

Orthonormal rational function vectors

Steven Delvaux Marc Van Barel

Report TW 390, May 2004



Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200A – B-3001 Heverlee (Belgium)

Orthonormal rational function vectors

Steven Delvaux *Marc Van Barel*

Report TW 390, May 2004

Department of Computer Science, K.U.Leuven

Abstract

In this paper, we develop a matrix framework to solve the problem of finding orthonormal rational function vectors with prescribed poles $y_{k,l} \in \mathbb{C}$, with respect to a certain discrete inner product that is defined by a set of data points $z_{i,j} \in \mathbb{C}$ and corresponding weight vectors $\vec{w}_{i,j}$. Our algorithm for solving the problem is recursive, and it is of complexity $\mathcal{O}(n^3)$. If all data points are real or lie on the unit circle, then the complexity is reduced by an order of magnitude.

Keywords : rational function vectors, prescribed poles, discrete inner product, (block) inverse eigenvalue problems, (block) lower semiseparable plus diagonal matrices

AMS(MOS) Classification : Primary : 42C05, Secondary : 65F18, 65D15.

Orthonormal rational function vectors*

Steven Delvaux Marc Van Barel

5th May 2004

Abstract

In this paper, we develop a matrix framework to solve the problem of finding orthonormal rational function vectors with prescribed poles $y_{k,l} \in \mathbb{C}$, with respect to a certain discrete inner product that is defined by a set of data points $z_{i,j} \in \mathbb{C}$ and corresponding weight vectors $\bar{w}_{i,j}$. Our algorithm for solving the problem is recursive, and it is of complexity $O(n^3)$. If all data points are real or lie on the unit circle, then the complexity is reduced by an order of magnitude.

Keywords: rational function vectors, prescribed poles, discrete inner product, (block) inverse eigenvalue problems, (block) lower semiseparable plus diagonal matrices

1 Introduction.

Let us consider a vector space \mathcal{F}_n of functions and a discrete inner product $\langle \cdot, \cdot \rangle$. In several applications, it is desired to compute an orthonormal basis for \mathcal{F}_n w.r.t. this discrete inner product. For several choices of $(\mathcal{F}_n, \langle \cdot, \cdot \rangle)$, this reduces to solving an inverse eigenvalue problem, where the resulting matrix contains the parameters for the recurrence relation for the orthonormal basis elements.

Let us give an overview of the different choices for $(\mathcal{F}_n, \langle \cdot, \cdot \rangle)$ that were investigated in the literature. Let \mathcal{F}_n be the vector space of polynomials of degree $\leq n$, with inner product defined as $\langle \phi, \psi \rangle = \sum_{i=0}^n |w_i|^2 \overline{\phi(z_i)} \psi(z_i)$ for any $\phi, \psi \in \mathcal{F}_n$ and $w_i, z_i \in \mathbb{C}$, $i = 0, \dots, n$. The parameters for the recurrence relation for the orthonormal polynomials are the elements of the upper Hessenberg matrix H that solves the inverse eigenvalue problem

$$Q^H D_z Q = H,$$

*The research was partially supported by the Research Council K.U.Leuven, project OT/00/16 (SLAP: Structured Linear Algebra Package), by the Fund for Scientific Research–Flanders (Belgium), projects G.0078.01 (SMA: Structured Matrices and their Applications), G.0176.02 (ANCILA: Asymptotic analysis of the Convergence behavior of Iterative methods in numerical Linear Algebra), G.0184.02 (CORFU: Constructive study of Orthogonal Functions) and G.0455.0 (RHPH: Riemann-Hilbert problems, random matrices and Padé-Hermite approximation), and by the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister’s Office for Science, Technology and Culture, project IUAP V-22 (Dynamical Systems and Control: Computation, Identification and Modelling). The scientific responsibility rests with the authors.

where D_z is the diagonal matrix containing the data points z_i , together with a condition to determine the first column of Q based on the weights w_i . For an overview on inverse eigenvalue problems, we refer the interested reader to [3]. When all data points z_i are real, the Hessenberg matrix reduces to a tridiagonal one [9]. When all data points z_i are on the unit circle, the Hessenberg matrix H is itself unitary. This allows to represent the matrix using $O(n)$ parameters, the so called Schur parameters [1].

Instead of polynomials, one can consider polynomial vectors having m components with a corresponding discrete inner product [14, 6, 17]. In this case the solution matrix H of the inverse eigenvalue problem is a generalized Hessenberg matrix having m nonzero subdiagonals. Similarly as in the scalar case, when all z_i are real, the generalized Hessenberg becomes a banded matrix [2, 15], and when all z_i are on the unit circle, H can be parametrized using block Schur parameters [16].

So far, we have only considered polynomial functions. When \mathcal{F}_n is the vector space of proper rational functions with prescribed poles y_k , $k = 1, \dots, n$, the inverse eigenvalue problem becomes

$$Q^H D_z Q = S + D_y, \quad (1)$$

where D_y is the diagonal matrix based on the poles y_k (with an arbitrary value for y_0), and where S has to be lower semiseparable, i.e. all submatrices that can be taken out of the lowertriangular part of S have rank at most 1. Also here, when all z_i are real, S becomes a symmetric semiseparable matrix and when all z_i lie on the unit circle, S has to be of lower as well as upper semiseparable form [18].

The techniques described above can be used in several applications in which polynomial or rational functions play an important role: linear system theory, control theory, system identification [4, 11], data fitting [7], (trigonometric) polynomial least squares approximation [12, 13], and so on. For a comprehensive overview of orthogonal rational functions, the interested reader can consult [5].

In this paper, we give a generalization of [18] to the vector case, i.e. to the computation of an orthonormal basis of rational function vectors. However, we follow a different path, in the sense that most of our results were not mentioned as such in [18], and we explicitly state which algebraic conditions are needed and which are not. The basis is again the inverse eigenvalue problem (1), but this time with $S \in \mathcal{S}$, the class of ‘block lower semiseparable’ matrices, together with an initial condition to determine the first block column of the unitary matrix Q . Emphasis is laid on the partition $\mathcal{S} = \mathcal{S}_{uv} \cup (\mathcal{S} \setminus \mathcal{S}_{uv})$, where $\mathcal{S}_{uv} \subset \mathcal{S}$ is a topologically dense subclass, consisting of the ‘ uv -representable’ block lower semiseparable matrices. It turns out that, depending on whether the matrix S for solving (1) belongs to \mathcal{S}_{uv} or its complement $\mathcal{S} \setminus \mathcal{S}_{uv}$, we can efficiently obtain the required orthonormal basis, or we can prove that no such orthonormal basis exists, respectively. Moreover, we do not assume S to be invertible, as was done in [18]. In the same article a recursive algorithm was developed to solve the inverse eigenvalue problem (1). We generalize this algorithm to the vector case and derive it in an alternative way.

The paper is organized as follows. In Section 2, the main concepts are defined, such as the vector space \mathcal{R}_v of rational functions with prescribed poles, the discrete inner product $\langle \cdot, \cdot \rangle$ and the problem formulation. Section 3 deals with a linear algebra treatment of the solvability conditions of the inverse eigenvalue problem (1), together

with a certain initial condition. Section 4 explains how from the solution of the inverse eigenvalue problem, the orthonormal basis of rational function vectors can be obtained, in the uv -representable case, or an element of norm zero (implying that no such basis exists) in the other case. Section 5 gives an efficient, recursive solution for the inverse eigenvalue problem. Section 6 handles some special configurations of the data points $z_{i,j}$ that allow a speed-up of the algorithm, for example when all the $z_{i,j}$ lie on the real line or on the unit circle.

2 Orthonormal rational function vectors.

In this section we will formulate the main concepts of the paper. Our main purpose is to find orthonormal rational function vectors with respect to a certain discrete inner product. The problem is one which can be easily solved, at least in principle, by the Gram-Schmidt orthogonalization procedure, but we will devise a better method by reducing the orthogonalization problem into a matrix context, for which we will then describe an efficient recursive algorithm in Section 5.

We start with some definitions.

Definition 1 (*Vector space.*) Let $y_{k,l} \in \mathbb{C}$, $k = 0, \dots, n$, $l = 1, 2$ be a given set of numbers satisfying $y_{k,l} \neq y_{\tilde{k},l}$ for all k, \tilde{k}, l . We define the complex vector space

$$\mathcal{R}_u = \text{range}_{\mathbb{C}} \begin{bmatrix} 1 & 0 & \frac{1}{z-y_{1,1}} & 0 & \dots & \frac{1}{z-y_{n,1}} & 0 \\ 0 & 1 & 0 & \frac{1}{z-y_{1,2}} & 0 & \dots & \frac{1}{z-y_{n,2}} \end{bmatrix} \subset \mathbb{C}(z)^2. \quad (2)$$

The $y_{k,l}$ are called the poles of \mathcal{R}_u , except for the $y_{0,l}$ which we call pseudo-poles.

The values of the pseudo-poles $y_{0,l}$ may be chosen arbitrarily since they are not relevant for the definition of \mathcal{R}_u . They are introduced for consistency in later theorems.

It follows from (2) that the elements of \mathcal{R}_u are rational function vectors $\vec{\phi}(z) \in \mathbb{C}(z)^2$ with components of the form $\frac{\sum_{k=0}^n c_{k,l} z^k}{\prod_{k=1}^n (z-y_{k,l})}$, $l = 1, 2$, and conversely every such vector belongs to \mathcal{R}_u by using partial fraction decomposition.

We can already mention that for most theorems in this paper to be valid, we will need the extra condition $y_{k,1} = y_{k,2}$ for all k .

Remark 2 (*Block size.*) The definition in (2) was formulated for vectors in \mathbb{C}^2 , but can be easily adapted for the case of vectors in \mathbb{C}^m , $m \geq 1$. This can be realized by allowing the index l to run between 1 and m . The number m is called the vector size or block size of the problem. During this section, we will use $m = 2$ for notational reasons.

Note that for $m = 1$, we will get the ‘scalar’ definitions of [18], and in this case the indices l (and further on, j) are not relevant.

Definition 3 (*(Pseudo-)inner product.*) Let $z_{i,j} \in \mathbb{C}$, $i = 0, \dots, n$, $j = 1, 2$ be a set of numbers satisfying $y_{k,l} \neq z_{i,j}$ for all i, j, k, l , and for each of them let $\vec{w}_{i,j} \in \mathbb{C}^2$ be a

corresponding weight vector, not equal to the zero vector. Then for $\vec{\phi}(z), \vec{\psi}(z) \in \mathcal{R}_n$ we define the discrete inner product

$$\langle \vec{\phi}(z), \vec{\psi}(z) \rangle = \sum_{i=0}^n \sum_{j=1}^2 \overline{\vec{\phi}(z_{i,j})}^T \vec{w}_{i,j} \vec{w}_{i,j}^T \vec{\psi}(z_{i,j}), \quad (3)$$

where the bar denotes complex conjugation. The $z_{i,j}$ are called the data points.

Note that the two factors $\overline{\vec{\phi}(z_{i,j})}^T \vec{w}_{i,j}$ and $\vec{w}_{i,j}^T \vec{\psi}(z_{i,j})$ in this definition are ordinary complex numbers.

As a special case, we can define the *norm* of $\vec{\phi}(z)$ by $\|\vec{\phi}(z)\| = \sqrt{\langle \vec{\phi}(z), \vec{\phi}(z) \rangle}$. Loosely speaking, this norm will be a kind of ‘weighted evaluation’ of $\vec{\phi}(z)$ in the data points $z_{i,j}$, and it is equal to zero if and only if in each data point $z_{i,j}$, the tangential interpolation condition $\vec{w}_{i,j}^T \vec{\phi}(z_{i,j}) = 0$ is satisfied.

One may ask whether (3) defines a *true* inner product on \mathcal{R}_n , i.e. whether $\|\vec{\psi}(z)\| \neq 0$ for all nonzero $\vec{\psi}(z) \in \mathcal{R}_n$. For the moment we will not worry about this question, but necessary and sufficient conditions will be given later in Theorems 18, 19.

Problem 1 *The problem we will investigate is to search an orthonormal basis $\{\vec{\phi}_{k,l}(z)\}$, $k = 0, \dots, n$, $l = 1, 2$, with respect to the inner product (3), such that with the notation $\phi_k(z) = \begin{bmatrix} \vec{\phi}_{k,1}(z) & \vec{\phi}_{k,2}(z) \end{bmatrix}$ we have that*

$$\text{range}_{\mathbb{C}} \begin{bmatrix} \phi_0(z) & \dots & \phi_n(z) \end{bmatrix} = \mathcal{R}_k \quad (4)$$

for every $k = 0, \dots, n$.

This means that we are going to search $2n + 2$ orthonormal vectors $\vec{\phi}_{k,l}(z) \in \mathbb{C}(z)^2$, $k = 0, \dots, n$, $l = 1, 2$, but for the degree structure implied in (4) we only imply $n + 1$ conditions which we formulate in terms of 2 by 2 *blocks* of basis vectors. Stated in a different way, this means that we build in one degree of freedom for each block $\phi_k(z) \in \mathbb{C}(z)^{2 \times 2}$.

Of course we could easily solve this orthogonalization problem by applying the Gram-Schmidt orthogonalization procedure. This would yield a solution which does not even need this extra degree of freedom for each block ϕ_k . We will try however to find another, better suited way for solving the problem.

For this it will be useful to consider an auxiliary matrix \tilde{Q} .

Definition 4 *Given a set of orthonormal rational function vectors $\vec{\phi}_{k,l}$ solving Problem (1). Then we define the matrix $\tilde{Q} \in \mathbb{C}^{(2n+2) \times (2n+2)}$ with $(i, j; k, l)$ th element*

$$\tilde{Q}_{i,j;k,l} = \vec{w}_{i,j}^T \vec{\phi}_{k,l}(z_{i,j}), \quad (5)$$

$i = 0, \dots, n$, $k = 0, \dots, n$, $j = 1, 2$, $l = 1, 2$. Note that we prefer to work with double indices to denote row and column numbers; a more precise meaning of this is that \tilde{Q} is a block matrix, such that for every i, k the block element $\tilde{Q}_{i,k}$ is of size 2 by 2, with individual elements parametrized by j, l .

Thus by definition, each individual column of \tilde{Q} corresponds to a basis vector $\vec{\phi}_{k,l}$ while each individual row corresponds to an evaluation point $z_{i,j}$ where this vector has to be evaluated, with a weighted evaluation corresponding to $\vec{w}_{i,j}$.

The reason for introducing \tilde{Q} is the following: the $\vec{\phi}_{k,l}(z)$ form an orthonormal basis for \mathcal{R}_x if and only if their corresponding matrix \tilde{Q} is a unitary matrix. Indeed, this is a direct consequence of our definition of inner product!

Note however that such a unitary matrix \tilde{Q} is not unique, since also the orthonormal rational function vectors $\vec{\phi}_{k,l}$ defined by (4) are not unique. But we will show later that \tilde{Q} is ‘essentially unique’ anyway.

3 An associated inverse eigenvalue problem.

To find a unitary matrix \tilde{Q} as in the previous section, we could try to deduce from the degree structure (4) a recursion for the columns of \tilde{Q} and then deduce a matrix equation from this. This would lead to a kind of *inverse eigenvalue problem*. In this section we will start by investigating the algebraic solvability properties of the inverse eigenvalue problem.

We introduce some terminology. We say of a matrix $A \in \mathbb{C}^{n \times n}$ to be *lower semiseparable* if each matrix that can be taken out of the part of A that is lying on and beneath the diagonal of A , is of rank at most 1. We will denote with \mathcal{S} the class of lower semiseparable matrices; here the size n of the matrices is assumed to be fixed.

An important subclass of \mathcal{S} is the class \mathcal{S}_{uv} which consists of those matrices $A \in \mathbb{C}^{n \times n}$ which can be written as $A = \text{Rk } 1 + R_{\text{strict}}$, where $\text{Rk } 1$ is a rank at most 1 matrix and R_{strict} a strictly uppertriangular matrix. The name \mathcal{S}_{uv} indicates that for such matrices, the lowertriangular part can be represented as an outer product uv^T with $u, v \in \mathbb{C}^{n \times 1}$. For this, we could call them *uv-representable*. In general, for $n \geq 2$ the inclusion $\mathcal{S}_{uv} \subset \mathcal{S}$ is strict since for example diagonal matrices belong to \mathcal{S} , but not each diagonal matrix belongs to \mathcal{S}_{uv} . In fact, \mathcal{S} can be characterized as the topological closure of \mathcal{S}_{uv} , with respect to some arbitrary matrix norm: see [19].

We have that $\mathcal{S} = \mathcal{H}^{-1}$, i.e. the invertible elements of the class \mathcal{S} are precisely the inverses of the invertible elements of the class \mathcal{H} of Hessenberg matrices; see for example [19]. However, we will not need this property in the paper.

The class \mathcal{S} can be characterized by its QR and RQ decompositions. For example, we have that $A \in \mathcal{S}$ if and only if there exists an RQ decomposition $A = RG$ for which the G factor consists of a product of $n - 1$ (embedded) Givens transformations

$$G = G^{(n-1,n)} \dots G^{(2,3)} G^{(1,2)}, \quad (6)$$

with each $G^{(k,k+1)}$ acting on columns $k, k + 1$. If A is singular, it may happen that some of the Givens transformations $G^{(k,k+1)}$ are not essentially unique (i.e., that some of the $G^{(k,k+1)}$ can be chosen to be an arbitrary unitary 2 by 2 matrix; see for example when A is the zero matrix), but in this case we agree to choose $G^{(k,k+1)} = I_2$.

Suppose $A \in \mathcal{S}$ and $A = RG$ with G as in (6), satisfying the convention about I_2 that we just mentioned. Then the following are equivalent:

- (i) $\exists k : G^{(k,k+1)}$ is diagonal.

(ii) A is block uppertriangular, i.e. $A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ 0 & A_{2,2} \end{bmatrix}$ with $A_{1,1}$ square and of size $k \times k$.

(iii) Either $A \in \mathcal{S} \setminus \mathcal{S}_{uv}$, or $A \in \mathcal{S}_{uv}$, but with last row or first column entirely zero.

For a proof, we refer to Lemma 7 where we will prove an even more general result about block matrices. Note that the equivalence (i) \Leftrightarrow (ii) is basically trivial, while (ii) \Leftrightarrow (iii) is merely a characterization of how elements of $\mathcal{S} \setminus \mathcal{S}_{uv}$ look like, with diagonal (or uppertriangular) matrices as an extreme case.

In the sequel, we will actually work with block versions of the classes \mathcal{S} , \mathcal{S}_{uv} . Let us define these more precisely:

Definition 5 (The (block) classes \mathcal{S} , \mathcal{S}_{uv} .) Let $A \in \mathbb{C}^{mn \times mn}$ be a matrix which is considered to be a block matrix of size $n \times n$, with block elements of fixed block size m . Then we call A block lower semiseparable if every matrix, consisting of blocks of A that can be chosen on and beneath the block diagonal of A , has block rank at most 1, i.e. scalar rank at most m . We denote this class with \mathcal{S} , where we assume the matrix dimension n and block size m to be known. Similarly, we denote with $\mathcal{S}_{uv} \subset \mathcal{S}$ the subclass of block matrices $A \in \mathbb{C}^{mn \times mn}$ which can be written as $A = Rk \ 1 + R_{\text{strict}}$, where $Rk \ 1$ is of block rank at most 1 and R_{strict} is strictly block uppertriangular.

The previously mentioned properties will essentially remain valid, replacing the scalar notions by the appropriate block notions.

To avoid confusion, let us be somewhat more detailed about the block QR and block RQ decompositions.

Definition 6 We say $A = QR$ to be a block QR decomposition of the block matrix A if Q is unitary (in the usual sense) and R is block uppertriangular. The decomposition is said to be essentially unique if Q is unique up to the right multiplication with a unitary, block diagonal matrix. Similar definitions are made for a block RQ decomposition $A = RQ$. The block size m is assumed to be known.

As in the scalar case, if A is nonsingular then the Q and R factors of $A = QR$ are essentially unique.

The class \mathcal{S} can be characterized by its block QR or RQ decomposition. For example, we have that $A \in \mathcal{S}$ if and only if there exists a block RQ decomposition $A = RG$ where

$$G = G^{(n-1,n)} \dots G^{(2,3)} G^{(1,2)} \quad (7)$$

is a product of (embedded) block Givens transformations with each $G^{(k,k+1)}$ acting on block columns $k, k+1$. Here with a ‘block Givens transformation’, we mean just a $2m$ times $2m$ unitary matrix (or the embedding of such a matrix; this will be clear from the context). If $A \in \mathcal{S}$ is singular, it may happen that some of the $G^{(k,k+1)}$ are not essentially unique (i.e. not unique up to multiplication on the right with a $2m$ times $2m$ unitary, block diagonal matrix; think for example of A being the zero matrix). But then we agree to choose the $(2, 1)$ block element $(G^{(k,k+1)})^H(2, 1)$ to be singular.

Let us explain this last convention about the $(2, 1)$ block element in somewhat more detail. For this, we will use the relation $AG^H = R$, and for simplicity of notation we restrict ourselves to the action of $(G^{(1,2)})^H$, thus $k = 1$. By abuse of notation, let us use the same notation $(G^{(1,2)})^H \in \mathbb{C}^{2m \times 2m}$ for the non-embedded form of this Givens transformation. Let $\tilde{A} \in \mathbb{C}^{(mn-m) \times 2m}$ be the matrix formed by the first 2 block columns of A , with first block row skipped. Then $H := G^{(1,2)}$ must satisfy $\tilde{A}H^H = \begin{bmatrix} 0 & X \end{bmatrix}$, where the blocks are of appropriate dimension. Partitioning $H^H = \begin{bmatrix} H_1^H & H_2^H \end{bmatrix}$, it follows that

$$\tilde{A}H_1^H = 0.$$

This means that $H := G^{(1,2)}$ is completely determined by the condition that the columns of H_1^H form an orthonormal basis for the right null space \mathcal{N} of \tilde{A} . (The columns of H_2^H can then be obtained by completing these to an orthonormal basis of the entire space \mathbb{C}^{2m} .) From the fact that $A \in \mathcal{S}$ is block lower semiseparable, it follows that $\dim \mathcal{N} \geq m$, proving that indeed m vectors in \mathcal{N} can be chosen to form the columns of H_1^H (existence), a fact we already knew of course. Then there are two possibilities. The first case is when $\dim \mathcal{N} = m$, and then it is easy to see that the matrix H^H will be essentially unique (uniqueness), i.e. up to the right multiplication with a $2m \times 2m$ unitary, block diagonal matrix. Indeed: this follows since the m columns of H_1^H form an orthonormal basis for \mathcal{N} , which must then be unique up to a unitary transformation, and since the m columns of H_2^H form an orthonormal basis for \mathcal{N}^\perp , which will also be unique up to such a unitary transformation. The second case is when $\dim \mathcal{N} > m$, and then there is no such uniqueness property. But it is easy to see that in this case, we can always choose the first scalar column of H_1^H to be of the form $\begin{bmatrix} \bar{x}^T & \bar{0}_m^T \end{bmatrix}^T$, with $\bar{0}_m$ being a zero vector of length m . This proves our claim that in the latter case, the block element $H^H(2, 1)$ can be chosen to be singular.

For further reference (Theorem 18), we prove

Lemma 7 *Suppose $A \in \mathcal{S}$ and $A = RG$ with G as in (7), satisfying the convention about the $(2, 1)$ block element that we just described. By abuse of notation, we will use the same notation $G^{(k,k+1)} \in \mathbb{C}^{2m \times 2m}$ to denote the non-embedded form of the block Givens transformations. Then the following are equivalent:*

- (i) $\exists k: G^{(k,k+1)}$ has a singular $(1,2)$ block element.
- (ii) $\exists k: G^{(k,k+1)}$ has a singular $(2,1)$ block element.
- (iii) A is 'almost' block uppertriangular, in the sense that $A = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$ where $A_{1,1}$ is square and of block size $k \times k$, and where $A_{2,1}$ has scalar rank $< m$.
- (iv) Either $A \in \mathcal{S} \setminus \mathcal{S}_{uv}$, or $A \in \mathcal{S}_{uv}$, but with last block row or first block column having scalar rank $< m$.

Remark. Again the equivalence between (i) and (iii) (or alternatively, (ii) and (iii)) will be rather straightforward, while the equivalence between (iii), (iv) is merely a characterization of how the elements of $\mathcal{S} \setminus \mathcal{S}_{uv}$ look like, with block diagonal (or block

uppertriangular) matrices as an extreme case.

PROOF. We first prove (i) \Leftrightarrow (ii), and our proof will in fact be valid for *any* unitary matrix $H \in \mathbb{C}^{2m \times 2m}$. First we prove the implication (i) \Rightarrow (ii). Let $H := G^{(k,k+1)}$ and define a block partitioning

$$H = \begin{bmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{bmatrix},$$

$H_{i,k} \in \mathbb{C}^{m \times m}$. Since H is unitary, we have $H_{1,1}^H H_{1,2} + H_{2,1}^H H_{2,2} = 0$. Supposing $H_{1,2}$ to be singular, then by multiplying on the right with a nonzero vector \vec{c} in its right null space, we obtain $H_{2,1}^H H_{2,2} \vec{c} = 0$. Then since $H_{2,2} \vec{c} \neq \vec{0}$ (because this would imply $\begin{bmatrix} H_{1,2}^T & H_{2,2}^T \end{bmatrix}^T \vec{c} = \vec{0}$, implying Q to be singular), we conclude that $H_{2,1}$ is singular.

The implication (ii) \Rightarrow (i) follows in exactly the same way, or by applying the result of (i) \Rightarrow (ii) on the unitary matrix $H^H := (G^{(k,k+1)})^H$.

Next we prove the implication (ii) \Rightarrow (iii). Suppose that $G^{(k,k+1)}$ has a singular (2,1) block element. For every $1 \leq i \leq n$, denote

$$A^{(i)} = R G^{(n-1,n)} \dots G^{(i,i+1)}.$$

Choosing $i = k + 1$, we obtain $A^{(k+1)}$ and define a partitioning

$$A^{(k+1)} = \begin{bmatrix} A_{(1,1)}^{(k+1)} & A_{(1,2)}^{(k+1)} \\ 0 & A_{(2,2)}^{(k+1)} \end{bmatrix} \quad (8)$$

with $A_{(1,1)}^{(k+1)}$ block uppertriangular of block size $k \times k$. Then choosing $i = k$, we obtain $A^{(k)}$ and write

$$A^{(k)} = A^{(k+1)} G^{(k,k+1)},$$

which we partition in exactly the same way, and with the same sizes of subblocks as in (8). Due to the singular (2,1) block element of $G^{(k,k+1)}$, it follows that the (2,1) block element $A_{(2,1)}^{(k)}$ (which can now be nonzero in its last block column), has rank $< m$, and then obviously the same must hold for the final matrix $A = A^{(1)} = A^{(k)} G^{(k-1,k)} \dots G^{(1,2)}$.

To prove the implication (iii) \Rightarrow (i), we use a similar reasoning but this time using the equation $AG^H = R$ instead of $A = RG$. Note that the result holds also in the singular case since we agreed to choose the (2,1) block element of $(G^{(k,k+1)})^H$, or equivalently the (1,2) block element of $G^{(k,k+1)}$ singular when it would not be essentially unique.

Next we prove (iii) \Rightarrow (iv). Suppose that A has a block of rank $< m$ below its diagonal. Assume then that A is uv -representable, thus $A = UV^T + R_{\text{strict}}$, with $U, V \in \mathbb{C}^{m \times m}$ and R_{strict} strictly block uppertriangular, and suppose by contradiction that the last block row and first block column of A have maximal rank, i.e. rank m . It follows that the block elements U_n, V_1^T must both be nonsingular. But then the bottom left block element $U_n V_1^T$ should be nonsingular too, contradicting the fact that A has a block of rank $< m$ below its diagonal.

Finally we prove the negation \neg (iii) \Rightarrow \neg (iv). Thus suppose that A does *not* have a block of scalar rank $< m$ below its diagonal. We will first show that the block element $A(n, 1)$ is nonsingular. Suppose by contradiction that $1 \leq k < n$ is the largest integer for

which the block vector $A(k : n, 1)$ has its maximal rank, i.e. scalar rank m . The fact that $A \in \mathcal{S}$ is block lower semiseparable implies the scalar rank of $A(k : n, 1 : k)$ to be $\leq m$. But since by the choice of k , $A(k : n, 1)$ is already of rank m , it follows that each column of $A(k : n, 1 : k)$ can be written as a linear combination of the columns of $A(k : n, 1)$. Then by the maximality of k , it follows that $A(k + 1 : n, 1)$ and hence $A(k + 1 : n, 1 : k)$ must be of rank $< m$. Thus A has a block of rank $< m$ below its diagonal, yielding a contradiction.

Thus we proved that $k = n$, i.e. the block element $A(n, 1)$ is nonsingular. Then we can consider UV^T with

$$U = A(:, 1), \quad V^T = A(n, 1)^{-1}A(n, :).$$

Now we claim that $A = UV^T + R_{\text{strict}}$ for a certain strictly block uppertriangular R_{strict} . Note that by construction, UV^T equals A in the first block column and the last block row. Then let $A(i, j; k, l)$ be a subdiagonal element of A , not lying in the first block column or the last block row, and write $A(i, j; k, l) = U_{i,j}V_{k,l}^T + c$ for some $c \in \mathbb{C}$. The fact that $A \in \mathcal{S}$ implies the $m + 1$ by $m + 1$ determinant

$$\begin{vmatrix} A(i, j; 1) & A(i, j; k, l) \\ A(n; 1) & A(n; k, l) \end{vmatrix}$$

to be zero. Since $A(n; 1)$ is of maximal rank m , only one value for $c \in \mathbb{C}$ will satisfy this condition, and this must then necessarily be the value $c = 0$. Thus we proved that $A = UV^T + R_{\text{strict}}$ and hence A is uv -representable. \square

We return to the data points $z_{i,j}$ and poles $y_{k,l}$ of the previous section.

Definition 8 (D_z, D_y .) We define D_z, D_y to be the diagonal matrices which are built from data points and poles, respectively. Thus D_z is a block diagonal matrix with i th block diagonal element equal to $\text{diag}(z_{i,j})_{j=1}^m$, $i = 0, \dots, n$ and D_y is block diagonal with k th block diagonal element equal to $y_k I_m$, $k = 0, \dots, n$. Note that we imposed here an additional assumption, saying that $y_{k,1} = \dots = y_{k,m} =: y_k$ for all k .

This assumption that $y_{k,1} = \dots = y_{k,m} =: y_k$ will be needed in all the following theorems. We will use this assumption throughout the rest of this paper.

It is well known that for every matrix A , the equation $Q^H A Q = H$ (a so called *inverse eigenvalue problem*) always has a solution (Q, H) with Q unitary, with fixed first (block) column, and $H \in \mathcal{H}$, the class of (block) Hessenberg matrices. Since block Hessenberg matrices are precisely the inverses of block lower semiseparable matrices, loosely speaking the classes \mathcal{H} and \mathcal{S} will have the same dimension, i.e. the same number of parameters, and hence we can expect something similar to be true for the class \mathcal{S} . We will prove this now.

Theorem 9 *The inverse eigenvalue problem*

$$Q^H D_z Q = D_y + S \tag{9}$$

always has a solution (Q, S) where Q is unitary, with first block column fixed to be $q_1 \in \mathbb{C}^{m \times m}$ (the individual columns of q_1 are supposed to be orthonormal to each other), and $S \in \mathcal{S}$.

Remark. Following Definition 8, it would be more correct to write the initial condition as $q_0 \in \mathbb{C}^{m(n+1) \times m}$ instead of $q_1 \in \mathbb{C}^{mn \times m}$. But to avoid unnecessary notational complexity, we prefer for the rest of this section to redefine $n := n + 1$, and to label the block columns of D_z, D_y with usual indices $1, \dots, n$.

PROOF. Let us put $A^{(0)} := D_z$. We will then construct a sequence of matrices $A^{(1)}, \dots, A^{(n)}$ where each $A^{(k)}$ satisfies the block lower semiseparable matrix plus diagonal structure in its first k block columns (see Definition 5; with fixed diagonal matrix D_y), and $A^{(k)} = (Q^{(k)})^H D_z Q^{(k)}$ for a unitary matrix $Q^{(k)}$ which has the desired first block column q_1 .

First we construct $A^{(1)} = (Q^{(1)})^H D_z Q^{(1)}$. Since there is no structure yet to be satisfied, the only requirement for this matrix is that the first block column of $Q^{(1)}$ has the desired value q_1 , and this can be easily achieved.

Suppose now that we found $A^{(k-1)} = (Q^{(k-1)})^H D_z Q^{(k-1)}$, $2 \leq k \leq n - 2$. By assumption, $A^{(k-1)}$ satisfies the block lower semiseparable plus diagonal structure up to and including its $(k - 1)$ th block column. Defining a partitioning

$$A^{(k-1)} = \begin{bmatrix} A_{(1,1)}^{(k-1)} & A_{(1,2)}^{(k-1)} \\ A_{(2,1)}^{(k-1)} & A_{(2,2)}^{(k-1)} \end{bmatrix},$$

with $A_{(1,1)}^{(k-1)}$ being square of block size $k - 1$ by $k - 1$, the structure implies that $A_{(2,1)}^{(k-1)}$ has block rank at most 1. This means that $A_{(2,1)}^{(k-1)}$ has *scalar* rank at most m ; let $r \leq m$ be the *exact* scalar rank. Now we choose a set of r independent scalar row indices i_1, \dots, i_r for $A_{(2,1)}^{(k-1)}$. Then we can always write

$$\begin{bmatrix} A_{(2,1)}^{(k-1)} & A_{(2,2)}^{(k-1)} \end{bmatrix} = \text{Rk } r + C \quad (10)$$

where $\text{Rk } r$ is a matrix of scalar rank equal to r , and C is a correction matrix satisfying

- C is zero in its first $k - 1$ block columns;
- the remaining entries of C , lying in rows with index i_1, \dots, i_r can be chosen *completely ad random* (by appropriate choice of $\text{Rk } r$). In particular, they can all be chosen to be zero except for entries lying on the diagonal of A , which we choose equal to y_k .

To show this structure more explicitly, let us write $C = \begin{bmatrix} 0 & C_{\text{sq}} \end{bmatrix}$ with C_{sq} the square right submatrix of C , having the form

$$C_{\text{sq}} = P^H \begin{bmatrix} y_k I_r & 0 \\ X & X \end{bmatrix} P,$$

with P an appropriate permutation matrix. Now since y_k is an r -fold eigenvalue of C_{sq} , we can always find a unitary matrix Q such that

$$Q^H C_{\text{sq}} Q = \begin{bmatrix} y_k I_r & X \\ 0 & X \end{bmatrix}, \quad (11)$$

which can be obtained by choosing the first r scalar columns of Q to be an orthonormal basis for the eigenspace of y_k . Then denoting with $\tilde{Q} = \begin{bmatrix} I_{m^{(k-1)}} & 0 \\ 0 & Q \end{bmatrix}$ the embedded form of Q , we define the new iterate

$$A^{(k)} := \tilde{Q}^H A^{(k-1)} \tilde{Q} = \begin{bmatrix} A_{(1,1)}^{(k-1)} & A_{(1,2)}^{(k-1)} Q \\ Q^H A_{(2,1)}^{(k-1)} & Q^H A_{(2,2)}^{(k-1)} Q \end{bmatrix}.$$

This matrix will satisfy the block lower semiseparable plus diagonal structure in its first k block columns, as required. Indeed, by (10) the bottom part of $A^{(k)}$ can be written as $Q^H \begin{bmatrix} A_{(2,1)}^{(k-1)} & A_{(2,2)}^{(k-1)} \end{bmatrix} \tilde{Q} = Q(\text{Rk } r + C)\tilde{Q}$, or equivalently

$$\widetilde{\text{Rk } r} + \begin{bmatrix} 0 & Q^H C_{\text{sq}} Q \end{bmatrix},$$

which has the desired structure since $\widetilde{\text{Rk } r} := Q^H(\text{Rk } r)Q$ is again of scalar rank r , and due to the form of the first block column of $Q^H C_{\text{sq}} Q$ in (11).

To be more precise, remark that only the first $r \leq m$ scalar columns of the first block column of $Q^H C_{\text{sq}} Q$ in (11) have the required form yet. To achieve this also for the remaining $m - r$ scalar columns, it suffices to write $\widetilde{\text{Rk } r} + \begin{bmatrix} 0 & Q^H C_{\text{sq}} Q \end{bmatrix} = \widetilde{\text{Rk } r} - E + (E + \begin{bmatrix} 0 & Q^H C_{\text{sq}} Q \end{bmatrix})$ where E is a correction matrix to bring also the remaining $m - r$ scalar columns of $Q^H C_{\text{sq}} Q$ in the required form. Then the matrix $\widetilde{\text{Rk } r} - E$ will be of scalar rank at most $r + (m - r) = m$, which is still of block rank 1, as required.

This ends the description of the k th recursive step of the algorithm. Following these steps, we end up with a matrix $A^{(n-1)}$ which satisfies the block lower semiseparable plus diagonal structure in its first $n - 1$ columns. But then $A^{(n-1)} \in \mathcal{S} + D_y$, and we can conclude by setting $A^{(n)} := A^{(n-1)}$. \square

Now we restrict ourselves to the class $\mathcal{S}_{uv} \subset \mathcal{S}$, and we want to prove a theorem which is similar to Theorem 9. We can expect some problems, however, since $\mathcal{S}_{uv} \subsetneq \mathcal{S}$ is a strict subclass which is not topologically closed. First we prove a lemma.

Lemma 10 *The matrix equation*

$$D_z C + C D_y = s t^T \tag{12}$$

always has a unique solution matrix C ; here

$$s = \begin{bmatrix} s_1^T & \dots & s_n^T \end{bmatrix}^T, \quad t = \begin{bmatrix} t_1^T & \dots & t_n^T \end{bmatrix}^T \tag{13}$$

are arbitrary block vectors with block elements $s_k, t_k \in \mathbb{C}^{m \times m}$. Moreover, if C is nonsingular, then necessarily all the t_k have to be nonsingular.

Remark. Matrix equations of the form $AC + CB = \text{Rk } r$, with $\text{Rk } r$ a low rank matrix and with C unknown, are usually called *displacement equations*. In our case where $A = D_z$ and $B = D_y$ are diagonal matrices, the resulting matrix C is called a *Cauchy-like*

matrix: see ([10]).

PROOF. From general displacement theory ([10]), it follows that the solution matrix C of Equation (12), $D_z C + C D_y = s t^T$ will certainly exist and be unique: the reason for this is that the eigenvalues of the coefficient matrices D_z and D_y are two by two different from each other, thus $z_{i,j} \neq y_k$ for all indices.

From this same assumption, $z_{i,j} \neq y_k$, we can prove also the second statement. Thus suppose that the solution C of Equation (12) is nonsingular. We will prove that all block elements t_k of t have to be nonsingular. Indeed: suppose by contradiction that a nonzero vector $\vec{c} \in \mathbb{C}^m$ exists in the right null space of some t_k^T . Then by evaluating the k th block column of (12) we obtain $D_z C_k - y_k C_k = s t_k^T$ (with C_k denoting the k th block column of C .) Multiplying on the right with \vec{c} yields $D_z C_k \vec{c} - y_k C_k \vec{c} = 0$, and thus $(D_z - y_k I) C_k \vec{c} = 0$. But by the nonsingularity of C , the vector $C_k \vec{c}$ must be different from zero, from which it follows that the matrix $D_z - y_k I$ must be singular. This yields a contradiction since we assumed $y_k \neq z_{i,j}$ for all indices. \square

The following theorem generalizes a result in [18] (see Corollary 14 further on), which was itself a generalization of results proved in [8].

Theorem 11 *Let $q_1 \in \mathbb{C}^{m \times m}$ be fixed (the individual columns of q_1 are supposed to be orthonormal to each other), then the following are equivalent:*

(i) *The inverse eigenvalue problem*

$$Q^H D_z Q = D_y + S_{uv} \quad (14)$$

has a solution (Q, S_{uv}) where $S_{uv} \in \mathcal{S}_{uv}$ and Q is unitary, with first block column fixed to be q_1 .

(ii) *The Cauchy-like matrix C given by the displacement equation*

$$D_z C + C D_y = s t^T \quad (15)$$

is nonsingular; here s and t are block vectors of the same dimension as q_1 , more specifically $s = (D_z - y_1 I) q_1$ and $t = [I_m \ \dots \ I_m]^T$.

Moreover, if one of these conditions is satisfied, then denoting with C the solution to (15), we will have that the first block column of C equals q_1 , and the solutions to (14) are precisely those unitary matrices Q for which $C = QR$ for a certain block uppertriangular matrix R with first block diagonal element being the identity matrix I_m .

PROOF. To establish the connection between Equations (14) and (15), we will use the following series of equivalences:

$$Q^H D_z Q = D_y + S_{uv} \quad (16)$$

$$\Leftrightarrow Q^H D_z Q = D_y + R_{\text{strict}} + \tilde{s} \tilde{t}^T \quad (17)$$

$$\Leftrightarrow Q^H D_z Q = R D_y R^{-1} + \tilde{s} \tilde{t}^T \quad (18)$$

$$\Leftrightarrow D_z(QR) = (QR) D_y + \tilde{\tilde{s}} \tilde{\tilde{t}}^T \quad (R \text{ nonsingular}). \quad (19)$$

Here Q is unitary, R block uppertriangular, R_{strict} strictly block uppertriangular and \tilde{s} , \tilde{t} ; $\tilde{\tilde{s}}, \tilde{\tilde{t}}$ are certain block vectors of the same dimension as q_1 .

Let us prove these equivalences: (16) \Leftrightarrow (17) follows by the definition of S_{uv} . The implication (17) \Rightarrow (18) follows by taking an eigendecomposition $D_y + R_{\text{strict}} = RD_yR^{-1}$. Indeed, the fact that $D_y + R_{\text{strict}}$ is diagonalizable, with the required eigenvalues follows since all its eigenspaces are maximal, due to the fact that the diagonal blocks of D_y (and hence of $D_y + R_{\text{strict}}$) are by definition of the form $y_k I$. The fact that the eigenvectors of $D_y + R_{\text{strict}}$ constitute a block uppertriangular matrix R (which can even be chosen to be scalar uppertriangular), follows easily by inspection. Conversely, the implication (18) \Rightarrow (17) states that RD_yR^{-1} can always be written in the form $D_y + R_{\text{strict}}$. This is possible since D_y has all its block diagonal elements of the form $y_k I_m$, and it is easy to check that this must then also be true for RD_yR^{-1} . Finally, the equivalence (18) \Leftrightarrow (19) follows by some easy left and right matrix multiplications.

Now let us take care of the extra conditions in the statement of the theorem. On the one hand, together with (16) we impose the condition that $Q_1 = q_1$ (with Q_1 denoting the first block column of Q). On the other hand, together with (19) we impose the condition that the matrix $\tilde{\tilde{t}}^T$ can be written as st^T with s and t as specified in the theorem. We will show that these extra conditions are compatible with each other under the equivalences (16) $\Leftrightarrow \dots \Leftrightarrow$ (19) established above.

\Rightarrow Suppose that together with (16) we have the condition that $Q_1 = q_1$. Following the above implications, we already obtained a nonsingular matrix QR satisfying

$$D_z(QR) - (QR)D_y = st^T$$

(we skipped the tildes). But we still have to bring the generators s and t in the form mentioned in the theorem. By multiplying on the right with a *nonsingular* block diagonal matrix D , we obtain

$$D_z(QRD) - (QRD)D_y = st^T D =: st^T, \quad (20)$$

where we redefined $t^T := t^T D$, and where we used the commutativity of D and D_y (since the latter has each block diagonal element a multiple of the identity matrix I_m). Now by appropriately choosing the first block diagonal element of D , we can always make the first block diagonal element of RD to be the identity matrix I_m (remark that R was nonsingular). Then since by assumption Q has its first block column equal to the given value q_1 , this must also hold for $Q(RD)$. By Lemma 10 we know that the block element t_1 is nonsingular, so that we can write

$$st^T = (st_1^T)(t_1^{-T}t^T) =: \tilde{\tilde{t}}^T, \quad (21)$$

where now the first block element of the generator \tilde{t} has become the identity matrix I_m . It follows by evaluating the first block column of (20) and using (21) that \tilde{s} must have the given value $\tilde{s} = (D_z - y_1 I)q_1$, mentioned in the theorem. Finally, by appropriate choice of the $(2, 2), \dots, (n, n)$ block elements of D , also the other block elements of \tilde{t} can be made equal to the identity matrix I_m . (We used here again the previous lemma, guaranteeing that all block elements \tilde{t}_k are nonsingular and can thus be inverted). Now both \tilde{s} and \tilde{t} are brought in the required form. By the uniqueness statement of the

previous lemma, we must then have that

$$C = QRD,$$

which is indeed nonsingular.

\Leftarrow Suppose conversely that the solution matrix C of equation (15), with right hand side st^T as mentioned in the theorem, is nonsingular. It is easy to check that the first block column of C must be equal to q_1 , which are already m orthonormal columns. Taking an arbitrary block QR decomposition $C = QR$ with first block diagonal element of R being the identity matrix I_m , we obviously have that $Q_1 = q_1$ (with Q_1 denoting the first block column of Q), and finally the construction of a matrix $S_{uv} \in \mathcal{S}_{uv}$ such that (Q, S_{uv}) solves Equation (14) follows identically as in the equivalences \Leftrightarrow of above. \square

Remark 12 *Theorem 9, Lemma 10 and Theorem 11 all remain valid if D_z is not necessarily a diagonal matrix, but an arbitrary matrix A satisfying $y_k \notin \text{Spec } A$, the spectrum of A . The matrix C will then not be Cauchy-like, but it is characterized instead by the displacement equation $AC - CD_y = st^T$.*

Corollary 13 *(Essential uniqueness.) If the Cauchy-like matrix C of the previous theorem is nonsingular, then each matrix Q for solving (14) can (in principle) be found by computing the Q factor of a block QR factorization of the matrix C . In particular, due to the essential uniqueness of block QR decomposition, it follows then that Q must be unique up to the right multiplication with a unitary, block diagonal matrix with first block element being the identity matrix I_m .*

Corollary 14 *(Scalar case.) In the scalar case, i.e. block size $m = 1$, the determinant of C can be shown to be zero if and only if either not all the values $\{z_i, y_k\}$ are two by two different from each other, or a certain weight number w_i is equal to zero. Now all these situations are ruled out by our working assumptions of Section 2, except for the condition that $z_i \neq z_{\tilde{i}}$. Thus if we additionally suppose this last condition to be satisfied, i.e. $z_i \neq z_{\tilde{i}}$ for all $i \neq \tilde{i}$, (which is reasonable, since it implies (3) to be not a true inner product), it follows that the inverse eigenvalue problem (9) always has a solution with $S \in \mathcal{S}_{uv}$. This result was also shown in [18].*

For block size $m \geq 2$, the explicitization of Theorem 11 is not so easy. For example, by using the theorem it can be shown that the inverse eigenvalue problem (9) having not a uv -representable solution $S_{uv} \in \mathcal{S}_{uv}$, is not an ‘irreversible’ situation, in the sense that it may be possible by introducing new data points $z_{n+1,j}$, a new pole y_{n+1} and new weight vectors $\vec{w}_{n+1,j}$ to obtain an updated problem which does have a uv -representable solution, even if this was not the case for the original problem. Of course, by Corollary 14 it follows that this situation can not occur in the scalar case. We may also note that this ‘reversibility’ property can be restated in terms of the orthonormal rational function vectors by using Theorems 18, 19 further on.

4 Recursion for the orthonormal rational function vectors.

Convention 15 *From now on we will systematically label the indices of the block rows and columns of an $n + 1$ times $n + 1$ block matrix as $0, 1, \dots, n$. We do this to have the correspondence with the indices $0 \leq i, k \leq n$ of poles, data points and weights.*

In this section we will explain how, from the solution of the inverse eigenvalue problem (9), and using an appropriate initial condition for the first block column of Q , the desired orthonormal rational function vectors can be found.

Let the poles $y_{k,l} = y_k$, data points $z_{i,j}$ and corresponding weight vectors $\vec{w}_{i,j}$ be given. As in Definition 8, we denote with D_y, D_z the diagonal matrices containing poles and data points, respectively. Following the analogy with D_z and D_y , we define

Definition 16 (W). *We define the block column vector $W = [W_0^T \ \dots \ W_n^T]^T \in \mathbb{C}^{(mn+m) \times m}$ by the blocks $W_k^T = [\vec{w}_{k,1} \ \dots \ \vec{w}_{k,m}]$.*

Suppose that (Q, S) is a given solution of Equation (9), i.e. $Q^H D_z Q = D_y + S$, or equivalently

$$D_z Q - Q D_y = Q S, \quad (22)$$

with appropriate initial condition on Q specified by

$$Q^H W = [X^T \ 0 \ \dots \ 0]^T, \quad (23)$$

for a certain block $X \in \mathbb{C}^{m \times m}$. Note that this initial condition on Q can always be realized. In the sequel, it will be used to guarantee that $\phi_{0,l}(z) \in \mathcal{R}_0$, and we will usually suppose that the corresponding coefficient block X is nonsingular, or equivalently that the columns of the block weight vector W are linearly independent.

This condition that W has maximal column rank, is not really a restriction. This is because in the other case, denoting with $\vec{c} \in \mathbb{C}^m$ a vector belonging to the right null space of W , then $\vec{c} \in \mathcal{R}_0 \subseteq \mathcal{R}_v$ and obviously all the weighted evaluations of this constant vector \vec{c} are equal to zero. Thus $\|\vec{c}\| = 0$, i.e. the inner product (3) is not a true inner product on \mathcal{R}_v , and hence the search for an orthonormal basis of \mathcal{R}_v makes no sense.

Now by evaluating the (i, j) th row of (22), we obtain a relation for the (i, j) th scalar row $Q(i, j; :)$

$$Q(i, j; :)(z_{i,j} I_{mn+m} - D_y) = Q(i, j; :) S. \quad (24)$$

Now our intention is to identify Q with a matrix \tilde{Q} as in Definition 4, having as its $(i, j; k, l)$ th element the weighted evaluation of $\vec{\phi}_{k,l}(z)$ in the point $z_{i,j}$. This condition would imply that for each (i, j) ,

$$Q(i, j; :) = \tilde{Q}(i, j; :) = \vec{w}_{i,j}^T \phi(z_{i,j}), \quad (25)$$

with $\phi(z_{i,j})$ the evaluation in $z_{i,j}$ of the 'block row vector' $\phi(z)$ which will be introduced in Definition 17. Thus for each fixed (i, j) , we should be able to identify $Q(i, j; :)$ and

$\phi(z_{i,j})$, except for a constant factor $\vec{w}_{i,j}^T$ on the left. In particular, we can hope to obtain the $\phi(z_{i,j})$ by the same recursion (24).

We will use this observation to determine the desired orthonormal rational function vectors.

Definition 17 *Recalling that $\phi_k(z) = [\vec{\phi}_{k,1}(z) \ \dots \ \vec{\phi}_{k,m}(z)]$, and setting the ‘block row vector’*

$$\phi(z) = [\phi_0(z) \ \dots \ \phi_n(z)] \in \mathbb{C}(z)^{m \times (mn+m)},$$

we define $\phi(z)$ by the equation

$$\phi(z)(zI_{mn+m} - D_y) = (\phi(z) + [0 \ \dots \ 0 \ X(z)])S, \quad (26)$$

where $X(z) \in \mathbb{C}(z)^{m \times m}$ is an extra coefficient block which is added for correctness, but which will be irrelevant for the computation of $\phi(z)$. We also fix the initial block element $\phi_{0,l}(z)$ of $\phi(z)$ by defining $\phi_0(z) = X^{-1}$ where X is the coefficient block defined by (23).

Let us prove that this definition yields us the desired orthonormal rational function vectors. For this, we will have to assume that $S \in S_{uv}$ is uv -representable.

Theorem 18 *(Orthonormal rational function vectors.) Suppose (Q, S_{uv}) is a solution of the inverse eigenvalue problem (9) with first block column of Q defined as in (23), and with $S_{uv} \in S_{uv}$ uv -representable. The value of $y_0 \in \mathbb{C}$ may be chosen arbitrarily. Then defining $\phi(z)$ as in (26) (with $S = S_{uv}$), with starting value $\phi_0(z) = X^{-1} \in \mathbb{C}^{m \times m}$ we have that:*

1. *The block elements $\phi_k(z)$ of $\phi(z)$ can be solved recursively from Equation (26), starting from the known value for $\phi_0(z)$.*
2. *The $\vec{\phi}_{k,l}(z)$ satisfy the desired degree structure, i.e. $\vec{\phi}_{k,l}(z) \in \mathcal{R}_k$ for each k .*
3. *For the associated matrix \tilde{Q} defined by $\tilde{Q}_{i,j;k,l} = \vec{w}_{i,j}^T \vec{\phi}_{k,l}(z_{i,j})$ (Definition 4), we have that $\tilde{Q} = Q$, which is a unitary matrix. (Note that in particular, this implies (25) to be satisfied for each (i, j)).*

Hence the $\vec{\phi}_{k,l}(z)$ form the desired orthonormal basis for \mathcal{R}_Q .

PROOF. Since $S_{uv} \in S_{uv} \subset S$ is block lower semiseparable, we know that it has an RQ factorization $S = RG$ where the G factor consists of a product of n block Givens transformations

$$G = G^{(n-1,n)} \dots G^{(2,3)} G^{(0,1)} \quad (27)$$

with each $G^{(k,k+1)}$ acting on block columns $k, k+1$. (Note that we count now from zero instead of one). Thus Equation (26) is equivalent with

$$\phi(z)(zI_{mn+m} - D_y)G^H = (\phi(z) + [0 \ \dots \ 0 \ X(z)])R. \quad (28)$$

The reader can check that the matrix

$$H(z) := (zI_{mn+m} - D_y)G^H \quad (29)$$

is *block Hessenberg*, due to the order of the block Givens transformations in G , and that its elements on block row k are scalar multiples of $z - y_k$. A sketch of this Hessenberg matrix looks as follows:

$$H(z) = \begin{bmatrix} (z-y_0)H_{0,0} & & \cdots & & (z-y_0)H_{0,n} \\ (z-y_1)H_{1,0} & & & & \\ 0 & & & & \vdots \\ \vdots & \ddots & & & \\ 0 & \cdots & 0 & (z-y_n)H_{n,n-1} & (z-y_n)H_{n,n} \end{bmatrix},$$

with the $H_{i,k} \in \mathbb{C}^{m \times m}$ appropriate coefficient blocks. Similarly, we can write the matrix R in the right hand side of (28) as

$$R = \begin{bmatrix} R_{0,0} & \cdots & R_{0,n} \\ 0 & \ddots & \vdots \\ \vdots & \ddots & \\ 0 & \cdots & 0 & R_{n,n} \end{bmatrix},$$

with the $R_{i,k} \in \mathbb{C}^{m \times m}$.

1. From the above discussion it follows that, starting with the initial value $\phi_0(z)$ we can find the subsequent $\phi_{k+1}(z)$, $k = 0, 1, \dots, n-1$ by solving each time block column k of the Hessenberg system (28).

The only problem that could occur in this way is when one of the subdiagonal blocks of $H(z) = (zI_{mn+m} - D_y)G^H$ does not have full rank, since then it is not sure that we can solve the block $\phi_{k+1}(z)$ from this equation. But the latter can only happen if one of the Givens transformations of G has a singular (1,2) block element. By Lemma 7, the latter would mean in turn that either S_{uv} would not be uv -representable, which is a contradiction, or that S_{uv} would have its last block row or first block column of scalar rank $< m$. Let us show that under our working assumptions, the latter is impossible.

Suppose by contradiction that $S_{uv,0}$, the first block column of S_{uv} , would be of scalar rank $< m$. Thus there exists a nonzero vector $\vec{c} \in \mathbb{C}^m$ in the right null space of $S_{uv,0}$. Recalling that S_{uv} was introduced as the solution of the inverse eigenvalue problem (9), or equivalently (22), $D_z Q = Q D_y + Q S_{uv}$, it follows by evaluating the first block column that $D_z Q_0 - y_0 Q_0 = Q S_{uv,0}$ (with Q_0 denoting the first block column of Q .) Multiplying on the right with \vec{c} yields $D_z Q_0 \vec{c} - y_0 Q_0 \vec{c} = 0$, and thus $(D_z - y_0 I) Q_0 \vec{c} = 0$. But by the nonsingularity of Q , the vector $Q_0 \vec{c}$ must be different from zero, from which it follows that the matrix $D_z - y_0 I$ must be singular. This yields a contradiction since we assumed $y_k \neq z_{i,j}$ for all indices. The case where the last block row of S_{uv} would have scalar rank $< m$, can be proven in the same way, this time by transforming (9) into the equivalent equation $Q^H D_z = D_y Q^H + S_{uv} Q^H$ and then evaluating the last block row.

2. To prove the desired degree structure, note that the subdiagonal block $H_{k+1,k}$ in the block column k of the block Hessenberg matrix $H(z)$ can be written as $(z - y_{k+1})I_m$, multiplied with the (2,1) block element of the unitary transformation $(G^{(k,k+1)})^H$. Then using block column k of (28) to solve for the block $\phi_{k+1}(z)$, we will have to invert this matrix with as consequence an extra pole y_{k+1} being introduced. (This is one of the points where we use our assumption $y_{k,1} = \dots = y_{k,m} =: y_k$ for each k !) By induction, it follows that $\vec{\phi}_{k,l}(z) \in \mathcal{R}_k$, $k = 0, 1, \dots, n$, $l = 1, \dots, m$.
3. We will prove that $\tilde{Q} = Q$. More precisely: let us fix indices i, j , then we will prove that the (i, j) th row of \tilde{Q} and Q are equal to each other. Premultiplying the Hessenberg system (28) with the weight vector $\vec{w}_{i,j}$ and evaluating in the corresponding data point $z_{i,j}$, we find that the row vector $\tilde{Q}(i, j; :)$ must satisfy

$$\tilde{Q}(i, j; :)(z_{i,j}I_{mn+m} - D_y)G^H = (\tilde{Q}(i, j; :) + [0 \quad \dots \quad 0 \quad \vec{w}_{i,j}X(z_i)])R. \quad (30)$$

Comparing this with (24), we see that this is exactly the same recursion as the one satisfied by the (i, j) th row of Q . (Except for the term $\vec{w}_{i,j}X(z_i)R$ in the last block column of (30), but this term is only added for correctness and will not be needed to actually determine $\tilde{Q}(i, j; :)$). Now since $S \in \mathcal{S}_{uv}$ is uv -representable, the subdiagonal blocks of the Hessenberg matrix G^H are invertible, and thus by solving the subsequent columns of (30) we can find the subsequent block columns of Q , \tilde{Q} . Since also the first block columns of Q and \tilde{Q} are equal, by construction (compare (23) and the definition of $\phi_0(z)$ in Definition 17), we conclude that $\tilde{Q} = Q$.

□

We now proved that in the uv -representable case, Equation (9) together with the initial condition (23) always yields us the required orthonormal rational function vectors. Next we investigate what could be called the *singular case*, i.e. when Equation (9) and initial condition (23) do not have a solution (Q, \mathcal{S}_{uv}) . By Theorem 9, this implies that there exists a solution of the form (Q, S) , where $S \in \mathcal{S} \setminus \mathcal{S}_{uv}$ is *not* uv -representable. We then obtain the following

Theorem 19 (*Singular case.*) *Suppose (Q, S) is a solution of Equation (9), with initial condition on Q specified by (23) and with $S \in \mathcal{S} \setminus \mathcal{S}_{uv}$ being not uv -representable. Let $S = RG$ be a block RQ decomposition, with G as in (27), thus counting from zero. Let $0 \leq k \leq n-1$ be the smallest index such that the block Givens transformation $G^{(k,k+1)}$ is singular. (See Lemma 7; moreover we implicitly used here the convention about the choice of the (2,1) block element which is used in the formulation of Lemma 7). Then there exists a nonzero element $\vec{\Psi}(z) \in \mathcal{R}_{k+1}$ such that $\|\vec{\Psi}(z)\| = 0$. This means that $\langle \cdot, \cdot \rangle$ is not a true inner product for \mathcal{R}_n , and hence that no orthonormal basis exists for \mathcal{R}_n .*

PROOF. Remark that we will have no problems to find the blocks $\phi_{\tilde{k}}$, $\tilde{k} = 0, \dots, k$. For this we can just take over the proof of the previous theorem, by solving the block columns $0, \dots, k-1$ of the Hessenberg system (28), together with the given value for

ϕ_0 . In the same way, we will have no problem to find the block columns $0, \dots, k$ of \tilde{Q} , defined by $\tilde{Q}_{i,j;\bar{k},l} = \bar{w}_{i,j}^T \tilde{\phi}_{\bar{k},l}(z_{i,j})$ (Definition 4). It follows in the same way as in the previous theorem that the first block columns $0, \dots, k$ of \tilde{Q} are equal to those of Q , and hence orthonormal to each other. Also (25) will then be valid.

The block column with index k of the Hessenberg system (28) can not be solved for $\phi_{k+1}(z)$, however.

On the other hand, this insolvability must have a consequence for the matrix Q . Given row indices (i, j) , we recall the recursion (24), $Q(i, j; :)(z_{i,j}I_{mn+m} - D_y) = Q(i, j; :S)$ or equivalently

$$Q(i, j; :H(z_{i,j})) = Q(i, j; :R), \quad (31)$$

where $H(z_{i,j})$ is the evaluation in the data point $z_{i,j}$ of the Hessenberg matrix $H(z) = (zI_{mn+m} - D_y)G^H$ defined in (29). Recalling also the relation (25),

$$Q(i, j; :) = \bar{w}_{i,j}^T \phi(z_{i,j}),$$

which we already remarked to be valid for the block columns $0, \dots, k$ of Q , we obtain from (31) that

$$\bar{w}_{i,j}^T \tilde{\Psi}(z_{i,j}) = 0$$

where

$$\begin{aligned} \tilde{\Psi}(z) := & \phi_0(z)(z - y_0)H_{0,k}\bar{c} + \dots + \phi_k(z)(z - y_k)H_{k,k}\bar{c} + \vec{0} \\ & - \phi_0(z)R_{0,k}\bar{c} - \dots - \phi_k(z)R_{k,k}\bar{c}, \end{aligned} \quad (32)$$

with $\bar{c} \in \mathbb{C}^m$ a nonzero vector belonging to the right null space of $H_{k+1,k}$.

Hence $\tilde{\Psi}(z) \in \mathcal{R}_k + z\mathcal{R}_k$ is an element for which all the weighted evaluations are zero, which implies $\tilde{\Psi}(z)$ to have norm equal to zero. Equivalently, the element $\tilde{\Psi}(z)/(z - y_{k+1}) \in \mathcal{R}_{k+1}$ has norm equal to zero. If we can prove that $\tilde{\Psi}(z) \neq 0$, the theorem is proved.

To conclude the proof, let us show that indeed $\tilde{\Psi}(z) \neq 0$. By postmultiplying block column k of the Hessenberg system (28) by a nonsingular m by m matrix of the form $\begin{bmatrix} \bar{c} & \vec{X} \end{bmatrix}$, with \bar{c} the null space vector of above, we may assume that $H_{k+1,k} = \begin{bmatrix} \vec{0}_m & \vec{X} \end{bmatrix}$, and then the $(k, 1)$ th column of the Hessenberg system (28) will exactly be: ' $\tilde{\Psi}(z) = 0$ '. We will thus try to obtain a contradiction, thereby showing in fact that the Hessenberg system defining the orthonormal rational function vectors is not solvable.

Denoting with $\vec{h} := H(:, k, 1)$ the $(k, 1)$ th column of H , we have by assumption that $\vec{h}_{k+1} = \vec{0}_m$. Since we assumed all the previous subdiagonal blocks of H to be nonsingular, we can apply an elementary, nonsingular column transformation T on H that further eliminates all the entries of \vec{h} , except for the first block element. Updating

$$H := HT, \quad R := RT, \quad \vec{h} := H(:, k, 1), \quad \vec{r} := R(:, k, 1), \quad (33)$$

then there are two possibilities: (a) $\vec{h}_0 = \vec{0}_m$. But then H would be singular, which is impossible since H has been defined in (29) as a product of nonsingular matrices. (b) $\vec{h}_0 \neq \vec{0}_m$. But then we get a contradiction since by evaluating the $(k, 1)$ th column of

(28), it would follow that $(z-y_0)\phi_0(z)\vec{h}_0 = \phi_0(z)\vec{r}_0 + \dots + \phi_k(z)\vec{r}_k$, which is impossible since the degree of the numerator of the right hand side can never exceed the degree of its denominator. Thus we obtain a contradiction, which proves that indeed $\vec{\Psi}(z) \neq 0$. \square

We have now established a theoretical connection between orthonormal rational function vectors and the matrix problem (9). In the next section, we will return to the practical aspects of solving this linear algebra problem.

5 A recursive solution for the inverse eigenvalue problem.

In this section we devise an efficient recursive procedure for the construction of the matrices Q and S solving the inverse eigenvalue problem (9) together with the initial condition (23), assuming that an uv -representable solution $S_{uv} \in \mathcal{S}_{uv}$ exists.

Our procedure for solving this problem is recursive. This means that information about poles, data points and weights can be introduced, and the problem can be solved in a stepwise way.

Following Definitions 8 and 16, we write

$$D_z = \begin{bmatrix} D_{z,0:n-1} & 0 \\ 0 & D_{z,n} \end{bmatrix}, \quad D_y = \begin{bmatrix} D_{y,0:n-1} & 0 \\ 0 & D_{y,n} \end{bmatrix}, \quad W = \begin{bmatrix} W_{0:n-1} \\ W_n^T \end{bmatrix},$$

where, with an obvious notation, we have written $D_{z,0:n-1}$, $D_{y,0:n-1}$ and $W_{0:n-1}$ for the principal submatrices containing the information about data points $z_{i,j}$, poles y_k and weight vectors $\vec{w}_{i,j}$ with indices $0 \leq i, k \leq n-1$. The m by m blocks $D_{z,n}$, $D_{y,n}$ and W_n contain the information corresponding to the indices $i = n$ and $k = n$.

Suppose now that we solved the problem for these submatrices $D_{z,0:n-1}$, $D_{y,0:n-1}$, $W_{0:n-1}$, i.e. that we found matrices $Q_{0:n-1}$, $S_{0:n-1} = R_{\text{strict},0:n-1} + \text{Rk } 1_{0:n-1} \in \mathcal{S}_{uv}$ solving (9),

$$Q_{0:n-1}^H D_{z,0:n-1} Q_{0:n-1} = D_{y,0:n-1} + R_{\text{strict},0:n-1} + \text{Rk } 1_{0:n-1}, \quad (34)$$

together with the initial condition (23),

$$Q_{0:n-1}^H W_{0:n-1} = [X^T \quad 0 \quad \dots \quad 0]^T. \quad (35)$$

Defining a permutation matrix $P = \begin{bmatrix} 0 & I_m \\ I_m & 0 \end{bmatrix}$, note that a right multiplication with P brings the right most block column of a matrix completely to the left, and left multiplication with P^H brings the bottom block row completely to the top. Defining the embedded matrices

$$Q = \begin{bmatrix} Q_{0:n-1} & 0 \\ 0 & I_m \end{bmatrix} P, \quad D_y^{(-1,0)} = P^H \begin{bmatrix} D_{y,0:n-1} & 0 \\ 0 & D_{z,n} \end{bmatrix} P,$$

$$R_{\text{strict}} = P^H \begin{bmatrix} R_{\text{strict},0:n-1} & 0 \\ 0 & 0_m \end{bmatrix} P, \quad \text{Rk } 1 = P^H \begin{bmatrix} \text{Rk } 1_{0:n-1} & 0 \\ 0 & 0_m \end{bmatrix} P,$$

Equations (34) and (35) can be easily embedded as

$$Q^H D_z Q = D_y^{(-1,0)} + R_{\text{strict}} + \text{Rk } 1 \quad (36)$$

and

$$Q^H W = [W_n^T \quad X^T \quad 0 \quad \dots \quad 0]^T. \quad (37)$$

Indeed: this can be directly seen by just factoring out the factors P^H, P . Note however that the matrix $D_y^{(-1,0)}$ is still far from being D_y , because of two reasons: (i) it has a disturbed (0,0) block element $D_{z,n}$, (ii) the other diagonal elements are in a wrong, permuted order. For this, we will now start to restore the right hand sides of these two equations in the form that is desired by (9) and (23).

First we bring the right hand side of (37) in the form required by (23). It is obvious that we can do this by applying a unitary transformation $(G^{(0,1)})^H$ on the *first two block rows* of (37).

Then we would like to update Q by absorbing the factor $G^{(0,1)}$ into it; let us denote the new Q factor by $Q^{(0,1)} = QG^{(0,1)}$. Of course, as a by product this updating will also influence relation (36). But this influence will be limited. To see this, note that

$$\text{Rk } 1^{(0,1)} := (G^{(0,1)})^H \text{Rk } 1 G^{(0,1)} \quad (38)$$

is again of block rank 1 and

$$R_{\text{strict}}^{(0,1)} := (G^{(0,1)})^H R_{\text{strict}} G^{(0,1)}$$

is again strictly block uppertriangular. It follows that (36) implies the new equation

$$(Q^{(0,1)})^H D_z Q^{(0,1)} = D_y^{(0,1)} + R_{\text{strict}}^{(0,1)} + \text{Rk } 1^{(0,1)}, \quad (39)$$

where $R_{\text{strict}}^{(0,1)} + \text{Rk } 1^{(0,1)} \in \mathcal{S}_{uv}$, and where

$$D_y^{(0,1)} := (G^{(0,1)})^H D_y^{(-1,0)} G^{(0,1)} + C$$

differs from the desired diagonal matrix D_y because of two reasons: (i) the diagonal elements y_k are in the wrong order, and (ii) it contains additional disturbances in its (0,0), (1,0) and (1,1) block elements. Here C is a correction matrix for the (0,1) block element of $D_y^{(0,1)}$, and its presence can be justified by similarly updating $R_{\text{strict}}^{(0,1)} := R_{\text{strict}}^{(0,1)} - C$.

To achieve the desired form for $D_y^{(0,1)}$, we will first restore the (1,0) and (0,0) block elements. Since the matrix $\text{Rk } 1^{(0,1)}$ is of block rank 1, let us write it as

$$\text{Rk } 1^{(0,1)} = UV^T = [U_0^T \quad \dots \quad U_n^T]^T [V_0^T \quad \dots \quad V_n^T]. \quad (40)$$

We will now assume that V_0^T has full rank. (The case where V_0^T does not have full rank is investigated in Theorem 21). Then we can find solutions U_0', U_1' of the equations $U_0 V_0^T - U_0' V_0^T = y_0 I_m - D_y^{(0,1)}(0,0)$ and $U_1 V_0^T - U_1' V_0^T = -D_y^{(0,1)}(1,0)$, respectively.

Now we *redefine the matrix* $\text{Rk } 1^{(0,1)}$ by taking U'_0, U'_1 as new values for U_0, U_1 , i.e. $\text{Rk } 1^{(0,1)} := \text{Rk } 1^{(0,1)} + C$ where

$$C = \begin{bmatrix} (U'_0 V^T - U_0 V^T)^T & (U'_1 V^T - U_1 V^T)^T & 0^T \end{bmatrix}^T$$

is a correction matrix for the first two block rows. Then to keep (39) valid, we can tackle this correction by a ‘complementary’ update

$$D_y^{(0,1)} := D_y^{(0,1)} - C,$$

and due to the form of C it follows that the $(0,0)$ and $(1,0)$ block elements of $D_y^{(0,1)}$ will be restored in their desired form.

Thus we may assume that $D_y^{(0,1)}$ only differs from the desired diagonal matrix D_y because of its *diagonal blocks*: all (k,k) block elements should be moved upwards to block position $(k-1, k-1)$, $k = 2, 3, \dots, n$, and the $(1,1)$ block element contains some ‘random’ value which should be removed.

To solve this, we will use a kind of chasing technique to chase the $(1,1)$ block element to the bottom of the matrix, thereby shifting upwards the lower block elements. We will next describe one step in the chasing procedure.

Suppose that we applied the chasing algorithm down until block row k , resulting in a matrix $D_y^{(k-1,k)}$. Then let us define A as the matrix consisting of the (k,k) , $(k, k+1)$, $(k+1, k)$ and $(k+1, k+1)$ block elements of $R_{\text{strict}}^{(k-1,k)} + D_y^{(k-1,k)}$. Then

$$A = \begin{bmatrix} X & X \\ 0 & y_k I_m \end{bmatrix}.$$

It follows that y_k is an m -fold eigenvalue of A , and moreover it has an m -dimensional eigenspace. Hence by choosing an orthonormal basis for this eigenspace and completing it in an arbitrary way to an orthonormal basis of whole \mathbb{C}^{2m} , we obtain a unitary matrix $G^{(k,k+1)}$ such that

$$(G^{(k,k+1)})^H A G^{(k,k+1)} = \begin{bmatrix} y_k I_m & X \\ 0 & X \end{bmatrix}.$$

Then if we define $Q^{(k,k+1)} = Q^{(k-1,k)} G^{(k,k+1)}$ (where $G^{(k,k+1)}$ is in fact the embedded form of the above $2m$ by $2m$ matrix, by slight abuse of notation), we obtain the relation

$$(Q^{(k,k+1)})^H D_z Q^{(k,k+1)} = D_y^{(k,k+1)} + R_{\text{strict}}^{(k,k+1)} + \text{Rk } 1^{(k,k+1)}, \quad (41)$$

where $R_{\text{strict}}^{(k,k+1)}$, $\text{Rk } 1^{(k,k+1)}$ are again strictly block uppertriangular and of block rank 1, respectively, and $D_y^{(k,k+1)}$ only differs from the desired diagonal matrix D_y by its block diagonal elements on positions $k+1, \dots, n+1$.

By performing all these chasing steps, at the end we will finish with a matrix $D_y^{(n-1,n)}$ which is strictly block uppertriangular except for its (n,n) block element. But then we can use exactly the same method as we used in the beginning of the procedure: let us write

$$\text{Rk } 1^{(n-1,n)} = \tilde{U} \tilde{V}^T = \begin{bmatrix} \tilde{U}_0^T & \dots & \tilde{U}_n^T \end{bmatrix}^T \begin{bmatrix} \tilde{V}_0^T & \dots & \tilde{V}_n^T \end{bmatrix} \quad (42)$$

(we used tildes to distinguish from (40)). We will now assume that \tilde{U}_n has full rank. (The case where \tilde{U}_n does not have full rank is investigated in Theorem 21). Then we can always find a solution \tilde{V}'_n of the equation $\tilde{U}_n \tilde{V}'_n{}^T - \tilde{U}_n (\tilde{V}'_n)'^T = y_n I_m - D_y^{(n-1,n)}(n, n)$. Now we *redefine the matrix* $\text{Rk } 1^{(n-1,n)}$ by taking \tilde{V}'_n as the new value for \tilde{V}_n , i.e. $\text{Rk } 1^{(n-1,n)} := \text{Rk } 1^{(n-1,n)} + C$ where

$$C = \begin{bmatrix} 0 & U(\tilde{V}'_n)^T - U\tilde{V}_n^T \end{bmatrix}$$

is a correction matrix for the last block column. Then to keep (41) valid, we can tackle this correction by a ‘complementary’ update

$$D_y^{(n-1,n)} := D_y^{(n-1,n)} - C,$$

and due to the form of C it follows that the (n, n) block element of $D_y^{(n-1,n)}$ will be restored in its desired form.

This ends our description of an inductive step of the algorithm.

Remark 20 (a). *In this way we can achieve even more: the freedom that we had in choosing the orthonormal base of eigenvectors for the $2m$ by $2m$ matrix A , earlier in this algorithm, can be used to make R_{strict} not only block strictly uppertriangular, but also scalar uppertriangular with all elements on the diagonal and first $m - 1$ superdiagonals equal to zero.*

(b). *The procedure used in this section has some connections with the general procedure in the proof of Theorem 9. The difference is that in the latter proof, we used a kind of eigenvalue placing technique, while in the current section, we merely use an eigenvalue chasing technique. However, the algorithm of this section is more efficient to solve our linear algebra problem, and it is recursive. As a drawback, recall that the algorithm of this section can only find an uv-representable solution (if it exists).*

(c). *The procedure will still work if D_z is not necessarily block diagonal, but also when $H := D_z$ may be block lower Hessenberg. The solution will then still be ‘recursive’ in the sense that, to update the solution, the last block row and the last block column of the matrix H are added (the latter having only two nonzero block elements). Indeed: recall that in the beginning, a permutation P is applied to bring these new data to the left and to the top. The crucial point is then that after applying the first block Givens transformation $G^{(0,1)}$, the matrix $D_y^{(0,1)}$ will still deviate from the desired D_y because of two reasons: (i) the diagonal elements are in a permuted order, and (ii) it contains additional disturbances in its $(0,0)$, $(1,0)$ and $(1,1)$ block elements — and nowhere else, because of the Hessenberg structure of H .*

Next we give a breakdown analysis of the algorithm.

Theorem 21 (Breakdown analysis). *Suppose that the solution of the inverse eigenvalue problem (9) with initial condition (23) has been found successfully for the data points $z_{i,j}$, poles y_k and weight vectors $\vec{w}_{i,j}$ with indices $0 \leq i, k \leq n - 1$, and that we use the algorithm of the present section to expand this solution for the new indices $i = n$, $k = n$. Then theoretically, the algorithm can break down at two places:*

1. A breakdown in the beginning can occur when the block V_0^T is singular. But this is rather unlikely to happen; for example it implies the weight block vector $W_{0:n-1}$ to be rankdeficient.
2. A breakdown at the end can occur when the block \tilde{U}_n is singular. In this case, if we agree to break down the algorithm at this point, we will have obtained a solution (Q, S) to (9) where S is not uv -representable. However when block size $m \geq 2$, such a breakdown is not ‘irreversible’ in the sense that it may be possible to repair the breakdown by applying the algorithm again, taking into consideration a new set of data points $z_{n+1,j}$, a new pole y_{n+1} and new weight vectors $\vec{w}_{n+1,j}$.
3. In the scalar case (block size $m = 1$), we can say more: if data points and poles z_i, y_k are two by two different from each other, and $w_i \neq 0$ for all i , then no breakdown can occur during the algorithm.

PROOF.

1. Let us recall the following steps of the algorithm: $\text{Rk } 1 = \begin{bmatrix} 0_2 & 0 \\ 0 & \text{Rk } 1_{0:n-1} \end{bmatrix}$ (initialization), $\text{Rk } 1^{(0,1)} := (G^{(0,1)})^H \text{Rk } 1 G^{(0,1)}$ and $\text{Rk } 1^{(0,1)} = UV^T$. Partitioning the first Givens transformation

$$H := G^{(0,1)} = \begin{bmatrix} H_{0,0} & H_{0,1} \\ H_{1,0} & H_{1,1} \end{bmatrix},$$

it follows that (with an obvious notation)

$$\begin{bmatrix} V_0^T & V_1^T \end{bmatrix} = \begin{bmatrix} 0 & V_{0,0:n-1}^T \end{bmatrix} \begin{bmatrix} H_{0,0} & H_{0,1} \\ H_{1,0} & H_{1,1} \end{bmatrix}.$$

Hence singularity of V_0^T can only occur if the first block column of this equation is singular. But since we assumed a successful application of the algorithm for all indices $0 \leq i, k \leq n-1$, the block $V_{0,0:n-1}^T$ on the right hand side must be nonsingular. It follows that $H_{1,0}$ is singular, hence due to the fact that H is unitary, $H_{0,1}$ is singular too: see the proof of Lemma 7 (i) \Leftrightarrow (ii). But recalling the definition of $H = G^{(0,1)}$,

$$\begin{bmatrix} H_{0,0}^H & H_{1,0}^H \\ H_{0,1}^H & H_{1,1}^H \end{bmatrix} \begin{bmatrix} W_n^T & X_{0:n-1}^T \end{bmatrix}^T = \begin{bmatrix} X^T & 0 \end{bmatrix}^T,$$

then evaluating the last block row of this equation yields that $H_{0,1}^H W_n + H_{1,1}^H X_{0:n-1} = 0$. By multiplying on the left with a nonzero vector \vec{c}^T in the left null space of $H_{0,1}^H$, we obtain $\vec{c}^T H_{1,1}^H X_{0:n-1}$. Then since $\vec{c}^T H_{1,1}^H \neq 0$ (because this would imply $\vec{c}^T \begin{bmatrix} H_{0,1}^H & H_{1,1}^H \end{bmatrix} = \vec{0}$, implying H to be singular), we conclude that the block $X_{0:n-1}$ must be singular. But since $X_{0:n-1}$ has been obtained in the previous applications of the algorithm by ‘compressing’ the weight block vector $W_{0:n-1}$ by unitary row operations, it follows that $W_{0:n-1}$ must be singular.

2. Suppose that the block \tilde{U}_n is singular. We agree to break down the algorithm at this point, i.e. we will not try to bring the last block diagonal element of $D_y^{(n-1,n)}$ in the required form $y_n I_m$. Writing

$$D_y^{(n-1,n)} = D_y + C$$

with C a correction matrix for the (n, n) element, it can be easily checked that

$$S := R_{\text{strict}}^{(n-1,n)} + \text{Rk } 1_{n-1,n} + C$$

belongs to $\mathcal{S} \setminus \mathcal{S}_{uv}$.

Now suppose the block size $m \geq 2$, and suppose that we have new data points $z_{n+1,j}$, a new pole y_{n+1} and new weight vectors $\vec{w}_{n+1,j}$ being introduced. We can then apply the chasing steps of the algorithm to update our solution of (9), (23). We do not carry out the ‘bottom’ chasing step, since this step would be incorrect anyway because our starting matrix $D_y^{(n-1,n)}$ (of the previous step) contained a wrong element on its last block diagonal position. Thus we are not finished yet, since the last two block elements of the D_y matrix are not in the required form. But we can always try to find a way out by using the last ‘eigenvalue placing’ step of the constructive proof of Theorem 9, i.e. the theorem constructing a solution of $Q^H D_z Q = D_y + S$ with $S \in \mathcal{S}$ not necessarily uv -representable. It turns out that, since the block size $m \geq 2$, we have a good chance for the obtained S to be uv -representable.

3. Suppose the block size m is 1. A first breakdown possibility is when the element v_0 would be zero. But then from part 1 it follows that the weight vector $\vec{w}_{0:n-1}$ must be the zero vector, which contradicts our assumption that $w_i \neq 0$ for all i . A second breakdown possibility is when the element \tilde{u}_n would be zero. We proved in part 2 that this yields a solution $Q^H D_z Q = D_y + S$ with S not uv -representable. Then it follows from Theorem 19 and Remark 14 that either one of the weights $w_i = 0$, or the numbers z_i, y_k are not two by two different from each other. This yields a contradiction with our assumptions.

□

6 Special configurations of the data points

For certain configurations of the data points $z_{i,j}$, it turns out that the matrix $S \in \mathcal{S}$ for solving the inverse eigenvalue problem (9) is not only block lower semiseparable, but will additionally be *almost block upper semiseparable* in the sense that its Hermitian S^H satisfies Definition 5, but with the adaption that the ‘block lowertriangular part’ should be replaced by the ‘strictly block lowertriangular part’ of S^H .

The special configuration for the $z_{i,j}$ is that they lie on a *generalized circle* in the complex plane, i.e. that the $z_{i,j}$ satisfy a relation $f(z, \bar{z})$ describing either a circle, or a straight line in \mathbb{C} .

The following theorem is based on a result that was proved in [18], the latter using a general result about Möbius transformations, lending itself also to implement the actual computations. Rather than generalizing this, we prefer to give a direct (but easy to understand) proof, not taking care of implementational aspects.

Theorem 22 *Suppose all data points $z_{i,j}$ are lying on a generalized circle of the complex plane, as just defined. Suppose we have an arbitrary solution (Q,S) of the inverse eigenvalue problem (9), i.e.*

$$Q^H D_z Q = D_y + S, \quad (43)$$

with $S \in \mathcal{S}$ block lower semiseparable and without initial condition on the matrix Q . Then partitioning

$$S = \begin{bmatrix} S_{1,1} & S_{1,2} \\ S_{2,1} & S_{2,2} \end{bmatrix}, \quad (44)$$

with $S_{1,1}$ square of block size k , it follows that both $S_{2,1}$ and $S_{1,2}$ must be of block rank at most 1. (Even more: since $S \in \mathcal{S}$, this will hold even if $S_{2,1}$ is extended with one block row on its top). In particular, S can be represented by $O(n)$ information.

Remark. A matrix S that satisfies the conditions on the $S_{1,2}$ and (extended) $S_{2,1}$ block in the partitioning (44), for every k , is often called a *block semiseparable* matrix (without adjective lower or upper). Note that this definition implies mostly, but not always that $S^H - D \in \mathcal{S}$ for some correcting diagonal matrix D . A counterexample for the latter is when S is block upper bidiagonal, for in this case the matrix S will satisfy the definition based on (44), but it will usually not allow a diagonal matrix D as just described, the underlying reason being that its block diagonal elements are ‘tending to infinity’.

PROOF. Note that (43) implies that $Q^H D_z Q$ and S only differ by a diagonal correction, and so it is sufficient to prove that by partitioning

$$Q^H D_z Q = \begin{bmatrix} H_{1,1} & H_{1,2} \\ H_{2,1} & H_{2,2} \end{bmatrix} \quad (45)$$

with $H_{1,1}$ square of block size k , as in (44), then both $H_{1,2}$ and $H_{2,1}$ will be of block rank at most 1.

First we prove the case where all $z_{i,j} \in \mathbb{R}$. Then obviously

$$(Q^H D_z Q)^H = Q^H D_z^H Q = Q^H D_z Q$$

for any unitary Q , and so this matrix is Hermitian. Furthermore, partitioning $Q^H D_z Q$ as in (45), it follows from (43) that $H_{2,1}$ has block rank at most 1, for every k . But then since $Q^H D_z Q$ is Hermitian, obviously the same must hold for $H_{1,2}$.

Next suppose that all $z_{i,j}$ lie on the unit circle, thus $|z_{i,j}| = 1$. Then obviously

$$(Q^H D_z Q)^{-1} = Q^H D_z^{-1} Q = Q^H D_z^H Q = (Q^H D_z Q)^H$$

for any unitary Q , and so this matrix is unitary. Furthermore, partitioning $Q^H D_z Q$ as in (45), it follows again from (43) that $H_{2,1}$ has block rank at most 1. But then since

$Q^H D_z Q$ is unitary, we can apply Lemma 7(i) \Leftrightarrow (ii) (which can be easily extended to prove that the *ranks* must be exactly equal), and it follows that the same must hold for $H_{1,2}$.

Next suppose that we apply a translation, i.e. $\tilde{z}_{i,j} = z_0 + z_{i,j}$ for a certain $z_0 \in \mathbb{C}$. Then

$$Q^H D_{\tilde{z}} Q = Q^H (z_0 I + D_z) Q = Q^H D_z Q + z_0 I.$$

Thus this matrix only differs from $Q^H D_z Q$ by some block diagonal correction, and if $z_{i,j} \in \mathbb{R}$ or $|z_{i,j}| = 1$, the same conclusions will hold for the $\tilde{z}_{i,j}$.

Finally, suppose that we apply a scaling, i.e. $\tilde{z}_{i,j} = r_0 z_{i,j}$ for a certain $r_0 \in \mathbb{C}$. Then

$$Q^H D_{\tilde{z}} Q = Q^H (r_0 D_z) Q = r_0 (Q^H D_z Q).$$

Thus this matrix only differs from $Q^H D_z Q$ by some scaling factor, and if $z_{i,j} \in \mathbb{R}$, $|z_{i,j}| = 1$, or some translation of one of these cases, the same conclusions will hold for the $\tilde{z}_{i,j}$. Since any generalized circle can be obtained as a composition of the above cases, the theorem is proved. \square

Recalling that in the recursive algorithm of Section 5, the intermediate matrices (Q, S) satisfied a relation of the form $Q^H D_z Q = \tilde{D}_y + S$, with \tilde{D}_y still block diagonal, but deviating from the desired D_y because its block diagonal elements are in a permuted order, it follows that we can apply Theorem 22 for these intermediate matrices too. Thus all the intermediate matrices allow a representation with only $O(n)$ parameters, instead of $O(n^2)$. It follows that the global algorithm can be implemented an order of magnitude faster than for the generic case, i.e. with $O(n^2)$ operations instead of $O(n^3)$. However, it is not the aim of this paper to go into the technical details of achieving this implementation.

7 Conclusion

In this paper we solved the problem of obtaining an orthonormal basis $\{\phi_k(z)\} \subset \mathcal{R}_w$, with \mathcal{R}_w the space of proper rational function vectors having certain, prescribed poles. The problem was reduced to an inverse eigenvalue problem involving a lower semiseparable matrix, and we described an efficient, recursive algorithm for solving this inverse eigenvalue problem. The paper was merely of a theoretical kind: a detailed implementation of the algorithm, and numerical experiments to illustrate its effectiveness, especially when the data points are on a generalized circle, will be the subject of further research.

References

- [1] G. Ammar, W. Gragg, and L. Reichel. Constructing a unitary Hessenberg matrix from spectral data. In G. H. Golub and P. Van Dooren, editors, *Numerical Linear Algebra, Digital Signal Processing and Parallel Algorithms*, volume 70 of *Computer and Systems Sciences*, pages 385–395. Springer-Verlag, 1991.

- [2] G. S. Ammar and W. B. Gragg. $o(n^2)$ reduction algorithms for the construction of a band matrix from spectral data. *SIAM Journal on Matrix Analysis and its Applications*, 12(3):426–432, 1991.
- [3] D. Boley and G.H. Golub. A survey of matrix inverse eigenvalue problems. *Inverse problems*, 3:595–622, 1987.
- [4] A. Bultheel and B. De Moor. Rational approximation in linear systems and control. *J. Comput. Appl. Math.*, 121:355–378, 2000.
- [5] A. Bultheel, P. González-Vera, E. Hendriksen, and O. Njåstad. *Orthogonal rational functions*, volume 5 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 1999.
- [6] A. Bultheel and M. Van Barel. Vector orthogonal polynomials and least squares approximation. *SIAM Journal on Matrix Analysis and its Applications*, 16(3):863–885, 1995.
- [7] S. Elhay, G. H. Golub, and J. Kautsky. Updating and downdating of orthogonal polynomials with data fitting applications. *SIAM Journal on Matrix Analysis and its Applications*, 12(2):327–353, 1991.
- [8] D. Fasino and L. Gemignani. A Lanczos type algorithm for the QR -factorization of regular Cauchy matrices. *Numerical Linear Algebra with Applications*, 9:305–319, 2002.
- [9] W. B. Gragg and W. J. Harrod. The numerically stable reconstruction of Jacobi matrices from spectral data. *Numerische Mathematik*, 44:317–335, 1984.
- [10] T. Kailath and A.H. Sayed, editors. *Fast reliable algorithms for matrices with structure*. SIAM, PA, 1999.
- [11] B. Ninnes and F. Gustafsson. A unifying construction of orthonormal bases for system identification. *IEEE Trans. Automat. Control*, 42:515–521, 1997.
- [12] L. Reichel. Fast QR decomposition of Vandermonde-like matrices and polynomial least squares approximation. *SIAM Journal on Matrix Analysis and its Applications*, 12:552–564, 1991.
- [13] L. Reichel, G.S. Ammar, and W.B. Gragg. Discrete least squares approximation by trigonometric polynomials. *Math. Comp.*, 57:273–289, 1991.
- [14] M. Van Barel and A. Bultheel. A new approach to the rational interpolation problem: the vector case. *Journal of Computational and Applied Mathematics*, 33(3):331–346, 1990.
- [15] M. Van Barel and A. Bultheel. A parallel algorithm for discrete least squares rational approximation. *Numerische Mathematik*, 63:99–121, 1992.

- [16] M. Van Barel and A. Bultheel. Discrete linearized least squares approximation on the unit circle. *Journal of Computational and Applied Mathematics*, 50:545–563, 1994.
- [17] M. Van Barel and A. Bultheel. Orthonormal polynomial vectors and least squares approximation for a discrete inner product. *Electronic Transactions on Numerical Analysis*, 3:1–23, March 1995.
- [18] M. Van Barel, D. Fasino, L. Gemignani, and N. Mastronardi. Orthogonal rational functions and structured matrices. Report TW 350, Department of Computer Science, K.U.Leuven, Belgium, November 2002.
- [19] R. Vandebril, M. Van Barel, and N. Mastronardi. A note on the representation and definition of semiseparable matrices. Report TW 368, Department of Computer Science, K.U.Leuven, Leuven, Belgium, October 2003.