagonal operators. Let $V \in \mathcal{D}(\mathcal{M}^n, (\mathcal{M}^n)^{(1)})$ be a diagonal operator with matrix entries V_k $(k = -\infty, \cdots, \infty)$, given by

$$V_k = \begin{bmatrix} (v_1)_k & & & \\ & (v_2)_k & & \\ & & \ddots & \\ & & & (v_n)_k \end{bmatrix}. \tag{5.41}$$

Likewise, define diagonal operators α and β by

$$\begin{aligned} \alpha_k &= [\; I \quad \cdots \quad I \;] \\ \beta_k &= [\; (s_1^*)_k \quad \cdots \quad (s_n^*)_k \;]. \end{aligned} \tag{5.42}$$

Then the set of n interpolation conditions in equation (5.40) becomes a single condition, and the interpolation problem can be stated as [16]

The 'fundamental' interpolation problem. Given operators $\alpha, \beta, V \in \mathcal{D}$, with $\ell_V < 1$, does a contractive operator $S \in \mathcal{U}$ exist such that

$$\mathbf{P}(S^* \alpha(I - ZV^*)^{-1} E) = \beta(I - ZV^*)^{-1} E, \qquad (\text{all } E \in \mathcal{D}_2). \tag{5.43}$$

In fact, written this way, the interpolation problem is more general than the Nevanlinna-Pick interpolation problem by which we started. Other classical interpolation problems, such as the Carathéodory-Fejér problem (multiple interpolation conditions — also on derivatives — in a single point), and mixed cases as well, can be described by the same equation by taking V, α and β different from the structure in (5.41) and (5.42).

Solution of the fundamental interpolation problem

In order to find solutions to the interpolation problem, we make yet another translation from the interpolation condition. Write

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} (I - ZV^*)^{-1} = B(I - ZA)^{-1} = \mathbf{F}^* Z^* \tag{5.44}$$

for obvious choices of A and B. Then the interpolation condition on S is equal to

$$\begin{aligned} & \mathbf{P}(S^* \alpha (I - ZV^*)^{-1} E) &&= \beta (I - ZV^*)^{-1} E \\ \Leftrightarrow \quad & \mathbf{P}([S^* \quad -I] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} (I - ZV^*)^{-1} E) &&= 0 \\ \Leftrightarrow \quad & \mathbf{P}([S^* \quad -I] \mathbf{F}^* Z^* E) &&= 0, \qquad (\text{all } E \in \mathcal{D}_2), \end{aligned}$$

so that S is an interpolant if

$$\mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{D}_2 \mathbf{F} \begin{bmatrix} S \\ -I \end{bmatrix}) = 0.$$

As usual, $\mathcal{D}_2\mathbf{F}$ describes some linear manifold in $\mathcal{L}_2 Z^{-1}$. Define the signature operator J_B conformably to the partitioning of B into α and β, and let $\Lambda = \Lambda_{\mathbf{F}}^J = \mathbf{P}_0(\mathbf{F} J \mathbf{F}^*)$. If Λ is boundedly invertible, then $\mathcal{H} := \mathcal{D}_2 \mathbf{F} J$ is a closed (regular) subspace, and $S \in \mathcal{U}$ is an interpolant if it is contractive and

$$\mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{H} J \begin{bmatrix} S \\ -I \end{bmatrix}) = 0.$$

Since Λ is boundedly invertible, there is, by theorem 5.19, a J-unitary operator Θ such that $\mathcal{H} = \mathcal{H}(\Theta)$. The solution of the interpolation problem reduces to the construction of Θ. We have thus made a (well-known) connection of interpolation problems to J-unitary operator theory.

The following theorem concerns the case where the interpolation data is such that $\Lambda \gg 0$, which is a sufficient condition for the existence of an interpolant S. It appears in [16] as theorem 1.3. The proof here is however different, and does not explicitly use reproducing kernel properties as in [16]. (It is also slightly less general, in the sense that in the current description we only consider locally finite operators and uniformly positive Λ.) A third proof is given in [11].

THEOREM 5.21. *Let be given the interpolation data (5.43), and define* $\mathbf{F}$ *as in (5.44),* $\Lambda = \Lambda_{\mathbf{F}}^J$, *and* $\mathcal{H} = \mathcal{D}_2 \mathbf{F} J$. *If* $\Lambda \gg 0$, *then the fundamental interpolation problem has a solution* $S \in \mathcal{U}$, *and* $S \in \mathcal{U}$ *is a solution to the interpolation problem if and only if*

$$S = T_\Theta[S_L]$$

where Θ is any J-inner operator such that $\mathcal{H} = \mathcal{H}(\Theta)$, and S_L is some (resp. any) contractive operator in $\mathcal{U}$.

PROOF If $\Lambda \gg 0$, then one can construct a J-unitary operator Θ such that $\mathcal{H} = \mathcal{H}(\Theta)$ (theorem 5.19) and because $J_{\mathcal{B}} = I$, we have $\Theta_{22}^{-1} \in \mathcal{U}$ (proposition 5.13).

($\Leftarrow$) If $S = T_\Theta[S_L]$, $\|S_L\| \leq 1$, then equation (5.6) holds:

$$\begin{bmatrix} S \\ -I \end{bmatrix} = \Theta \begin{bmatrix} S_L \\ -I \end{bmatrix} \Phi^{-1}, \qquad \Phi = \Theta_{22} - \Theta_{21} S_L .$$

Since S_L is contractive and $\Theta_{22}^{-1} \in \mathcal{U}$ because $\Lambda \gg 0$, we have $\Phi^{-1} = (I - \Theta_{22}^{-1}\Theta_{21}S_L)^{-1}\Theta_{22}^{-1}$ is upper too. Hence $S = T_\Theta[S_L]$ implies

$$\mathcal{H}(\Theta) J \begin{bmatrix} S \\ -I \end{bmatrix} = \mathcal{H}(\Theta) J \Theta \begin{bmatrix} S_L \\ -I \end{bmatrix} \Phi^{-1} = \mathcal{H}_0(\Theta) \begin{bmatrix} S_L \\ -I \end{bmatrix} \Phi^{-1} \quad \in \mathcal{U}$$

is upper, so that

$$\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left(\mathcal{H}(\Theta) J \begin{bmatrix} S \\ -I \end{bmatrix} \right) = 0,$$

that is, S is an interpolant.

($\Rightarrow$) If S is an interpolant, then $\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left(\mathcal{H}(\Theta) J \begin{bmatrix} S \\ -I \end{bmatrix} \right) = 0$, and we have to show that there is some contractive operator $S_L \in \mathcal{U}$ such that $S = T_\Theta[S_L]$. The proof consists of four steps.

1. $X := \Theta^{-1} \begin{bmatrix} S \\ -I \end{bmatrix}$ *is upper.*

PROOF
$$\begin{aligned} \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 X) &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left(\mathcal{U}_2 \Theta^{-1} \begin{bmatrix} S \\ -I \end{bmatrix} \right) \\ &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left(\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left[\mathcal{U}_2 \Theta^* \right] J \begin{bmatrix} S \\ -I \end{bmatrix} \right) \qquad \text{[since } S \in \mathcal{U}\text{]} \\ &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}} \left(\mathcal{H}(\Theta) J \begin{bmatrix} S \\ -I \end{bmatrix} \right) \\ &= 0. \end{aligned}$$

2. Decomposing X into two upper operators G_1 and G_2, we can write

$$\begin{bmatrix} S \\ -I \end{bmatrix} = \Theta \begin{bmatrix} G_1 \\ -G_2 \end{bmatrix} \qquad (G_1, G_2 \in \mathcal{U}). \tag{5.45}$$

G_2 is boundedly invertible, and $S_L := G_1 G_2^{-1}$ is well defined and contractive. In addition, $S = (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1} = T_\Theta[S_L]$, as required.

PROOF Θ is boundedly invertible: $\Theta^{-1} = J\Theta^* J$ with $\|\Theta^{-1}\| = \|\Theta\|$. Hence $\Theta\Theta^* \geq \varepsilon I$ (for some $\varepsilon > 0$) and

$$\begin{aligned} G_1^* G_1 + G_2^* G_2 &= [S^* \ I]\, \Theta\Theta^* \begin{bmatrix} S \\ I \end{bmatrix} \\ &\geq \varepsilon(S^* S + I) \\ &\geq \varepsilon I. \end{aligned}$$

We also have from the J-unitarity of Θ and the contractivity of S that

$$G_1^* G_1 \leq G_2^* G_2 \tag{5.46}$$

Together, this shows that $G_2 \geq 1/2\, \varepsilon I$, and hence G_2 is boundedly invertible (but at this point potentially not in $\mathcal{U}$). With $S_L = G_1 G_2^{-1}$, equation (5.46) shows that $S_L^* S_L \leq 1$, and hence $\|S_L\| \leq 1$. Evaluating equation (5.45) gives

$$\begin{aligned} G_2^{-1} &= \Theta_{22} - \Theta_{21} S_L \\ S G_2^{-1} &= \Theta_{11} S_L - \Theta_{12} \end{aligned} \tag{5.47}$$

and hence $S = (\Theta_{11} S_L - \Theta_{12})(\Theta_{22} - \Theta_{21} S_L)^{-1}$.

3. *Let $X \in \mathcal{X}$ be a strictly contractive operator. Then $(I - X)^{-1} \in \mathcal{U} \Leftrightarrow X \in \mathcal{U}$.*

   PROOF ($\Rightarrow$) is clear. ($\Leftarrow$): Proposition 6.11 (which can be independently read at this point) claims that $(I-X)^{-1}$ is a Hilbert space isomorphism. Hence $\mathcal{U}_2(I-X)^{-1} = \mathcal{U}_2$, so that $\mathcal{U}_2 = \mathcal{U}_2(I-X)$, which implies $X \in \mathcal{U}$.

4. *S_L is upper.*

   PROOF From (5.47), we have $G_2\Theta_{22} = (I - \Theta_{22}^{-1}\Theta_{21} S_L)^{-1}$, where it is known that the left-hand side is upper. Hence, by step 3, $\Theta_{22}^{-1}\Theta_{21} S_L$ is upper, and since $\Theta_{22} \in \mathcal{U}$, $\Theta_{21} S_L$ is upper, and $G_2^{-1} = \Theta_{22} - \Theta_{21} S_L$ is upper, too, so that $S_L = G_1 G_2^{-1} \in \mathcal{U}$. □

If we take $S_L = 0$, then we obtain the solution $S = -\Theta_{12}\Theta_{22}^{-1}$, which is sometimes called the "maximum entropy" or central solution of the interpolation problem. In this case, $S = \Sigma_{12}$, so that

$$\Sigma = \begin{bmatrix} \Sigma_{11} & S \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Σ is a unitary embedding of S. Σ_{11} and Σ_{22} contain the "interpolation transmission zeros," since they are such that S interpolates independently of the choice of S_L. In the time-invariant case, these transmission zeros are equal to the specified points ω_i. In the time-varying setting, $S = \Sigma_{12} + \Sigma_{11} S_L (I - \Sigma_{21} S_L)^{-1} \Sigma_{22}$, and to be independent of S_L, it follows that we must have $\hat{\Sigma}_{11}(\nu_i) = 0$. With Σ known, the discussion in chapter 9 has shown how a cascade factorization of Σ can be obtained. Each of the factors in the cascade has an

encoding of one of the v_i, but unlike in the time-invariant case, not only the interpolation values but also the v_i in the second and later sections are transformed by earlier sections, because multiplications are no longer commutative. A recursive construction of Σ should in principle be possible, but is problematic if the v_i are not boundedly invertible because the transformation of the v_i involves v_i^{-1} (similar problems have been reported by Kamen [27]). It is however possible to obtain an explicit formula for Θ, see [16].

In chapter 6, we extend the above definite interpolation theorem to apply to non-definite interpolation problems, too. Then, Λ is no longer required to be definite (but we still require Λ to be boundedly invertible), and hence the resulting interpolant S need no longer be in $\mathcal{U}$. We do not discuss cases where Λ is not invertible; some aspects of this (for the case where Λ is positive semi-definite) are treated in [16].

Bibliography

[1] J.A. Ball and J.W. Helton, "Lie groups over the field of rational functions, signed spectral factorization, and amplifier design," *J. Operator Theory*, vol. 8, pp. 19–64, 1982.

[2] J.A. Ball and J.W. Helton, "Interpolation problems of Pick-Nevanlinna and Loewner types for meromorphic matrix functions: Parametrization of the set of all solutions," *Integral Eq. Operator Th.*, vol. 9, pp. 155–203, 1986.

[3] J.W. Helton e.a., *Operator Theory, Analytic Functions, Matrices, and Electrical Engineering*, vol. 68 of *CBMS regional conference series*. Providence: American Math. Soc., 1987.

[4] L. de Branges and J. Rovnyak, *Canonical Models in Quantum Scattering Theory*, pp. 295–392. New York: Wiley, 1966.

[5] H. Dym, *J-Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation*. No. 71 in CBMS regional conference series, Providence: American Math. Soc., 1989.

[6] J.A. Ball, I. Gohberg, and L. Rodman, *Interpolation of Rational Matrix Functions*, vol. 45 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, 1990.

[7] P. Dewilde and H. Dym, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Informat. Th.*, vol. 27, pp. 446–461, July 1981.

[8] P. Dewilde and H. Dym, "Lossless chain scattering matrices and optimum linear prediction: The vector case," *Circuit Theory and Appl.*, vol. 9, pp. 135–175, 1981.

[9] P. Dewilde and H. Dym, "Lossless inverse scattering, digital filters, and estimation theory," *IEEE Trans. Informat. Th.*, vol. 30, pp. 644–662, July 1984.

[10] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Automat. Control*, vol. 26, pp. 301–320, Apr. 1981.

[11] J.A. Ball, I. Gohberg, and M.A. Kaashoek, "Nevanlinna-Pick interpolation for time-varying input-output maps: the discrete case," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 1–51, Birkhäuser Verlag, 1992.

[12] J. Bognar, *Indefinite Inner Product Spaces*. Springer Verlag, 1974.

[13] M.G. Krein, "Introduction to the geometry of indefinite J-spaces and to the theory of operators in those spaces," *Amer. Math. Soc. Transl.*, vol. 93, pp. 103–176, 1970.

[14] T.Ya. Azizov and I.S. Yokhvidov, *Linear Operators in Spaces with an Indefinite Metric*. Pure and Applied Mathematics, John Wiley & Sons, 1989.

[15] D. Alpay, P. Dewilde, and H. Dym, "Lossless Inverse Scattering and reproducing kernels for upper triangular operators," in *Extension and Interpolation of Linear Operators and Matrix Functions* (I. Gohberg, ed.), vol. 47 of *Operator Theory, Advances and Applications*, pp. 61–135, Birkhäuser Verlag, 1990.

[16] P. Dewilde and H. Dym, "Interpolation for upper triangular operators," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 153–260, Birkhäuser Verlag, 1992.

[17] D. Alpay and P. Dewilde, "Time-varying signal approximation and estimation," in *Signal Processing, Scattering and Operator Theory, and Numerical Methods* (M.A. Kaashoek, J.H. van Schuppen, and A.C.M. Ran, eds.), vol. III of *Proc. Int. Symp. MTNS-89*, pp. 1–22, Birkhäuser Verlag, 1990.

[18] P.M. Dewilde, "A course on the algebraic Schur and Nevanlinna-Pick interpolation problems," in *Algorithms and Parallel VLSI Architectures* (Ed. F. Deprettere and A.J. van der Veen, eds.), Elsevier, 1991.

[19] J.A. Ball, I. Gohberg, and M.A. Gohberg, "Time-varying systems: Nevanlinna-Pick interpolation and sensitivity minimization," in *Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing I (Proc. Int. Symp. MTNS-91)* (H. Kimura and S. Kodama, eds.), pp. 53–58, MITA Press, Japan, 1992.

[20] H. Woerdeman, *Matrix and Operator Extensions*. PhD thesis, Dept. Math. Comp Sci., Free University, Amsterdam, The Netherlands, 1989.

[21] I. Gohberg, M.J. Kaashoek, and H.J. Woerdeman, "The band method for positive and strictly contractive extension problems: an alternative version and new applications," *Integral Eq. Operator Th.*, vol. 12, pp. 343–382, 1989.

[22] I. Gohberg, M.A. Kaashoek, and H.J. Woerdeman, "A maximum entropy principle in the general framework of the band method," *J. Functional Anal.*, vol. 95, pp. 231–254, Feb. 1991.

[23] I. Gohberg, M.A.Kaashoek, and H.J. Woerdeman, "Time variant extension problems of Nehari type and the band method," in *H^∞-Control Theory (lectures given at the 2nd session of C.I.M.E., Como, June 18-26, 1990)* (C. Foias, B. Francis, and J.W. Helton, eds.), Lecture Notes Math. 1496, pp. 309–323, Springer Verlag, 1991.

[24] H. Lev-Ari and T. Kailath, "State-space approach to factorization of lossless transfer functions and structured matrices," *Lin. Alg. Appl.*, vol. 162, pp. 273–295, Feb. 1992.

[25] A.H. Sayed, T. Kailath, H. Lev-Ari, and T. Constantinescu, "Efficient recursive solutions of rational interpolation problems," *subm. Integral Equations and Operator Theory*, 1992.

[26] A.H. Sayed, T. Constantinescu, and T. Kailath, "Lattice structures for time-variant interpolation problems," in *Proc. 31-st IEEE Conf. on Decision and Control*, (Tuscon, AZ), Dec. 1992.

[27] E.W. Kamen, "The poles and zeros of a linear time-varying system," *Lin. Alg. Appl.*, vol. 98, pp. 263–289, 1988.

Chapter 6

HANKEL-NORM MODEL REDUCTION

6.1 INTRODUCTION

In the previous chapters, we assumed that a given upper operator or matrix T has a computational model of a sufficiently low order to warrant the (computationally expensive) step of deriving for it a computational model in the form of a state realization. Once a state model is known, we showed how multiplication by T or its inverse can be done efficiently, using the model rather than the entries of T. We also derived some useful factorizations, such as the external and inner-outer ($\sim QR$) factorization. A spectral factorization/Cholesky factorization result is given in chapter 8.

However, if the ranks of the sequence of Hankel matrices of T are not sufficiently low, then the system order of the computational model will be large. This can already happen if T is modified only slightly, *e.g.*, caused by numerical imprecisions, as the rank of a matrix is a very sensitive (ill-conditioned) parameter. Hence one wonders whether, for a given $T \in \mathcal{U}$, there is an approximating system T_a close to it such that T_a has a low system order. Such an approximation is useful also when T is known exactly, but if for analysis purposes one would like to work with a low complexity, yet accurate approximating model.

One standard way to find an approximant of a matrix (A, say) goes via the singular value decomposition (SVD). This decomposition yields a diagonal matrix of singular values. Setting those singular values that are smaller than some tolerance level ε equal to zero produces an approximant $\hat{A}$ such that $\| A - \hat{A} \| < \varepsilon$ and rank(A) is equal to the remaining number of non-zero singular values. One can show that the thus-obtained approximant is optimal in the operator norm (matrix 2-norm), and also in the Hilbert-Schmidt norm (matrix Frobenius norm). Since the state complexity of the operator/matrix T is given by the rank sequence of H_T rather than the rank of T itself (corollary 3.14), it seems logical to approximate each H_k by some $\hat{H}_k$ of lower rank. However, the Hankel matrices have many entries in common, and approximating one of them by a matrix of low rank might make it impossible for all other $\hat{H}_k$ to acquire a low rank: a local optimum might prevent

a global one. In this respect, the approximation error norm used is also of importance: the Hilbert-Schmidt (Frobenius) norm is rather strong:

$$\min_{\operatorname{rank}\hat{A}\leq d} \|A-\hat{A}\|_{HS}$$

has only one (unique) solution $\hat{A}$, obtained by setting all but the first d singular values equal to zero, and keeping the first d untouched. The operator norm approximation problem

$$\min_{\operatorname{rank}\hat{A}\leq d} \|A-\hat{A}\|$$

has many solutions, since only the largest singular value of the difference $E = A - \hat{A}$ is minimized, and $d-1$ others are free, as long as they remain smaller. For sequences of Hankel matrices, the extra freedom in each of the $\hat{H}_k$ can be used to reduce the rank of the other H_k. The problem can be described in two ways: by

$$\min_{\operatorname{rank}\hat{H}_k\leq d_k} \|H_k-\hat{H}_k\|, \qquad \text{(for all } k\text{)},$$

which is the *model error reduction problem* for given target ranks d_k, and by

$$\min\{\operatorname{rank}\hat{H}_k : \|H_k-\hat{H}_k\| \leq \varepsilon_k\}, \qquad \text{(for all } k\text{)}, \tag{6.1}$$

the *model order reduction problem* for given tolerance levels ε_k. The latter problem description is the one which we take up in this chapter. The error criterion (6.1) leads to the definition of the *Hankel norm*, which is a generalization of the Hankel norm for time-invariant systems:

$$\|T\|_H = \|H_T\|. \tag{6.2}$$

$\|T\|_H$ is the supremum over the operator norm of each individual Hankel matrix H_k. It is a reasonably strong norm: if T is a strictly upper triangular matrix and $\|T\|_H \leq 1$, then each row and column of T has vector norm smaller than 1. The main approximation theorem that we derive can be stated as follows.

THEOREM 6.1. *Let $T \in \mathcal{U}$, and let $\Gamma = \operatorname{diag}(\gamma_i) \in \mathcal{D}$ be a Hermitian operator. Let H_k be the Hankel operator of $\Gamma^{-1}T$ at stage k, and suppose that an $\varepsilon > 0$ exists such that, for each k, none of the singular values of H_k are in the interval $[1-\varepsilon, 1+\varepsilon]$. Then there exists a strictly upper triangular operator T_a with system order at stage k at most equal to the number of singular values of H_k that are larger than 1, such that*

$$\|\Gamma^{-1}(T-T_a)\|_H \leq 1. \tag{6.3}$$

The error tolerance diagonal Γ parametrizes the problem. As ε in (6.1), it can be used to influence the local approximation error: if $\Gamma = \gamma I$, then $\|T-T_a\|_H \leq \gamma$ and the approximation

error is uniformly distributed over T. If one of the γ_i is made larger than γ, then the error at the i-th row of T can become larger also, which might result in an approximant T_a to take on fewer states. Hence Γ can be chosen to yield an approximant that is accurate at certain points but less tight at others, and whose complexity is minimal.

Although we have seen that, with the same tolerance level, the operator norm allows more freedom than the Hilbert-Schmidt norm, the computational task still seems formidable: there is an infinity of minimization problems, all coupled to each other. It is remarkable that the problem allows a clean and straightforward solution (as we show in this chapter), which can even be obtained in a non-iterative way. The clue is in the fact that the condition (6.3) translates to the computation of contractive operators E, which, as we saw in chapter 5, are automatically obtained by 'loading' a J-unitary operator Θ with a contractive operator S_L. This is the way that J-unitary systems enter into the picture. The general solution using this approach is published in [1], and specializations to finite matrices are made in [2].

Hankel norm approximation theory originates as a special case of the solution to the Schur-Takagi interpolation problem in the context of complex function theory. Suppose that a number of complex values are given at a set of points in the interior of the unit disc of the complex plane, then this problem consists in finding a complex function (*a*) which interpolates these values at the given points (multiplicities counted), (*b*) which is meromorphic with at most k poles inside the unit disc, and (*c*) whose restriction to the unit circle (if necessary via a limiting procedure from inside the unit disc) belongs to L_∞ with minimal norm. The Schur-Takagi problem can be seen as an extension problem whereby the "conjugate-analytic" or anti-causal part of a function is given, and it is desired to extend it to a function which is meromorphic with at most k poles inside the unit disc, and belongs to L_∞ with minimal norm. (Translated into our context, the objective would be to determine an extension of an operator $T^* \in \mathcal{L}\mathcal{Z}^{-1}$ to an operator $T' \in \mathcal{X}$, such that T' is contractive and has an upper part with state dimension sequence smaller than a given sequence.) The solution was formulated by Adamjan, Arov and Krein (AAK)[3], who studied properties of the SVD of infinite dimensional Hankel matrices with Hankel structure, and associated approximation problems of bounded analytical functions by rational functions.

It was remarked in Bultheel-Dewilde [4] and subsequently worked out by a number of authors (Kung-Lin [5], Genin-Kung [6], Ball-Helton [7], Glover [8]) that the procedure of AAK could be utilized to solve the problem of optimal model-order reduction of a dynamical time-invariant system. The computational problem with the general theory is that it involves an operator which maps a Hilbert space of input sequences to a Hilbert space of output sequences, and which is thus intrinsically non-finite. In [4] it was shown that the computations are finite if one assumes the context of a system of finite (but possibly large) degree, *i.e.*, an approximant to the original system of high order. The resulting computations involve only the realization matrices $\{A, B, C, D\}$ of the approxi-

mating system and can be done using classical matrix calculus. They can also be done in a recursive fashion, see Limebeer-Green [9] as a pioneering paper in this respect. The recursive method is based on the interpolation theory of the Schur-Takagi type.

For time-invariant systems, the Hankel-norm model reduction method may be compared with another popular method for model reduction, known as the balanced model reduction method. In these methods, a reduced-order model is obtained by setting all small singular values of the Hankel matrix equal to zero, and using the resulting truncated column space and row space in the construction of a state model [5]. Alternatively, one may start from a high-order balanced model (one for which the controllability and observability Gramians are diagonal and equal to each other), and delete all states variables that correspond to small entries in the Gramians [10, 11]. These methods also give good approximation results, although no tight upper bounds on the modeling error have been derived. An extensive study on error bounds was made by Glover [8], and by Glover-Curtain-Partington [12] for the infinite-dimensional time-invariant case.

Connections between circuit and system theory problems and the mathematical techniques around interpolation, reproducing kernels and the lifting of a contractive operator had been obtained a decade earlier by Helton [13] in the pursuit of a solution to the broadband matching problem (see also [14]). Subsequently, more connections between (definite) interpolation problems, reproducing kernel Hilbert spaces and the global and recursive solutions to the Lossless Inverse Scattering problem were studied in [15, 16, 17, 18], and collected in the monograph [19] by Dym. The recursive solution of the Schur-Takagi problem by Limebeer and Green [9] can be viewed as an extension of such results to meromorphic (indefinite) interpolation problems. In a parallel development, the state space theory for the interpolation problem was extensively studied in the book [20] by Ball, Gohberg and Rodman. The wide interest in this type of problems was kindled by one of its many applications: the robust (H_∞-) control problem formulated by Zames in [21] and brought into the context of scattering and interpolation theory by Helton [22].

Numerical example

As an example of the use of theorem 6.1, we consider a matrix T and determine an approximant T_a. Let the matrix to be approximated be

$$T = \left[\begin{array}{cccc|cc} 0 & .800 & .200 & .050 & .013 & .003 \\ 0 & 0 & .600 & .240 & .096 & .038 \\ 0 & 0 & 0 & .500 & .250 & .125 \\ \hline 0 & 0 & 0 & 0 & .400 & .240 \\ 0 & 0 & 0 & 0 & 0 & .300 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

The position of the Hankel matrix H_4 is indicated. Taking $\Gamma = 0.1\,I$, the non-zero singular values of the Hankel operators of $\Gamma^{-1}T$ are

H_1	H_2	H_3	H_4	H_5	H_6
	8.26	6.85	6.31	5.53	4.06
		0.33	0.29	0.23	
			0.01		

Hence T has a state-space realization which grows from zero states ($i = 1$) to a maximum of 3 states ($i = 4$), and then shrinks back to 0 states ($i > 6$). The number of Hankel singular values of $\Gamma^{-1}T$ that are larger than one is 1 ($i = 2, \cdots, 6$). This is to correspond to the number of states of the approximant at each point. Using the technique detailed in this chapter, we obtain

$$T_a = \begin{bmatrix} 0 & .790 & .183 & .066 & .030 & .016 \\ 0 & 0 & .594 & .215 & .098 & .052 \\ 0 & 0 & 0 & .499 & .227 & .121 \\ 0 & 0 & 0 & 0 & .402 & .214 \\ 0 & 0 & 0 & 0 & 0 & .287 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

with non-zero Hankel singular values (scaled by Γ)

H_1	H_2	H_3	H_4	H_5	H_6
	8.15	6.71	6.16	5.36	3.82

The number of non-zero singular values indeed corresponds to the number of Hankel singular values of $\Gamma^{-1}T$ that are larger than 1. The modeling error is

$$T - T_a = \begin{bmatrix} 0 & .010 & .017 & -.016 & -.017 & -.013 \\ 0 & 0 & .006 & .025 & -.002 & -.014 \\ 0 & 0 & 0 & .001 & .023 & .004 \\ 0 & 0 & 0 & 0 & -.002 & .026 \\ 0 & 0 & 0 & 0 & 0 & .013 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

and indeed, the Hankel norm of $\Gamma^{-1}(T - T_a)$ is less than 1:

$$\| \Gamma^{-1}(T - T_a) \|_H = \sup\{0.334,\ 0.328,\ 0.338,\ 0.351,\ 0.347\} = 0.351$$

The realization algorithm (algorithm 3.1) yields as realization for T

$$\mathbf{T}_1 = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline -.826 & 0 \end{array}\right] \qquad \mathbf{T}_2 = \left[\begin{array}{cc|c} .246 & -.041 & -.968 \\ \hline -.654 & -.00 & 0 \end{array}\right]$$

$$\mathbf{T}_3 = \left[\begin{array}{ccc|c} .397 & -.044 & .000 & -.917 \\ .910 & .140 & .040 & .388 \\ \hline -.573 & .00 & .00 & 0 \end{array}\right] \qquad \mathbf{T}_4 = \left[\begin{array}{cc|c} .487 & .037 & -.873 \\ .853 & -.237 & .465 \\ .189 & .971 & .147 \\ \hline -.466 & .00 & 0 \end{array}\right]$$

$$\mathbf{T}_5 = \left[\begin{array}{c|c} -.515 & -.858 \\ .858 & -.515 \\ \hline .300 & 0 \end{array}\right] \qquad \mathbf{T}_6 = \left[\begin{array}{c|c} \cdot & 1 \\ \hline \cdot & 0 \end{array}\right]$$

A realization of the approximant is determined via algorithm 6.2 in section 6.3 as

$$\mathbf{T}_{a,1} = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline -.993 & 0 \end{array}\right] \qquad \mathbf{T}_{a,2} = \left[\begin{array}{c|c} .293 & -.795 \\ \hline -.946 & 0 \end{array}\right]$$

$$\mathbf{T}_{a,3} = \left[\begin{array}{c|c} .410 & -.629 \\ \hline -.901 & 0 \end{array}\right] \qquad \mathbf{T}_{a,4} = \left[\begin{array}{c|c} .525 & -.554 \\ \hline -.837 & 0 \end{array}\right]$$

$$\mathbf{T}_{a,5} = \left[\begin{array}{c|c} -.651 & -.480 \\ \hline .729 & 0 \end{array}\right] \qquad \mathbf{T}_{a,6} = \left[\begin{array}{c|c} \cdot & .393 \\ \hline \cdot & 0 \end{array}\right]$$

The corresponding computational schemes are depicted in figure 6.1, to show the effect that a small change in T can lead to a significant reduction in the complexity of the computations.

Hankel norm

As mentioned in the introduction, we compute approximants which are optimal in the Hankel norm, defined as

$$\| T \|_H = \| H_T \| .$$

It is a norm on $\mathcal{U}$, a semi-norm on $\mathcal{Y}$. Since this is not such a familiar norm as, for example, the operator norm of T, we first determine its relation to this norm. The Hankel norm can also be compared to another norm, the *diagonal 2-norm*. Let T_i be the i-th row of a block matrix representation of $T \in \mathcal{Y}$, then

$$D \in \mathcal{D} : \ \| D \|_{\mathcal{D}2} = \| D \| = \sup_i \| D_i \|,$$
$$T \in \mathcal{Y} : \ \| T \|_{\mathcal{D}2}^2 = \| \mathbf{P}_0(TT^*) \|_{\mathcal{D}2} = \sup_i \| T_i T_i^* \| .$$

For diagonals, it is equal to the operator norm, but for more general operators, it is the supremum over the vector 2-norms of each row of T.

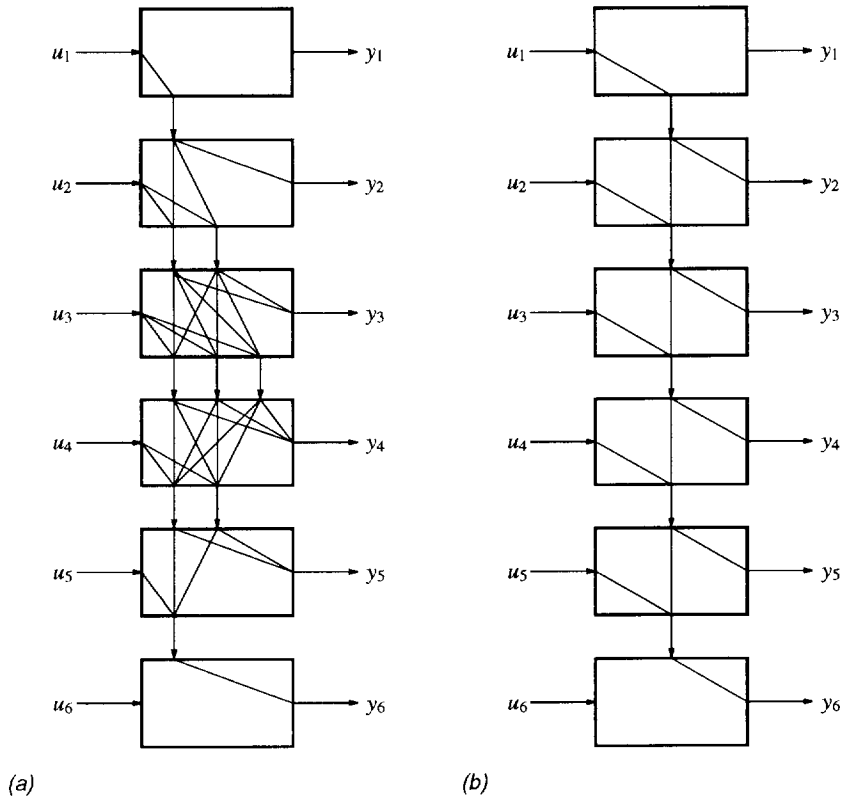


Figure 6.1. Computational scheme (a) of T and (b) of T_a.

PROPOSITION 6.2. *The Hankel norm satisfies the following ordering:*

$$T \in \mathcal{X}: \qquad \|T\|_H \leq \|T\| \tag{6.4}$$

$$T \in \mathcal{ZU}: \qquad \|T\|_{\mathcal{D}2} \leq \|T\|_H. \tag{6.5}$$

PROOF The first norm inequality is proven by

$$\begin{aligned} \|T\|_H &= \sup_{u \in \mathcal{L}_2 Z^{-1}, \|u\|_{HS} \leq 1} \|\mathbf{P}(uT)\|_{HS} \\ &\leq \sup_{u \in \mathcal{L}_2 Z^{-1}, \|u\|_{HS} \leq 1} \|uT\|_{HS} \\ &\leq \sup_{u \in \mathcal{X}_2, \|u\|_{HS} \leq 1} \|uT\|_{HS} = \|T\|. \end{aligned}$$

For the second norm inequality, we first prove $\|T\|_{\mathcal{D}2}^2 \leq \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|DTT^*D^*\|_{HS}$. Indeed,

$$\begin{aligned} \|T\|_{\mathcal{D}2}^2 &= \|\mathbf{P}_0(TT^*)\|_{\mathcal{D}2}^2 \\ &= \sup_{D \in \mathcal{D}_2, \|D\|_{\mathcal{D}2} \leq 1} \|D\mathbf{P}_0(TT^*)D^*\|_{\mathcal{D}2} \\ &= \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|D\mathbf{P}_0(TT^*)D^*\|_{HS} \\ &\leq \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|DTT^*D^*\|_{HS}. \end{aligned}$$

Then (6.5) follows, with use of the fact that $T \in \mathcal{ZU}$, by

$$\begin{aligned} \|T\|_{\mathcal{D}2}^2 &\leq \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|DTT^*D^*\|_{HS} \\ &= \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|DZ^*TT^*ZD^*\|_{HS} \\ &= \sup_{D \in \mathcal{D}_2, \|D\|_{HS} \leq 1} \|\mathbf{P}(DZ^*T)\left[\mathbf{P}(DZ^*T)\right]^*\|_{HS} \\ &\leq \sup_{u \in \mathcal{L}_2 Z^{-1}, \|u\|_{HS} \leq 1} \|\mathbf{P}(uT)\left[\mathbf{P}(uT)\right]^*\|_{HS} \\ &= \|T\|_H^2. \end{aligned}$$

□

We see that the Hankel norm is not as strong as the operator norm, but is stronger than the row-wise uniform least square norm.

6.2 APPROXIMATION VIA INDEFINITE INTERPOLATION

Approximation recipe

In the present section we outline a procedure to obtain a reduced-order approximant, and put the various relevant facts in perspective. Details are proven in subsequent sections.

Let $T \in \mathcal{U}$ be a given bounded, locally finite, strictly upper operator. The decision to assume T strictly upper is made for the sake of convenience and is without serious consequences: $D = \mathbf{P}_0(T)$ has no influence on the Hankel (semi-)norm, so that there are no conditions on the D operator of the approximant. Let $\Gamma \in \mathcal{D}$ be a diagonal and Hermitian operator. As discussed in the introduction, the objective is to determine an

operator $T_a \in \mathcal{U}$ such that $\| \Gamma^{-1}(T - T_a) \|_H \leq 1$. Instead of working with T_a directly, we look for a bounded operator $T' \in \mathcal{X}$ such that

$$\| \Gamma^{-1}(T - T') \| \leq 1, \tag{6.6}$$

and such that the strictly upper part of T' has state-space dimensions of low order — as low as possible for a given Γ. Let T_a be the strictly causal part of T'. Proposition 6.2 showed that

$$\begin{aligned} \| \Gamma^{-1}(T - T_a) \|_H &= \| \Gamma^{-1}(T - T') \|_H \\ &\leq \| \Gamma^{-1}(T - T') \| \leq 1, \end{aligned} \tag{6.7}$$

so that T_a is a Hankel-norm approximant of T (parametrized by Γ) whenever T' is an operator-norm approximant. T' can be viewed as an extension of T_a which is such that $\| \Gamma^{-1}(T - T_a) \|_H \leq \| \Gamma^{-1}(T - T') \|$. A generalization of Nehari's theorem to the present setting would state that $\inf \| E \|$ over all possible extensions $E \in \mathcal{X}$ of a given part $E_a \in \mathcal{U}$ actually equals $\| E_a \|_H$ (see section 6.5).

The construction of an operator T' satisfying (6.6) consists of three steps, specified in the following lemma. (The definitions and notation in this lemma will be kept throughout the rest of the section.)

LEMMA 6.3. (RECIPE FOR A HANKEL-NORM APPROXIMANT) *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be strictly upper, and let $\Gamma \in \mathcal{D}(\mathcal{M}, \mathcal{M})$ be a given diagonal Hermitian operator. Then, provided the indicated factorizations exist, an operator $T' \in \mathcal{X}$ such that $\| \Gamma^{-1}(T - T') \| \leq 1$ is obtained by performing the following steps:*

1. *an external factorization (inner-coprime factorization; theorem 4.9):*

$$T = \Delta^* U \qquad (U, \Delta \in \mathcal{U}, U \text{ unitary}), \tag{6.8}$$

2. *a J-inner coprime factorization (corollary 5.20):*

$$[U^* \quad -T^*\Gamma^{-1}]\Theta = [A' \quad -B'] \in [\mathcal{U} \quad \mathcal{U}] \qquad (\Theta \in \mathcal{U}, \ J\text{-unitary}), \tag{6.9}$$

3. *with a block-decomposition of Θ as in (5.2),*

$$T'^* = B'\Theta_{22}^{-1}\Gamma. \tag{6.10}$$

PROOF If the factorizations exist, then Θ_{22} is boundedly invertible so that $\Sigma_{12} = -\Theta_{12}\Theta_{22}^{-1}$ exists and is contractive (theorem 5.2). From (6.9) we have $B' = -U^*\Theta_{12} + T^*\Gamma^{-1}\Theta_{22}$. Substitution of (6.10) leads to

$$\begin{aligned} T'^*\Gamma^{-1} &= T^*\Gamma^{-1} - U^*\Theta_{12}\Theta_{22}^{-1} \\ &= T^*\Gamma^{-1} - U^*\Sigma_{12} \end{aligned}$$

and it follows that $(T^* - T'^*)\Gamma^{-1} = -U^*\Sigma_{12}$. Because Σ_{12} is contractive and U unitary,

$$\begin{aligned} \| (T^* - T'^*)\Gamma^{-1} \| &= \| -U^*\Sigma_{12} \| \\ &= \| \Sigma_{12} \| \leq 1 , \end{aligned}$$

so that $T' = \left(B'\Theta_{22}^{-1}\Gamma\right)^*$ is indeed an approximant with an admissible modeling error. □

In anticipation of a proof of theorem 6.1, it remains to show that the strictly upper part T_a of T' has at most the specified number of states, and to verify the relation with the Hankel singular values of $\Gamma^{-1}T$. This is done in the remaining part of this section. The definition of T' in (6.10) can be generalized by the introduction of a contractive operator S_L that parametrizes the possible approximants, which is the subject of section 6.4. The crucial step in the procedure is step 2. As discussed in section 5.3, the computation of Θ can be viewed as the solution of an interpolation problem

$$U^*[I \quad S]\Theta \in [\mathcal{U} \quad \mathcal{U}], \qquad S = -UT^*\Gamma^{-1} = -\Delta\Gamma^{-1}, \tag{6.11}$$

where the interpolation subspace is determined by U. If $\Theta_{22}^{-1} \in \mathcal{U}$, then an exact representation of S in Θ is obtained as $S = -\Theta_{12}\Theta_{22}^{-1}$. In this case, the interpolation problem is definite: the relevant J-Gramian is positive definite, which happens if $\Gamma^{-1}T$ is strictly contractive. In addition, $T'^* = B'\Theta_{22}^{-1}\Gamma$ is upper, and the approximant T_a is zero, which matches one's expectation in view of $\| \Gamma^{-1}T \| < 1$. If $\Gamma^{-1}T$ is not contractive then Θ_{22}^{-1} is not upper, and this is the situation which leads to approximations and which is considered in this chapter.

Construction of Θ

We now determine sufficient conditions on a state-space realization $\{A, B, C, 0\}$ of T for the existence of the two factorizations in the above lemma. Assuming $\ell_A < 1$, the external factorization in the first step can be computed from the given realization if it is uniformly observable (theorem 4.9). Without loss of generality, we can (and do) assume that such a realization has been normalized, so that $AA^* + CC^* = I$. Then, a realization for the inner factor U of the external factorization is given by

$$\mathbf{U} = \left[\begin{array}{cc} A & C \\ B_U & D_U \end{array} \right]$$

where B_U and D_U are obtained by locally completing $[A_k \quad C_k]$ to a square and unitary matrix.

The second step is to derive expressions for Θ to satisfy the interpolation condition (6.9). $[U^* \quad -T^*\Gamma^{-1}]^*$ has a realization

$$\left[\begin{array}{c} U \\ -\Gamma^{-1}T \end{array} \right] = \left[\begin{array}{c} D_U \\ 0 \end{array} \right] + \left[\begin{array}{c} B_U \\ -\Gamma^{-1}B \end{array} \right] Z(I - AZ)^{-1}C ,$$

so that, according to corollary 5.20, there exists a J-unitary operator Θ mapping $[U^* \quad -T^*\Gamma^{-1}]$ to upper if the relevant J-Gramian $\Lambda := \Lambda^J$ (as defined in (5.8)) is boundedly invertible. With the above realization of $[U^* \quad -T^*\Gamma^{-1}]^*$, Λ satisfies the J-Lyapunov equation (*cf.* equation (5.36))

$$\Lambda^{(-1)} = A^*\Lambda A + B_U^* B_U - B^*\Gamma^{-2}B.$$

Substituting the relation $A^*A + B_U^*B_U = I$ yields $I - \Lambda^{(-1)} = A^*(I-\Lambda)A + B^*\Gamma^{-2}B$. With the additional definition of $M = I - \Lambda$, it is seen that M satisfies

$$M^{(-1)} = A^*MA + B^*\Gamma^{-2}B$$

so that M is the controllability Gramian of the given realization of $\Gamma^{-1}T$. It follows that the J-inner coprime factorization exists if $I - M$ is boundedly invertible, that is, if 1 is a regular point for the operator M [23]. With M known (and hence Λ), Θ is determined along the lines of the proof of theorem 5.19. In particular, the input state space of Θ is defined by

$$\mathcal{H}(\Theta) = \mathcal{D}_2^{\mathcal{B}}\,(I - Z^*A^*)^{-1}Z^* \begin{bmatrix} B_U^* & B^*\Gamma^{-1} \end{bmatrix}. \tag{6.12}$$

Let $\Lambda = R^* J_{\mathcal{B}} R$ be a factorization of Λ, then

$$\begin{bmatrix} A_\Theta \\ B_\Theta \end{bmatrix} = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} A \\ B_U \\ \Gamma^{-1}B \end{bmatrix} R^{(-1)}$$

is J-isometric, and a J-unitary realization for Θ is of the form

$$\boldsymbol{\Theta} = \begin{bmatrix} A_\Theta & C_\Theta \\ B_\Theta & D_\Theta \end{bmatrix} = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} A & C_1 & C_2 \\ B_U & D_{11} & D_{12} \\ \Gamma^{-1}B & D_{21} & D_{22} \end{bmatrix} \begin{bmatrix} R^{(-1)} & & \\ & I & \\ & & I \end{bmatrix} \tag{6.13}$$

and is obtained by completing A_Θ and B_Θ with certain diagonal operators C_Θ and D_Θ to a square J-unitary matrix. Corollary 5.20 claims that this is always possible under the present conditions (Λ boundedly invertible), and the procedure to do so is given in lemma 5.18. Since the realization $\boldsymbol{\Theta}$ is J-unitary, the corresponding transfer operator Θ is also J-unitary and has the specified input state space. The third step in lemma 6.3 is always possible (*cf.* theorem 5.2).

We have proven the following lemma:

LEMMA 6.4. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ be a strictly upper locally finite operator, with output normal realization $\{A, B, C, 0\}$ such that $\ell_A < 1$, and let Γ be a Hermitian diagonal operator. If the solution M of the Lyapunov equation*

$$M^{(-1)} = A^*MA + B^*\Gamma^{-2}B \tag{6.14}$$

is such that $\Lambda = I - M$ *is boundedly invertible, then the conditions mentioned in lemma 6.3 are satisfied: there exists an external factorization* $T = \Delta^* U$*, a J-unitary block upper operator* Θ *such that*

$$[U^* \quad -T^*\Gamma^{-1}]\Theta \in [\mathcal{U} \quad \mathcal{U}],$$

and an operator $T' \in \mathcal{X}$ *such that* $\|\Gamma^{-1}(T - T')\| \leq 1$*, according to the recipe in lemma 6.3.*

Let $\mathcal{M}$, $\mathcal{N}$ and $\mathcal{B}$ be the input, output and state space sequences of T and its realization, and let $\mathcal{M}_U$ be the input space sequence for U: its index sequence is specified by

$$\#\mathcal{M}_U = \#\mathcal{B}^{(-1)} - \#\mathcal{B} + \#\mathcal{N}.$$

The signature $J_\mathcal{B}$ of Λ determines a decomposition of $\mathcal{B}$ into $\mathcal{B} = \mathcal{B}_+ \times \mathcal{B}_-$. Let $\Theta^* J_1 \Theta = J_2$, $\Theta J_2 \Theta^* = J_1$, where J_1 and J_2 are shorthand for $J_1 = J_{\mathcal{M}_\Theta}$ and $J_2 = J_{\mathcal{N}_\Theta}$. The space sequence $\mathcal{M}_\Theta$ is equal to $\mathcal{M}_\Theta = \mathcal{M}_U \times \mathcal{M}$, and the corresponding signature operator J_1 follows this partitioning. The dimensions of the positive and negative parts of the output sequence space of Θ, and hence the signature J_2, are then given by inertia rules as (*cf.* corollary 5.20)

$$\begin{aligned} \#(\mathcal{N}_\Theta)_+ &= \#\mathcal{B}_+ - \#\mathcal{B}_+^{(-1)} + \#\mathcal{M}_U \\ \#(\mathcal{N}_\Theta)_- &= \#\mathcal{B}_- - \#\mathcal{B}_-^{(-1)} + \#\mathcal{M} \end{aligned}$$

Algorithm 6.1 summarizes the construction in lemma 6.4 and can be used to compute Θ satisfying equation (6.9). The inner factor U of T is computed *en passant.*

Connection with the Hankel operator

We continue by establishing the link between the Lyapunov equation (6.14) and the Hankel operator of $\Gamma^{-1}T$.

LEMMA 6.5. *Let* $T \in \mathcal{U}$ *be a locally finite strictly upper operator, with strictly stable realization* $\{A, B, C, 0\}$ *in output normal form. Let* H_k *be the Hankel operator of* $\Gamma^{-1}T$ *at stage k, and suppose that an* $\varepsilon > 0$ *exists such that, for each k, none of the singular values of* H_k *are in the interval* $[1 - \varepsilon, 1 + \varepsilon]$*. Let* N_k *be equal to the number of singular values of* H_k *that are larger than 1. Then the solution M of the Lyapunov equation*

$$M^{(-1)} = A^* M A + B^* \Gamma^{-2} B \tag{6.15}$$

is such that $\Lambda = I - M$ *is boundedly invertible and has a signature operator* $J_\mathcal{B}$ *with* N_k *negative entries at point k.*

PROOF The solutions of the two Lyapunov equations associated to the realization of $\Gamma^{-1}T$ (corresponding to the controllability and observability Gramians),

$$\begin{aligned} M^{(-1)} &= A^* M A + B^* \Gamma^{-2} B \\ Q &= A Q^{(-1)} A^* + C C^* \end{aligned}$$

In: $\mathbf{T}$ (model in output normal form for a strictly upper matrix T)
Γ (approximation parameters)
Out: $\boldsymbol{\Theta}$ (realization for Θ satisfying (6.9))

$$M_1 = [\cdot]$$
$$R_1 = [\cdot]$$
$$J_{\mathcal{B}_1} = [\cdot]$$

for $k = 1, \cdots, n$

$$M_{k+1} = A_k^* M_k A_k + B_k^* \Gamma_k^{-2} B_k$$
$$R_{k+1}^* J_{\mathcal{B}_{k+1}} R_{k+1} := I - M_{k+1}$$
$$[B_{U,k} \;\; D_{U,k}] = [A_k \;\; C_k]^\perp$$
$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} R_k A_k \\ \hline B_{U,k} \\ \Gamma_k^{-1} B_k \end{bmatrix} R_{k+1}^{-1}$$
$$\begin{bmatrix} c \\ d \end{bmatrix} = \begin{bmatrix} J_{\mathcal{B}_k} \alpha \\ J_1 \beta \end{bmatrix}^\perp$$
$$r^* J_2 r := [c \;\; d]^* \begin{bmatrix} J_{\mathcal{B}_k} & \\ & J_1 \end{bmatrix} \begin{bmatrix} c \\ d \end{bmatrix}$$
$$\begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \begin{bmatrix} c \\ d \end{bmatrix} r^{-1}$$
$$\boldsymbol{\Theta}_k = \begin{bmatrix} \alpha & \gamma \\ \beta & \delta \end{bmatrix}$$

end

Algorithm 6.1. Indefinite interpolation: step 1 and 2 of lemma 6.3.

may be expressed in terms of the controllability and observability operators of $\Gamma^{-1}T$,

$$\mathcal{C} := \begin{bmatrix} (\Gamma^{-1}B)^{(+1)} \\ (\Gamma^{-1}B)^{(+2)}A^{(+1)} \\ (\Gamma^{-1}B)^{(+3)}A^{(+2)}A^{(+1)} \\ \vdots \end{bmatrix} \qquad \mathcal{O} := \begin{bmatrix} C & AC^{(-1)} & AA^{(-1)}C^{(-2)} & \cdots \end{bmatrix}$$

as $M = \mathcal{C}^*\mathcal{C}$, $Q = \mathcal{O}\mathcal{O}^*$. The Hankel operator H_k of $\Gamma^{-1}T$ at time instant k satisfies the decomposition $H_k = \mathcal{C}_k\mathcal{O}_k$. Hence

$$H_kH_k^* = \mathcal{C}_k\mathcal{O}_k\mathcal{O}_k^*\mathcal{C}_k^*.$$

The state realization of T is assumed to be in output normal form, so that $Q_k = \mathcal{O}_k\mathcal{O}_k^* = I$. With the current locally finiteness assumption, the non-zero eigenvalues of $H_kH_k^* = \mathcal{C}_k\mathcal{C}_k^*$ are the same as those of $\mathcal{C}_k^*\mathcal{C}_k = M_k$. In particular, the number of singular values of H_k that are larger than 1 is equal to the number of eigenvalues of M_k that are larger than 1. Writing $\Lambda_k = I - M_k$, this is in turn equal to the number of negative eigenvalues of Λ_k. □

Figure 6.2 shows a simple instance of the application of the theory developed in this section, emphasizing the dimensions of the input, output and state space sequences related to the Θ operator. We assume in the figure that one singular value of the Hankel operator of $\Gamma^{-1}T$ at time 1 is larger than 1, so that the state signature $J_{\mathcal{B}}$ of Θ has one negative entry in total. We known from equation (5.16) that the negative entries of $J_{\mathcal{B}}$ determine the number of upward arrows in the diagram of the unitary scattering operator $\mathbf{\Sigma}$. We show, in the following subsection, that this number also determines the number of states of the Hankel-norm approximant T_a of T.

Complexity of the approximant

At this point we have proven the first part of theorem 6.1: we have constructed a J-unitary operator Θ and from it an operator T' with strictly upper part T_a which is a Hankel-norm approximant of T. It remains to verify the complexity assertion, which stated that the sequence of dimensions of the state space of T_a is at most equal to the sequence N: the number of Hankel singular values of $\Gamma^{-1}T$ that are larger than 1. In view of lemmas 6.4 and 6.5, N is equal to the number of negative entries in the state signature $J_{\mathcal{B}}$ of Θ. We now show that the state dimension sequence of T_a is smaller than or equal to N. (Later, we will show that equality holds.) The proof is, again, based on the determination of the natural input state space for T_a, which can be derived in terms of the realization of the scattering operator Σ that is connected to Θ.

Suppose that the conditions of lemma 6.3 are fulfilled so that Θ satisfies

$$\begin{bmatrix} U^* & -T^*\Gamma^{-1} \end{bmatrix} \Theta = \begin{bmatrix} A' & -B' \end{bmatrix}$$

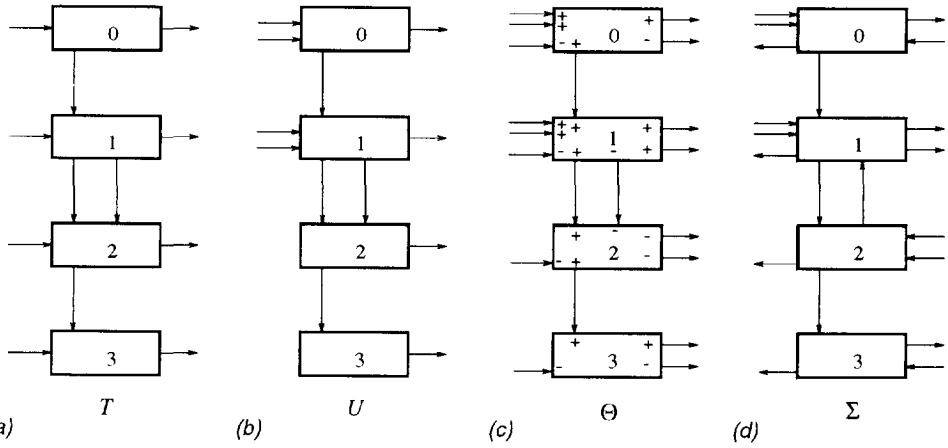


Figure 6.2. (*a*) State-space realization scheme for T and (*b*) for U. (*c*) State-space realization scheme for a possible Θ, where it is assumed that one singular value of the Hankel operator of $\Gamma^{-1}T$ at time 1 is larger than 1, and (*d*) for the corresponding scattering operator Σ.

with A', $B' \in \mathcal{U}$. Let $T'^*\Gamma^{-1} = B'\Theta_{22}^{-1}$ as in lemma 6.3. The approximating transfer function T_a is, in principle, given by the strictly upper part of T' (see lemma 6.3 for the summary of the procedure). It might not be a bounded operator, since operators in $\mathcal{X}$ do not necessarily have a decomposition into an upper and lower part in $\mathcal{X}$. However, its extension T' is bounded, and hence its Hankel operator $H_{T_a} = H_{T'}$ is well defined and bounded. We have the following lemma.

LEMMA 6.6. *Under the conditions of lemma 6.4, the natural input state space of* $\Gamma^{-1}T_a$ *satisfies*

$$\mathcal{H}(\Gamma^{-1}T_a) \subset \mathcal{H}(\Theta_{22}^{-*}). \tag{6.16}$$

PROOF From the definition of $\mathcal{H}$ in equation (3.45) and the operators we have

$$\begin{aligned}
\mathcal{H}(\Gamma^{-1}T_a) &= \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T_a^*\Gamma^{-1}) \\
&= \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T'^*\Gamma^{-1}) \\
&= \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2B'\Theta_{22}^{-1}) \\
&\subset \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2\Theta_{22}^{-1}) \qquad [\text{since } B' \in \mathcal{U}] \\
&= \mathcal{H}(\Theta_{22}^{-*}).
\end{aligned}$$

□

Hence the sequence of dimensions of the subspace $\mathcal{H}(\Theta_{22}^{-*})$ is of interest. According to proposition 5.16, this dimension sequence is equal to $N = \#(\mathcal{B}_-)$, *i.e.*, the number of negative entries in the state signature sequence $J_{\mathcal{B}}$ of Θ. Combining this result with the lemmas in this section proves the model reduction theorem, theorem 6.1, repeated below:

THEOREM 6.7. *Let $T \in \mathcal{U}$ be a locally finite strictly upper operator with a uniformly observable strictly stable realization, and let $\Gamma = \mathrm{diag}(\gamma) \in \mathcal{D}$ be a Hermitian operator. Let H_k be the Hankel operator of $\Gamma^{-1}T$ at stage k, and suppose that an $\varepsilon > 0$ exists such that, for each k, none of the singular values of H_k are in the interval $[1-\varepsilon, 1+\varepsilon]$. Then there exists a strictly upper triangular operator T_a with system order at stage k at most equal to the number of singular values of H_k that are larger than 1, such that*

$$\| \Gamma^{-1}(T - T_a) \|_H \leq 1 .$$

PROOF Under the present conditions on T, lemma 6.3 can be applied. Indeed, lemma 6.5 claims that the controllability Gramian M of the realization (normalized to output normal form) is such that $\Lambda = I - M$ is boundedly invertible, where Λ satisfies the same J-Lyapunov equation as in lemma 6.4. This lemma showed that the necessary conditions to apply the procedure in lemma 6.3 are satisfied. Thus construct T' and T_a using lemma 6.3, so that $\| \Gamma^{-1}(T - T_a) \|_H \leq 1$. According to lemma 6.6, the state dimension sequence of T_a is less than or equal to the state dimension sequence of the causal part of Θ_{22}^{-*}, which is equal to the number of negative entries of the state signature sequence $J_{\mathcal{B}}$ (proposition 5.16), in turn equal to N (lemma 6.5). Hence T_a has the claimed state complexity, so that it is a Hankel norm approximant of T for the given Γ. □

6.3 STATE REALIZATION OF THE APPROXIMANT

Theorem 6.7 shows the existence of a Hankel norm approximant T_a under certain conditions. The proof uses a construction of this approximant (lemma 6.3), but this construction is at the operator level. However, it is also possible to obtain a *state realization* for T_a directly. We will derive this result in the present section.

Throughout this section, we take signals a_1, a_2, b_1, b_2 to be elements of $\mathcal{X}_2$, generically related by

$$[a_1 \quad b_1]\Theta = [a_2 \quad b_2]$$

where Θ is as constructed in the previous section. In particular, Θ is a bounded operator, and Θ_{22}^{-1} exists and is bounded. In section 6.2 we constructed Θ via a J-unitary realization $\boldsymbol{\Theta}$, with state signature matrix $J_{\mathcal{B}}$. $\boldsymbol{\Theta}$ is bounded by construction (because of the assumption that none of the Hankel singular values of $\Gamma^{-1}T$ are equal or 'asymptotically

close' to 1), and is strictly stable because T is assumed to be so. As before, the part of an operator $u \in \mathcal{X}_2$ that is in $\mathcal{L}_2 Z^{-1}$ is denoted by $u_p = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(u)$, and the part in $\mathcal{U}_2$ is $u_f = \mathbf{P}(u)$. Associated to the transfer operator Θ is the scattering operator Σ which relates

$$[a_1 \quad b_1]\Theta = [a_2 \quad b_2] \quad \Leftrightarrow \quad [a_1 \quad b_2]\Sigma = [a_2 \quad b_1].$$

We have derived in theorem 5.2 a representation $\mathbf{\Sigma} = \{F, G, H, K\}$ in terms of entries $\{A_\Theta, B_\Theta, C_\Theta, D_\Theta\}$ in $\mathbf{\Theta}$, according to the relation

$$\begin{aligned} [x_+ \quad x_- \quad a_1 \quad b_1]\mathbf{\Theta} &= [x_+ Z^{-1} \quad x_- Z^{-1} \quad a_2 \quad b_2] \\ [x_+ \quad x_- Z^{-1} \quad a_1 \quad b_2]\mathbf{\Sigma} &= [x_+ Z^{-1} \quad x_- \quad a_2 \quad b_1]. \end{aligned}$$

The above realizations act on operators in $\mathcal{X}_2$. Taking the k-th diagonal of each operator yields the following state recursions on diagonals, which we use throughout the section:

$$\begin{aligned} [x_{+[k]} \quad x_{-[k]} \quad a_{1[k]} \quad b_{1[k]}]\mathbf{\Theta} &= [x^{(-1)}_{+[k+1]} \quad x^{(-1)}_{-[k+1]} \quad a_{2[k]} \quad b_{2[k]}] \\ [x_{+[k]} \quad x^{(-1)}_{-[k+1]} \quad a_{1[k]} \quad b_{2[k]}]\mathbf{\Sigma} &= [x^{(-1)}_{+[k+1]} \quad x_{-[k]} \quad a_{2[k]} \quad b_{1[k]}]. \end{aligned}$$

In order to compute a realization of T_a, we first determine a model for the strictly upper part of Θ_{22}^* from the model $\mathbf{\Sigma}$. It is given in terms of operators S and R defined as[1]

$$\begin{aligned} x_{-[0]}S &= x_{+[0]} \quad \text{when } a_{1p} = 0,\ b_{2p} = 0 \\ x_{+[0]}R &= x_{-[0]} \quad \text{when } a_{1f} = 0,\ b_{2f} = 0, \end{aligned} \tag{6.17}$$

which can be obtained from $\mathbf{\Sigma}$ in terms of two recursive equations. S is, for example, obtained as the input scattering matrix of a ladder network consisting of a semi-infinite chain of contractive (*i.e.*, lossy) scattering matrices F_{ij}.

LEMMA 6.8. *The relations*

$$\begin{aligned} x_{-[0]}S &= x_{+[0]} \quad \textit{when } a_{1p} = 0,\ b_{2p} = 0 \\ x_{+[0]}R &= x_{-[0]} \quad \textit{when } a_{1f} = 0,\ b_{2f} = 0, \end{aligned} \tag{6.18}$$

define bounded maps which are strictly contractive: $\|S\| < 1$, $\|R\| < 1$.

PROOF S exists as a partial map of Σ_p, taking $a_{1p} = 0$, $b_{2p} = 0$. In this situation,

$$[0 \quad b_{1p}]\Theta_p = [x_{+[0]} \quad x_{-[0]} \quad a_{2p} \quad 0],$$

and we have

$$\|x_{-[0]}\|^2 = \|x_{+[0]}\|^2 + \|b_{1p}\|^2 + \|a_{2p}\|^2.$$

[1] Here, S is not the same as S in (6.11), and no connection is intended.

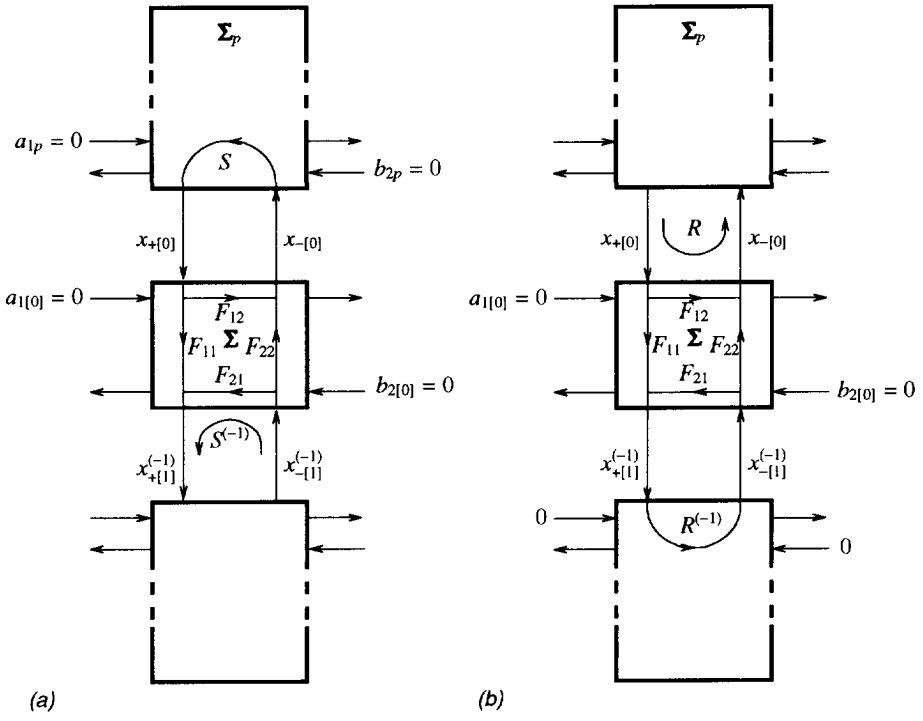


Figure 6.3. (*a*) The propagation of S, (*b*) the propagation of R.

According to proposition 5.15, there is an ε, $0 < \varepsilon \leq 1$, such that $\| b_{1p} \|^2 \geq \varepsilon^2 \| x_{-[0]} \|^2$, and hence

$$\| x_{-[0]} \|^2 \geq \| x_{+[0]} \|^2 + \varepsilon^2 \| x_{-[0]} \|^2 .$$

Consequently, there is a constant μ $(0 \leq \mu < 1)$ such that $\| x_{+[0]} \|^2 \leq \mu^2 \| x_{-[0]} \|^2$ (take $\mu = \sqrt{1-\varepsilon^2}$). This shows that $\| S \| < 1$. A similar argument holds for R. □

PROPOSITION 6.9. *The operators S and R defined in (6.18) are determined in terms of $\boldsymbol{\Sigma}$ (with block entries as in 5.17) by the following recursions:*

$$\begin{aligned} S &= \left(F_{21} + F_{22}(I - SF_{12})^{-1} SF_{11} \right)^{(+1)} \\ R &= F_{12} + F_{11}(I - R^{(-1)}F_{21})^{-1} R^{(-1)} F_{22} . \end{aligned} \tag{6.19}$$

A state-space model $\{A_a, B_a, C_r\}$ of the strictly upper part of Θ_{22}^{-*} is given in terms of S,

R by

$$\begin{aligned} A_a &= \left(F_{22}(I-SF_{12})^{-1}\right)^* \\ B_a &= \left(H_{22}+F_{22}(I-SF_{12})^{-1}SH_{12}\right)^* \\ C_r &= (I-SR)^{-*}\left[G_{22}+G_{21}(I-R^{(-1)}F_{21})^{-1}R^{(-1)}F_{22}\right]^* . \end{aligned} \tag{6.20}$$

This model is uniformly minimal, with contractive controllability and observability Gramians.

PROOF The existence and contractivity of $S \in \mathcal{D}$ and $R \in \mathcal{D}$ has already been determined (lemma 6.8). First observe that although S satisfies by definition $x_{-[0]}S = x_{+[0]}$ ($a_{1p} = b_{2p} = 0$), it also satisfies $x_{-[1]}S = x_{+[1]}$ ($a_{1p} = b_{2p} = 0$ and $a_{1[0]} = b_{2[0]} = 0$), etc. This is readily obtained by applying inputs $Z^{-1}a_1$, etc., so that we get states $Z^{-1}x_+$ and $Z^{-1}x_-$. If $(Z^{-1}a_1)_p = Z^{-1}a_{1p} + Z^{-1}a_{1[0]} = 0$, then $(Z^{-1}x_-)_{[0]}S = (Z^{-1}x_+)_{[0]}$. But $(Z^{-1}x_-)_{[0]} = x_{-[1]}$, and likewise $(Z^{-1}x_+)_{[0]} = x_{+[1]}$. Hence $x_{-[1]}S = x_{+[1]}$.

To determine a state realization for the strictly upper part of $\Sigma_{22}^* = \Theta_{22}^{-*}$, we start from the definition of $\mathbf{\Sigma}$ (5.16), and specialize to the 0-th diagonal to obtain

$$[x_{+[0]} \quad x_{-[1]}^{(-1)} \quad a_{1[0]} \quad b_{2[0]}]\,\mathbf{\Sigma} = [x_{+[1]}^{(-1)} \quad x_{-[0]} \quad a_{2[0]} \quad b_{1[0]}].$$

Taking $a_1 = 0$ throughout this proof, inserting the partitioning of $\mathbf{\Sigma}$ in (5.17) gives

$$\left\{\begin{array}{lcl} x_{+[1]}^{(-1)} &=& x_{+[0]}F_{11} + x_{-[1]}^{(-1)}F_{21} + b_{2[0]}G_{21} \\ x_{-[0]} &=& x_{+[0]}F_{12} + x_{-[1]}^{(-1)}F_{22} + b_{2[0]}G_{22} \\ b_{1[0]} &=& x_{+[0]}H_{12} + x_{-[1]}^{(-1)}H_{22} + b_{2[0]}K_{21} \end{array}\right. \tag{6.21}$$

With $b_{2p} = 0$ and $b_{2[0]} = 0$, these equations yield an expression for $S^{(-1)}$:

$$\left\{\begin{array}{lcl} x_{+[1]}^{(-1)} = x_{-[1]}^{(-1)}S^{(-1)} &=& x_{-[0]}SF_{11} + x_{-[1]}^{(-1)}F_{21} \\ x_{-[0]} &=& x_{-[0]}SF_{12} + x_{-[1]}^{(-1)}F_{22} \end{array}\right.$$

$$\Leftrightarrow$$

$$\left\{\begin{array}{lcl} x_{-[0]} &=& x_{-[1]}^{(-1)}F_{22}(I-SF_{12})^{-1} \\ x_{-[1]}^{(-1)}S^{(-1)} &=& x_{-[1]}^{(-1)}\left\{F_{22}(I-SF_{12})^{-1}SF_{11}+F_{21}\right\} \end{array}\right. \tag{6.22}$$

(note that $(I-SF_{12})^{-1}$ is bounded because $\|S\| < 1$ and $\|F_{12}\| \le 1$), and hence S satisfies the indicated recursive relations (see also figure 6.3). The recursion for R is determined likewise.

In view of proposition 5.15, we can take x_- as the state of a minimal realization of the strictly upper part of Θ_{22}^{-*}. Let $\{A_a, B_a, C_r\}$ be a corresponding state realization, so that the strictly lower part of Θ_{22}^{-1} has an anti-causal state realization

$$\left\{\begin{array}{lcl} x_{-[0]} &=& x_{-[1]}^{(-1)}A_a^* + b_{2[0]}C_r^* \\ b_{1[0]} &=& x_{-[1]}^{(-1)}B_a^* \end{array}\right. .$$

The unknowns A_a, B_a and C_r can be expressed in terms of F, G, H by substitution in equations (6.21), and using S and R as intermediate quantities. Doing so with $b_2 = 0$, the first equation in (6.22) yields the expression for A_a in (6.20) and B_a can be determined in terms of S from the last equation in (6.21). C_r^* is obtained as the transfer $b_{2[0]} \mapsto x_{-[0]}$ for $a_1 = 0$ and $b_2 = b_{2[0]} \in \mathcal{D}_2$, so that $x_{-[0]}S = x_{+[0]}$ and $x_{-[1]}^{(-1)} = x_{+[1]}^{(-1)}R^{(-1)}$. Inserting the latter expression into the first equation of (6.21) twice yields

$$x_{-[1]}^{(-1)} = x_{+[0]}F_{11}(I - R^{(-1)}F_{21})^{-1}R^{(-1)} + b_{2[0]}G_{21}(I - R^{(-1)}F_{21})^{-1}R^{(-1)}.$$

Inserting this in the second equation of (6.21), and using $x_{+[0]} = x_{-[0]}S$ results in

$$\begin{aligned} x_{-[0]} &= x_{-[0]}SF_{12} + x_{-[0]}SF_{11}(I - R^{(-1)}F_{21})^{-1}R^{(-1)}F_{22} \\ &\quad + b_{2[0]}G_{21}(I - R^{(-1)}F_{21})^{-1}R^{(-1)}F_{22} + b_{2[0]}G_{22} \\ \Rightarrow & \\ x_{-[0]}(I - SR) &= b_{2[0]}(G_{22} + G_{21}(I - R^{(-1)}F_{21})^{-1}R^{(-1)}F_{22}) \end{aligned}$$

which gives the expression for C_r.

We have defined, in equation (5.23), the conjugate-Hankel operator $H' = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\cdot\,\Theta_{22}^{-1})|_{\mathcal{U}_2}$. In proposition 5.15 we showed that H' has a factorization $H' = \sigma\tau$, where the maps $\sigma : b_{2f} \mapsto x_{-[0]}$ and $\tau : x_{-[0]} \mapsto b_{1p}$ are onto and one-to-one, respectively, and both contractive. In particular, we can write $H' = \mathbf{P}_0(\cdot\,\mathbf{F}_r^*)\mathbf{F}_a$, where $\tau = \mathbf{F}_a = [B_a Z(I - A_a Z)^{-1}]^*$ (if $\ell_{A_a} < 1$) and $\sigma = \mathbf{P}_0(\cdot\,\mathbf{F}_r^*)$ with $\mathbf{F}_r = (I - A_a Z)^{-1}C_r$ (if $\ell_{A_a} < 1$). The properties of σ and τ imply that the derived model $\{A_a, B_a, C_r\}$ is uniformly minimal, with contractive controllability/observability Gramians. □

The second step in the construction of a realization for T_a is to determine a state realization for B'. This is done by evaluating $[U^* \quad -T^*\Gamma^{-1}]\Theta = [A' \quad -B']$. This has already been done in equation (5.37), which gives, with the state model for Θ written as

$$\Theta = \begin{bmatrix} A_\Theta & C_\Theta \\ B_\Theta & D_\Theta \end{bmatrix} = \left[\begin{array}{c|c} R & \\ \hline & I \end{array}\right] \left[\begin{array}{c|cc} A & C_1 & C_2 \\ \hline B_U & D_{11} & D_{12} \\ \Gamma^{-1}B & D_{21} & D_{22} \end{array}\right] \left[\begin{array}{c|c} R^{(-1)} & \\ \hline & I \end{array}\right],$$

$$\begin{aligned} [U^* \quad -T^*\Gamma^{-1}]\Theta &= \{[D_U^* \quad 0]D_\Theta + C^*\Lambda[C_1 \quad C_2]\} + \\ &\quad + \left\{[D_U^* \quad 0]\begin{bmatrix} B_U \\ \Gamma^{-1}B \end{bmatrix} + C^*\Lambda A\right\} Z(I - AZ)^{-1}[C_1 \quad C_2] \\ &= \{[D_U^* \quad 0]D_\Theta + C^*\Lambda[C_1 \quad C_2]\} + C^*(\Lambda - I)AZ(I - AZ)^{-1}[C_1 \quad C_2] \end{aligned}$$

(in which we used $C^*A + D_U^*B_U = 0$). Since this expression is equal to $[A' \quad -B']$ and $M = I - \Lambda$, we obtain a state-space model for B' as

$$B' = \{-D_U^*D_{12} - C^*(I - M)C_2\} + C^*MAZ(I - AZ)^{-1}C_2. \tag{6.23}$$

We are now in a position to determine a state realization for T_a.

THEOREM 6.10. *Let T, Γ, U and Θ be as in lemma 6.3, so that $[U^* \ -T^*\Gamma^{-1}]\Theta = [A' \ -B']$. Let $\{A, B, C, 0\}$ be an output normal strictly stable state realization for T, let M satisfy the Lyapunov equation (6.14), and let $\{A, B_U, C, D_U\}$ be a realization for U. Denote the block entries of $\boldsymbol{\Theta}$ as in (6.13), and let $\boldsymbol{\Sigma}$ corresponding to $\boldsymbol{\Theta}$ have a partitioning (5.17).*

Then the approximant T_a, defined as the strictly upper part of $T' = \Gamma\Theta_{22}^{-*}B'^*$, has a state realization $\{A_a, \Gamma B_a, C_a, 0\}$, where A_a, $B_a \in \mathcal{D}$ are defined by (6.20), and C_a is given by

$$C_a = C_r\left[-D_{12}^*D_U - C_2^*(I-M)C\right] + A_aY^{(-1)}A^*MC, \tag{6.24}$$

where C_r is defined in (6.20), and $Y \in \mathcal{D}$ satisfies the recursion $Y = A_aY^{(-1)}A^* + C_rC_2^*$.

PROOF The state realization for T_a is obtained by multiplying the model for B' in (6.23) by the model $\{A_a, B_a, C_r\}$ of the strictly upper part of Θ_{22}^{-*} as obtained in proposition 6.9. From this proposition, we have a state model of Θ_{22}^{-1} as

$$\Theta_{22}^{-1} \ = \ [\text{upper}] + C_r^*\mathbf{F}_a$$

where $\mathbf{F}_a$ is the selected basis representation of $\mathcal{H}(\Theta_{22}^{-*})$, satisfying $\mathbf{F}_a = (I-A_aZ)^{-*}Z^*B_a^* \in \mathcal{L}Z^{-1}$ when $\ell_{A_a} < 1$, and more in general the recursive equation

$$\mathbf{F}_a \ = \ Z^*B_a^* + Z^*A_a^*\mathbf{F}_a.$$

The model of B' is given in (6.23) as $B' \ = \ D' + C^*MAZ\mathbf{F}_0$, where

$$\begin{aligned} D' \ &= \ -D_{12}^*D_U - C_2^*(I-M)C, \\ \mathbf{F}_0 \ &= \ (I-AZ)^{-1}C_2, \qquad \mathbf{F}_0 = C_2 + AZ\mathbf{F}_0. \end{aligned}$$

Hence T_a is given by

$$\begin{aligned} T_a^*\Gamma^{-1} \ &= \ \mathbf{P}_{\mathcal{L}Z^{-1}}(B'\Theta_{22}^{-1}) \\ &= \ D'C_r^*\mathbf{F}_a + C^*MA\,\mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{F}_0\Theta_{22}^{-1}). \end{aligned}$$

It remains to evaluate $\mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{F}_0\Theta_{22}^{-1})$. Since $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{D}_2\mathbf{F}_0\Theta_{22}^{-1}) \in \mathcal{H}(\Theta_{22}^{-*})$, we can write $\mathbf{P}_{\mathcal{L}Z^{-1}}(\mathbf{F}_0\Theta_{22}^{-1}) = Y^*\mathbf{F}_a$, for some $Y \in \mathcal{D}$. Consequently,

$$\begin{aligned} \mathbf{P}_{\mathcal{L}Z^{-1}}(ZY^*\mathbf{F}_a) \ &= \ Y^{*(-1)}\mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{F}_a) \\ &= \ Y^{*(-1)}A_a^*\mathbf{F}_a. \end{aligned}$$

Because also $\mathbf{P}_{\mathcal{L}Z^{-1}}(ZY^*\mathbf{F}_a) = \mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{P}_{\mathcal{L}Z^{-1}}(Y^*\mathbf{F}_a)) = \mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{F}_0\Theta_{22}^{-1})$, we obtain

$$T_a^*\Gamma^{-1} \ = \ \left\{D'C_r^* + C^*MAY^{*(-1)}A_a^*\right\}\mathbf{F}_a,$$

which gives the expression for C_a in (6.24). Finally, the indicated recursion for Y follows via

$$\begin{aligned} A\mathbf{P}_{\mathcal{L}Z^{-1}}(Z\mathbf{F}_0\Theta_{22}^{-1}) &= \mathbf{P}_{\mathcal{L}Z^{-1}}(AZ\mathbf{F}_0\Theta_{22}^{-1}) \\ &= \mathbf{P}_{\mathcal{L}Z^{-1}}(\mathbf{F}_0\Theta_{22}^{-1}) - \mathbf{P}_{\mathcal{L}Z^{-1}}(C_2\Theta_{22}^{-1}) \\ \Leftrightarrow \quad AY^{*(-1)}A_a^*\mathbf{F}_a &= Y^*\mathbf{F}_a - C_2C_r^*\mathbf{F}_a \\ \Leftrightarrow \quad AY^{*(-1)}A_a^* &= Y^* - C_2C_r^*, \end{aligned}$$

where in the last step we used that $\mathbf{F}_a$ is a strong basis representation (proposition 6.9). □

A check on the dimensions of A_a reveals that this state realization for T_a has indeed a state dimension sequence given by $N = \#(\mathcal{B}_-)$: at each point in time it is equal to the number of Hankel singular values of T at that point that are larger than 1. The realization is given in terms of four recursions: two for M and S that run forward in time, the other two for R and Y that run backward in time and depend on S. One implication of this is that it is not possible to compute part of an optimal approximant of T if the model of T is known only partly, say up to time instant k. Algorithm 6.2 shows the computations involved in theorem 6.10. It computes a model $\{A_a, B_a, C_a, 0\}$ for the Hankel norm approximant T_a in terms of Γ and a model $\{A, B, C, 0\}$ for T.

There are a few remaining issues. T_a, as the strictly upper part of some operator in $\mathcal{X}$, is possibly unbounded. This occurs if the strictly upper part of Θ_{22}^{-*} is unbounded. We do not known whether this can occur. The realization of T_a is well defined, because Θ_{22}^{-1} is bounded, as well as projections of the kind $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot\,\Theta_{22}^{-1})$, so that in particular the Hankel operator H' which defines that realization is bounded. [In fact, one could probably set up a realization theory for unbounded operators with bounded Hankel operators.] A related second issue is that possibly $\ell_{A_a} = 1$. Although this seems unlikely in view of the assumptions on ℓ_A and the singular values of H_T that we have made, we have no proof yet that this cannot occur. Note that the proof of theorem 6.10 is not dependent on ℓ_{A_a} being strictly smaller than 1. Finally, an alternative derivation of a model for T_a goes via an inner-outer factorization of Θ_{22}. This gives rise to different expressions but still produces a two-directional recursive algorithm.

6.4 PARAMETRIZATION OF ALL APPROXIMANTS

Section 6.4 is devoted to the description of all possible solutions to the Hankel norm approximation problem that have order smaller than or equal to N, where $N = \text{s-dim } \mathcal{H}(\Theta_{22}^{-*})$ is the sequence of dimensions of the input state space of Θ_{22}^{-*}. We determine all possible bounded operators of mixed causality type T' for which it is true that

(1) $\|\Gamma^{-1}(T - T')\| = \|S^*U\| \leq 1$,

and (2) the state dimension sequence of T_a = (upper part of T') is at most equal to N.

In: $\mathbf{T}$ (model in output normal form for a strictly upper matrix T)
Γ (approximation parameters)
Out: $\mathbf{T}_a$ (model for Hankel norm approximant T_a)

do algorithm 6.1: gives M_k, $\mathbf{\Theta}_k$, $J_{\mathcal{B}_k}$, $C_{2,k}$, $D_{12,k}$, $D_{U,k}$ $(k = 1, \cdots, n)$
$S_1 = [\,\cdot\,]$

for $k = 1, \cdots, n$
[Compute $\mathbf{\Sigma}_k$ from $\mathbf{\Theta}_k$ using (5.15): gives F_{ij}, G_{ij}, H_{ij}
$S_{k+1} = F_{21,k} + F_{22,k}(I - S_k F_{12,k})^{-1} S_k F_{11,k}$
end

$R_{n+1} = [\,\cdot\,]$
$Y_{n+1} = [\,\cdot\,]$

for $k = n, \cdots, 1$

$$\begin{bmatrix} R_k &= F_{12,k} + F_{11,k}(I - R_{k+1}F_{21,k})^{-1} R_{k+1} F_{22,k} \\ C_{r,k}^* &= \left\{G_{22,k} + G_{21,k}(I - R_{k+1}F_{21,k})^{-1} R_{k+1} F_{22,k}\right\} (I - S_k R_k)^{-1} \\ A_{a,k} &= \left\{F_{22,k}(I - S_k F_{12,k})^{-1}\right\}^* \\ B_{a,k} &= \left\{H_{22,k} + F_{22,k}(I - S_k F_{12,k})^{-1} S_k H_{12,k}\right\}^* \\ Y_k &= A_{a,k} Y_{k+1} A_k^* + C_{r,k} C_{2,k}^* \\ C_{a,k} &= C_{r,k}\left\{-D_{12,k}^* D_{U,k} - C_{2,k}^*(I - M_k) C_k\right\} + A_{a,k} Y_{k+1} A_k^* M_k C_k \end{bmatrix}$$

end

Algorithm 6.2. The approximation algorithm.

(Note that we do not assume boundedness of T_a.) As we show in theorem 6.17 below, there are no Hankel norm approximants with state dimension lower than N. The result is that all solutions are obtained by a linear fractional transform (chain scattering transformation) of Θ with an upper and contractive parameter S_L. That this procedure effectively generates all approximants of locally finite type of s-degree at most equal to the sequence N can be seen from the fact that if $\| \Gamma^{-1}(T-T_a) \|_H \leq 1$, then an extension T' of T_a must exist such that $\| \Gamma^{-1}(T-T') \| \leq 1$. This is a consequence of a theorem on the Nehari problem (see section 6.5).

The notation is as in the previous sections. We started out with an operator $T \in Z\mathcal{U}$, and we assumed it to be locally finite, with a state realization in output normal form and related factorization $T = \Delta^* U$, where $\Delta \in \mathcal{U}$ and $U \in \mathcal{U}$, unitary and locally finite. Then we proceeded to solve the interpolation problem $[U^* \ -T^*\Gamma^{-1}]\Theta = [A' \ -B'] \in [\mathcal{U} \ \mathcal{U}]$, and we saw that the problem was solvable provided a related Lyapunov-Stein equation had a boundedly invertible solution. The solution was given in terms of an operator $T' = \Gamma^{-1}\Theta_{22}^{-*}B'^*$ in $\mathcal{X}$ of mixed causality type, and the approximant T_a of low order was given by the strictly upper part of T'. In the present section we shall first show that a large class of Hankel-norm approximants can be given in terms of the same J-unitary operator Θ and an arbitrary upper, contractive parameter S_L. Our previous result is the special case $S_L = 0$. Then we move on to show that all approximants of s-degree at most N are obtained in this way.

We first derive a number of preliminary facts which allow us to determine the state dimension sequence of a product of certain matrices.

Preliminary facts

PROPOSITION 6.11. *Let $B = I - X$, where $X \in \mathcal{X}$ and $\|X\| < 1$. Then $\mathbf{P}(\cdot B)\big|_{\mathcal{U}_2}$ and $\mathbf{P}(\cdot B^{-1})\big|_{\mathcal{U}_2}$ are Hilbert space isomorphisms on $\mathcal{U}_2$. Likewise, $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B)\big|_{\mathcal{L}_2Z^{-1}}$ and $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B^{-1})\big|_{\mathcal{L}_2Z^{-1}}$ are isomorphisms on $\mathcal{L}_2Z^{-1}$.*

PROOF B is invertible because $\|X\| < 1$. Since also

$$X_p := \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot X)\big|_{\mathcal{L}_2Z^{-1}}, \qquad X_f := \mathbf{P}(\cdot X)\big|_{\mathcal{U}_2}$$

are strictly contractive: $\|X_p\| < 1$, $\|X_f\| < 1$, it follows that $B_p = I - X_p = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B^{-1})\big|_{\mathcal{L}_2Z^{-1}}$ is invertible in $\mathcal{L}$, and $B_f = I - X_f$ is invertible in $\mathcal{U}$. In particular, for $u \in \mathcal{L}_2Z^{-1}$, the decomposition $uB =: y_1 + u_1$ (with $y_1 \in \mathcal{U}_2$, $u_1 = uB_p \in \mathcal{L}_2Z^{-1}$) satisfies

$$\|u_1\| \geq \varepsilon \|u\|, \qquad \text{some } \varepsilon > 0. \tag{6.25}$$

Take $y \in \mathcal{U}_2$, $y \neq 0$. To show that $\mathbf{P}(\cdot B^{-1})\big|_{\mathcal{U}_2}$ is one-to-one, we will show that the upper part of yB^{-1} is uniformly positive: $y_2 := \mathbf{P}(yB^{-1})$ has $\|y_2\| \geq \varepsilon_1 \|y\|$ (with $\varepsilon_1 > 0$).

Indeed, put $yB^{-1} =: y_2+u_2$ ($y_2 \in \mathcal{U}_2$, $u_2 \in \mathcal{L}_2Z^{-1}$). Since $u_2B = y-y_2B$, and B_p is invertible, we can apply the relation (6.25) proven above, in the form $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(u_2B) \geq \varepsilon_2 \| u_2 \|$, to obtain

$$\| \mathbf{P}_{\mathcal{L}_2Z^{-1}}(y_2B) \| = \| \mathbf{P}_{\mathcal{L}_2Z^{-1}}(u_2B) \| \geq \varepsilon_2 \| u_2 \| \qquad (\varepsilon_2 > 0).$$

Because B is bounded: $\|B\| < 2$, it follows that $\| y_2 \| > 1/2 \| y_2B \| > 1/2\varepsilon_2 \| u_2 \|$, or

$$\| y_2 \| > \varepsilon_3 \| u_2 \|, \qquad \varepsilon_3 = 1/2\varepsilon_2 > 0.$$

Hence, at this point we have $yB^{-1} = y_2 + u_2$ with $\| y_2 \| > \varepsilon_3 \| u_2 \|$ ($\varepsilon_3 > 0$). Because B^{-1} is boundedly invertible, there exists $\varepsilon_4 > 0$ such that $\| yB^{-1} \| \geq \varepsilon_4 \| y \|$, and we have

$$\| y_2 \| (1 + \frac{1}{\varepsilon_3}) > \| y_2 \| + \| u_2 \| > \| y_2 + u_2 \| > \varepsilon_4 \| y \|.$$

We finally obtain that

$$\| y_2 \| > \frac{\varepsilon_4}{1 + 1/\varepsilon_3} \| y \| =: \varepsilon_1 \| y \|$$

so that $\mathbf{P}(\cdot B^{-1})\big|_{\mathcal{U}_2}$ is one-to-one.

To show that $\mathbf{P}(\cdot B^{-1})\big|_{\mathcal{U}_2}$ is onto: $\mathbf{P}(\mathcal{U}_2 B^{-1}) = \mathcal{U}_2$, we have to show that for all $y_2 \in \mathcal{U}_2$, there exists an $y \in \mathcal{U}_2$ such that

$$\mathbf{P}(yB^{-1}) = y_2,$$

i.e., given $y_2 \in \mathcal{U}_2$ find $y \in \mathcal{U}_2$ such that $yB^{-1} = y_2+u_2$ (some $u_2 \in \mathcal{L}_2Z^{-1}$), or equivalently, $y_2B = y + u_2B$. We will use the fact that $B_p = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B)\big|_{\mathcal{L}_2Z^{-1}}$ is invertible so that $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(u_2B) = u_2B_p$ uniquely determines u_2. Indeed, given y_2, u_2 is computed as $u_2 = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(y_2B)B_p^{-1}$, and then $y \in \mathcal{U}_2$ is given by $y = (u_2 + y_2)B$.

The property on $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B^{-1})\big|_{\mathcal{L}_2Z^{-1}}$ is proven in a similar manner. □

Proposition 6.11 allows us to conclude, in particular, that if $\mathcal{A}$ is a subspace in $\mathcal{U}_2$, then

$$\text{s-dim}\, \mathbf{P}(\mathcal{A}B^{-1}) = \text{s-dim}\, \mathcal{A}$$

and if $\mathcal{B}$ is another subspace in $\mathcal{U}_2$, then $\mathcal{B} \subset \mathcal{A} \Leftrightarrow \mathbf{P}(\mathcal{B}B^{-1}) \subset \mathbf{P}(\mathcal{A}B^{-1})$.

PROPOSITION 6.12. *If $B = I - X$, $X \in \mathcal{X}$ and $\| X \| < 1$, and if $\mathcal{B} = \mathbf{P}(\mathcal{L}_2Z^{-1}\, B)$, then*

$$\mathbf{P}(\mathcal{B}B^{-1}) = \mathbf{P}(\mathcal{L}_2Z^{-1}\, B^{-1}).$$

PROOF We show mutual inclusion.

(1) $\mathbf{P}(\mathcal{B}B^{-1}) \subset \mathbf{P}(\mathcal{L}_2Z^{-1}B^{-1})$. Let $y \in \mathbf{P}(\mathcal{B}B^{-1})$. Then there exist $u \in \mathcal{L}_2Z^{-1}$ and $u_1 \in \mathcal{L}_2Z^{-1}$ such that $y = \mathbf{P}\left((uB + u_1)B^{-1}\right) = \mathbf{P}(u_1B^{-1})$. Hence $y \in \mathbf{P}(\mathcal{L}_2Z^{-1}B^{-1})$.

(2) $\mathbf{P}(\mathcal{L}_2Z^{-1}B^{-1}) \subset \mathbf{P}(\mathcal{B}B^{-1})$. Assume $y = \mathbf{P}(u_1B^{-1})$ for some $u_1 \in \mathcal{L}_2Z^{-1}$. Since $B_p = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\cdot B)\big|_{\mathcal{L}_2Z^{-1}}$ is an isomorphism (proposition 6.11), a $u \in \mathcal{L}_2Z^{-1}$ exists such that $\mathbf{P}_{\mathcal{L}_2Z^{-1}}(uB) = -u_1$. It follows that

$$\begin{aligned} y &= \mathbf{P}(u_1B^{-1}) \\ &= \mathbf{P}\left((uB+u_1)B^{-1}\right) \\ &= \mathbf{P}\left((uB-\mathbf{P}_{\mathcal{L}_2Z^{-1}}(uB))B^{-1}\right) \\ &= \mathbf{P}\left(\mathbf{P}(uB)B^{-1}\right) \quad \in \mathbf{P}(\mathcal{B}B^{-1}). \end{aligned}$$

□

PROPOSITION 6.13. *If $A \in \mathcal{L}$ and $A^{-1} \in \mathcal{X}$ and if $\mathcal{A} = \mathbf{P}(\mathcal{L}_2Z^{-1}A^{-1})$, then*

$$\mathcal{L}_2Z^{-1}A^{-1} = \overline{\mathcal{A}} \oplus \mathcal{L}_2Z^{-1}.$$

PROOF (Note that $\mathcal{A}$, as the range of a Hankel operator, need not be closed.) The left-to-right inclusion is obvious. To show the right-to-left inclusion, we show that $\mathcal{L}_2Z^{-1} \subset \mathcal{L}_2Z^{-1}A^{-1}$. Assume that $u \in \mathcal{L}_2Z^{-1}$. Then $u = (uA)A^{-1}$. But since $A \in \mathcal{L}$, we have $uA \in \mathcal{L}_2Z^{-1}$, and $u \in \mathcal{L}_2Z^{-1}A^{-1}$. The fact that $\overline{\mathcal{A}}$ is also in the image follows by complementation: $\mathcal{L}_2Z^{-1}A^{-1} \ominus \mathcal{L}_2Z^{-1} = \overline{\mathbf{P}(\mathcal{L}_2Z^{-1}A^{-1})}$. □

THEOREM 6.14. *Let $A \in \mathcal{L}$, $A^{-1} \in \mathcal{X}$, and suppose that the space $\mathcal{A} = \mathbf{P}(\mathcal{L}_2Z^{-1}A^{-1})$ is locally finite of s-dimension N. Let $B = I - X$ with $X \in \mathcal{X}$ and $\|X\| < 1$. Then*

$$\text{s-dim}\,\mathbf{P}(\mathcal{L}_2Z^{-1}A^{-1}B^{-1}) = N + p \quad \Rightarrow \quad \text{s-dim}\,\mathbf{P}(\mathcal{L}_2Z^{-1}BA) = p.$$

PROOF

$$\begin{aligned} \mathbf{P}(\mathcal{L}_2Z^{-1}A^{-1}B^{-1}) &= \mathbf{P}\left((\mathcal{L}_2Z^{-1} \oplus \overline{\mathcal{A}})B^{-1}\right) && \text{[prop. 6.13]} \\ &= \mathbf{P}(\mathcal{L}_2Z^{-1}B^{-1}) + \mathbf{P}(\overline{\mathcal{A}}B^{-1}) && \text{[linearity]} \\ &= \mathbf{P}(\mathcal{B}B^{-1}) + \mathbf{P}(\overline{\mathcal{A}}B^{-1}) && \text{[prop. 6.12]} \end{aligned}$$

where $\mathcal{B} = \mathbf{P}(\mathcal{L}_2Z^{-1}B)$.

In the sequel of the proof, we use the following two properties. The closure of a D-invariant locally finite linear manifold $\mathcal{H}$ yields a locally finite D-invariant subspace $\overline{\mathcal{H}}$ with the same s-dim. Secondly, let $\mathcal{M}$ be another locally finite D-invariant subspace and let X be a bounded operator on $\mathcal{U}_2$, then $\mathcal{H}X = [\mathbf{P}_{\mathcal{M}}(\mathcal{H})]X$ if $\mathcal{M}^{\perp}X = 0$.

Since $\overline{\mathcal{A}}$ and $\mathcal{B}$ are spaces in $\mathcal{U}_2$, and since according to proposition 6.11, $\mathbf{P}(\cdot B^{-1})\big|_{\mathcal{U}_2}$ is an isomorphism mapping $\overline{\mathcal{A}}$ and $\mathcal{B}$ to $\mathbf{P}(\overline{\mathcal{A}}B^{-1})$ and $\mathbf{P}(\mathcal{B}B^{-1})$, respectively, we obtain that s-dim$(\overline{\mathcal{A}} + \mathcal{B}) = N + p$. With $\mathcal{A}^{\perp} = \mathcal{U}_2 \ominus \overline{\mathcal{A}}$, it follows that $\mathbf{P}_{\mathcal{A}^{\perp}}(\mathcal{B})$ has s-dim equal to p, because s-dim$\overline{\mathcal{A}} = N$. The proof terminates by showing that

(1) $\mathbf{P}(\mathcal{L}_2 Z^{-1} BA) = \mathbf{P}(\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B})A)$, for

$$\begin{aligned}\mathbf{P}(\mathcal{L}_2 Z^{-1} BA) &= \mathbf{P}(\mathbf{P}(\mathcal{L}_2 Z^{-1} B)A) \\ &= \mathbf{P}(\mathcal{B}A) \\ &= \mathbf{P}(\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B})A),\end{aligned}$$

because $\overline{\mathcal{A}}A \subset \mathcal{L}_2 Z^{-1}$.

(2) $\mathbf{P}(\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B})A)$ is isomorphic to $\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B})$, which follows from the fact that the map $\mathbf{P}(\cdot A)\big|_{\mathcal{A}^\perp}$ is one-to-one, for $\mathbf{P}(xA) = 0 \Rightarrow x \in \overline{\mathcal{A}} \oplus \mathcal{L}_2 Z^{-1}$, and the kernel of $\mathbf{P}(\cdot A)\big|_{\mathcal{A}^\perp}$ is thus just $\{0\}$.

Consequently, s-dim $\mathbf{P}(\mathcal{L}_2 Z^{-1} BA) =$ s-dim $\mathbf{P}(\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B})A) =$ s-dim $\mathbf{P}_{\mathcal{A}^\perp}(\mathcal{B}) = p$. □

In the above theorem, we had $A \in \mathcal{L}$. A comparable result for $A \in \mathcal{U}$ follows directly by considering a duality property, and yields the corollary below.

COROLLARY 6.15. *Let $A \in \mathcal{U}$, $X \in \mathcal{X}$, $B = I - X$ and $\|X\| < 1$, and let A be invertible in $\mathcal{X}$. Suppose that $\mathcal{A} = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 A^{-1})$ has s-dimension N. Then*

$$\text{s-dim}\, \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 B^{-1} A^{-1}) = N + p \qquad \Rightarrow \qquad \text{s-dim}\, \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 AB) = p.$$

PROOF For any bounded operator, the dimension of its range is equal to the dimension of its co-range. Hence for $T \in \mathcal{X}$, we have that s-dim ran(H_T) = s-dim ran(H_T^*), or

$$\text{s-dim}\, \mathbf{P}(\mathcal{L}_2 Z^{-1} T) = \text{s-dim}\, \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 T^*).$$

□

Generating new solutions of the interpolation problem

Throughout the remainder of the section we use the notion of *causal state dimension sequence* of an operator $T \in \mathcal{X}$ as the s-dimension N of the space $\mathcal{H}(T) = \mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 T^*)$. N is thus a sequence of numbers $\{N_i : i \in \mathbb{Z}\}$ where all N_i in our case are finite. Dually, the s-dimension of $\mathbf{P}_{\mathcal{L}_2 Z^{-1}}(\mathcal{U}_2 T)$ is described as the *anti-causal state dimension sequence.* We use the following lemma, in which we must assume that Θ is constructed according to the recipe given in corollary 5.20, so that its input state space $\mathcal{H}(\Theta)$ is generated by (*viz.* equation (6.12))

$$\mathcal{H}(\Theta) = \mathcal{D}_2^{\mathcal{B}} (I - Z^* A^*)^{-1} Z^* \begin{bmatrix} B_U^* & B^* \Gamma^{-1} \end{bmatrix}.$$

LEMMA 6.16. *Let T, Γ and Θ be as in lemma 6.4, such that $T = \Delta^* U$ is a factorization of T with $\Delta \in \mathcal{U}$ and $U \in \mathcal{U}$ is inner, and Θ is the J-unitary operator with input state space given by (6.12) and defined by the realization (6.13). Then*

$$\begin{array}{rcl} [U^* \quad 0]\Theta & \in & [\mathcal{L} \quad \mathcal{L}] \\ [-\Delta^* \quad \Gamma]\Theta & \in & [\mathcal{L} \quad \mathcal{L}]. \end{array}$$

PROOF We prove this by brute-force calculations on the realizations of U and Θ, as in (6.13):

$$\begin{aligned} [U^* \quad 0]\Theta &= \left\{D_U^* + C^*(I - Z^*A^*)^{-1}Z^*B_U^*\right\}\left\{[D_{11} \quad D_{12}] + B_U Z(I - AZ)^{-1}[C_1 \quad C_2]\right\} \\ &= D_U^*[D_{11} \quad D_{12}] + D_U^* B_U Z(I - AZ)^{-1}[C_1 \quad C_2] + \\ &\quad + C^*(I - Z^*A^*)^{-1}Z^*B_U^*[D_{11} \quad D_{12}] + \\ &\quad + C^*(I - Z^*A^*)^{-1}Z^*B_U^* B_U Z(I - AZ)^{-1}[C_1 \quad C_2]. \end{aligned}$$

Upon using the identities $D_U^* B_U + C^*A = 0$, $B_U^* B_U + A^*A = I$, and

$$(I - Z^*A^*)^{-1}Z^*(I - A^*A)Z(I - AZ)^{-1} = AZ(I - AZ)^{-1} + (I - Z^*A^*)^{-1},$$

it is seen that the terms with $(I - AZ)^{-1}$ cancel each other, so that

$$\begin{aligned} [U^* \quad 0]\Theta &= D_U^*[D_{11} \quad D_{12}] + C^*[C_1 \quad C_2] + \\ &\quad + C^*(I - Z^*A^*)^{-1}Z^*\left\{A^*[C_1 \quad C_2] + B_U^*[D_{11} \quad D_{12}]\right\} \\ &\in [\mathcal{L} \quad \mathcal{L}]. \end{aligned}$$

In much the same way,

$$\begin{aligned} [-\Delta^* \quad \Gamma]\Theta &= \left[\left\{-DD_U^* - BB_U^* - (DC^* + BA^*)Z^*(I - A^*Z^*)^{-1}B_U^*\right\} \quad \Gamma\right] \times \\ &\quad \times \left\{\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} + \begin{bmatrix} B_U \\ \Gamma^{-1}B \end{bmatrix} Z(I - AZ)^{-1}[C_1 \quad C_2]\right\} \\ &= (\text{lower}) + \left\{(-DD_U^* - BB_U^*)B_U + B\right\} Z(I - AZ)^{-1}[C_1 \quad C_2] + \\ &\quad + (-DC^* - BA^*)Z^*(I - A^*Z^*)^{-1}B_U^* B_U Z(I - AZ)^{-1}[C_1 \quad C_2] \\ &= (\text{lower}) + \left\{-DD_U^* B_U - BB_U^* B_U + B - DC^*A - BA^*A\right\} Z(I - AZ)^{-1}[C_1 \quad C_2] \\ &= (\text{lower}) + \left\{DC^*A - B + BA^*A + B - DC^*A - BA^*A\right\} Z(I - AZ)^{-1}[C_1 \quad C_2] \\ &= (\text{lower}) + 0. \end{aligned}$$

□

THEOREM 6.17. *Let $T \in Z\mathcal{U}$ be a locally finite operator with strictly stable output normal realization $\{A, B, C, 0\}$, let Γ be an invertible Hermitian diagonal operator. Let H_k be the Hankel operator of $\Gamma^{-1}T$ at time point k, and suppose that an $\varepsilon > 0$ exists such that, for each k, none of the singular values of H_k are in the interval $[1 - \varepsilon, 1 + \varepsilon]$. Let N be the sequence of the numbers N_k of singular values of H_k that are larger than 1.*

Let U be the inner factor of an external factorization (theorem 4.9), with unitary realization $\{A, B_U, C, D_U\}$, and let Θ be a J-unitary block-upper operator such that its input state space $\mathcal{H}(\Theta)$ is given by (6.12).

(1) *If $S_L \in \mathcal{U}$ is contractive, then $\Theta_{22} - \Theta_{21}S_L$ is boundedly invertible, and*

$$S = (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1}$$

is contractive.

(2) *Let, further, $T' = T + \Gamma S^* U$. Then*

(*a*) $\| \Gamma^{-1}(T - T') \| = \| S^* U \| \leq 1$,

(*b*) *the causal state dimension sequence of T_a = (upper part of T') is precisely equal to N.*

That is, T_a is a Hankel norm approximant of T.

PROOF (1) By the J-unitarity of Θ, Θ_{22} is boundedly invertible and $\| \Theta_{22}^{-1}\Theta_{21} \| < 1$, whence $\Theta_{22} - \Theta_{21}S_L = \Theta_{22}(I - \Theta_{22}^{-1}\Theta_{21}S_L)$ is boundedly invertible. Hence S exists. Its contractivity follows by the usual direct calculation (see *e.g.*, [24]).

(2*a*) follows immediately since $\Gamma^{-1}(T - T') = S^* U$ and U is unitary.

(2*b*) The proof uses the following equality:

$$\begin{aligned}
T'^* \Gamma^{-1} &= [U^* \quad -T^*\Gamma^{-1}] \begin{bmatrix} S \\ -I \end{bmatrix} \\
&= [U^* \quad -T^*\Gamma^{-1}] \begin{bmatrix} \Theta_{11} & \Theta_{12} \\ \Theta_{21} & \Theta_{22} \end{bmatrix} \begin{bmatrix} S_L \\ -I \end{bmatrix} (\Theta_{22} - \Theta_{21}S_L)^{-1} \\
&= [A' \quad -B'] \begin{bmatrix} S_L \\ -I \end{bmatrix} (\Theta_{22} - \Theta_{21}S_L)^{-1} \\
&= (A'S_L + B')(\Theta_{22} - \Theta_{21}S_L)^{-1} .
\end{aligned}$$

Since $(A'S_L + B') \in \mathcal{U}$, the anti-causal state dimension sequence of T'^* is at each point in time *at most* equal to the number of anti-causal states of $(\Theta_{22} - \Theta_{21}S_L)^{-1}$ at that point. Because the latter expression is equal to $(I - \Theta_{22}^{-1}\Theta_{21}S_L)^{-1}\Theta_{22}^{-1}$, and $\|\Theta_{22}^{-1}\Theta_{21}S_L\| < 1$, application of corollary 6.15 with $A = \Theta_{22}$ and $B = I - \Theta_{22}^{-1}\Theta_{21}S_L$ shows that this sequence is equal to the anti-causal state dimension sequence of Θ_{22}^{-1}, *i.e.*, equal to N. Hence s-dim $\mathcal{H}(T') \leq N$ (pointwise).

The proof terminates by showing that also s-dim $\mathcal{H}(T') \geq N$, so that in fact s-dim $\mathcal{H}(T') = N$. Define

$$\begin{cases} G_2 = (\Theta_{22} - \Theta_{21}S_L)^{-1} \\ G_1 = S_L G_2 \end{cases}$$

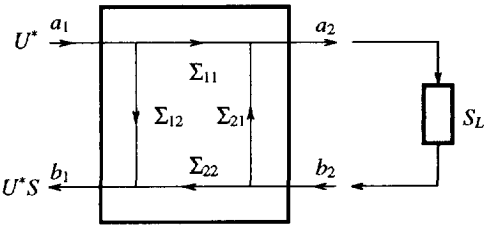


Figure 6.4. Θ (or Σ) generates Hankel norm approximants via S and parametrized by S_L.

so that

$$\begin{bmatrix} S \\ -I \end{bmatrix} = \Theta \begin{bmatrix} G_1 \\ -G_2 \end{bmatrix}.$$

Because Θ is J-inner: $\Theta^* J \Theta = J$, this is equivalent to $[G_1^* \quad G_2^*] := [S^* \quad I]\Theta$, and using $S = -\Delta\Gamma^{-1} + UT'^*\Gamma^{-1}$ we obtain

$$\Gamma[G_1^* \quad G_2^*] = T'[U^* \quad 0]\Theta + [-\Delta^* \quad \Gamma]\Theta \tag{6.26}$$

However, according to lemma 6.16,

$$\begin{aligned} [U^* \quad 0]\Theta &\in [\mathcal{L} \quad \mathcal{L}] \\ [-\Delta^* \quad \Gamma]\Theta &\in [\mathcal{L} \quad \mathcal{L}]. \end{aligned}$$

This implies $\mathcal{H}(G_2^*) \subset \mathcal{H}(T')$ (same proof as in lemma 6.6). Hence s-dim$\mathcal{H}(T') \geq$ s-dim$\mathcal{H}(G_2^*) = N$. □

Thus, all S of the form $S = (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1}$ with $S_L \in \mathcal{U}$, $\|S_L\| \leq 1$ give rise to Hankel norm approximants of T. We encountered this expression earlier in chapter 5: it is a chain-scattering transformation of S_L by Θ. Consequently, S is the transfer of port a_1 to b_1 if $b_2 = a_2 S_L$, as in figure 6.4.

The reverse question is: are all Hankel norm approximants obtained this way? That is, given some T' whose strictly upper part is a Hankel norm approximant of T, is there a corresponding upper and contractive S_L such that T' is given by $T' = T + \Gamma S^* U$, with S as above. This problem is addressed in the following theorem. The main issue is to prove that S_L as defined by the equations is upper; the proof is an extension of the proof that S_L generated all interpolants in the definite interpolation problem in section 5.3 (theorem 5.21), although some of the items are now more complicated.

Generating all approximants

THEOREM 6.18. *Let T, Γ, U and Θ be as in theorem 6.17, and let N be the number of Hankel singular values of $\Gamma^{-1}T$ that are larger than 1. Let be given a bounded operator $T' \in \mathcal{X}$ such that*

(1) $\|\Gamma^{-1}(T-T')\| \leq 1$,
(2) *the state dimension sequence of T_a = (upper part of T') is at most equal to N.*

Define $S = U(T'^* - T^*)\Gamma^{-1}$. Then there is an operator S_L with ($S_L \in \mathcal{U}$, $\|S_L\| \leq 1$) such that

$$S = (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1}$$

(i.e., Θ generates all Hankel-norm approximants). The state dimension of T_a is precisely equal to N.

PROOF The main line of the proof runs in parallel with [20], but differs in detail. In particular, the 'winding number' argument to determine state dimensions is replaced by theorem 6.14 and its corollary 6.15. The proof consists of five steps.

1. From the definition of S, and using the factorization $T = \Delta^* U$, we know that

$$\|S\| = \|U(T'^* - T^*)\Gamma^{-1}\| = \|\Gamma^{-1}(T' - T)\| \leq 1$$

so S is contractive. Since $S = -\Delta\Gamma^{-1} + UT'^*\Gamma^{-1}$, where Δ and U are upper, the anti-causal state dimension sequence of S is at most equal to N, since it depends exclusively on T'^*, for which this is given.

2. Define

$$[G_1^* \quad G_2^*] := [S^* \quad I]\Theta. \tag{6.27}$$

Then $\mathcal{H}(G_1^*) \subset \mathcal{H}(T')$ and $\mathcal{H}(G_2^*) \subset \mathcal{H}(T')$.

PROOF Using $S = -\Delta\Gamma^{-1} + UT'^*\Gamma^{-1}$, equation (6.27) can be rewritten as

$$\Gamma[G_1^* \quad G_2^*] = T'[U^* \quad 0]\Theta + [-\Delta^* \quad \Gamma]\Theta$$

According to lemma 6.16,

$$\begin{aligned} [U^* \quad 0]\Theta &\in [\mathcal{L} \quad \mathcal{L}] \\ [-\Delta^* \quad \Gamma]\Theta &\in [\mathcal{L} \quad \mathcal{L}]. \end{aligned}$$

As in the proof of theorem 6.17, this implies $\mathcal{H}(G_1^*) \subset \mathcal{H}(T')$ and $\mathcal{H}(G_2^*) \subset \mathcal{H}(T')$.

3. Equation (6.27) can be rewritten using $\Theta^{-1} = J\Theta^* J$ as

$$\begin{bmatrix} S \\ -I \end{bmatrix} = \Theta \begin{bmatrix} G_1 \\ -G_2 \end{bmatrix} \tag{6.28}$$

G_2 is boundedly invertible, and S_L defined by $S_L = G_1 G_2^{-1}$ is well defined and contractive: $\| S_L \| \leq 1$. In addition, S satisfies $S = (\Theta_{11} S_L - \Theta_{12})(\Theta_{22} - \Theta_{21} S_L)^{-1}$ as required.

PROOF As in the proof of theorem 5.21, step 2, we have, for some $\varepsilon > 0$,

$$G_1^* G_1 + G_2^* G_2 \geq \varepsilon I, \qquad G_1^* G_1 \leq G_2^* G_2 . \tag{6.29}$$

Together, this shows that $G_2 \geq 1/2\, \varepsilon I$, and hence G_2 is boundedly invertible (but not necessarily in $\mathcal{U}$). With $S_L = G_1 G_2^{-1}$, equation (6.29) shows that $S_L^* S_L \leq 1$, and hence $\| S_L \| \leq 1$. Evaluating equation (6.28) gives

$$\begin{aligned} G_2^{-1} &= \Theta_{22} - \Theta_{21} S_L \\ S G_2^{-1} &= \Theta_{11} S_L - \Theta_{12} \end{aligned} \tag{6.30}$$

and hence $S = (\Theta_{11} S_L - \Theta_{12})(\Theta_{22} - \Theta_{21} S_L)^{-1}$.

4. $G_2^{-1} \in \mathcal{U}$, the space $\mathcal{H}(T')$ has the same dimension as $\mathcal{H}(\Theta_{22}^*)$, and $\mathcal{H}(G_1^*) \subset \mathcal{H}(G_2^*)$.

PROOF According to equation (6.30), G_2^{-1} satisfies

$$\begin{aligned} G_2^{-1} &= \Theta_{22}(I - \Theta_{22}^{-1}\Theta_{21} S_L) \\ G_2 &= (I - \Theta_{22}^{-1}\Theta_{21} S_L)^{-1} \Theta_{22}^{-1} . \end{aligned}$$

Let p be the dimension sequence of anti-causal states of G_2^{-1}, and $N_2 \leq N$ be the number of anti-causal states of G_2, with N the number of anti-causal states of Θ_{22}^{-1}. Application of corollary 6.15 with $A = \Theta_{22}$ and $B = (I - \Theta_{22}^{-1}\Theta_{21} S_L)$ shows that $N_2 = N + p$, and hence $N_2 = N$ and $p = 0$: $G_2^{-1} \in \mathcal{U}$, and $\mathcal{H}(G_2^*)$ has dimension N. Step 2 claimed $\mathcal{H}(G_2^*) \subset \mathcal{H}(T')$, and because T' has at most N anti-causal states, we must have that in fact $\mathcal{H}(G_2^*) = \mathcal{H}(T')$, and hence $\mathcal{H}(G_1^*) \subset \mathcal{H}(G_2^*)$.

5. $S_L \in \mathcal{U}$.

PROOF This can be inferred from $G_2^{-1} \in \mathcal{U}$, and $\mathcal{H}(G_1^*) \subset \mathcal{H}(G_2^*)$, as follows. $S_L \in \mathcal{U}$ is equivalent to $\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 S_L) = 0$, and

$$\begin{aligned} \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 S_L) &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_1 G_2^{-1}) \\ &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}\left(\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_1) G_2^{-1} \right) \end{aligned}$$

since $G_2^{-1} \in \mathcal{U}$. Using $\mathcal{H}(G_1^*) \subset \mathcal{H}(G_2^*)$, or $\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_1) \subset \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_2)$ we obtain that

$$\begin{aligned} \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 S_L) &\subset \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}\left(\mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_2) G_2^{-1} \right) \\ &= \mathbf{P}_{\mathcal{L}_2 \mathcal{Z}^{-1}}(\mathcal{U}_2 G_2 G_2^{-1}) \qquad \text{(since } G_2^{-1} \in \mathcal{U}\text{)} \\ &= 0 . \end{aligned}$$

□

6.5 THE NEHARI PROBLEM

The theorems given in section 6.2 contain an implicit proof of an equivalent of Nehari's theorem in the present setting, for operators T which have a strictly stable, uniformly observable realization. If Γ in (6.7) is chosen such that all local Hankel singular values are uniformly smaller than one, then $T' = (B'\Theta_{22}^{-1}\Gamma)^*$ obtained through lemma 6.3 is a lower ($\in \mathcal{L}$) operator and the state sequence x_- is of dimension zero: $\#(\mathcal{B}_-) = 0$ and $J_{\mathcal{B}} = I$. Such a T' is known as the Nehari extension of T: it is such that $\| \Gamma^{-1}(T-T') \| \leq 1$ so that, when $\| \Gamma^{-1}T \|_H < 1$, there exists an extension $E \in \mathcal{X}$ such that the upper part of E is equal to $\Gamma^{-1}T$ and E is contractive. The Nehari problem is to find E or, equivalently, T'. This problem can also be viewed as a distance problem: given $T \in \mathcal{U}$, find an operator $T' \in \mathcal{L}$ that is closest to it, in the sense that $\| T-T' \|$ is minimized.

THEOREM 6.19. *If T is a bounded upper operator which has a locally finite strictly stable and uniformly observable realization, then*

$$\| T \|_H = \inf_{T' \in \mathcal{L}} \| T-T' \|. \tag{6.31}$$

PROOF Let $d = \| T \|_H$ and consider the operator $(d+\varepsilon)^{-1}T$ for some $\varepsilon > 0$. Then, with $\Gamma = d+\varepsilon$, $r := \| (d+\varepsilon)^{-1}\Gamma^{-1}T \|_H < 1$ and lemma 6.4 applies. Since the largest singular value of any local Hankel operator of $(d+\varepsilon)^{-1}T$ is majorized by r, we have that the sequence of singular values larger than one is zero, so that $\Theta_{22}^{-1} \in \mathcal{U}$ and $T' = (B'\Theta_{22}^{-1}(d+\varepsilon))^*$ is a lower operator. Lemma 6.4 ensures that

$$\| (d+\varepsilon)^{-1}(T-T') \| \leq 1$$

by construction, and hence

$$\| T-T' \| \leq d+\varepsilon.$$

Letting $\varepsilon \downarrow 0$ achieves (6.31). The reverse inequality is obvious from proposition 6.2. □

Again, one can show that all possible Nehari extensions are parameterized by the set of contractive upper operators S_L.

For time-invariant systems, the Nehari problem is one of the classical extension problems with well-known solutions [25, 3], which are derived using interpolation or Beurling-Lax representation theory. For time-varying systems, an early statement and proof appears in the work of Arveson [26, thm. 1.1] on operators in a nest algebra. A comparable result has been obtained by Gohberg, Kaashoek and Woerdeman [27, 28, 29] in the context of block matrix and operator matrix extensions. Their solutions are recursive on the entries of the block matrix: it is possible to work from top to bottom, adding rows to the extension found so far, in such a way that the resulting matrices remain contractive. The recursion

scheme is a generalized Schur recursion, similar to the recursive solution to the indefinite interpolation problem to be described in section 6.6 below, but specialized to apply to the definite case.

A state-space realization of the 'maximum entropy' or 'central' Nehari extension T' can be obtained as a special instance of the method presented in section 6.3, and does not need the upward recursions because the dimension of x_- is zero. The result is a closed-form solution: it is specified solely in terms of the given state realization operators of T.

THEOREM 6.20. *Let $T \in \mathcal{U}$ be a strictly upper locally finite operator with realization $\{A, B, C, 0\}$ in output normal form. If $\|T\|_H < 1$ then T has a Nehari extension $E = T - T' \in \mathcal{X}$ such that E is contractive and the strictly upper part of E is equal to T (i.e., $T'^* \in \mathcal{U}$. A realization of T'^*, i.e., the upper part of $-E^*$, is given by*

$$\begin{aligned} A_e &= A(I-(I-A^*MA)^{-1}B^*B) \\ B_e &= -C^*MA(I-(I-A^*MA)^{-1}B^*B) \\ C_e &= A(I-A^*MA)^{-1}B^* \\ D_e &= -C^*MA(I-A^*MA)^{-1}B^* \end{aligned} \tag{6.32}$$

where M satisfies $M^{(-1)} = A^*MA + B^*B$.

PROOF The existence of the Nehari extension has already been proven: it suffices to take $T'^* = B'\Theta_{22}^{-1}$, where B' and Θ are as in lemma 6.3 and 6.4. Let B_U and D_U be such that

$$\mathbf{U} = \begin{bmatrix} A & C \\ B_U & D_U \end{bmatrix}$$

is a unitary realization of the inner factor U of the external factorization of T. The realization $\boldsymbol{\Theta}$ has the general form of equation (6.13) (with $\Gamma = I$), but since $J_B = I$, all negative signature is associated with D_{22}, which implies that D_{22}^{-1} exists and is bounded, and also that D_{21} can be chosen equal to zero (as in [24, thm. 3.1]). Hence we consider Θ of the form

$$\boldsymbol{\Theta} = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{c|cc} A & C_1 & C_2 \\ \hline B_U & D_{11} & D_{12} \\ B & 0 & D_{22} \end{array}\right] \begin{bmatrix} R^{-(-1)} & & \\ & I & \\ & & I \end{bmatrix}$$

where the first column of the operator matrix in the middle is specified, and an extension by a second and third column is to be determined, as well as a state transformation R, such that $\boldsymbol{\Theta}$ is J-unitary. We use the fact that $\mathbf{U}$ is unitary to derive expressions for entries in $\boldsymbol{\Theta}$. Let, as before, $\Lambda = R^*R$ (recall that $J_B = I$). The remainder of the proof consists of 6 steps.

1. $C_1 = \Lambda^{-1}C\alpha$,
   $D_{11} = D_U\alpha$, where $\alpha = (C^*\Lambda^{-1}C + D_U^*D_U)^{1/2}$.

PROOF The J-unitarity relations between the first and second block column of $\boldsymbol{\Theta}$ lead to

$$\begin{aligned} A^*\Lambda C_1 + B_U^* D_{11} &= 0 \\ C_1^*\Lambda C_1 + D_{11}^* D_{11} &= I. \end{aligned}$$

The first equation shows that, for some scaling α,

$$\begin{bmatrix} \Lambda C_1 \\ D_{11} \end{bmatrix} = \begin{bmatrix} A \\ B_U \end{bmatrix}^{\perp} \alpha = \begin{bmatrix} C \\ D_U \end{bmatrix} \alpha$$

The scaling α follows from the second equation.

2. $C_2^* C + D_{12}^* D_U = 0$.

PROOF The J-unitarity conditions between the second and third column lead to

$$\begin{aligned} & C_1^*\Lambda C_2 + D_{11}^* D_{12} &&= 0 \\ \Rightarrow \quad & \alpha^* C^* C_2 + \alpha^* D_U^* D_{12} &&= 0. \end{aligned}$$

3. $B' = C^* M (I - AZ)^{-1} C_2$.

PROOF A state-space model of B' was given in equation (6.23) as

$$B' = \left\{ -D_U^* D_{12} - C^*(I - M) C_2 \right\} + C^* MAZ (I - AZ)^{-1} C_2 .$$

Using the result of step 2 gives the intended simplification.

4. $T'^* = B' \Theta_{22}^{-1} = C^* M \left(I - [A - C_2 D_{22}^{-1} B] Z\right)^{-1} C_2 D_{22}^{-1}$.

PROOF Let $A_e = A - C_2 D_{22}^{-1} B$. Then

$$\begin{aligned} T'^* = B' \Theta_{22}^{-1} &= \left[C^* M (I - AZ)^{-1} C_2\right] \left[D_{22}^{-1} - D_{22}^{-1} BZ (I - A_e Z)^{-1} C_2 D_{22}^{-1}\right] \\ &= C^* M (I - AZ)^{-1} \left[I - C_2 D_{22}^{-1} BZ (I - A_e Z)^{-1}\right] C_2 D_{22}^{-1} \\ &= C^* M (I - AZ)^{-1} \left[(I - A_e Z) - C_2 D_{22}^{-1} BZ\right] (I - A_e Z)^{-1} C_2 D_{22}^{-1} \\ &= C^* M (I - AZ)^{-1} (I - AZ)(I - A_e Z)^{-1} C_2 D_{22}^{-1} . \end{aligned}$$

5. $C_2 D_{22}^{-1} = A (I - A^* MA)^{-1} B^*$.

PROOF The J-unitarity conditions imply

$$\begin{bmatrix} A & C_1 \\ B_U & D_{11} \\ B & 0 \end{bmatrix}^* \begin{bmatrix} \Lambda & & \\ & I & \\ & & -I \end{bmatrix} \begin{bmatrix} C_2 \\ D_{12} \\ D_{22} \end{bmatrix} = 0$$

$$\Rightarrow \quad \begin{bmatrix} A & C_1 \\ B_U & D_{11} \end{bmatrix}^* \begin{bmatrix} \Lambda & \\ & I \end{bmatrix} \begin{bmatrix} C_2 \\ D_{12} \end{bmatrix} = \begin{bmatrix} B^* \\ 0 \end{bmatrix} D_{22}$$

$$\Rightarrow \quad \begin{bmatrix} C_2D_{22}^{-1} \\ D_{12}D_{22}^{-1} \end{bmatrix} = \begin{bmatrix} \Lambda^{-1} & \\ & I \end{bmatrix} \begin{bmatrix} A & C_1 \\ B_U & D_{11} \end{bmatrix}^{-*} \begin{bmatrix} B^* \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} \Lambda^{-1} & \\ & I \end{bmatrix} \begin{bmatrix} \Lambda A & C\alpha \\ B_U & D_U\alpha \end{bmatrix} \begin{bmatrix} (\Lambda^{(-1)} + B^*B)^{-1} & \\ & I \end{bmatrix} \begin{bmatrix} B^* \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} A(\Lambda^{(-1)} + B^*B)^{-1}B^* \\ B_U(\Lambda^{(-1)} + B^*B)^{-1}B^* \end{bmatrix}$$

where we have used the fact that

$$\begin{bmatrix} A^* & B_U^* \\ C_1^* & D_{11}^* \end{bmatrix} \begin{bmatrix} \Lambda A & C\alpha \\ B_U & D_U\alpha \end{bmatrix} = \begin{bmatrix} \Lambda^{(-1)} + B^*B & \\ & I \end{bmatrix}$$

Finally, using $M = I - \Lambda$, where M satisfies $M^{(-1)} = A^*MA + B^*B$ gives $\Lambda^{(-1)} + B^*B = I - A^*MA$.

6. $T'^* = D_e + B_eZ(I - A_eZ)^{-1}C_e$, where $\{A_e, B_e, C_e, D_e\}$ are as in equation (6.32).

PROOF From step 4,

$$\begin{aligned} T'^* &= C^*M(I - A_eZ)^{-1}C_2D_{22}^{-1} \\ &= C^*MC_2D_{22}^{-1} + C^*MA_eZ(I - A_eZ)^{-1}C_2D_{22}^{-1} \end{aligned}$$

where $A_e = A - C_2D_{22}^{-1}B$. It remains to make the substitution $C_2D_{22}^{-1} = A(I - A^*MA)^{-1}B^*$. □

Numerical example

We illustrate theorem 6.20 with a numerical example. Let T be given by

$$T = \begin{bmatrix} 0 & .326 & .566 & .334 & .078 & -.008 & -.012 & -.003 \\ 0 & 0 & .326 & .566 & .334 & .078 & -.008 & -.012 \\ 0 & 0 & 0 & .326 & .566 & .334 & .078 & -.008 \\ 0 & 0 & 0 & 0 & .326 & .566 & .334 & .078 \\ 0 & 0 & 0 & 0 & 0 & .326 & .566 & .334 \\ 0 & 0 & 0 & 0 & 0 & 0 & .326 & .566 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & .326 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

The norm of T is computed as $\|T\| = 1.215$, and T has Hankel singular values equal to

H_1	H_2	H_3	H_4	H_5	H_6	H_7	H_8
	.7385	.9463	.9856	.9866	.9856	.9463	.7385
		.2980	.3605	.3661	.3605	.2980	
			.0256	.0284	.0256		

so that $\|T\|_H = .9866 < 1$. A realization for T is obtained via algorithm 3.1 as

$$\mathbf{T}_1 = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline -.739 & .000 \end{array}\right] \qquad \mathbf{T}_2 = \left[\begin{array}{cc|c} .733 & -.517 & -.442 \\ \hline -.738 & .000 & .000 \end{array}\right]$$

$$\mathbf{T}_3 = \left[\begin{array}{ccc|c} .733 & -.517 & .000 & -.442 \\ .508 & -.012 & -.084 & .857 \\ \hline -.738 & -.000 & .000 & .000 \end{array}\right] \qquad \mathbf{T}_4 = \left[\begin{array}{ccc|c} .733 & -.517 & -.000 & -.442 \\ .508 & -.012 & -.084 & .857 \\ .430 & .836 & -.212 & -.265 \\ \hline -.738 & -.000 & .000 & .000 \end{array}\right]$$

$$\mathbf{T}_5 = \left[\begin{array}{ccc|c} .738 & -.509 & .000 & -.442 \\ .509 & -.005 & .076 & .857 \\ .424 & .845 & .192 & -.264 \\ \hline -.734 & .000 & .000 & .000 \end{array}\right] \qquad \mathbf{T}_6 = \left[\begin{array}{cc|c} .780 & .441 & -.444 \\ .506 & -.026 & .862 \\ -.369 & .897 & .244 \\ \hline -.654 & -.000 & .000 \end{array}\right]$$

$$\mathbf{T}_7 = \left[\begin{array}{c|c} -.867 & -.499 \\ .499 & -.867 \\ \hline .326 & .000 \end{array}\right] \qquad \mathbf{T}_8 = \left[\begin{array}{c|c} \cdot & 1.000 \\ \hline \cdot & .000 \end{array}\right]$$

Theorem 6.20 gives a realization of T'^{*} as

$$\mathbf{T}_1' = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline .000 & .000 \end{array}\right] \qquad \mathbf{T}_2' = \left[\begin{array}{cc|c} .021 & -.517 & -.965 \\ \hline -.005 & .125 & .233 \end{array}\right]$$

$$\mathbf{T}_3' = \left[\begin{array}{ccc|c} -.124 & -.517 & .000 & -1.161 \\ .025 & -.012 & -.084 & -.654 \\ \hline .074 & .281 & -.018 & .494 \end{array}\right] \qquad \mathbf{T}_4' = \left[\begin{array}{ccc|c} -.130 & -.517 & -.000 & -1.168 \\ .023 & -.012 & -.084 & -.657 \\ .296 & .836 & -.212 & -.183 \\ \hline .084 & .306 & -.024 & .519 \end{array}\right]$$

$$\mathbf{T}_5' = \left[\begin{array}{ccc|c} -.121 & -.509 & -.000 & -1.171 \\ .027 & -.005 & .076 & -.656 \\ .295 & .845 & .192 & -.176 \\ \hline .080 & .303 & .021 & .519 \end{array}\right] \qquad \mathbf{T}_6' = \left[\begin{array}{cc|c} .062 & .441 & -1.098 \\ .114 & -.026 & -.601 \\ -.289 & .897 & .122 \\ \hline -.006 & -.272 & .494 \end{array}\right]$$

$$\mathbf{T}_7' = \left[\begin{array}{c|c} -.702 & -.504 \\ .404 & .290 \\ \hline .324 & .233 \end{array}\right] \qquad \mathbf{T}_8' = \left[\begin{array}{c|c} \cdot & .000 \\ \hline \cdot & .000 \end{array}\right]$$

and the resulting Nehari extension follows as

$$E = T - T' = \begin{bmatrix} 0 & .326 & .566 & .334 & .078 & -.008 & -.012 & -.003 \\ 0 & -.233 & .326 & .566 & .334 & .078 & -.008 & -.012 \\ 0 & .076 & -.494 & .326 & .566 & .334 & .078 & -.008 \\ 0 & .003 & .267 & -.519 & .326 & .566 & .334 & .078 \\ 0 & -.011 & -.050 & .295 & -.519 & .326 & .566 & .334 \\ 0 & .003 & -.013 & -.050 & .267 & -.494 & .326 & .566 \\ 0 & .000 & .003 & -.011 & .003 & .076 & -.233 & .326 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

E is indeed contractive: $\|E\| = .9932$.

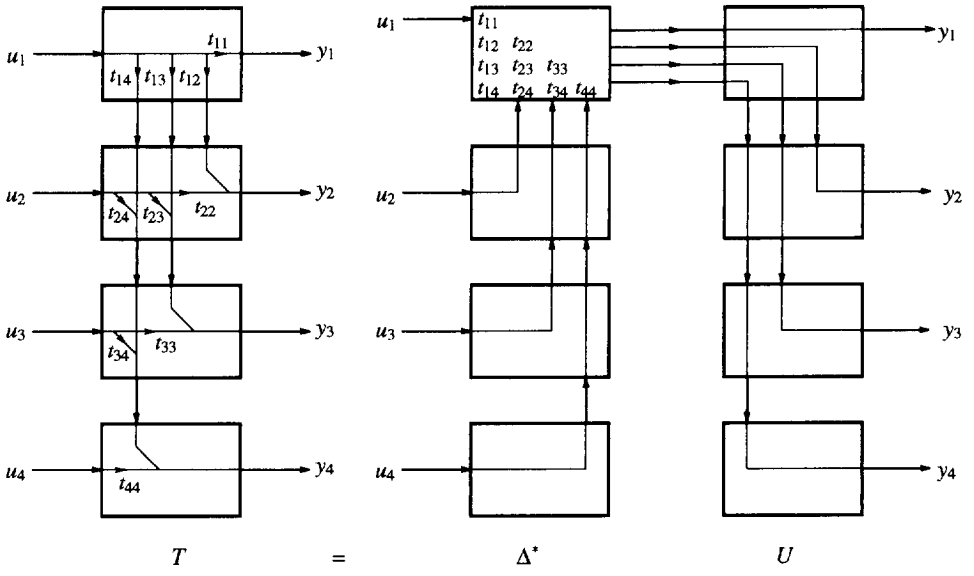


Figure 6.5. Trivial external factorization of T.

6.6 ORDER-RECURSIVE INTERPOLATION

The global state-space procedure of the previous sections constructs, for a given $T \in \mathcal{U}$, an inner factor U and an interpolating operator Θ. The procedure can be specialized and applied to the case where T is a general upper triangular matrix without an a priori known state structure. The specialization produces a generalized Schur recursion, which we derive for an example T.

Consider a 4×4 strictly upper triangular matrix T,

$$T = \begin{bmatrix} \boxed{0} & t_{12} & t_{13} & t_{14} \\ & \underline{0} & t_{23} & t_{24} \\ & & \underline{0} & t_{34} \\ & & & \underline{0} \end{bmatrix},$$

where the (1, 1) entry is indicated by a square and the main diagonal by underscores. For convenience of notation, and without loss of generality, we may take $\Gamma = I$, and thus seek for T_a (a 4×4 matrix) such that $\| T - T_a \| \leq 1$. A trivial (but non-minimal) state realization for T that has $AA^* + CC^* = I$ is obtained by selecting $\{[0\ 0\ 1], [0\ 1\ 0], [1\ 0\ 0]\}$ as a basis for the row space of the second Hankel matrix $H_2 = [t_{12}\ t_{13}\ t_{14}]$, and likewise we select trivial bases for H_3 and H_4. Omitting the details, the realizations for T and an

inner factor U that result from this choice turn out to be

$$\mathbf{T}_1 = \left[\begin{array}{ccc|c} \cdot & \cdot & \cdot & \cdot \\ \hline t_{14} & t_{13} & t_{12} & 0 \end{array}\right] \qquad \mathbf{U}_1 = \left[\begin{array}{ccc|c} \cdot & \cdot & \cdot & \cdot \\ \hline 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{array}\right]$$

$$\mathbf{T}_2 = \left[\begin{array}{cc|c} 1 & & \\ & 1 & \\ \hline & & 1 \\ t_{24} & t_{23} & 0 \end{array}\right] \qquad \mathbf{U}_2 = \left[\begin{array}{cc|c} 1 & & \\ & 1 & \\ \hline & & 1 \\ \cdot & \cdot & \cdot \end{array}\right]$$

$$\mathbf{T}_3 = \left[\begin{array}{c|c} 1 & \\ \hline & 1 \\ t_{34} & 0 \end{array}\right] \qquad \mathbf{U}_3 = \left[\begin{array}{c|c} 1 & \\ \hline & 1 \\ \cdot & \cdot \end{array}\right]$$

$$\mathbf{T}_4 = \left[\begin{array}{c|c} \cdot & 1 \\ \hline \cdot & 0 \end{array}\right] \qquad \mathbf{U}_4 = \left[\begin{array}{c|c} \cdot & 1 \\ \hline \cdot & \cdot \end{array}\right]$$

('·' stands for an entry with zero dimensions). The corresponding matrices U and $\Delta = UT^*$ are

$$U = \begin{bmatrix} \boxed{1} & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix} \qquad \Delta = \begin{bmatrix} \boxed{\begin{array}{c} 0 \\ t_{12}^* \\ t_{13}^* \\ t_{14}^* \end{array}} & \begin{array}{c} 0 \\ t_{23}^* \\ t_{24}^* \end{array} & \begin{array}{c} 0 \\ t_{34}^* \end{array} & 0 \end{bmatrix}$$

with input space sequence $\mathbb{C}^4 \times \emptyset \times \emptyset \times \emptyset$, and output space sequence $\mathbb{C}^1 \times \mathbb{C}^1 \times \mathbb{C}^1 \times \mathbb{C}^1$. All inputs of U and Δ are concentrated at point 1, and hence the causality requirement is always satisfied: $U \in \mathcal{U}$ and $\Delta \in \mathcal{U}$. The structure of Δ and U is clarified by figure 6.5.

The global realization procedure would continue by computing a sequence M

$$M_{k+1} = A_k^* M_k A + B_k^* B_k, \qquad M_1 = [\,\cdot\,]$$

and using this to derive Θ as demonstrated in section 6.2. Note that it is not necessary to have a *minimal* realization for T (or U). The extra states correspond to eigenvalues of M that are zero, and hence are of no influence on the negative signature of $\Lambda = I - M$ (independently of Γ). Hence our non-minimal choice of the realization for T does not influence the complexity of the resulting approximant T_a. For a recursive derivation of an interpolating matrix Θ, however, we proceed as follows. The (trivial) state realizations $\mathbf{T}$ and $\mathbf{U}$ are not needed, but the resulting U is used. The interpolation problem is to determine a J-unitary and causal Θ (whose signature will be determined by the construction)

such that

$$[U^* \quad -T^*]\Theta \in [\mathcal{U} \quad \mathcal{U}].$$

Assume that $\Theta \in \mathcal{U}(\mathcal{M}_\Theta, \mathcal{N}_\Theta)$. The signature matrix $J_1 := J_{\mathcal{M}_\Theta}$ is known from the outset and is according to the decomposition $[U^* \quad -T^*]$. Although the signature $J_2 := J_{\mathcal{N}_\Theta}$ is not yet known at this point, the number of outputs of Θ (*i.e.*, the space sequence $\mathcal{N}_\Theta$) is already determined by the condition that each $\mathbf{\Theta}_k$ is a square matrix. With the above (trivial) realizations of T and U, it follows that Θ has a constant number of two outputs at each point in time. The signature of each output (+1 or −1) is determined in the process of constructing Θ, which is done in two steps: $\Theta = \Theta'\Pi$. Here, Θ' is such that $[U^* \quad -T^*]\Theta' \in [\mathcal{U} \quad \mathcal{U}]$, where the dimension sequences of each $\mathcal{U}$ are constant and equal to 1 at each point; for example

$$\begin{array}{c} + \; + \; + \; + \quad - \quad - \quad - \quad - \\ \left[\begin{array}{cccc|cccc} \boxed{1} & & & & \boxed{-t_{11}^*} & & & \\ & 1 & & & -t_{12}^* & -t_{22}^* & & \\ & & 1 & & -t_{12}^* & -t_{23}^* & -t_{33}^* & \\ & & & 1 & -t_{14}^* & -t_{24}^* & -t_{34}^* & -t_{44}^* \end{array}\right] \end{array} \Theta' = \begin{array}{c} + \; + \quad - \; - \quad + \; + \quad - \; - \\ \left[\begin{array}{cccc|cccc} \boxed{*} & * & * & * & \boxed{*} & * & * & * \\ & \underline{*} & * & * & & \underline{*} & * & * \\ & & \underline{*} & * & & & \underline{*} & * \\ & & & \underline{*} & & & & \underline{*} \end{array}\right] \end{array}$$

where the first upper triangular matrix at the right-hand side corresponds to the first output of each section of Θ', and the second to the second output. At this point, the signature of each column at the right-hand side can be positive of negative: the output signature matrix of Θ' is J_2', which is an *unsorted* signature matrix such that $\Theta' J_2' \Theta'^* = J_1$ (the signature of the right-hand side in the equation above is just an example). See also figure 6.6. The second step is to sort the columns according to their signature, by introducing a permutation matrix $\Pi \in \mathcal{P}$, such that $J_2 = \Pi^* J_2' \Pi$ is a conventional (sorted) signature matrix. The permutation does not change the fact that $[U^* \quad -T^*]\Theta \in [\mathcal{U} \quad \mathcal{U}]$, but the output dimension sequences of each $\mathcal{U}$ are different now, and are in general no longer constant. For the above example signature, $[A' \quad -B']$ has the form

$$\begin{array}{c} + \; + \; + \; + \quad - \quad - \quad - \quad - \\ \left[\begin{array}{cccc|cccc} \boxed{1} & & & & \boxed{-t_{11}^*} & & & \\ & 1 & & & -t_{12}^* & -t_{22}^* & & \\ & & 1 & & -t_{12}^* & -t_{23}^* & -t_{33}^* & \\ & & & 1 & -t_{14}^* & -t_{24}^* & -t_{34}^* & -t_{44}^* \end{array}\right] \end{array} \Theta = \begin{array}{c} + \; + \; + \; + \qquad\qquad - \; - \; - \; - \\ \left[\begin{array}{cccc|cccc} \boxed{*} & * & * & * & \boxed{\cdot} & \cdot & * & * & * & * \\ & \underline{*} & \underline{*} & * & & \doteq & * & * & * & * \\ & & & \doteq & & & * & * & * & * \\ & & & \doteq & & & & \underline{*} & \underline{*} & \end{array}\right] \end{array}$$

$$= [A' \quad -B']$$

where A' has as output sequence $\mathbb{C}^2 \times \mathbb{C}^2 \times \varnothing \times \varnothing$, and B' has as output sequence $\varnothing \times \varnothing \times \mathbb{C}^2 \times \mathbb{C}^2$. We now consider these operations in more detail.

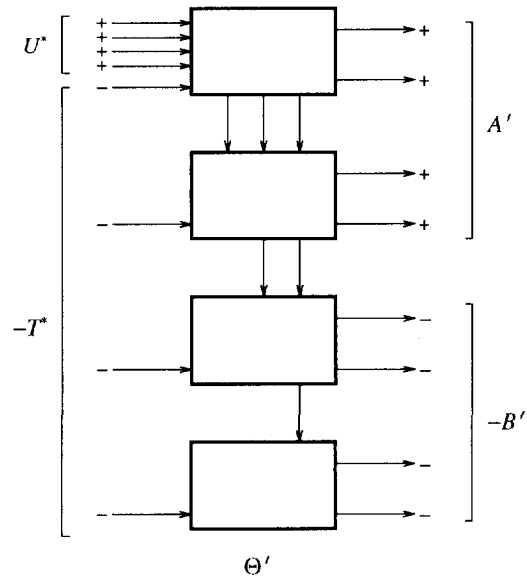


Figure 6.6. Computational structure of Θ', with example signature at the outputs.

Computational structure

Θ' can be determined recursively in $2n$ steps: $\Theta' = \Theta_{(11)}\Theta_{(12)} \cdot \Theta_{(21)}\Theta_{(22)} \cdots \Theta_{(n1)}\Theta_{(n2)}$, in the following way. The columns of Θ' act on the columns of U^* and $-T^*$. Its operations on U^* are always causal because all columns of U^* correspond to the first point of the recursion ($k = 1$). However, for Θ to be causal, the k-th column of Θ can act only on the first k columns of T^*. Taking this into consideration, we are led to a recursive algorithm of the form $[A_{(k)} \;\; B_{(k)}]\Theta_{(k1)}\Theta_{(k2)} = [A_{(k+1)} \;\; B_{(k+1)}]$, initialized by $A_{(1)} = U^*$, $B_{(1)} = -T^*$, and where $\Theta_{(k1)}\Theta_{(k2)}$ involves two actions:

a. Using columns $k+1, \cdots, n$ of $A_{(k)}$, make the last $(n-k)$ entries of the k-th column of $A_{(k)}$ equal to 0. In particular, the $(k+i)$-th column of $A_{(k)}$ is used to make the $(k+i)$-th entry of the k-th column of $A_{(k)}$ equal to zero.

b. Make the last $(n-k)$ entries of the k-th column of $B_{(k)}$ equal to 0, again using columns $k+1, \cdots, n$ of $A_{(k)}$.

The operations required to carry out each of these steps are elementary unitary (Givens) or J-unitary rotations that act on two columns at a time and make a selected entry of the second column equal to zero. The precise nature of a rotation depends on the correspond-

ing signature and is in turn dependent on the data — this will be detailed later. We first verify that this recursion leads to a solution of the interpolation problem.

$k = 1$ *a.* no action is needed: the first column of U^* has already the required form.

b. using 3 elementary rotations, the entries t^*_{12}, t^*_{13}, t^*_{14} are subsequently zeroed. This produces

$$\left[\begin{array}{cccc|c|ccc} 1 & * & * & * & \boxed{*} & & & \\ 0 & * & & & 0 & -t^*_{22} & & \\ 0 & * & * & & 0 & -t^*_{23} & -t^*_{33} & \\ 0 & * & * & * & 0 & -t^*_{24} & -t^*_{34} & -t^*_{44} \end{array}\right]$$

$k = 2$ *a.*

$$\left[\begin{array}{cccc|c|ccc} 1 & * & * & * & \boxed{*} & & & \\ 0 & * & * & * & 0 & -t^*_{22} & & \\ 0 & 0 & * & 0 & 0 & -t^*_{23} & -t^*_{33} & \\ 0 & 0 & * & * & 0 & -t^*_{24} & -t^*_{34} & -t^*_{44} \end{array}\right]$$

b.

$$\left[\begin{array}{cccc|c|cc|c} 1 & * & * & * & \boxed{*} & * & & \\ 0 & * & * & * & 0 & * & & \\ 0 & 0 & * & 0 & 0 & 0 & -t^*_{33} & \\ 0 & 0 & * & * & 0 & 0 & -t^*_{34} & -t^*_{44} \end{array}\right]$$

$k = 3$ *a.*

$$\left[\begin{array}{cccc|c|cc|c} 1 & * & * & * & \boxed{*} & * & & \\ 0 & * & * & * & 0 & * & & \\ 0 & 0 & * & * & 0 & 0 & -t^*_{33} & \\ 0 & 0 & 0 & * & 0 & 0 & -t^*_{34} & -t^*_{44} \end{array}\right]$$

b.

$$\left[\begin{array}{cccc|c|ccc} 1 & * & * & * & \boxed{*} & * & * & \\ 0 & * & * & * & 0 & * & * & \\ 0 & 0 & * & * & 0 & 0 & * & \\ 0 & 0 & 0 & * & 0 & 0 & 0 & -t^*_{44} \end{array}\right]$$

$k = 4$: no rotations are required.

The resulting matrices are upper triangular. The signal flow corresponding to this computational scheme is outlined in figure 6.7(*a*). Note that the computations have introduced an implicit notion of state, formed by the arrows that cross a dotted line between two stages, so that a (non-minimal) realization of Θ can be inferred from the elementary operations.

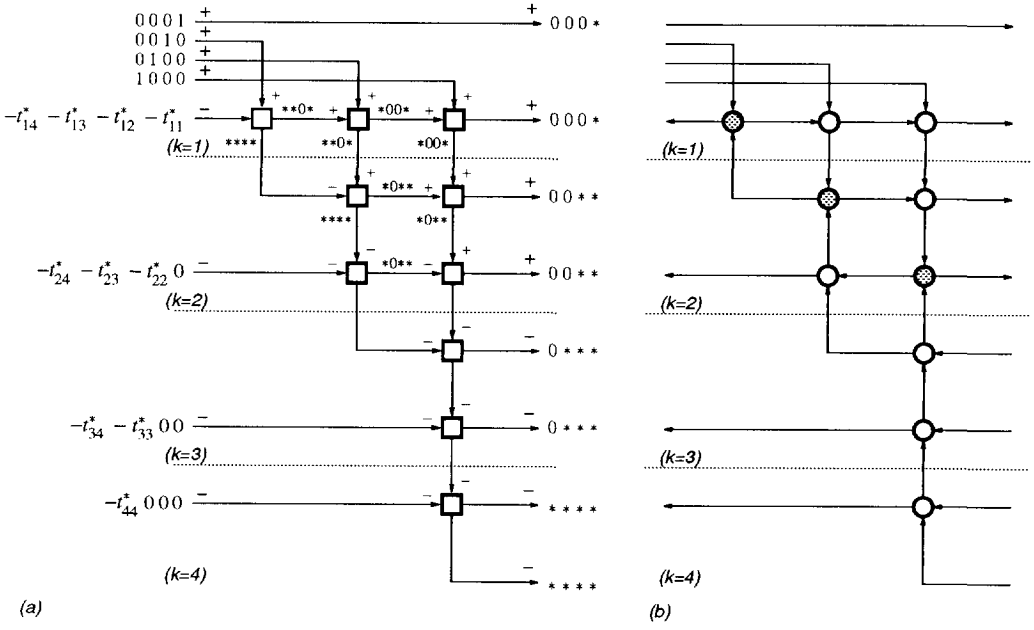


Figure 6.7. Computational structure of a recursive solution to the interpolating problem. (*a*) Θ', with elementary rotations of mixed type (both circular and hyperbolic); (*b*) one possible corresponding Σ', with circular elementary rotations. The type of sections in (*a*) and the signal flow in (*b*) depend on the data of the interpolation problem. The rotations which cause an upward arrow (ultimately: a state for T_a) are shaded.

Elementary rotations: keeping track of signatures

We now consider the elementary operations in the above recursions. An elementary rotation θ such that $\theta^* j_1 \theta = j_2$ (j_1 and j_2 are 2×2 signature matrices) is defined by

$$[u \quad t]\theta = [* \quad 0],$$

where u, t are scalars, and where '$*$' stands for some resulting scalar. Initially, one would consider θ of a traditional J-unitary form:

$$\theta_1 = \begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix} \frac{1}{c^*}, \qquad cc^* + ss^* = 1,\ c \neq 0$$

which satisfies

$$\theta_1^* \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \theta_1 = \begin{bmatrix} 1 & \\ & -1 \end{bmatrix}.$$

However, since $|s| < 1$, a rotation of this form is appropriate only if $|u| > |t|$. In the recursive algorithm, this is the case only if $TT^* < I$ which corresponds to a 'definite' interpolation problem and leads to an approximant $T_a = 0$. Our situation is more general. If $|u| < |t|$, we require a rotational section of the form

$$\theta_2 = \begin{bmatrix} -s & 1 \\ 1 & -s^* \end{bmatrix} \frac{1}{c^*},$$

resulting in $[u \ \ t]\theta_2 = [* \ \ 0]$. θ_2 has signature pairs determined by

$$\theta_2^* \begin{bmatrix} 1 & \\ & -1 \end{bmatrix} \theta_2 = \begin{bmatrix} -1 & \\ & 1 \end{bmatrix}.$$

This shows that the signature of the 'energy' of the output vector of such a section is reversed: if $[a_1 \ \ b_1]\theta_2 = [a_2 \ \ b_2]$, then $a_1 a_1^* - b_1 b_1^* = -a_2 a_2^* + b_2 b_2^*$. Because this signature can be reversed at each elementary step, we have to keep track of it to ensure that the resulting global Θ-matrix is J-unitary with respect to a certain signature. Thus assign to each column in $[U^* \ \ -T^*]$ a signature (+1 or −1), which is updated after each elementary operation, in accordance to the type of rotation. Initially, the signature of the columns of U^* is chosen +1, and those of $-T^*$ are chosen −1. Because $\Theta' = \Theta_{(11)}\Theta_{(12)} \cdots \Theta_{(n1)}\Theta_{(n2)}$, where $\Theta_{(i)}$ is an embedding of the i-th elementary rotation $\theta_{(i)}$ into one of full size, it is seen that keeping track of the signature at each intermediate step ensures that

$$\Theta^* \begin{bmatrix} I & \\ & -I \end{bmatrix} \Theta = J_2',$$

where J_2' is the unsorted signature matrix given by the signatures of the columns of the final resulting upper triangular matrices. The types of signatures that can occur, and the

appropriate elementary rotations to use, are listed below. These form the processors in figure 6.7(*a*).

$$1.\quad \begin{matrix}+ & - \\ u & t\end{matrix}\begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix}\frac{1}{c^*} = \begin{matrix}+ & - \\ * & 0\end{matrix}, \qquad \text{if } |u|>|t|$$

$$2.\quad \begin{matrix}+ & - \\ u & t\end{matrix}\begin{bmatrix} -s & 1 \\ 1 & -s^* \end{bmatrix}\frac{1}{c^*} = \begin{matrix}- & + \\ * & 0\end{matrix}, \qquad \text{if } |u|<|t|$$

$$3.\quad \begin{matrix}- & + \\ u & t\end{matrix}\begin{bmatrix} -s & 1 \\ 1 & -s^* \end{bmatrix}\frac{1}{c^*} = \begin{matrix}+ & - \\ * & 0\end{matrix}, \qquad \text{if } |u|>|t|$$

$$4.\quad \begin{matrix}- & + \\ u & t\end{matrix}\begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix}\frac{1}{c^*} = \begin{matrix}- & + \\ * & 0\end{matrix}, \qquad \text{if } |u|<|t|$$

$$5.\quad \begin{matrix}+ & + \\ u & t\end{matrix}\begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} = \begin{matrix}+ & + \\ * & 0\end{matrix}$$

$$6.\quad \begin{matrix}- & - \\ u & t\end{matrix}\begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix} = \begin{matrix}- & - \\ * & 0\end{matrix}$$

We can associate, as usual, with each J-unitary rotation a corresponding unitary rotation, which is obtained by rewriting the corresponding equations such that the '+' quantities appear on the left-hand side and the '–' quantities on the right-hand side. The last two sections are already circular rotation matrices. By replacing each of the sections of Θ by the corresponding unitary section, a unitary Σ matrix that corresponds to Θ is obtained. A signal flow scheme of a possible Σ in our 4×4 example is depicted in figure 6.7(*b*). The matching of signatures at each elementary rotation in the algorithm effects in figure 6.7(*b*) that the signal flow is well defined: an arrow leaving some section will not bounce into a signal flow arrow that leaves a neighboring section.

Finally, a solution to the interpolation problem $[U^* \quad -T^*]\Theta = [A' \quad -B']$ is obtained by *sorting* the columns of the resulting upper triangular matrices obtained by the above procedure according to their signature, such that all positive signs correspond to A' and all negative signs to B'. The columns of Θ are sorted likewise. The solution that is obtained this way is reminiscent of the state-space solution in the previous section, and in fact can be derived from it by factoring Θ into elementary operations as above. Again, the network of Σ is not computable since it contains loops.

To give an example of the foregoing, suppose that T is a band matrix. It may be verified that computations on entries off the band reduce to identity operations and can therefore be omitted. The corresponding computational scheme is, for a typical example, depicted in figure 6.8. A number of '0' entries that are needed to match the sequences in the correct way have been suppressed in the figure: a sequence that is introduced at level k

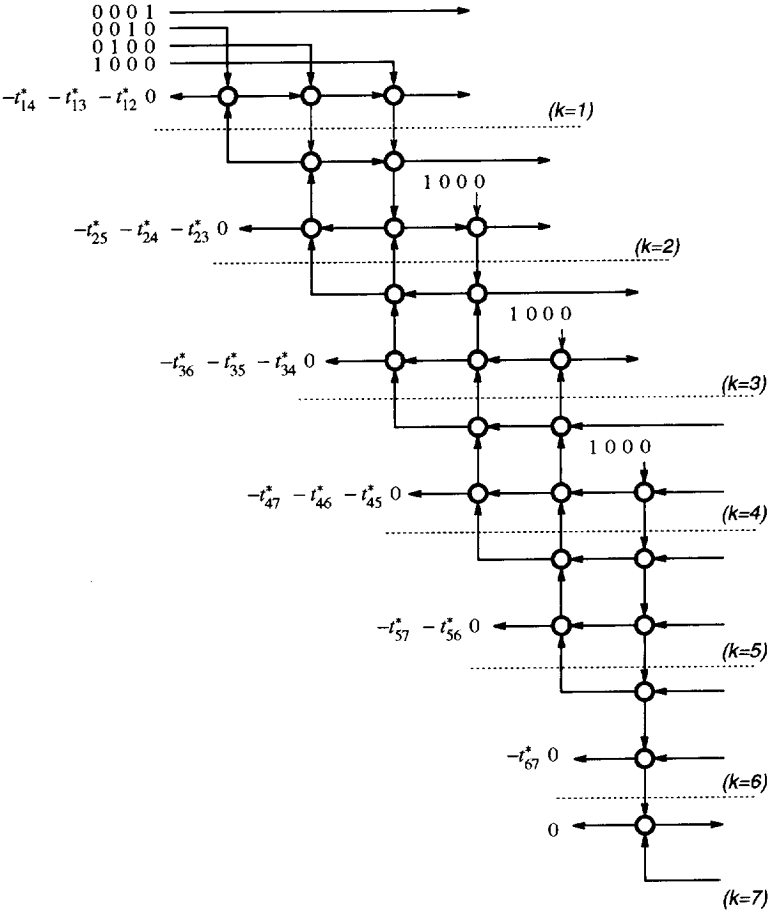


Figure 6.8. Computational network of an interpolating Σ-matrix of a band-matrix (7×7 matrix, band width 3).

must in fact be delayed by $k-1$ delays (equivalent to prepending $k-1$ '0's), while as many trailing '0's as needed must be postpended to make each sequence have length 7. The recursive procedure can be specialized even further to handle staircase matrices as well, for which even more of the elementary computations are rendered trivial and can be omitted. The structure of the diagram will reflect the structure of the staircase.

The recursion and the resulting computational network is a further generalization (to include indefinite interpolation) of the generalized Schur algorithm introduced in [30]. However, the formalism by which the matrices are set up to initiate the algorithm is new.

Computation of the approximant

With Θ and B' available, there are various ways to obtain the Hankel norm approximant T_a. The basic relations are given in terms of T' (the upper triangular part of which is equal to T_a) and the operator Σ associated to Θ:

$$\begin{aligned} T'^* &= T^* + U^*\Sigma_{12} \\ T'^* &= B'\Theta_{22}^{-1}, \qquad \Theta_{22}^{-1} = \Sigma_{22}. \end{aligned}$$

Ideally, one would want to use the computational network of Σ to derive either $U^*\Sigma_{12}$ or $B'\Theta_{22}^{-1}$. However, the network that has been constructed in the previous step of the algorithm is not *computable*: it contains delay-free loops, and hence it cannot be used directly. A straightforward alternative is to extract Θ_{22} from the network of Θ (by applying an input of the form $[0\ I]$), and subsequently use any technique to invert this matrix and apply it to B'. A second alternative is to compute a (non-causal) state realization for Σ from its network. This is a local operation: it can be done independently for each stage. From this realization, one can derive a realization for the upper triangular part of Θ_{22}^{-*}, by using the recursions given in section 6.3.

The first solution can be made more or less 'in style' with the way Θ has been constructed, to the level that only elementary, unitary operations are used. However, the overall solution is a bit crude: after extracting the matrix Θ_{22}, the computational network of Θ is discarded, although it reveals the structure of Θ_{22} and Θ_{22}^{-1}, and the algorithm continues with a matrix inversion technique that is not very specific to its current application. The state-space technique, on the other hand, uses half of the computational network structure of Θ (the 'vertical' segmentation into stages), but does not use the structure within a stage. The algorithm operates on (state-space) matrices, rather than at the elementary level, and is in this respect 'out of style' with the recursive computation of Θ. It is as yet unclear whether an algorithm can be devised that acts directly on the computational network of Θ using elementary operations.

6.7 CONCLUSIONS

In this chapter, we have presented an approximation scheme to derive, for a given upper triangular matrix T, a Hankel-norm approximant T_a of lower complexity. A model of T_a can be computed starting from a high-order model of T (obtained *e.g.*, by algorithm 3.1) by applying algorithm 6.2. However, the derivation of a model for T can be computationally intensive: it involves a sequence of SVDs to compute the relevant subspaces. An alternative approach is via the algorithm discussed in section 6.6, which acts directly on the entries of T. Only local computations are needed to obtain Θ and B'. However, further research is required to efficiently compute T_a as the upper part of $(B'\Theta_{22}^{-1})^*$: a direct computation is not really satisfactory in view of the fact that Θ is obtained in a factored form.

A second open problem is the selection of a suitable error tolerance matrix Γ. At present, one has to select some Γ, which then results in an approximant with a certain complexity. It is, as yet, unclear how to obtain the reverse, *i.e.*, how to derive, for a given desired complexity of the approximant, the tolerance Γ that will achieve this complexity.

Bibliography

[1] P.M. Dewilde and A.J. van der Veen, "On the Hankel-norm approximation of upper-triangular operators and matrices," *to appear in Integral Equations and Operator Theory*, 1993.

[2] A.J. van der Veen and P.M. Dewilde, "On low-complexity approximation of matrices," *subm. Linear Algebra and its Applications*, 1992.

[3] V.M. Adamjan, D.Z. Arov, and M.G. Krein, "Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur-Takagi problem," *Math. USSR Sbornik*, vol. 15, no. 1, pp. 31–73, 1971. (transl. of *Iz. Akad. Nauk Armjan. SSR Ser. Mat. 6 (1971)).*

[4] A. Bultheel and P.M. Dewilde, "On the Adamjan-Arov-Krein approximation, identification, and balanced realization," in *Proc. 1980 Eur. Conf. on Circ. Th. and Design*, vol. 2, pp. 186–191, 1980.

[5] S.Y. Kung and D.W. Lin, "Optimal Hankel norm model reductions: Multi-variable systems," *IEEE Trans. Automat. Control*, vol. 26, pp. 832–852, Aug. 1981.

[6] Y.V. Genin and S.Y. Kung, "A two-variable approach to the model reduction problem with Hankel norm criterion," *IEEE Trans. Circuits Syst.*, vol. 28, no. 9, pp. 912–924, 1981.

[7] J.A. Ball and J.W. Helton, "A Beurling-Lax theorem for the Lie group $U(m, n)$ which contains most classical interpolation theory," *J. Operator Theory*, vol. 9, pp. 107–142, 1983.

[8] K. Glover, "All optimal Hankel norm approximations of linear multi-variable systems and their L^∞-error bounds," *Int. J. Control*, vol. 39, no. 6, pp. 1115–1193, 1984.

[9] D.J.N. Limebeer and M. Green, "Parametric interpolation, H_∞-control and model reduction," *Int. J. Control*, vol. 52, no. 2, pp. 293–318, 1990.

[10] L. Pernebo and L.M. Silverman, "Model reduction via balanced state space representations," *IEEE Trans. Automat. Control*, vol. 27, pp. 382–387, Apr. 1982.

[11] B.C. Moore, "Principal component analysis in linear systems: Controllability, observability and model reduction," *IEEE Trans. Automat. Control*, vol. 26, pp. 17–32, Feb. 1981.

[12] K. Glover, R.F. Curtain, and J.R. Partington, "Realization and approximation of linear infinite-dimensional systems with error bounds," *SIAM J. Control and Optimization*, vol. 26, no. 4, pp. 863–898, 1988.

[13] J.W. Helton, "Orbit structure of the Möbius transformation semigroup acting on H_∞ (broadband matching)," in *Topics in Functional Analysis*, vol. 3 of *Adv. in Math. Suppl. Studies*, pp. 129–133, Academic Press, 1978.

[14] J.W. Helton e.a., *Operator Theory, Analytic Functions, Matrices, and Electrical Engineering*, vol. 68 of *CBMS regional conference series*. Providence: American Math. Soc., 1987.

[15] P. Dewilde, A.C. Vieira, and T. Kailath, "On a generalized Szegö-Levinson realization algorithm for optimal linear predictors based on a network synthesis approach," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 663–675, Sept. 1978.

[16] P. Dewilde and H. Dym, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Informat. Th.*, vol. 27, pp. 446–461, July 1981.

[17] P. Dewilde and H. Dym, "Lossless chain scattering matrices and optimum linear prediction: The vector case," *Circuit Theory and Appl.*, vol. 9, pp. 135–175, 1981.

[18] P. Dewilde and H. Dym, "Lossless inverse scattering, digital filters, and estimation theory," *IEEE Trans. Informat. Th.*, vol. 30, pp. 644–662, July 1984.

[19] H. Dym, *J-Contractive Matrix Functions, Reproducing Kernel Hilbert Spaces and Interpolation*. No. 71 in CBMS regional conference series, Providence: American Math. Soc., 1989.

[20] J.A. Ball, I. Gohberg, and L. Rodman, *Interpolation of Rational Matrix Functions*, vol. 45 of *Operator Theory: Advances and Applications*. Birkhäuser Verlag, 1990.

[21] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Automat. Control*, vol. 26, pp. 301–320, Apr. 1981.

[22] J.W. Helton, "Non-Euclidean functional analysis and electronics," *Bull. of the AMS*, vol. 7, no. 1, pp. 1–64, 1982.

[23] N.I. Akhiezer and I.M. Glazman, *Theory of Linear Operators in Hilbert Space*, vol. I and II. Pitman Publishing Ltd, London, 1981.

[24] P. Dewilde and H. Dym, "Interpolation for upper triangular operators," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 153–260, Birkhäuser Verlag, 1992.

[25] Z. Nehari, "On bounded bilinear forms," *Ann. of Math.*, vol. 65, no. 2, pp. 153–162, 1957.

[26] W. Arveson, "Interpolation problems in nest algebras," *J. Functional Anal.*, vol. 20, pp. 208–233, 1975.

[27] H. Woerdeman, *Matrix and Operator Extensions*. PhD thesis, Dept. Math. Comp Sci., Free University, Amsterdam, The Netherlands, 1989.

[28] I. Gohberg, M.J. Kaashoek, and H.J. Woerdeman, "The band method for positive and strictly contractive extension problems: an alternative version and new applications," *Integral Eq. Operator Th.*, vol. 12, pp. 343–382, 1989.

[29] I. Gohberg, M.A.Kaashoek, and H.J. Woerdeman, "Time variant extension problems of Nehari type and the band method," in *H^∞-Control Theory (lectures given at the 2nd session of C.I.M.E., Como, June 18-26, 1990)* (C. Foias, B. Francis, and J.W. Helton, eds.), Lecture Notes Math. 1496, pp. 309–323, Springer Verlag, 1991.

[30] P. Dewilde and E. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy," in *Operator Theory: Advances and Applications*, vol. 29, pp. 97–116, Birkhäuser Verlag, 1988.

Chapter 7

ORTHOGONAL EMBEDDING

In chapter 3, we saw how a state realization of a time-varying transfer operator T can be computed. The realizations which we obtained were in principle either in input normal form ($A^*A + B^*B = I$) or in output normal form ($AA^* + CC^* = I$). In chapter 4, we considered unitary systems V with unitary realizations. Such realizations are both in input normal form and in output normal form, and satisfy the additional property that both $\|V\| = 1$ and $\|\mathbf{V}\| = 1$, while for $\mathbf{T}$ in either normal form, we have $\|\mathbf{T}\| \geq 1$, whether $\|T\|$ is small or not. Since $\|\mathbf{T}\|$ tells something about the sensitivity of the realization, *i.e.*, the transfer of errors in either the input or the current state to the output and the next state, it is interesting to know whether it is possible to have a realization of T for which $\|\mathbf{T}\| \leq 1$ when $\|T\| \leq 1$. This issue can directly be phrased in terms of the problem which is the topic in this chapter: the *orthogonal embedding problem.* This problem is, given a transfer operator $T \in \mathcal{U}$, to extend this system by adding more inputs and outputs to it such that the resulting system Σ, a 2×2 block operator with entries in $\mathcal{U}$,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

is inner and has T as its partial transfer when the extra inputs are forced to zero: $T = \Sigma_{11}$. See figure 7.1. Since the unitarity of Σ implies $T^*T + T_c^*T_c = I$, (where $T_c = \Sigma_{21}$), it will be possible to find solutions to the embedding problem only if T is contractive: $I - T^*T \geq 0$, so that $\|T\| \leq 1$. Since Σ is inner, it has a unitary realization $\mathbf{\Sigma}$, and a possible realization $\mathbf{T}$ of T is at each point k in time a submatrix of $\mathbf{\Sigma}_k$ (with the same A_k, and smaller dimensional B_k, C_k, D_k), and hence $\mathbf{T}$ is a contractive realization.

The orthogonal embedding problem, and algorithms to solve it, are the central issues in this chapter. The orthogonal embedding problem is known in other fields as well: it is called the unitary extension problem in operator theory, and the equations governing its solution (in a state-space context) are known in control theory as the discrete-time bounded real lemma.

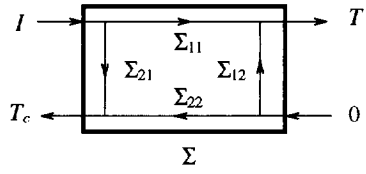


Figure 7.1. Embedding of a contractive time-varying operator T.

7.1 INTRODUCTION AND CONNECTIONS

In this chapter, we present a constructive solution to the embedding problem, under the assumption that the number of states of T is finite at any point in time (locally finite systems). The construction is done in a state-space context and gives rise to (again) a time-varying Riccati equation. While it is clear that the contractivity of T is a necessary condition for the existence of an embedding, we show in the sequel that contractivity of T is, also in the time-varying context, sufficient to construct a solution when T is locally finite. (It is known that not all contractive transfer operators have an orthogonal embedding, see Dewilde [1].) We first derive such a solution for the case where T is strictly contractive. This result has been reported in condensed form in [2, 3]. The extension to the boundary case invokes some mathematical complications but in the end, almost the same algorithm is obtained [4].

Besides the above application, the orthogonal embedding problem is typically the first step in digital filter synthesis problems in which filters (contractive operators) are realized as the partial transfer operator of a lossless multi-port filter Σ. Once such a Σ is obtained, it can be factored into various kinds of "ladder" or "lattice" cascade realizations consisting of elementary lossless degree-1 sections. Such a factorization is known in classical (time-invariant) circuit theory as a Darlington synthesis [5, 6], and provides a structured way to realize a given operator ('filter') in elementary components (in the circuit case, gyrators and a single resistor). In our case, each section is constructed with two elementary (Givens) rotors which have time-varying rotation angles, and the network that is obtained can, for example, be of the form depicted in figure 1.4. In this figure, the transfer function T is from (block) input u_1 to output y_1 if the secondary input u_2 is made equal to zero (the secondary output y_2 is not used). The structural factorization is the topic of chapter 9.

An application of the embedding problem in an operator or linear algebra context is the (Cholesky or spectral) factorization of a positive definite operator Ω into factors $\Omega = W^*W$, where W is an upper operator. The transition to the embedding problem is obtained by a Cayley transformation, which transforms $\Omega > 0$ to an upper strictly contractive operator T: a scattering operator. From the orthogonal embedding Σ, a factor W can be derived

via a few straightforward manipulations. This subsumes the generalized Schur method [7] that has also been used for this application, and in which an embedding Σ is obtained in cascaded form. However, the Schur method is order recursive, and can indeed give rise to a fairly large order, whereas the embedding procedure in this chapter can be used to obtain an embedding Σ and a factor W of minimal order. This connection is described in chapter 8.

The time-invariant orthogonal embedding problem in its simplest form acts on transfer functions $T(z)$ and uses a spectral factorization: with

$$T(z) = \frac{h(z)}{f(z)}, \quad T_c(z) = \frac{g(z)}{f(z)} \tag{7.1}$$

where f, g, h are polynomials of finite degree, it is derived that $g(z)$ (and hence $T_c(z)$) can be determined from a spectral factorization of

$$g(z)g_*(z) = f(z)f_*(z) - h(z)h_*(z)$$

where $f_*(z) = \overline{f(\bar{z}^{-1})}$ [8]. The solution of the spectral factorization problem involves finding the zeros of $g(z)g_*(z)$. Note that in equation (7.1) we use the knowledge that T_c can have the same poles as T.

Polynomial spectral factorization for multi-input/multi-output systems is rather complicated, see *e.g.*, [1]. A solution strategy that is easier to handle (and that carries over to the time-varying case too) is obtained when the problem is cast into a state space context. Such an approach is discussed in [6] for continuous-time systems, and implies what is called the bounded real lemma. This lemma states that $T(s)$ is contractive if and only if certain conditions on the state-space matrices are fulfilled. If this is the case, the conditions are such that they imply a realization for $T_c(s)$ such that $[T(s) \;\; T_c(s)]$ is lossless and has the same A and C matrices as the realization of T. To determine this solution, a Riccati equation has to be solved. The bounded real lemma can without much effort be stated in the discrete-time context by means of a bilinear transformation [9]. A derivation based on the conservation of energy appears in [10], and a proof independent of a continuous-time equivalent is given in [11]. A Riccati equation which describes the problem is stated in [12], which forms the basis of a cascade factorization. Control applications of the bounded real lemma include H_∞-optimal state regulation and state estimation [13].

In the present chapter, the aim is to extend the above classical time-invariant theory to the time-varying context. To introduce the strategy for solving the time-varying embedding problem in a state-space context, consider the following simplified problem. Let T be a single-input, single-output system, with state-space realization $\mathbf{T}$ of constant dimensions. The objective is to determine a lossless embedding system Σ, having two inputs and two

outputs, and with state-space realization $\mathbf{\Sigma}$ of the form

$$\mathbf{\Sigma} = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{cc|c} A & C & C_2 \\ B & D & D_{12} \\ \hline B_2 & D_{21} & D_{22} \end{array}\right] \begin{bmatrix} [R^{(-1)}]^{-1} & & \\ & I & \\ & & I \end{bmatrix},$$

(all entries in this expression are diagonals). $\mathbf{\Sigma}$ contains the given realization $\mathbf{T}$, suitably state-space transformed by some boundedly invertible $R = \mathrm{diag}(R_i)$, which does not alter the input-output characteristics, hence Σ_{11} is equal to the given T. $\mathbf{\Sigma}$ is extended by matrix operators B_2, C_2, D_{21}, D_{12}, D_{22} corresponding to the second input and output. Because Σ is inner, it has a unitary realization $\mathbf{\Sigma}$ (theorem 4.5). Conversely, if $\mathbf{\Sigma}$ is unitary, then the corresponding transfer operator Σ is inner (if $\ell_A < 1$, anyway; see theorem 4.6), and hence a way to solve the embedding problem using state-space methods is to require $\mathbf{\Sigma}$ to be unitary.

The embedding problem is thus reduced to finding the state transformation R, and the embedding matrices B_2 etc., such that $\mathbf{\Sigma}$ is unitary. The problem can be split into two parts:

1. Determine R, B_2, D_{21} to make the columns of $\mathbf{\Sigma}_a$ isometric and orthogonal to each other, with

$$\mathbf{\Sigma}_a = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{cc} A & C \\ B & D \\ \hline B_2 & D_{21} \end{array}\right] \begin{bmatrix} [R^{(-1)}]^{-1} & \\ & I \end{bmatrix}.$$

   That is, $(\mathbf{\Sigma}_a)^* \mathbf{\Sigma}_a = I$.

2. Add one orthonormal column $\mathbf{\Sigma}_b$ to $\mathbf{\Sigma}_a$ to make $\mathbf{\Sigma} = [\, \mathbf{\Sigma}_a \quad \mathbf{\Sigma}_b \,]$ unitary. The realization $\mathbf{\Sigma}$ that is obtained consists of a diagonal sequence of square finite-dimensional matrices, hence this can always be done.

The key step in the above construction is step 1. With the proper attention as to the dimensions of the embedding, it is always possible to find solutions to step 2 since in general, $\mathbf{\Sigma}_b$ is just the orthogonal complement of the columns of $\mathbf{\Sigma}_a$.

The orthonormality conditions of step 1 translate to a set of equations whose solution depends at each time instant i on the (strict) positivity of a matrix $M_i = R_i^* R_i$, which, as we will show, can be computed recursively from the given state-space realization as

$$\begin{aligned} M_{i+1} \;=\; & A_i^* M_i A_i + B_i^* B_i + \\ & \left[A_i^* M_i C_i + B_i^* D_i\right] (I - D_i^* D_i - C_i^* M_i C_i)^{-1} \left[D_i^* B_i + C_i^* M_i A_i\right]. \end{aligned} \tag{7.2}$$

This recursion is again a Riccati-type recursion. The problem with such recursions is the term $(I - D_i^* D_i - C_i^* M_i C_i)$, which can potentially become negative and cause M_{i+1} to

become negative (or indefinite) too. The main contribution of the theory given in the rest of the chapter is to show that the recursion does not break down (*i.e.*, all M_i are uniformly positive, hence we can find a sequence of invertible state-space transformations R_i), under the condition that T is strictly contractive and the given realization for T is uniformly controllable. Subsequently, we show in section 7.5 that a slightly altered recursion also does not break down if T is contractive (but not necessarily in the strict sense), but then we have to impose more requirements on $\mathbf{T}$, for example that it be uniformly observable. These requirements are sufficient but possibly too restrictive.

Preliminary relations

We recall some notations and definitions from chapter 2 and 3, and define some additional ones as well. Let $T \in \mathcal{U}$. We will use the following partial transfer operators on a restricted domain and range (*cf.* equation (3.22)):

$$\begin{array}{lll} H_T: & \mathcal{L}_2Z^{-1} \to \mathcal{U}_2, & u H_T = \mathbf{P}(uT) \\ K_T: & \mathcal{L}_2Z^{-1} \to \mathcal{L}_2Z^{-1}, & uK_T = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(uT) \\ V_T: & \mathcal{L}_2Z^{-1} \to \mathcal{D}_2, & uV_T = \mathbf{P}_0(uT). \end{array}$$

For $u \in \mathcal{L}_2Z^{-1}$ we have that $uT = uK_T + uH_T$. V_T is a further restriction of H_T.

We have already used the fact that H_T is a left D-invariant operator, and hence has 'snapshots' H_i (equation (2.44)), which can be viewed as a sequence of *time-varying* matrices that would have a Hankel structure in the time-invariant case. In the same way, matrix representations are obtained for K_i and vector representations for V_i:

$$H_i = \begin{bmatrix} T_{i-1,i} & T_{i-1,i+1} & T_{i-1,i+2} & \cdots \\ T_{i-2,i} & T_{i-2,i+1} & & \\ T_{i-3,i} & & \ddots & \\ \vdots & & & \end{bmatrix}$$

$$V_i = \begin{bmatrix} T_{i-1,i} \\ T_{i-2,i} \\ T_{i-3,i} \\ \vdots \end{bmatrix} \qquad K_i = \begin{bmatrix} T_{i-1,i-1} & & \mathbf{0} & \\ T_{i-2,i-1} & T_{i-2,i-2} & & \\ T_{i-3,i-1} & T_{i-3,i-2} & T_{i-3,i-3} & \\ \vdots & & \vdots & \ddots \end{bmatrix}.$$

Again because H_T, K_T and V_T are D invariant, they also have diagonal expansions $\tilde{H}_T$, $\tilde{K}_T$ and $\tilde{V}_T$. The general definition of this notion appeared in (2.43), but as in chapter 3, it can be specialized because the domains and ranges are one-sided rather than $\mathcal{X}_2$. Define the diagonal expansions of signals u in $\mathcal{L}_2Z^{-1}$ and y in $\mathcal{U}_2$ as in equations (3.25) and (3.26):

$$\begin{aligned} u &= Z^{-1}u_{[-1]} + Z^{-2}u_{[-2]} + \cdots = u^{(+1)}_{[-1]}Z^{-1} + u^{(+2)}_{[-2]}Z^{-2} + \cdots \\ \tilde{u} &= \begin{bmatrix} u^{(+1)}_{[-1]} & u^{(+2)}_{[-2]} & \cdots \end{bmatrix} \in \ell_2^-(\mathcal{D}). \end{aligned}$$

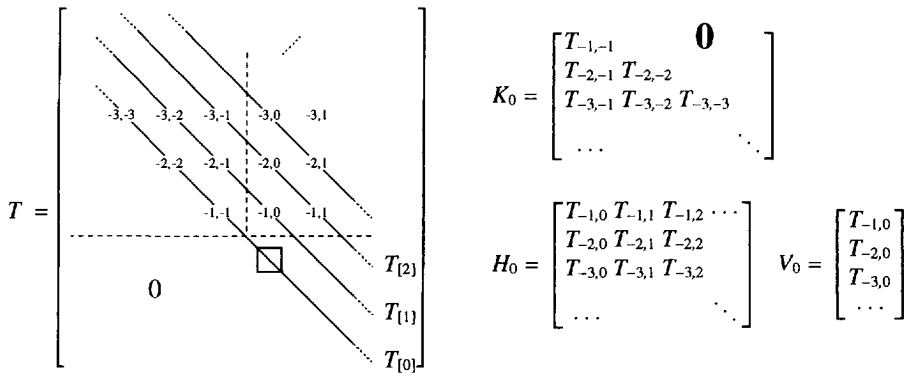


Figure 7.2. K_i, H_i and V_i are submatrices of T.

$$y = y_{[0]} + Zy_{[1]} + Z^2 y_{[2]} + \cdots = y_{[0]} + y_{[1]}^{(-1)} Z + y_{[2]}^{(-2)} Z^2 + \cdots$$
$$\tilde{y} = \begin{bmatrix} y_{[0]} & y_{[1]}^{(-1)} & y_{[2]}^{(-2)} & \cdots \end{bmatrix} \in \ell_2^+(\mathcal{D}).$$

Induced by this isomorphy, the definitions

$$\begin{array}{lcll} y_f = uH_T & \in & \mathcal{U}_2 & \Leftrightarrow \quad \tilde{y}_f = \tilde{u}\tilde{H}_T \\ y_p = uK_T & \in & \mathcal{L}_2 Z^{-1} & \Leftrightarrow \quad \tilde{y}_p = \tilde{u}\tilde{K}_T \\ D = uV_T & \in & \mathcal{D}_2 & \Leftrightarrow \quad D = \tilde{u}\tilde{V}_T. \end{array}$$

lead to

$$\tilde{H}_T = \begin{bmatrix} T_{[1]} & T_{[2]}^{(-1)} & T_{[3]}^{(-2)} & \cdots \\ T_{[2]} & T_{[3]}^{(-1)} & & \\ T_{[3]} & & \ddots & \\ \vdots & & & \end{bmatrix}$$
$$\tilde{V}_T = \begin{bmatrix} T_{[1]} \\ T_{[2]} \\ T_{[3]} \\ \vdots \end{bmatrix} \qquad \tilde{K}_T = \begin{bmatrix} T_{[0]}^{(+1)} & & \mathbf{0} & \\ T_{[1]}^{(+1)} & T_{[0]}^{(+2)} & & \\ T_{[2]}^{(+1)} & T_{[1]}^{(+2)} & T_{[0]}^{(+3)} & \\ \vdots & & \vdots & \ddots \end{bmatrix}. \tag{7.3}$$

As discussed in chapter 3, the connection of $\tilde{H}_T$ with H_i is obtained by selecting the i-th entry of each diagonal in $\tilde{H}_T$ and constructing a matrix from it. Similarly, the sequence K_i forms a matrix representation of the operator K_T and likewise V_i is the vector representation of the operator V_T, obtained by selecting the i-th entry of each diagonal in the representation of $\tilde{V}_T$.

In chapter 3, we frequently made use of the fact that $\tilde{H}_T$ and H_i have decompositions $\tilde{H}_T = \mathcal{C}\mathcal{O}$ and $H_i = \mathcal{C}_i\mathcal{O}_i$. $\mathcal{C}$ and $\mathcal{O}$ are the controllability and observability operators as

defined in (3.28):

$$\mathcal{C} := \begin{bmatrix} B^{(1)} \\ B^{(2)}A^{(1)} \\ B^{(3)}A^{(2)}A^{(1)} \\ \vdots \end{bmatrix} \qquad \mathcal{O} := \begin{bmatrix} C & AC^{(-1)} & AA^{(-1)}C^{(-2)} & \cdots \end{bmatrix}.$$

Since $\tilde{V}_T$ is the first column of $\tilde{H}_T$, we have that $\tilde{V}_T$ has a decomposition

$$\tilde{V}_T = \mathcal{C} \cdot C. \tag{7.4}$$

Finally, it is clear from equation (7.3) that $\tilde{K}_T$ satisfies the relation

$$\tilde{K}_T^{(-1)} = \begin{bmatrix} T_{[0]} & 0 \\ \tilde{V}_T & \tilde{K}_T \end{bmatrix}. \tag{7.5}$$

This relation is seen to correspond to a recursive relation: it specifies that

$$K_{i+1} = \begin{bmatrix} T_{ii} & 0\,0\cdots \\ V_i & K_i \end{bmatrix}$$

for all time instants i. K_i 'grows' when i increases as the history of the system grows — in particular, K_∞ is just a mirrored version of T.

7.2 STRICTLY CONTRACTIVE SYSTEMS

As indicated in the introduction, an orthogonal embedding of a transfer operator $T \in \mathcal{U}$ is possible only if T is at least contractive. In this section, we explore the consequences of assuming the strict contractivity of T, to determine sufficient conditions for an embedding to exist if T is strictly contractive. This is done in two steps. Lemma 7.3 derives a general relation in terms of $\tilde{V}_T$ and $\tilde{K}_T$ which is a direct consequence of the strict contractivity of T. Theorem 7.4 uses this relation to show that some quantity $M \in \mathcal{D}$, defined by $M = \mathcal{C}^*(I - \tilde{K}_T\tilde{K}_T^*)^{-1}\mathcal{C}$, is strictly positive definite, and gives a recursion for this M in terms of state-space quantities of T. The point is that this recursion is precisely the same as the recursion for M in the embedding problem (*viz.* equation (7.2)). This proves the essential step in the embedding problem for strictly contractive operators (section 7.3). The case where T is contractive, but not necessarily strictly contractive, is deferred to section 7.5.

Contractivity of a transfer operator

Recall proposition 2.3 on the positivity, respectively the strict positivity of a Hermitian operator $A \in \mathcal{X}$:

$$\begin{aligned} A \geq 0 \quad &\Leftrightarrow \quad \{uA, u\} \geq 0, \quad (\text{all } u \in \mathcal{X}_2) \\ A \gg 0 \quad &\Leftrightarrow \quad \exists\, \varepsilon > 0 : \{uA, u\} \geq \varepsilon\{u, u\}, \quad (\text{all } u \in \mathcal{X}_2). \end{aligned}$$

Let T be a transfer operator in $\mathcal{U}$. We have defined, in section 2.3, to call T contractive, respectively strictly contractive, if

$$I - TT^* \geq 0, \qquad \text{resp.} \qquad I - TT^* \gg 0.$$

In the latter case, $I - TT^*$ is boundedly invertible. In this section, our focus is on the case that T is strictly contractive. The more general case is treated in section 7.5. $I - TT^* \gg 0$ implies that $I - T^*T \gg 0$, because of the identity $I + T^*(I - TT^*)^{-1}T = (I - T^*T)^{-1}$.

LEMMA 7.1. *If T is strictly contractive, then K_T and $\tilde{K}_T$ are strictly contractive.*

PROOF Let $u \in \mathcal{L}_2\mathcal{Z}^{-1}$, and $y = uK_T$. Since T is strictly contractive, we have from the above definition that

$$\begin{aligned} \mathbf{P}_0(uu^*) - \mathbf{P}_0(yy^*) &= \mathbf{P}_0\left[u(I - K_TK_T^*)u^*\right] \\ &\geq \mathbf{P}_0\left[u(I - TT^*)u^*\right] \\ &\geq \varepsilon\mathbf{P}_0(uu^*) \qquad (\text{some } \varepsilon > 0). \end{aligned}$$

Since, by definition of the diagonal expansion, $\mathbf{P}_0(uu^*) = \tilde{u}\tilde{u}^*$ and $\mathbf{P}_0(yy^*) = \tilde{y}\tilde{y}^*$, and by definition of $\tilde{K}_T$, $\tilde{y} = \tilde{u}\tilde{K}_T$, we obtain that

$$\begin{aligned} \tilde{u}(I - \tilde{K}_T\tilde{K}_T^*)\tilde{u}^* &= \tilde{u}\tilde{u}^* - \tilde{y}\tilde{y}^* \\ &= \mathbf{P}_0(uu^*) - \mathbf{P}_0(yy^*) \\ &\geq \varepsilon\mathbf{P}_0(uu^*) = \varepsilon\tilde{u}\tilde{u}^* \qquad (\text{some } \varepsilon > 0), \end{aligned}$$

which shows that we also have that $\tilde{K}_T$ is strictly contractive: $I - \tilde{K}_T\tilde{K}_T^* \gg 0$, $I - \tilde{K}_T^*\tilde{K}_T \gg 0$. □

The fact that K_T is strictly contractive implies in turn that all K_i are strictly contractive.

Strict contractivity in terms of a state-space realization

The purpose of this section is to find conditions in state-space quantities on the contractivity of a transfer operator T. To this end, we use K_T rather than T, and in particular the fact that $I - K_TK_T^*$ is boundedly invertible and strictly positive when T is contractive. Since $\tilde{K}_T^{(-1)}$ can be specified in terms of $\tilde{K}_T$ and an extra column of diagonals (equation (7.5)), it is possible to derive a (recursive) formula for $(I - \tilde{K}_T\tilde{K}_T^*)^{(-1)}$ in terms of $\tilde{K}_T$ and the newly introduced column. The following lemma is standard and will be instrumental.

LEMMA 7.2. (SCHUR COMPLEMENTS/INVERSION FORMULA) *Let X be a block-partitioned operator,*

$$X = \begin{bmatrix} A & B^* \\ B & C \end{bmatrix},$$

where A, B and C are bounded operators on Hilbert spaces, and let A and C be self-adjoint. Then

$$X \gg 0 \quad \Leftrightarrow \quad \begin{cases} (1) & C \gg 0 \\ (2) & A - B^* C^{-1} B \gg 0. \end{cases}$$

If $X \gg 0$, then

$$\begin{bmatrix} A & B^* \\ B & C \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -C^{-1}B & I \end{bmatrix} \begin{bmatrix} (A - B^*C^{-1}B)^{-1} & 0 \\ 0 & C^{-1} \end{bmatrix} \begin{bmatrix} I & -B^*C^{-1} \\ 0 & I \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & C^{-1} \end{bmatrix} + \begin{bmatrix} I \\ -C^{-1}B \end{bmatrix} (A - B^*C^{-1}B)^{-1} \begin{bmatrix} I & -B^*C^{-1} \end{bmatrix}.$$

PROOF $X \gg 0$ implies that $C \gg 0$. If $C \gg 0$, then C^{-1} exists, and

$$\begin{bmatrix} A & B^* \\ B & C \end{bmatrix} = \begin{bmatrix} I & B^*C^{-1} \\ & I \end{bmatrix} \begin{bmatrix} A - B^*C^{-1}B & \\ & C \end{bmatrix} \begin{bmatrix} I & \\ C^{-1}B & I \end{bmatrix}$$

Because the first and third factors in this decomposition are invertible,

$$\begin{bmatrix} A & B^* \\ B & C \end{bmatrix} \gg 0 \quad \Leftrightarrow \quad \begin{bmatrix} A - B^*C^{-1}B & \\ & C \end{bmatrix} \gg 0$$

$$\Leftrightarrow \quad \begin{cases} (1) & C \gg 0 \\ (2) & A - B^*C^{-1}B \gg 0. \end{cases}$$

This proves the first part of the lemma. The second part is immediate from the above factorization of X. □

LEMMA 7.3. *Let be given a transfer operator $T \in \mathcal{U}$. If T is strictly contractive, then*

$$I - T_{[0]}^* T_{[0]} - \tilde{V}_T^* (I - \tilde{K}_T \tilde{K}_T^*)^{-1} \tilde{V}_T \gg 0.$$

PROOF Since T is strictly contractive, lemma 7.1 ensures that $\tilde{K}_T$ and $\tilde{K}_T^{(-1)}$ are also strictly contractive. Using equation (7.5), we have that

$$I - \tilde{K}_T^{(-1)*} \tilde{K}_T^{(-1)} = \begin{bmatrix} I - T_{[0]}^* T_{[0]} - \tilde{V}_T^* \tilde{V}_T & -\tilde{V}_T^* \tilde{K}_T \\ -\tilde{K}_T^* \tilde{V}_T & I - \tilde{K}_T^* \tilde{K}_T \end{bmatrix} \tag{7.6}$$

Now apply lemma 7.2. It is seen that this expression is positive definite if and only if

$$\begin{cases} (1) & I - \tilde{K}_T^* \tilde{K}_T \gg 0 \\ (2) & I - T_{[0]}^* T_{[0]} - \tilde{V}_T^* \tilde{V}_T - \tilde{V}_T^* \tilde{K}_T (I - \tilde{K}_T^* \tilde{K}_T)^{-1} \tilde{K}_T^* \tilde{V}_T \gg 0. \end{cases}$$

The first condition is satisfied because T is strictly contractive. The second condition is equal to the result, because of the equality $I + \tilde{K}_T (I - \tilde{K}_T^* \tilde{K}_T)^{-1} \tilde{K}_T^* = (I - \tilde{K}_T \tilde{K}_T^*)^{-1}$. □

THEOREM 7.4. *Let $T \in \mathcal{U}$ be a locally finite transfer operator with state realization $\{A, B, C, D\}$, where $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ is strictly stable ($\ell_A < 1$). If T is strictly contractive, then $M \in \mathcal{D}(\mathcal{B}, \mathcal{B})$, defined by*

$$M = \mathcal{C}^*(I - \tilde{K}_T\tilde{K}_T^*)^{-1}\mathcal{C}, \tag{7.7}$$

satisfies the relations $I - D^*D - C^*MC \gg 0$, and

$$M^{(-1)} = A^*MA + B^*B + \left[A^*MC + B^*D\right](I - D^*D - C^*MC)^{-1}\left[D^*B + C^*MA\right].$$

If in addition the state-space realization is uniformly controllable, then $M \gg 0$.

PROOF M is well defined if T is strictly contractive, which also implies that $M \geq 0$. If in addition the state-space realization is uniformly controllable, $\mathcal{C}^*\mathcal{C} \gg 0$, then $M \gg 0$ and hence M is invertible.

With the definition of M and using the fact that $D = T_{[0]}$ and $\tilde{V}_T = \mathcal{C} \cdot C$ (equation (7.4)), the positivity of $I - D^*D - C^*MC$ follows directly from lemma 7.3.

The recursive relation for M is obtained by an application of Schur's inversion formula (lemma 7.2) to equation (7.6), which gives

$$\left[I - \tilde{K}_T^{(-1)*}\tilde{K}_T^{(-1)}\right]^{-1} = \begin{bmatrix} 0 & \\ & (I - \tilde{K}_T^*\tilde{K}_T)^{-1} \end{bmatrix} + \begin{bmatrix} I \\ (I - \tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\tilde{V}_T \end{bmatrix} \Phi^{-2} \begin{bmatrix} I & \tilde{V}_T^*\tilde{K}_T(I - \tilde{K}_T^*\tilde{K}_T)^{-1} \end{bmatrix}$$

with

$$\begin{aligned} \Phi^2 &= I - T_{[0]}^*T_{[0]} - \tilde{V}_T^*\tilde{V}_T - \tilde{V}_T^*\tilde{K}_T(I - \tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\tilde{V}_T \\ &= I - D^*D - C^*MC. \end{aligned}$$

The invertibility of Φ^2 was already shown. Inserting this expression into the definition of $M^{(-1)}$, and using the relations that have been summarized above, $M^{(-1)}$ is obtained as

$$\begin{aligned} M^{(-1)} &= \mathcal{C}^{(-1)*}\left[I - \tilde{K}_T^{(-1)}\tilde{K}_T^{(-1)*}\right]^{-1}\mathcal{C}^{(-1)} \\ &= \mathcal{C}^{(-1)*}\left[I + \tilde{K}_T^{(-1)}\left(I - \tilde{K}_T^{(-1)*}\tilde{K}_T^{(-1)}\right)^{-1}\tilde{K}_T^{(-1)*}\right]\mathcal{C}^{(-1)} \\ &= \begin{bmatrix} B^* & A^*\mathcal{C}^* \end{bmatrix}\begin{bmatrix} B \\ \mathcal{C}A \end{bmatrix} + \begin{bmatrix} B^* & A^*\mathcal{C}^* \end{bmatrix}\begin{bmatrix} T_{[0]} & \\ \tilde{V}_T & \tilde{K}_T \end{bmatrix} \cdot \\ &\quad \cdot \left(\begin{bmatrix} 0 & \\ & (I - \tilde{K}_T^*\tilde{K}_T)^{-1} \end{bmatrix} + \begin{bmatrix} I \\ (I - \tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\tilde{V}_T \end{bmatrix} \Phi^{-2} \begin{bmatrix} I & \tilde{V}_T^*\tilde{K}_T(I - \tilde{K}_T^*\tilde{K}_T)^{-1} \end{bmatrix}\right) \cdot \\ &\quad \cdot \begin{bmatrix} T_{[0]} & \\ \tilde{V}_T & \tilde{K}_T \end{bmatrix}^* \begin{bmatrix} B \\ \mathcal{C}A \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&= B^*B + A^*\mathcal{C}^*\mathcal{C}A \quad + \quad A^*\mathcal{C}^*\tilde{K}_T(I-\tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\mathcal{C}A + \\
&\quad + \left(B^*D + A^*\mathcal{C}^*\left[I + \tilde{K}_T(I-\tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\right]\mathcal{C}C\right)\cdot\Phi^{-2}\cdot \\
&\quad \cdot\left(D^*B + C^*\mathcal{C}^*\left[I + \tilde{K}_T^*(I-\tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T\right]\mathcal{C}A\right) \\
&= B^*B + A^*MA \quad + \quad (A^*MC + B^*D)\Phi^{-2}(D^*B + C^*MA).
\end{aligned}
$$

□

The above theorem gives sufficient conditions for the existence of an orthogonal embedding for the case of strictly contractive operators (theorem 7.5): M plays a crucial role in the construction of such an embedding. It also furnishes part of the proof of the bounded real lemma, which appears in section 7.5.

7.3 STRICTLY CONTRACTIVE SYSTEMS

In this section, we solve the embedding problem as defined in the introduction: given a bounded causal transfer operator of a locally finite system T, the objective is to determine a lossless system Σ such that $\Sigma_{11} = T$. The strategy is as outlined in the introduction: the prime quantity to be determined is a state transformation operator R such that the transformed realization of T is part of the realization of Σ. We consider the case where T is strictly contractive in this section. The more general case appears at the end of section 7.5 (theorem 7.15).

THEOREM 7.5. (EMBEDDING THEOREM, I) *Let $T \in \mathcal{U}(\mathcal{M}_1, \mathcal{N}_1)$ be a locally finite transfer operator with state realization $\mathbf{T} = \{A, B, C, D\}$ such that $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ is strictly stable: $\ell_A < 1$. If T is strictly contractive and $\mathbf{T}$ is uniformly controllable, then the orthogonal embedding problem has a solution $\Sigma \in \mathcal{U}(\mathcal{M}_1 \times \mathcal{M}_2, \mathcal{N}_1 \times \mathcal{N}_2)$ such that Σ is inner and $\Sigma_{11} = T$. The newly introduced dimension sequences of Σ are specified by*

$$
\begin{aligned}
\#(\mathcal{M}_2) &= \#(\mathcal{N}_1), \\
\#(\mathcal{N}_2) &= \#(\mathcal{B}) - \#(\mathcal{B}^{(-1)}) + \#(\mathcal{M}_1).
\end{aligned}
$$

Σ has a unitary realization $\mathbf{\Sigma} = \{A_\Sigma, B_\Sigma, C_\Sigma, D_\Sigma\}$ where A_Σ is state equivalent to A by a boundedly invertible state transformation.

PROOF The proof is by construction, and follows the outline in two steps as presented in section 7.1, but suitably adapted for the general case. Let $\mathbf{\Sigma}$ be of the form

$$
\begin{aligned}
\mathbf{\Sigma} &= \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{cc|c} A & C & C_2 \\ B & D & D_{12} \\ \hline B_2 & D_{21} & D_{22} \end{array}\right] \begin{bmatrix} R^{(-1)} & & \\ & I & \\ & & I \end{bmatrix} \qquad (7.8) \\
&= [\mathbf{\Sigma}_a \;\; \mathbf{\Sigma}_b]
\end{aligned}
$$

in which $R \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ is a boundedly invertible state transformation, and R, B_2, D_{12}, D_{21}, D_{22} are to be determined such that $\mathbf{\Sigma}$ is unitary: $\mathbf{\Sigma}^*\mathbf{\Sigma} = I$, $\mathbf{\Sigma}\mathbf{\Sigma}^* = I$.

The first step is to determine R, B_2 and D_{21} such that $(\mathbf{\Sigma}_a)^*\mathbf{\Sigma}_a = I$, where

$$\mathbf{\Sigma}_a = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{cc} A & C \\ B & D \\ \hline B_2 & D_{21} \end{array}\right] \begin{bmatrix} R^{(-1)} & \\ & I \end{bmatrix}.$$

Putting $M = R^*R$, the orthogonality conditions become

$$\begin{cases} A^*MA + B^*B + B_2^*B_2 &= M^{(-1)} \\ C^*MC + D^*D + D_{21}^*D_{21} &= I \\ A^*MC + B^*D + B_2^*D_{21} &= 0. \end{cases} \tag{7.9}$$

The previous section has provided the material necessary to prove that there is a strictly positive M such that the above equations have a solution, for the case where T is strictly contractive.

Step 1. If T is strictly contractive and $\mathbf{T}$ is uniformly controllable, then a solution of the equations (7.9) is given by M in equation (7.7). M satisfies the recursion

$$\begin{aligned} M^{(-1)} &= A^*MA + B^*B \; + \\ &+ \left[A^*MC + B^*D\right](I - D^*D - C^*MC)^{-1}\left[D^*B + C^*MA\right]. \end{aligned} \tag{7.10}$$

and is strictly positive definite. $B_2 \in \mathcal{D}(\mathcal{N}_1, \mathcal{B}^{(-1)})$ and $D_{21} \in \mathcal{D}(\mathcal{N}_1, \mathcal{N}_1)$ are given by

$$\begin{cases} D_{21} &= (I - D^*D - C^*MC)^{\frac{1}{2}} \\ B_2 &= -(I - D^*D - C^*MC)^{-\frac{1}{2}}\left[D^*B + C^*MA\right] \end{cases} \tag{7.11}$$

Proof of step 1. Put $M = C^*(I - \tilde{K}_T\tilde{K}_T^*)^{-1}C$. Then theorem 7.4 ensures that $M \gg 0$, $I - D^*D - C^*MC \gg 0$, and that M satisfies (7.10). Hence D_{21} and B_2 as defined in equations (7.11) are well defined. It remains to verify that this M, D_{21} and B_2 satisfy equations (7.9). But this is immediate.

Step 2. Augment $\mathbf{\Sigma}_a$ with $\mathbf{\Sigma}_b$, such that the resulting $\mathbf{\Sigma}$ is a diagonal operator whose diagonal entries are square matrices. This is always possible, and can be done independently for each time instant. The index sequence of the number of columns that must be added is equal to

$$\#(\mathcal{N}_2) = \#(\mathcal{B}) - \#(\mathcal{B}^{(-1)}) + \#(\mathcal{M}_1). \tag{7.12}$$

Proof of step 2. The extension of a rectangular isometric matrix to a unitary matrix by adding columns is a standard linear algebra procedure that always has a solution. The 'target size' of $\mathbf{\Sigma}$ is given by $[\#(\mathcal{B}) + \#(\mathcal{M}_1) + \#(\mathcal{N}_1)]$, and the number of columns of $\mathbf{\Sigma}_a$ is $\left[\#(\mathcal{B}^{(-1)}) + \#(\mathcal{N}_1)\right]$, hence the number of columns to be added is equal to $\#(\mathcal{N}_2)$ as given in (7.12). This number is non-negative because the columns of $\mathbf{\Sigma}_a$ are linearly independent. □

7.4 NUMERICAL ISSUES

Initial point for the recursion

Suppose that we are given a realization of a system T that meets the requirements of the embedding theorem. How do we go about determining a realization of Σ? The embedding theorem is constructive, and $\mathbf{\Sigma}_i$ (the realization of Σ at time instant i) can be determined from knowledge of $\mathbf{T}_i$ and both M_i and M_{i+1}. In addition, equation (7.10) can be used to determine M_{i+1} from M_i:

$$M_{i+1} = A_i^* M_i A_i + B_i^* B_i + \left[A_i^* M_i C_i + B_i^* D_i\right] (I - D_i^* D_i - C_i^* M_i C_i)^{-1} \left[D_i^* B_i + C_i^* M_i A_i\right] , \quad (7.13)$$

and this is the only recursive aspect of the procedure. The single missing item is the initial point of this recursion: the value of $M_{-\infty}$, or rather M_{k_0}, where k_0 is the point in time at which the solution of the embedding problem starts to be of interest.

It is possible to find an initial value for the recursion for certain specific time-varying systems. The first (and simplest) class is the case where the state dimension of T is zero at a certain point in time k_0. Consider, for example, a finite $n \times n$ upper triangular (block)-matrix T, then the input space sequence is

$$\mathcal{M}_1 = \cdots \times \varnothing \times \varnothing \times \underbrace{\boxed{\mathbf{C}} \times \mathbf{C} \times \cdots \times \mathbf{C}}_{n} \times \varnothing \times \varnothing \times \cdots$$

and output space sequence $\mathcal{N}_1 = \mathcal{M}_1$. A controllable realization of T obviously has a state-space sequence $\mathcal{B}$ also with $\mathcal{B}_i = \varnothing$ for $(i < 0, i \geq n)$, and hence an initial value of the recursion for M is $M_0 = [\,\cdot\,]$.

A second example is the case where T is time invariant before a certain point in time ($i = 0$ say). T has a time-invariant realization $\{a, b, c, d\}$ for $i < 0$, and there is a time-invariant solution for M also: $M_{i+1} = M_i =: m$ $(i < 0)$. The recursion (7.13) becomes an eigenvalue (Riccati) equation

$$m = a^* m a + b^* b + \left[a^* m c + b^* d\right] (I - d^* d - c^* m c)^{-1} \left[d^* b + c^* m a\right] . \quad (7.14)$$

This equation has exact solutions m which can be obtained in a number of ways: either analytically from the eigenvectors of a corresponding (Hamiltonian) matrix (*cf.* [6]), or numerically by using a Newton-Raphson iteration. An overview of these and other methods can be found in the collection [14]. It is well known that the solution via a Hamiltonian equation usually gives more than one solution that satisfies the Riccati equation; the solution $M = \mathcal{C}^*(I - \tilde{K}_T \tilde{K}_T^*)^{-1}\mathcal{C}$ corresponds to the 'stable' solution (corresponding to eigenvalues of the Hamiltonian matrix that are smaller than 1). The stable solution is also the only solution of the Riccati equation that is stable to a small perturbation when it is plugged in the Riccati recursion (7.13). In fact, one way to solve (7.14) is to use

the *recursion* (7.13) for an initial value of $M_{-\infty} = 0$, and to iterate till convergence. It is known that this occurs if the eigenvalues of a are strictly smaller than 1, and that the recursion will monotonically converge to the 'stable' solution of the Riccati equation.

We can do the same for time-varying systems, which will then apply to other specific situations as well, such as periodic systems. The claim is that if $M_0' = 0$ is taken as the initial value of the recursion (7.13) which gives a sequence M_i', then $M_i' \to M_i$ as $i \to \infty$. An elegant proof is possible, not based on numerical properties of the Riccati equation but rather on the knowledge that $M = \mathcal{C}^*(I - \tilde{K}_T\tilde{K}_T^*)\mathcal{C}$ is the solution of the recursion that we are looking. Details of this proof are however cumbersome because many time indices will appear, but we give an outline of it below. (A formal proof of convergence of a related Riccati equation appears in section 8.4.

PROPOSITION 7.6. *Let $\{A, B, C, D\}$ be a strictly stable realization ($\ell_A < 1$) of a locally finite strictly contractive transfer operator $T \in \mathcal{U}$. Let $M_i = \mathcal{C}_i^*(I - K_iK_i^*)\mathcal{C}_i$ be the exact solution of the Riccati equation (7.13), and let M_i' be the solution, obtained by starting the recursion with $M_0' = 0$. Then $M_i' \to M_i$ for $i \to \infty$ (strong convergence).*

PROOF (outline). The initial value $M_0' = 0$ is the *exact* initial point of a recursion for M' of a system T' which is related to T: $T_{ij}' = 0$ for $i < 0$, and $T_{ij}' = T_{ij}$ for $i \geq 0$. The sequence M_i corresponds to T and is at each point i in time given by $M_i = \mathcal{C}_i^*(I - K_iK_i^*)^{-1}\mathcal{C}_i$. For $i \geq 0$, we can define a partitioning of K_i and $\mathcal{C}_i$ as

$$K_i = \begin{bmatrix} K_i' & 0 \\ H_0^r & K_0 \end{bmatrix} \qquad \mathcal{C}_i = \begin{bmatrix} \mathcal{C}_i' \\ \mathcal{C}_0 A^{[0..i-1]} \end{bmatrix}$$

where K_i' is an $(i \times i)$ matrix, $\mathcal{C}_i'$ is equal to the first i rows of $\mathcal{C}_i$,

$$A^{[0..n-1]} := A_0 A_1 \cdots A_{n-1},$$

and H_0^r is related to the Hankel operator H_0, but has a finite number (i) of columns, which are in reversed order in comparison with H_0. In terms of these quantities, M' is given at time $i \geq 0$ by $M_i' = \mathcal{C}_i'^*(I - K_i'K_i'^*)^{-1}\mathcal{C}_i'$. Using this decomposition of K_i, and a variant of Schur's inversion lemma (lemma 7.2), one can derive that, for $i \geq 0$,

$$\begin{aligned}(I - K_iK_i^*)^{-1} &= \begin{bmatrix} (I - K_i'K_i'^*)^{-1} & \\ & 0 \end{bmatrix} + \\ &+ \begin{bmatrix} (I - K_i'K_i'^*)^{-1}K_i'(H_0^r)^* \\ I \end{bmatrix} \Phi^{-2} \begin{bmatrix} H_0^r K_i'^*(I - K_i'K_i'^*)^{-1} & I \end{bmatrix}\end{aligned}$$

where

$$\Phi^2 = I - K_0K_0^* - H_0^r(I - K_i'^*K_i')^{-1}(H_0^r)^* \gg 0$$

In: $\{\mathbf{T}_k\}$ (a controllable realization of T, $\|T\| < 1$)
Out: $\{\mathbf{\Sigma}_k\}$ (a unitary realization of embedding Σ)

$R_1 = [\,\cdot\,]$

for $k = 1, \cdots, n$

$$\mathbf{T}_{e,k} = \begin{bmatrix} R_k & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} A_k & C_k \\ B_k & D_k \\ 0 & I \end{bmatrix}$$

$$\mathbf{T}'_{e,k} := \Theta_k \mathbf{T}_{e,k}, \quad \Theta_k \text{ such that } \mathbf{T}'_{e,k}(2,2) = \mathbf{T}'_{e,k}(1,2) = \mathbf{T}'_{e,k}(2,1) = 0$$

$$\mathbf{T}'_{e,k} =: \begin{bmatrix} R_{k+1} & 0 \\ 0 & 0 \\ B_{2,k} & D_{21,k} \end{bmatrix}$$

$$\mathbf{\Sigma}_{1,k} = \begin{bmatrix} R_k & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{cc} A_k & C_k \\ B_k & D_k \\ \hline B_{2,k} & D_{21,k} \end{array}\right] \begin{bmatrix} R_{k+1}^{-1} & \\ & I \end{bmatrix}$$

$$\mathbf{\Sigma}_k = \begin{bmatrix} \mathbf{\Sigma}_{1,k} & \mathbf{\Sigma}_{1,k}^{\perp} \end{bmatrix}$$

end

Algorithm 7.1. The embedding algorithm for finite $n \times n$ matrices.

and hence its inverse is bounded. Inserting the expression for $\mathcal{C}_i$ and defining $H_0^r = \mathcal{C}_0 \mathcal{O}_0^r$ yields

$$M_i = M_i' + \left[\mathcal{C}_i'^*(I - K_i' K_i'^*)^{-1} K_i' (\mathcal{O}_0^r)^* + (A^{[0..i-1]})^*\right] \mathcal{C}_0^* \Phi^{-2} \mathcal{C}_0 \left[\mathcal{O}_0^r K_i'^* (I - K_i' K_i'^*)^{-1} \mathcal{C}_i' + A^{[0..i-1]}\right] .$$

An examination of the term $\mathcal{O}_0^r K_i'^* (I - K_i' K_i'^*)^{-1} \mathcal{C}_i'$ that is more detailed than we wish to include at this point reveals that it consists of a summation of i terms, each of which has a factor $A^{[0..k-1]}$ and $A^{[k+1..i-1]}$ (for $0 \le k \le i$). The stability condition $\ell_A < 1$ implies that $\varepsilon > 0$ exists such that, in the limit, products of the form $A^{[k..k+n]}$ are bounded in norm by $(1-\varepsilon)^n$ which goes to 0 strongly and uniformly in k as $n \to \infty$. Since Φ^{-2} is bounded, this equation gives $M_i' \to M_i$ as $i \to \infty$. □

"Square-root" solution of the Riccati equation

The embedding algorithm can be implemented along the lines of the proof of the embedding theorem. However, as was the case with the solution of the inner-outer factorization problem in chapter 4, the Riccati recursions on M_i can be replaced by more efficient algorithms that recursively compute the square root of M_i, *i.e.*, R_i, instead of M_i itself.

These square-root algorithms have been known for a long time; see *e.g.*, Morf [15] for a list of pre-1975 references. The square-root algorithm is given in algorithm 7.1. The algorithm acts on data known at the k-th step: the state matrices A_k, B_k, C_k, D_k, and the state transformation R_k obtained at the previous step. This data is collected in a matrix $\mathbf{T}_{e,k}$:

$$\mathbf{T}_e = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \begin{bmatrix} A & C \\ B & D \\ 0 & I \end{bmatrix} \tag{7.15}$$

The key of the algorithm is the construction of a J-unitary operator $\Theta \in \mathcal{D}^{3\times 3}$, satisfying $\Theta^* J \Theta = J$, where

$$\Theta = \begin{bmatrix} \Theta_{11} & \Theta_{12} & \Theta_{13} \\ \Theta_{21} & \Theta_{22} & \Theta_{23} \\ \Theta_{31} & \Theta_{32} & \Theta_{33} \end{bmatrix} \qquad J = \begin{bmatrix} I & & \\ & I & \\ & & -I \end{bmatrix},$$

such that certain entries of $\mathbf{T}'_e := \Theta \mathbf{T}_e$ are zero. (We will be brief on the properties of J-unitary operators at this point; more general J-unitary operators are the subject of chapter 5.) It turns out that, because Θ is J-unitarity, we have that $\mathbf{T}'^*_e J \mathbf{T}_e = \mathbf{T}^*_e J \mathbf{T}_e$; writing these equations out, it follows that the remaining non-zero entries of $\mathbf{T}'_e$ are precisely the unknowns $R^{(-1)}$, B_2 and D_{21}.

PROPOSITION 7.7. *Let $T \in \mathcal{U}$ be a strictly contractive operator, and let $\{A, B, C, D\}$ be a uniformly controllable realization of T. Define $\mathbf{T}_e$ as in equation (7.15).*

Then there is a J-unitary operator $\Theta \in \mathcal{D}^{3\times 3}$ such that $\mathbf{T}'_e := \Theta \mathbf{T}_e$ has zeros at the entries (2, 2), (1, 2) *and* (2, 1). $\mathbf{T}'_e$ *is of the form*

$$\mathbf{T}'_e = \Theta T_e = \begin{bmatrix} R^{(-1)} & 0 \\ 0 & 0 \\ B_2 & D_{21} \end{bmatrix}$$

where $M = R^*R$, B_2, D_{21} satisfy the embedding equations (7.9).

PROOF Assume first that such an operator Θ exists. A direct computation reveals that (with $M = R^*R$)

$$\mathbf{T}^* J \mathbf{T} = \begin{bmatrix} A^*MA + B^*B & A^*MC + B^*D \\ (A^*MC + B^*D)^* & -(I - D^*D - C^*MC) \end{bmatrix}$$

$$\mathbf{T}'^* J \mathbf{T}' = \begin{bmatrix} M^{(-1)} - B_2^* B_2 & -D_{21}^* B_2 \\ -B_2^* D_{21} & -D_{21}^* D_{21} \end{bmatrix}$$

Since Θ is J-unitary, we must have $\mathbf{T}^*J\mathbf{T} = \mathbf{T}'^*J\mathbf{T}'$, which produces the relations (7.9):

$$\begin{cases} A^*MA + B^*B + B_2^*B_2 &= M^{(-1)} \\ C^*MC + D^*D + D_{21}^*D_{21} &= I \\ A^*MC + B^*D + B_2^*D_{21} &= 0 \end{cases}$$

i.e., the equations that constituted the Riccati equations. It remains to verify the existence of a J-unitary Θ such that $\mathbf{T}'_e$ has zeros at the entries (2, 2), (1, 2) and (2, 1). Choose Θ of the form

$$\Theta = \Theta_3\Theta_2\Theta_1 = \begin{bmatrix} \Theta_{11}^3 & \Theta_{12}^3 & 0 \\ \Theta_{21}^3 & \Theta_{22}^3 & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} \Theta_{11}^2 & 0 & \Theta_{13}^2 \\ 0 & I & 0 \\ \Theta_{31}^2 & 0 & \Theta_{33}^2 \end{bmatrix} \begin{bmatrix} I & 0 & 0 \\ 0 & \Theta_{22}^1 & \Theta_{23}^1 \\ 0 & \Theta_{32}^1 & \Theta_{33}^1 \end{bmatrix}$$

where the submatrix $\{\Theta_{ij}^3\}_{i,j=1}^2$ is unitary, while the submatrices $\{\Theta_{ij}^2\}$ and $\{\Theta_{ij}^1\}$ are J-unitary with with signature matrix $J_1 = \left[\begin{smallmatrix} I & 0 \\ 0 & -I \end{smallmatrix}\right]$. The submatrices are determined by the requirements

$$\begin{bmatrix} \Theta_{22}^1 & \Theta_{23}^1 \\ \Theta_{32}^1 & \Theta_{33}^1 \end{bmatrix} \begin{bmatrix} D \\ I \end{bmatrix} = \begin{bmatrix} 0 \\ (I - D^*D)^{1/2} \end{bmatrix}$$

$$\begin{bmatrix} \Theta_{11}^2 & \Theta_{13}^2 \\ \Theta_{31}^2 & \Theta_{33}^2 \end{bmatrix} \begin{bmatrix} RC \\ (I - D^*D)^{1/2} \end{bmatrix} = \begin{bmatrix} 0 \\ (I - D^*D - C^*MC)^{1/2} \end{bmatrix}$$

$$\begin{bmatrix} \Theta_{11}^3 & \Theta_{12}^3 \\ \Theta_{21}^3 & \Theta_{22}^3 \end{bmatrix} \cdot \begin{bmatrix} \Theta_{11}^2 & 0 & \Theta_{13}^2 \\ 0 & I & 0 \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \Theta_{22}^1 \\ 0 & \Theta_{32}^1 \end{bmatrix} \begin{bmatrix} RA \\ B \end{bmatrix} = \begin{bmatrix} * \\ 0 \end{bmatrix}.$$

Hence necessary requirements are $I - D^*D \geq 0$ and $I - D^*D - C^*MC \geq 0$, respectively. In the present case, because T is strictly contractive, we know that $I - D^*D \gg 0$ and $I - D^*D - C^*MC \gg 0$, and these conditions ensure that the J-unitary submatrices $\{\Theta_{ij}^1\}$ and $\{\Theta_{ij}^2\}$ are well defined, and for example, of the form of a Halmos extension [16]

$$H(K) = \begin{bmatrix} (I - KK^*)^{-1/2} & 0 \\ 0 & (I - K^*K)^{-1/2} \end{bmatrix} \begin{bmatrix} I & K \\ K^* & I \end{bmatrix}.$$

The unitary submatrix $\{\Theta_{ij}^3\}$ is always well defined. □

It is also a standard technique to factor Θ even further down into elementary (J)-unitary operations that each act on only two scalar entries of $\mathbf{T}_e$, and zero one of them by applying an elementary J-unitary rotation of the form

$$\theta = \frac{1}{c}\begin{bmatrix} 1 & s \\ s & 1 \end{bmatrix}, \qquad c^*c + s^*s = 1.$$

With B_2 and D_{21} known, it is conjectured that it is not really necessary to apply the state transformation by R and to determine the orthogonal complement of $\mathbf{\Sigma}_1$ if, in the end, only a cascade factorization of T is required, much as in [17]. Cascade factorizations are the subject of chapter 9.

7.5 THE BOUNDARY CASE

We now derive an equivalent of theorem 7.4 and 7.5 for the case where T is contractive but not necessarily strictly contractive: $I - TT^* \geq 0$. While the mathematical motivation is more complicated now, the resulting theorem is only slightly altered. In the present case, K_T is not strictly contractive, so that, instead of $(I - \tilde{K}_T\tilde{K}_T^*)^{-1}$, we have to use the pseudo-inverse of $(I - \tilde{K}_T^*\tilde{K}_T)$. Mathematical complications arise because the range of $(I - \tilde{K}_T^*\tilde{K}_T)$ is not necessarily closed, so that its pseudo-inverse is defined only on a dense domain, and is possibly unbounded. However, the application of an unbounded pseudo-inverse can yield a bounded result if its operand is in the domain on which the pseudo-inverse is bounded, and the main thrust of this section is to make sure that we are in this situation and hence obtain results that are meaningful (*i.e.*, bounded).

Schur inversion formulas for positive semi-definite operators

Let be given some operator A on a Hilbert space $\mathcal{H}$. For better correspondence with results from other papers, as well as for historical reasons, we work in this section with operators written from the right to the left, and thus denote the 'left' range of A as $\mathcal{R}(A) = \{Ax : x \in \mathcal{H}\}$, and its null space as $\mathcal{N}(A) = \{x : Ax = 0\}$, which is a closed subspace. An orthogonal complement is denoted by $\perp$. The operator pseudo-inverse of A is defined as follows (following Beutler and Root [18]).

DEFINITION 7.8. *Let $\mathcal{H}$ be a Hilbert space, and A be a bounded linear operator defined on $\mathcal{H}$. The linear operator $A^\dagger : \mathcal{H} \to \mathcal{H}$ is a pseudo-inverse of A if and only if it is defined on $\mathcal{R}(A) \oplus \mathcal{R}(A)^\perp$ (which is dense in $\mathcal{H}$) and satisfies the following conditions:*

$$\begin{array}{lrcl} (1) & \mathcal{N}(A^\dagger) & = & \mathcal{R}(A)^\perp \\ (2) & \overline{\mathcal{R}(A^\dagger)} & = & \mathcal{N}(A)^\perp \quad (= \overline{\mathcal{R}(A^*)}) \\ (3) & AA^\dagger x & = & x \quad \text{for all } x \in \mathcal{R}(A). \end{array}$$

It is proven in [18] that $(A^\dagger)^\dagger = A$, $(A^\dagger)^* = (A^*)^\dagger$, $(A^*A)^\dagger = A^\dagger A^{*\dagger}$, and that $A^\dagger$ is bounded if and only if $\mathcal{R}(A)$ is closed.

We apply the following result by Douglas [19] on the majorization of operators on Hilbert spaces:

THEOREM 7.9. *Let A and B be bounded operators on a Hilbert space $\mathcal{H}$. The following are equivalent:*

$$\begin{array}{lrcl} (1) & AA^* & \leq & \lambda^2 BB^* \quad (\text{some } \lambda > 0), \\ (2) & \mathcal{R}(A) & \subset & \mathcal{R}(B), \\ (3) & A & = & BC \quad \text{for some bounded operator } C \text{ on } \mathcal{H}. \end{array}$$

If (1)-(3) *are valid, then a unique operator* C *exists such that*

$$\begin{array}{lll} (a) & \|C\| & = \inf\{\mu : AA^* \le \mu BB^*\}, \\ (b) & \mathcal{N}(A) & = \mathcal{N}(C), \\ (c) & \mathcal{R}(C) & \subset \overline{\mathcal{R}(B^*)}. \end{array}$$

The 'unique operator C' in this theorem is in fact $C = B^\dagger A$, since also $B^\dagger$ is uniquely defined and $B^\dagger A$ qualifies for C. Consequently, if $AA^* \le BB^*$, then this C satisfies $\|C\| \le 1$.

The lemma on Schur complements, and on the existence of a pseudo-inverse of a block partitioned operator which is based on Schur complements is more complicated now, because a range condition enters. (We show that in our case the range condition is automatically fulfilled.) The following lemmas replace lemma 7.2 and are suitable for our purpose.

LEMMA 7.10. *With* $\mathcal{H}_1$ *and* $\mathcal{H}_2$ *Hilbert spaces, let* $A : \mathcal{H}_1 \to \mathcal{H}_1$, $B : \mathcal{H}_1 \to \mathcal{H}_2$, $C : \mathcal{H}_2 \to \mathcal{H}_2$ *be bounded operators, and let* A *and* C *be self-adjoint. Consider the block operator* $X : \mathcal{H}_1 \times \mathcal{H}_2 \to \mathcal{H}_1 \times \mathcal{H}_2$,

$$X = \begin{bmatrix} A & B^* \\ B & C \end{bmatrix}.$$

Then $X \ge 0$ *if and only if*

$$\begin{array}{lll} (1) & C \ge 0, & \\ (2) & \mathcal{R}(B) \subset \mathcal{R}(C^{1/2}); & \text{or, equivalently, } B_1 = C^{\dagger/2}B \text{ is bounded,} \\ (3) & A - B_1^* B_1 \ge 0. & \end{array}$$

LEMMA 7.11. *Let* A, B, C, X *be as in lemma 7.10. Let* $X \ge 0$ *and write* $B_1 = C^{\dagger/2}B$. *Define the operator* $W^\ddagger$:

$$W^\ddagger = \begin{bmatrix} (A - B_1^* B_1)^{\dagger/2} & \\ & I \end{bmatrix} \begin{bmatrix} I & -B_1^* \\ & I \end{bmatrix} \begin{bmatrix} I & \\ & C^{\dagger/2} \end{bmatrix}$$

Then $W^\ddagger$ *is well defined and bounded on* $\mathcal{R}(X^{1/2})$. *If* v *is some bounded operator with range in* $\mathcal{R}(X^{1/2})$, *and if*

$$v_1 = X^{\dagger/2}v, \qquad v_2 = W^\ddagger v$$

then v_1 *and* v_2 *are bounded, and* $v_1^* v_1 = v_2^* v_2$.

The proof of both lemmas appears as an appendix. Note that $W^\ddagger \neq X^{\dagger/2}$, but rather $W^\ddagger = UX^{\dagger/2}$ on $\mathcal{R}(X^{1/2})$, where U is some Hilbert space isometry such that $U^*U = I$. The point is that $W^\ddagger$ is specified in terms of A, B, C, whereas it is hard to do so for $X^{\dagger/2}$.

Contractivity in terms of a state-space realization

We are now ready to derive a solution to the embedding problem along the lines of section 7.2 for the case where T is contractive, but not necessarily strictly contractive. Recall the definition of H_T and K_T of section 7.1.

LEMMA 7.12. *Let T be a system transfer operator in $\mathcal{U}$. If T is contractive, then*

$$I - K_T K_T^* \geq H_T H_T^* \geq 0 \tag{7.16}$$

and hence K_T and $\tilde{K}_T$ are contractive.

PROOF Let $u \in \mathcal{L}_2 Z^{-1}$, and put $y = uT = uK_T + uH_T$. The contractivity of T implies

$$\begin{aligned} & \mathbf{P}_0(uu^*) - \mathbf{P}_0(yy^*) \geq 0 \\ \Leftrightarrow\ & \mathbf{P}_0\left(u[I - TT^*]u^*\right) \geq 0 \\ \Leftrightarrow\ & \mathbf{P}_0\left(u[I - K_T K_T^* - H_T H_T^*]u^*\right) \geq 0 \\ \Leftrightarrow\ & \mathbf{P}_0\left(u[I - K_T K_T^*]u^*\right) \geq \mathbf{P}_0\left(uH_T H_T^* u^*\right) \geq 0 \end{aligned}$$

Hence $I - K_T K_T^* \geq 0$ on $\mathcal{L}_2 Z^{-1}$. $\tilde{K}_T$ represents K_T and is hence also contractive. □

COROLLARY 7.13. *If $\tilde{H}_T = \mathcal{C}\mathcal{O}$ is a decomposition of $\tilde{H}_T$ such that $\mathcal{O}\mathcal{O}^* \gg 0$ (**T** is uniformly observable), then $\mathcal{R}(\tilde{K}_T^* \mathcal{C}) \subset \mathcal{R}(I - \tilde{K}_T^* \tilde{K}_T)^{1/2}$ and hence $\mathcal{C}_1$ defined by*

$$\mathcal{C}_1 = (I - \tilde{K}_T^* \tilde{K}_T)^{\dagger/2} \tilde{K}_T^* \mathcal{C} \tag{7.17}$$

is bounded.

PROOF Apply theorem 7.9 to (7.16). From $I - K_T K_T^* \geq H_T H_T^*$ it follows that

$$H_T = (I - K_T K_T^*)^{1/2} N,$$

for some operator N with $\|N\| \leq 1$. Taking diagonal expansions, we have that $\tilde{H}_T = (I - \tilde{K}_T \tilde{K}_T^*)^{1/2} \tilde{N}$, and with $\tilde{H}_T = \mathcal{C}\mathcal{O}$ such that $\mathcal{O}\mathcal{O}^* \gg 0$, we obtain

$$\begin{aligned} \tilde{K}_T^* \mathcal{C} &= \tilde{K}_T^* \mathcal{C}\mathcal{O}\mathcal{O}^*(\mathcal{O}\mathcal{O}^*)^{-1} \\ &= \tilde{K}_T^* \tilde{H}_T \mathcal{O}^*(\mathcal{O}\mathcal{O}^*)^{-1} \\ &= \tilde{K}_T^* (I - \tilde{K}_T \tilde{K}_T^*)^{1/2} \tilde{N} \mathcal{O}^*(\mathcal{O}\mathcal{O}^*)^{-1} \\ &= (I - \tilde{K}_T^* \tilde{K}_T)^{1/2} \mathcal{C}_1 \end{aligned}$$

where $\mathcal{C}_1 = \tilde{K}_T^* \tilde{N} \cdot \mathcal{O}^*(\mathcal{O}\mathcal{O}^*)^{-1}$ is bounded. □

For $\mathcal{C}_1$ defined in (7.17), define the operator $M \in \mathcal{D}$ by

$$M = \mathcal{C}^*\mathcal{C} + \mathcal{C}_1^*\mathcal{C}_1 . \tag{7.18}$$

M is bounded, and $M \gg 0$ if $\mathcal{C}^*\mathcal{C} \gg 0$, *i.e.*, if the realization is uniformly controllable. This definition of M is compatible with the definition of M in (7.7) if T is strictly contractive, viz. $M = \mathcal{C}^*(I-\tilde{K}_T\tilde{K}_T^*)^{-1}\mathcal{C}$, because then $\mathcal{C}_1^*\mathcal{C}_1 = \mathcal{C}^*\tilde{K}_T(I-\tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^*\mathcal{C}$, and $I+\tilde{K}_T(I-\tilde{K}_T^*\tilde{K}_T)^{-1}\tilde{K}_T^* = (I-\tilde{K}_T\tilde{K}_T^*)^{-1}$. The latter relation is however not necessarily valid if a pseudo-inverse is used.

The following theorem subsumes theorem 7.4.

THEOREM 7.14. *Let $T \in \mathcal{U}$ be a system transfer operator with a strictly stable state-space realization $\{A, B, C, D\}$. If T is contractive and the realization is uniformly observable, then M defined by*

$$\begin{aligned} L &= (I-\tilde{K}_T^*\tilde{K}_T)^{1/2} \\ \mathcal{C}_1 &= L^\dagger\tilde{K}_T^*\mathcal{C} \\ M &= \mathcal{C}^*\mathcal{C} + \mathcal{C}_1^*\mathcal{C}_1 \end{aligned} \tag{7.19}$$

is bounded and satisfies the relation

$$M^{(-1)} = A^*MA + B^*B + \left([A^*MC + B^*D]\Phi^\dagger\right)\cdot\left(\Phi^\dagger[D^*B + C^*MA]\right) \tag{7.20}$$

with $\Phi = (I - D^*D - C^*MC)^{1/2}$. If, in addition, the state-space realization is uniformly controllable (hence uniformly minimal), then $M \gg 0$.

PROOF The proof uses equations (7.4), (7.5):

$$\tilde{K}_T^{(-1)} = \begin{bmatrix} T_{[0]} & \\ \tilde{V}_T & \tilde{K}_T \end{bmatrix}, \quad \mathcal{C}^{(-1)} = \begin{bmatrix} B \\ \mathcal{C}A \end{bmatrix}, \quad \tilde{V}_T = \mathcal{C}C, \quad T_{[0]} = D.$$

To find an expression for $M^{(-1)}$, put

$$X = (I-\tilde{K}_T^*\tilde{K}_T)^{(-1)} = \begin{bmatrix} I - T_{[0]}^*T_{[0]} - \tilde{V}_T^*\tilde{V}_T & -\tilde{V}_T^*\tilde{K}_T \\ -\tilde{K}_T^*\tilde{V}_T & I-\tilde{K}_T^*\tilde{K}_T \end{bmatrix}$$

According to lemma 7.12, $X \geq 0$. Lemma 7.10 then implies that $\mathcal{R}(\tilde{K}_T^*\tilde{V}_T) \subset \mathcal{R}(I-\tilde{K}_T^*\tilde{K}_T)^{1/2}$ so that $(I-\tilde{K}_T^*\tilde{K}_T)^{\dagger/2}\,\tilde{K}_T^*\tilde{V}_T = \mathcal{C}_1C$ is bounded. (This result would also follow from corollary 7.13 because $\mathcal{R}(\tilde{K}_T^*\tilde{V}_T) = \mathcal{R}(\tilde{K}_T^*\mathcal{C}C) \subset \mathcal{R}(\tilde{K}_T^*\mathcal{C})$.) Let

$$\begin{aligned} \Phi &= \left[I - T_{[0]}^*T_{[0]} - \tilde{V}_T^*\tilde{V}_T - C^*\mathcal{C}_1^*\mathcal{C}_1C\right]^{1/2} \\ &= \left[I - D^*D - C^*(\mathcal{C}^*\mathcal{C} + \mathcal{C}_1^*\mathcal{C}_1)C\right]^{1/2} \\ &= (I - D^*D - C^*MC)^{1/2} \end{aligned}$$

Hence the third item of lemma 7.10 implies that $I - D^*D - C^*MC \geq 0$. Put

$$\begin{aligned} W^\ddagger &= \begin{bmatrix} \Phi^\dagger & \\ & I \end{bmatrix}\begin{bmatrix} I & C^*\mathcal{C}_1^* \\ & I \end{bmatrix}\begin{bmatrix} I & \\ & (I-\tilde{K}_T^*\tilde{K}_T)^{\dagger/2} \end{bmatrix} \\ v &= \left[\tilde{K}_T^*\mathcal{C}\right]^{(-1)} = \tilde{K}_T^{*(-1)}\begin{bmatrix} B \\ \mathcal{C}A \end{bmatrix} = \begin{bmatrix} D^*B + C^*\mathcal{C}^*\mathcal{C}A \\ \tilde{K}_T^*\mathcal{C}A \end{bmatrix}. \end{aligned}$$

Then lemma 7.11 yields that the operator $v_1 = X^{\dagger/2}v = \mathcal{C}_1^{(-1)}$ is bounded, and $v_2 = W^{\dagger}v$ is such that $v_1^* v_1 = v_2^* v_2$. Evaluation of v_2 gives

$$\begin{aligned} v_2 = W^{\dagger}v &= \begin{bmatrix} \Phi^{\dagger} & \\ & I \end{bmatrix} \begin{bmatrix} I & C^*\mathcal{C}_1^* \\ & I \end{bmatrix} \begin{bmatrix} I & \\ & (I - \tilde{K}_T^* \tilde{K}_T)^{\dagger/2} \end{bmatrix} \begin{bmatrix} D^*B + C^*\mathcal{C}^*\mathcal{C}A \\ \tilde{K}_T^*\mathcal{C}A \end{bmatrix} \\ &= \begin{bmatrix} \Phi^{\dagger} & \\ & I \end{bmatrix} \begin{bmatrix} I & C^*\mathcal{C}_1^* \\ & I \end{bmatrix} \begin{bmatrix} D^*B + C^*\mathcal{C}^*\mathcal{C}A \\ \mathcal{C}_1 A \end{bmatrix} \\ &= \begin{bmatrix} \Phi^{\dagger}(D^*B + C^*MA) \\ \mathcal{C}_1 A \end{bmatrix}. \end{aligned}$$

Hence

$$\begin{aligned} \left[\mathcal{C}_1^*\mathcal{C}_1\right]^{(-1)} &= v_1^* v_1 = v_2^* v_2 = \\ &= A^*\mathcal{C}_1^*\mathcal{C}_1 A + \left([B^*D + A^*MC]\Phi^{\dagger}\right) \cdot \left(\Phi^{\dagger}[D^*B + C^*MA]\right) \end{aligned}$$

and with $\mathcal{C}^{(-1)} = \left[\begin{smallmatrix} B \\ \mathcal{C}A \end{smallmatrix}\right]$ we finally obtain

$$\begin{aligned} M^{(-1)} &= \left[\mathcal{C}^*\mathcal{C}\right]^{(-1)} + \left[\mathcal{C}_1^*\mathcal{C}_1\right]^{(-1)} \\ &= B^*B + A^*\mathcal{C}^*\mathcal{C}A + A^*\mathcal{C}_1^*\mathcal{C}_1 A + \left([B^*D + A^*MC]\Phi^{\dagger}\right) \cdot \left(\Phi^{\dagger}[D^*B + C^*MA]\right) \\ &= A^*MA + B^*B + \left([B^*D + A^*MC]\Phi^{\dagger}\right) \cdot \left(\Phi^{\dagger}[D^*B + C^*MA]\right). \end{aligned}$$

□

The result of this section is thus a relatively simple extension of theorem 7.4: in the case that T is not strictly contractive, we can use the recursion

$$\begin{aligned} \Phi &= (I - D^*D - C^*MC)^{1/2} \\ M^{(-1)} &= A^*MA + B^*B + \left[A^*MC + B^*D\right]\Phi^{\dagger} \cdot \Phi^{\dagger}\left[D^*B + C^*MA\right] \end{aligned}$$

although we required the given realization to be uniformly minimal this time to have $M \gg 0$. This condition is sufficient, but too strong (for time-invariant systems, the usual condition is that the realization should be 'stabilizable'), which is a bit worrying because we know from chapter 3 that not every time-varying system admits a uniformly minimal realization, not even if it has a finite state dimension. (The condition is that the operator must have closed state spaces $\mathcal{H}$ and $\mathcal{H}_0$, *i.e.*, the range of H_T must be closed. See proposition 3.13.) The recursion for M is very close to (and encompasses) the expression that we have obtained before in the strictly contractive case. Note that we know only that $\Phi^{\dagger}(D^*B + C^*MA)$ is bounded, but not necessarily $\Phi^{\dagger}\Phi^{\dagger}(D^*B + C^*MA)$: we have to evaluate $\Phi^{\dagger}(D^*B + C^*MA)$, and then square this expression in order to get a correct answer.

With the above theorem, an extension of the step 1. of the embedding theorem 7.5, to include contractive systems that need not be strictly contractive, is straightforward. The results are the same.

THEOREM 7.15. (EMBEDDING THEOREM, II) *Let $T \in \mathcal{U}(\mathcal{M}_1, \mathcal{N}_1)$ be a locally finite transfer operator with state realization $\mathbf{T} = \{A, B, C, D\}$ such that $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ is strictly stable: $\ell_A < 1$. If T is strictly contractive and $\mathbf{T}$ is uniformly controllable or if T is contractive and $\mathbf{T}$ is both uniformly controllable and uniformly observable, then the orthogonal embedding problem has a solution $\Sigma \in \mathcal{U}(\mathcal{M}_1 \times \mathcal{M}_2, \mathcal{N}_1 \times \mathcal{N}_2)$ such that Σ is inner and $\Sigma_{11} = T$. The dimension sequences of Σ are*

$$\begin{aligned} \#(\mathcal{M}_2) &= \#(\mathcal{N}_1), \\ \#(\mathcal{N}_2) &= \#(\mathcal{B}) - \#(\mathcal{B}^{(-1)}) + \#(\mathcal{M}_1). \end{aligned}$$

Σ has a unitary realization $\mathbf{\Sigma} = \{A_\Sigma, B_\Sigma, C_\Sigma, D_\Sigma\}$ where A_Σ is state equivalent to A by a boundedly invertible state transformation R.

PROOF The proof is the same as the proof of theorem 7.5, except that step 1. is now reformulated as follows:

A solution to step 1. of the embedding procedure (theorem 7.5, equation (7.9)) is given by M in (7.19). M satisfies the recursion

$$M^{(-1)} = A^*MA + B^*B + \left([A^*MC + B^*D]\Phi^\dagger\right) \cdot \left(\Phi^\dagger[D^*B + C^*MA]\right) .$$

with $\Phi = (I - D^*D - C^*MC)^{1/2}$. A bounded solution M exists under the conditions [T is contractive] and [$\mathbf{T}$ is uniformly observable] and is strictly positive definite if [$\mathbf{T}$ is uniformly controllable]. $B_2 \in \mathcal{D}(\mathcal{N}_1, \mathcal{B}^{(-1)})$ and $D_{21} \in \mathcal{D}(\mathcal{N}_1, \mathcal{N}_1)$ are determined as

$$\begin{bmatrix} D_{21} &= \Phi \\ B_2 &= -\Phi^\dagger \left[D^*B + C^*MA\right] \end{bmatrix}$$

The proof of step 1. is as before, but now uses theorem 7.14. □

Bounded real lemma

The embedding problem is intimately connected to the bounded real lemma [6, 11] which is of some interest in system and control theory. For discrete time systems, it can be formulated as follows.

THEOREM 7.16. (BOUNDED REAL LEMMA) *Let $T \in \mathcal{U}(\mathcal{M}_1, \mathcal{N}_1)$ be a bounded causal locally finite transfer operator, with state-space realization $\mathbf{T} = \{A, B, C, D\}$. Suppose $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ is strictly stable: $\ell_A < 1$, and that $\mathbf{T}$ is uniformly controllable.*

(i) If $\mathbf{T}$ is uniformly observable, then T is contractive if and only if the set of equations

$$\left\{ \begin{array}{lllll} A^*MA & + & B^*B & + & B_2^*B_2 &= M^{(-1)} \\ C^*MC & + & D^*D & + & D_{21}^*D_{21} &= I \\ A^*MC & + & B^*D & + & B_2^*D_{21} &= 0 \end{array} \right. \tag{7.21}$$

has a solution $M \in \mathcal{D}(\mathcal{B}, \mathcal{B})$, $B_2 \in \mathcal{D}(\mathcal{N}_1, \mathcal{B}^{(-1)})$, $D_{21} \in \mathcal{D}(\mathcal{N}_1, \mathcal{N}_1)$ *such that* $M \gg 0$.

(ii) If T is strictly contractive, then there exists a uniformly positive M and operators B_2, D_{21} *such that the equations (7.21) are satisfied.*

PROOF *(ii)* and the 'only if' part of *(i)* are corollaries of theorems 7.4 and 7.14, if we take M as given by equation (7.19). M satisfies the recursion (7.20) with $\Phi = (I - D^*D - C^*MC)^{1/2}$. Taking $B_2 \in \mathcal{D}(\mathcal{N}_1, \mathcal{B}^{(-1)})$ and $D_{21} \in \mathcal{D}(\mathcal{N}_1, \mathcal{N}_1)$ to be

$$\begin{bmatrix} D_{21} &= \Phi \\ B_2 &= -\Phi^\dagger \left[D^*B + C^*MA\right] \end{bmatrix}$$

then it is immediately verified that equations (7.21) are satisfied.

It remains to prove the 'if' part of *(i)*, which is in fact a corollary to the embedding theorem. With M, B_2 and D_{21} satisfying (7.21), we can construct a lossless realization $\mathbf{\Sigma}$ of the form (7.8), and it has been proven that a system Σ corresponding to a lossless realization $\mathbf{\Sigma}$ is inner. Because T is equal to Σ_{11} and hence is a partial transfer of Σ, it must be contractive. □

Appendix: derivation of lemmas 7.10 and 7.11

The contents of lemmas 7.10 and 7.11 are well known for finite matrices (see *e.g.*, [20, 21]) for generalized inverse formulas involving Schur complements). The matrix case is readily extended to operators if the operators are assumed to have closed range. Without this condition, complications arise because the pseudo-inverses that are involved are unbounded operators.

We will repeatedly use theorem 7.9 in the following form. Let $X \geq 0$ be a bounded operator on a Hilbert space $\mathcal{H}$. If v is a bounded operator whose range is in $\mathcal{R}(X)$, then $v = Xv_1$, for some bounded $v_1 \in \overline{\mathcal{R}(X^*)}$ for which we can take $v_1 = X^\dagger v$.

A second fact that is used in the proof of lemma 7.11 is that $X^\dagger X = \mathbf{P}_{\overline{\mathcal{R}(X^*)}}$: the orthogonal projector onto $\overline{\mathcal{R}(X^*)}$, with domain $\mathcal{H}$ [18].

Proof of lemma 7.10

Suppose first that $X \geq 0$; we show that (1), (2), (3) hold. It is immediate that $A \geq 0$, $C \geq 0$.

$\mathcal{R}(B) \subset \mathcal{R}(C^{1/2})$ is proven by showing that there exists λ such that $BB^* \leq \lambda C$; Douglas' theorem then implies the result. The proof is by contradiction. Suppose that there is not such a λ. Then there exists a sequence $\{x_n : n \in \mathbb{N}\}$ such that

$$(BB^*x_n, x_n) \geq n(Cx_n, x_n) > 0. \tag{7.22}$$

where $(\cdot,\cdot)$ denotes the inner product in $\mathcal{H}$. In particular, $\|B^*x_n\| > 0$ (all n). For any u_n, $X \geq 0$ implies

$$\left(\begin{bmatrix} A & B^* \\ B & C \end{bmatrix}\begin{bmatrix} u_n \\ x_n \end{bmatrix}, \begin{bmatrix} u_n \\ x_n \end{bmatrix}\right) \geq 0$$

i.e., $(Au_n, u_n)+(B^*x_n, u_n)+(Bu_n, x_n)+(Cx_n, x_n) \geq 0$. Choose $u_n = -\frac{1}{\sqrt{n}}B^*x_n$. Using (7.22), we obtain

$$\left(B\left\{\frac{A}{n} - \frac{2}{\sqrt{n}} + \frac{I}{n}\right\}B^*x_n, x_n\right) \geq 0.$$

But if $n > \|I+A\|^2$, the term in braces is smaller than $-1/\sqrt{n}$, which gives a contradiction. Hence $\mathcal{R}(B) \subset \mathcal{R}(C^{1/2})$.

Define $L = C^{1/2}$ (although $L = L^*$, we will not use this), and let $B_1 = L^\dagger B$. Then B_1 is bounded, and $B = LB_1$ with $\mathcal{R}(B_1) \subset \overline{\mathcal{R}(L^*)}$, which implies

$$\mathcal{N}(B_1^*) \supset \mathcal{N}(L). \tag{7.23}$$

To prove $A - B_1^*B_1 \geq 0$, we will show that

$$X = \begin{bmatrix} A & B_1^*L^* \\ LB_1 & LL^* \end{bmatrix} \geq 0 \quad \Rightarrow \quad \begin{bmatrix} A & B_1^* \\ B_1 & I \end{bmatrix} \geq 0 \tag{7.24}$$

from which $A - B_1^*B_1 \geq 0$ follows directly by applying vectors of the form $\begin{bmatrix} I \\ -B_1 \end{bmatrix} a$.

Thus for $x \in \mathcal{H}_1 \oplus \mathcal{H}_2$, take x of the form

$$x = \begin{bmatrix} u \\ x_1 + x_2 \end{bmatrix} \in \begin{bmatrix} \mathcal{H}_1 \\ \mathcal{N}(L) \oplus \mathcal{R}(L^*) \end{bmatrix}$$

where $x_1 \in \mathcal{N}(L)$ and $x_2 \in \mathcal{R}(L^*)$. Note that $\mathcal{N}(L) \oplus \mathcal{R}(L^*)$ is dense in $\mathcal{H}_2$. Then $\mathcal{N}(B_1^*) \supset \mathcal{N}(L)$ implies $B_1^*x_1 = 0$, while $x_2 \in \mathcal{R}(L^*)$ implies that $x_2 = L^*x_2'$, for some bounded x_2'. Using these observations, it follows that

$$\begin{aligned}
&\left(\begin{bmatrix} A & B_1^* \\ B_1 & I \end{bmatrix}\begin{bmatrix} u \\ x_1 + x_2 \end{bmatrix}, \begin{bmatrix} u \\ x_1 + x_2 \end{bmatrix}\right) \\
&\quad = (Au, u) + (B_1^*x_1, u) + (B_1u, x_1) + (x_1, x_1) + (B_2^*x_2, u) + (B_1u, x_2) + (x_2, x_2) \\
&\quad \geq (Au, u) + (B_1^*x_2, u) + (B_1u, x_2) + (x_2, x_2) \\
&\quad = (Au, u) + (B^*x_2', u) + (B_1u, x_2') + (x_2', x_2') \\
&\quad = \left(X\begin{bmatrix} u \\ x_2' \end{bmatrix}, \begin{bmatrix} u \\ x_2' \end{bmatrix}\right) \geq 0.
\end{aligned}$$

Hence relation (7.24) holds on a dense subset of $\mathcal{H}_1 \oplus \mathcal{H}_2$. By continuity, it holds everywhere, and consequently $A - B_1^*B_1 \geq 0$.

It remains to prove the reverse implication: $X \geq 0$ if the three conditions are satisfied. Because $C \geq 0$ a decomposition of C as $C = LL^*$ is defined. Using this decomposition and $B = LB_1$,

$$X = \begin{bmatrix} A & B_1^*L^* \\ LB_1 & LL^* \end{bmatrix} = \begin{bmatrix} I & B_1^* \\ & L \end{bmatrix} \begin{bmatrix} A - B_1^*B_1 & \\ & I \end{bmatrix} \begin{bmatrix} I & \\ B_1 & L^* \end{bmatrix}.$$

Under the stated conditions, the operator

$$W = \begin{bmatrix} I & \\ & L \end{bmatrix} \begin{bmatrix} I & B_1^* \\ & I \end{bmatrix} \begin{bmatrix} (A - B_1^*B_1)^{1/2} & \\ & I \end{bmatrix} \tag{7.25}$$

is well defined, and is a factor of X such that $X = WW^*$. Hence $X \geq 0$. □

Proof of lemma 7.11

Let $X \geq 0$ have a factorization $X = WW^*$, then $\mathcal{R}(X^{1/2}) = \mathcal{R}(W)$ (again by theorem 7.9). It can be inferred from Beutler and Root [18] that

$$X^\dagger = W^{*\dagger}W^\dagger = X^{\dagger/2}X^{\dagger/2},$$

hence if $\mathcal{R}(v) \subset \mathcal{R}(X^{1/2}) = \mathcal{R}(W)$, then v_1 and v_2 defined by

$$\begin{aligned} v_1 &= X^{\dagger/2}v, & \mathcal{R}(v_1) &\subset \overline{\mathcal{R}(X^{1/2})} \\ v_2 &= W^\dagger v, & \mathcal{R}(v_2) &\subset \overline{\mathcal{R}(W^*)} \end{aligned}$$

are bounded, and[1] $v_1^*v_1 = v_2^*v_2$.

Let $L = C^{1/2}$, $B_1 = L^\dagger B$ and put W as in (7.25), so that $X = WW^*$. Define the operator $W^\ddagger$ by

$$W^\ddagger = \begin{bmatrix} (A - B_1^*B_1)^{\dagger/2} & \\ & I \end{bmatrix} \begin{bmatrix} I & -B_1^* \\ & I \end{bmatrix} \begin{bmatrix} I & \\ & L^\dagger \end{bmatrix}.$$

We prove that $W^\ddagger = W^\dagger$ on $\mathcal{R}(W)$. The result will be, for a bounded operator v with $\mathcal{R}(v) \subset \mathcal{R}(X^{1/2}) = \mathcal{R}(W)$, that $W^\dagger v = W^\ddagger v$, so that $v_1 := X^{\dagger/2}v$ and $v_2 := W^\ddagger v$ are bounded and satisfy $v_1^*v_1 = v_2^*v_2$.

For any v with range in $\mathcal{R}(W)$ we have that the operator $v_1 = W^\dagger v$ is bounded and such that $v = Wv_1$. Hence $W^\ddagger v = W^\ddagger Wv_1 = W^\dagger Wv_1 = W^\dagger v$, so that $W^\ddagger = W^\dagger$ on $\mathcal{R}(W)$ if and only if

$$W^\ddagger W = W^\dagger W \quad \text{on } \overline{\mathcal{R}(W^*)}.$$

[1] We are careful here not to write $X^\dagger v$. Although $\overline{\mathcal{R}(X)} = \overline{\mathcal{R}(X^{1/2})}$, we only have that $\mathcal{R}(X) \subset \mathcal{R}(X^{1/2})$, and hence $X^\dagger v$ can be unbounded with $\mathcal{R}(v) \in \mathcal{R}(X^{1/2})$.

To analyze $W^\ddagger W$, we first prove that $B_1^* - B_1^* L^\dagger L = 0$. Indeed, if $x \in \mathcal{N}(L)$ then $x \in \mathcal{N}(B_1^*)$ (by equation (7.23)), and hence both $B_1^* x = 0$ and $Lx = 0$. If, on the other hand, $x \in \mathcal{N}(L)^\perp$, then $L^\dagger L x = x$ since $L^\dagger L$ is the projector onto $\mathcal{N}(L)^\perp$, and hence $B_1^* L^\dagger L x = B_1^* x$.

With the definition of $W^\ddagger$ and the above result,

$$\begin{aligned} W^\ddagger W &= \begin{bmatrix} (A - B_1^* B_1)^{\dagger/2} & \\ & I \end{bmatrix} \begin{bmatrix} I & -B_1^* \\ & I \end{bmatrix} \begin{bmatrix} I & \\ & L^\dagger \end{bmatrix} \cdot \\ &\quad \cdot \begin{bmatrix} I & \\ & L \end{bmatrix} \begin{bmatrix} I & B_1^* \\ & I \end{bmatrix} \begin{bmatrix} (A - B_1^* B_1)^{1/2} & \\ & I \end{bmatrix} \\ &= \begin{bmatrix} (A - B_1^* B_1)^{\dagger/2} & \\ & I \end{bmatrix} \begin{bmatrix} I & B_1^* - B_1^* L^\dagger L \\ & L^\dagger L \end{bmatrix} \begin{bmatrix} (A - B_1^* B_1)^{1/2} & \\ & I \end{bmatrix} \\ &= \begin{bmatrix} (A - B_1^* B_1)^{\dagger/2} (A - B_1^* B_1)^{1/2} & \\ & L^\dagger L \end{bmatrix} =: \begin{bmatrix} \mathbf{P}_1 & \\ & \mathbf{P}_2 \end{bmatrix} \end{aligned}$$

$\mathbf{P}_1$ and $\mathbf{P}_2$ are projectors onto $\overline{\mathcal{R}(A - B_1^* B_1)^{1/2}}$ and $\overline{\mathcal{R}(L^*)}$, respectively. Now, using

$$W^* = \begin{bmatrix} (A - B_1^* B_1)^{1/2} & \\ & I \end{bmatrix} \cdot \begin{bmatrix} I & \\ B_1 & L^* \end{bmatrix}$$

and $\mathcal{R}(B_1) \subset \overline{\mathcal{R}(L^*)}$, we have that

$$\overline{\mathcal{R}(W^*)} \subset \overline{\mathcal{R}} \begin{bmatrix} (A - B_1^* B_1)^{1/2} & \\ & L^* \end{bmatrix}$$

Since $W^\dagger W$ is the projector onto $\overline{\mathcal{R}(W^*)}$, and $W^\ddagger W$ is the projector onto the range at the right-hand side of the expression, this proves that $W^\ddagger W = W^\dagger W$ on $\overline{\mathcal{R}(W^*)}$, as required. Hence $W^\ddagger = W^\dagger$ on $\mathcal{R}(W)$, which also implies that $W^\ddagger$ is well defined on $\mathcal{R}(W)$. □

Bibliography

[1] P. Dewilde, "Input-output description of roomy systems," *SIAM J. Control and Optimization*, vol. 14, pp. 712–736, July 1976.

[2] A.J. van der Veen and P.M. Dewilde, "Orthogonal embedding theory for contractive time-varying systems," in *Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing II (Proc. Int. Symp. MTNS-91)* (H. Kimura and S. Kodama, eds.), pp. 513–518, MITA Press, Japan, 1992.

[3] A.J. van der Veen and P.M. Dewilde, "Orthogonal embedding theory for contractive time-varying systems," in *Proc. IEEE ISCAS*, pp. 693–696, 1992.

[4] A.J. van der Veen and P.M. Dewilde, "Embedding of time-varying contractive systems in lossless realizations," *subm. Math. Control Signals Systems*, July 1992.

[5] S. Darlington, "Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics," *J. Math. Phys.*, vol. 18, pp. 257–355, 1939.

[6] B.D.O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*. Prentice Hall, 1973.

[7] P. Dewilde and E. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy," in *Operator Theory: Advances and Applications*, vol. 29, pp. 97–116, Birkhäuser Verlag, 1988.

[8] A.V. Belevitch, *Classical Network Theory*. San Francisco: Holden Day, 1968.

[9] B.D.O. Anderson, K.L. Hitz, and N.D. Diem, "Recursive algorithm for spectral factorization," *IEEE Trans. Circuits Syst.*, vol. 21, no. 6, pp. 742–750, 1974.

[10] G.C. Goodwin and K.S. Sin, *Adaptive Filtering, Prediction and Control*. Englewood Cliffs, NJ: Prentice Hall, 1984.

[11] P.P. Vaidyanathan, "The discrete-time Bounded-Real Lemma in digital filtering," *IEEE Trans. Circuits Syst.*, vol. 32, no. 9, pp. 918–924, 1985.

[12] U.B. Desai, "A state-space approach to orthogonal digital filters," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 160–169, Feb. 1991.

[13] I. Yaesh and U. Shaked, "A transfer function approach to the problems of discrete-time systems: H_∞-optimal linear control and filtering," *IEEE Trans. Automat. Control*, vol. 36, pp. 1264–1271, Nov. 1991.

[14] S. Bittanti, A.J. Laub, and J.C. Willems, eds., *The Riccati Equation*. Comm. Control Eng. Series, Springer Verlag, 1991.

[15] M. Morf and T. Kailath, "Square-root algorithms for Least-Squares Estimation," *IEEE Trans. Automat. Control*, vol. 20, no. 4, pp. 487–497, 1975.

[16] P. Dewilde and H. Dym, "Interpolation for upper triangular operators," in *Time-Variant Systems and Interpolation* (I. Gohberg, ed.), vol. 56 of *Operator Theory: Advances and Applications*, pp. 153–260, Birkhäuser Verlag, 1992.

[17] H. Lev-Ari and T. Kailath, "State-space approach to factorization of lossless transfer functions and structured matrices," *Lin. Alg. Appl.*, vol. 162, pp. 273–295, Feb. 1992.

[18] F.J. Beutler and W.L Root, "The operator pseudo-inverse in control and systems identification," in *Generalized Inverses and Applications* (M. Zuhair Nashed, ed.), pp. 397–494, Academic Press, 1976.

[19] R.G. Douglas, "On majorization, factorization and range inclusion of operators on Hilbert space," *Proc. Amer. Math. Soc.*, vol. 17, pp. 413–415, 1966.

[20] D. Carlson, E. Haynsworth, and T. Markham, "A generalization of the Schur complement by means of the Moore-Penrose inverse," *SIAM J. Applied Math.*, vol. 26, pp. 169–175, 1974.

[21] F. Burns, D. Carlson, E. Haynsworth, and T. Markham, "Generalized inverse formulas using Schur complements," *SIAM J. Applied Math.*, vol. 26, pp. 254–259, 1974.

Chapter 8

SPECTRAL FACTORIZATION

It is known that positive operators Ω on a Hilbert space admit a factorization of the form $\Omega = W^*W$, where W is an outer operator. This Hilbert space result also proves the existence of a spectral factorization for time-varying systems. In this chapter, we translate the above mathematical theorem into system theory language by deriving how such a factorization can be actually computed if a state realization of the upper part of Ω is known. The crucial step in this algorithm is again the solution of a Riccati recursion with time-varying coefficients. We show that, under certain conditions, a positive solution exists, which produces a factor which is outer. The same can also be formulated in terms of a time-varying positive real lemma. Finally, we provide some connections with related problems in which Riccati equations occurred such as inner-outer factorization, orthogonal embedding and the time-varying bounded-real lemma.[1]

8.1 INTRODUCTION

In the discussion on the inner-outer factorization problem and the embedding problem, we have obtained solutions governed by Riccati equations. In many other problems in time-invariant system and (H_∞) control theory, for example linear quadratic optimal control, optimal filtering and sensitivity minimization, Riccati equations play an important role. There is a family of related forms of this equation, and the precise form depends on the application. Underlying these problems is typically a spectral factorization problem, and the discrete-time Riccati equation corresponding to this problem was originally studied in [2, 3]. The equation usually has more than one solution, and important issues are the existence and computation of Hermitian solutions which are maximally positive or negative, as these conditions imply minimal-phase properties of spectral factors, or the stability of (closed-loop) transfer operators constructed from the solution. Such solutions

[1] The contents of this chapter has been submitted as a paper [1].

are, for time-invariant systems, obtained by an analysis of the eigenvalues and invariant subspaces of an associated (Hamiltonian) matrix. A recent overview of solution methods, as well as many references to older literature, can be found in the collection [4].

For general time-varying systems, the Riccati equation becomes a recursion with time-varying coefficients that can also have time-varying dimensions. For such equations, much less is known on the structure of solutions. One reason for this is that the usual eigenvalue analysis to classify stable and unstable systems is no longer applicable, *e.g.*, because A_k need not be square. Some results, *e.g.*, on the convergence of solutions starting from an approximate initial point, had already been obtained in the solution of the embedding problem (chapter 7). Rather than directly analyzing the recursion, a relatively simple derivation in that chapter used the 'physical' properties of the problem underlying the recursion (i.e. the embedding problem). However, the time-invariant equation has a number of other interesting and important features that could not yet be generalized along the lines of chapter 7. For example, it is known that the result of the embedding (or actually, of the spectral factorization of $I - T^*T$ underlying the embedding) yields extensions which are actually *outer*, a property which we use later in chapter 9.

In this chapter, we approach the time-varying Riccati equation from a different angle, by starting from the spectral factorization problem. The same approach is followed in [5] although, in that paper, the starting point is the existence of the Cholesky factor of a positive definite, finite size matrix. The Riccati recursion in these factorization problems emerges once a state realization for the operator is assumed. Solutions of the spectral factorization and inner-outer factorization problems are known also in the more general case of Hilbert space nest algebras (see the work of Arveson [6]), and this context applies to time-varying systems, too. For example, a bounded positive operator Ω has a factorization into

$$\Omega = W^*W$$

where W can be chosen to be outer. We show how, from this property of W, properties on the related time-varying Riccati equation can be derived. In particular, the fact that an outer factor W exists implies the existence of a 'maximal positive solution' of the Riccati equation, provided certain additional conditions with respect to stability and controllability are satisfied.

In this chapter, we consider only the 'easy' case where the inverted term in the Riccati equation exists and is bounded, and in particular where this term is definite. Generalizations are still possible but are analytically more difficult to treat as they lead to generalized inverses with range conditions. The spectral factorization problem is treated in section 8.2, where also a (related) time-varying version of the positive real lemma is formulated. Some computational issues are discussed in section 8.3. It is argued in section 8.4 that under certain conditions the Riccati recursion converges to the exact solution even if the recursion is started from an approximate initial point. This allows us to compute spectral factors of more general time-varying positive operators, even if they are not constant or

periodically varying before some point in time. Finally, in section 8.5, we discuss some connections of the spectral factorization theory with related problems in which a Riccati equation occurs, in particular the orthogonal embedding problem of contractive operators and the inner-outer factorization problem.

8.2 SPECTRAL FACTORIZATION

We recall the definitions of outer operators from chapter 4. An operator $W_0 \in \mathcal{U}(\mathcal{M}, \mathcal{N})$ is defined to be (left) outer if

$$\overline{\mathcal{U}_2^{\mathcal{M}} W_0} = \mathcal{U}_2^{\mathcal{N}}.$$

W_0 is (right) outer if

$$\overline{\mathcal{L}_2 Z^{-1} W_0^*} = \mathcal{L}_2 Z^{-1}.$$

Arveson [6] has shown, in the general context of nest algebras which also applies to our model of time-varying systems, that if $\Omega \in \mathcal{X}$ is a positive operator, an operator $W \in \mathcal{U}$ exists such that

$$\Omega = W^* W.$$

W can be chosen to be outer, in which case the factorization is called a spectral factorization. Related to this fact is another theorem by Arveson in the same paper, which claims that operators in a Hilbert space have an inner-outer factorization

$$W = UW_0$$

where U is a co-isometry ($U^*U = I$) and W_0 is (right) outer.[2] Moreover, if Ω is uniformly positive definite, then Ω has a factorization $\Omega = W_0^* W_0$ where W_0 is (left and right) outer and invertible, and hence $\mathcal{L}_2 Z^{-1} W_0^* = \mathcal{L}_2 Z^{-1}$ (no closure is needed) and $W_0^{-1} \in \mathcal{U}$. Any other invertible factor W can be written as $W = UW_0$, where U is now invertible and hence inner.

In this section, we derive an algorithm to compute a time-varying spectral factorization of operators with a state-space realization. The computation amounts to the (recursive) solution of a Riccati equation. Such equations have in general a collection of solutions. We show that in order to obtain an outer spectral factor, one must select a uniformly positive solution of the Riccati equation, and that this solution is unique. We need a number of preliminary results.

[2] Actually, Arveson uses a slightly different definition of outerness (not requiring $\ker (\cdot W_0)\big|_{\mathcal{L}_2 Z^{-1}} = 0$), so that U can be chosen inner. The resulting inner-outer factorizations are the same when W is invertible. See chapter 4.)

Realization for T^*T

We first derive a formula to compute a realization of the upper part of the operator T^*T, when a realization of $T \in \mathcal{U}$ is given.

LEMMA 8.1. *Let $T \in \mathcal{U}$ be given by the state realization $T = D + BZ(I-AZ)^{-1}C$, where $\ell_A < 1$. Then a state realization of the upper part of T^*T is*

$$\begin{bmatrix} A & C \\ D^*B + C^*\Lambda A & D^*D + C^*\Lambda C \end{bmatrix}$$

where $\Lambda \in \mathcal{D}$ is the (unique) operator satisfying the Lyapunov equation $\Lambda^{(-1)} = A^*\Lambda A + B^*B$.

PROOF Evaluation of T^*T gives

$$\begin{aligned} T^*T &= \left[D^* + C^*(I-Z^*A^*)^{-1}Z^*B^*\right]\left[D + BZ(I-AZ)^{-1}C\right] \\ &= D^*D + C^*(I-Z^*A^*)^{-1}Z^*B^*D + D^*BZ(I-AZ)^{-1}C + \\ &\quad + C^*(I-Z^*A^*)^{-1}Z^*B^*BZ(I-AZ)^{-1}C. \end{aligned}$$

The expression $(I-Z^*A^*)^{-1}Z^*B^*BZ(I-AZ)^{-1}$ evaluates as

$$(I-Z^*A^*)^{-1}Z^*B^*BZ(I-AZ)^{-1} = (I-Z^*A^*)^{-1}Z^*X + \Lambda(I-AZ)^{-1}$$

where $X = A^*\Lambda$, and Λ is given by the Lyapunov equation $\Lambda^{(-1)} = A^*\Lambda A + B^*B$. Λ is unique if $\ell_A < 1$, and

$$T^*T = \left[D^*D + C^*\Lambda C\right] + \left[D^*B + C^*\Lambda A\right]Z(I-AZ)^{-1}C + C^*(I-Z^*A^*)^{-1}Z^*\left[A^*\Lambda C + B^*D\right].$$

□

Properties of outer factors

The outer factor in a spectral factorization of a positive operator has certain characteristic properties of its input and output state spaces. This is formulated in proposition 8.3. The recursive version of these properties then produces a Riccati recursive equation, and the existence of the outer factor implies the existence of a (positive) solution to this equation. Other properties of the equation can be derived from the link with outer factors as well. Recall the definitions of input and output state spaces of T as $\mathcal{H}(T) = \mathbf{P}_{\mathcal{L}_2Z^{-1}}(\mathcal{U}_2T)$, $\mathcal{H}_0(T) = \mathbf{P}(\mathcal{L}_2Z^{-1}T)$, *viz.* equations (3.45), (3.47).

LEMMA 8.2. *Let $W \in \mathcal{U}$ be boundedly invertible (in $\mathcal{X}$), with inner-outer factorization $W = UW_0$. Then U is inner, and*

$$\overline{\mathcal{H}}(W) = \overline{\mathcal{H}}(W_0)U^* \oplus \mathcal{H}(U).$$

PROOF According to theorem 4.17, U is inner if and only if $\ker(\cdot W^*) = 0$, which is the case here since W is invertible. Because W_0 is boundedly invertible, $\mathcal{X}_2 W_0^* = \mathcal{X}_2$, and because it is outer, $\mathcal{L}_2 Z^{-1} W_0^* = \mathcal{L}_2 Z^{-1}$. Hence $\mathcal{U}_2 W_0^* = \overline{\mathcal{H}}(W_0) \oplus \mathcal{U}_2$: all of $\mathcal{U}_2$ is reached. Applying this result gives

$$\begin{aligned} \mathcal{U}_2 W^* &= \mathcal{U}_2 W_0^* U^* \\ &= \overline{\mathcal{H}}(W_0) U^* \oplus \mathcal{U}_2 U^* \\ &= \overline{\mathcal{H}}(W_0) U^* \oplus \mathcal{H}(U) \oplus \mathcal{U}_2 \qquad \text{[prop. 4.1]} \\ \Rightarrow \overline{\mathcal{H}}(W) &= \overline{\mathcal{H}}(W_0) U^* \oplus \mathcal{H}(U). \end{aligned}$$

□

The following proposition is of crucial importance in proving that there is a solution to the Riccati equation associated to the time-varying spectral factorization problem which gives an outer factor W, and in characterizing this solution.

PROPOSITION 8.3. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be such that $T^* + T \gg 0$. In addition, let $W \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be an invertible factor of $T^* + T = W^* W$. Then $\mathcal{H}_0(T) \subset \mathcal{H}_0(W)$, and*

$$\begin{aligned} W \text{ outer} &\Leftrightarrow \mathcal{H}_0(T) = \mathcal{H}_0(W) \\ &\Leftrightarrow \overline{\mathcal{H}}_0(T) = \overline{\mathcal{H}}_0(W). \end{aligned}$$

If T has a realization $\{A, B, C, D\}$ (with $\ell_A < 1$) that is controllable, then W has a controllable realization with the same A and C, if and only if W is outer. Moreover, if the realization of T is uniformly controllable, then this realization of W is also uniformly controllable.

PROOF According to Arveson [6], an invertible operator $W \in \mathcal{U}$ exists such that

$$T^* + T = W^* W.$$

In general, $\mathcal{L}_2 Z^{-1} W^* \subset \mathcal{L}_2 Z^{-1}$, and $\mathcal{L}_2 Z^{-1} W^* = \mathcal{L}_2 Z^{-1}$ if and only if W is outer. Thus

$$\begin{aligned} \mathcal{H}_0(T) &= \mathbf{P}(\mathcal{L}_2 Z^{-1} T) \\ &= \mathbf{P}(\mathcal{L}_2 Z^{-1} [T + T^*]) \qquad \text{[since } T^* \in \mathcal{L} Z^{-1}\text{]} \\ &= \mathbf{P}(\mathcal{L}_2 Z^{-1} W^* W) \\ &\subset \mathbf{P}(\mathcal{L}_2 Z^{-1} W) = \mathcal{H}_0(W). \end{aligned}$$

If W is outer, then $\mathcal{L}_2 Z^{-1} W^* = \mathcal{L}_2 Z^{-1}$ and the inclusion in the above derivation becomes an identity: W outer $\Rightarrow \mathcal{H}_0(T) = \mathcal{H}_0(W)$. In this case, it is clear that W is locally finite if T is locally finite.

Suppose that W is a boundedly invertible factor of $T + T^*$, but is not outer. Then let $W = UW_0$, where U is inner and W_0 is outer. Lemma 8.2 applies: $\overline{\mathcal{H}}(W) = \overline{\mathcal{H}}(W_0) U^* \oplus \mathcal{H}(U)$.

Note that s-dim$\overline{\mathcal{H}}(W_0)U^* =$ s-dim$\overline{\mathcal{H}}(W_0)$ since U is unitary. Hence if W is not outer, that is, if $\mathcal{H}(U)\neq 0$, then

$$\text{s-dim}\,\overline{\mathcal{H}}(W) > \text{s-dim}\,\overline{\mathcal{H}}(W_0) = \text{s-dim}\,\overline{\mathcal{H}}_0(W_0) = \text{s-dim}\,\overline{\mathcal{H}}_0(T),$$

so that if W is not outer, $\overline{\mathcal{H}}_0(T) \subset \overline{\mathcal{H}}_0(W)$ (strict inclusion), *i.e.*, $\overline{\mathcal{H}}_0(T) = \overline{\mathcal{H}}_0(W) \Rightarrow$ W outer.

If T has a realization $\{A, B, C, D\}$ such that $\ell_A < 1$ and the realization is controllable, then $\overline{\mathcal{H}}_0(T) = \overline{\mathcal{D}_2(I-AZ)^{-1}C}$ (proposition 3.12). If W is outer, then $\overline{\mathcal{H}}_0(W) = \overline{\mathcal{H}}_0(T)$, so that W has a controllable realization with the same A, C as well. If W is not outer, then $\overline{\mathcal{H}}_0(T) \subset \overline{\mathcal{H}}_0(W)$ (strict inclusion), and W cannot have a controllable realization with the same A, C. If the realization of T is uniformly controllable, then W outer $\Rightarrow \mathcal{H}_0(T) = \mathcal{H}_0(W)$ and $\mathcal{H}_0(T) = \mathcal{D}_2(I-AZ)^{-1}C$ implies that the realization of W is uniformly controllable as well. □

Note that we do not require a minimal realization of T; controllability is sufficient. The proposition assures that an outer factor W of $T + T^* \gg 0$ is obtained by forcing W to have the same output state space as T, that is, the same (A, C) pair if the realization is controllable. This observation forms the main part of the proof of the following theorem, which can be used to actually compute the realization of the outer factor if a realization of T is given.

THEOREM 8.4. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be a locally finite operator with an observable and [uniformly] controllable state realization $\{A, B, C, D\}$ such that $\ell_A < 1$. Then $T^* + T \gg 0$ if and only if a solution $\Lambda \in \mathcal{D}$ exists of*

$$\Lambda^{(-1)} = A^*\Lambda A + \left[B^* - A^*\Lambda C\right](D + D^* - C^*\Lambda C)^{-1}\left[B - C^*\Lambda A\right]$$

such that $D + D^* - C^*\Lambda C \gg 0$ and $\Lambda > 0$ [$\Lambda \gg 0$]. Such a solution Λ is unique.

If $T^* + T \gg 0$, let $W \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be an invertible factor of $T^* + T = W^*W$. A realization $\{A, B_W, C, D_W\}$ for W such that W is outer is then given by the solution $\Lambda > 0$ [$\Lambda \gg 0$] of the above equation, and solutions D_W, B_W of

$$\begin{cases} D_W^* D_W &= D + D^* - C^*\Lambda C \\ B_W &= D_W^{-*}\left[B - C^*\Lambda A\right]. \end{cases}$$

PROOF Let the realization of T satisfy the given requirements, and let $W = D_W + B_W Z(I - AZ)^{-1}C$ be an operator with some [uniformly] controllable realization and the same (A, C)-pair as the realization of T, and with controllability Gramian denoted by Λ. Then, with aid of lemma 8.1, proposition 8.3 ensures that

$$\begin{cases} T + T^* = W^*W \gg 0 \\ W \text{ outer, invertible} \end{cases} \Leftrightarrow \begin{aligned} D + D^* &= D_W^* D_W + C^*\Lambda C, \qquad D_W^* D_W \gg 0 \\ BZ(I-AZ)^{-1}C &= \left[D_W^* B_W + C^*\Lambda A\right] Z(I-AZ)^{-1}C \\ \Lambda^{(-1)} &= A^*\Lambda A + B_W^* B_W, \qquad \Lambda > 0 \ [\Lambda \gg 0]. \end{aligned}$$

Here, we used in particular the fact that a realization of W can have the same (A, C) pair as the (minimal) realization of T if and only if W is outer. We also used that W invertible with W^{-1} again upper implies that D_W has an inverse D_W^{-1}, so that $D_W^* D_W \gg 0$. Because the realization of T is observable, the operator $\cdot\,(I - AZ)^{-1}C$ is one-to-one (by definition 3.5), and the above set of equations reduce to

$$\left\{\begin{array}{l} T + T^* = W^* W \gg 0 \\ W \text{ outer, invertible} \end{array}\right. \quad \Leftrightarrow \quad \begin{array}{rcll} D + D^* & = & D_W^* D_W + C^* \Lambda C, & D_W^* D_W \gg 0 \\ B & = & D_W^* B_W + C^* \Lambda A & \\ \Lambda^{(-1)} & = & A^* \Lambda A + B_W^* B_W, & \Lambda > 0 \ [\Lambda \gg 0]. \end{array}$$

Let $T + T^* \gg 0$, and let W be an invertible outer operator such that $T + T^* = W^* W$. Then D_W is invertible, and B_W can be determined from the above equations as

$$\begin{array}{rcl} D_W^* D_W & = & D + D^* - C^* \Lambda C \\ B_W & = & D_W^{-*} \left[B - C^* \Lambda A\right] \\ \Lambda^{(-1)} & = & A^* \Lambda A + \left[B^* - A^* \Lambda C\right] (D + D^* - C^* \Lambda C)^{-1} \left[B - C^* \Lambda A\right], \end{array}$$

so that Λ satisfies the given Riccati equation. In fact, we showed that if $T + T^* \gg 0$, the existence of an outer factor implies that there is a solution Λ of the Riccati equation which is [uniformly] positive, and such that also $D + D^* - C^* \Lambda C \gg 0$. The converse, to show that $T + T^* \gg 0$ if these quantities are positive, respectively uniformly positive, is obvious at this point, since such a solution directly specifies a realization of an outer and invertible factor W of $T + T^*$.

Finally, a solution $\Lambda > 0$ must be unique, because it is the controllability Gramian of the realization of the outer factor W, which is unique up to a left diagonal unitary factor [6]. Such a factor has no influence on the controllability Gramian. □

The above theorem can be extended to observable realizations without controllability constraint.

THEOREM 8.5. *Let $T \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be a locally finite operator with an observable state realization $\{A, B, C, D\}$ such that $\ell_A < 1$. Then $T^* + T \gg 0$ if and only if a solution $\Lambda \in \mathcal{P}$ exists of*

$$\Lambda^{(-1)} = A^* \Lambda A + \left[B^* - A^* \Lambda C\right] (D + D^* - C^* \Lambda C)^{-1} \left[B - C^* \Lambda A\right] \tag{8.1}$$

such that $D + D^* - C^* \Lambda C \gg 0$ and $\Lambda \geq 0$. Such a solution Λ is unique.

If $T^* + T \gg 0$, let $W \in \mathcal{U}(\mathcal{M}, \mathcal{M})$ be a factor of $T^* + T = W^* W$. A realization $\{A, B_W, C, D_W\}$ for W such that W is outer is then given by the solution $\Lambda \geq 0$ of the above equation, and solutions D_W, B_W of

$$\left\{\begin{array}{rcl} D_W^* D_W & = & D + D^* - C^* \Lambda C \\ B_W & = & D_W^{-*} \left[B - C^* \Lambda A\right]. \end{array}\right.$$

PROOF Let $\Lambda_{\mathbf{F}}$ be the controllability Gramian of the given realization of T. Since $\Lambda_{\mathbf{F}} \geq 0$, it has a factorization

$$\Lambda_{\mathbf{F}} = X^* \begin{bmatrix} \Lambda_{11} & \\ & 0 \end{bmatrix} X, \qquad X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

where $\Lambda_{11} > 0$ and X is an invertible operator (*e.g.*, X can be chosen unitary). X has the indicated block decomposition, satisfying $\operatorname{ran}(\cdot X_2) = \ker(\cdot \Lambda_{\mathbf{F}})$, $\operatorname{ran}(\cdot X_1) = \overline{\operatorname{ran}}(\cdot \Lambda_{\mathbf{F}})$. Applying X as state transformation to $\mathbf{T}$ leads to a realization $\mathbf{T}' = \{A', B', C', D\}$ given by

$$\begin{bmatrix} A' & C' \\ B' & D \end{bmatrix} = \begin{bmatrix} X & \\ & I \end{bmatrix} \begin{bmatrix} A & C \\ B & D \end{bmatrix} \begin{bmatrix} X^{-(-1)} & \\ & I \end{bmatrix}.$$

$\Lambda_{\mathbf{F}'} := \begin{bmatrix} \Lambda_{11} & \\ & 0 \end{bmatrix}$ is the controllability Gramian of $\mathbf{T}'$, and satisfies the Lyapunov equation $\Lambda_{\mathbf{F}'}^{(-1)} = A'^* \Lambda_{\mathbf{F}'} A' + B'^* B'$. Partition A', B', C' conformably to the partitioning of X. Then

$$A' = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}, \qquad B' = [B_1 \quad 0], \qquad C' = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix}, \tag{8.2}$$

because the above Lyapunov equation leads, in particular, to $0 = B_2^* B_2 + A_{12}^* \Lambda_{11} A_{12}$, so that, since $\Lambda_{11} > 0$, we must have that $B_2 = 0$ and $A_{12} = 0$.

It follows that $\{A_{11}, B_1, C_1, D\}$ is a realization of T which is both observable and controllable. By theorem 8.4, $T + T^* \gg 0$ if and only if an operator $P \in \mathcal{D}$, $P > 0$ exists such that

$$\begin{aligned} P^{(-1)} &= A_{11}^* P A_{11} + \left[B_1^* - A_{11}^* P C_1\right] (D + D^* - C_1^* P C_1)^{-1} \left[B_1 - C_1^* P A_{11}\right] \\ D + D^* - C_1^* P C_1 &\gg 0. \end{aligned}$$

The solution P is unique given the realization $\{A_{11}, B_1, C_1, D\}$. Define $P' = \begin{bmatrix} P & \\ & 0 \end{bmatrix}$, $\Lambda = X^* P' X = X_1^* P X_1$, then $\Lambda \geq 0$. Because of the structure of A', B' in (8.2), P' satisfies

$$\begin{aligned} P'^{(-1)} &= A'^* P' A' + \left[B'^* - A'^* P' C'\right] (D + D^* - C'^* P' C')^{-1} \left[B' - C'^* P' A'\right] \\ D + D^* - C'^* P' C' &\gg 0, \end{aligned}$$

so that Λ satisfies (8.1). The uniqueness of P implies the uniqueness of Λ, given the state transformation X. Moreover, Λ is independent of the actual choice of the state transformation X: if any other X_1' such that $\operatorname{ran}(\cdot X_1') = \overline{\operatorname{ran}}(\cdot \Lambda_{\mathbf{F}})$ is given by $X_1' = RX_1$, where $R \in \mathcal{D}$ is some invertible operator, and the corresponding $\Lambda' = X_1^* R^* \cdot R^{-*} P R^{-1} \cdot R X_1 = \Lambda$. Hence the solution $\Lambda \geq 0$ to (8.1) is unique. The first part of the corollary is proven.

By the second part of theorem 8.4, the outer factor W of $T^* + T = W^* W$ has a realization $\{A_{11}, B_{1W}, C_1, D_W\}$ with D_W and B_{1W} given by

$$\begin{aligned} D_W^* D_W &= D + D^* - C_1^* P C_1 \\ B_{1W} &= D_W^{-*} \left[B_1 - C_1^* P A_{11}\right], \end{aligned}$$

so that, in terms of P',

$$\begin{aligned} D_W^* D_W &= D + D^* - C'^* P' C' \\ [B_{1W} \quad 0] &= D_W^*(B' - C'^* P' A'). \end{aligned}$$

With $\Lambda = X^* P' X$ and $B_W := [B_{1W} \quad 0] X^{(-1)}$, we obtain a realization of W as $\{A, B_W, C, D_W\}$, where

$$\begin{aligned} D_W^* D_W &= D + D^* - C^* \Lambda C \\ B_W &= D_W^{-*} \left[B - C^* \Lambda A\right] . \end{aligned}$$

□

Theorems 8.4 and 8.5 can also be specified in two alternate forms, familiar from the time-invariant context [7, 8]:

COROLLARY 8.6. (POSITIVE REAL LEMMA) *Let $T \in \mathcal{U}$ be a locally finite operator such that $T = D + BZ(I - AZ)^{-1}C$, where $\ell_A < 1$ and $\{A, B, C, D\}$ is an observable realization of T.*

Then $T^* + T \gg 0$ if and only if there exist diagonal operators Λ, Q, B'_W with $\Lambda \geq 0$ and $Q \gg 0$ satisfying the following relationships:

$$\begin{aligned} \Lambda^{(-1)} &= A^* \Lambda A + B_W'^* Q B'_W \\ B_W'^* Q &= B^* - A^* \Lambda C \\ Q &= D + D^* - C^* \Lambda C. \end{aligned}$$

If the given realization of T is controllable, respectively uniformly controllable, then the above condition on Λ narrows down to $\Lambda > 0$, respectively $\Lambda \gg 0$.

PROOF In view of theorem 8.4, it suffices to make the connection $Q = D_W^* D_W$ and $B_W = D_W B'_W$. □

COROLLARY 8.7. (SPECTRAL FACTORIZATION) *Let $\Omega \in \mathcal{X}$ be a Hermitian operator with locally finite upper part, and given by*

$$\Omega = D + BZ(I - AZ)^{-1}C + C^*(I - Z^*A^*)^{-1}Z^*B^*,$$

where $\{A, B, C, D\}$ is an observable realization with $\ell_A < 1$. Then $\Omega \gg 0$ if and only if a solution $\Lambda \in \mathcal{D}$ exists of

$$\Lambda^{(-1)} = A^* \Lambda A + \left[B^* - A^* \Lambda C\right] (D - C^* \Lambda C)^{-1} \left[B - C^* \Lambda A\right] , \tag{8.3}$$

such that $\Lambda \geq 0$ and $D - C^* \Lambda C \gg 0$. A solution $\Lambda \geq 0$ is unique.

If $\Omega \gg 0$ and Λ is such a solution, then a realization $\{A, B_W, C, D_W\}$ for an outer factor W of Ω is given by solutions D_W, B_W of

$$\begin{aligned} D_W^* D_W &= D - C^* \Lambda C \\ B_W &= D_W^{-*} \left[B - C^* \Lambda A \right] . \end{aligned}$$

If the realization $\{A, B, C, D\}$ is controllable, respectively uniformly controllable, then $\Lambda > 0$, respectively $\Lambda \gg 0$: the realization for W is [uniformly] controllable.

8.3 COMPUTATIONAL ISSUES

In this section, we consider some computational issues that play a role in actually computing a spectral factorization of a uniformly positive operator Ω with a locally finite observable realization given as in (8.3). First, note that taking the k-th entry along each diagonal of (8.3) leads to the Riccati *recursion*

$$\Lambda_{k+1} = A_k^* \Lambda_k A_k + \left[B_k^* - A_k^* \Lambda_k C_k \right] (D_k - C_k^* \Lambda_k C_k)^{-1} \left[B_k - C_k^* \Lambda_k A_k \right] , \qquad (8.4)$$

and with Λ_k known, $(B_W)_k$, $(D_W)_k$ also follow locally:

$$\begin{aligned} (D_W)_k^* (D_W)_k &= D_k - C_k^* \Lambda_k C_k \\ (B_W)_k &= (D_W^{-*})_k \left[B_k - C_k^* \Lambda_k A_k \right] . \end{aligned}$$

Hence all that is needed in practical computations is an initial point for the recursion of Λ_k. Special cases where such an initial point can indeed be obtained are familiar from previous chapters.

Finite matrices

One case in which exact initial conditions can be obtained is the case where $\Omega \in \mathcal{X}(\mathcal{M}, \mathcal{M})$ is actually a finite matrix, *i.e.*, where

$$\mathcal{M} = \cdots \times \varnothing \times \varnothing \times \mathcal{M}_1 \times \mathcal{M}_2 \times \cdots \times \mathcal{M}_n \times \varnothing \times \cdots .$$

In this case, Ω is a finite $n \times n$ (block) matrix, and a realization for Ω can start off with no states at point 1 in time. Since the dimension of Λ follows that of A, an exact initial point for the recursion is $\Lambda_1 = [\,\cdot\,]$ (a 0×0 matrix). The spectral factorization reduces for finite matrices to a Cholesky factorization, and the resulting algorithm is an efficient way to compute Cholesky factorizations for (large) matrices with a sparse state space.

Initial time-invariance

A second class of systems are systems which are time invariant before some point in time, say $k = 1$. Then, before point $k = 1$, all Λ_k are equal to each other, and in particular

$\Lambda_0 = \Lambda_1$. Hence the recursion for Λ reduces to an algebraic equation

$$\Lambda_0 = A_0^*\Lambda_0 A_0 + \left[B_0^* - A_0^*\Lambda C_0\right](D_0 - C_0^*\Lambda_0 C_0)^{-1}\left[B_0 - C_0^*\Lambda_0 A_0\right],$$

which is the classical time-invariant Riccati equation. A solution to this equation can be obtained in one of the classical ways, *e.g.*, as the solution of a Hamiltonian equation. Multiple solutions exist, and in order to obtain an outer spectral factor W, the 'maximal positive' solution of the above equation must be chosen: $\Lambda_0 \geq 0$. Because the Λ_k $(k > 0)$ are determined by Λ_0 via the recursion (8.4), and because the solution Λ is unique, it follows that the requirement to choose this Λ_0 is also sufficient to have Λ positive.

Periodic systems

If Ω is periodically time varying, with period n say, then one can apply the usual time-invariance transformation, by considering a block system consisting of n consecutive state realization sections. Since the block-system is time invariant, one can compute Λ_1 from the resulting block-Riccati equation with the classical techniques, and Λ_1 is an exact initial condition to compute the realization of the spectral factor for time points $2, \cdots, n$. As usual, such a technique may not be attractive if the period is large.

Unknown initial conditions

Finally, we consider the more general case where Ω is not completely specified but only, say, the submatrix $[\Omega_{i,j}]_0^\infty$ is known. The 'past' of Ω is assumed to be such that $\Omega \gg 0$. In this case, the exact initial point for the recursion of Λ_k is unknown. It is possible to start the recursion (8.4) from an approximate initial point, for which typically $\hat{\Lambda}_0 = 0$ is chosen. The convergence of this choice is investigated in the following section. It is shown in proposition 8.9 that when the realization $\{A, B, C, D\}$ is observable and has $\ell_A < 1$, then $\hat{\Lambda}_k$ (corresponding to the recursion (8.4) with initial point $\hat{\Lambda}_0 = 0$) converges to Λ_k, the exact solution obtained with the correct initial point Λ_0.

8.4 CONVERGENCE OF THE RICCATI RECURSION

In this section, we study the convergence of an approximate solution $\hat{\Lambda}_k$ $(k \geq 0)$ to the Riccati recursion (8.4), if the recursion is started with $\hat{\Lambda}_0 = 0$ rather than the exact initial point Λ_0. It is shown that $\hat{\Lambda}_k \to \Lambda_k$ for $k \to \infty$, when $\Omega \gg 0$, $\ell_A < 1$ and the given realization is observable. Similar results are well known for the time-invariant case, and for the time-varying case some results are known from the connection of the Riccati recursion with Kalman filtering (*cf.* [9, 10]). However, the derivation given below is more general because state dimensions are allowed to vary, and hence A_k cannot be assumed to be square and invertible, as required in [9].

Consider the following block decomposition of the matrix representation of $\Omega = W^*W$, and a related operator $\hat{\Omega} = \hat{W}^*\hat{W}$:

$$\Omega = \left[\begin{array}{c|cc} \underline{\Omega}_{11} & \underline{\Omega}_{12} & \underline{\Omega}_{13} \\ \hline \underline{\Omega}_{12}^* & \underline{\Omega}_{22} & \underline{\Omega}_{23} \\ \underline{\Omega}_{13}^* & \underline{\Omega}_{23}^* & \underline{\Omega}_{33} \end{array}\right], \qquad W = \left[\begin{array}{c|cc} W_{11} & W_{12} & W_{13} \\ \hline & W_{22} & W_{23} \\ & & W_{33} \end{array}\right]$$
$$\hat{\Omega} = \left[\begin{array}{c|cc} \underline{\Omega}_{11} & 0 & 0 \\ \hline 0 & \underline{\Omega}_{22} & \underline{\Omega}_{23} \\ 0 & \underline{\Omega}_{23}^* & \underline{\Omega}_{33} \end{array}\right], \qquad \hat{W} = \left[\begin{array}{c|cc} \hat{W}_{11} & 0 & 0 \\ \hline & \hat{W}_{22} & \hat{W}_{23} \\ & & \hat{W}_{33} \end{array}\right]. \tag{8.5}$$

In these decompositions,[3] $\underline{\Omega}_{11}$ corresponds to $[\Omega_{i,j}]_{-\infty}^{-1}$, $\underline{\Omega}_{22} = [\Omega_{i,j}]_0^{n-1}$ is a finite $n \times n$ matrix (where n is some integer to be specified later), and $\underline{\Omega}_{33}$ corresponds to $[\Omega_{i,j}]_n^{\infty}$. The point of introducing the operator $\hat{\Omega}$ is that $\hat{\Lambda}_0$ is the *exact* initial point of the Riccati recursion (8.4) for a spectral factorization of the lower right part of $\hat{\Omega}$, and leads to a spectral factor $\hat{W}$ such that $\hat{\Omega} = \hat{W}^*\hat{W}$, of which only the lower right part is computed. This is seen by putting $A_{-1} = 0$, $B_{-1} = 0$ in the Riccati recursion for Λ, which leads to $\hat{\Lambda}_0 = 0$. The convergence of $\hat{\Lambda}_k$ to Λ_k is studied from this observation.

As a preliminary step, the following lemma considers a special case of the above Ω.

LEMMA 8.8. *Let be given an operator $\Omega \in \mathcal{X}$, $\Omega \gg 0$, with block decomposition*

$$\Omega = \begin{bmatrix} \underline{\Omega}_{11} & \underline{\Omega}_{12} & 0 \\ \underline{\Omega}_{12}^* & \underline{\Omega}_{22} & \underline{\Omega}_{23} \\ 0 & \underline{\Omega}_{23}^* & \underline{\Omega}_{33} \end{bmatrix}$$

where $\underline{\Omega}_{22}$ is an $n \times n$ matrix. Let the upper triangular part of Ω be locally finite and strictly stable. Then

$$(\Omega^{-1})_{33} \to (\underline{\Omega}_{33} - \underline{\Omega}_{23}^*\underline{\Omega}_{22}^{-1}\underline{\Omega}_{23})^{-1} \quad \text{as } n \to \infty$$

(strong convergence). Hence $(\Omega^{-1})_{33} \to (\hat{\Omega}^{-1})_{33}$, where $\hat{\Omega}$ is equal to Ω, but with $\underline{\hat{\Omega}}_{12} = 0$.

PROOF Let $\{A, B, C, D\}$ be a realization of the upper triangular part of Ω with $\ell_A < 1$. Let $\underline{\Omega}_{12} = \underline{\mathcal{C}}_1\underline{\mathcal{O}}_1$, $\underline{\Omega}_{23} = \underline{\mathcal{C}}_2\underline{\mathcal{O}}_2$, where

$$\underline{\mathcal{C}}_1 = \begin{bmatrix} \vdots \\ B_{-3}A_{-2}A_{-1} \\ B_{-2}A_{-1} \\ B_{-1} \end{bmatrix}, \qquad \underline{\mathcal{C}}_2 = \begin{bmatrix} B_0A_1 \cdots A_{n-1} \\ \vdots \\ B_{n-3}A_{n-2}A_{n-1} \\ B_{n-2}A_{n-1} \\ B_{n-1} \end{bmatrix},$$
$$\underline{\mathcal{O}}_1 = [\, C_0 \quad A_0C_1 \quad A_0A_1C_2 \quad \cdots \quad A_0 \cdots A_{n-2}C_{n-1} \,]$$
$$\underline{\mathcal{O}}_2 = [\, C_n \quad A_nC_{n+1} \quad A_nA_{n+1}C_{n+2} \quad \cdots \,].$$

[3]The underscore is used in this section to denote that we take block submatrices rather than entries of Ω.

Then $\underline{\mathcal{O}}_1\underline{\mathcal{C}}_2$ is a summation of n terms, each containing a product $A_0 \cdots A_{i-1}$ and a product $A_{i+1} \cdots A_{n-1}$. Because $\ell_A < 1$ implies that products of the form $A_k \cdots A_{k+n} \to 0$ as $n \to \infty$ strongly and uniformly in k, we obtain $\underline{\mathcal{O}}_1\underline{\mathcal{C}}_2 \to 0$ if $n \to \infty$.

Write $X_3 = (\Omega^{-1})_{33}$. By repeated use of Schur's inversion formula (lemma 7.2), X_3 is given by the recursion

$$X_1 = \underline{\Omega}_{11}^{-1}, \qquad X_{k+1} = (\underline{\Omega}_{k+1,k+1} - \underline{\Omega}_{k,k+1}^* X_k \underline{\Omega}_{k,k+1})^{-1}. \tag{8.6}$$

We first consider a special case, where $\underline{\Omega}_{k,k} = I$ $(k = 1, 2, 3)$. In the derivation below, we, for ease of discussion, assume that also $\underline{\mathcal{O}}_k\underline{\mathcal{O}}_k^* = I$, *i.e.*, the realization is uniformly observable and in output normal form, although this is not an essential requirement. The recursion (8.6) becomes

$$\begin{aligned} Y_k &= \underline{\mathcal{C}}_k^* X_k \underline{\mathcal{C}}_k \\ X_{k+1} &= (I - \underline{\mathcal{O}}_k^* Y_k \underline{\mathcal{O}}_k)^{-1} = I + \underline{\mathcal{O}}_k^* \left[Y_k + Y_k^2 + \cdots \right] \underline{\mathcal{O}}_k, \end{aligned}$$

so that, in particular,

$$Y_2 = \underline{\mathcal{C}}_2^*\underline{\mathcal{C}}_2 + \underline{\mathcal{C}}_2^*\underline{\mathcal{O}}_1^* \left[Y_1(I - Y_1)^{-1} \right] \underline{\mathcal{O}}_1\underline{\mathcal{C}}_2.$$

For large n, $Y_2 \to \underline{\mathcal{C}}_2^*\underline{\mathcal{C}}_2$ and becomes independent of Y_1 and $\underline{\mathcal{C}}_1$, and

$$X_3 \to (I - \underline{\mathcal{O}}_2^*\underline{\mathcal{C}}_2^*\underline{\mathcal{C}}_2\underline{\mathcal{O}}_2)^{-1} = (\underline{\Omega}_{33} - \underline{\Omega}_{23}^*\underline{\Omega}_{22}^{-1}\underline{\Omega}_{23})^{-1}$$

independently of $\underline{\mathcal{C}}_1$. The expression on the right-hand side is the same as the value obtained for $\underline{\mathcal{C}}_1 = 0$, *i.e.*, $\underline{\Omega}_{12} = 0$.

The general case reduces to the above special case by a pre- and post-multiplication by

$$\begin{bmatrix} \underline{\Omega}_{11}^{-1/2} & & \\ & \underline{\Omega}_{22}^{-1/2} & \\ & & \underline{\Omega}_{33}^{-1/2} \end{bmatrix}.$$

This maps $\underline{\Omega}_{k,k}$ to I, $\underline{\mathcal{C}}_k$ to $\underline{\Omega}_{k,k}^{-1/2}\underline{\mathcal{C}}_k$, and $\underline{\mathcal{O}}_k$ to $\underline{\mathcal{O}}_k\underline{\Omega}_{k+1,k+1}^{-1/2}$. The latter two mappings lead to realizations with different B_i and C_i, but the A_i remain the same, and in particular the convergence properties of $\underline{\mathcal{C}}_2\underline{\mathcal{O}}_1$ remain unchanged. It follows that $(\Omega^{-1})_{33} \to (\underline{\Omega}_{33} - \underline{\Omega}_{23}^*\underline{\Omega}_{22}^{-1}\underline{\Omega}_{23})^{-1}$ also in the general case. □

We now return to the spectral factorization problem, with Ω given as in (8.5).

PROPOSITION 8.9. *Let $\Omega \in \mathcal{X}$, $\Omega \gg 0$ have an upper triangular part which is locally finite and given by an observable realization $\{A, B, C, D\}$ where $\ell_A < 1$. Let $\Lambda \in \mathcal{D}$ be the unique solution of (8.3) so that its entries Λ_n satisfy the recursive Riccati equation (8.4). Let $\hat{\Lambda}_n$ $(n \geq 0)$ be the sequence obtained from the same recursion, but starting from $\hat{\Lambda}_0 = 0$.*

Then $\hat{\Lambda}_n \to \Lambda_n$ as $n \to \infty$ (strong convergence).

PROOF Let Ω, $\hat{\Omega}$ have block decompositions as in (8.5), where $\underline{\Omega}_{22}$ is an $n \times n$ matrix. Let $\Omega = W^*W$, $\hat{\Omega} = \hat{W}^*\hat{W}$, where W, $\hat{W}$ are outer spectral factors, then Λ, $\hat{\Lambda}$ are the controllability Gramians of the realization of W, $\hat{W}$ given in corollary 8.7. Denote

$$\begin{aligned} W_{12} &= \underline{\mathcal{C}}_{W,1}\underline{\mathcal{O}}_1 \\ W_{23} &= \underline{\mathcal{C}}_{W,2}\underline{\mathcal{O}}_2 \\ W_{13} &= \underline{\mathcal{C}}_{W,1}A_0A_1\cdots A_{n-1}\underline{\mathcal{O}}_2 . \end{aligned}$$

Because $\ell_A < 1$, we have that $W_{13} \to 0$ as $n \to \infty$ (strongly), so that for large enough n, $\Lambda_n \approx \underline{\mathcal{C}}^*_{W,2}\underline{\mathcal{C}}_{W,2}$ and hence

$$\begin{aligned} \underline{\Omega}_{33} &= W^*_{33}W_{33} + W^*_{23}W_{23} + W^*_{13}W_{13} \\ &\approx W^*_{33}W_{33} + \underline{\mathcal{O}}^*_2\Lambda_n\underline{\mathcal{O}}_2 . \end{aligned}$$

Consequently, $\underline{\mathcal{O}}^*_2(\Lambda_n - \hat{\Lambda}_n)\underline{\mathcal{O}}_2 \approx \hat{W}^*_{33}\hat{W}_{33} - W^*_{33}W_{33}$. The next step is to show that $\hat{W}^*_{33}\hat{W}_{33} - W^*_{33}W_{33} \to 0$ for large n, so that, if the realization is observable, $\hat{\Lambda}_n \to \Lambda_n$.

Let $X_3 = (W^*_{33}W_{33})^{-1}$, and $\hat{X}_3 = (\hat{W}^*_{33}\hat{W}_{33})^{-1}$. Since $\Omega^{-1} = W^{-1}W^{-*}$, and W is outer so that $W^{-1} \in \mathcal{U}$, it follows that $X_3 = (\Omega^{-1})_{33}$ and $\hat{X}_3 = (\hat{\Omega}^{-1})_{33}$. Lemma 8.8 proves that, if $\ell_A < 1$, then $(\Omega^{-1})_{33} \to (\hat{\Omega}^{-1})_{33}$ as $n \to \infty$, so that $X_3 \to \hat{X}_3$, and hence $\hat{\Lambda}_n \to \Lambda_n$. □

Note that in the premisses of the above convergence proposition, we required the given realization to be observable. This condition was not present in the convergence proof of the Riccati equation that occurred in the solution of the embedding problem (*cf.* proposition 7.6). It is conjectured that the observability condition is not necessary for convergence, but is only necessary to obtain outer factors.

Finally, we remark that always $\hat{\Lambda}_k \leq \Lambda_k$. This is a consequence of the fact that

$$\hat{\Lambda}_k \leq \Lambda_k \quad \Rightarrow \quad \hat{\Lambda}_{k+1} \leq \Lambda_{k+1} , \tag{8.7}$$

which can be proven directly from the Riccati recursion (8.4) in a way similar to [10, ch. 9]. Indeed, let the matrix G_{X,Λ_k} be given by

$$G_{X,\Lambda_k} = \begin{bmatrix} X - A_k^*\Lambda_kA_k & B_k - C_k^*\Lambda_kA_k \\ B_k^* - A_k^*\Lambda_kC_k & D_k - C_k^*\Lambda_kC_k \end{bmatrix} = \begin{bmatrix} X & B_k \\ B_k^* & D_k \end{bmatrix} - \begin{bmatrix} A_k^* \\ C_k^* \end{bmatrix} \Lambda_k \begin{bmatrix} A_k & C_k \end{bmatrix} ,$$

parameterized by some matrix $X = X^*$. Using Schur's complements, it follows that, if $D_k - C_k^*\Lambda_kC_k > 0$, then

$$G_{X,\Lambda_k} \geq 0 \quad \Rightarrow \quad X - A_k^*\Lambda_kA_k - \left[B_k^* - A_k^*\Lambda C_k\right](D_k - C_k^*\Lambda_kC_k)^{-1}\left[B_k - C_k^*\Lambda_kA_k\right] \geq 0 .$$

Hence $\Lambda_{k+1} = \min\{X : G_{X,\Lambda_k} \geq 0\}$. But if $\hat{\Lambda}_k \leq \Lambda_k$, then $G_{\Lambda_{k+1},\hat{\Lambda}_k} \geq G_{\Lambda_{k+1},\Lambda_k} \geq 0$. It follows that $\Lambda_{k+1} \geq \hat{\Lambda}_{k+1}$, since $\hat{\Lambda}_{k+1}$ is the smallest matrix X for which $G_{X,\hat{\Lambda}_k} \geq 0$.

8.5 CONNECTIONS

In this section, we point out some of the connections between the spectral factorization results of the preceding sections, and the incarnations of the time-varying Riccati equation that we encountered earlier in the solution of the orthogonal embedding problem (chapter 7) and inner-outer factorizations (chapter 4).

Orthogonal embedding

Recall the orthogonal embedding problem: given a transfer operator T of a causal bounded discrete-time linear system, extend this system by adding more inputs and outputs to it such that the resulting system Σ,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

is inner and has $T = \Sigma_{11}$. The embedding problem can be viewed as a spectral factorization problem of $\Sigma_{12}^*\Sigma_{12} = I - T^*T$, which gives Σ_{21}. Via this connection, the solution of the embedding problem can also be obtained starting from the spectral factorization theorems 8.4 and 8.5. This leads to a variant of the embedding theorem 7.5 which is more general, as it also follows that Σ_{21} is in fact an outer operator.

THEOREM 8.10. *Let $T \in \mathcal{U}(\mathcal{M}_1, \mathcal{N}_1)$ be a locally finite operator with an observable state realization $\{A, B, C, D\}$ such that $\ell_A < 1$. Then $I - T^*T \gg 0$ if and only if a solution $M \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ exists of*

$$M^{(-1)} = A^*MA + B^*B + \left[A^*MC + B^*D\right](I - D^*D - C^*MC)^{-1}\left[D^*B + C^*MA\right]. \quad (8.8)$$

such that $I - D^*D - C^*MC \gg 0$ and $M \geq 0$. This M is unique. If in addition the realization of T is [uniformly] controllable, then M is [uniformly] positive.

If $I - T^*T \gg 0$, let $W \in \mathcal{U}(\mathcal{N}_1, \mathcal{N}_1)$ be a factor of $I - T^*T = W^*W$. A realization $\{A, B_W, C, D_W\}$ for W such that W is outer is then given by the solution M of the above equation, and solutions D_W, B_W of

$$\begin{cases} D_W^*D_W &= I - D^*D - C^*MC \\ B_W &= -D_W^{-*}\left[D^*B + C^*MA\right]. \end{cases} \quad (8.9)$$

PROOF Since $\ell_A < 1$, the Lyapunov equation

$$\Lambda^{(-1)} = A^*\Lambda A + B^*B$$

has a unique solution $\Lambda \geq 0$. By lemma 8.1, an expression for $I - T^*T$ is

$$I - T^*T = (I - D^*D - C^*\Lambda C) - \left[D^*B + C^*\Lambda A\right] Z(I - AZ)^{-1}C - C^*(I - Z^*A^*)^{-1}Z^*\left[B^*D + A^*\Lambda C\right].$$

The implied realization for the upper part of $I - T^*T$ need not be controllable. Theorem 8.5 claims that $I - T^*T \gg 0$ if and only if a solution $P \in \mathcal{P}$ exists of

$$P^{(-1)} = A^*PA + \left[B^*D + A^*(\Lambda + P)C\right](I - D^*D - C^*(\Lambda + P)C)^{-1}\left[D^*B + C^*(\Lambda + P)A\right]$$

such that $I - D^*D - C^*(\Lambda + P)C \gg 0$ and $P \geq 0$. This P is unique. As a consequence, the operator $M = \Lambda + P$ is positive semi-definite, unique and satisfies equation (8.8). If the realization of T is [uniformly] controllable, then $\Lambda > 0$ [$\Lambda \gg 0$], and the same holds for M.

Theorem 8.5 in addition shows that the realization $\{A, B_W, C, D_W\}$, with D_W, B_W as given in (8.9), defines an outer factor W of $I - T^*T = W^*W$. □

COROLLARY 8.11. *If, in theorem 7.5, the realization of T is observable, then Σ_{21} is an outer operator.*

Inner-outer factorization

A realization of the outer factor T_0 in an inner-outer factorization of T can also be computed via a Riccati equation, as was shown in theorem 4.19. A realization of the outer factor followed from a observable realization $\{A, B, C, D\}$ of T as

$$\mathbf{T}_0 = \begin{bmatrix} I & \\ & R^* \end{bmatrix} \begin{bmatrix} A & C \\ C^*MA + D^*B & C^*MC + D^*D \end{bmatrix} \tag{8.10}$$

where $M \geq 0$ is the solution of maximal rank of

$$M^{(-1)} = A^*MA + B^*B - \left[A^*MC + B^*D\right](D^*D + C^*MC)^{\dagger}\left[D^*B + C^*MA\right]$$

and R is a minimal (full range) factor of $RR^* = (D^*D + C^*MC)^{\dagger}$. Let T_0 be invertible, so that the pseudo-inverse becomes an ordinary inverse. Using lemma 8.1, one can verify that, indeed, $T^*T = T_0^*T_0$, by deriving that the realizations of the upper parts are equal. With lemma 8.1, the realization of the upper part of $T_0^*T_0$ is obtained from (4.19) as

$$\begin{bmatrix} A & C \\ (D^*B + C^*MA) + C^*\Lambda'A & (D^*D + C^*MC) + C^*\Lambda'C \end{bmatrix} \tag{8.11}$$

where Λ' is the unique operator satisfying the Lyapunov equation

$$\Lambda' = A^*\Lambda'A + \left[B^*D + A^*MC\right](D^*D + C^*MC)^{-1}\left[D^*B + C^*MA\right].$$

Consequently, $(\Lambda' + M)^{(-1)} = A^*(\Lambda' + M)A + B^*B$, so that $\Lambda = \Lambda' + M$ satisfies the Lyapunov equation $\Lambda^{(-1)} = A^*\Lambda A + B^*B$. With Λ, the realization (8.11) becomes

$$\begin{bmatrix} A & C \\ B^*D + C^*\Lambda A & D^*D + C^*\Lambda C \end{bmatrix},$$

which is the same realization as that of T^*T in lemma 8.1. Conversely, one can try to derive theorem 4.19 from the spectral factorization theorem in this manner, for the special case where T^*T is invertible (theorem 4.19 is more general).

Cholesky factorization

As noted before, spectral factorization of a finite positive matrix reduces to Cholesky factorization. For time-invariant systems (Toeplitz operators), one efficient technique to compute a Cholesky factorization makes use of Schur recursions [11, 12]. The Schur algorithm can be generalized in various ways to apply to triangular factorizations of general matrices [13], structured operators which have a displacement structure [14, 15, 16, 17], and approximate factorizations on a staircase band [18]. See [19] for an overview.

The key step in the traditional and also generalized Schur and Levinson algorithms is the translation of the original context (Ω, with $\Omega > 0$) to a scattering context (contractive operators). A standard transition to the scattering context is obtained by finding upper triangular operators Γ, Δ, such that $\Omega = \Gamma\Gamma^* - \Delta\Delta^*$. Using $\mathbf{P}(\Omega)$, the upper triangular part of Ω, possible Γ and Δ are defined by

$$\begin{array}{lcll} \Omega_1 & = & 2\mathbf{P}(\Omega) - I & \\ \Gamma & = & \frac{1}{2}(\Omega_1 + I) & = \mathbf{P}(\Omega) \\ \Delta & = & \frac{1}{2}(\Omega_1 - I) & = \mathbf{P}(\Omega) - I \end{array}$$

It is readily verified that, indeed, $\Omega = \Gamma\Gamma^* - \Delta\Delta^*$, and because $\Omega > 0$, Γ is boundedly invertible and $S := \Gamma^{-1}\Delta = (\Omega_1 + I)^{-1}(\Omega_1 - I)$ is a well-defined and contractive operator: $\|S\| < 1$. The definition of S may be recognized as a Cayley transformation of Ω_1. It has a direct relation with Ω:

$$\mathbf{P}(\Omega) = (I - S)^{-1}; \qquad S = I - [\mathbf{P}(\Omega)]^{-1}.$$

Since S is strictly contractive and $\mathbf{P}(\Omega)$ is upper triangular, the first expression ensures that S is upper triangular. S is even strictly upper triangular because Δ is so. Also the state structure is preserved: S has the same number of states as $\mathbf{P}(\Omega)$, and its model can be directly derived from the model of $\mathbf{P}(\Omega)$ using equation (1.3).

The standard way to obtain a Cholesky factorization of Ω continues as follows. Compute any J-unitary matrix Θ such that

$$[\Gamma \quad \Delta]\Theta = [A_1 \quad 0], \tag{8.12}$$

A consequence of the J-unitarity of Θ is that

$$A_1 A_1^* = [\Gamma \quad \Delta]\Theta J \Theta \begin{bmatrix} \Gamma^* \\ \Delta^* \end{bmatrix} = \Gamma\Gamma^* - \Delta\Delta^* = \Omega.$$

Hence A_1 is a factor of Ω: $\Omega = A_1 A_1^*$. With Θ, a factor of Ω^{-1} is obtained by computing

$$[I \quad I]\Theta = [A_2 \quad B_2], \tag{8.13}$$

as it is readily verified using (8.12) and the J-unitarity of Θ that $\Omega^{-1} = A_2 A_2^* = B_2 B_2^*$. Hence knowledge of Θ provides both a factorization of Ω and of its inverse. Θ can be computed recursively using a generalized Schur algorithm (as *e.g.*, in [18]) which amounts to a repetition of (*i*) shifting the rows of Γ one position to the right to align with Δ (*i.e.*, postmultiplication by Z), and (*ii*) using an elementary Θ 'section' to cancel the front diagonal of Δ against the corresponding diagonal of Γ. It is thus an order-recursive algorithm. For finite upper triangular matrices of size $n \times n$, the algorithm can be carried out in a finite number of steps and yields a Θ-matrix having at most $n-1$ states. It is possible to obtain an approximate factor by making Δ zero only on a staircase band. This leads to approximate factors A_2' of Ω^{-1} that are zero outside the staircase band, and whose inverse matches the factor A_1 of Ω on the band [18, 20].

The above algorithm is just one way to compute a Cholesky factorization of a given positive matrix Ω. Efficient ('fast') algorithms are based on exploiting knowledge on the structure of Ω. For example, if Ω is a Toeplitz matrix, then Θ can be computed using the same algorithm but now acting only on the top row of Γ and the top row of Δ (the 'generators' of Γ and Δ). This yields the traditional Schur method. More general displacement structures obeying a relation as in (1.1) are treated in much the same way [19, 17].

Using the embedding technique given in chapter 7, one other possibility to compute the Cholesky factor via Θ is the following. Assume that a computational model for $\mathbf{P}(\Omega)$, the upper triangular part of Ω, is known. We have already noted that, since S is also upper triangular, a computational model for the associated scattering operator S follows without special effort. The next step is to do an embedding: using theorem 7.5, construct a lossless embedding matrix Σ for S, which is a unitary (2×2) block matrix computed such that $\Sigma_{12} = S$. The J-unitary Θ-matrix associated to Σ is defined as usual by

$$\Theta = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1} \end{bmatrix}$$

Because of corollary 8.11, it is known at this point that Σ_{22} is outer, so that Σ_{22}^{-1} and hence Θ are again upper. Σ and Θ satisfy by construction the relations (for some $A_1' \in \mathcal{U}$)

$$[I \quad 0]\Sigma = [A_1' \quad S] \qquad \Leftrightarrow \qquad [I \quad S]\Theta = [A_1' \quad 0]$$

and since $S = \Gamma^{-1}\Delta$, multiplication by Γ shows that Θ indeed satisfies

$$[\Gamma \quad \Delta]\Theta = [A_1 \quad 0].$$

From the model of Θ, factors B_2 and $W = B_2^{-1}$ of Ω^{-1} and Ω, respectively, follow using equation (8.13). The whole algorithm can be put into a single recursion. Not surprisingly, the resulting recursion for W is precisely the Riccati equation in corollary 8.7.

Bibliography

[1] A.J. van der Veen and M.H.G. Verhaegen, "On spectral factorization and Riccati equations for time-varying systems in discrete time," *subm. IEEE Trans. Automat. Control*, Feb. 1993.

[2] B.D.O. Anderson, "An algebraic solution to the spectral factorization problem," *IEEE Trans. Automat. Contr.*, vol. 12, pp. 410–414, 1967.

[3] B.D.O. Anderson, K.L. Hitz, and N.D. Diem, "Recursive algorithm for spectral factorization," *IEEE Trans. Circuits Syst.*, vol. 21, no. 6, pp. 742–750, 1974.

[4] S. Bittanti, A.J. Laub, and J.C. Willems, eds., *The Riccati Equation*. Comm. Control Eng. Series, Springer Verlag, 1991.

[5] L.H. Son and B.D.O. Anderson, "Design of Kalman filters using signal model output statistics," *Proc. IEE*, vol. 120, no. 2, pp. 312–318, 1973.

[6] W. Arveson, "Interpolation problems in nest algebras," *J. Functional Anal.*, vol. 20, pp. 208–233, 1975.

[7] B.D.O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis*. Prentice Hall, 1973.

[8] M.J. Denham, "On the factorization of discrete-time rational spectral density matrices," *IEEE Trans. Automat. Control*, pp. 535–537, 1975.

[9] H.B. Aasnaes and T. Kailath, "Initial-condition robustness of Linear Least-Squares filtering algorithms," *IEEE Trans. Automat. Control*, vol. 19, pp. 393–397, Aug. 1974.

[10] B.D.O. Anderson and J.B. Moore, *Optimal Filtering*. Prentice Hall, 1979.

[11] I. Schur, "Uber Potenzreihen, die im Innern des Einheitskreises beschränkt sind, I," *J. Reine Angew. Math.*, vol. 147, pp. 205–232, 1917. Eng. Transl. *Operator Theory: Adv. Appl.*, vol. 18, pp. 31-59, Birkhäuser Verlag, 1986.

[12] T. Kailath, "A theorem of I. Schur and its impact on modern signal processing," in *Operator Theory: Advances and Applications*, vol. 18, pp. 9–30, Basel: Birkhäuser Verlag, 1986.

[13] H. Ahmed, J. Delosme, and M. Morf, "Highly concurrent computing structures for matrix arithmetic and signal processing," *Computer*, pp. 65–82, Jan. 1982.

[14] T. Kailath, S.Y. Kung, and M. Morf, "Displacement ranks of matrices and linear equations," *J. Math. Anal. Appl.*, vol. 68, no. 2, pp. 395–407, 1979.

[15] H. Lev-Ari and T. Kailath, "Lattice filter parametrization and modeling of nonstationary processes," *IEEE Trans. Informat. Th.*, vol. 30, pp. 2–16, Jan. 1984.

[16] H. Lev-Ari and T. Kailath, "Triangular factorizations of structured Hermitian matrices," in *Operator Theory: Advances and Applications*, vol. 18, pp. 301–324, Birkhäuser Verlag, 1986.

[17] H. Lev-Ari and T. Kailath, "Lossless arrays and fast algorithms for structured matrices," in *Algorithms and Parallel VLSI Architectures* (Ed. F. Deprettere and A.J. van der Veen, eds.), vol. A, pp. 97–112, Elsevier, 1991.

[18] P. Dewilde and E. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy," in *Operator Theory: Advances and Applications*, vol. 29, pp. 97–116, Birkhäuser Verlag, 1988.

[19] J. Chun, *Fast Array Algorithms for Structured Matrices*. PhD thesis, Stanford Univ., Stanford, CA, 1989.

[20] H. Nelis, *Sparse Approximations of Inverse Matrices*. PhD thesis, Delft Univ. Techn., The Netherlands, 1989.

Chapter 9

LOSSLESS CASCADE FACTORIZATIONS

In chapter 7, we showed how a contractive transfer operator T can be embedded into an inner operator Σ. We now derive minimal structural factorizations of locally finite inner transfer operators into elementary inner operators of degree one. The resulting lossless cascade networks provide a canonical realization of T into a network of minimal degree and with a minimal number of coefficients. For a better understanding of the problem, we will first review some aspects of cascade factorizations for time-invariant systems.

9.1 TIME-INVARIANT CASCADE FACTORIZATIONS

Overview

An important and recurring subject in network theory concerns the synthesis (implementation, or actual realization) of a desired transfer function using elementary components. For continuous-time systems, these components would be resistors, capacitances, inductors and transformers. In the discrete-time context, the elementary operator is the basic processor which performs the actual calculations on the digital signals: typically a multiplier-adder, but other elementary processors are certainly possible. While one can directly use the given $\{A, B, C, D\}$ realization as the actual realization of the transfer operator, doing so is often unsatisfactory. The number of multiplications in an arbitrary state realization of the given system is not minimal: a single-input single output system with n states would require $(n+1)^2$ multiplications. Typically, such an implementation is also rather sensitive to small changes in the values of the coefficients: a small change (*e.g.*, because of finite word length effects) can sometimes even make the modified system unstable. For digital filters, a third issue is the occurrence of limit cycles and register overflow. The above-mentioned effects are mitigated by a deliberate use of the freedom of state transformations on the given state realization. By selecting certain canonical forms of the A matrix, such as a companion form or a diagonal form (which is not always possible), filters specified

by a minimal number of coefficients are obtained [1].

The coefficient sensitivity issue is a more complicated matter. The central idea is that one of the few ways to make the locations of poles and zeros of the resulting system well defined is to factor the given transfer operator into a cascade of elementary (degree one) transfer operators:

$$T(z) = T_1(z) \cdot T_2(z) \cdots T_n(z). \tag{9.1}$$

Each elementary transfer operator realizes a pole and a zero of $T(z)$. For an n-th order system $T(z)$, the factorization is minimal if it consists of n degree one sections. In this case, the factorization into n elementary factors is canonical and leads to a minimum of coefficients, for SISO systems $2n+1$, *i.e.*, n coefficients for the poles, n for the zeros, and one coefficient for the overall scaling.

The synthesis of passive transfer operators via cascade factorizations has a long history in classical network theory. The first results were concerned with the factorization of a lossless (inner) transfer operator of degree n into a product of n degree-1 lossless transfer operators, by recursively extracting a degree-1 subnetwork. This procedure is known as Darlington synthesis of lossless multiports [2], and produces ladder filters with well-known properties [3]. The use of a lossless (unitary) state realization of the inner operator gave the synthesis procedures by Youla and Tissi [4], while the synthesis of more general J-unitary operators was considered by Fettweis [5] in connection with wave-digital filters.

The cascade realization of inner operators leads to a realization procedure of any passive (contractive) rational transfer function, via a lossless embedding of the contractive transfer function $T(z)$. Thus, one obtains a realization of $T(z)$ in which either the poles or the zeros of $T(z)$ are localized in the elementary sections. State-space versions of this procedure are discussed in Roberts and Mullis' book [6].

Although it is more general, the Darlington synthesis procedure is closely connected to the Levinson algorithm, used in estimation filter theory of stationary stochastic processes [7]. The estimation filters are prediction (AR) filters with their transmission zeros at infinity, but the filter structure that is obtained is also a ladder filter which can be derived recursively from the covariance matrix of the stochastic process. The synthesis procedure thus constitutes a recursive Cholesky factorization of positive Toeplitz matrices. The Toeplitz matrices can be generalized to include the covariance matrices of more general α-stationary processes [8, 9], and leads to a generalized Schur parametrization of structured (α-stationary) matrices, *i.e.*, matrices with a low displacement rank [10]. The paper by Genin *et al.*[11] explored the relation between lossless state realizations and the characterization of structured matrices via a cascade of elementary lossless sections. Finally, there are many parallel results in operator theory: Potapov [12] obtained a complete description of (not necessarily rational) J-unitary and J-contractive matrix functions in terms of general cascade decompositions, while the lossless embedding and subsequent factorization of contractive functions in the setting of colligations was considered by Livsic and Brodskii

[13, 14]. The Darlington synthesis procedure is also closely connected, via the Lossless Inverse Scattering problem, to classical interpolation problems of the Nevanlinna-Pick type; see [7, 15, 16, 17].

Rather than a factorization of a lossless embedding of T, it is also possible to determine a direct factorization (9.1) [18, 19, 20, 21, 22]. Such factorizations realize both a zero and a pole of T in each elementary section, which makes them attractive in some applications, but they are also more complicated to derive. One can act directly on the transfer function $T(z)$, and in this case the complication is that non-square factors can occur [19], giving rise to a plethora of possible elementary sections. The situation is easier to describe in state-space terms. Let $T(z)$ be a bounded system, and suppose that it has a factorization $T = T_1 T_2$, where T_1, T_2 are again bounded systems, with minimal realizations $\mathbf{T}_1 = \{A_1, B_1, C_1, D_1\}, \mathbf{T}_2 = \{A_2, B_2, C_2, D_2\}$. A realization for T is thus given by

$$\mathbf{T} = \left[\begin{array}{cc|c} A_1 & & C_1 \\ & I & \\ \hline B_1 & & D_1 \end{array}\right] \left[\begin{array}{cc|c} I & & \\ & A_2 & C_2 \\ \hline & B_2 & D_2 \end{array}\right] = \left[\begin{array}{ccc|c} A_1 & C_1 B_2 & C_1 D_2 \\ 0 & A_2 & C_2 \\ \hline B_1 & D_1 B_2 & D_1 D_2 \end{array}\right]. \tag{9.2}$$

Note that $A = A_T$ is block upper triangular. If D_1 and D_2 are both invertible, then T^{-1} has a realization given by the product of the realizations of T_1^{-1} and T_2^{-1}, which turns out to have

$$A^\times = A_{T^{-1}} = \begin{bmatrix} A_1^\times & 0 \\ -C_2 D_2^{-1} D_1^{-1} B_1 & A_2^\times \end{bmatrix},$$

(where $A^\times := A - BD^{-1}C$ is the A-matrix of the inverse system, whose eigenvalues are the zeros of T). This matrix is block lower triangular. It can be shown, see *e.g.*, [23, 21] that T can be factorized minimally into factors T_1, T_2 if and only if it has a minimal realization $\mathbf{T}$ in which A_T is block upper triangular and $A_T^\times$ is block lower triangular. The factorization problem is thus reduced to finding a state-space transformation acting on a given realization of T and a partitioning into 2×2 blocks such that A_T and $A_T^\times$ have the required forms. To this end, one has either to determine the solutions of a certain Riccati equation (this replaces the Riccati equation that occurs in the embedding step), or to compute eigenvalue decompositions (Schur decompositions) of both A and $A^\times$, describing the poles and zeros of the given transfer function. However, in the subsequent factorization procedure, the conditioning of certain inverses can be problematic [23]. Such problems do not occur with the factorization of inner or J-inner functions, as in this case the poles of the system also determine the zeros: for inner functions Σ with unitary realizations, $\mathbf{\Sigma}^*$ is a realization of $\Sigma^{-1} = \Sigma^*$, and hence $A^\times = A^*$. We only consider the cascade realization of inner functions Σ from now on.

Repetition of the above factorization into two systems leads to a factorization of a degree-n system into n systems of degree 1: the elementary sections. A particular realization of the elementary sections produces orthogonal digital filters. Here, the elementary operator is not a multiplication, but a plane rotation, where the rotation angle is the coefficient of the

section. The advantage that such filters have is that (with ideal rotors) they are inherently lossless and stable, even if the coefficients are imprecise, and that no limit cycles or overflow oscillations can occur. Another advantage is that the filters are typically cascade arrays of identical processors with only nearest neighbor connections, which allows for VLSI implementation. Some other issues to be considered are the pipelinability and computability of the array, which are not always guaranteed. A number of orthogonal filter structures are possible, depending on the precise factorization of the inner transfer operator, and on whether a factorization of Σ, or its associated J-unitary operator Θ is performed. The factorization can also be done directly on the transfer function $T(z)$, if it is specified as a ratio of two polynomials, or on the state-space matrices. In both cases, a preliminary embedding step is necessary. The main reference papers on orthogonal filter realizations are by Deprettere, Dewilde, P. Rao and Nouta [24, 25, 26, 27], S.K. Rao and Kailath [28], Vaidyanathan [29], Regalia, Mitra and Vaidyanathan [30], and Roberts and Mullis' book [6]. More recent references are [31, 32].

Givens rotations

We say that $\hat{\Sigma}$ is an elementary orthogonal rotation if $\hat{\Sigma}$ is a 2×2 unitary matrix (with scalar entries) of the form

$$\hat{\Sigma} = \begin{bmatrix} c^* & s \\ -s^* & c \end{bmatrix}, \tag{9.3}$$

with $c^*c + s^*s = 1$. An important property of elementary rotations is that they can be used to zero a selected entry of a given operator: for given a and b, an elementary orthogonal rotation $\hat{\Sigma}$ exists such that

$$\hat{\Sigma}^* \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a' \\ 0 \end{bmatrix}, \tag{9.4}$$

i.e., such that $s^*a + c^*b = 0$ and $a' = (a^*a + b^*b)^{1/2}$. In this case, $\hat{\Sigma}$ is called a Givens rotation, and we write $\hat{\Sigma} = \text{givens}[a; b]$ in algorithms. Givens rotations are used to factor a given state realization into elementary rotations, or certain generic groups of rotations called elementary sections. Acting on state realizations, the 2×2 elementary rotation matrix is typically extended by identity matrices, say

$$\mathbf{\Sigma}_i = \left[\begin{array}{ccc|c} I_{i-1} & & & \\ & \times & & \times \\ & & I_{n-i} & \\ \hline & \times & & \times \end{array}\right], \tag{9.5}$$

where the four '$\times$'-s together form the 2×2 unitary matrix. We use a hat symbol to denote this elementary 2×2-matrix, *i.e.*, we write it as $\hat{\mathbf{\Sigma}}_i$.

An elementary J-unitary rotation $\hat{\Theta}$ can be obtained from $\hat{\Sigma}$ in (9.3) if $c \neq 0$ as

$$\hat{\Theta} = \begin{bmatrix} 1 & -s \\ -s^* & 1 \end{bmatrix} \frac{1}{c}$$

It can be used to zero entries of vectors:

$$\hat{\Theta}^{-1}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a' \\ 0 \end{bmatrix},$$

only if $a^*a - b^*b = a'^*a' > 0$, by setting $s^* = -b/a$.

Orthogonal digital filter synthesis

Assume that Σ is known, along with a unitary realization $\mathbf{\Sigma}$. As was shown in equation (9.2), a necessary condition for factorization of Σ is that A_Σ is upper triangular. From the given realization, this can be ensured via a unitary state-space transformation Q obtained from a Schur decomposition of the given A-matrix:

$$Q A_\Sigma Q^* = R,$$

where R is upper triangular. This decomposition always exists, and amounts to a computation of the poles of the system. With A_Σ upper triangular, the second phase of the factorization procedure is the factoring of $\mathbf{\Sigma}$ into a minimal number of elementary (degree-1) factors. Here, one makes use of the fact that the product of two unitary matrices is again unitary. A consequence of this fact is that, in equation (9.2) (where all matrices are unitary now), any $\mathbf{\Sigma}_1$ such that $\mathbf{\Sigma}_1^*\mathbf{\Sigma}$ has zero block entries (2, 1) and (3, 1) leads to $\mathbf{\Sigma}_2$ of the required form. Since the (2, 1) entry is already equal to zero, it follows that $\mathbf{\Sigma}_1$ can be of the form indicated in (9.2): using $\mathbf{\Sigma}_1^*$, one only has to cancel entry (3, 1) using entry (1, 1). The unitarity of the product $\mathbf{\Sigma}_1^*\mathbf{\Sigma}$ ensures that also its entries (1, 2) and (1, 3) are zero. Upon factoring $\mathbf{\Sigma}$ down to the scalar level, it follows that the elementary unitary factors have the form $\mathbf{\Sigma}_i$ in (9.5). If Σ is of degree n, then the factorization consists of n degree-1 factors and is of the form $\mathbf{\Sigma} = \mathbf{\Sigma}_1 \cdots \mathbf{\Sigma}_n$, where

$$\mathbf{\Sigma} = \left[\begin{array}{cccc} a_{11} & \times & \times & \times \\ & a_{22} & \times & \times \\ & & a_{nn} & \times \\ \hline \times & \times & \times & \times \end{array}\right] = \left[\begin{array}{ccc|c} a_{11} & & & \times \\ & 1 & & \\ & & 1 & \\ \hline \times & & & \times \end{array}\right] \left[\begin{array}{ccc|c} 1 & & & \\ & a_{22} & & \times \\ & & 1 & \\ \hline & \times & & \times \end{array}\right] \cdots \left[\begin{array}{ccc|c} 1 & & & \\ & 1 & & \\ & & a_{nn} & \times \\ \hline & & \times & \times \end{array}\right]. \tag{9.6}$$

The a_{ii} are the diagonal entries of A_Σ, which are the poles of the system. Hence, each elementary section realizes a pole of Σ. In (9.6), we assumed that Σ is a SISO system. For multi-input multi-output systems, the procedure is an extension of the above, and gives (for an example of a system with two inputs and two outputs)

$$\mathbf{\Sigma} = \left[\mathbf{\Sigma}_{1,1}\mathbf{\Sigma}_{1,2}\right] \cdot \left[\mathbf{\Sigma}_{2,1}\mathbf{\Sigma}_{2,2}\right] \cdots \left[\mathbf{\Sigma}_{n,1}\mathbf{\Sigma}_{n,2}\right] \mathbf{\Sigma}' \tag{9.7}$$

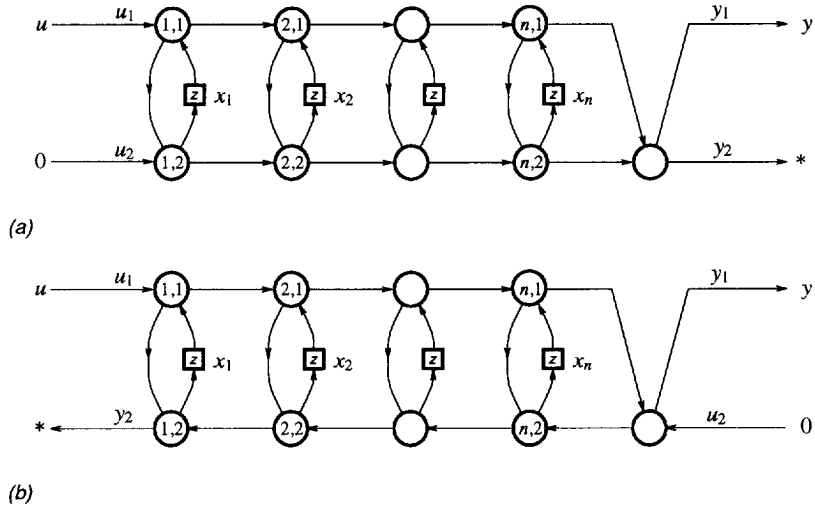


Figure 9.1. (*a*) Σ-based cascade factorization, based on a Schur decomposition of A_Σ. Σ is a unitary embedding of $T: u \to y$ which is the transfer of u_1 to y_1 if $u_2 = 0$. (*b*) Θ-based cascade factorization, based on a Schur decomposition of A_Θ, where Θ is the J-unitary chain scattering operator associated to Σ.

$$= \left[\begin{array}{cc|c} \times & & \times \\ & 1 & \\ & & 1 \\ \hline \times & & \times \\ & & & 1 \end{array}\right] \left[\begin{array}{cc|c} \times & & \times \\ & 1 & \\ & & 1 \\ \hline & \times & & 1 \\ \times & & & \times \end{array}\right] \left[\begin{array}{cc|c} 1 & & \\ & \times & \times \\ & & 1 \\ \hline & \times & \times \\ & & & 1 \end{array}\right] \left[\begin{array}{cc|c} 1 & & \\ & \times & \times \\ & & 1 \\ \hline & & & 1 \\ & \times & & \times \end{array}\right] \cdots$$

$$\cdots \left[\begin{array}{cc|c} 1 & & \\ & 1 & \\ & & \times & \times \\ \hline & \times & \times \\ & & & 1 \end{array}\right] \left[\begin{array}{cc|c} 1 & & \\ & 1 & \\ & & \times & \times \\ \hline & & & 1 \\ & \times & & \times \end{array}\right] \left[\begin{array}{cc|c} 1 & & \\ & 1 & \\ & & 1 & \\ \hline & & & \times \ \times \\ & & & \times \ \times \end{array}\right].$$

$\mathbf{\Sigma}'$ is the terminating section of degree 0. It is in general a unitary matrix itself, which can also be factored into elementary Givens rotations, and finally a unit-norm scaling. The network structure that is obtained is drawn in figure 9.1, which is straightforwardly derived from (9.7) by considering how a vector $[x_1 \ x_2 \ \cdots x_n \ u_1 \ u_2]$ is transformed in elementary steps to $[x_1' \ x_2' \ \cdots x_n' \ y_1 \ y_2]$. The network is pipelinable: the signal flow is strictly unidirectional (from the left to the right). It is also computable: given the current values of the inputs and of the states, the outputs and the next states can be computed. The network is specified by a minimal number of $2n + 1$ coefficients (rotation angles). Any strictly contractive LTI system T can be realized in this way, by embedding T into

an inner system Σ such that $T = \Sigma_{11}$. As a matter of fact, it is not necessary to compute the embedding completely: if Σ has a realization as in (7.8), viz.

$$\mathbf{\Sigma} = \begin{bmatrix} R & & \\ & I & \\ & & I \end{bmatrix} \left[\begin{array}{ccc} A & C & C_2 \\ B & D & D_{12} \\ B_2 & D_{21} & D_{22} \end{array}\right] \begin{bmatrix} R^{-1} & & \\ & I & \\ & & I \end{bmatrix},$$

where $\{A, B, C, D\}$ is the given realization of T, and R, B_2, D_{22} are computed via a Riccati equation, then only A, B and B_2 determine the factors $\mathbf{\Sigma}_{i,j}$ ($i = 1, \cdots, n$, $j = 1, 2$), and C_2, D_{12} and D_{22} are not needed. As far as the cascade factorization is concerned, it is even possible to omit the state transformation by R [33], although this is at the expense of a number of other matrix inversions, and we still have to compute R to determine the extension by B_2, D_{21} anyway. As an alternative to the above factorization of Σ, one can convert Σ to a J-unitary Θ operator with realization $\mathbf{\Theta}$ (cf theorem 5.2), factor $\mathbf{\Theta}$ in a comparable way as done for $\mathbf{\Sigma}$, and convert the factors back to the scattering domain. This gives network structures as depicted in figure 9.1(*b*).

In the above two solutions to the factorization problem, the trick to determine a minimal factorization was to compute a Schur decomposition of A_Σ (or A_Θ), which introduced as many zero entries in $\mathbf{\Sigma}$ as possible. The remaining $2n + 1$ non-zero entries below the main diagonal of $\mathbf{\Sigma}$ induced a factorization of $\mathbf{\Sigma}$ into $2n + 1$ elementary factors. There are other structures of $\mathbf{\Sigma}$, not requiring an (expensive) Schur decomposition step, which still result in a factorization of $\mathbf{\Sigma}$ into $2n + 1$ elementary factors. However, this time we do not obtain a factorization of Σ itself into a product $\Sigma_1 \cdots \Sigma_n$, so that the individual elementary sections do not realize poles and zeros of Σ, and the implementation is not truly a cascade factorization in the sense used before. One possible structure that can be obtained via a unitary state transformation is a Hessenberg structure of A and the first row of B, which can be computed non-recursively:

$$\mathbf{\Sigma} = \left[\begin{array}{cccc|cc} \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times \\ \hline & & & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \end{array}\right]$$

$\mathbf{\Sigma}$ can be factored for example by acting on its columns, zeroing entries of rows $n + 1$ till row 1 in turn. In that case, the first two steps are

$$\mathbf{\Sigma} = \left[\begin{array}{cccc|cc} \times & \times & \times & 0 & \times & \times \\ \times & \times & \times & 0 & \times & \times \\ & \times & \times & 0 & \times & \times \\ & & \times & 0 & \times & \times \\ \hline & & & 1 & 0 & 0 \\ \times & \times & \times & 0 & \times & \times \end{array}\right] \left[\begin{array}{cccc|cc} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \times & \times & \\ \hline & & & \times & \times & \\ & & & & & 1 \end{array}\right] \left[\begin{array}{cccc|cc} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & \times & & \times \\ \hline & & & & 1 & \\ & & & \times & & \times \end{array}\right].$$

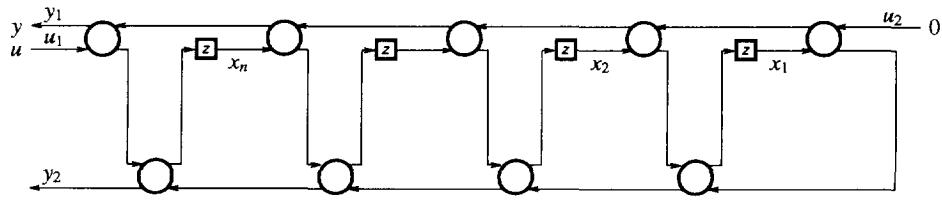


Figure 9.2. Hessenberg lossless filter structure.

The factorization continues in a similar fashion, and the resulting network structure is as depicted in figure 9.2 (viz. [28; 27, 32]). The network is again pipelinable. With more work one can also act on realizations that are not normalized to be unitary, and again, only a partial solution of the embedding has to be computed, since the last row of $\mathbf{\Sigma}$ is not used in the factorization. In particular, if $T(z)$ is given as a ratio

$$T(z) = \frac{B(z)}{A(z)} = \frac{b_0 + b_1 z + \cdots + b_{n-1} z^{n-1}}{1 + a_1 z + \cdots + a_{n-1} z^{n-1}}$$

and an extension $T_e(z) = B_e(z)/A(z)$ is computed such that $[T \;\; T_e]$ is an isometry, then a non-unitary realization of $\mathbf{\Sigma}$ in Hessenberg form can directly be determined as the companion form

$$\mathbf{\Sigma} = \left[\begin{array}{cccc|cc} & & & -a_4 & b_4 - a_4 b_0 & b_{e,4} - a_4 b_{e,0} \\ 1 & 0 & & -a_3 & b_3 - a_3 b_0 & b_{e,3} - a_3 b_{e,0} \\ & 1 & 0 & -a_2 & b_2 - a_2 b_0 & b_{e,2} - a_2 b_{e,0} \\ & & 1 & -a_1 & b_1 - a_1 b_0 & b_{e,1} - a_1 b_{e,0} \\ \hline & & & 1 & b_0 & b_{e,0} \\ \times & \times & \times & \times & \times & \times \end{array}\right].$$

The cascade factorization can be computed from a factorization of $\mathbf{\Sigma}$. This yields the same network as obtained earlier by S.K. Rao and Kailath [28], who derived a simple and straightforward method for computing the factorization, based on the application of Schur's algorithm to an array containing the coefficients of the polynomials $A(z)$, $B(z)$ and $B_e(z)$.

9.2 TIME-VARYING Σ-BASED CASCADE FACTORIZATION

In this section, we extend the cascade factorization results based on realizations of Σ in Schur form to the context of time-varying systems. The network synthesis procedure is roughly the same two-stage algorithm as for the time-invariant case:

1. Using unitary state transformations, bring $\mathbf{\Sigma}$ into a form that allows a minimal factorization (*i.e.*, a minimal number of degree-1 factors). We choose to make the A matrix of $\mathbf{\Sigma}$ upper triangular. This leads to a QR iteration on the $\{A_k\}$ and is the equivalent of the Schur decomposition (eigenvalue computations) of A that would be required for time-invariant systems.

2. Using Givens rotations extended by I to the correct size, factor $\mathbf{\Sigma}$ into a product of such elementary sections. From this factorization, the lossless cascade network follows directly.

For time-invariant systems, cascade factorizations based on a state transformation to Hessenberg form have been considered to avoid eigenvalue computations. In the time-varying setting, eigenvalue computations are in a natural way replaced by recursions consisting of QR factorizations, so this seems no longer to be an issue. The actual factorization (step 2) is similar to the time-invariant procedure, and can be carried out locally. The main difference is that for time-varying systems, the dimensions of the state-space matrices need not be constant, and a distinction has to be made between shrinking and growing state-space dimensions. We first derive a global procedure for factorization into two lossless factors, then look at the stages that are produced, and finally consider the factorization into elementary sections of local degree at most equal to 1. It is shown that it is still possible to obtain a factorization $\Sigma = \Sigma_1 \cdots \Sigma_n \Sigma'$, where n is the maximal local state dimension over all stages, and each Σ_i is a section of local degree at most equal to 1. In a sense, the result is evident: by adding extra inputs and outputs, it is possible to expand the realization of Σ to a non-minimal realization which has n states at each point. However, the theorem is more specific: the local state dimensions of the factors add up to the local degree of Σ, and we obtain a cascade network with a minimal number of coefficients as well.

Time-varying 'Schur decomposition'

Let be given a unitary realization $\mathbf{\Sigma}$ of a locally finite inner operator Σ. Let $A = A_\Sigma \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ be the A operator of $\mathbf{\Sigma}$. The first step in the factorization algorithm is to find a locally square unitary state transformation $Q \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ such that

$$QAQ^{(-1)*} = R, \tag{9.8}$$

where $R \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ has R_k upper triangular. If A_k is not square, say of size $d_k \times d_{k+1}$, then R_k will be of the same size and also be rectangular. In this case, 'upper triangular' is to be made more precise: it means $(R_k)_{i,j} = 0$ for $i > j + (d_k - d_{k+1})$ (figure 9.3). In the case where $d_{k+1} > d_k$ (figure 9.3(*c*)), and if the increase in the number of states is 2 or more, it is possible to introduce extra zero entries in B too, as indicated in the figure. These play a role later in this chapter. In the time-invariant case, expression (9.8) would read $QAQ^* = R$, and the solution is then precisely the Schur decomposition of A. In this context, the main diagonal of A consists of its eigenvalues, which are the (inverses of the)

poles of the system. In the present context, relation (9.8) is effectively the (unshifted) QR iteration algorithm that is sometimes used to compute the eigenvalues of $A = A_k$, if all A_k are the same [34]. The iteration (or rather recursion) is obtained by expanding the diagonal relation into its entries: $Q_k A_k Q_{k+1}^* = R_k$, or

$$\begin{array}{lll} \vdots & & \\ Q_1 A_1 =: R_1 Q_2 & \rightarrow Q_2, R_1 & \\ Q_2 A_2 =: R_2 Q_3 & \rightarrow Q_3, R_2 & \\ Q_3 A_3 =: R_3 Q_4 & & \\ \vdots & & \end{array} \tag{9.9}$$

Each step in the computation amounts to a multiplication by the previously computed Q_k, followed by a QR factorization of the result, yielding Q_{k+1} and R_k. Given an initial Q_{k_0}, *e.g.*, $Q_{k_0} = I$, the above recursion can be carried out in two directions, both forward and backward in time. For example, take $k_0 = 1$, then the forward recursion is given by (9.9), while the backward decomposition is

$$\begin{array}{lll} A_0 Q_1^* = Q_0^* R_0 & \rightarrow Q_0, R_0 & \\ A_{-1} Q_0^* = Q_{-1}^* R_{-1} & \rightarrow Q_{-1}, R_{-1} & \\ A_{-2} Q_{-1}^* = Q_{-2}^* R_{-2} & & \\ \vdots & & \end{array}$$

Since we can start at any k_0 with any unitary Q_{k_0}, the decomposition (9.8) is not unique, although it always exists. For later reference, we formulate this result in the following proposition.

PROPOSITION 9.1. *Let $A \in \mathcal{D}(\mathcal{B}, \mathcal{B}^{(-1)})$ be locally finite. Then a unitary state transformation $Q \in \mathcal{D}(\mathcal{B}, \mathcal{B})$ exist such that $QAQ^{(-1)*} = R$ is a diagonal operator with all R_k is upper triangular: if A_k has size $d_k \times d_{k+1}$, then $(R_k)_{i,j} = 0$ for $i > j + (d_k - d_{k+1})$.*

PROOF The proof has been given in the text preceding the proposition. □

In the context of finite upper triangular matrices whose state realization starts with 0 states at instant $k = 1$, we can take as initial transformation $Q_1 = [\,\cdot\,]$. If the A_k are equal to each other, then the above recursion is precisely the (unshifted) QR iteration for computing the eigenvalues (or Schur decomposition) of A. It is known (see [34]) that the unshifted QR iteration will converge if the absolute values of the eigenvalues of A are unequal to each other, and that the rate of convergence is dependent on the smallest ratio between those absolute eigenvalues. For periodically time-varying systems, with period n say, an initial state transformation Q_1 such that $Q_k = Q_{n+k}$ is also periodical can be computed by considering the conjunction of n consecutive stages. Writing $A_p = A_1 A_2 \cdots A_n$, the

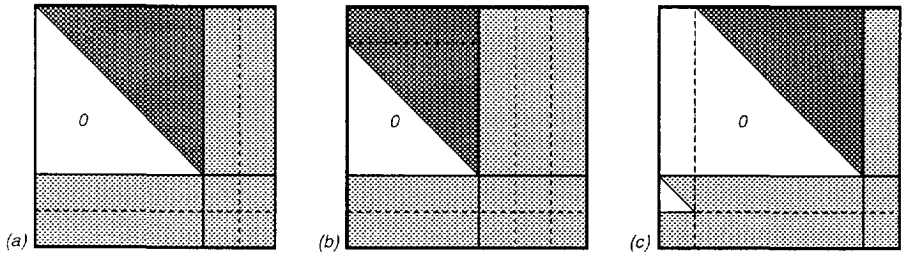


Figure 9.3. Schur forms of $\mathbf{\Sigma}_k$. (*a*) Constant state dimension, (*b*) shrinking state dimension, (*c*) growing state dimension.

Schur decomposition of A_p ($Q_1 A_p Q_1^* = R_p$) gives Q_1, while (9.8) gives $Q_2, \cdots, Q_n$ in turn. Recent investigations show that one can compute the Schur decomposition of a product of matrices without ever explicitly evaluating the product [35]. The procedure is called the periodic QR algorithm, and consists basically of an implicit shifted QR algorithm acting over a sequence of matrices, rather than just one matrix. It starts with a preliminary step, in which $A_1, \cdots, A_{n-1}$ are made upper triangular, and A_n is put in Hessenberg form (upper triangular, plus one diagonal below the main diagonal). This step consists of QR factorizations only. In subsequent steps, a shifted QR technique is applied, where the shift is computed from the product of the lower right 2×2 submatrices of the A_i, which is an approximation of one of the eigenvalues of A_p. The use of a shift greatly increases the convergence rate of the algorithm, and after some iterations one of the below-diagonal entries of A_n is reduced to zero. At this point, the problem deflates to one of lower dimensions.

Factorization into two factors

The factorization result (equation (9.2)), which stated that a time-invariant rational transfer operator T has a factorization $T = T_1 T_2$ if and only if its realization has a certain structure, admits a straightforward generalization to time-varying inner systems.

PROPOSITION 9.2. *Let $\mathbf{\Sigma} \in \mathcal{D}(\mathcal{B} \times \mathcal{M}, \mathcal{B}^{(-1)} \times \mathcal{N})$ be unitary, with locally finite dimensions, and have a block partitioning as*

$$\mathbf{\Sigma} = \left[\begin{array}{cc|c} A_{11} & A_{12} & C_1 \\ 0 & A_{22} & C_2 \\ \hline B_1 & B_2 & D \end{array}\right] \tag{9.10}$$

where $A_{11} \in \mathcal{D}(\mathcal{B}_1, \mathcal{B}_1^{(-1)})$ for some state-space sequence $\mathcal{B}_1 \subset \mathcal{B}$. Define the space sequences $\mathcal{N}_1$ and $\mathcal{B}_2$ by the relations $\mathcal{B}_1 \times \mathcal{M} = \mathcal{B}_1^{(-1)} \times \mathcal{N}_1$, and $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$.

1. *Then unitary operators* $\hat{\mathbf{\Sigma}}_1, \hat{\mathbf{\Sigma}}_2$ *exist, with* $\hat{\mathbf{\Sigma}}_1 = \{A_{11}, B_1, C_1', D_1\} \in \mathcal{D}(\mathcal{B}_1 \times \mathcal{M}, \mathcal{B}_1^{(-1)} \times \mathcal{N}_1)$, $\hat{\mathbf{\Sigma}}_2 = \{A_{22}, B_2', C_2, D_2\} \in \mathcal{D}(\mathcal{B}_2 \times \mathcal{N}_1, \mathcal{B}_2^{(-1)} \times \mathcal{N})$, *such that*

$$\mathbf{\Sigma} = \left[\begin{array}{cc|c} A_{11} & & C_1' \\ & I & \\ \hline B_1 & & D_1 \end{array}\right] \left[\begin{array}{c|cc} I & & \\ & A_{22} & C_2 \\ \hline & B_2' & D_2 \end{array}\right] =: \mathbf{\Sigma}_1 \mathbf{\Sigma}_2 . \tag{9.11}$$

2. *If* Σ *is an inner operator with unitary realization* $\mathbf{\Sigma}$ *of the form (9.10), with* $\ell_{A_\Sigma} < 1$, *then* $\Sigma = \Sigma_1 \Sigma_2$, *where* Σ_1, Σ_2 *are inner operators with unitary realizations given by* $\hat{\mathbf{\Sigma}}_1, \hat{\mathbf{\Sigma}}_2$, *with* $\ell_{A_{11}} < 1$, $\ell_{A_{22}} < 1$. *The sequence of state dimensions of* Σ_1, Σ_2 *add up to the sequence of state dimensions of* Σ: *the factorization is minimal.*

PROOF

1. Consider $[A_{11}^* \;\; B_1^*]^*$. It is an isometry in $\mathcal{D}$ because $A_{11}^* A_{11} + B_1^* B_1 = I$. Choose $C_1', D_1 \in \mathcal{D}$ such that, for each point k,

$$(\mathbf{\Sigma}_1)_k = \left[\begin{array}{cc} (A_{11})_k & (C_1')_k \\ (B_1)_k & (D_1)_k \end{array}\right]$$

   is a unitary matrix. Then $\mathbf{\Sigma}_1$ is a unitary operator in $\mathcal{D}$ as required, and the number of added outputs is $\#(\mathcal{N}_1) = \#(\mathcal{B}_1) - \#(\mathcal{B}_1^{(-1)}) + \#(\mathcal{M})$. Because $[A_{11}^* \; 0 \; B_1^*]^*$ is also the first column of $\mathbf{\Sigma}$, it directly follows that $\mathbf{\Sigma}_1^* \mathbf{\Sigma} = \mathbf{\Sigma}_2$ has the form specified in (9.10).

2. The fact $\ell_{A_\Sigma} < 1 \;\Rightarrow \ell_{A_{11}} < 1, \ell_{A_{22}} < 1$ is straightforward to show. With $\ell_{A_{11}} < 1, \ell_{A_{22}} < 1$, the unitary realizations $\hat{\mathbf{\Sigma}}_1$, $\hat{\mathbf{\Sigma}}_2$ define inner operators Σ_1, Σ_2 (theorem 4.6). The cascade $\Sigma_1 \Sigma_2$ has a realization $\mathbf{\Sigma}_1 \mathbf{\Sigma}_2 = \mathbf{\Sigma}$ as in (9.10), and hence $\Sigma = \Sigma_1 \Sigma_2$. The factorization is minimal because (with $\ell_A < 1$) $\hat{\mathbf{\Sigma}}_1$, $\hat{\mathbf{\Sigma}}_2$ are minimal realizations, whose degrees add up to the degree of $\mathbf{\Sigma}$. □

Some remarks are apposite here. First note that if $\ell_{A_\Sigma} = 1$, and $\mathbf{\Sigma}$ is a unitary realization with controllability and observability Gramians equal to the identity, then $\hat{\mathbf{\Sigma}}_1$ inherits the fact that the controllability Gramian is I, but if $\ell_{A_{11}} = 1$, then nothing can be said, at first sight, of its observability Gramian, and hence the fact that Σ_1 is inner is not proven in this case. Second, note that all computations can be carried out locally (separately) for each stage k. The state dimension sequence $\mathcal{B}_1$ determines the degree of the factors, and also the number of outputs (inputs) of Σ_1 (Σ_2). The choice of $\mathcal{B}_1$ is restricted by the required form of (9.10), *i.e.*, the fact that $A_{21} = 0$.

The above proposition can be formulated in a different way that provides some additional (more fundamental) insight.

PROPOSITION 9.3. *Let Σ be a locally finite inner operator. Then*

$$\Sigma = \Sigma_1\Sigma_2 \quad \Rightarrow \quad \mathcal{H}(\Sigma) = \mathcal{H}(\Sigma_1) \oplus \mathcal{H}(\Sigma_2)\Sigma_1^*,$$

where Σ_1 and Σ_2 are inner operators. Conversely, let Σ_1 be an inner operator, then

$$\mathcal{H}(\Sigma_1) \subset \mathcal{H}(\Sigma) \quad \Rightarrow \quad \Sigma = \Sigma_1\Sigma_2,$$

where Σ_2 is an inner operator.

PROOF For an inner operator Σ_2, we have that $\mathcal{U}_2\Sigma_2^* = \mathcal{U}_2 \oplus \mathcal{H}(\Sigma_2)$ (proposition 4.1). Consequently, $\mathcal{U}_2\Sigma^* = \mathcal{U}_2\Sigma_1^* \oplus \mathcal{H}(\Sigma_2)\Sigma_1^*$, and because $\Sigma_1^* \in \mathcal{L}$,

$$\begin{aligned}\mathcal{H}(\Sigma) &= \mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{U}_2\Sigma^*)\\ &= \mathcal{H}(\Sigma_1) \oplus \mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{H}(\Sigma_2)\Sigma_1^*)\\ &= \mathcal{H}(\Sigma_1) \oplus \mathcal{H}(\Sigma_2)\Sigma_1^*.\end{aligned}$$

Conversely, the fact that $\Sigma_2 = \Sigma_1^*\Sigma$ is a unitary operator is clear, and we have to show that it is in fact upper. Indeed, since $\Sigma \in \mathcal{U}$,

$$\begin{aligned}\mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{U}_2\Sigma_2) &= \mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{U}_2\Sigma_1^*\Sigma)\\ &= \mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{H}(\Sigma_1)\Sigma)\\ &\subset \mathbf{P}_{\ell_2 Z^{-1}}(\mathcal{H}(\Sigma)\Sigma) \quad = 0 \qquad \text{[prop. 4.1]}\end{aligned}$$

so that the lower triangular part of Σ_2 is zero. □

Hence, in order to obtain a factorization of Σ, we can select any inner Σ_1 such that $\mathcal{H}(\Sigma_1) \subset \mathcal{H}(\Sigma)$. A suitable Σ_1 is again obtained from equation (9.10): a minimal realization based on A_{11} and B_1 has $\mathcal{H}(\Sigma_1) = \mathcal{D}_2\left[B_1Z(I-A_{11}Z)^{-1}\right]^* = [\mathcal{D}_2 \;\; 0]\left[[B_1 \;\; B_2]Z(I-AZ)^{-1}\right]^*$ because $A_{21} = 0$, so that indeed $\mathcal{H}(\Sigma_1) \subset \mathcal{H}(\Sigma)$. Σ_1 is obtained, as in the proof of proposition 9.2, by extending $[A_{11}^* \;\; B_1^*]^*$ to a unitary state-space operator. Special cases occur if $(\mathcal{B}_1)_k = 0$ for some k, although the propositions remains valid. The following two situations are typical.

- If $\#(\mathcal{B}_1)_{k+1} = 0$, with $\#(\mathcal{B}_1)_k = d \geq 0$, then $(A_{11})_k$ is a $(d\times 0)$-matrix. In this case, $\mathbf{\Sigma}_k$ has the form

$$\mathbf{\Sigma}_k = \left[\begin{array}{cc|c} \cdot & A_{12} & C_1 \\ \cdot & A_{22} & C_2 \\ \hline \cdot & B_2 & D \end{array}\right]$$

(as before, '·' stands for an entry of zero dimensions) so that

$$(\hat{\mathbf{\Sigma}}_1)_k = \left[\begin{array}{c|cc} \cdot & I_d & 0 \\ \hline \cdot & 0 & I \end{array}\right], \qquad (\mathbf{\Sigma}_1)_k = \left[\begin{array}{cc|cc} \cdot & 0 & I_d & 0 \\ \cdot & I & 0 & 0 \\ \hline \cdot & 0 & 0 & I \end{array}\right].$$

$\mathbf{\Sigma}_1$ is a trivial state-space operator mapping its first d states to d outputs. If $d = 0$, then $(\mathbf{\Sigma}_1)_k = I$.

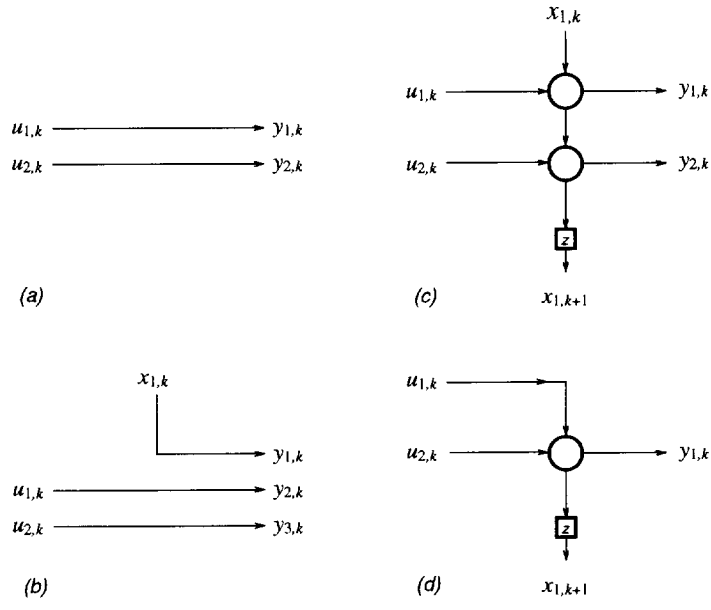


Figure 9.4. Elementary sections in a stage. (*a*) $C(0)$ constant section with zero states, (*b*) S section, going from 1 state to 0, (*c*) $C(1)$ section with a constant number of 1 states, (*d*) G section, going from 0 to 1 state. The number of inputs/outputs have arbitrarily been set to 2.

- If $\#(\mathcal{B}_1)_k = 0$, $\#(\mathcal{B}_1)_{k+1} = d \geq 0$, then $(\hat{\mathbf{\Sigma}}_1)_k$ is obtained as the extension of $(B_1)_k$ to a unitary matrix:

$$(\hat{\mathbf{\Sigma}}_1)_k = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline (B_1)_k & (D_1)_k \end{array}\right].$$

  Note that this case can only happen if $(A_\Sigma)_k$ has its first d columns equal to zero:

$$(A_\Sigma)_k = \begin{bmatrix} \cdot & \cdot \\ 0 & (A_{22})_k \end{bmatrix},$$

  that is, in view of figure 9.3, this can only happen at points where the state dimension of Σ grows with at least d states.

Elementary lossless stage sections

We apply proposition 9.2 to the most elementary type of state dimension sequence $\mathcal{B}_1$: $\mathcal{B}_1$ with entries having dimensions $\#(\mathcal{B}_1)_k \in \{0, 1\}$. In a later section, we discuss the

choice of $\mathcal{B}_1$; here, we consider the factorization of a single stage of $\boldsymbol{\Sigma}$, and pay detailed attention to the fact that input/output and state dimensions can be time varying. With a partitioning of $\boldsymbol{\Sigma}$ as before in (9.10), a factor $\hat{\boldsymbol{\Sigma}}_1$ of $\boldsymbol{\Sigma}$ is determined by finding a unitary extension of the matrices $(A_{11})_k$ and $(B_1)_k$. The purpose of this section is to show how an extension can be obtained in factored form using elementary Givens rotations. With $\#(\mathcal{B}_1)_k \in \{0, 1\}$ and $\#(\mathcal{B}_1)_{k+1} \in \{0, 1\}$, the submatrix $(A_{11})_k$ can have only the following sizes:

$$\begin{cases} C(0): & 0\times 0, & S: & 1\times 0, \\ C(1): & 1\times 1, & G: & 0\times 1. \end{cases}$$

The cases $C(0)$ and $C(1)$ describe sections with a constant state dimension, while G, S stand for sections with growing and shrinking state dimensions, respectively. We discuss these sections in turn.

$C(0)$: $(\hat{\boldsymbol{\Sigma}}_1)_k$ has the form $(\hat{\boldsymbol{\Sigma}}_1)_k = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline \cdot & I \end{array}\right]$. See figure 9.4($a$). Obviously, a $C(0)$ section can always be extracted, but doing so does not lead to a degree reduction. Nonetheless, it plays a role as padding section in the description of a global factorization of Σ into a constant number of sections, later in this chapter.

S: $(\hat{\boldsymbol{\Sigma}}_1)_k$ has the form $(\hat{\boldsymbol{\Sigma}}_1)_k = \left[\begin{array}{c|cc} \cdot & 1 & 0 \\ \hline \cdot & 0 & I \end{array}\right]$. See figure 9.4($b$).

$C(1)$: Let $a = (A_{11})_k$, and suppose that Σ has n inputs at point k, so that $b = (B_1)_k$ is an $n \times 1$ vector. Then $(\hat{\boldsymbol{\Sigma}}_1)_k$ is a unitary extension of the vector $[a^* \; b_1^* \; \cdots \; b_n^*]^*$. Of the many possible extensions, one that results in a minimal number of coefficients is obtained using Givens rotations, which gives the extension directly in factored form:

$$\begin{aligned} (\hat{\boldsymbol{\Sigma}}_1)_k &= (\hat{\boldsymbol{\Sigma}}_1)_{1,k} \cdots (\hat{\boldsymbol{\Sigma}}_1)_{n,k} \qquad (9.12) \\ &= \left[\begin{array}{c|cc} c_1^* & -s_1 & \\ \hline s_1^* & c_1 & \\ & & \begin{array}{cc} 1 & \\ & 1 \end{array} \end{array}\right] \left[\begin{array}{c|cc} c_2^* & & -s_2 \\ \hline & \begin{array}{c} 1 \end{array} & \\ s_2^* & & c_2 \\ & & \begin{array}{c} 1 \end{array} \end{array}\right] \cdots \left[\begin{array}{c|cc} c_n^* & & -s_n \\ \hline & \begin{array}{c} 1 \end{array} & \\ & & \begin{array}{cc} 1 & \\ & c_n \end{array} \\ s_n^* & & \end{array}\right] \end{aligned}$$

where $(\hat{\boldsymbol{\Sigma}}_1)_{i,k}$ is used to zero the $(i+1)$-st entry of the vector $(\hat{\boldsymbol{\Sigma}}_1)_{i-1,k}^* \cdots (\hat{\boldsymbol{\Sigma}}_1)_{1,k}^* \left[\begin{smallmatrix} a \\ b \end{smallmatrix}\right]$ against the first entry. The computational structure (for $n = 2$) is shown in figure 9.4(c).

G: In this case, $(A_{11})_k = [\,\cdot\,]$, and $(\hat{\boldsymbol{\Sigma}}_1)_k$ is a unitary extension of the vector $b = (B_1)_k$. Again, the extension can be found in factored form, now requiring $n-1$ Givens rotations. See figure 9.4(d).

The four types of elementary stage sections in figure 9.4 form the building blocks of the cascade network realizations based on the Schur form. General structures are obtained

by connecting these sections horizontally (realizing a single stage in factored form) and vertically (realizing an elementary degree-1 factor of Σ). These are discussed in turn.

Structure of a factored lossless stage

A single stage $\mathbf{\Sigma}_k$ of $\mathbf{\Sigma}$ has, after transformation to Schur form, one of the three structures displayed in figure 9.3, depending on whether the state dimension of $\mathbf{\Sigma}$ is constant, shrinking or growing at point k. The factorization procedure is to recursively extract elementary factors of one of the above types. If the state dimension of Σ_k is constant, then its factorization is precisely the same as in the time-invariant case (equation (9.7)): suppose

$$\mathbf{\Sigma}_k = \left[\begin{array}{cccc|cc} a & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times \\ & & & \times & \times & \times \\ \hline b_1 & \times & \times & \times & \times & \times \\ b_2 & \times & \times & \times & \times & \times \end{array}\right]$$

then factoring of a $C(1)$ section $(\mathbf{\Sigma}_1)_k$ leads to

$$(\mathbf{\Sigma}_1^*)_k\,\mathbf{\Sigma}_k = \left[\begin{array}{cccc|cc} 1 & 0 & 0 & 0 & 0 & 0 \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times \\ & & & \times & \times & \times \\ \hline 0 & \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times & \times \end{array}\right].$$

Continuing recursively, we obtain a factorization as $\mathbf{\Sigma}_k = \mathbf{\Sigma}_{1,k} \cdots \mathbf{\Sigma}_{n,k}\mathbf{\Sigma}'_k$, where each degree-1 elementary section $\mathbf{\Sigma}_{i,k}$ is of $C(1)$-type and has a factorization as in (9.12). $\mathbf{\Sigma}'_k$ is the residue $\left[\begin{array}{c|c} I & \\ \hline & D'_k \end{array}\right]$ where D'_k is a unitary matrix. $\mathbf{\Sigma}'_k$ plays the role of a constant (non-dynamic) termination section, and can also be factored into elementary operations (using G-like sections, which result in a Gentleman-Kung triangular array). The corresponding network structure of a single stage is depicted in figure 9.5(a).

In the case of a shrinking state dimension, we have for example

$$\mathbf{\Sigma}_k = \left[\begin{array}{cccc|ccc} \cdot & \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times & \times \\ & & & \times & \times & \times & \times \\ \hline \cdot & \times & \times & \times & \times & \times & \times \\ \cdot & \times & \times & \times & \times & \times & \times \end{array}\right].$$

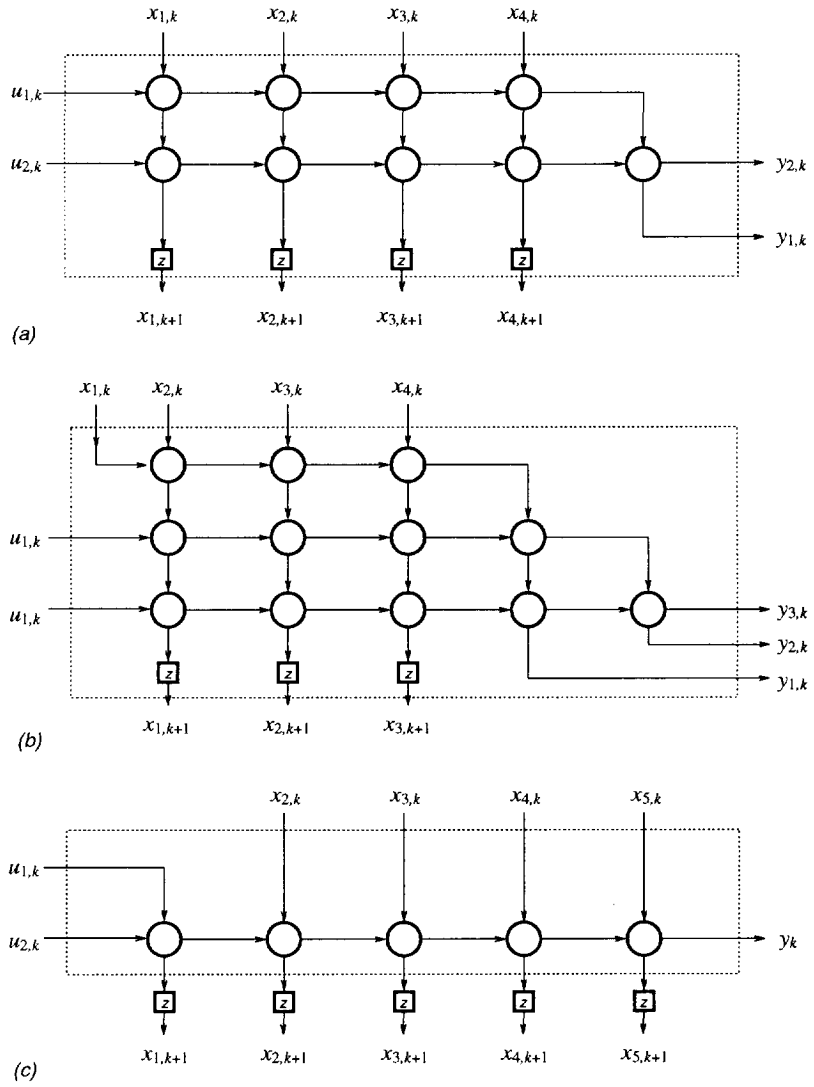


Figure 9.5. Lossless cascade realizations of a contractive system T, stage k. (a) Constant state dimension, (b) shrinking state dimension, (c) growing state dimension.

Of necessity, the state dimension of the first factor has to be $\#(\mathcal{B}_1)_{k+1} = 0$, for else $\#(\mathcal{B}_1)_k > 1$. With an elementary section of S-type as the first factor, we obtain

$$(\mathbf{\Sigma}_1^*)_k \mathbf{\Sigma}_k = \left[\begin{array}{ccc|ccc} \times & \times & \times & \times & \times & \times \\ & \times & \times & \times & \times & \times \\ & & \times & \times & \times & \times \\ \hline \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \\ \times & \times & \times & \times & \times & \times \end{array}\right],$$

so that, effectively, the first state has become an input of the subsequent factors. At this point, the factorization continues with the factorization of a realization with constant state dimension, which case has been covered above. The resulting network structure of the lossless stage is shown in figure 9.5(b). If the state dimension of $\mathbf{\Sigma}_k$ shrinks by d states, then a number of d S sections can (and have to) be extracted.

Finally, if the state dimension of $\mathbf{\Sigma}_k$ grows, for example

$$\mathbf{\Sigma}_k = \left[\begin{array}{ccccc} 0 & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ \hline b_1 & \times & \times & \times & \times \\ b_2 & \times & \times & \times & \times \end{array}\right],$$

then the first section is necessarily a G section. It cancels b_2 against b_1, after which

$$(\mathbf{\Sigma}_1^*)_k \mathbf{\Sigma}_k = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ & \times & \times & \times & \times \\ & & \times & \times & \times \\ & & & \times & \times \\ \hline 0 & \times & \times & \times & \times \end{array}\right].$$

The first input $u_{1,k}$ has effectively been mapped to a new state $x_{1,k+1}$ by $(\mathbf{\Sigma}_1)_k$. The first column and the first row of B do not play a role in the subsequent factorization, which again is reduced to a factorization of a realization with constant state dimension, as above. The corresponding network is depicted in figure 9.5(c). If, more in general, the first d columns of A would have been zero, then the first d sections of the factorization would have been G sections. For minimality of the factorization, we must require that the top left $d \times d$ submatrix of B_Σ has been made upper triangular by suitable unitary state transformations, in the process of the transformation to Schur form.

Algorithm 9.1 summarizes the procedure for the general case where A_k is of size $d_k \times d_{k+1}$. When the state dimension shrinks, *i.e.*, $d_k > d_{k+1}$, then the first $d_k - d_{k+1}$ states are treated as inputs rather than as states, but the actual factorization algorithm remains the same.

In: $\boldsymbol{\Sigma} = \{A, B, C, D\}$ ($A_k : d_k \times d_{k+1}$, M_k inputs, N_k outputs)
Out: $\{(\boldsymbol{\Sigma}_{ij})_k\}$, $\{(\boldsymbol{\Sigma}'_{ij})_k\}$ (elementary rotations: factors of $\boldsymbol{\Sigma}_k$)

$Q_1 = [\,\cdot\,]$

for $k = 1, \cdots, n$

$$\boldsymbol{\Sigma}_k := \begin{bmatrix} Q_k & \\ & I \end{bmatrix} \boldsymbol{\Sigma}_k$$

– if $d_k > d_{k+1}$ ('shrink'), move first $d_k - d_{k+1}$ rows of $[A_k \;\; C_k]$ to $[B_k \;\; D_k]$.
– if $d_k < d_{k+1}$ ('grow'), move first $d_{k+1} - d_k$ rows of $[B_k \;\; D_k]$ to $[A_k \;\; C_k]$.

$A_k =: R_{k+1} Q_{k+1}$ RQ-decomposition of A_k

$$\boldsymbol{\Sigma}_k := \boldsymbol{\Sigma}_k \begin{bmatrix} Q^*_{k+1} & \\ & I \end{bmatrix}$$

for $i = 1, \cdots, d_k$
  for $j = 1, \cdots, N_k$
    $(\hat{\boldsymbol{\Sigma}}_{ij})_k$ = givens$^\dagger[A_k(i, i); B_k(j, i)]$
    $\boldsymbol{\Sigma}_k$:= $(\boldsymbol{\Sigma}_{ij})^*_k \boldsymbol{\Sigma}_k$
  end
end

$\boldsymbol{\Sigma}'_k = D_{\Sigma_k}$ (also factor termination section)
for $i = 1, \cdots, N_k$
  for $j = 1, \cdots, N_k$
    $(\hat{\boldsymbol{\Sigma}}'_{ij})_k$ = givens$[\boldsymbol{\Sigma}'(i, i)_k; \boldsymbol{\Sigma}'(j, i)_k]$
    $\boldsymbol{\Sigma}'_k$:= $(\boldsymbol{\Sigma}'^*_{ij})_k \boldsymbol{\Sigma}'_k$
  end
end
end

† givens$[a; b]$: see equation (9.4).

Algorithm 9.1. The factorization algorithm.

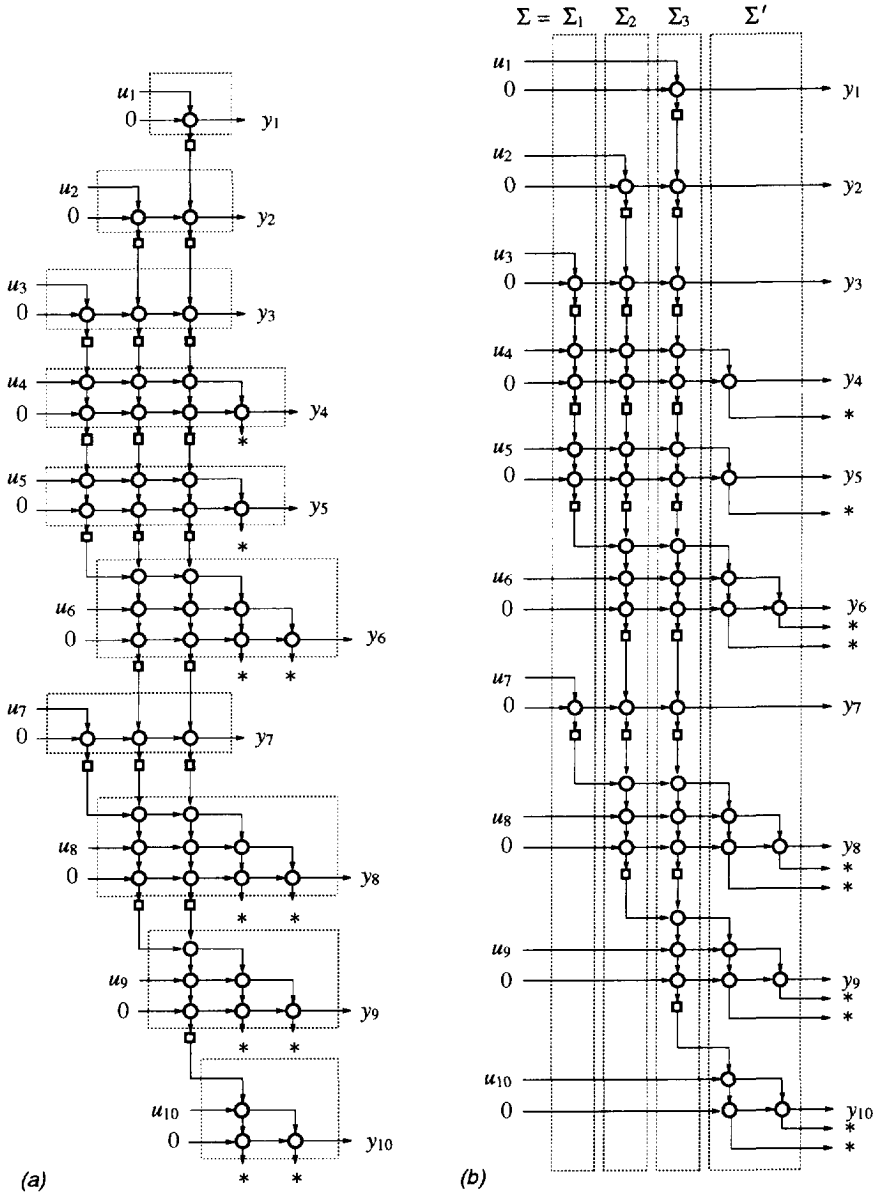


Figure 9.6. (*a*) Lossless embedding and cascaded network structure of $T: u \rightarrow y$, a 10×10 upper triangular matrix with local state dimension ≤ 3. Outputs marked by '*' are ignored. (*b*) Same as (*a*), but now displayed as a factorization of Σ into three degree-1 sections and a 'constant' termination section.

However, if the state dimension grows ($d_k < d_{k+1}$), then some inputs are treated as extra states in the factorization algorithm.

With the three types of stages given in figure 9.5, we can describe all possible stages that occur in locally finite unitary realizations that are in Schur form. It has already been mentioned that the stages can be factored independently of each other. The cascade network structure of the complete state realization $\mathbf{\Sigma}$ then follows by piecing together the results of the individual stages. An example network is shown in figure 9.6(a). In the example, we consider a 10×10 strictly contractive upper triangular matrix T, with 1 input and 1 output at each point, and a state dimension sequence $\mathcal{B}$ given by

$$\#\mathcal{B} = [0, 1, 2, 3, 3, 3, 2, 3, 2, 1, 0].$$

T has an embedding into an inner operator Σ, with block decomposition

$$\Sigma = \begin{bmatrix} \Sigma_{11} & T \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Hence T is the partial transfer operator of Σ from the first input to the second output when the secondary input is put to zero.

Factorization into degree-1 lossless sections

At this point, we have described a global cascade factorization procedure (proposition 9.2) of a locally finite inner operator Σ into two factors Σ_1 and Σ_2, and we have derived some of its implications with respect to the factorization at the local (stage) level. It now remains to make precise the remark at the end of the previous subsection that the local stages can be 'pieced together' to yield a global cascade factorization.

Let be given a locally finite inner operator Σ, with state dimension sequence $\mathcal{B}$. The objective is to compute a factorization $\Sigma = \Sigma_1 \cdots \Sigma_n \Sigma'$ into a minimal number of n degree-1 sections, and a terminating diagonal unitary operator Σ' (a 'constant' section). A related question is: what is the minimal value of n? It is clear that n is at least equal to the maximal number $\max_k \#(\mathcal{B})_k$ of states of Σ that are present at any stage. We show that n can in fact be equal to this number.

In view of proposition 9.2, there remains to determine a possible state sequence $\mathcal{B}_1$ of the first factor Σ_1. The other factors are then obtained recursively, by factoring $\Sigma_1^* \Sigma$, until the state dimension has been reduced to zero. The remainder $\Sigma_n^* \cdots \Sigma_1^*$ is then equal to the constant section Σ'. The number of states $\#(\mathcal{B}_1)_k$ of the first factor is allowed to be at most equal to 1 at each stage k, in order to obtain a degree-1 section. The other constraint on $\mathcal{B}_1$ is the fact that $(A_{21})_k$ in (9.10) must be equal to zero (or have vanishing dimensions) for each k. The discussions in the previous paragraph have shown that, as a consequence, within a stage it is not possible to extract a $C(1)$ section before an S section or a G section. A trivial $C(0)$ section can always be extracted.

The following choice of $\mathcal{B}_1$ satisfies the constraints. Let $n = \max_k \#(\mathcal{B})_k$. Then $\mathcal{B}_1$ is given by

$$\#(\mathcal{B}_1)_k = \begin{cases} 1, & \text{if } \#(\mathcal{B})_k = n, \\ 0, & \text{otherwise.} \end{cases} \tag{9.13}$$

Indeed, with this $\mathcal{B}_1$, we extract as many stages with $C(0)$ sections as possible (which do not have constraints), and only extract other sections where factors Σ_2 till Σ_n must have states anyway. At the same time, $\mathcal{B}_1$ is such that it reduces the degree of Σ: $\Sigma_1^* \Sigma$ has a maximal state dimension $n-1$. Acting recursively, we obtain a factorization of Σ into n sections, each of which has local degree at most 1. The results are summarized in the following theorem.

THEOREM 9.4. *Let Σ be an inner operator which is locally finite with state dimension sequence $\mathcal{B}$, and strictly stable. Let $n = \max_k \#(\mathcal{B})_k$. Then Σ has a factorization*

$$\Sigma = \Sigma_1 \cdots \Sigma_n \Sigma',$$

where each Σ_i is a strictly stable inner section of local degree at most 1 ($\max_k \#(\mathcal{B}_i)_k = 1$), and whose local degrees add up to the local degree of Σ ($\sum_i \#(\mathcal{B}_i)_k = \#(\mathcal{B})_k$). Σ' is a unitary diagonal operator.

PROOF According to theorem 4.5, Σ has a unitary realization $\mathbf{\Sigma}$. The realization can be put into Schur form by unitary state transformations (proposition 9.1). Next, choose $\mathcal{B}_1$ according to equation (9.13). We first show that $\mathcal{B}_1$ generates a partitioning of $A = A_\Sigma$ such that, for all k, $(A_{21})_k = 0$ or has vanishing dimensions. Indeed, as long as $\#(\mathcal{B})_k < n$ and $\#(\mathcal{B})_{k+1} < n$, we have $\#(\mathcal{B}_1)_k = 0$ and $\#(\mathcal{B}_1)_{k+1} = 0$ so that $(A_{21})_k = [\cdot]$. At a certain point k, $\#(\mathcal{B})_k < n$ and $\#(\mathcal{B})_{k+1} = n$, and figure 9.3($c$) shows that in this case we can put $\#(\mathcal{B}_1)_{k+1} = 1$, which makes $(A_{21})_k$ equal to the first column, consisting only of zero entries. While $\#(\mathcal{B})_k = n$ and $\#(\mathcal{B})_{k+1} = n$, A_k is an upper triangular matrix, so that we can put $\#(\mathcal{B}_1)_k = 1$, $\#(\mathcal{B}_1)_{k+1} = 1$ to obtain $(A_{21})_k = 0$. Finally, when $\#(\mathcal{B})_k = n$ and $\#(\mathcal{B})_{k+1} < n$, A_k has the form shown in figure 9.3(b), so that we have to put $\#(\mathcal{B}_1)_{k+1} = 0$, which gives $(A_{21})_k = [\cdot]$. Hence $\mathcal{B}_1$ satisfies the requirements, so that, according to proposition 9.2, we can extract a factor Σ_1. We can continue in the same way with $\Sigma_1^* \Sigma$, which has a maximal state dimension equal to $n-1$. This degree reduction is because we had $\#(\mathcal{B}_1)_k = 1$ whenever $\#(\mathcal{B})_k = n$. Acting recursively, we end with $\Sigma' = \Sigma_n^* \cdots \Sigma_1^* \Sigma$ having 0 states, and hence Σ' is a unitary diagonal constant. □

We can write the 10×10 example in figure 9.6(a) in factored form, as obtained by the above theorem. The resulting cascade factorization is displayed in figure 9.6(b). The actual structure is the same as given in figure 9.6(a), but the elementary stage sections are now grouped vertically into sections, rather than horizontally into stages.

Computational complexity

The computational complexity of the cascade network is, at each stage, linear in the number of elementary operations. This is in contrast to a direct network realization of a given state realization $\{A, B, C, D\}$, which would have quadratical complexity. If the network consists of N stages and if the average number of states in a stage is d, then the number of elementary operations required for a vector-matrix multiplication using the cascade network is of order $O(2dN)$ rotations, rather than $O(\frac{1}{2}N^2)$ multiplications for a direct vector-matrix multiplication. (The complexity of a rotation operation is 4 multiplications for a direct implementation, but less if a special hardware implementation (CORDIC) is used.) Hence, if $d \ll N$, a considerable reduction in the number of operations is achieved. In addition, the network is numerically stable. All elementary operations are rotations, which means that the network is lossless and does not amplify numerical errors introduced at any point in the computation.

9.3 TIME-VARYING Θ-BASED CASCADE FACTORIZATION

In the previous section, we embedded the given contractive operator T in a unitary operator Σ, and subsequently factored this operator into elementary sections. The result was a computational network consisting of unitary Givens rotations, with a data flow strictly from the left to the right, and from the top to the bottom. An alternative cascade factorization is obtained by computing the J-unitary operator Θ associated with Σ,[1] factoring Θ into elementary J-unitary sections Θ_i, and converting each of the sections to their unitary equivalent. The result is again a minimal factorization of the unitary realization $\mathbf{\Sigma}$ of Σ into degree-1 realizations, although the factorization is different from the one we obtained earlier. The order of the computations in this factorization is such that the corresponding cascade factorization of Σ can no longer be written as a product of elementary unitary sections.

The reason for studying Θ-based factorizations is at least twofold. Firstly, they lead to different realizations of unitary operators Σ, also specified by a minimal number of parameters. These realizations may have different numerical properties with respect to parameter sensitivity (although we do not go to that level of detail). Secondly, the same type of networks are obtained in the solution of a number of other problems. For example, the solution of certain constrained interpolation problems, such as the Nevanlinna-Pick interpolation problem [36], or the solution of the Nehari problem and (more in general) the model approximation problem in chapter 6, leads to Θ-based cascade networks. This is of course not coincidental: the description of the solution of these interpolation problems also gives rise to J-unitary operators Θ. Upon factorization of Θ, each factor contains a single interpolation constraint of the original problem. Other problems where networks of the

[1] In this section, we assume that the reader has knowledge of the contents of section 5.1.

same type occur are in the Generalized Schur algorithm for inverse Cholesky factorization [37, 38], and (time-varying) prediction-error filters and RLS adaptive filters [39].

We will first derive some factorization results for J-unitary upper operators, and then specialize to the case where the state signature sequence equals $J_{\mathcal{B}} = I$. Subsequently, we derive the corresponding factorization of $\mathbf{\Sigma}$, and the computational network that this factorization of $\mathbf{\Sigma}$ represents.

Factorization into J-unitary elementary sections

The J-unitary factorization into elementary sections is again straightforward once a general factorization into two J-unitary factors has been derived. The latter is formulated in the following proposition, comparable to proposition 9.2.

PROPOSITION 9.5. *Let $\mathbf{\Theta} \in \mathcal{U}(\mathcal{B} \times \mathcal{M}, \mathcal{B}^{(-1)} \times \mathcal{N})$ be J-unitary, and and have a block partitioning as*

$$\mathbf{\Theta} = \left[\begin{array}{cc|c} A_{11} & A_{12} & C_1 \\ 0 & A_{22} & C_2 \\ \hline B_1 & B_2 & D \end{array}\right] \tag{9.14}$$

where $A_{11} \in \mathcal{U}(\mathcal{B}_1, \mathcal{B}_1^{(-1)})$ for some state-space sequence $\mathcal{B}_1 \subset \mathcal{B}$. Define the space sequences $\mathcal{N}_1$ and $\mathcal{B}_2$ by the relations $\mathcal{B}_1 \times \mathcal{M} = \mathcal{B}_1^{(-1)} \times \mathcal{N}_1$, and $\mathcal{B} = \mathcal{B}_1 \times \mathcal{B}_2$.

1. *Then J-unitary operators $\hat{\mathbf{\Theta}}_1, \hat{\mathbf{\Theta}}_2$ exist, with $\hat{\mathbf{\Theta}}_1 = \{A_{11}, B_1, C_1', D_1\} \in \mathcal{U}(\mathcal{B}_1 \times \mathcal{M}, \mathcal{B}_1^{(-1)} \times \mathcal{N}_1)$, $\hat{\mathbf{\Theta}}_2 = \{A_{22}, B_2', C_2, D_2\} \in \mathcal{U}(\mathcal{B}_2 \times \mathcal{N}_1, \mathcal{B}_2^{(-1)} \times \mathcal{N})$, such that*

$$\mathbf{\Theta} = \left[\begin{array}{cc|c} A_{11} & & C_1' \\ & I & \\ \hline B_1 & & D_1 \end{array}\right] \left[\begin{array}{cc|c} I & & \\ & A_{22} & C_2 \\ \hline & B_2' & D_2 \end{array}\right] = \mathbf{\Theta}_1 \mathbf{\Theta}_2 . \tag{9.15}$$

2. *If $\Theta \in \mathcal{U}$ is a J-unitary operator with a J-unitary realization $\mathbf{\Theta}$ of the form (9.14), and if $\ell_{A_\Theta} < 1$, then $\Theta = \Theta_1 \Theta_2$, where Θ_1, Θ_2 are J-unitary operators with J-unitary realizations given by $\hat{\mathbf{\Theta}}_1, \hat{\mathbf{\Theta}}_2$, with $\ell_{A_{11}} < 1$, $\ell_{A_{22}} < 1$. The factorization is minimal.*

PROOF The proof is the same as in proposition 9.2, except that now a J-unitary extension of $[A_{11}^* \;\; B_1^*]^*$ must be found. The existence of such an extension was proven in lemma 5.18. The extension yields $\hat{\mathbf{\Theta}}_1$, and $\hat{\mathbf{\Theta}}_2$ then follows from $\mathbf{\Theta}_1^{-1}\mathbf{\Theta} = \mathbf{\Theta}_2$, which has the form specified in (9.14). □

In order to obtain a factorization into elementary sections of local degree ≤ 1, we choose $\mathcal{B}_1$ as in equation (9.13), viz.

$$\#(\mathcal{B}_1)_k = \begin{cases} 1 & \text{if } \#(\mathcal{B})_k = n, \\ 0 & \text{otherwise.} \end{cases}$$

With this choice, theorem 9.4 can be adapted to J-unitary operators:

THEOREM 9.6. *Let $\Theta \in \mathcal{U}$ be a J-unitary operator which is locally finite with state dimension sequence $\mathcal{B}$, and strictly stable. Let $n = \max_k \#(\mathcal{B})_k$. Then Θ has a factorization*

$$\Theta = \Theta_1 \cdots \Theta_n \Theta',$$

where each Θ_i is a strictly stable J-unitary section of local degree ≤ 1 ($\max_k \#(\mathcal{B}_i)_k = 1$), and the local degrees of the Θ_i add up to the local degree of Θ ($\sum_i \#(\mathcal{B}_i)_k = \#(\mathcal{B})_k$). Θ' is a J-unitary diagonal operator.

PROOF The proof is the same as that of theorem 9.4, but now refers to proposition 9.5. □

There remains to investigate the structure of an elementary J-unitary section.

Elementary Θ sections

We now describe the factorization of an elementary J-unitary section of local degree at most equal to 1 into (J-unitary) Givens rotations. The resulting *structure* of the factored section is the same as in the unitary case, because the same sequence of operations is used to do the factorization. However, the *type* of each elementary operation is now either a unitary or a J-unitary Givens rotation. To keep the discussion manageable, we assume from now on that all state signatures are positive, as this will be the case in our future application.

As in the unitary case, we assume that a J-isometric column $[A_{11}^* \;\; B_1^*]^* \in \mathcal{D}$ is given, where each matrix $(A_{11})_k$ of the diagonal has dimensions at most equal to 1. This column is extended to a J-unitary realization $\hat{\mathbf{\Theta}}_1$, to be obtained in factored form. It is sufficient at this point to look only at the factorization of a single stage of the degree-1 section. With $\#(\mathcal{B}_1)_k \in \{0, 1\}$ and $\#(\mathcal{B}_1)_{k+1} \in \{0, 1\}$, the four possible sections in a stage are again described by the dimension of $(A_{11})_k$ as

$$\begin{cases} C(0): & 0\times 0, & S: & 1\times 0, \\ C(1): & 1\times 1, & G: & 0\times 1. \end{cases}$$

The cases $C(0)$ and S result in the same (trivial) sections as before:

$$(\hat{\mathbf{\Theta}}_1)_k = \left[\begin{array}{c|c} \cdot & \cdot \\ \hline \cdot & I \end{array}\right] \qquad \text{resp.} \qquad (\hat{\mathbf{\Theta}}_1)_k = \left[\begin{array}{c|cc} \cdot & 1 & 0 \\ \hline \cdot & 0 & I \end{array}\right]$$

(see figure 9.7(*a*),(*b*)). The case $C(1)$ is more interesting and follows from a factorization with Givens rotations of vectors of the form

$$\left[\begin{array}{c} a \\ \hline b_+ \\ b_- \end{array}\right] = \left[\begin{array}{c} (A_{11})_k \\ (B_1)_k \end{array}\right],$$

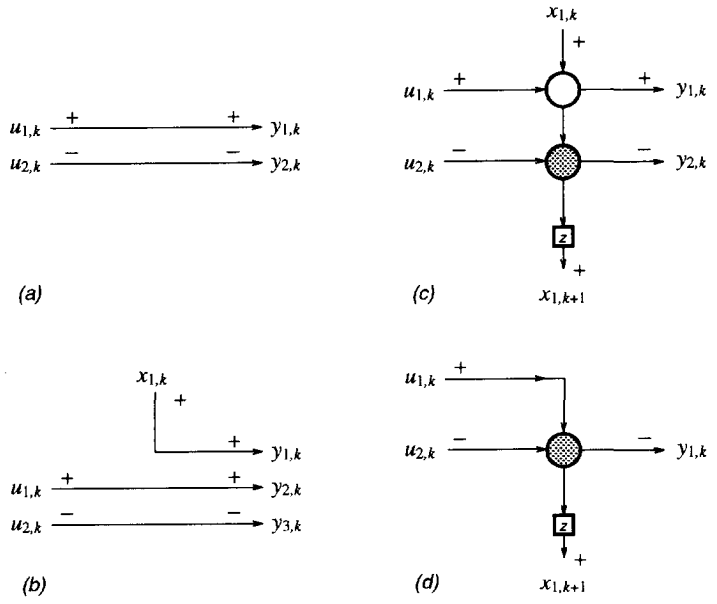


Figure 9.7. Elementary J-unitary sections in a stage. (a) $C(0)$ constant section with zero states, (b) S section, going from 1 state to 0, (c) $C(1)$ section with a constant number of 1 states, (d) G section, going from 0 to 1 state. The number of inputs/outputs have arbitrarily been set to 2: one with positive signature, the other with negative signature. The shaded circles represent J-unitary Givens rotations.

where a is a scalar and $b = \left[\begin{smallmatrix} b_+ \\ b_- \end{smallmatrix}\right]$ is partitioned according to the signature of the inputs at that point. The factorization is obtained in two steps,

$$(\hat{\boldsymbol{\Theta}}_{1,1}^{-1})_k \begin{bmatrix} a \\ \hline b_+ \\ b_- \end{bmatrix} = \begin{bmatrix} a' \\ \hline 0 \\ b_- \end{bmatrix}, \qquad (\hat{\boldsymbol{\Theta}}_{1,2}^{-1})_k \begin{bmatrix} a' \\ \hline 0 \\ b_- \end{bmatrix} = \begin{bmatrix} a'' \\ \hline 0 \\ 0 \end{bmatrix}.$$

Here, $(\hat{\boldsymbol{\Theta}}_{1,1})_k$ consists solely of unitary Givens rotations, used to cancel the entries of b_+ against a, while $(\hat{\boldsymbol{\Theta}}_{1,2})_k$ consists only of J-unitary Givens rotations. See figure 9.7(c). Note that the unitary scattering operator $(\hat{\boldsymbol{\Sigma}}_{1,1})_k$ corresponding to $(\hat{\boldsymbol{\Theta}}_{1,1})_k$ is the same because it is already unitary: $(\hat{\boldsymbol{\Sigma}}_{1,1})_k = (\hat{\boldsymbol{\Theta}}_{1,1})_k$. The factorization of a G section results in a comparable structure, and can also be described as $(\hat{\boldsymbol{\Theta}}_1)_k = (\hat{\boldsymbol{\Theta}}_{1,1})_k(\hat{\boldsymbol{\Theta}}_{1,2})_k = (\hat{\boldsymbol{\Sigma}}_{1,1})_k(\hat{\boldsymbol{\Theta}}_{1,2})_k$. As the same can obviously be done for the $C(0)$- and the S sections, the overall result is as follows.

LEMMA 9.7. *Let* $[A_{11}^* \;\; B_1^*]^* \in \mathcal{U}(\mathcal{B}_1 \times \mathcal{M}, \mathcal{B}_1^{(-1)})$ *be* $\{I, J_{\mathcal{M}}\}$*-isometric:*

$$[A_{11}^* \;\; B_1^*] \begin{bmatrix} I & \\ & J_{\mathcal{M}} \end{bmatrix} \begin{bmatrix} A_{11} \\ B_1 \end{bmatrix} = I,$$

and assume that its state dimension sequence $\mathcal{B}_1$ *has dimension at most equal to* 1 *at each point. Then this column has a* J*-unitary extension to* $\hat{\boldsymbol{\Theta}}_1 \in \mathcal{U}(\mathcal{B}_1 \times \mathcal{M}, \mathcal{B}_1^{(-1)} \times \mathcal{N})$ *such that*

$$\hat{\boldsymbol{\Theta}}_1 = \hat{\boldsymbol{\Theta}}_{1,1}\hat{\boldsymbol{\Theta}}_{1,2} = \hat{\boldsymbol{\Sigma}}_{1,1}\hat{\boldsymbol{\Theta}}_{1,2}$$

$$= \left[\begin{array}{c|cc} \times & \times & \\ \hline \times & \times & \\ & & I \end{array}\right] \left[\begin{array}{c|cc} \times & & \times \\ \hline & I & \\ \times & & \times \end{array}\right]$$

(where partitionings are according to $J_{\mathcal{M}}$*).*

With theorem 9.6, the result is that if Θ is a J-unitary operator which has a J-unitary realization $\boldsymbol{\Theta}$ with state signature sequence $J_{\mathcal{B}} = I$, then $\boldsymbol{\Theta}$ has a factorization into unitary and J-unitary factors as

$$\boldsymbol{\Theta} = [\boldsymbol{\Sigma}_{1,1}\boldsymbol{\Theta}_{1,2}] \cdot [\boldsymbol{\Sigma}_{2,1}\boldsymbol{\Theta}_{2,2}] \cdots [\boldsymbol{\Sigma}_{n,1}\boldsymbol{\Theta}_{n,2}] \cdot \boldsymbol{\Theta}'. \tag{9.16}$$

LEMMA 9.8. *If* $\boldsymbol{\Theta}$ *has factorization (9.16), then the corresponding* $\boldsymbol{\Sigma}$ *has factorization*

$$\boldsymbol{\Sigma} = [\boldsymbol{\Sigma}_{1,1}\boldsymbol{\Sigma}_{2,1} \cdots \boldsymbol{\Sigma}_{n,1}] \; \boldsymbol{\Sigma}' \; [\boldsymbol{\Sigma}_{n,2} \cdots \boldsymbol{\Sigma}_{2,2}\boldsymbol{\Sigma}_{1,2}] \tag{9.17}$$

in which $\boldsymbol{\Theta}_{i,2} \leftrightarrow \boldsymbol{\Sigma}_{i,2}$, $\boldsymbol{\Theta}' \leftrightarrow \boldsymbol{\Sigma}'$.

PROOF We first argue that $\boldsymbol{\Theta}$ in (9.16) can be written as

$$\boldsymbol{\Theta} = [\boldsymbol{\Sigma}_{1,1}\boldsymbol{\Sigma}_{2,1} \cdots \boldsymbol{\Sigma}_{n,1}] \cdot [\boldsymbol{\Theta}_{1,2}\boldsymbol{\Theta}_{2,2} \cdots \boldsymbol{\Theta}_{n,2}] \cdot \boldsymbol{\Theta}' \tag{9.18}$$

Indeed, because $\boldsymbol{\Sigma}_{i,1}$ and $\boldsymbol{\Theta}_{j,2}$, for $i \neq j$, act on different state variables and on different inputs, their order of application may be reversed: $\boldsymbol{\Theta}_{j,2}\boldsymbol{\Sigma}_{i,1} = \boldsymbol{\Sigma}_{i,1}\boldsymbol{\Theta}_{j,2}$. This allows to transform (9.16) into (9.18). Omitting the details, we mention that the transition from a Θ-representation to a Σ-representation is obtained by reversing the computational direction of the secondary inputs and outputs. This does not affect $[\boldsymbol{\Sigma}_{1,1}\boldsymbol{\Sigma}_{2,1} \cdots \boldsymbol{\Sigma}_{n,1}]$ as only the primary inputs and outputs are involved, while $[\boldsymbol{\Theta}_{1,2}\boldsymbol{\Theta}_{2,2} \cdots \boldsymbol{\Theta}_{n,2}] \cdot \boldsymbol{\Theta}' \leftrightarrow \boldsymbol{\Sigma}' \cdot [\boldsymbol{\Sigma}_{n,2} \cdots \boldsymbol{\Sigma}_{2,2}\boldsymbol{\Sigma}_{1,2}]$. This leads to equation (9.17). □

The structure of Θ according to the above factorization of $\boldsymbol{\Theta}$ is depicted in figure 9.8(*a*). It is the same as the structure of the network of Σ given in figure 9.6(*b*), but contains both unitary and J-unitary rotations (represented by shaded circles). The structure of Σ corresponding to this factorization of $\boldsymbol{\Theta}$ (figure 9.8(*b*)) is again the same, but the order in which computations are done is not only from left to right, but partially also from

right to left. Within a single stage, suppose that the inputs and the current state variables are known. In order to compute the next states and the outputs, first all rotations going from left to right have to be performed, and only then the next state variables and the output at the left can be computed. The network is said to be *non-pipelinable*, and the computational dependency, going from the left to the right and back to the left again, is said to be the computational bottleneck. This bottleneck is not present in the network in figure 9.6, and hence, from a computational point of view, a direct factorization of $\mathbf{\Sigma}$ yields a more attractive network.

Note that this network of Σ is a special case of the type of networks that has been obtained in the model reduction problem (*cf.* figure 6.8). In chapter 6, more general networks were obtained because the state signature of Θ was allowed to contain negative entries too.

Θ-based cascade factorization of T

Let $T \in \mathcal{U}$ be a given strictly contractive locally finite transfer operator. The process of realizing T via a Θ-based cascade starts with the orthogonal embedding of T in a unitary operator Σ, such that

$$\Sigma = \begin{bmatrix} \Sigma_{11} & T \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \tag{9.19}$$

where we have set $\Sigma_{12} = T$. The next step is to convert Σ to Θ, which requires the invertibility of Σ_{22}:

$$\Theta = \begin{bmatrix} \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} & -\Sigma_{12}\Sigma_{22}^{-1} \\ \Sigma_{22}^{-1}\Sigma_{21} & \Sigma_{22}^{-1} \end{bmatrix}$$

Θ is an upper operator only if Σ_{22}^{-1} is upper. As the factorization of Θ in the previous subsection required Θ to be upper (so that it has a causal realization), we see that Σ_{22} should be outer and invertible in order to obtain a Θ-based cascade factorization of Σ. If this requirement is satisfied, then a J-unitary realization $\mathbf{\Theta}$ of Θ is obtained in terms of a unitary realization $\mathbf{\Sigma}$ of Σ as

$$\mathbf{\Sigma} = \left[\begin{array}{c|cc} A & C_1 & C_2 \\ \hline B_1 & D_{11} & D_{12} \\ B_2 & D_{21} & D_{22} \end{array}\right] \quad \Rightarrow \quad \mathbf{\Theta} = \left[\begin{array}{c|cc} A - C_2D_{22}^{-1}B_2 & C_1 - C_2D_{22}^{-1}D_{21} & -C_2D_{22}^{-1} \\ \hline B_1 - D_{12}D_{22}^{-1}B_2 & D_{11} - D_{12}D_{22}^{-1}D_{21} & -D_{12}D_{22}^{-1} \\ D_{22}^{-1}B_2 & D_{22}^{-1}D_{21} & D_{22}^{-1} \end{array}\right]. \tag{9.20}$$

Note that if Σ_{22}^{-1} would not be upper, then we would by necessity obtain $\ell_{A_\Theta} > 1$ at this point. The factorization proceeds with a state transformation to make A_Θ upper triangular at each stage, which requires the time-varying Schur decomposition discussed in section 9.2. $\mathbf{\Theta}$ is subsequently factored into elementary sections, and conversion to scattering operators finally produces a factorization of $\mathbf{\Sigma}$ as in equation (9.17), and in a computational network as in figure 9.8(*b*). In this figure, T is the transfer operator $u \rightarrow y$ if the inputs at the right are put to zero.

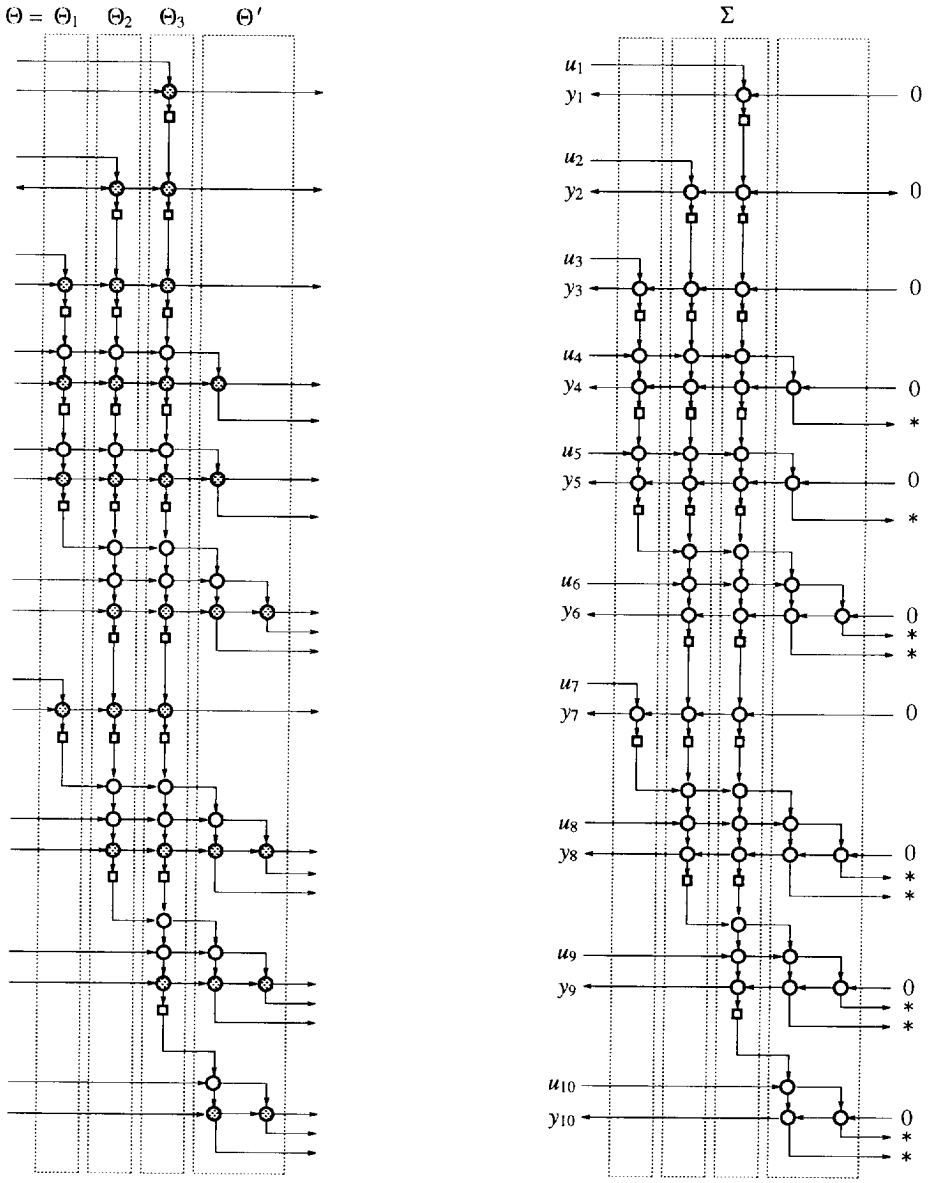


Figure 9.8. (*a*) a *J*-unitary cascade factorization has the same structure as a unitary cascade factorization, but contains *J*-unitary rotations (shaded circles), (*b*) Lossless embedding and Θ-cascade factorization of a strictly contractive upper operator $T: u \to y$.

However, the above is only possible when Σ_{22}^{-1} is outer and invertible. With Σ given as (9.19), when is this the case? A necessary condition for invertibility is that $\Sigma_{22}^{*}\Sigma_{22} \gg 0$, and since $\Sigma_{22}^{*}\Sigma_{22} = I - T^{*}T$, it follows that T must be strictly contractive. In this case, and if in addition the realization of T is observable, corollary 8.11 has shown that the embedding algorithm (theorem 7.5) yields Σ_{22} as an outer spectral factor of $I - T^{*}T$. Hence, if T is strictly contractive, Σ_{22} is outer and invertible automatically, and T has a Θ-based cascade realization. This is the reason why we have put $\Sigma_{12} = T$ in equation (9.19).

The Θ-based cascade network of Σ represents a filter structure which is well known in its time-invariant incarnation. In this context, one typically chooses $\Sigma_{11}(z) = T(z)$, because then the transmission zeros of $\Sigma(z)$, the zeros of $\Sigma_{11}(z)$, are equal to those of $T(z)$. Simultaneously, the zeros of $\Sigma_{22}(z)$ are directly related to those of $\Sigma_{11}(z)$ (they are 'reflected' in the unit circle). The point of using this filter structure is that these zeros appear as the zeros of the individual sections of the cascade, and hence they are individually determined directly by the parameters of the corresponding section, rather than by the combined effect of all parameters. It follows that the zeros of $T(z)$ are highly insensitive to parameter changes of the cascade, which makes the construction of filters with a well-defined stopband possible, even if approximate parameters (finite word-length implementations) are used.

However, note that in the time-varying case, using the above-described procedure, it is not possible to choose $\Sigma_{11} = T$, because Σ_{22} will in general not be outer and in this case A_Θ in (9.20) is not stable: $\ell_{A_\Theta} > 1$. In the time-invariant case, this does not pose real problems: even with the eigenvalues of A_Θ larger than 1, it is possible to factor Θ in the same way as before, which ultimately results in a stable cascade filter back in the scattering domain. There is no apparent reason why the same would not work in the time-varying domain: currently, the limitation seems to lie in the fact that we always require our realizations to be stable, in order to associate a transfer operator to it via $(I - AZ)^{-1}$. The foregoing factors provide reason to investigate (in other research) cases where the A-matrix contains both a stable and an anti-stable part. Because of state transformations, these parts can become mixed, and one of the first issues to address would be, given an A operator, to decouple it into stable and anti-stable parts.

Bibliography

[1] T. Kailath, *Linear Systems*. Prentice Hall, Englewood Cliffs, NJ, 1980.

[2] S. Darlington, "Synthesis of reactance 4-poles which produce prescribed insertion loss characteristics," *J. Math. Phys.*, vol. 18, pp. 257–355, 1939.

[3] A.V. Belevitch, *Classical Network Theory*. San Francisco: Holden Day, 1968.

[4] D.C. Youla and P. Tissi, "n-Port synthesis via reactance extraction–part I," *IEEE Int. Conf. Rec.*, vol. 14, no. 7, pp. 183–205, 1966.

[5] A. Fettweis, "Factorization of transfer matrices of lossless two-ports," *IEEE Trans. Circuit Th.*, vol. 17, pp. 86–94, 1970.

[6] R.A. Roberts and C.T. Mullis, *Digital Signal Processing*. Addison-Wesley, 1987.

[7] P. Dewilde, A.C. Vieira, and T. Kailath, "On a generalized Szegö-Levinson realization algorithm for optimal linear predictors based on a network synthesis approach," *IEEE Trans. Circuits Syst.*, vol. 25, pp. 663–675, Sept. 1978.

[8] T. Kailath, S.Y. Kung, and M. Morf, "Displacement ranks of matrices and linear equations," *J. Math. Anal. Appl.*, vol. 68, no. 2, pp. 395–407, 1979.

[9] B. Friedlander, M. Morf, T. Kailath, and L. Ljung, "New inversion formulas for matrices classified in terms of their distance from Toeplitz matrices," *Lin. Alg. Appl.*, vol. 23, pp. 31–60, 1979.

[10] H. Lev-Ari and T. Kailath, "Lattice filter parametrization and modeling of nonstationary processes," *IEEE Trans. Informat. Th.*, vol. 30, pp. 2–16, Jan. 1984.

[11] Y. Genin, P. Van Dooren, T. Kailath, J.M. Delosme, and M. Morf, "On Σ-lossless transfer functions and related questions," *Lin. Alg. Appl.*, vol. 50, pp. 251–275, 1983.

[12] V.P. Potapov, "The multiplicative structure of J-contractive matrix functions," *Amer. Math. Soc. Transl. Ser. 2*, vol. 15, pp. 131–243, 1960.

[13] M. Brodskii and M.S. Livsic, "Spectral analysis of non-self adjoint operators and intermediate systems," *Amer. Math. Soc. Transl. Ser. 2*, vol. 13, pp. 265–346, 1958.

[14] M.S. Livsic, *Operators, Oscillations, Waves (Open Systems)*, vol. 34. Providence: Amer. Math. Soc. Transl. Math. Monographs, 1972.

[15] P. Dewilde and H. Dym, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Informat. Th.*, vol. 27, pp. 446–461, July 1981.

[16] P. Dewilde and H. Dym, "Lossless chain scattering matrices and optimum linear prediction: The vector case," *Circuit Theory and Appl.*, vol. 9, pp. 135–175, 1981.

[17] P. Dewilde and H. Dym, "Lossless inverse scattering, digital filters, and estimation theory," *IEEE Trans. Informat. Th.*, vol. 30, pp. 644–662, July 1984.

[18] P.M. Dewilde, A.V. Belevitch, and R. Newcomb, "On the problem of degree reduction of a scattering matrix by factorization," *J. Franklin Inst.*, vol. 291, pp. 387–401, May 1971.

[19] J. Vandewalle and P.M. Dewilde, "On the irreducible cascade synthesis of a system with a real rational transfer matrix," *IEEE Trans. Circuits Syst.*, vol. 24, pp. 481–494, Sept. 1977.

[20] H. Bart, I. Gohberg, and M.A. Kaashoek, *Minimal Factorization of Matrix and Operator Functions*. Basel: Birkhäuser Verlag, 1979.

[21] P. Van Dooren and P.M. Dewilde, "Minimal cascade factorization of real and complex rational transfer matrices," *IEEE Trans. Circuits Syst.*, vol. 28, pp. 390–400, May 1981.

[22] M. Rakowski, "Minimal factorization of rational matrix functions," *IEEE Trans. Circuits and Systems–I: Fund. Th. Appl.*, vol. 39, pp. 440–445, June 1992.

[23] H. Bart, I. Gohberg, M.A. Kaashoek, and P. Van Dooren, "Factorization of transfer functions," *SIAM J. Control and Optimization*, vol. 18, pp. 675–696, Nov. 1980.

[24] E. Deprettere and P. Dewilde, "Orthogonal cascade realization of real multiport digital filters," *Circuit Theory and Appl.*, vol. 8, pp. 245–272, 1980.

[25] Ed. F. Deprettere, P.M. Dewilde, and P. Rao, "Orthogonal filter design and VLSI impementation," in *Proc. Int. Conf. Computers, Systems, and Signal Proc.*, (Bangalore, India), pp. 779–790, 1984.

[26] P. Dewilde, E.F. Deprettere, and R. Nouta, "Parallel and pipelined VLSI implementation of signal processing algorithms," in *VLSI and Moder Signal Processing* (H.J. Whitehouse S.Y. Kung and T. Kailath, eds.), Englewood Cliffs, NJ: Prentice-Hall, 1984.

[27] P. Dewilde, "Advanced digital filters," in *Modern Signal Processing* (T. Kailath, ed.), pp. 169–209, Springer Verlag, 1985.

[28] S.K. Rao and T. Kailath, "Orthogonal digital filters for VLSI implementation," *IEEE Trans. Circuits Syst.*, vol. 31, pp. 933–945, Nov. 1984.

[29] P.P. Vaidyanathan, "A unified approach to orthogonal digital filters and wave digital filters, based on LBR two-pair extraction," *IEEE Trans. Circuits Syst.*, vol. 32, pp. 673–686, July 1985.

[30] P.A. Regalia, S.K. Mitra, and P.P. Vaidyanathan, "The digital all-pass filter: A versatile signal processing building block," *Proc. IEEE*, vol. 76, pp. 19–37, Jan. 1988.

[31] M.R. Jarmasz and G.O. Martens, "A simplified synthesis of lossless cascade analog and digital two-port networks," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 1501–1516, Dec. 1991.

[32] U.B. Desai, "A state-space approach to orthogonal digital filters," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 160–169, Feb. 1991.

[33] H. Lev-Ari and T. Kailath, "State-space approach to factorization of lossless transfer functions and structured matrices," *Lin. Alg. Appl.*, vol. 162, pp. 273–295, Feb. 1992.

[34] G. Golub and C.F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1984.

[35] A. Bojanczyk, G. Golub, and P. Van Dooren, "The periodic Schur decomposition. Algorithms and applications," in *Proc. SPIE, "Advanced Signal Processing Algorithms, Architectures, and Implementations", III* (F.T. Luk, ed.), vol. 1770, (San Diego), pp. 31–42, July 1992.

[36] P.M. Dewilde, "A course on the algebraic Schur and Nevanlinna-Pick interpolation problems,". in *Algorithms and Parallel VLSI Architectures* (Ed. F. Deprettere and A.J. van der Veen, eds.), vol. A, pp. 13–69, Elsevier, 1991.

[37] E. Deprettere, "Mixed-form time-variant lattice recursions," in *Outils et Modèles Mathématiques pour l'Automatique, l'Analyse de Systèmes et le Traitement du Signal*, (Paris), CNRS, 1981.

[38] P. Dewilde and E. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy," in *Operator Theory: Advances and Applications*, vol. 29, pp. 97–116, Birkhäuser Verlag, 1988.

[39] S. Haykin, *Adaptive Filter Theory*. Prentice-Hall, 1991.

Chapter 10

CONCLUSION

At the end of this thesis, we have reached the point to reflect on what has been achieved. We indicate some potential applications and point out directions in which additional research is required. Once more, the application of computational linear algebra is considered, and in particular how the theory for upper triangular matrices can be generalized to apply to the usual type of matrices that are 'mixed' upper and lower. A second application area concerns the control of time-varying systems. We discuss the generalization of the H_∞ control problem of time-invariant systems to the time-varying context. Although the control of such systems in practice is still out of reach, with the theory in this thesis the operator-theoretic results of Feintuch and Francis [1] can now be put into a computable algorithm acting on state-space matrices.

10.1 APPLICATIONS TO COMPUTATIONAL LINEAR ALGEBRA

Most of the algorithms derived in the previous chapters were specified in terms of 'upper' matrices or operators. Because dimensions are permitted to be non-uniform, it is perhaps best to stress again that upper in this context really refers to matrices that are block upper, as shown in figure 10.1(a), and it may well be that such generalized matrices are not truly upper triangular in the ordinary linear algebra sense. Computations on these matrices can be done more efficiently if the ranks of the Hankel matrices, the top-left submatrices, are all relatively small compared to the size of the full matrix. If this is the case, several orders of magnitude of computational effort may be saved if a state realization of the matrix is used instead of the matrix itself: for $N \times N$ matrices with less than d states at each stage, algorithms that require $O(N^p)$ (for some p) operations are typically replaced by algorithms of order $O(d^p N)$. This is because all information in the matrix is condensed in N matrices whose size is of the order of $d \times d$ rather than $N \times N$. Because a realization is unique only up to a state transformation, this transformation can be used to derive 'sparse' realizations, which permits us to reduce the number of operations even further. For example, the

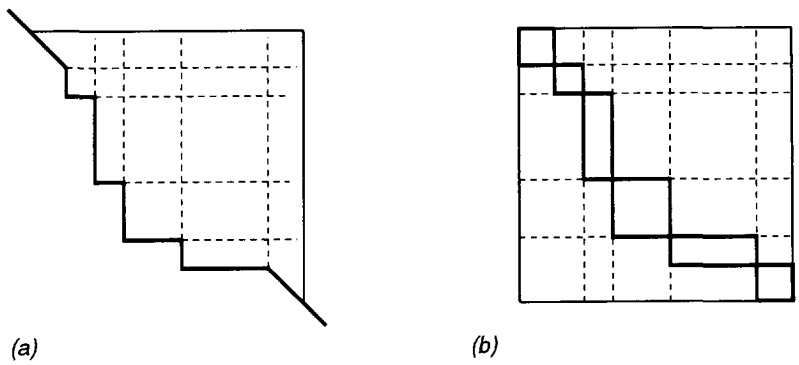


Figure 10.1. (*a*) Block-upper matrix, (*b*) mixed upper-lower matrix.

cascade factorization in chapter 9 reduces the multiplication of a vector by a contractive matrix to $O(2dN)$ rotations, which is comparable to $O(4dN)$ multiplications if specialized rotation modules are used.

However, the computational advantages are not limited to block-upper matrices. Matrices which are of mixed upper-lower type (that is, ordinary block-matrices, such as depicted in figure 10.1(*b*)) can also be considered, by converting them to block-upper matrices. This may be done in several ways. For example, such a matrix X can be decomposed as the sum of an upper and a lower matrix: $X = T_U + T_L$, in which both T_U and T_L^* are required to have a sparse state structure. X can also be factored as the product of a lower and an upper matrix: *e.g.*, $X = U^*T$, where U is inner (upper and unitary) and T is upper. Such a matrix U can be computed from a coprime factorization of T_L^*: Let $T_L^* = \Delta^* U$ be a factorization as in section 4.2, where U is inner and Δ is upper, then $UT_L = \Delta \in \mathcal{U}$, and hence $UX =: T$ is upper. A construction as in equation (4.12) provides a state realization of U in terms of one of T_L^*, and a realization of T is computed by composing a realization of Δ (as in equation (4.13)) with one of UT_U.

Using these decompositions, one immediately obtains efficient ways to do calculations with the general block-matrices, by reducing them to calculations on upper (or lower) matrices with sparse realizations.

1. Multiplication of a vector u by X can be done in two ways: $uX = uT_U + uT_L$ (parallel), or $uX = (uU^*)T$ (cascade). Both computations involve a forward recursion and a backward recursion. The first way is to be preferred because the recursions are decoupled, and because the total number of operations is lower than in the second case.

2. A QR factorization of X can be obtained in two steps. The first is, again, a fac-

torization $X = U^*T$. Using the observation made in section 4.2 that an inner-outer factorization of a block-upper matrix is in fact a QR factorization of this matrix, an inner-outer factorization $T = VT_0$ is computed using algorithm 4.1, so that a QR factorization of X is obtained as $X = (U^*V)T_0$. Note that $X = UT$ is not yet a QR factorization of X, because T is a block-upper matrix rather than a 'truly' upper triangular matrix: if the number of states in the lower part of X is time-varying, then the dimension sequences of T are non-uniform even if X has uniform dimensions, so that the inner-outer factorization of T is required in order to obtain a T_0 which is 'truly' upper triangular.

3. If X is invertible, then X^{-1} is obtained from the above QR factorization as $X = T_0^{-1}V^*U$. The role played by the factorization into inner and outer factors is essential in cases where T is a matrix with non-uniform dimension sequences, as the inverse of T need not be upper in this case (as exemplified in equation (2.16)). Since T_0 is outer, its inverse is known to be upper and hence there is an easy formula for its state realization in terms of the realization of T_0, see equation (1.3).

4. A spectral factorization can be used to obtain the Cholesky factor of a positive Hermitian matrix. This has already been described in chapter 8.

Generalizations are obtained by considering 'infinite' matrices, or operators. As with upper matrices, one can consider infinite matrices with borders that are constant along the diagonals (this corresponds to systems which are time varying only on a finite interval), periodical cases, and, more in general, infinite matrices which have strictly stable realizations, so that the associated Riccati recursions converge from approximate initial points. In the latter case, the matrix is 'diagonally dominant' at the borders, and only the interval in which the matrix is of interest needs to be considered in the computations.

One application in which such general matrices of very large dimensions arise occurs in the modeling of parasitic capacitances of a VLSI circuit [2, 3, 4, 5]. A discretization of the physical relations lead to a finite element matrix, where the (i,j)-th entry describes the potential induced at a point in space labeled j, caused by a point charge at position i. The matrix is positive definite, and has a banded structure. If the circuit consists of a repetition of a certain cell (*e.g.*, a memory cell), then the matrix is periodical, and if the substrate is much larger than the circuit ('infinitely' large), then the matrix is constant outside a finite region. The objective is to compute the inverse of the matrix, as the entries of the inverse are the capacitances between points in the circuit. It is sufficient to obtain an approximation of the inverse, as only the larger capacitances are of interest. The technique described in [2, 3] consists essentially of the Generalized Schur method as discussed in section 8.5, and will determine a spectral factor of an approximant of the inverse which has zero entries outside a band of interest. To determine this factor, only the entries of the original matrix on the same band need to be known. Matrices specified on multiple bands are considered in [5].

There are several issues in the above computational schemes which require further attention. For example, given some matrix, efficient algorithms are possible if it has a sparse state structure, that is, if the sequence of Hankel matrices of X and X^* have low rank. Since we allow non-uniform dimension sequences, these ranks can be different for different matrix partitionings, and the first problem is to determine a suitable block-partitioning of the matrix such that all Hankel matrices have low rank (assuming this is possible). A second, related, step is to capture the structure, *i.e.*, to determine a realization for X. Algorithm 3.1 can be used to this end, but the SVD occurring in the inner loop of this algorithm might make the algorithm computationally prohibitive. Since the Hankel matrices differ only by one (block) column and row from each other, their SVDs are related. Which SVD updating schemes are effective in this case? As a start, some possibilities are mentioned at the end of chapter 3, but for large matrices, a combination with some approximation scheme is also required, *e.g.*, by only taking a band of the matrix into account. (In the system theory context, this approach is connected to what is known as the partial realization problem.)

With the high-order realization obtained in one way or the other, the subsequent step is typically a Hankel-norm approximation to obtain an 'optimal' low-order approximating system. Here, an open problem is to determine an appropriate tolerance (scaling matrix) Γ, which should be such that the required number of states is obtained. Unlike the time-invariant case (where Γ is a scalar), how to do this is, as yet, unknown. The approximation step may be combined with the realization scheme. For example, a subspace decomposition algorithm with 'cheap' rank estimation (a QR factorization ?) might be used to estimate the state spaces, which would overdetermine the rank and yield realizations that are too large. The extra states can subsequently be removed by the approximation step.

The realization and approximation step can be combined into a single algorithm, utilizing the 'order-recursive interpolation' step detailed in section 6.6 to compute the required Θ-matrix. Again, for large matrices the attention has to be restricted to a band. The embedding of this step into an overall satisfactory algorithm is still to be investigated. For mixed upper-lower matrices, the upper and lower part can be approximated separately, leading to some kind of 'double Hankel-norm' approximant. The question arises as to how optimal this approximant is, considering that alternative approximation schemes are possible. For example, one may first factor X into $X = U^*T$, and subsequently approximate U and T separately.

Concluding, we might say that the system approach to computational linear algebra yields promising techniques in which first principles have become clear but several details still need to be sorted out. On a higher level, the techniques can probably be combined with other sparse matrix computational methods as well. It will, in particular, be interesting to consider links with α-stationary techniques.

10.2 APPLICATION TO AN H_∞-CONTROL PROBLEM

In the context of control of time-varying systems, a number of the standard tools that are typically required have been described in the previous chapters. The main results in this area are the availability of state-space algorithms for a number of factorization problems (coprime factorization, inner-outer factorization and spectral factorization), the solution of the Nehari problem in state-space terms, and the computation of optimal Hankel-norm approximants. One control application in which a number of these results are instrumental is given by the standard optimal control problem, known as the H_∞ control problem. This problem was introduced by Zames [6] and has received considerable interest in the past decade (see *e.g.*, Francis [7]). For time-varying systems, the problem of uniformly optimal control of time-varying systems was discussed by Feintuch and Francis [8, 1]. In this problem, a plant G is given, and a causal regulator K is to be designed such that the closed-loop system is stable and the noise terms v and w, acting on the inputs of the plant and the regulator, have minimal effect on the outputs u, y of the closed-loop system. See figure 10.2. A mathematical formulation of the latter problem is obtained by defining $\| [u \quad y] \|$ as the cost function, to be minimized by design of K and subject to the worst possible noise input with joint energy bounded by 1:

$$\min_K \{ \| [u \quad y] \| : \| [v \quad w] \| \leq 1 \}$$

One thus wants to minimize the norm of a certain transfer operator, at the same time satisfying a stability requirement (this is also known as the *sensitivity minimization problem*). By introducing a Youla reparametrization [9, 10], the problem can be reformulated as

$$\min_{Q \in \mathcal{U}} \| R - SQT \| \tag{10.1}$$

where R, S, T are known causal bounded linear operators, and Q is a causal operator to be designed. The above problem formulation is known as the *model matching problem.* Another control problem which reduces to the same formulation is *robust stabilization*: given a set of plants, specified as a nominal plant and a 'radius' (upper bound on the deviation of the actual plant from the nominal plant), find a stable feedback operator K that stabilizes all plants in the set [7]. Time-varying sensitivity minimization is also discussed by Ball, Gohberg and Kaashoek [11]. Their solution required the inverse of an operator $T \in \mathcal{U}$, which calls for an inner-outer factorization.

The solution of the minimization problem (10.1) is readily described using inner-outer factorizations $S = S_i S_o$, $T = T_o T_i$. For the sake of exposition, first assume that S, T are boundedly invertible. Because $\| R - SQT \| = \| S_i^*(R - SQT)T_i^* \| = \| S_i^* R T_i^* - S_o Q T_o \|$, the problem reduces to a Nehari problem: with $R' = S_i^* R T_i^* \in \mathcal{X}$, $Q' = S_o Q T_o \in \mathcal{U}$, Q' is determined from

$$\min_{Q' \in \mathcal{U}} \| R' - Q' \|,$$

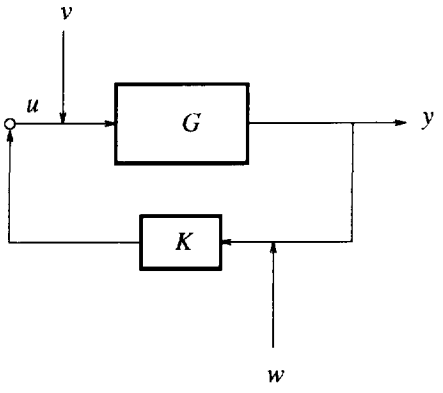


Figure 10.2. The 'standard' control configuration.

and Q follows as $Q = S_o^{-1} Q' T_o^{-1}$. Note that Q is causal as required because S_o and T_o are outer. Hence all that is required in the simplified case is two inner-outer factorizations and the solution of a Nehari problem.

In the real control application, however, S and T are not invertible. The solution to the general problem was formulated by Feintuch and Francis [1], and involves, besides the inner-outer factorizations and the solution to the Nehari problem, also two spectral factorizations. Their solution was described at the operator level, and it was remarked that "at present, computation of uniformly optimal controllers for time-varying systems is not feasible". Currently however, with algorithms for the factorizations and the Nehari problem available at the state-space level, computation of the optimal controller is certainly possible once a state realization of the plant is known.

It should be remarked, however, that the optimal controller can only be obtained as the solution of the above (theoretical) problem formulation, and that application of the time-varying system theory to practical problems in control is largely an open research area. Problems already arise at the first step: the identification of the system. The input and output state spaces $\mathcal{H}$ and $\mathcal{H}_0$ can only be determined from a large (in theory infinite) collection of pairs of input sequences with their corresponding output sequences, which this is not feasible. For time-varying systems of which only one copy is available and which can be exited only once, it is even impossible to obtain more than one input-output pair, and hence an exact realization cannot be obtained without making further assumptions. Standard techniques use parametric models ('templates') of possible system responses, or assume that the system is only slowly time-varying.

With a realization available, the next obstacle is that, in the solution of the optimal control problem, recursions are required which go both forward and backward in time. For

example, the inner-outer factorization required for T is actually an outer-inner factorization: $T = T_o T_i$, which in the state-space algorithm results in a backward recursion, going from $k = +\infty$ to $k = -\infty$. Hence, the optimal controller can only be computed if the realization matrix of the plant is available for all points in time. One way out is to apply windowing techniques: with the assumption that the system is adequately damped, it is sufficient, for the computation of variables at time point k, to start backward recursions at point $k + N$ for some finite integer N (the window size). However, the computational overhead in this scheme can be quite high (it is increased by a factor N), and one still has to know realization matrices N time-steps in advance. Note that no backward recursions occur if T is not present in the optimization problem, or if T is already outer. Hence, in an attempt to match theory with practise, it might be useful to derive a modified optimization problem in which T is already of this form.

Bibliography

[1] A. Feintuch and B.A. Francis, "Uniformly optimal control of linear feedback systems," *Automatica*, vol. 21, no. 5, pp. 563–574, 1985.

[2] P. Dewilde and E. Deprettere, "The generalized Schur algorithm: Approximation and hierarchy," in *Operator Theory: Advances and Applications*, vol. 29, pp. 97–116, Birkhäuser Verlag, 1988.

[3] P. Dewilde, "New algebraic methods for modeling large-scale integrated circuits," *Circuit Theory and Appl.*, vol. 16, pp. 473–503, 1988.

[4] Z.-Q. Ning, *Accurate and Efficient Modeling of Global Circuit Behaviour in VLSI Layouts*. PhD thesis, Delft Univ. Techn., The Netherlands, 1989.

[5] H. Nelis, *Sparse Approximations of Inverse Matrices*. PhD thesis, Delft Univ. Techn., The Netherlands, 1989.

[6] G. Zames, "Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms, and approximate inverses," *IEEE Trans. Automat. Control*, vol. 26, pp. 301–320, Apr. 1981.

[7] B.A. Francis, *A Course in H_∞ Control Theory*, vol. 88 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, 1987.

[8] A. Feintuch and B.A. Francis, "Uniformly optimal control of linear time-varying systems," *Syst. Control Letters*, vol. 5, p. 67, 1984.

[9] D.C. Youla, H.A. Jabr, and J.J. Bongiorno, "Modern Wiener-Hopf design of optimal controllers: Part II," *IEEE Trans. Automat. Contr.*, vol. 21, p. 319, 1976.

[10] A. Feintuch and R. Saeks, *System Theory: A Hilbert Space Approach.* Academic Press, 1982.

[11] J.A. Ball, I. Gohberg, and M.A. Gohberg, "Time-varying systems: Nevanlinna-Pick interpolation and sensitivity minimization," in *Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing I (Proc. Int. Symp. MTNS-91)* (H. Kimura and S. Kodama, eds.), pp. 53–58, MITA Press, Japan, 1992.

SAMENVATTING

Tijdvariërende systeemtheorie en rekenschema's: toestandsmodellen, benaderingen, en factorisaties

Tijdvariërende lineaire systemen vormen een belangrijke veralgemenisering van het tijdsinvariante geval, waar veel meer over bekend is. In dit proefschrift worden tijddiscrete tijdvariërende systemen beschreven aan de hand van hun overdrachts-operatoren, waarbij signalen voorgesteld worden als oneindige reeksen (vectoren) met begrensde energie, de zogenaamde ℓ_2-reeksen. Operatoren op zulke reeksen hebben matrixrepresentaties, en in de gebruikte notatie zullen causale systemen overeenkomen met bovendriehoeks-matrices. Het aantal in- en uitgangen van een systeem hoeft niet noodzakelijk constant te zijn, met als gevolg dat de operator in feite een blok-matrixrepresentatie heeft, waarbij de matrixelementen op hun beurt weer matrices kunnen zijn, mogelijk met onderling verschillende dimensies.

Als speciaal geval beschouwen we de situatie waarbij alle behalve een eindig aantal van deze dimensies nul zijn, zodat de operator reduceert tot een eindige matrix. Het toepassen van het systeem op een signaal correspondeert met een matrix-vector vermenigvuldiging, zodat op deze manier een verband gelegd kan worden tussen tijdvariërende systeemtheorie en lineaire algebra. Het blijkt dat, voor een bepaalde klasse van gestructureerde matrices, dit verband met de systeemtheorie leidt tot efficiënte rekenmethoden voor problemen uit de lineaire algebra.

Tijdvariërende lineaire systemen kunnen gewoonlijk beschreven worden door een toestandsmodel. Dit model kan gezien worden als een rekenschema waarmee het systeem uit een gegeven ingangssignaal het bijbehorende uitgangssignaal bepaalt; de interne tussenresultaten in de berekening van het matrix-vector produkt worden bewaard in de toestandsvariabelen van het systeem. Efficiënte rekenmethoden worden verkregen als het aantal toestandsvariabelen klein is in vergelijking met de afmetingen van de matrix zelf.

De volgende aspecten komen aan de orde.

- *Realisatie-theorie:* gegeven een begrensd en causaal systeem, bepaal een toestandsmodel van zo klein mogelijke dimensies dat dit systeem representeert. Een belangrijke rol wordt gespeeld door een veralgemenisering van de Hankel-operator.
- *Optimale modelreductie:* als de toestandsdimensie van het systeem te groot is, dient een goede benadering gevonden te worden die wel een klein aantal toestandsvariabelen toelaat. De norm waarin dit probleem wordt opgelost is een veralgemenisering van de Hankel-norm zoals ingevoerd door Adamjan, Arov en Krein. Het is mogelijk om een parametrisatie te geven van alle systemen die dichter dan een gegeven tolerantie van de originele matrix afliggen, en om van een bepaalde benadering met minimale toestandsdimensies een expliciet toestandsmodel te vinden.
- *Inner-outer factorisatie, spectrale factorisatie, en embedding in een verliesvrij systeem* spelen een belangrijke rol in de tijdsinvariante systeemtheorie bij het oplossen van allerhande problemen, zoals bijvoorbeeld het 'robust-control' probleem. Het is mogelijk om (onder condities) deze factorisaties ook te bepalen voor tijdvariërende systemen, en diverse algoritmen om dit daadwerkelijk te doen worden afgeleid. De algoritmen werken met toestandsmodellen en geven aanleiding tot Riccati-vergelijkingen met tijdvariërende coëfficiënten. Over deze vergelijkingen is nog niet veel bekend (in tegenstelling tot het tijdsinvariante geval); enige eigenschappen worden aangetoond.
- *Cascade-factorisaties van inner systemen.* Een numeriek stabiele implementatie van een tijdvariërend toestandsmodel wordt verkregen door een cascade-factorisatie in elementaire (eerste-graads) overdrachts-operatoren.

Eindige (blok-)matrices zijn speciale gevallen van tijdvariërende overdrachtsoperatoren, en de bovenstaande resultaten geven aanleiding tot nieuwe rekenschema's voor problemen uit de lineaire algebra. Zoals hierboven al is opgemerkt, levert een toestandsmodel een efficiënt rekenschema voor het bepalen van matrix-vector vermenigvuldigingen. De inner-outer factorisatie blijkt te reduceren tot een QR-factorisatie, en spectrale factorisatie is gerelateerd aan Cholesky decomposities. Tenslotte levert de theorie voor optimale Hankel-norm modelreductie nieuwe matrixbenaderingen op die geschikt zijn voor deze efficiënte rekenschema's.

BIOGRAPHY

Alle-Jan van der Veen was born in The Netherlands in 1966. He graduated (cum laude) from Delft University of Technology, Department of Electrical Engineering, in 1988. As part of the studies, he spent three months at Philips Research Laboratories, Eindhoven, The Netherlands, participating in a computer graphics project. His master's thesis research was performed at the Network Theory section, Department of Electrical Engineering, Delft University of Technology, with Dr. Ed Deprettere as advisor.

In 1989, he started his work towards a Ph.D. in the same section, with Prof. Patrick Dewilde as principal advisor. Part of the research was performed during a three-month visit to the group of Prof. Harry Dym at the Department of Theoretical Mathematics, Weizmann Institute of Science, Israel, in the summer of 1990.

His research interests are in the general area of system theory, in particular system identification, model reduction and time-varying system theory, and in numerical methods and parallel algorithms for linear algebra problems. In 1989, he was chairman of the IEEE student branch Delft. He has organized two workshops in the area of signal processing, and is the co-editor of the book *Algorithms and Parallel VLSI Architectures.*

Publications

Books

[1] A.J. van der Veen and M.A.J. Bloemendaal, eds., *IEEE Proceedings Symposium on Computer Architecture & Real Time Graphics.* Delft, The Netherlands, 1989.

[2] E.F. Deprettere and A.J. van der Veen, eds., *Algorithms and Parallel VLSI Architectures*, vol. A and B. Elsevier, 1991.

Journal articles

[3] A.J. van der Veen and Ed. F. Deprettere, "Parallel VLSI matrix pencil algorithm for high resolution direction finding," *IEEE Trans. Signal Processing*, vol. 39, pp. 383–394, Feb. 1991.

[4] A.J. van der Veen, P.B. Ober, and E.F. Deprettere, "Azimuth and elevation computation in high resolution DOA estimation," *IEEE Trans. Signal Processing*, vol. 40, pp. 1828–1832, July 1992.

[5] A.J. van der Veen, E.F. Deprettere, and A.L. Swindlehurst, "Subspace based signal analysis using singular value decomposition," *to appear in Proceedings of the IEEE*, 1993.

[6] P. Dewilde and A.J. van der Veen, "On the Hankel-norm approximation of upper-triangular operators and matrices," *to appear in Integral Equations and Operator Theory*, 1993.

[7] A.J. van der Veen and P.M. Dewilde, "Embedding of time-varying contractive systems in lossless realizations," *subm. to Math. Control Signals Systems*, July 1992.

[8] A.J. van der Veen and P.M. Dewilde, "On low-complexity approximation of matrices," *subm. to Linear Algebra and its Applications*, June 1992.

[9] A.J. van der Veen and P.M. Dewilde, "Time-varying computational networks: Realization, orthogonal embedding and structural factorization," *subm. to Integration, the VLSI journal*, Sept. 1992.

[10] A.J. van der Veen and M.G. Verhaegen, "On spectral factorization and Riccati equations for time-varying systems in discrete time," *submitted to IEEE Trans. Automat. Control*, Feb. 1993.

Chapters in books

[11] A.J. van der Veen and E.F. Deprettere, "SVD-based low-rank approximations of rational models," in *SVD and Signal Processing: Algorithms, Analysis and Applications* (R.J. Vaccaro, ed.), vol. II, pp. 431–454, Elsevier, 1991.

[12] A.J. van der Veen, Ed. F. Deprettere, and A.L. Swindlehurst, "SVD-based estimation of low-rank system parameters," in *Algorithms and Parallel VLSI Architectures* (Ed. F. Deprettere and A.J. van der Veen, eds.), vol. A, pp. 203–228, Elsevier, 1991.

[13] A.J. van der Veen and P.M. Dewilde, "Time-varying system theory for computational networks," in *Algorithms and Parallel VLSI Architectures, II* (P. Quinton and Y. Robert, eds.), pp. 103–127, Elsevier, 1992.

[14] P.M. Dewilde and A.J. van der Veen, "Reduction and approximation of linear computational circuits," in *NATO ASI series, ser. E, vol. 232* (M. Moonen et al., ed.), pp .109–135, Sept. 1992.

[15] A.J. van der Veen, "Computation of the inner-outer factorization for time-varying systems," in *Challenges of a Generalized System Theory* (M. Verhaegen et al., ed.), Essays of the Royal Dutch Academy of Sciences, 1993.

Conference proceedings

[16] A.J. van der Veen and Ed. F. Deprettere, "A parallel VLSI matrix algorithm for high resolution direction finding," in *Proc. SPIE, "Advanced Algorithms and Architectures for Signal Processing III"* (F.T. Luk, ed.), vol. 975, San Diego, CA, pp. 289–299, 1988.

[17] A.J. van der Veen, "Intersection test for NURBS," in *Proc. IEEE Symp. on Computer Architecture & Real Time Graphics* (A.J. van der Veen and M.A.J. Bloemendaal, eds.), Delft, The Netherlands, pp. 101–114, 1989.

[18] A.J. van der Veen and P.M. Dewilde, "Orthogonal embedding theory for contractive time-varying systems," in *Recent Advances in Mathematical Theory of Systems, Control, Networks and Signal Processing (Proc. Int. Symp. MTNS-91)* (H. Kimura and S. Kodama, eds.), vol. II, pp. 513–518, MITA Press, Japan, 1992.

[19] A.J. van der Veen and P.M. Dewilde, "Orthogonal embedding theory for contractive time-varying systems," in *Proc. IEEE ISCAS*, (San Diego, CA), pp. 693–696, May 1992.

[20] A.J. van der Veen and P.M. Dewilde, "Time-varying computational networks: Realization, orthogonal embedding and structural factorization," in *Proc. SPIE, "Advanced Signal Processing Algorithms, Architectures, and Implementations III"* (F.T. Luk, ed.), vol. 1770, (San Diego, CA), pp. 164–177, July 1992. Also presented at SIAM 40th anniversary meeting (Los Angeles, July 1992).

[21] A.J. van der Veen and P.M. Dewilde, "AAK model reduction for time-varying systems," in *Eusipco-92 Conference* (J. Vandewalle e.a., ed.), pp. 901–904, Elsevier Science Publishers, Aug. 1992.

[22] A.J. van der Veen and P.M. Dewilde, "AAK model reduction for time-varying systems," in *IEEE 31st Conf. on Decision and Control*, (Tucson, AZ), pp. 3076–3081, Dec. 1992.

[23] A.J. van der Veen and P.M. Dewilde, "Parametrization of Hankel-norm approximants of time-varying systems," in *Proc. Int. Symposium MTNS-93*, (Regensburg, Germany), July 1993.

[24] A.J. van der Veen and P.M. Dewilde, "Generalized systems theory and computational linear algebra," in *Proc. ECCTD*, (Davos, Switzerland), Aug. 1993.

GLOSSARY OF NOTATION

Diagonal algebra

$\mathcal{N} = \mathbb{C}^N$: space of (non-uniform) sequences with i-th entry in $\mathbb{C}^{N_i}$ (p. 28).

$N = \#\mathcal{N}$: the sequence of dimensions of $\mathcal{N}$ (p. 28).

$\ell_2^{\mathcal{N}}$: space of bounded (non-uniform) sequences in $\mathcal{N}$ (p. 29).

$\mathcal{X}(\mathcal{M}, \mathcal{N})$: space of bounded operators $\ell_2^{\mathcal{M}} \to \ell_2^{\mathcal{N}}$ and $\mathcal{X}_2^{\mathcal{M}} \to \mathcal{X}_2\mathcal{N}$ (p. 30).

$\mathcal{U}, \mathcal{L}, \mathcal{D}$: upper/lower/diagonal bounded operators in $\mathcal{X}$ (p. 30).

$\mathcal{X}_2, \mathcal{U}_2, \mathcal{L}_2, \mathcal{D}_2$: (Hilbert) spaces of operators in $\mathcal{X}, \mathcal{U}, \mathcal{L}, \mathcal{D}$ with bounded HS-norm (p. 32).

π_i: sequence constructor. $A \in \mathcal{X}$ has entries $A_{ij} = \pi_i A \pi_j^*$ (p. 28).

Z: bilateral causal shift operator (p. 34).

$T^{(k)}$: diagonal shift of $T \in \mathcal{X}$ over k positions into south-west direction (p. 35).

$r(X)$: spectral radius of X (p. 32).

$\mathbf{P}_{\mathcal{H}}$: projection onto a subspace $\mathcal{H} \subset \mathcal{X}_2$ (p. 33).

$\mathbf{P}, \mathbf{P}_0$: projection onto $\mathcal{U}_2$, $\mathcal{D}_2$ (p. 33).

$\{A, B\}$ $= \mathbf{P}_0(AB)$: diagonal inner product (p. 38).

$[A, B]$ $= \mathbf{P}_0(AJB^*)$: indefinite diagonal inner product (p. 153).

$A \gg 0$: A is uniformly strictly positive definite (p. 39).

$A^{\{k\}}$ $= A^{(k)}A^{(k-1)} \cdots A^{(1)}$ (p. 35).

$A^{[k]}$ $= AA^{(1)} \cdots A^{(k-1)}$ (p. 34).

$T_{[k]}$ $= \mathbf{P}_0(Z^{-k}T)$: the k-th diagonal above the main (0-th) diagonal of T (p. 36).

$\Lambda_{\mathbf{F}}$ $= \mathbf{P}_0(\mathbf{F}\mathbf{F}^*)$: the Gram operator associated to a basis representation $\mathbf{F}$ (p. 45).

s-dim$(\cdot)$: the sequence of dimensions of a left D-invariant subspace (p. 41).

$\ell_2^{\mathcal{D}}$: the space of bounded sequences with entries in $\mathcal{D}$ (p. 48).

$\tilde{U}$: the diagonal expansion of U to a sequence in $\ell_2^{\mathcal{D}}$. For operators: the associated operator, via $UT = Y \Leftrightarrow \tilde{U}\tilde{T} = \tilde{Y}$ (p. 48).

$(\cdot)^\dagger$: the pseudo-inverse (Moore-Penrose generalized inverse) (p. 252).

System theory

$\mathbf{T}$: realization matrix. $\mathbf{T} = \{A, B, C, D\}$ stands for the matrix $\mathbf{T} = \left[\begin{smallmatrix} A & C \\ B & D \end{smallmatrix}\right]$ (p. 56).

ℓ_A: the spectral radius of AZ (p. 57).

$\mathcal{H}(T), \mathcal{H}_0(T), \mathcal{K}(T), \mathcal{K}_0(T)$: input state space, output state space, input null space, output null space of an operator $T \in \mathcal{X}$ (p. 79).

$\mathbf{Q}, \mathbf{F}, \mathbf{G}, \mathbf{F}_0$: typically, $\mathbf{Q}$ and $\mathbf{G}$ are orthonormal basis representations of the input and output state space. $\mathbf{F}$, $\mathbf{F}_0$ are strong basis representations of these spaces (p. 86 *ff.*).

$\mathcal{C}, \mathcal{O}$: controllability, observability operators (diagonal expansions of $\mathbf{F}$ and $\mathbf{F}_0$) (p. 72).

H_T, K_T, E_T: the operator T on restricted domains and ranges. $H_T : \mathcal{L}_2\mathcal{Z}^{-1} \to \mathcal{U}_2$ is the Hankel operator. $K_T : \mathcal{L}_2\mathcal{Z}^{-1} \to \mathcal{L}_2\mathcal{Z}^{-1}$, $E_T : \mathcal{U}_2 \to \mathcal{U}_2$ (p. 69).

$T_\Theta[S_L]$ $= (\Theta_{11}S_L - \Theta_{12})(\Theta_{22} - \Theta_{21}S_L)^{-1}$ (p. 156).

INDEX