Semi- and Quasi-separable Systems

36

Patrick Dewilde and Alle-Jan Van der Veen

Contents

Abstract

The main objects of this chapter are "semi-separable systems," sometimes called "quasi-separable systems." These are systems of equations, in which the operator has a special structure, called "semi-separable" in this chapter. By this is meant that the operator, although typically infinite dimensional, has a recursive structure determined by sequences of finite matrices, called transition matrices. This type

P. Dewilde (✉)
Technische Universität München, Institute for Advanced Study, München, Germany
e-mail: p.dewilde@me.com

A.-J. Van der Veen
Circuits and Systems Section, Delft University of Technology, Delft, The Netherlands
e-mail: a.j.vanderveen@tudelft.nl

D. Alpay (ed.), *Operator Theory*,
DOI 10.1007/978-3-0348-0667-1_52

of operator occurs commonly in Dynamical System Theory for systems with a finite dimensional state space and/or in systems that arise from discretization of continuous time and space. They form a natural generalization of finite matrices and a complete theory based on sequences of finite matrices is available for them. The chapter concentrates on the invertibility of such systems: either the computation of inverses when they exist, or the computation of approximate inverses of the Moore–Penrose type when not. Semi-separable systems depend on a single principal variable (often identified with time or a single dimension in space). Although there are several types of semi-separable systems depending on the continuity of that principal variable, the present chapter concentrates on indexed systems (so-called discrete-time systems). This is the most straightforward and most appealing type for an introductory text. The main workhorse is "inner–outer factorization," a technique that goes back to Hardy space theory and generalizes to any context of nest algebras, as is the one considered here. It is based on the definition of appropriate invariant subspaces in the range and co-range of the operator. It translates to attractive numerical algorithms, such as the celebrated "square-root algorithm," which uses proven numerically stable operations such as QR-factorization and singular value decomposition (SVD).

Introduction

What is the inverse of the (singly infinite dimensional) lower bi-diagonal (so-called Toeplitz) matrix

$$\begin{bmatrix} 1 & & & \\ -1/2 & 1 & & \\ & -1/2 & 1 & \\ & & \ddots & \ddots \end{bmatrix} ? \tag{36.1}$$

Using analogy to the inversion of doubly infinite Toeplitz matrices and their well-known z-transform theory, one easily finds for the inverse (a direct check is immediate):

$$\begin{bmatrix} 1 & & & & \\ 1/2 & 1 & & & \\ 1/4 & 1/2 & 1 & & \\ 1/8 & 1/4 & 1/2 & 1 & \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} . \tag{36.2}$$

What about

$$\begin{bmatrix} 1 & & & \\ -2 & 1 & & \\ & -2 & 1 & \\ & & \ddots & \ddots \end{bmatrix} ? \tag{36.3}$$

It turns out that this matrix is not invertible, it has a co-kernel found by left multiplication with $\left[1\ 1/2\ 1/4 \cdots\right]$, as can be checked directly. Yet, it also has a (bounded) left inverse given by

$$\begin{bmatrix} 0 & -1/2 & -1/4 & -1/8 & \cdots \\ 0 & 0 & -1/2 & -1/4 & \ddots \\ 0 & 0 & 0 & -1/2 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}. \tag{36.4}$$

The matrix actually has a nice Moore–Penrose pseudo-inverse (given at the end of this chapter). So far, the examples just given are all half-infinite Toeplitz (meaning elements on diagonals are equal), but this turns out not be essential at all, the only really important thing about these infinite-dimensional matrices is boundedness. This chapter deals with matrices that represent operators between "ℓ_2" spaces. Also, scalar entries are not important, all entries can be matrices, provided dimensions remain consistent, i.e., all matrices on the same (block-)row must have the same row dimension, and all matrices on the same (block-)column must have the same column dimension. Dimensions may change from row to row or column to column. In this way a sequence of indices arises: $\mathbf{m} := \{m_k\}_{k=-\infty:\infty}$ (using a MATLAB-like notation) for the columns and $\mathbf{n} := \{n_k\}_{k=-\infty:\infty}$ for the rows, the matrix in position (j,k) having dimensions $n_j \times m_k$. Zero dimensions are allowed (in that case the entry at that index point just disappears) and the indexing may run from $-\infty$ to $+\infty$. In the case of doubly infinitely indexed objects, one needs to identify the entry of index zero, which one does with a surrounding box: $\boxed{T_{0,0}}$ for the entry with indices $(0,0)$ in a doubly infinite operator matrix T. Typically, a bounded operator T will map an "input space" $\ell_2^{\mathbf{m}}$ to an "output" space $\ell_2^{\mathbf{n}}$, where $\ell_2^{\mathbf{m}}$ is, e.g., the natural Hilbert space of real or complex sequences of type $\{u_k\}_{k=-\infty:\infty}$ with $u_k \in \mathcal{R}^{m_k}$ (respect. $\in \mathcal{C}^{m_k}$) and $\mathcal{R}$ the real (respect. $\mathcal{C}$ the complex) numbers. Matrix transpose (respect. hermitian transpose) is denoted with an accent: $[A']_{j,k} = A'_{k,j}$. Zero-dimensional indices indicate just a "place-holder" at the respective index. Some new calculus rules with zero-indexed entries consistent with regular matrix calculus is therefore called for. A zero-row, one column matrix is denoted by a horizontal dash (—), while a zero-column, one-row matrix is represented by a vertical dash (|) and a zero row, zero column matrix by a dot (·). (New) *multiplication rules with dashes* then work as follows (":=" is used throughout to define a quantity, "×" to indicate multiplication explicitly):

$$| \times - := [0], \quad - \times | := \cdot \tag{36.5}$$

With these simple rules, finite and half infinitely indexed matrices are naturally embedded in doubly infinitely indexed ones. The following sections will soon demonstrate the necessity for such conventions.

Semi- or Quasi-Separability

A lower block-triangular system of equations $Tu = y$ with bounded operator $T \in \ell_2^{\mathbf{m}} \to \ell_2^{\mathbf{n}}$ is *semi-separable* iff there exist a series of indices $\mathbf{b} = \{b_k\}_{k=-\infty}^{\infty}$, a uniformly bounded sequence of (complex) vectors $x_k \in \mathcal{C}^{b_k}$ (or $\mathcal{R}^{b_k}$ in case of real arithmetic) and sequences of uniformly bounded matrices $\{A_k, B_k, C_k, D_k\}$ such that the following recursion holds for all indices k:

$$\begin{cases} x_{k+1} = A_k x_k + B_k u_k \\ \quad y_k \;\; = C_k x_k + D_k u_k \end{cases} \tag{36.6}$$

or in matrix notation:

$$\begin{bmatrix} x_{k+1} \\ y_k \end{bmatrix} = \begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} \begin{bmatrix} x_k \\ u_k \end{bmatrix}. \tag{36.7}$$

This is called a *(causal) state space realization* of the operator T, with state transition matrix A_k, input operator B_k, output operator C_k and feed-through D_k. All these matrices have variable dimensions depending on the sequences $\mathbf{m}, \mathbf{n}$, and $\mathbf{b}$. An issue is whether the recursive representation actually defines a bounded operator. Sufficient for this (but there are important exceptions, see further) is that not only the matrices A_k, B_k, C_k, D_k are uniformly bounded but that the sequence of the so-called state transitions A_k is also *uniformly exponentially stable (denoted u.e.s.)*, i.e., that for k,

$$\sigma = \limsup_k [\sup_{\ell} \|A_{k+\ell-1} \cdots A_{\ell+1} A_\ell)\|]^{1/k} < 1 \tag{36.8}$$

i.e., the continuous product $A_{k+\ell-1} \cdots A_{\ell+1} A_\ell$ gets eventually majorized by $\sigma^{k+\epsilon}$ in norm for any small $\epsilon > 0$, uniformly over ℓ.

The operator T then has the matrix representation

$$T = \begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \ddots & C_{-1}A_{-2}B_{-3} & C_{-1}B_{-2} & D_{-1} & 0 & 0 & \ddots \\ \ddots & C_0A_{-1}A_{-2}B_{-3} & C_0A_{-1}B_{-2} & C_0B_{-1} & \boxed{D_0} & 0 & \ddots \\ \ddots & C_1A_0A_{-1}A_{-2}B_{-3} & C_1A_0A_{-1}B_{-2} & C_1A_0B_{-1} & C_1B_0 & D_1 & \ddots \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{36.9}$$

the general term of which is, for $j > k$: $T_{j,k} = C_j A_{j-1} \cdots A_{k+1} B_k$. In this term the increasing continuous product of state-transition matrices appears, which is sometimes denoted as $A_{j,k}^{>\times} := A_{j-1} \cdots A_{k+1}$. This chapter adopts a different and more compact notation: for a sequence of matrices $\{A_k\}_{k=-\infty\cdots\infty}$ the *constructor* "diag" threads them into a block-diagonal operator: $A = \text{diag}[A_k]$. Let, in addition, Z be the forward or "causal" shift: $(Zx)_k := x_{k-1}$, with of course $(Z^{-1}x)_k = x_{k+1}$, then the global, now anti-causal, state-space equations simply become:

$$\begin{cases} Z^{-1}x = Ax + Bu \\ \quad y \quad = Cx + Du \end{cases}. \tag{36.10}$$

These equations can formally be solved to produce $T = D + C(I - ZA)^{-1}ZB$, a form that certainly makes sense when the operator $(I - ZA)$ is bounded invertible. One verifies that this is the case when A is u.e.s., by Neumann series expansion (σ is the spectral radius of ZA). However that be, one may always write T as a unilateral expansion of diagonals: $T = D + CZB + CZAZB + \cdots$, the general term $C(ZA)^{k-2}ZB,\ k \geq 2$ of which defines the kth sub-diagonal of T with a finite product, an expression that makes sense whenever the matrix representation of T does – a strategy that can be used to represent unbounded or numerically unstable operators.

The shift operator Z does not normally commute with other operators. Let $T^{<+1>} := ZTZ^{-1}$ denote the *diagonal* shift in the South-East direction. Then $ZT = T^{<+1>}Z$. Similarly, $T^{<-1>} = Z^{-1}TZ$ is a diagonal upward shift. A word of caution: the dimensions of Z are variable, e.g., one has $Z_{k+1,k} = I_{m_k}$ in the product Zx with $x \in \mathcal{R}^{\mathbf{m}}$, and $Z_{j,k} = 0$ for $j \neq k + 1$ (the symbol Z actually represents a collection of operators). In the present theory, adjoints will coincide with matrix transposition (for real matrices), or hermitian transposes (for complex matrices).

Since operators act on a Hilbert space, they have adjoints. Abstract operator adjoints are typically denoted by a $\cdot^*$. As most operations in this chapter are matrix operations, there is no need here to consider more general adjoints and most operations are in real arithmetic, the notion of matrix transpose suffices. It is simply denoted with a prime: $[T']_{j,k} = T'_{j,k}$ – in particular $Z' = Z^{-1}$. For the case of complex arithmetic, the prime denotes the hermitian conjugate. (The theory is even valid for more general fields, but that will not be of concern in this chapter.) Upper operators are dual to lower operators, and in a similar vein as before, an upper semi-separable operator has a representation

$$\begin{cases} x_{k-1} = A_k x_k + B_k u_k \\ \quad y_k \quad = C_k x_k + D_k u_k \end{cases}, \quad \begin{cases} Zx = Ax + Bu \\ \ y \ = Cx + Du \end{cases}, \tag{36.11}$$

in which case (upper or anti-causal) $T = D + C(I - Z'A)^{-1}Z'B$ (notice that the "incoming" state has index k in both the lower and the upper realization, hence takes place at different locations). This leads to the final definition:

Definition 1. A semi-separable operator $T : \ell_2^{\mathbf{m}} \to \ell_2^{\mathbf{n}}$ is a (bounded) operator that possesses (potentially different) state-space realizations for its lower (causal) and its upper (anti-causal) part:

$$T = C_c(I - ZA_c)^{-1}ZB_c + D + C_a(I - Z'A_a)^{-1}Z'B_a \tag{36.12}$$

in which the operators $\{A_c, B_c, C_c, D, A_a, B_a, D_a\}$ (sometimes called "generators") are all bounded block diagonal operators and the inverses in the expression are unilateral expansions in respect. Z and Z'. It is called *u.e.s.* when A_c as well as A_a are u.e.s.

The boundedness conditions stated in the definition can be relaxed, but that goes beyond the present chapter. Finite (block-)matrices are automatically semi-separable (see the next section and the notes at the end), but for them the definition only makes sense when the respective state dimensions are small compared to the overall dimension of the matrix.

Realization Theory

Many state-space realizations are possible for a given semi-separable transfer operator T. An important class of realizations are the *minimal*. This is obtained when the state dimension at each index point is minimal. Realization theory (which is only summarized here) states that this minimal dimension is actually equal to the rank of the so-called Hankel operator at that index point. An arbitrary minimal factorization of that Hankel operator produces moreover a specific realization. Here is how that works.

Definition 2. Let T be a lower semi-separable operator. Its kth Hankel operator is the matrix

$$H_k = \begin{bmatrix} \text{‘}T_{k,k-1} & T_{k,k-2} & T_{k,k-3} & \cdots \\ T_{k+1,k-1} & T_{k+1,k-2} & T_{k+1,k-3} & \cdots \\ T_{k+2,k-1} & T_{k+1,k-2} & T_{k+1,k-3} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \tag{36.13}$$

H_k maps the "strict past" of the input vector to the "present and future" of the output vector at index point k. In any realization one has

$$H_k = \begin{bmatrix} C_k \\ C_{k+1}A_k \\ C_{k+2}A_{k+1}A_k \\ \vdots \end{bmatrix} \begin{bmatrix} B_{k-1} & A_{k-1}B_{k-2} & A_{k-1}A_{k-2}B_{k-3} & \cdots \end{bmatrix} \tag{36.14}$$

hence H_k factors into a *reachability operator*

$$\mathbf{R}_k := \begin{bmatrix} B_{k-1} & A_{k-1}B_{k-2} & A_{k-1}A_{k-2}B_{k-3} & \cdots \end{bmatrix} \tag{36.15}$$

and an *observability operator* (using the "col" constructor that makes a column out of the list of matrices)

$$\mathbf{O}_k := \text{col}\left\{ C_k \ C_{k+1}A_k \ C_{k+2}A_{k+1}A_k \ \cdots \right\}. \tag{36.16}$$

The converse works equally well and is the basis for realization theory: every (reasonably bounded) factorization of the Hankel operator will produce a realization. Minimal realizations are obtained when each factorization (i.e., for each index k) is minimal, i.e., when the rows of each $\mathbf{R}_k$ and the columns of each $\mathbf{O}_k$ form a basis (are linearly independent). The columns of $\mathbf{O}_k$ then form a basis for the range of H_k, while the transpose of the rows of $\mathbf{R}_k$ form a basis for the co-range of H_k – i.e., the range of H_k'. It also follows that the vectors in any such base belong to ℓ_2 (of appropriate dimensions) and form bounded operators $\mathbf{O}_k$ and $\mathbf{R}_k'$ with closed range and zero co-kernel.

From any minimal factorization one can derive a realization, as follows. Using a Matlab-like notation to single out sub-matrices, one chooses $B_{k-1} = [\mathbf{R}_k]_1$ and $C_k = [\mathbf{O}_k]_1$. Furthermore:

$$[\mathbf{O}_k]_{2:\infty} = [\mathbf{O}_{k+1}]A_k \tag{36.17}$$

and, if the columns of $\mathbf{O}_{k+1}$ form a basis, then it has a (actually many) left bounded (pseudo-) inverse $\mathbf{O}_{k+1}^{\dagger}$ (one can take $\mathbf{O}_{k+1}^{\dagger} = (\mathbf{O}_{k+1}'\mathbf{O}_{k+1})^{-1}\mathbf{O}_{k+1}'$) and one must have $A_k = \mathbf{O}_{k+1}^{\dagger}[\mathbf{O}_k]_{2:\infty}$. It turns out that this definition of A_k is actually independent of the choice of left pseudo-inverse, and whether one has worked on the observability or reachability operators.

Canonical Forms

In particular, one can select an orthonormal basis for all the observability operators, and put $A_k = \mathbf{O}_{k+1}'[\mathbf{O}_k]_{2:\infty}$. In that case each $\begin{bmatrix} A_k \\ C_k \end{bmatrix}$ is isometric (i.e., $A_k'A_k + C_k'C_k = I$) for all k. The realization is then in *output normal form*. Dually, one can choose an orthonormal basis for each reachability operator, in which case the realization will be in *input normal form* and $\begin{bmatrix} A_k & B_k \end{bmatrix}$ is co-isometric for all k (i.e., $A_kA_k' + B_kB_k' = I$). Another interesting form is when a realization is chosen *balanced*. To obtain it, one performs a reduced Singular Value Decomposition of each Hankel ($H_k = U_k\Sigma_kV_k'$) with U_k and V_k isometric and Σ_k square non-singular, and puts $\mathbf{O}_k := U_k\Sigma_k^{1/2}$, $\mathbf{R}_k := \Sigma_k^{1/2}V_k'$. Corresponding to these choices of basis, there are gramians, which typically are then called *observability*, respect. *reachability* gramians. In the case of the balanced realization, both are diagonal and equal to Σ_k at index k. Minimal realizations are both *reachable* and *observable*. Reachability means that any state x_k can be generated by an input in the strict past of the system, i.e., some u_p in $\ell_2^{\mathbf{m}_{-\infty:k-1}}$. Observability, on the other hand, means that there is a one-to-one relation between a state x_k and the zero-input future response $y_f \in \ell_2^{\mathbf{n}_k:\infty}$ it produces (alternatively, it is reachability of the adjoint system realization).

State Equivalence

All minimal realizations define bases for both the reachability and observability operators at each point k. As a result, they all relate to each other via a basis transformation, which is actually a basis transformation on the state. Let $x_k = R_k \hat{x}_k$ be such a transformation with each R_k square non-singular, then the realization (of a lower system) transforms as

$$\begin{bmatrix} A_k & B_k \\ C_k & D_k \end{bmatrix} \mapsto \begin{bmatrix} R_{k+1}^{-1} A_k R_k & R_{k+1}^{-1} B_k \\ C_k R_k & D_k \end{bmatrix}. \tag{36.18}$$

One can of course use such a transformation to convert a system to any of the canonical forms described above. In particular, if one has a realization with reachability data $\begin{bmatrix} A_k & B_k \end{bmatrix}$, which one wants to convert to input normal form, then one has to find R_k's such that the transformed realization has $\begin{bmatrix} \hat{A}_k & \hat{B}_k \end{bmatrix} := \begin{bmatrix} R_{k+1}^{-1} A_k R_k & R_{k+1}^{-1} B_k \end{bmatrix}$ co-isometric. Putting $M_k := R_k R_k'$, this means finding (non-singular) M_k's such that

$$M_{k+1} = A_k M_k A_k' + B_k B_k'. \tag{36.19}$$

This is a famous forward-recursive *Lyapunov–Stein equation*, and it will have a numerically stable solution when the operator A is u.e.s. All M_k will be non-singular, provided the original system is reachable, because M_k is actually the reachability gramian of the original realization at index k. It is numerically not advisable to solve the Lyapunov–Stein equation directly, because the numerical conditioning of M is square that of R. A direct method to compute the R_k is the so-called square-root algorithm, which in this case is the recursion:

$$\begin{bmatrix} A_k R_k & B_k \end{bmatrix} = \begin{bmatrix} R_{k+1} & 0 \end{bmatrix} V_k, \tag{36.20}$$

in which R_{k+1} is square non-singular and V_k an orthogonal matrix (the columns of R_{k+1} form a basis for the range of $\begin{bmatrix} A_k R_k & B_k \end{bmatrix}$). The recursion assumes knowledge of R_k and then computes R_{k+1} and V_k by column reduction. This is an example of a so-called R-Q factorization; R_{k+1} can typically be obtained either in lower triangular or in upper triangular form, and because of the minimality conditions, it is guaranteed to be square non-singular (in the case of balanced realizations one would resort to SVDs). The unknown R_{k+1} and orthogonal matrix V_k are computed from the left-hand side data (this is "array processing": a lot of new data directly computed from a source, without a closed mathematical formula). As an added benefit, V_k contains the new reachability data, i.e.,

$$V_k = \begin{bmatrix} \hat{A}_k & \hat{B}_k \\ C_{V,k} & D_{V,k} \end{bmatrix} \tag{36.21}$$

in which $C_{V,k}$ and $D_{V,k}$ complete the orthogonal matrix (see further the discussion on canonical factorizations for their significance). The transformed realization for T at stage k is then $\begin{bmatrix} \hat{A}_k & \hat{B}_k \\ C_k R_k & D_k \end{bmatrix}$, which, with $\hat{C}_k = C_k R_k$ and using the diagonal notation, can be written globally as a matrix of diagonal operators $\begin{bmatrix} \hat{A}_k & \hat{B}_k \\ \hat{C}_k & D_k \end{bmatrix}$, with $T = D + \hat{C}(I - Z\hat{A})^{-1} Z\hat{B}$ as well. Several issues are now in order.

First, there is the tricky question of the boundedness of R and R^{-1}. For good results, the global operator $R = \text{diag} R_k$ should be restricted to being bounded with bounded inverse, whenever possible. This is achieved by requiring the existence of semi-separable realizations in which both the reachability and the observability gramians are strictly positive (i.e., the inverse M^{-1} of the respective gramian M is bounded). In that case, both the input and the output normal forms of the system at hand have state transition matrices that are u.e.s. This is certainly not always the case and is important for how the system behaves at infinity. Under the condition of a strictly positive reachability gramian, there exists a semi-separable output normal form with state transition matrix $\hat{A}$ that is u.e.s. (and conversely). Dually, the output normal form representation will also possess a state transition matrix that is u.e.s. iff the observability gramian is strictly positive definite.

Next, there is the issue of starting the recursion, in the case of the input normal form discussed so far, the recursion goes forward (from k to $k+1$). An initial value is needed. This requires some knowledge of the system around $-\infty$. For example, the system may be originally time-invariant, in which case there is a fixed-point solution to the recursion that can be obtained directly. In many cases the behavior at earlier times is unknown. One may then assume an arbitrary initial value to start up the recursion. It turns out that because of the u.e.s. property, the error made will die out, at a rate given by σ^k (see the u.e.s. definition (36.8)). This is true for numerical errors made during the computation as well, both the Lyapunov–Stein and the square-root recursion are extremely stable numerically. The counterpart of this is that the Lyapunov–Stein equation cannot be inverted: the inversion will be extremely unstable and will produce incorrect results. The observability recursion starts at $+\infty$ and runs backwards (from k to $k-1$). Also this recursion can not be reversed, for the same numerical stability reasons, now in reverse order.

The operator $V = D_V + C_V(I - Z\hat{A})^{-1} Z\hat{B}$ has a unitary realization with $\hat{A}$ u.e.s. (which will be the case if the original operator has a uniformly reachable realization that is also u.e.s). One shows easily that V is then a (global) unitary operator as well. The converse is also true: a unitary and lower semi-separable operator has a unitary realization with A u.e.s. One word of caution here: the qualification $\hat{A}$ u.e.s. is essential. It is easy to produce unitary realizations that do not lead to a unitary operator, but this is only possible with state transition matrices that are not u.e.s.

Canonical (Co-prime) External Forms

Let T and V be as in the previous subsection, and consider the product

$$TV' = (D + \hat{C}(I - Z\hat{A})^{-1}Z\hat{B}) \times (D_V' + \hat{B}'Z'(I - \hat{A}'Z')^{-1}C_V'). \tag{36.22}$$

One checks easily that

$$(I - Z\hat{A})^{-1}Z\hat{B}\hat{B}'Z'(I - \hat{A}'Z')^{-1} = (I - Z\hat{A})^{-1}Z\hat{A} + I + \hat{A}'Z'(I - \hat{A}'Z')^{-1} \tag{36.23}$$

(because $\hat{A}\hat{A}' + \hat{B}\hat{B}' = I$), so that

$$\begin{array}{l} TV' = (DD_V + \hat{C}C_V) + \hat{C}(I - Z\hat{A})^{-1}Z(\hat{A}C_V' + \hat{B}D_V') \\ \qquad +(\hat{C}\hat{A}' + D\hat{B}')Z'(I - \hat{A}'Z')^{-1}C_V' \end{array}. \tag{36.24}$$

Next, $\hat{A}C_V' + \hat{B}D_V' = 0$, again because of orthogonality of the realization for V, and

$$\Delta' := TV' = (DD_V' + \hat{C}C_V') + (\hat{C}\hat{A}' + D\hat{B}')Z'(I - \hat{A}'Z')^{-1}C_V' \tag{36.25}$$

turns out to be upper (anti-causal). Finally, as V is unitary, one has $T = \Delta' V (= \Delta'(V')^{-1})$, and a representation for (causal) T results as the ratio of two anti-causal operators. Such a factorization will be called a *right external factorization* – the case considered here is where the right factor is unitary. It turns out that it is also co-prime (see further the section on geometry), with as a consequence that it cannot be further reduced. Δ and V are uniquely determined by T, up to left unitary equivalence by a unitary diagonal operator (these are the units of the present theory). In the section on geometry, it will appear that V characterizes the kernel of the global Hankel operator.

Dually, the output normal form leads to an external co-prime factorization of the type $T = W\Delta_r'$ again with Δ_r' anti-causal and W unitary.

Isometric and Unitary Operators

Proposition 1. *A semi-separable causal isometric (respect. co-isometric) operator V has an isometric (respect. co-isometric) realization.*

Proof. A realization in output normal form derived from an orthonormal basis for each observability operator $\mathcal{O}_k$ already has $\begin{bmatrix} A_{Vk} \\ C_{Vk} \end{bmatrix}$ isometric. Remains to show that the resulting realization $\begin{bmatrix} A_{Vk} & B_{Vk} \\ C_{Vk} & D_{Vk} \end{bmatrix}$ is isometric as well. This follows from the isometry of V. At any index k, any input or output can be orthogonally decomposed into a strict past component $u_{p,k}$ (respect. $y_{p,k}$) with support $(-\infty : k-1]$ and a "future" component $u_{f,k}$ (respect. $y_{f,k}$ with support $[k : \infty)$: $u = u_{p,k} + u_{f,k}$ (respect. $y = y_{p,k} + y_{f,k}$). The isometry then forces $\|u_{p,k}\|^2 + \|u_{f,k}\|^2 = \|y_{p,k}\|^2 + \|y_{f,k}\|^2$ for all inputs u and $y = Vu$. Consider now an input with support $(-\infty, k+1]$, but otherwise arbitrary. At index point k, and with the given output normal form realization, it generates the state $x_{u,k}$ and at index point $k+1$, the state x_{k+1}. Let $y = Vu$. Because of the isometry of the observability operator $\mathbf{O}_k$, we have $\|x_{u,k}\|^2 = \|y_{f,k}\|^2$ and $\|x_{u,k+1}\|^2 = \|y_{f,k+1}\|^2$. Because also $u_{p,k+1} = u_{p,k} + u_k$, $\|u_{p,k+1}\|^2 = \|u_{p,k}\|^2 + \|u_k\|^2$, $y_{f,k} = y_k + y_{f,k+1}$, $\|y_{f,k}\|^2 = \|y_k\|^2 + \|y_{f,k+1}\|^2$ it follows that $\|x_{u,k}\|^2 + \|u_k\|^2 = \|x_{u,k+1}\|^2 + \|y_k\|^2$, and the state-space realization is isometric for any reachable state $x_{u,k}$ and any input u_k. As the realization is minimal, any state x_k is reachable, because the Hankel operator $H_k = \mathbf{O}_k\mathbf{R}_k$, the factorization is minimal, the co-kernel $\ker(\mathbf{R}_k') = 0$, and hence $\mathbf{R}_k$ is onto as a consequence (it being finite dimensional and hence necessarily closed). $\square$

Much more tricky is whether the resulting realization is u.e.s. A semi-separable causal unitary operator V has of course a (causal) unitary realization, and it turns out to be automatically u.e.s. The proof is pretty technical and given in the literature (see e.g., Dewilde and Van der Veen [6]). An important element in the proof is the fact that the range and co-range of a unitary operator are closed spaces. When V is merely isometric (respect. co-isometric), there is no guarantee that its range (respect. co-range) is indeed closed. When V is (causal) semi-separable and isometric, then its isometric realization is uniformly observable by construction, but there is no reason why its state transition operator A_V should be u.e.s. One shows (again a technical proof) that A_V is u.e.s. iff the range of V is closed. Suppose now that $\left[\begin{array}{c|c} A_V & B_V \\ \hline C_V & D_V \end{array}\right]$ is an isometric realization for V. Such a realization can always be completed to unitary: compute $\left[\begin{array}{c|c} C_W & D_W \end{array}\right]$ such that

$$\left[\begin{array}{c|c} A_V & B_V \\ \hline C_V & D_V \\ C_W & D_W \end{array}\right] \tag{36.26}$$

is unitary, and it will be the realization of a unitary operator $\begin{bmatrix} V \\ W \end{bmatrix}$ with $W = D_W + C_W(I - ZA_V)^{-1}ZB_V$ iff A_V is u.e.s. When A_V is not u.e.s., the resulting operator will not be unitary, even though it has a unitary realization. There is a good "physical" interpretation of what happens. When square norms on inputs, outputs, and states are interpreted as "energy," then some of it may leak to infinity. When A_V is u.e.s., then this guarantees that all inputed energy is eventually transferred to the output.

Hankel Geometry

Each Hankel operator H_k related to a lower semi-separable operator T at index point k maps $\ell_2^{\mathbf{m}-\infty:k-1}$ to $\ell_2^{\mathbf{n}_{k:\infty}}$ (in the matrix notation of Eq. (36.13) the input vector is put in reverse order so that H_k looks like a regular matrix, here the normal order of the input vector is assumed). The global Hankel map can then be viewed as the direct sum of these maps. This is consistent with the fact that information on the system's behavior is needed at each index to determine its "internal state structure" at that index from its input–output behavior. More precisely, let $\mathcal{X}_2^{\mathbf{m}} = \bigoplus_{k=-\infty}^{+\infty} \ell_2^{\mathbf{m}}$ be the space of "stacks of inputs," one for each index point, endowed with a Hilbert–Schmidt inner product (one has $U \in \mathcal{X}_2^{\mathbf{m}}$ when $U = \text{row}[u_{j,:}]_{j=-\infty:+\infty}, u_{j,:} \in \ell_2^{\mathbf{m}}$ and $\sum_{k=-\infty}^{+\infty} \|u_{j,k}\|^2 < \infty$). Each column of U provides an input, for each index point one. As inputs to the global Hankel map, one restricts the input U at index k to $\ell_2^{\mathbf{m}-\infty:k-1}$ (the strictly upper part of $\mathcal{X}_2^{\mathbf{m}}$) and the output $Y = HU$ to the lower part of $\mathcal{X}_2^{\mathbf{n}}$. Let $\mathcal{U}_2^{\mathbf{m}}$ denote the natural embedding of $\bigoplus_{k=-\infty}^{\infty} \ell_2^{\mathbf{m}-\infty:k}$ into $\mathcal{X}_2^{\mathbf{m}}$ (i.e., the upper part of $\mathcal{X}_2^{\mathbf{m}}$), then the strictly upper part of $\mathcal{X}_2^{\mathbf{m}}$, $\mathcal{U}_2^{\mathbf{m}} Z'$, carries the input space of the global Hankel operator, and it maps to $\mathcal{L}_2^{\mathbf{n}} := \bigoplus_{k=-\infty}^{\infty} \ell_2^{\mathbf{n}_{k:\infty}}$, which also naturally embeds in $\mathcal{X}^{\mathbf{n}}$. Similarly, let $\mathcal{D}_2^{\mathbf{m}} := \mathcal{U}_2^{\mathbf{m}} \cap \mathcal{L}_2^{\mathbf{m}}$ denote the diagonals in $\mathcal{X}_2^{\mathbf{m}}$.

The operator T itself extends in a natural way to stacks: (formally $TU := [Tu_{:,k}]_{k=-\infty}^{\infty}$ where $u_{:,k}$ is the input sequence of the kth system) and using (Hilbert–Schmidt) orthogonal projection operators $\Pi_\cdot$, the (embedded) global Hankel operator connected to T becomes

$$H = \Pi_{\mathcal{L}_2^{\mathbf{n}}} T \, \Pi_{\mathcal{U}_2^{\mathbf{m}} Z'} \tag{36.27}$$

mapping strictly upper stacks of inputs to lower stacks of outputs. The interesting (geometric) properties of H concern its kernel, range, co-kernel, and co-range (the latter being the kernel and the range of H'). Consider first the kernel $\mathcal{K}$. Let D be an arbitrary bounded diagonal operator (consisting of scalar elements), if $U \in \mathcal{K}$, then evidently also $UD \in \mathcal{K}$, one says that $\mathcal{K}$ is *right D-invariant*. Moreover, $\mathcal{K}$ is invariant for shifts Z', indeed, if $U \in \mathcal{U}_2^{\mathbf{m}}$, then also $UZ' \in \mathcal{U}_2^{\mathbf{m}}$ and $HUZ' = 0$ whenever $HU = 0$, hence $\mathcal{K}$ is *right-Z'-invariant*. Shift-invariant spaces have special properties, and that is the case even for semi-separable matrices, although they do not fit traditional algebraic structures like Hardy spaces or modules. Traditionally one likes to work with Z-invariant spaces, and the generalization of the classical Beurling–Lax theorem to the present case (it is actually an example of a *nest algebra*) is:

Theorem 1. *For any right D-Z-invariant subspace $\mathcal{K}$ of $\mathcal{L}_2^{\mathbf{m}}$ there exists an index sequence $\mathbf{k}$ with for each j $\mathbf{k}_j \leq \mathbf{m}_j$ and an isometric semi-separable operator $V \in \mathcal{L}_2^{\mathbf{k}}$ such that $\mathcal{K} = V\mathcal{L}_2^{\mathbf{k}}$.*

The construction of V in the proof of the theorem (which is in [6]) follows the classical Beurling–Lax argument: one considers the "wandering subspace" $\mathcal{K} \ominus \mathcal{K}Z$ and constructs an orthonormal basis for it.

This generalized Beurling–Lax theorem provides for a geometric interpretation of the external factorization of the previous subsection. Consider the Hankel operator H related to T, and let $\mathcal{K}$ be its kernel. As indicated before, it is a right D-Z'-invariant subspace of $\mathcal{U}_2^{\mathbf{m}} Z'$, hence there is a sequence $\mathbf{k}$ and an isometric V' such that $\mathcal{K} = V'\mathcal{U}_2^{\mathbf{k}} Z'$. It follows, because of the definition of the Hankel operator, that $TV' = \Delta'$ for some lower Δ. From the computation in the previous section we already had a unitary and lower V such that TV' is upper, it follows immediately that $V'\mathcal{U}_2^{\mathbf{m}} Z' \in \mathcal{K}$, and hence that $\mathbf{k} = \mathbf{m}$ as well.

However, the main application of the Hankel geometry is in the next section and will give the key to system inversion theory.

Inner–Outer Factorization

Let T be a lower semi-separable operator, and consider $\mathcal{M} = T\mathcal{L}_2^{\mathbf{m}}$, i.e., the range of T for lower (causal) inputs, and $\mathcal{N} := \mathcal{L}_2^{\mathbf{n}} T$. The notation $\overline{\mathcal{M}}$ indicates closure of the space $\mathcal{M}$ in the Hilbert–Schmidt metric.

Definition 3. T is right-outer (has a lower right inverse) iff $\overline{\mathcal{M}} = \mathcal{L}_2^{\mathbf{n}}$. It is left-outer (has a lower left inverse) iff $\overline{\mathcal{N}} = \mathcal{L}_2^{\mathbf{m}}$. It is outer when both are the case.

($\mathcal{M}$ and $\mathcal{N}$ are not necessarily closed!) Remark that T (lower) is right-outer iff $\ker(T') = 0$ and left-outer iff $\ker(T) = 0$.

When T is outer, then necessarily $\mathbf{n} = \mathbf{m}$ (the proof is based on arguing that D has to be square and invertible). When it is left-outer, then only $\ker(D_T) = 0$, i.e., each diagonal block D_{Tk} of T has a left inverse, but D_T may only have dense co-range. Outerness is a tricky property, because the respective spaces $\mathcal{M}$ or $\mathcal{N}$ are not necessarily closed. When $\mathcal{M}$ is actually a closed subspace, then T has a bounded right lower (pseudo-)inverse. When $\mathcal{M}$ is not closed, then one can only assert the existence of an approximate right lower (pseudo-)inverse, as the inverse only exists on the dense range of T, and is then also necessarily unbounded. In the semi-separable case, a lower semi-separable representation of such inverses exists (see further how it is computed in the section on the square-root algorithm), but it may produce an unbounded result for some inputs, and will be unstable in a weak sense (its analysis goes beyond this treatment). This situation is unavoidable: e.g., an operator such as $I - Z$ is outer with unbounded inverse. Unbounded outer inverses are very common and have important implications.

Clearly, $\mathcal{M}$ is right D-Z-invariant. Because of the generalized Beurling–Lax theorem, there is a sequence $\mathbf{k}$ and a lower, isometric V such that $\overline{\mathcal{M}} = V\mathcal{L}_2^{\mathbf{k}}$. Hence, $V\mathcal{L}_2^{\mathbf{k}} = \overline{T\mathcal{L}_2^{\mathbf{m}}}$. Let $T_o := V'T$, then $VT_o = VV'T$. It turns out that $VV'T = T$, because $T\mathcal{L}_2^{\mathbf{m}} = \mathcal{M}$, so that for all lower U's, $TU \in \mathcal{M}$ and VV'

is a projection operator on $\mathcal{M}$. Hence $VT_o = T$ on $\mathcal{L}_2^{\mathbf{m}}$. This argument extends to the full space $\mathcal{X}_2^{\mathbf{m}}$, because it is also evidently true that $\overline{T\mathcal{X}_2^{\mathbf{m}}} = V\mathcal{X}_2^{\mathbf{m}}$. Moreover, T_o will be right-outer, because $\overline{\mathcal{L}_2^{\mathbf{k}} T_o} = \overline{V'T\mathcal{L}_2^{\mathcal{M}}} = \overline{V'\mathcal{M}} = \mathcal{L}_2^{\mathbf{k}}$. This development gives rise to further definitions:

Definition 4. A lower (causal) isometric operator V is said to be left-inner ($V'V = I$). A lower (causal) co-isometric operator is said to be right-inner ($VV' = I$). A lower (causal) operator is said to be inner (equivalently: bi-inner), when it is unitary.

Such definitions are of course also valid in the context of the upper shift (here Z'), and even in more general nest algebras, but the context should always be clearly defined. The treatment in this section then leads to the next theorem.

Theorem 2. *Given a lower (causal) semi-separator operator T, then there exist a left-inner operator V and a right-outer operator T_o such that $T = VT_o$ (inner–outer factorization). These operators are uniquely defined except for a unitary diagonal left factor on T_o (right factor on V).*

The inner–outer factorization $T = VT_{or}$ already produces a pseudo-inverse $V'T_{or}^{\dagger}$ in which $T_{or}^{\dagger}$ is a right inverse of T_{or}. This is not yet a Moore–Penrose inverse, except in the case where $\ker(T') = 0$. Another factorization, this time an outer–inner factorization on T_{or} is needed to produce the Moore-Penrose inverse: $T_{or} = T_o W$ for a right-inner W and a left-outer T_o (the dual case of before). It turns out that T_o is fully outer, $T = VT_oW$ and the Moore–Penrose inverse is $T^{\dagger} = W'T_o^{-1}V'$.

An additional benefit of the inner–outer factorization is the fact that $\ker T'|_{\mathcal{L}_2^{\mathbf{n}}} = \ker V'|_{\mathcal{L}_2^{\mathbf{n}}}$ as well as $\ker T'|_{\mathcal{X}_2^{\mathbf{n}}} = \ker V'|_{\mathcal{X}_2^{\mathbf{n}}}$, an important property for inversion theory. This follows immediately from $T = VT_o$ and $V = TT_o^{\dagger}$, where $T_o^{\dagger}$ is any pseudo-inverse of T_o (the property remains valid even when $T_o^{\dagger}$ is unbounded with dense domain).

Consider now $\mathcal{K} := \ker(T|_{\mathcal{L}_2^{\mathbf{m}}})$. Trivially, $\mathcal{K} \in \ker(T|_{\mathcal{X}_2^{\mathbf{m}}})$, as well as all its anti-causal right shifts: $\mathcal{K}(Z')^k \in \ker(T|_{\mathcal{X}_2^{\mathbf{m}}})$ for any $k \geq 0$. Let $\mathcal{K}_{\text{in}} = \text{span}(\mathcal{K}(Z')^k)_{k=0}^{\infty}$, then also $\mathcal{K}_{\text{in}} \in \ker(T|_{\mathcal{X}_2^{\mathbf{m}}})$. Remarkably, it may happen (and often happens) that $\ker(T|_{\mathcal{X}_2^{\mathbf{m}}}) \neq \mathcal{K}_{\text{in}}$. In that case $\ker(T|_{\mathcal{X}_2^{\mathbf{m}}}) = \mathcal{K}_{\text{in}} \oplus \mathcal{K}_{\text{in}}''$, where $\mathcal{K}_{\text{in}}''$ is a (doubly) right-invariant D-Z-Z' subspace of $\mathcal{X}_2^{\mathbf{m}}$ – i.e., $\mathcal{K}_{\text{in}}''Z \subset \mathcal{K}_{\text{in}}''$ as well as $\mathcal{K}_{\text{in}}''Z' \subset \mathcal{K}_{\text{in}}''$. $\mathcal{K}_{\text{in}}''$ cannot belong to $\mathcal{L}_2$ nor to $\mathcal{U}_2$ except in very trivial contexts. This issue is the topic of the chapter on invertibility, where also an example is given.

Finally, suppose that an isometric realization has been chosen for V and let $\mathbf{O}_T$ and $\mathbf{O}_V$ be the observability operators of respect. T and V, then also $\ker(\mathbf{O}_T'\mathbf{O}_V)' = 0$, because $x\mathbf{O}_V'\mathbf{O}_T = 0 \Longrightarrow x\mathbf{O}_V' \in \ker(T'|_{\mathcal{L}_2^{\mathbf{n}}}) = \ker(V'|_{\mathcal{L}_2^{\mathbf{n}}})$. But V has an isometric realization, and hence $\mathbf{O}_V'\mathbf{O}_V = I$ and $x = 0$. It follows that $M^{<+1>} :=$

$\mathbf{O}'_V\mathbf{O}_T$ (which plays an important role in the next section) is a locally left-invertible diagonal operator.

The Square-Root Algorithm

An outer–inner factorization (respect. inner–outer) is easy to compute in the semi-separable case. The strategy followed here is somewhat heuristic: the approach is to find the solution by induction and then to check it to be correct. It has the advantage to be intuitive and computational, for a more formal approach one should check the literature (see the notes at the end of the chapter). The expression $V'T = T_o$ may be seen as defining a maximal left-inner V whose transpose "pushes" T to upper (anti-causal) without destroying its causality. A first consequence of the relation is that the reachability space of T_o must be contained in the reachability space of T, since the reachability space of T_o is the range of the Hankel operator related to $T_o' = T'V$. Hence one may look for a (potentially non-minimal) realization for T_o that borrows the reachability data $\begin{bmatrix} A & B \end{bmatrix}$ from T. Posing realizations for the unknowns $V := D_V + C_V(I - ZA_V)^{-1}ZB_V$ and $T_o := D_o + C_o(I - ZA)^{-1}ZB$, with $\begin{bmatrix} A_V & B_V \\ C_V & D_V \end{bmatrix}$ isometric, $V'T = T_o$ translates to

$$T_o = D_o + C_o(I - ZA)^{-1}ZB = (D_V' + B_V'Z'(I - A_V'Z')^{-1}C_V') \\ *(D + C(I - ZA)^{-1}ZB. \qquad (36.28)$$

As in the section on external factorizations, the main difficulty with this expression is the occurrence of a "quadratic term" in the product, and as before, one checks that it can be split:

$$Z'(I - A_V'Z')^{-1}C_V'C(I - ZA)^{-1}Z = Z'(I - A_V'Z')^{-1}A_V'M \\ +M + MA(I - ZA)^{-1}Z \qquad (36.29)$$

in which M satisfies a forward Lyapunov–Stein equation

$$M^{<+1>} = A_V'MA + C_V'C \qquad (36.30)$$

the difference with before being that the equation now contains the unknowns A_V and B_V as well as M. Introducing the split, one obtains

$$\begin{array}{l} (D_V'D + B_V'MB) + B_V'Z'(I - A_V'Z')^{-1}(C_V'D + A_V'MB) \\ \quad +(D_V'C + C_V'MA)(I - ZA)^{-1}ZB \\ \quad ? = ?D_o + C_o(I - AZ)^{-1}ZB \end{array} . \qquad (36.31)$$

A first requirement is: keeping $V'T$ lower; hence one must require $C_V'D + A_V'MB = 0$. Next, the expression confirms the contention that $V'T$ shares the reachability data with T. Finally: $D_o = D_V'D + B_V'MB$ and $C_o = D_V'C + B_V'MA$ suffice to satisfy the equation. Summarizing:

$$\begin{bmatrix} A_V' & C_V' \\ B_V' & D_V' \end{bmatrix} * \begin{bmatrix} MA & MB \\ C & D \end{bmatrix} = \begin{bmatrix} M^{<+1>} & 0 \\ C_o & D_o \end{bmatrix}. \tag{36.32}$$

At index k the equation produces the forward recursion

$$\begin{bmatrix} A_{Vk}' & C_{Vk}' \\ B_{Vk}' & D_{Vk}' \end{bmatrix} * \begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix} = \begin{bmatrix} M_{k+1} & 0 \\ C_{ok} & D_{ok} \end{bmatrix}. \tag{36.33}$$

To solve this recursion, knowledge of M_k is assumed, and the computation of the local realizations of V and T_o is attempted, as well as the computation of the next M_{k+1}. The right outerness of T_o requires coker$(D_{ok}) = 0$, or, in words, the rows of D_{ok} have to be linearly independent. Similarly, the fact that $M^{<+1>} = \mathbf{O}_V'\mathbf{O}_T$ forces M_{k+1} to have a right inverse (see the previous subsection). If V has to be as large as possible, then the best one can do is have the rows of the right-hand side span the co-range of $\begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix}$. This observation connects immediately with QL-factorization. Suppose Q and L are such that

$$\begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix} = \begin{bmatrix} Q_{11} & Q_{12} & Q_{13} \\ Q_{21} & Q_{22} & Q_{23} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ L_{21} & 0 \\ L_{31} & L_{32} \end{bmatrix} \tag{36.34}$$

with Q unitary and L right invertible. The columns of $\begin{bmatrix} Q_{13} \\ Q_{23} \end{bmatrix}$ then form an orthonormal basis for the range of $\begin{bmatrix} M_k B_k \\ D_k \end{bmatrix}$, and L_{32} a basis for its co-range. Hence one identifies $L_{32} = D_{ok}$. The next step produces $\begin{bmatrix} Q_{12} \\ Q_{22} \end{bmatrix}$ as a basis for ran $\begin{bmatrix} M_k A_k \\ C_k \end{bmatrix} \ominus$ ran $\begin{bmatrix} M_k B_k \\ D_k \end{bmatrix}$, for whose co-range the rows of L_{21} then provide a basis. Hence $L_{21} = M_{k+1}$. Finally, $\begin{bmatrix} Q_{11} \\ Q_{21} \end{bmatrix}$ will span the co-kernel of the original. (The QL-factorization starts out with reducing the last column to the right-bottom element and then proceeds to the next column to the left.).

One easily identifies the block entries in Q and L with the realizations of V, W, and T_o, here is the final result:

Proposition 2. *The Q-L factorization of* $\begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix}$ *produces realizations for* V, W *and* T_o *as follows:*

$$\begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix} = \begin{bmatrix} B_{Wk} & A_{Vk} & B_{Vk} \\ D_{Wk} & C_{Vk} & D_{Vk} \end{bmatrix} \begin{bmatrix} 0 & 0 \\ M_{k+1} & 0 \\ C_{ok} & D_{ok} \end{bmatrix} \tag{36.35}$$

for which

$$\begin{aligned} V &= D_V + C_V(I - ZA_V)^{-1} Z B_V \\ W &= D_W + C_W(I - ZA_V)^{-1} Z B_V \, . \\ T_o &= D_o + C_o(I - ZA)^{-1} Z B \end{aligned} \tag{36.36}$$

(the check has to be done, but it is straightforward). This is the square-root algorithm, and as before, it is numerically stable, meaning that both an erroneous choice for M_k to start up the recursion and numerical errors incurred during the recursion will die out exponentially fast.

The Moore–Penrose Inverse of a General Semi-separable Operator

If T is lower (causal) semi-separable, then a Moore–Penrose inverse for T is obtained from an inner–outer and an outer–inner factorization, as explained in the previous section. $T = VT_oW$ and hence $T^\dagger = W' T_o^{-1} V'$. In this expression V is semi-separable isometric, W is semi-separable and co-isometric, and T_o is outer. T_o^{-1} is not necessarily bounded, but it has a causal realization and exists on a dense subset of the output space for T_o. The state-space dimensions of all these operators are equal or smaller than the state-space dimension of the original T at each index point k. Typically, one would not execute the product to find a solution to the Moore–Penrose minimization problem, which formulates briefly as: given y find

$$x = \operatorname{argmin}_{u \in \operatorname{argmin}_v (\|Tv - y\|_2)} \|u\|_2, \tag{36.37}$$

whose solution is $x = T^\dagger y$; but one would leave $T^\dagger$ as a product of three operators, two of which are semi-separable (V' and W') and one (T_o) may have an unbounded inverse, which has a more or less decent state-space representation. If T is known to have a bounded inverse, then T_o will of course have a bounded inverse as well, and with some operator theoretic arguments one can show that the realization obtained through inner–outer factorizations is u.e.s.

The next step is how to handle a full semi-separable operator, given by the realization

$$T = C_c(I - ZA_c)^{-1}ZB_c + D + C_a(I - Z'A_a)^{-1}Z'B_a \tag{36.38}$$

(in which A_c and A_a are u.e.s.). Assume the realizations to be minimal (if not: make them minimal!) and put the anti-causal (upper) part in input normal form – i.e. $\begin{bmatrix} A_a & B_a \end{bmatrix}$ is co-isometric. Let then B_W and D_W form a unitary completion:

$$\begin{bmatrix} A'_a & B_W \\ B'_a & D_W \end{bmatrix} \tag{36.39}$$

is unitary. Let $W = D_W + B'_a(I - ZA'_a)^{-1}ZB_W$, then, as in the section on external factorization, TW will be lower. As in that section, a realization for $T_u := TW$ is obtained as $T_u = D_u + C_u(I - ZA_u)^{-1}ZB_u$ with

$$\begin{bmatrix} A_u & B_u \\ C_u & D_u \end{bmatrix} = \left[\begin{array}{cc|c} A_c & B_cB'_a & B_cD_W \\ 0 & A'_a & B_W \\ \hline C_c & C_aA'_a + DB'_a & DD_W + C_aB_W \end{array}\right]. \tag{36.40}$$

This realization may not be minimal (e.g., if $T = W'$ one would have $T_u = I$), but it is reachable, the reachability gramian is simply $\begin{bmatrix} G_c & \\ & I \end{bmatrix}$, in which G_c is the reachability gramian of the lower part. The next step is now to perform inner–outer decompositions on $T_u = V_1T_oV_2$, potentially after a minimalization of T_u (and then, later, of T_o). This then produces $T = V_1T_oV_2W'$ and finally the Moore–Penrose inverse

$$T^\dagger = WV'_2T_o^{-1}V'_1 \tag{36.41}$$

in which all factors have realizations that are smaller than the original, and can hence be called "efficient."

LU and Spectral Factorization

An interesting question with many applications is whether there exists a factorization $T = LU$ with L a lower and lower invertible operator (i.e., L outer) and U an upper and upper invertible operator (i.e., U outer in the Z'-context). This is generally called*spectral factorization*, a key step in solving Fredholm equations. The problem is hard to solve when T itself is not bounded-invertible, so the assumption of such invertibility is commonly made. In the Hardy space context of the unit complex disc it is called *dichotomy*: no "zeros" of the system lie on the unit circle. If the collections of zeros and poles strictly inside the unit disc as well as that strictly outside are finite and the numbers of poles and zeros in the respective domains match (multiplicities counted), then the factorization exists. This would certainly be the case when T is a (strictly) positive rational operator, for in that case

T is bounded-invertible, and the matching condition necessarily holds. In the case of LU-factorization of finite matrices, the factorization does not necessarily exist, even when T has a bounded inverse. In this section, necessary and sufficient conditions for the existence of the LU-factorization will be derived under the condition of (bounded) invertibility of T, and it will be given in terms of characteristic inner factors that generalize the notion of "poles" and "zeros" to the semi-separable case.

The starting point is again a realization for a general semi-separable operator:

$$T = C_c(I - ZA_c)^{-1}ZB_c + D + C_a(I - Z'A_a)^{-1}Z'B_a \tag{36.42}$$

with the additional assumption that the anti-causal part has a uniformly reachable and u.e.s. realization. Hence it can be assumed in input normal form ($A_aA_a' + B_aB_a' = I$), with A_a u.e.s.

Two preliminary remarks are in order: (1) the factorization is not unique, but it is unique up to a right diagonal unitary factor on L and its conjugate as a left factor on U. This allows normalization of one of the factors to have unit main diagonal. Here, the main diagonal of U is taken to be $D_U = I$; (2) an LU-factorization is necessarily minimal, i.e., the minimal state space realization of U will have the same dimension as that of the upper part of T (i.e., $T_a := C_a(I - ZA_a)^{-1}ZB_a$), and likewise with L and the lower part of T. Actually, U may borrow the reachability pair $\begin{bmatrix} A_a & B_a \end{bmatrix}$ of T_a.

The first step is as before: let $W = D_W + C_W(I - ZA_a')^{-1}B_a'$ be an inner operator, obtained after unitary completion of $\begin{bmatrix} A_a & B_a \end{bmatrix}$, and consider now $T_u = TW$ with realization given by Eq. (36.40). Let $T_u = T_oV$ be an outer–inner factorization of T_u, on the basis of the given realization of T_u, which may be non-minimal, but, as shown in the previous section, is uniformly reachable. Under the given hypotheses, T_o is outer, but V may merely be right-inner (i.e., causal and co-isometric). It turns out that the LU-factorization exists if V is (fully) inner with appropriate dimensions. The full result, including formal expressions, is in the following theorem. The resulting algorithm to compute the factorization, with some further motivation, is given thereafter.

Theorem 3. *Let T be a semi-separable operator with bounded inverse and minimal realization given by (36.42), in which $\begin{bmatrix} A_a & B_a \end{bmatrix}$ is co-isometric and A_a is u.e.s. Let W be a minimal inner operator that makes $T_u := TW$ causal, and let $T_u = T_oV$ be an outer–inner factorization of T_u. Let then a unitary realization for W be given by $W = D_W + C_W(I - ZA_a')^{-1}ZB_a'$ and a co-isometric one for $V = D_V + C_V(I - ZA_V)^{-1}ZB_V$, and let R satisfy the Lyapunov–Stein recursion*

$$R^{<-1>} = A_a'RA_V' + B_a'B_V'. \tag{36.43}$$

Then the LU-factorization $T = LU$ exists iff R is bounded invertible (and hence square). In that case V is inner and the (normalized) upper factor U is given by

$$U = I + F(I - Z'A_a)^{-1}ZB_a, \tag{36.44}$$

with

$$F = -(C_W R A_V' + C_W B_V')(R^{<-1>})^{-1}. \tag{36.45}$$

Furthermore, a realization for the (anti-causal) inverse of U is given by

$$U^{-1} = I - FR^{<-1>}(I - Z'A_V')^{-1}Z'R^{-1}B_a \tag{36.46}$$

and for the outer left factor L by

$$L = (D + C_c M_1 R^{-1} C_W') + C_c(I - ZA_c)^{-1}Z(A_c M_1 R^{-1} C_W' + B_c). \tag{36.47}$$

It takes a bit of work to give full proof of the theorem (it is originally in Dewilde [5]), but the algorithm to compute U is straightforward. Since W is already known, one has to compute V as the right inner factor of T_u and solve the forward recursion for R. U is then expressed in these quantities. The outer–inner factorization of T_u follows the schema of the square-root algorithm established in the section "Inner–Outer Factorization," rewritten here in terms of T_u:

$$\left[\begin{array}{c|c} A_c M_1 + B_c B_a' M_2 & B_c D_W \\ A_a' M_2 & B_W \\ \hline C_c M_1 + (C_a A_a' + D B_a') M_2 & D D_W + C_a B_W \end{array}\right] \begin{bmatrix} A_V' & C_V' \\ B_V' & D_V' \end{bmatrix} = \left[\begin{array}{c|c} M_1^{<-1>} & B_{o1} \\ M_2^{<-1>} & B_{o2} \\ \hline 0 & D_o \end{array}\right] \tag{36.48}$$

in which one remarks that M splits into two blocks (because of the dimensions of T_u). The second block-row reduces to the recursion for R: $R := M_2$.

A further observation (this is the crucial element of the proof) concerns UW. This quantity happens to be a so-called maximal phase operator, i.e., a causal invertible operator, whose inverse is anti-causal. A (minimal) realization of $W^{-1}U^{-1}$ is easily determined by direct calculation (using the unknown F) and is $W^{-1}U^{-1} = D_W' + (B_W' - D_W' F)[I - Z'(A_W' - C_W' F)]^{-1} Z' C_W'$. Since $VW^{-1}U^{-1} = T_o^{-1}L^{-1}$ is upper, V has to be a (minimal) external left factor of $W^{-1}U^{-1}$ and V' therefore shares observability data with it (in the Z'-context). Hence there must be a state transformation R such that

$$\begin{bmatrix} R A_V' R^{-<-1>} \\ B_V' R^{-<-1>} \end{bmatrix} = \begin{bmatrix} A_W' - C_W' F \\ B_W' - D_W' F \end{bmatrix} = \begin{bmatrix} A_W' & C_W' \\ B_W' & D_W' \end{bmatrix} \begin{bmatrix} I \\ -F \end{bmatrix}. \tag{36.49}$$

Inverting the unitary realization of W' produces the equations for R and F sought. Once U is computed, the realization for L follows as well by direct calculation on

$L = T_u(W^{-1}U^{-1})$. Much of the proof of the theorem consists in checking all these contentions.

The main diagonal of L has an interesting interpretation as subsequent “pivots” in the LU-factorization (which they would indeed be in the case that T is a matrix with scalar elements). As can be seen from its expression, these pivots are obtained from a ratio of two quantities ($M_1 M_2^{-1} = M_1 R^{-1}$) which are computed from intrinsic orthogonal operations in the square root algorithm. It turns out that these quantities also exist, even when R is not invertible. This leads to potential extensions of the result, beyond the scope of the present chapter.

Example: Block-Tridiagonal System

An interesting new question is what happens when the system is more complex than just semi-separable, for example when it has a block-band structure, in which the blocks themselves are semi-separable. To conclude the section with a nice example, consider a half-infinite case of the form

$$T = \begin{bmatrix} \boxed{D_0} & N_0' & & & \\ N_0 & D_1 & N_1' & & \\ & N_1 & D_2 & N_2' & \\ & & \ddots & \ddots & \ddots \end{bmatrix}, \tag{36.50}$$

in which all D_k and N_k are square and banded matrices themselves (e.g., with three bands) and such that the N_k are non-singular (e.g., this would be the case in a simple 2D finite difference discretization of Poisson's equation). The previous theory clearly applies with the D_k and N_k blocks as the entries of the matrix. (A further issue is whether the sub-band structure of the entries in T can be exploited to achieve a higher order of numerical efficiency, it is discussed at the end of the section.) Using the previous notation and with the N_k's non-singular, the W operator is trivial, it is just $W = Z$ (with Z matching the dimensions of the blocks). A realization for T_u in input normal form is then given by

$$\operatorname{diag}\left(\left[\begin{array}{c|c} & I \\ \hline \cdot & - \end{array}\right], \boxed{\left[\begin{array}{c|c} I & 0 \\ 0 & 0 \\ \hline M_0 & N_0' \end{array}\right]}, \left[\begin{array}{cc|c} 0 & I & 0 \\ 0 & 0 & I \\ \hline N_0 & M_1 & N_1' \end{array}\right], \ldots \right). \tag{36.51}$$

Hence the square-root recursion to be solved (general term) becomes:

$$\left[\begin{array}{c|c} M_{k2} & 0 \\ 0 & I \\ \hline N_{k-1}M_{k1} + D_k M_{k2} & N_k' \end{array}\right] = \left[\begin{array}{c|c} M_{k+1,1} & B_{ok1} \\ M_{k+1,2} & B_{ok2} \\ \hline 0 & D_{ok} \end{array}\right] \begin{bmatrix} A_{Vk} & B_{Vk} \\ C_{Vk} & D_{Vk} \end{bmatrix}. \tag{36.52}$$

After inverting V, it holds that

$$\begin{bmatrix} M_{k+1,1} \\ M_{k+1,2} \end{bmatrix} = \begin{bmatrix} M_{k2}A'_{Vk} \\ B'_{Vk} \end{bmatrix} \tag{36.53}$$

and, in particular, B_V must be invertible for the factorization to exist. Once the recursion is obtained, it also follows from Eq. (36.49) that

$$\begin{bmatrix} R^{<-1>} \\ -FR^{<-1>} \end{bmatrix} = \begin{bmatrix} 0 & I \\ I & 0 \end{bmatrix} \begin{bmatrix} RA'_V \\ B'_V \end{bmatrix} = \begin{bmatrix} B'_V \\ RA'_V \end{bmatrix}. \tag{36.54}$$

This defines all quantities needed, since in particular $R_k = M_{k2}$ and $F = -RA'_V R^{-<-1>}$ in this case. Hence also $F_k = -R_k A'_{Vk} R_{k+1}^{-1} = -M_{k+1,1}M_{k+1,2}^{-1}$, while the pivot is given by $d = D + NM_1M_2^{-1} = M - NF^{<+1>}$. All this reduces to the key equation

$$\begin{bmatrix} N_k M_{k1} + D_k M_{k2} & N_k' \end{bmatrix} \begin{bmatrix} A'_V \\ B'_V \end{bmatrix} = 0 \tag{36.55}$$

with the latter factor isometric and $R_{k+1} = B'_{Vk}$ square non-singular. This in turn requires $N_k M_{k1} + D_k M_{k2}$ non-singular and

$$A'_{Vk} B_{Vk}^{-\prime} = -(N_k M_{k1} + D_k M_{k2})^{-1} N_k' = -R_k^{-1}(D_k - N_k F_{k-1})^{-1} N_k'. \tag{36.56}$$

It follows directly that

$$F_k = -R_k A'_{Vk} B_{Vk}^{-\prime} = (D_k - N_k F_{k-1})^{-1} N_k' \tag{36.57}$$

as could be expected from the classical Schur-complement formula, and one recognizes the pivots $d = D - NF^{<+1>}$ (which in this simple case can easily be computed directly). As only ratios appear in the recursion for F_k, an unnormalized recursion is maybe more comfortable. From the last equations it follows that

$$\begin{bmatrix} M_{k+1,1} \\ M_{k+1,2} \end{bmatrix} = \begin{bmatrix} -M_{k2} \\ N_k^{-\prime}(N_k M_{k1} + D_k M_{k2}) \end{bmatrix} x_k \tag{36.58}$$

for some x_k, hence the following, linear recursion will produce the same ratios:

$$\begin{bmatrix} \hat{M}_{k+1,1} \\ \hat{M}_{k+1,2} \end{bmatrix} = \begin{bmatrix} 0 & -I \\ N_k^{-\prime} N_k & N_k^{-\prime} D_k \end{bmatrix} \begin{bmatrix} \hat{M}_{k1} \\ \hat{M}_{k2} \end{bmatrix} \tag{36.59}$$

(i.e., $F_{k-1} = -M_{k1}M_{k2}^{-1} = -\hat{M}_{k1}\hat{M}_{k2}^{-1}$). In the most simple instance $N_k' = N_k$, in which case the linearized recursion simply becomes

$$\begin{bmatrix} \hat{M}_{k+1,1} \\ \hat{M}_{k+1,2} \end{bmatrix} = \begin{bmatrix} 0 & -I \\ I & N_k^{-\prime} D_k \end{bmatrix} \begin{bmatrix} \hat{M}_{k1} \\ \hat{M}_{k2} \end{bmatrix}. \tag{36.60}$$

When both N_k and M_k are tri-banded, then $N_k^{-\prime} M_k$ will have semi-separable order at most six at each sub-index. At each step in the recursion, the overall semi-separable order increases roughly with six, and after a few steps in the main recursion the low sub-order is destroyed. This phenomenon can easily be analyzed in detail, from which it appears that the recursion does not preserve the sub-band structure. This means that an exact solution of the LU-factorization that preserves the sub-band structure does not exist. Not only the sparsity in the entries is lost, but the entries of the blocks in the factorization do not even remain semi-separable. However, it remains possible to find low degree semi-separable approximate factorizations, which in many applications may suffice, in particular when a pre-conditioner is desired, which would allow to solve the system of equations iteratively and efficiently as well.

Limit Behavior

The term "limit behavior" concerns ranges and kernels of semi-separable operators as it differs from classical finite matrix theory. Let $T = C_c(I - ZA_c)^{-1}ZB_c + D + C_a(I - Z'A_a)^{-1}Z'B_a$ be a (double-sided) semi-separable operator as considered in the previous two sections, in which the realizations are just assumed to be minimal, but not necessarily uniformly reachable and observable (as was assumed in the last section). Without impairing generality, the anti-causal part may be assumed to be in input normal form (if not, it can be put in that form, using a backward recursion). Hence $\begin{bmatrix} A_a & B_a \end{bmatrix}$ is co-isometric (but A_a is not necessarily u.e.s.), and $W = D_W + C_W(I - ZA_a')ZB_a'$ is an isometric realization as well such that $T = T_u W$, with T_u lower, as before. Furthermore, let $T_u = T_o V$ be an outer–inner factorization of T_u, also as before. The kernel of T is described in terms of V and W by the following proposition:

Proposition 3.

$$\ker T = \ker W \oplus W' \ker V. \tag{36.61}$$

Proof. Clearly $\ker W \subset \ker T$. As T_o is left-outer, $\ker T_u = \ker V$. Let $y \in \ker V \cap \operatorname{ran} W$, then there is an input u such that $u = W'y$ and $y = Wu$ because $WW' = I$, and requiring $y \in \ker V'$ puts $u \in W' \ker V$. Orthogonality between $\ker W$ and $W' \ker V$ follows from $\operatorname{ran} W' \perp \ker W$. □

As a right inner factor, V is co-isometric as well. It follows that $\ker V' = 0$. Let $V : \mathcal{X}_2^{\mathbf{m}} \to \mathcal{X}_2^{\mathbf{k}}$, and Let $\mathcal{K}_{\text{in}} = \ker V|_{\mathcal{L}_2^{\mathbf{m}}}$. Then (as before)

$$\ker V = \operatorname{span}_{i=0}^{\infty} (\mathcal{K}_{\text{in}} Z'^i)^- + \mathcal{K}''_{\text{in}}, \tag{36.62}$$

in which $\mathcal{K}_{\text{in}}$ is a right D-Z invariant subspace of $\mathcal{L}_2^{\mathbf{m}}$ (it is the co-kernel of H_V and given by $U\mathcal{L}_2^{\mathbf{k}_1}$ for a co-isometric $U = D_U + C_U(I - ZA_V)^{-1}ZB_V$ with C_U and D_U complementing the co-isometric realization for V and $\mathbf{k}_1 = \mathbf{m} - \mathbf{k}$), and $\mathcal{K}''_{\text{in}}$ is a right D-Z-Z′ invariant subspace.

The kernel of T (and dually of T') can therefore be evaluated completely from the properties of W and V (respect. similar operators related to T'), which in turn follow mainly from the behavior of their transition operators A_V and A_W. In many applications the kernels of type $\mathcal{K}_{\text{in}}$, which are by definition infinite dimensional, are zero and only doubly invariant subspaces remain as kernels, one for T and one for T'. In the case of semi-separable systems, these kernels are finite dimensional and such systems are therefore of "Fredholm" type, with Fredholm index the difference between the two dimensions. Although a full treatment of this case is beyond the scope of this chapter, the dimensionality theorem is stated here and an example related to the introduction of the chapter is given.

Theorem 4. *For any semi-separable, co-isometric, and causal V whose state space dimensions are uniformly bounded, $\mathcal{K}''_{\text{in}}$ has finite dimension.*

Proof. Let $V_1 = \begin{bmatrix} V \\ U \end{bmatrix}$, with U as just define above, and let $\mathcal{H} = \operatorname{ran} H'_{V_1} \in \mathcal{U}_2^{\mathbf{m}}$ – the co-range of the Hankel operator H_{V_1}. Then $\mathcal{H} = \operatorname{ran} H'_V$ as well because the co-range of H_V is determined by the reachability pair $\begin{bmatrix} A_V & B_V \end{bmatrix}$. In addition, $\mathcal{U}_2^{\mathbf{m}} = \mathcal{H} \oplus V_1'\mathcal{U}_2$, by construction of V_1. Let now $u_{\text{in}} \in \mathcal{K}''_{\text{in}}$, and let Π_- be the orthogonal projection of $\mathcal{X}_2^{\mathbf{m}}$ onto $\mathcal{U}_2^{\mathbf{m}}$. Then $u := \Pi_- u_{\text{in}} \in \mathcal{H}$, because $u \perp V_1'\mathcal{U}_2^{\mathbf{m}}$, as can be checked directly (one has $V_1 u_{\text{in}} = 0$ and $u_{\text{in}} - u$ is in $\mathcal{L}_2^{\mathbf{m}} Z$ and hence orthogonal on $V_1'\mathcal{U}_2^{\mathbf{m}}$). Let $\mathcal{H}_k = \operatorname{ran}(H'_{Vk})$ be the range of the kth Hankel operator of V'. $\mathcal{H}_k$ is isomorphic to the minimal state space (by the realization theory). Let, moreover, π_{k-} be the projection of any $\ell_2(-\infty : \infty)$ on $\ell_2(-\infty : k)$. Then $\pi_{k-}\mathcal{K}''_{\text{in}} \in \mathcal{H}_k$, by specialization of the relation $\mathcal{H} = \operatorname{ran} H'_{V_1}$ to the index k, and there is a natural embedding of $\pi_{k-}\mathcal{K}''_{\text{in}}$ in $\pi_{j-}\mathcal{K}''_{\text{in}}$ when $j > k$. As the dimension of $\pi_{k-}\mathcal{K}''_{\text{in}}$ is uniformly bounded by assumption, and $\lim_{k\to\infty} \pi_{k-}(u_{\text{in}}) = u_{\text{in}}$, the dimension of $\mathcal{K}''_{\text{in}}$ cannot be larger than the bound (standard proof by contradiction). □

The construction in the proof of the theorem provides, with some work, for a concrete way to compute $\pi_{k-}\mathcal{K}''_{\text{in}}$, directly from A_V and B_V. Although this goes beyond the present chapter, it concludes with the calculation for one of the examples given in its introduction.

Example

Possibly the simplest (and very instructive) example is given by the half-infinite Toeplitz matrix (36.3). It clearly has a co-kernel (kernel of T') spanned by col$\left[\boxed{1}\ 1/2\ 1/4 \cdots\right]$, hence the matrix is not invertible. What is its Moore–Penrose inverse? The matrix has a left inverse given by (36.4) which is not the Moore–Penrose inverse, as its range is not orthogonal on the co-kernel. The answer is produced by the square-root algorithm for an outer–inner factorization (the left inner factor will be unity because there is a left inverse) – this is the dual of the case treated above and it will involve an "output" Fredholm space $\mathcal{K}_o''$. Before determining it and looking at its properties, we remark that the co-kernel of T in the relevant Hilbert–Schmidt space $\mathcal{X}_2$ is given by

$$\mathcal{K}_o'' = \begin{bmatrix} \cdots & — & — & — & \cdots \\ \cdots & 1 & \boxed{1} & 1 & \cdots \\ \cdots & 1/2 & 1/2 & 1/2 & \cdots \\ \cdots & 1/4 & 1/4 & 1/4 & \cdots \\ \cdots & \vdots & \vdots & \vdots & \cdots \end{bmatrix} \mathcal{D}_2 \tag{36.63}$$

which is obviously a right DZ *and* DZ' invariant subspace (it belongs neither to $\mathcal{U}_2$ nor $\mathcal{L}_2$!). The situation is in sharp contrast with the doubly infinite indexed Toeplitz case from classical LTI or Hardy space theory. With "To$[\cdots]$" a constructor that produces a doubly infinite block Toeplitz matrix out of the series in the argument, To$[\cdots, 0, -2, \boxed{1}, 0, \cdots]$ has a full, bounded, anti-causal inverse, namely To$[\cdots, 0, \boxed{0}, -1/2, -1/4, -1/8, \cdots]$. Doubly invariant subspaces cannot occur in the LTI rational case (e.g., see Helson [11]). This has great consequences for embedding and interpolation theory.

The inner–outer factorization for this example now proceeds as follows. First, as T has a causal and bounded left-inverse, it must have a trivial right inner factor: in $T = T_{or}V$ one may put $V = I$ (V is unique except for a diagonal unitary operator). This is because $\mathcal{L}_2 = \mathcal{L}_2 T^\dagger T \subset \mathcal{L}_2 V \subset \mathcal{L}_2$, hence $\mathcal{L}_2 V = \mathcal{L}_2$ and V must be unitary diagonal. Remains the left inner–outer factorization: $T = UT_o$ (T_o will now be both left and right outer, i.e., it has a causal (approximate) inverse). This factorization follows from a square root backward recursion. A causal realization of T is

$$\text{diag}\left(\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}, \boxed{\begin{bmatrix} | & 1 \\ | & 1 \end{bmatrix}}, \begin{bmatrix} 0 & 1 \\ -2 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ -2 & 1 \end{bmatrix}, \cdots\right), \tag{36.64}$$

where the series continues as a future LTI system with realization $\begin{bmatrix} 0 & 1 \\ -2 & 1 \end{bmatrix}$. The fixed point solution for the inner–outer factorization of the LTI system can be easily computed directly and is simply $\begin{bmatrix} A_U & B_U \\ C_U & D_U \end{bmatrix} = \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix}$ for U and

$\begin{bmatrix} A_o & B_o \\ C_o & D_o \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}$ for T_o, with $M = \sqrt{3}$ (for stable numerical methods to compute the fixed point solution, see e.g., Dewilde and van der Veen [7]). At step 0 we have (now in the variant

$$\begin{bmatrix} M_k A_k & M_k B_k \\ C_k & D_k \end{bmatrix} = \begin{bmatrix} B_{ak} & A_{Uk} & B_{Uk} \\ D_{ak} & C_{Uk} & D_{Uk} \end{bmatrix} \begin{bmatrix} ''0'' & ''0'' \\ M_{k-1} & 0 \\ C_{ok} & D_{ok} \end{bmatrix} \tag{36.65}$$

of the square-root algorithm, where the quoted zeros may disappear and with row compression to the South-East quarter):

$$\begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} | & \sqrt{3} \\ | & 1 \end{bmatrix} = \begin{bmatrix} | & 0 \\ | & 2 \end{bmatrix} \tag{36.66}$$

giving $M_{-1} = |$ and then from index -1 the recursion proceeds to $-\infty$ just matching dimensions: $\begin{bmatrix} | & | \\ \cdot & \cdot \end{bmatrix} = \begin{bmatrix} | & | \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}$. The result is in state space models:

$$\begin{aligned} \operatorname{diag} \begin{bmatrix} A_{Uk} & B_{Uk} \\ C_{Uk} & D_{Uk} \end{bmatrix} &= \operatorname{diag} \left(\begin{bmatrix} | & | \\ \cdot & \cdot \end{bmatrix}, \boxed{\begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix}}, \begin{bmatrix} 1/2 & \sqrt{3}/2 \\ -\sqrt{3}/2 & 1/2 \end{bmatrix}, \ldots \right) \\ \operatorname{diag} \begin{bmatrix} A_{ok} & B_{ok} \\ C_{ok} & D_{ok} \end{bmatrix} &= \operatorname{diag} \left(\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}, \boxed{\begin{bmatrix} | & 1 \\ | & 2 \end{bmatrix}}, \begin{bmatrix} 0 & 1 \\ -1 & 2 \end{bmatrix}, \ldots \right) \end{aligned} \tag{36.67}$$

and as input–output operators:

$$\begin{aligned} U &= \begin{bmatrix} \boxed{1/2} & & & \\ -3/4 & 1/2 & & \\ -3/8 & -3/4 & 1/2 & \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix}. \\ T_o &= \begin{bmatrix} \boxed{2} & & & \\ -1 & 2 & & \\ 0 & -1 & 2 & \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \end{aligned} \tag{36.68}$$

This produces the Moore–Penrose inverse as $T^\dagger = T_o^{-1}U'$:

$$T^\dagger = \frac{1}{4}\begin{bmatrix} 1 & -3/2 & -3/4 & -3/8 & \cdots \\ 1/2 & 1/4 & -15/8 & -15/16 & \cdots \\ 1/4 & 1/8 & 1/16 & -63/32 & \cdots \\ 1/8 & 1/16 & 1/32 & 1/64 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix} \tag{36.69}$$

a form for whose entries one can easily derive a closed form expression. More interesting than such an expression is to see that there is an efficient, be it mixed form state space realization for it, when one wants to compute $T^\dagger y$, one computes the intermediate $v = U'y$ via a simple stable backward recursion, and then the resulting $u = T_o^{-1}v$ via an equally stable forward recursion (conversion to the additive form is straightforward and interesting as well). Realizations for U' and T_o^{-1} are simply given by

$$\begin{aligned} \operatorname{diag}\begin{bmatrix} A_{U'k} & B_{U'k} \\ C_{U'k} & D_{U'k} \end{bmatrix} &= \operatorname{diag}\left(\begin{bmatrix} - & \cdot \\ - & \cdot \end{bmatrix}, \boxed{\begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}}, \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}, \ldots\right) \\ \operatorname{diag}\begin{bmatrix} A_{ok} - B_{ok}D_{ok}^{-1}C_{ok} & B_{ok}D_{ok}^{-1} \\ -D_{ok}^{-1}C_{ok} & D_{ok}^{-1} \end{bmatrix} &= \operatorname{diag}\left(\begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix}, \boxed{\begin{bmatrix} | & 1/2 \\ | & 1/2 \end{bmatrix}}, \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}, \ldots\right) \end{aligned} \tag{36.70}$$

The resulting U is not unitary, but merely isometric. Lacking is a basis for the co-kernel of T, namely the vector $\frac{\sqrt{3}}{2}\operatorname{col}\left\{\boxed{1}\ 1/2\ 1/4 \cdots\right\}$. When this column is added to U, e.g., as a first column, a unitary operator appears, showing the uni-dimensional co-kernel. This T has therefore Fredholm index: dim(kernel) $-$ dim(co-kernel) $=$ -1.

Notes

The idea behind solving discretized Fredholm equations by approximating the kernel with a vector outer product and then using that representation to derive an efficient numerical inverse goes back to, e.g., [10]. It gave rise to the term "semi-separable" operators and matrices, whereby, given the Fredholm kernel $K(t,s)$, the semi-separability refers to different (vector outer product) representations for the upper part of the kernel where $t < s$ and its lower part, where $t > s$ (usually the diagonal term where $t = s$ is just used as it is – $K(t,s)$ is often a matrix). The authors just mentioned realized that such an outer representation would give rise to "efficient" numerical calculations, where the numerical complexity is not any more cubic in the dimensions of the overall matrix (N^3, with N is the number of points in the discretization), but linear in N and at most cubic in the number of terms in the outer representation – a great gain in efficiency when the realizations have small dimensions. The main problem with this approach is its heavy emphasis on Gaussian elimination, which is not always applicable and struggles with numerical instabilities even in cases where it is.

In parallel to this, state space theory developed in the wake of Kalman's seminal papers on estimation and control theory (Kalman [12]), giving rise to full blown "state space models" and input–output operators of the Fredholm type derived from them. The connection between state space realizations, the semi-separable decomposition of a Fredholm operator and the potential for efficient matrix or operator inversion was not fully realized at first. A complete theory of time-varying systems, which parallels most of the results of time-invariant theory came into being (Alpay et al. [1]), a comprehensive treatment can be found in Dewilde and van der Veen [6]. A key element of this theory is the use of "canonical factorizations" with inner factors, both of the co-prime type (here called "external factorizations") and of the inner–outer type.

Numerically, such factorizations consist of sequences of orthogonal transformations, known as a "QR-algorithm" or an SVD. Such operations are intrinsically numerically stable, are applicable even when Gaussian elimination is not possible and can be used to compute generalized inverses as well. They have been exploited by a great number of authors to solve matrix problems and kernel problems both efficiently and accurately.

As in the case of a semi-separable decomposition, a different state space representation (usually called a "realization") would be used for the upper part of the kernel or the matrix and its lower part. It turns out that a semi-separable representation can be considered to be a non-minimal state space representation in which the state transition operator is restricted to being a unit matrix, but the opposite is not true, the state space representation is more general as it allows general state transition operators, so that it is easily possible that a system has a good state-space realization but no (meaningful) outer product representations (that is the case when the kernel has a band structure and the state transition operator is hence nilpotent). To deal with this situation, Gohberg and Eidelman [8] introduced the term "quasi-separable" system to characterize the more general type. However, there is a problem going that path. Most of the theory concerns finite dimensional matrices and efficient inversion methods for them, and a matrix would be called "semi-separable" if there is a low dimensional semi-separable representation for it, and of course similarly "quasi-separable" in the more general case. But any semi-separable matrix in this sense is also trivially quasi-separable, and, under some conditions, a quasi-separable matrix can be converted into a low order semi-separable one. It follows that the terms "semi-separable" or "quasi-separable" do not actually refer to the matrix but only to a representation for it.

In the infinite dimensional case, the situation is even more complex, and it may happen that matrices have a low degree quasi-separable representation but not a semi-separable one, while vice versa there will always be a quasi-separable representation when there is a semi-separable one. To make matters more complex, some authors make the distinction between the two and others use the term semi-separable indiscriminately for the whole class, as it is a logical extension of the original notion, and the original authors were not aware of the existence of the generalization. In the case of the inversion of matrices and Fredholm kernels, one would always go for the general representation, as it provides flexibility and numerical stability. However, there are cases where the existence of a semi-separable representation has an important meaning in itself, namely in the so-called factor analysis, where a minimal number of terms in a semi-separable (outer vector) representation corresponds to a minimal number of sources – namely the matrix representing the correlation of sources in a signal detection environment. As this case is not of interest here, and as the original problems that lead to the semi-separable methodology which is best implemented using state space models (as this chapter tries to exemplify!), the terms semi-separable and quasi-separable are used indiscriminately here.

After the establishment of the basic theory, a wealth of contributions came into being exploring various aspects and extensions of the semi- and quasi-separable approach. The method to find Moore–Penrose inverses using inner–outer decomposition was first presented in van der Veen [14]. The connection with standard QR-factorization for finite matrices is in Chandrasekaran [4]. The case of Gaussian elimination for the quasi-separable case is in Eidelman and Gohberg [9]. Spectral factorization is a method of choice to solve the special case of time-invariant Fredholm equations, often referred to as "Volterra equations," originally attributed to Wiener and Hopf [15]. Gohberg and Ben-Artzi extended this notion to the so-called dichotomy (Ben-Artzi and Gohberg [3]), applicable to the more general quasi-separable type. The method to do Gaussian elimination and spectral factorization using inner–outer decomposition came only pretty late (Dewilde [5]), but was preceded by a direct solution to the spectral factorization problem for the positive definite case (van der Veen [14]).

The great impact of inner–outer factorization deserves special mention. It goes back to Hardy space theory and the theory of invariant subspaces, see in particular (Helson [11]) for an attractive modern treatment of the classical approach. These results were greatly extended by Arveson to the so-called Nest Algebras (Arveson [2]). Semi- or quasi-separable operators form a special case of Nest Algebras. In parallel to these developments, Kailath and Morf discovered a particularly attractive way of dealing with the Kalman filter, called the "square-root algorithm" (Morf and Kailath [13]), which turns out to be an inner–outer decomposition of a special case, a fact that was only realized much later; see in this respect (Dewilde-van der Veen [7]).

References

1. Alpay, D., Dewilde, P., Dym, H.: Lossless inverse scattering and reproducing kernels for upper triangular operators. In: Gohberg, I. (ed.) Extension and Interpolation of Linear Operators and Matrix Functions. Operator Theory, Advances and Applications, vol. 47, pp. 61–135. Birkhäuser Verlag, Basel (1990)

2. Arveson, W.: Interpolation problems in nest algebras. J. Funct. Anal. **20**, 208–233 (1975)
3. Ben-Artzi, A., Gohberg, I.: Inertia theorems for nonstationary discrete systems and dichotomy. Linear Algebra Appl. **120**, 95–138 (1989)
4. Chandrasekaran, S., Dewilde, P., Gu, M., Pals, T., Sun, X., van der Veen, A.J., White, D.: Some fast algorithms for sequentially semi-separable representations. SIAM J. Matrix Anal. Appl. **27**(2), 341–364 (2005)
5. Dewilde, P.: On the lu factorization of infinite systems of semi-separable equations. Indag. Math. **23**, 1028–1052 (2012)
6. Dewilde, P., van der Veen, A.-J.: Time-Varying Systems and Computations. Kluwer, Boston (1998). https://dl.dropboxusercontent.com/u/46018027/Kluwer.pdf
7. Dewilde, P., van der Veen, A.-J.: Inner-outer factorization and the inversion of locally finite systems of equations. Linear Algebra Appl. **313**, 53–100 (2000)
8. Eidelman, Y., Gohberg, I.: On a new class of structured matrices. Integ. Equ. Oper. Theory **34**(3), 293–324 (1999)
9. Eidelman, Y., Gohberg, I.: Fast inversion algorithms for a class of block structured matrices. Contemp. Math. **281**, 17–38 (2001)
10. Gohberg, I., Kailath, T., Koltracht, I.: Linear complexity algorithms for semiseparable matrices. Integral Equ. Oper. Theory **8**, 780–804 (1985)
11. Helson, H.: Lectures on Invariant Subspaces. Academic, New York (1964)
12. Kalman, R.E.: Mathematical description of linear dynamical systems. SIAM J. Control **1**, 152–192 (1963)
13. Morf, M., Kailath, T.: Square-root algorithms for least-squares estimation. IEEE Trans. Automat. Control **20**(4), 487–497 (1975)
14. van der Veen, A.J.: Time-varying system theory and computational modeling: realization, approximation, and factorization. Ph.D. thesis, Delft University of Technology, Delft (1993)
15. Wiener, N., Hopf, E.: Ueber eine klasse singulärer integralgleichungen. Sitzungber. Akad. Wiss. Berlin, pp. 696–706 (1931)