Computation of matrix-valued formally orthogonal polynomials and applications ⋆

Roland W. Freund

Bell Laboratories, Lucent Technologies, Room 2C–525, 700 Mountain Avenue, Murray Hill, New Jersey 07974–0636, United States

Abstract

We present a computational procedure for generating formally orthogonal polynomials associated with a given bilinear Hankel form with rectangular matrix-valued moments. Our approach covers the most general case of moments of any size and is not restricted to square moments. Moreover, our algorithm has a built-in deflation procedure to handle linearly dependent or almost linearly dependent columns and rows of the block Hankel matrix associated with the bilinear form. Possible singular or close-to-singular leading principal submatrices of the deflated block Hankel matrix are avoided by means of look-ahead techniques. Applications of the computational procedure to eigenvalue computations, reduced-order modeling, the solution of multiple linear systems, and the fast solution of block Hankel systems are also briefly described.

Key words: Bilinear form, vector-valued polynomials, matrix-valued moments, block Hankel matrix, deflation, look-ahead, realization, block Krylov matrix, Lanczos-type algorithm, matrix-Padé approximation, fast block Hankel solver

1 Introduction

It has been known for a long time that many of the algebraic properties of scalar orthogonal polynomials on the real line carry over to the more general case of formally orthogonal polynomials induced by a given sequence of scalar moments; see, e.g., [3,4,8,15–17] and the references given there. For example, such formally orthogonal polynomials still satisfy three-term recurrences, as long as the scalar Hankel matrix $\mathbf{H}$ associated with the moment sequence

⋆ Numerical Analysis Manuscript No. 00–3–01, Bell Laboratories, Murray Hill, NJ, January 2000. Available online from `http://cm.bell-labs.com/cs/doc/00`.

is *strongly regular*, i.e., all leading principal submatrices of $\mathbf{H}$ are nonsingular. If $\mathbf{H}$ has some singular or in some sense nearly singular leading principal submatrices, then so-called look-ahead techniques can be used to jump over these submatrices, resulting in recurrence relations that connect the formally orthogonal polynomials corresponding to three consecutive look-ahead steps. In particular, these recurrences reduce to the standard three-term recurrences whenever three consecutive leading principal submatrices of $\mathbf{H}$ are nonsingular.

Scalar formally orthogonal polynomials are intimately connected with a number of algorithms for matrix computations. For example, the classical Lanczos process [19] for nonsymmetric matrices, fast solvers for linear systems with Hankel structure [14], and the computation of Padé approximants of transfer functions of single-input single-output linear dynamical systems [6,10] are all closely related to formally orthogonal polynomials. Furthermore, the theory of formally orthogonal polynomials has proven to be useful for developing more robust versions of these algorithms. For instance, the look-ahead variants [12,17,22] of the Lanczos process, which remedy possible breakdowns in the classical algorithm, are direct translations of the extended recurrences for formally orthogonal polynomials in the general case of Hankel matrices $\mathbf{H}$ with singular or nearly singular leading principal submatrices.

The concept of formally orthogonal polynomials can be extended to the case of arbitrary, in general rectangular, matrix-valued moments. However, except for special cases such as square orthogonal matrix polynomials on the line [25], the theory of the associated matrix-valued polynomials is a lot less developed than in the case of scalar moments. For example, a suitable extension of the classical Lanczos process, which is only applicable to single right and left starting vectors, to multiple, say m right and p left, starting vectors is intimately related to formally orthogonal polynomials associated with sequences of $p \times m$-matrix-valued moments that are given by a so-called realization. Such a Lanczos-type method for multiple starting vectors was developed only recently [1,7], motivated mainly by the need for such an algorithm for the computation of matrix-Padé approximants of transfer functions of m-input p-output linear dynamical systems [7,9,10].

There are two intrinsic difficulties that arise in the case of $p \times m$-matrix-valued moments, but not in the case of scalar moments. First, in the important case of matrix moments given by a realization, the block Hankel matrix $\mathbf{H}$ associated with these moments necessarily exhibits systematic singularities or ill-conditioning due to linearly dependent or nearly linearly dependent columns and rows. These linear or nearly linear dependencies imply that from a certain point on all leading principal submatrices of $\mathbf{H}$ are singular or nearly singular, although the moment information contained in $\mathbf{H}$ has not been fully exhausted yet. In particular, block Hankel matrices induced by a realization are not strongly regular, and singular or nearly singular submatrices caused by linearly dependent or nearly linearly dependent columns and rows cannot be avoided by means of look-ahead techniques. Instead, so-called deflation is needed in order to remove systematic singularities or ill-conditioning due to linearly dependent or nearly linearly dependent columns and rows of $\mathbf{H}$. Second, the fact that $m \neq p$ in general excludes the possibility of constructing the formally orthogonal polynomials directly as right $m \times m$-matrix-valued and left $p \times p$-matrix-valued polynomials. Moreover, each deflation of a column of $\mathbf{H}$ effectively reduces m by one and each deflation of a row of $\mathbf{H}$ effectively reduces p by one. Since deflations of columns and rows occur independently in general, this means that the "current" values of m and p in the course of the construction of formally orthogonal polynomials will be different in general, even if $m = p$ initially. The difficulties due to different m and p can be avoided by constructing the polynomials associated with $p \times m$-matrix-valued moments vector-wise, rather than matrix-wise.

In this paper, we present a computational procedure for generating right and left formally orthogonal polynomials associated with a given bilinear form induced by a sequence of general rectangular $p \times m$-matrix-valued moments. Our approach covers the most general case of arbitrary integers $m, p \geq 1$, and we need not assume that the block Hankel matrix $\mathbf{H}$ associated with the given bilinear form is strongly regular. Our algorithm has a built-in deflation procedure to handle linearly dependent or almost linearly dependent columns and rows of $\mathbf{H}$. Possible singular or close-to-singular leading principal submatrices of the deflated block Hankel matrix are avoided by means of look-ahead techniques. Applications of the computational procedure to eigenvalue computations, reduced-order modeling, the solution of multiple linear systems, and the fast solution of block Hankel systems are also briefly described.

We remark that our approach of constructing formally orthogonal polynomials induced by matrix moments in a vector-wise fashion is related to earlier work, such as [2,24,27]. However, in these papers, the assumption that $\mathbf{H}$ is strongly regular is made, and this excludes the intrinsic difficulties described above.

The remainder of this paper is organized as follows. In Section 2, we introduce some notation. Section 3 describes our general setting of bilinear Hankel forms and discusses the need for deflation and look-ahead. In Section 4, we present our notion of formally orthogonal polynomials associated with a given bilinear Hankel form. In Section 5, we explain the structure of the recurrence relations used in our construction of formally orthogonal polynomials. A complete statement of our algorithm for generating formally orthogonal polynomials is given in Section 6, and some properties of this algorithm are stated in Section 7. Applications of the algorithm are sketched in Section 8. Finally, in Section 9, we make some concluding remarks.

2 Notation and some preliminaries

In this section, we introduce some notation used throughout this paper.

2.1 Notation

All vectors and matrices are allowed to have real or complex entries. We use boldface letters to denote vectors and matrices. As usual, $\overline{\mathbf{M}} = [\, \overline{m_{jk}} \,]$, $\mathbf{M}^{\mathrm{T}} = [\, m_{kj} \,]$, and $\mathbf{M}^{\mathrm{H}} = \overline{\mathbf{M}}^{\mathrm{T}} = [\, \overline{m_{kj}} \,]$ denote the complex conjugate, transpose, and the conjugate transpose, respectively, of the matrix $\mathbf{M} = [\, m_{jk} \,]$. We use the notation $[\, x_j \,]_{j \in \mathcal{J}}$ for the subvector of $\mathbf{x} = [\, x_j \,]$ induced by the index set $\mathcal{J}$, and analogously, $[\, m_{jk} \,]_{j \in \mathcal{J},\, k \in \mathcal{K}}$ for the submatrix of $\mathbf{M} = [\, m_{jk} \,]$ induced by the row indices $\mathcal{J}$ and column indices $\mathcal{K}$. The vector norm $\|\mathbf{x}\| := \sqrt{\mathbf{x}^{\mathrm{H}}\mathbf{x}}$ is always the Euclidean norm, and $\|\mathbf{M}\| := \max_{\|\mathbf{x}\|=1} \|\mathbf{M}\,\mathbf{x}\|$ is the corresponding induced matrix norm.

The i-th unit vector of dimension j is denoted by $\mathbf{e}_i^{(j)}$. We use $\mathbf{I}_n$ to denote the $n \times n$ identity matrix, and we will simply write $\mathbf{I}$ if the actual dimension is apparent from the context.

The sets of real and complex numbers are denoted by $\mathbb{R}$ and $\mathbb{C}$, respectively. We use the symbols $\mathbb{N}$ for the set of positive integers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$ for the set of non-negative integers.

We denote by $\mathcal{P}^{(j)}$ the set of all vector-valued polynomials

$$\boldsymbol{\phi}(\lambda) \equiv \mathbf{a}_0 + \mathbf{a}_1\,\lambda + \cdots + \mathbf{a}_i\,\lambda^i, \quad \text{where} \quad \mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_i \in \mathbb{C}^j,\ i \in \mathbb{N}_0, \tag{1}$$

with coefficient vectors of dimension j, and by

$$\mathcal{P}^{(j \times k)} := \left\{ \boldsymbol{\Phi} = [\, \boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \cdots \quad \boldsymbol{\phi}_k \,] \;\middle|\; \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_k \in \mathcal{P}^{(j)} \right\}$$

the set of all matrix-valued polynomials with coefficient matrices of size $j \times k$.

We use the symbol 0 both for the number zero and for the scalar zero polynomial, and similarly, the symbol $\mathbf{0}$ for the $m \times n$ zero matrix and the zero polynomials in $\mathcal{P}^{(j)}$ and $\mathcal{P}^{(j \times k)}$. The actual dimension of $\mathbf{0}$ will always be apparent from the context.

2.2 The degree of a vector polynomial

Following [2,5,27], we can associate with any given vector polynomial (1), $\boldsymbol{\phi}$, the scalar polynomial

$$\varphi(\lambda) \equiv [\, 1 \quad \lambda \quad \cdots \quad \lambda^{j-1} \,] \cdot \boldsymbol{\phi}(\lambda^j) \equiv \sum_{k=0}^{(i+1)j-1} \alpha_k \, \lambda^k. \tag{2}$$

Here, the α_k's are just the coefficients of the *stacked* coefficient vector

$$\mathbf{a} = \begin{bmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_i \end{bmatrix} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{(i+1)j-1} \end{bmatrix} \in \mathbb{C}^{(i+1)j} \tag{3}$$

of (1). The (diagonal) *degree*, $\deg \boldsymbol{\phi}$, of $\boldsymbol{\phi}$ is then defined as the degree of the scalar polynomial (2), i.e.,

$$\deg \boldsymbol{\phi} := \begin{cases} \max\{\, k \mid \alpha_k \neq 0 \text{ in } (2) \,\} & \text{if } \boldsymbol{\phi} \neq \mathbf{0}, \\ -\infty & \text{if } \boldsymbol{\phi} = \mathbf{0}. \end{cases}$$

In the sequel, we will also use the notation

$$\operatorname{vec} \boldsymbol{\phi} := \begin{cases} [\, \alpha_k \,]_{0 \le k \le \deg \boldsymbol{\phi}} & \text{if } \boldsymbol{\phi} \neq \mathbf{0}, \\ 0 & \text{if } \boldsymbol{\phi} = \mathbf{0}, \end{cases}$$

for the vector that results from (3) by deleting any trailing zeros.

3 Bilinear Hankel forms and block Hankel matrices

In this section, we describe our general setting of bilinear Hankel forms and their associated infinite block Hankel matrices. We also discuss the need for deflation and look-ahead to avoid possible singular or ill-conditioned submatrices of the block Hankel matrices.

3.1 Bilinear Hankel forms

Let m, $p \geq 1$ be given integers. A complex-valued functional

$$\langle \cdot, \cdot \rangle : \ \mathcal{P}^{(p)} \times \mathcal{P}^{(m)} \mapsto \mathbb{C} \tag{4}$$

is called a *bilinear form* if

$$\begin{aligned} \langle \boldsymbol{\psi}, \sigma_1 \, \boldsymbol{\phi}_1 + \sigma_2 \, \boldsymbol{\phi}_2 \rangle &= \sigma_1 \, \langle \boldsymbol{\psi}, \boldsymbol{\phi}_1 \rangle + \sigma_2 \, \langle \boldsymbol{\psi}, \boldsymbol{\phi}_2 \rangle \\ \text{and} \quad \langle \sigma_1 \, \boldsymbol{\psi}_1 + \sigma_2 \, \boldsymbol{\psi}_2, \boldsymbol{\phi} \rangle &= \sigma_1 \, \langle \boldsymbol{\psi}_1, \boldsymbol{\phi} \rangle + \sigma_2 \, \langle \boldsymbol{\psi}_2, \boldsymbol{\phi} \rangle \end{aligned} \tag{5}$$

for all $\boldsymbol{\phi}$, $\boldsymbol{\phi}_1$, $\boldsymbol{\phi}_2 \in \mathcal{P}^{(m)}$, $\boldsymbol{\psi}$, $\boldsymbol{\psi}_1$, $\boldsymbol{\psi}_2 \in \mathcal{P}^{(p)}$, σ_1, $\sigma_2 \in \mathbb{C}$. We say that (4) is a *Hankel form* if the following shift property is satisfied:

$$\langle \boldsymbol{\psi}, \lambda \, \boldsymbol{\phi} \rangle = \langle \lambda \, \boldsymbol{\psi}, \boldsymbol{\phi} \rangle \quad \text{for all} \quad \boldsymbol{\phi} \in \mathcal{P}^{(m)}, \ \boldsymbol{\psi} \in \mathcal{P}^{(p)}. \tag{6}$$

In the sequel, it is always assumed that (4) is a given bilinear Hankel form. Furthermore, we extend (4) to matrix-valued polynomials by setting

$$\langle \boldsymbol{\Psi}, \boldsymbol{\Phi} \rangle := \begin{bmatrix} \langle \boldsymbol{\psi}_1, \boldsymbol{\phi}_1 \rangle & \langle \boldsymbol{\psi}_1, \boldsymbol{\phi}_2 \rangle & \cdots & \langle \boldsymbol{\psi}_1, \boldsymbol{\phi}_k \rangle \\ \langle \boldsymbol{\psi}_2, \boldsymbol{\phi}_1 \rangle & \langle \boldsymbol{\psi}_2, \boldsymbol{\phi}_2 \rangle & \cdots & \langle \boldsymbol{\psi}_2, \boldsymbol{\phi}_k \rangle \\ \vdots & \vdots & & \vdots \\ \langle \boldsymbol{\psi}_j, \boldsymbol{\phi}_1 \rangle & \langle \boldsymbol{\psi}_j, \boldsymbol{\phi}_2 \rangle & \cdots & \langle \boldsymbol{\psi}_j, \boldsymbol{\phi}_k \rangle \end{bmatrix} \in \mathbb{C}^{j \times k} \tag{7}$$

for all

$$\boldsymbol{\Phi} = [\, \boldsymbol{\phi}_1 \quad \boldsymbol{\phi}_2 \quad \cdots \quad \boldsymbol{\phi}_k \,] \in \mathcal{P}^{(m \times k)}, \quad \boldsymbol{\Psi} = [\, \boldsymbol{\psi}_1 \quad \boldsymbol{\psi}_2 \quad \cdots \quad \boldsymbol{\psi}_j \,] \in \mathcal{P}^{(p \times j)}.$$

In particular, using the notation (7), we define $p \times m$ (matrix-valued) *moments*

$$\mathbf{M}_{j,k} := \langle \mathbf{I}_p \, \lambda^j, \mathbf{I}_m \, \lambda^k \rangle \in \mathbb{C}^{p \times m} \quad \text{for all} \quad j, \, k \in \mathbb{N}_0 \,. \tag{8}$$

In view of the bilinearity (5), any bilinear form (4) is completely determined by its moments (8). Furthermore, the shift property (6) means that the moments $\mathbf{M}_{j,k}$ only depend on $j+k$, and we set $\mathbf{M}_{j+k} := \mathbf{M}_{j,k}$. Therefore, any bilinear Hankel form (4) is completely determined by the sequence of moments

$$\mathbf{M}_i := \langle \mathbf{I}_p, \mathbf{I}_m \, \lambda^i \rangle = \langle \mathbf{I}_p \, \lambda^i, \mathbf{I}_m \rangle, \quad i \in \mathbb{N}_0 \,.$$

The associated infinite block Hankel matrix

$$\mathbf{H} := [\,\mathbf{M}_{j+k}\,]_{j,k\geq 0} = \begin{bmatrix} \mathbf{M}_0 & \mathbf{M}_1 & \mathbf{M}_2 & \cdots \\ \mathbf{M}_1 & \mathbf{M}_2 & & \\ \mathbf{M}_2 & & & \\ \vdots & & & \end{bmatrix} \tag{9}$$

is called the *moment matrix* of the bilinear Hankel form (4). Although $\mathbf{H}$ has a block Hankel structure, we will also consider $\mathbf{H}$ as a scalar matrix with entries $h_{\nu,\mu}$, i.e.,

$$\mathbf{H} = [\,h_{\nu,\mu}\,]_{\nu,\mu\geq 0}\,, \quad \text{where} \quad h_{\nu,\mu} \in \mathbb{C} \quad \text{for all} \quad \nu,\ \mu \in \mathbb{N}_0\,.$$

Furthermore, for each $n,\ k \in \mathbb{N}_0$, we set

$$\mathbf{H}_{n,k} := [\,h_{\nu,\mu}\,]_{0\leq\nu\leq n,\,0\leq\mu\leq k} \quad \text{and} \quad \mathbf{H}_n := \mathbf{H}_{n,n} = [\,h_{\nu,\mu}\,]_{0\leq\nu,\mu\leq n}\,.$$

Note that $\mathbf{H}_n$ is the n-th scalar leading principal submatrix of $\mathbf{H}$.

With the notation just introduced, for any pair of vector-valued polynomials $\boldsymbol{\phi} \in \mathcal{P}^{(m)}$ and $\boldsymbol{\psi} \in \mathcal{P}^{(p)}$, we have

$$\langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle := \mathbf{b}^{\mathrm{T}}\,\mathbf{H}_{\deg \boldsymbol{\psi},\deg \boldsymbol{\phi}}\,\mathbf{a}, \quad \text{where} \quad \mathbf{a} = \operatorname{vec} \boldsymbol{\phi}, \quad \mathbf{b} = \operatorname{vec} \boldsymbol{\psi}. \tag{10}$$

3.2 Bilinear Hankel forms associated with realizations

An important special case is bilinear Hankel forms that are associated with so-called realizations of time-invariant linear dynamical systems. We refer the reader to [18, Chapter 10.11], [23], or [26, Chapter 5.5] for a discussion of the concepts and results from realization theory that we will use in this subsection.

Let

$$\mathbf{M}_i \in \mathbb{C}^{p\times m}, \quad i = 0, 1, \ldots, \tag{11}$$

be a given sequence of moments. A triple of matrices

$$\mathbf{A} \in \mathbb{C}^{N\times N}, \quad \mathbf{R} \in \mathbb{C}^{N\times m}, \quad \mathbf{L} \in \mathbb{C}^{N\times p} \tag{12}$$

is called a *realization* of the sequence (11) if

$$\mathbf{M}_i = \mathbf{L}^{\mathrm{T}}\,\mathbf{A}^i\,\mathbf{R} \quad \text{for all} \quad i \in \mathbb{N}_0.$$

The integer N is called the *dimension* of the realization (12). A realization (12) of a given sequence (11) is said to be *minimal* if its dimension N is as small as possible.

Not every given sequence (11) has a realization. The following well-known result (see, e.g., [26, Theorem 21]) gives a necessary and sufficient condition for the existence of a realization in terms of the infinite block Hankel matrix (9) with block entries (11).

Theorem A *A sequence* (11) *admits a realization if, and only if, the associated infinite block Hankel matrix* (9), $\mathbf{H}$, *has finite rank. Furthermore, if* (11) *has a realization, then* $N = \operatorname{rank} \mathbf{H}$ *is the dimension of a minimal realization.*

For the remainder of this subsection, we now assume that $\mathbf{H}$ has finite rank and that (12) is a given, not necessarily minimal, realization of the block entries (11) of $\mathbf{H}$.

Note that, in view of (9) and (12), the block Hankel matrix $\mathbf{H}$ can be factored into block Krylov matrices as follows:

$$\mathbf{H} = \begin{bmatrix} \mathbf{L}^{\mathrm{T}} \\ \mathbf{L}^{\mathrm{T}}\,\mathbf{A} \\ \mathbf{L}^{\mathrm{T}}\,\mathbf{A}^2 \\ \vdots \end{bmatrix} \cdot \begin{bmatrix} \mathbf{R} & \mathbf{A}\,\mathbf{R} & \mathbf{A}^2\,\mathbf{R} & \cdots \end{bmatrix}. \tag{13}$$

As a first application of (13), we have the following connection of vector-valued polynomials with vectors in $\mathbb{C}^N$.

Remark 1 *Let* $\boldsymbol{\phi} \in \mathcal{P}^{(m)}$ *and* $\boldsymbol{\psi} \in \mathcal{P}^{(p)}$ *be any pair of polynomials. Then, using the representations*

$$\begin{aligned} \boldsymbol{\phi}(\lambda) &\equiv \mathbf{a}_0 + \mathbf{a}_1\,\lambda + \cdots + \mathbf{a}_j\,\lambda^j, \\ \boldsymbol{\psi}(\lambda) &\equiv \mathbf{b}_0 + \mathbf{b}_1\,\lambda + \cdots + \mathbf{b}_k\,\lambda^k, \end{aligned}$$

we can associate with $\boldsymbol{\phi}$ *and* $\boldsymbol{\psi}$ *the pair of vectors*

$$\begin{aligned} \mathbf{v} = \boldsymbol{\phi}(\mathbf{A}) \circ \mathbf{R} &:= \mathbf{R}\,\mathbf{a}_0 + \mathbf{A}\,\mathbf{R}\,\mathbf{a}_1 + \cdots + \mathbf{A}^j\,\mathbf{R}\,\mathbf{a}_j \in \mathbb{C}^N, \\ \mathbf{w} = \boldsymbol{\psi}(\mathbf{A}^{\mathrm{T}}) \circ \mathbf{L} &:= \mathbf{L}\,\mathbf{b}_0 + \mathbf{A}^{\mathrm{T}}\,\mathbf{L}\,\mathbf{b}_1 + \cdots + (\mathbf{A}^{\mathrm{T}})^k\,\mathbf{L}\,\mathbf{b}_k \in \mathbb{C}^N. \end{aligned} \tag{14}$$

By (10), (13), *and* (14), *it follows that* $\langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle = \mathbf{w}^{\mathrm{T}}\,\mathbf{v}$.

The factorization (13) of $\mathbf{H}$ necessarily implies that from a certain n on, all leading principal submatrices $\mathbf{H}_n$ are singular. Indeed, consider the *right block Krylov matrix*

$$[\,\mathbf{R} \quad \mathbf{A}\,\mathbf{R} \quad \mathbf{A}^2\,\mathbf{R} \quad \cdots\,] \tag{15}$$

in (13). The columns of the matrix (15) are vectors in $\mathbb{C}^N$, and hence at most N of them are linearly independent. By scanning the columns of (15) from left to right and deleting each column that is linearly dependent on earlier columns, we obtain the *deflated* right block Krylov matrix

$$[\,\mathbf{R}_0 \quad \mathbf{A}\,\mathbf{R}_1 \quad \mathbf{A}^2\,\mathbf{R}_2 \quad \cdots \quad \mathbf{A}^{j_{\max}-1}\,\mathbf{R}_{j_{\max}}\,]\,. \tag{16}$$

By the structure of (15), a column $\mathbf{A}^{j-1}\,\mathbf{r}$ being linearly dependent on earlier columns implies that all columns $\mathbf{A}^i\mathbf{r}$, $i \geq j$, are also linearly dependent on earlier columns. This implies that, for each $j = 0, 1, \ldots, j_{\max}$, $\mathbf{R}_j$ is a submatrix of $\mathbf{R}_{j-1}$, where, for $j = 0$, we set $\mathbf{R}_{-1} := \mathbf{R}$. Similarly, by scanning the columns of the *left block Krylov matrix*

$$[\,\mathbf{L} \quad \mathbf{A}^{\mathrm{T}}\,\mathbf{L} \quad (\mathbf{A}^{\mathrm{T}})^2\,\mathbf{L} \quad \cdots\,]$$

from left to right and deleting each column that is linearly dependent on earlier columns, we obtain the *deflated* left block Krylov matrix

$$[\,\mathbf{L}_0 \quad \mathbf{A}^{\mathrm{T}}\,\mathbf{L}_1 \quad (\mathbf{A}^{\mathrm{T}})^2\,\mathbf{L}_2 \quad \cdots \quad (\mathbf{A}^{\mathrm{T}})^{k_{\max}-1}\,\mathbf{L}_{k_{\max}}\,]\,. \tag{17}$$

Here, for each $k = 0, 1, \ldots, k_{\max}$, $\mathbf{L}_k$ is a submatrix of $\mathbf{L}_{k-1}$, where, for $k = 0$, we set $\mathbf{L}_{-1} := \mathbf{L}$. Now let j_0 be the smallest integer such that $\mathbf{R}_{j_0} \neq \mathbf{R}$ and let k_0 be the smallest integer such that $\mathbf{L}_{k_0} \neq \mathbf{L}$. Then, by construction, it follows that the n-th leading principal submatrix

$$\mathbf{H}_n \quad \text{is singular for all} \quad n \geq \min\{\,(j_0+1)m, (k_0+1)p\,\} - 1.$$

Finally, we note that this systematic singularity of the submatrices $\mathbf{H}_n$ can be avoided by replacing the moment matrix $\mathbf{H}$ by the *deflated moment matrix*

$$\mathbf{H}^{\mathrm{defl}} := \begin{bmatrix} \mathbf{L}_0^{\mathrm{T}} \\ \mathbf{L}_1^{\mathrm{T}}\,\mathbf{A} \\ \vdots \\ \mathbf{L}_{k_{\max}}^{\mathrm{T}}\,\mathbf{A}^{k_{\max}-1} \end{bmatrix} \cdot [\,\mathbf{R}_0 \quad \mathbf{A}\,\mathbf{R}_1 \quad \cdots \quad \mathbf{A}^{j_{\max}-1}\,\mathbf{R}_{j_{\max}}\,]\,.$$

3.3 The need for deflation

We now return to general moment matrices $\mathbf{H}$, and we extend the procedure for avoiding systematic singularities of submatrices $\mathbf{H}_n$ to this general case.

Generalizing the procedure described in Section 3.2, we scan the columns, $[\, h_{i,\mu} \,]_{i\geq 0}$, $\mu = 0, 1, \ldots$, of $\mathbf{H}$, from left to right and delete each column that is linearly dependent, or in some sense "almost" linearly dependent, on earlier columns. Similarly, we scan the rows, $[\, h_{\nu,i} \,]_{i\geq 0}$, $\nu = 0, 1, \ldots$, of $\mathbf{H}$, from top to bottom and delete each row that is linearly dependent, or in some sense "almost" linearly dependent, on earlier rows. We refer to this process of deleting linearly dependent and "almost" linearly dependent columns and rows as *deflation*. Moreover, we say that *exact deflation* is performed when only the linearly dependent columns and rows are removed. Clearly, exact deflation is only possible in exact arithmetic, and deflation of "almost" linearly dependent columns and rows has to be included when actual computations are done in finite-precision arithmetic.

In the sequel, we denote by

$$0 \leq \mu_0 < \mu_1 < \cdots < \mu_n < \cdots \quad \text{and} \quad 0 \leq \nu_0 < \nu_1 < \cdots < \nu_n < \cdots \tag{18}$$

the sequences of indices of those columns and rows of $\mathbf{H}$, respectively, that are left after deflation has been performed. Moreover, we denote by

$$\mathcal{M} = \{\mu_n\}_{n=0}^{n_{\mathrm{r}}} \quad \text{and} \quad \mathcal{N} = \{\nu_n\}_{n=0}^{n_{\mathrm{l}}} \tag{19}$$

the sets of all the indices (18), and by

$$\mathbf{H}^{\mathrm{defl}} := [\, h_{\nu,\mu} \,]_{\nu\in\mathcal{N},\, \mu\in\mathcal{M}} \tag{20}$$

the corresponding *deflated* moment matrix. Note that, in general, each of the two sequences (18) may be finite of infinite, i.e., n_{r}, $n_{\mathrm{l}} \in \mathbb{N}_0 \cup \{\infty\}$. Hence $\mathbf{H}^{\mathrm{defl}}$ can have finitely or infinitely many rows or columns. However, in the special case of Hankel matrices $\mathbf{H}$ of finite rank discussed in Section 3.2, n_{r}, $n_{\mathrm{l}} < \infty$ and thus $\mathbf{H}^{\mathrm{defl}}$ is a finite matrix.

In view of the block Hankel structure (9) of $\mathbf{H}$, an exact deflation of a μ-th column of $\mathbf{H}$ implies that also all $(\mu + jm)$-th columns, where $j = 1, 2, \ldots$, need to be deflated. Similarly, an exact deflation of a ν-th row implies that also all $(\nu + jp)$-th rows, where $j = 1, 2, \ldots$, need to be deflated. We assume that the same rule is also applied in the case of general deflation, and so whenever a μ-th column or ν-th row is deflated, we also deflate all $(\mu + jm)$-th columns,

$j \in \mathbb{N}$, respectively all $(\nu + jp)$-th rows, $j \in \mathbb{N}$. This implies that the sets (19) always satisfy the following conditions:

$$\mu \notin \mathcal{M} \quad \Longrightarrow \quad \mu + jm \notin \mathcal{M} \quad \text{for all} \quad j \in \mathbb{N}_0,$$
$$\nu \notin \mathcal{N} \quad \Longrightarrow \quad \nu + jp \notin \mathcal{N} \quad \text{for all} \quad j \in \mathbb{N}_0.$$

Finally, we note that, by (18), the mappings $n \mapsto \mu_n$ and $n \mapsto \nu_n$ are both invertible, and we will use μ^{-1} and ν^{-1} to denote the inverse mappings defined by

$$n = \mu^{-1}(\mu_n), \quad \mu_n \in \mathcal{M}, \quad \text{and} \quad n = \nu^{-1}(\nu_n), \quad \nu_n \in \mathcal{N},$$

respectively.

3.4 The need for look-ahead

By replacing $\mathbf{H}$ with the deflated moment matrix $\mathbf{H}^{\text{defl}}$, we have removed any rank deficiencies due to linearly dependent columns and rows. Next, we discuss potential singularities of the leading principal submatrices of $\mathbf{H}^{\text{defl}}$.

By (19) and (20), all the leading principal submatrices of $\mathbf{H}^{\text{defl}}$ are given by

$$\mathbf{H}^{\text{defl}}_{n-1} := [\, h_{\nu_i,\mu_j} \,]_{0 \le i,j < n} \quad \text{for all} \quad n \in \mathbb{N} \quad \text{with} \quad n \le n_{\max}, \tag{21}$$

where

$$n_{\max} := 1 + \min\{\, n_{\mathrm{r}},\, n_{\mathrm{l}} \,\}. \tag{22}$$

Note that either $n_{\max} \in \mathbb{N}_0$ or $n_{\max} = \infty$. In the special case of Hankel matrices $\mathbf{H}$ of finite rank discussed in Section 3.2, $n_{\max}$ is a finite integer.

By construction, the necessary singularities implied by linearly dependent columns or rows of $\mathbf{H}$ have been removed from the submatrices (21). Obviously, as in the case of scalar Hankel matrices (see, e.g., [14] and the references given there), this construction alone is not sufficient to always guarantee that

$$\det \mathbf{H}^{\text{defl}}_{n-1} \neq 0 \quad \text{for all} \quad n \in \mathbb{N} \quad \text{with} \quad n \le n_{\max}. \tag{23}$$

However, the situation that (23) is satisfied is the *generic case*.

In the *general case*, some of the submatrices (21) may be singular or in some sense "close" to singular, and we employ so-called *look-ahead* techniques [22,28] to avoid these submatrices. We use the indices

$$n_0 := 0 < n_1 < n_2 < \cdots < n_k < \cdots \tag{24}$$

to mark those submatrices $\mathbf{H}_{n_k-1}$ that remain after any singular or close-to-singular submatrix has been removed from (21). In particular, by construction, we have

$$\det \mathbf{H}^{\mathrm{defl}}_{n_k-1} \neq 0 \quad \text{for all} \quad k \in \mathbb{N} \quad \text{with} \quad n_k \leq n_{\max}.$$

Finally, note that in the generic case when no look-ahead is necessary, the indices (24) are simply given by

$$n_k = k \quad \text{for all} \quad k \in \mathbb{N}_0 \quad \text{with} \quad k \leq n_{\max}. \tag{25}$$

4 Formally orthogonal polynomials

In this section, we present our notion of formally orthogonal polynomials associated with a given bilinear Hankel form $\langle \cdot, \cdot \rangle \colon\ \mathcal{P}^{(p)} \times \mathcal{P}^{(m)} \mapsto \mathbb{C}$.

4.1 Two sequences of polynomials

Let $\mathcal{M} = \{\mu_n\}_{n=0}^{n_{\mathrm{r}}}$ and $\mathcal{N} = \{\nu_n\}_{n=0}^{n_{\mathrm{l}}}$ be the column and row indices introduced in (18) and (19). Of course, in practice, these indices are not given beforehand, and instead, they have to be determined within our computational procedure for constructing vector-valued orthogonal polynomials. In Section 4.4 below, we will show how this can be done, but for now, we assume that $\mathcal{M}$ and $\mathcal{N}$ are given.

Our computational procedure generates two sequences of *right* and *left* polynomials,

$$\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n, \ldots \in \mathcal{P}^{(m)} \quad \text{and} \quad \boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n, \ldots \in \mathcal{P}^{(p)}, \tag{26}$$

respectively. Here, $n \in \mathbb{N}_0$ and $n \leq n_{\max}$, where $n_{\max}$ is given by (22). Furthermore, the polynomials (26) are constructed such that their degrees are just the indices $\mathcal{M}$ and $\mathcal{N}$, i.e.,

$$\deg \boldsymbol{\phi}_n = \mu_n \quad \text{and} \quad \deg \boldsymbol{\psi}_n = \nu_n \quad \text{for all} \quad n, \tag{27}$$

and their coefficient vectors,

$$\mathbf{a}^{(n)} := \operatorname{vec} \boldsymbol{\phi}_n = \begin{bmatrix} \alpha_0^{(n)} \\ \alpha_1^{(n)} \\ \vdots \\ \alpha_{\mu_n}^{(n)} \end{bmatrix} \quad \text{and} \quad \mathbf{b}^{(n)} := \operatorname{vec} \boldsymbol{\psi}_n = \begin{bmatrix} \beta_0^{(n)} \\ \beta_1^{(n)} \\ \vdots \\ \beta_{\nu_n}^{(n)} \end{bmatrix}, \tag{28}$$

have no nonzero entries outside $\mathcal{M}$ and $\mathcal{N}$, i.e., for all n,

$$\alpha_\mu^{(n)} = 0 \quad \text{if} \quad \mu \neq \mathcal{M}, \quad \text{and} \quad \beta_\nu^{(n)} = 0 \quad \text{if} \quad \nu \neq \mathcal{N}. \tag{29}$$

4.2 Orthogonality in the generic case

The goal is to construct the polynomials (26) such that they are regular formally orthogonal polynomials in the sense of the following definition.

Definition 2 *The polynomial $\boldsymbol{\phi}_n \in \mathcal{P}^{(m)}$ is said to be an n-th right formally orthogonal polynomial (RFOP) if $\deg \boldsymbol{\phi}_n = \mu_n$ and*

$$\langle \boldsymbol{\psi}, \boldsymbol{\phi}_n \rangle = 0 \quad \textit{for all} \quad \boldsymbol{\psi} \in \mathcal{P}^{(p)} \quad \textit{with} \quad \deg \boldsymbol{\psi} = \nu_i, \quad i < n. \tag{30}$$

The polynomial $\boldsymbol{\psi}_n \in \mathcal{P}^{(p)}$ is said to be an n-th left formally orthogonal polynomial (LFOP) if $\deg \boldsymbol{\psi}_n = \nu_n$ and

$$\langle \boldsymbol{\psi}_n, \boldsymbol{\phi} \rangle = 0 \quad \textit{for all} \quad \boldsymbol{\phi} \in \mathcal{P}^{(m)} \quad \textit{with} \quad \deg \boldsymbol{\phi} = \mu_i, \quad i < n. \tag{31}$$

Moreover, the RFOP $\boldsymbol{\phi}_n$ and the LFOP $\boldsymbol{\psi}_n$ are said to be regular if they are uniquely determined by (30) and (31), respectively, up to a nonzero scalar factor.

Using (10), (21), and (27)–(29), one readily verifies that the condition (30) is equivalent to the system of linear equations,

$$\mathbf{H}_{n-1}^{\text{defl}} \begin{bmatrix} \alpha_{\mu_0}^{(n)} \\ \alpha_{\mu_1}^{(n)} \\ \vdots \\ \alpha_{\mu_{n-1}}^{(n)} \end{bmatrix} = -\alpha_{\mu_n}^{(n)} \begin{bmatrix} h_{\nu_0,\mu_n} \\ h_{\nu_1,\mu_n} \\ \vdots \\ h_{\nu_{n-1},\mu_n} \end{bmatrix}, \quad \alpha_{\mu_n}^{(n)} \neq 0, \tag{32}$$

for the potentially nonzero coefficients of $\boldsymbol{\phi}_n$. Similarly, (31) is equivalent to the system of linear equations,

$$\left(\mathbf{H}_{n-1}^{\mathrm{defl}}\right)^{\mathrm{T}} \begin{bmatrix} \beta_{\nu_0}^{(n)} \\ \beta_{\nu_1}^{(n)} \\ \vdots \\ \beta_{\nu_{n-1}}^{(n)} \end{bmatrix} = -\beta_{\nu_n}^{(n)} \begin{bmatrix} h_{\nu_n,\mu_0} \\ h_{\nu_n,\mu_1} \\ \vdots \\ h_{\nu_n,\mu_{n-1}} \end{bmatrix}, \quad \beta_{\nu_n}^{(n)} \neq 0, \tag{33}$$

for the potentially nonzero coefficients of $\boldsymbol{\psi}_n$. In view of (32) and (33), $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ can be constructed as a regular RFOP and a regular LFOP, respectively, if, and only if, $\mathbf{H}_{n-1}^{\mathrm{defl}}$ is nonsingular. We thus have the following result on the existence of regular RFOPs and LFOPs in the generic case (23).

Theorem 3 *The polynomials* (26) *can all be constructed as regular RFOPs and LFOPs only in the generic case, i.e., if the matrices* $\mathbf{H}_{n-1}^{\mathrm{defl}}$ *are nonsingular for all* $n \in \mathbb{N}$ *with* $n \leq n_{\max}$.

In the generic case, we thus construct the polynomials (26) such that

$$\langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_n \rangle = 0 \quad \text{for all} \quad i \neq n, \quad i, n \in \mathbb{N}_0, \quad i, n \leq n_{\max}. \tag{34}$$

Using the notation

$$\boldsymbol{\Phi}_n := [\, \boldsymbol{\phi}_0 \quad \boldsymbol{\phi}_1 \quad \cdots \quad \boldsymbol{\phi}_n \,] \quad \text{and} \quad \boldsymbol{\Psi}_n := [\, \boldsymbol{\psi}_0 \quad \boldsymbol{\psi}_1 \quad \cdots \quad \boldsymbol{\psi}_n \,] \tag{35}$$

for the matrix polynomials whose columns are the first $n+1$ right and left polynomials (26), respectively, the orthogonality condition (34) can be stated as follows:

$$\langle \boldsymbol{\Psi}_n, \boldsymbol{\Phi}_n \rangle = \boldsymbol{\Delta}_n \quad \text{for all} \quad n \in \mathbb{N}_0, \quad n \leq n_{\max}. \tag{36}$$

Here,

$$\boldsymbol{\Delta}_n := \operatorname{diag}(\delta_0, \delta_1, \ldots, \delta_n), \quad \text{where} \quad \delta_i := \langle \boldsymbol{\psi}_i, \boldsymbol{\phi}_i \rangle \quad \text{for all} \quad i. \tag{37}$$

Finally, let

$$\mathbf{A}_n := [\, \alpha_{\mu_i}^{(j)} \,]_{0 \leq i,j \leq n} \quad \text{and} \quad \mathbf{B}_n := [\, \beta_{\nu_i}^{(j)} \,]_{0 \leq i,j \leq n} \tag{38}$$

denote the matrices of the potentially nonzero coefficients of $\boldsymbol{\Phi}_n$ and $\boldsymbol{\Psi}_n$, respectively. Then, by (10), (21), (28), (29), and (38), the condition (36) is equivalent to the matrix factorization

$$\mathbf{B}_n^{\mathrm{T}} \, \mathbf{H}_n^{\mathrm{defl}} \, \mathbf{A}_n = \boldsymbol{\Delta}_n, \tag{39}$$

where $\mathbf{A}_n$ and $\mathbf{B}_n$ are nonsingular upper triangular matrices. Thus, it follows from (37) and (39) that the condition (23) for the generic case is equivalent to

$$\delta_i \neq 0 \quad \text{for all} \quad 0 \leq i < n_{\max}. \tag{40}$$

4.3 Orthogonality in the case of look-ahead

We now turn to the general case where look-ahead is used to avoid singular or close-to-singular submatrices $\mathbf{H}_{n-1}^{\mathrm{defl}}$. In view of Theorem 3, the polynomials $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ cannot be constructed as a regular RFOP and LFOP if $\mathbf{H}_{n-1}^{\mathrm{defl}}$ is exactly singular. If $\mathbf{H}_{n-1}^{\mathrm{defl}}$ is nonsingular, but in some sense close to singular, then building $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ as a regular RFOP and LFOP will result in numerical instabilities in general. Therefore, we only construct the polynomials $\boldsymbol{\phi}_{n_k}$ and $\boldsymbol{\psi}_{n_k}$ corresponding to the index sequence (24) as regular RFOPs and LFOPs, while the remaining polynomials satisfy only a relaxed version of the orthogonality condition (34).

More precisely, based on the indices (24), we partition the right and left polynomials (26) into *clusters*

$$\begin{aligned} \mathbf{\Phi}^{(k)} &:= [\, \boldsymbol{\phi}_{n_k} \quad \boldsymbol{\phi}_{n_k+1} \quad \cdots \quad \boldsymbol{\phi}_{n_{k+1}-1} \,] \\ \text{and} \quad \mathbf{\Psi}^{(k)} &:= [\, \boldsymbol{\psi}_{n_k} \quad \boldsymbol{\psi}_{n_k+1} \quad \cdots \quad \boldsymbol{\psi}_{n_{k+1}-1} \,], \end{aligned} \tag{41}$$

respectively. The polynomials (26) are then constructed such that we have the cluster-wise orthogonality

$$\langle \mathbf{\Psi}^{(j)}, \mathbf{\Phi}^{(k)} \rangle = \mathbf{0} \quad \text{for all} \quad j \neq k, \quad j,\, k \in \mathbb{N}_0, \quad n_j,\, n_k \leq n_{\max}. \tag{42}$$

We remark that, by (42), the leading polynomials $\boldsymbol{\phi}_{n_k}$ and $\boldsymbol{\psi}_{n_k}$ of each k-th cluster (41) are regular RFOPs and LFOPs, respectively. Next, we set

$$\mathbf{\Delta}^{(k)} := \langle \mathbf{\Psi}^{(k)}, \mathbf{\Phi}^{(k)} \rangle \quad \text{for all} \quad k.$$

Then, using essentially the same argument that lead to condition (40) in the generic case, it follows that, in the general case,

$$\mathbf{\Delta}^{(k)} \quad \text{is nonsingular for all} \quad k \in \mathbb{N}_0 \quad \text{with} \quad n_k < n_{\max}.$$

Finally, note that each n-th pair of polynomials $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ in (26) is part of exactly one pair of clusters (41), namely those with index $k = \gamma(n)$. Here and in the sequel, we use the notation

$$\gamma(n) := \max\{\, j \in \mathbb{N}_0 \mid n_j \le n \,\} \tag{43}$$

for the function that determines the cluster index for the n-th pair of polynomials. Recall from (25) that in the case of no look-ahead, $n_k = k$ for all k. Thus, in the generic case, (43) reduces to

$$\gamma(n) = n \quad \text{for all} \quad n. \tag{44}$$

4.4 How deflation is done

In practice, the indices (18) and (19), which describe deflations, are not given beforehand, and instead, they have to be determined as part of the algorithm for constructing the polynomials (26). In this subsection, we describe how this is done.

In the algorithm, we keep track of the *current* block sizes m_{c} and p_{c}. Initially, $m_{\mathrm{c}} = m$ and $p_{\mathrm{c}} = p$. Every time a deflation of a right polynomial is performed, we set $m_c = m_c - 1$, and every time a deflation of a left polynomial is performed, we set $p_c = p_c - 1$. Thus, at any stage of the algorithm, $m - m_{\mathrm{c}}$, respectively $p - p_{\mathrm{c}}$, is just the number of deflations of right, respectively left, polynomials that have occurred so far.

Now assume that we already have constructed the right and left polynomials (26) up to index $n-1$. In addition to these polynomials, our algorithm has built m_{c} right and p_{c} left auxiliary polynomials,

$$\hat{\boldsymbol{\phi}}_n, \hat{\boldsymbol{\phi}}_{n+1}, \cdots, \hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}-1} \quad \text{and} \quad \hat{\boldsymbol{\psi}}_n, \hat{\boldsymbol{\psi}}_{n+1}, \cdots, \hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}-1}, \tag{45}$$

that satisfy the following "partial" orthogonality conditions:

$$\begin{aligned} \langle \boldsymbol{\Psi}^{(k)}, \hat{\boldsymbol{\phi}}_i \rangle &= \mathbf{0} \quad \text{for all} \quad 0 \le k < \gamma(n), \quad n \le i < n + m_{\mathrm{c}}, \\ \langle \hat{\boldsymbol{\psi}}_i, \boldsymbol{\Phi}^{(k)} \rangle &= \mathbf{0} \quad \text{for all} \quad 0 \le k < \gamma(n), \quad n \le i < n + p_{\mathrm{c}}. \end{aligned} \tag{46}$$

The polynomials (45) are the candidates for the next m_{c} right, respectively p_{c} left, polynomials in (26). In particular, in view of (46), the polynomials $\hat{\boldsymbol{\phi}}_n$ and $\hat{\boldsymbol{\psi}}_n$ already satisfy the necessary orthogonality conditions of $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$. It remains to decide if $\hat{\boldsymbol{\phi}}_n$ or $\hat{\boldsymbol{\psi}}_n$ should be deflated. If $\hat{\boldsymbol{\phi}}_n$ is deflated, it is deleted from (45), the indices of the remaining right polynomials in (45) are reduced by one, and m_c is reduced by one. If $\hat{\psi}_n$ is deflated, one proceeds analogously.

The decision if $\hat{\phi}_n$ or $\hat{\psi}_n$ needs to be deflated is relatively simple for the special case of bilinear Hankel forms given in terms of a realization (12). Indeed, it is easy to see that an exact deflation of $\hat{\phi}_n$, respectively $\hat{\psi}_n$, needs to be performed if, and only if, the associated vectors (see (14)) satisfy

$$\hat{\boldsymbol{\phi}}_n(\mathbf{A}) \circ \mathbf{R} = \mathbf{0}, \quad \text{respectively} \quad \hat{\boldsymbol{\psi}}_n(\mathbf{A}^{\mathrm{T}}) \circ \mathbf{L} = \mathbf{0}.$$

In practice, one thus deflates $\hat{\phi}_n$, respectively $\hat{\psi}_n$, if

$$\left\| \hat{\boldsymbol{\phi}}_n(\mathbf{A}) \circ \mathbf{R} \right\| \leq \mathtt{dtol}_n^{\mathtt{r}}, \quad \text{respectively} \quad \left\| \hat{\boldsymbol{\psi}}_n(\mathbf{A}^{\mathrm{T}}) \circ \mathbf{L} \right\| \leq \mathtt{dtol}_n^{\mathtt{l}}, \tag{47}$$

where $\mathtt{dtol}_n^{\mathtt{r}} \geq 0$ and $\mathtt{dtol}_n^{\mathtt{l}} \geq 0$ are suitably chosen small deflation tolerances. Note that the check (47) reduces to exact deflation only if $\mathtt{dtol}_n^{\mathtt{r}} = \mathtt{dtol}_n^{\mathtt{l}} = 0$.

In the general case of Hankel matrices $\mathbf{H}$ of not necessarily finite rank, one can show that exact deflation of $\hat{\phi}_n$, respectively $\hat{\psi}_n$, needs to be performed if, and only if,

$$\langle \mathbf{I}_p\, \lambda^k, \hat{\boldsymbol{\phi}}_n \rangle = \mathbf{0}, \quad \text{respectively} \quad \langle \hat{\boldsymbol{\psi}}_n, \mathbf{I}_m\, \lambda^k \rangle = \mathbf{0}, \quad \text{for all} \quad k \in \mathbb{N}_0. \tag{48}$$

Since (48) represents infinitely many conditions (representing the fact that $\mathbf{H}$ is an infinite matrix), in practice, one needs to replace (48) by some appropriate finite version. For example, imitating (47), one can check if, for all $0 \leq k \leq k(n)$,

$$\left\| \langle \mathbf{I}_p\, \lambda^k, \hat{\boldsymbol{\phi}}_n \rangle \right\| \leq \mathtt{dtol}_n^{\mathtt{r}}, \quad \text{respectively} \quad \left\| \langle \hat{\boldsymbol{\psi}}_n, \mathbf{I}_m\, \lambda^k \rangle \right\| \leq \mathtt{dtol}_n^{\mathtt{l}}, \tag{49}$$

where $k(n)$ is a sufficiently large, but finite integer.

We conclude this section with some comments on the choice of the deflation tolerances $\mathtt{dtol}_n^{\mathtt{r}}$ and $\mathtt{dtol}_n^{\mathtt{l}}$ in (47) and (49). Clearly, deflation of $\hat{\phi}_n$ or $\hat{\psi}_n$ should occur independent of the actual scaling of the problem. First, consider the special case of bilinear Hankel forms given in terms of a realization (12). To make the deflation check independent of the actual scaling of the columns $\mathbf{r}_j$

of $\mathbf{R}$, of the columns $\mathbf{l}_j$ of $\mathbf{L}$, and of the matrix $\mathbf{A}$, we use the tolerances

$$\begin{aligned} \mathtt{dtol}_n^{\mathtt{r}} &= \begin{cases} \mathtt{dtol} \cdot \left\| \mathbf{r}_{\deg \hat{\boldsymbol{\phi}}_n} \right\| & \text{if } \deg \hat{\boldsymbol{\phi}}_n \le m, \\ \mathtt{dtol} \cdot \operatorname{nest}(\mathbf{A}) & \text{if } \deg \hat{\boldsymbol{\phi}}_n > m, \end{cases} \\ \mathtt{dtol}_n^{\mathtt{l}} &= \begin{cases} \mathtt{dtol} \cdot \left\| \mathbf{l}_{\deg \hat{\boldsymbol{\psi}}_n} \right\| & \text{if } \deg \hat{\boldsymbol{\psi}}_n \le p, \\ \mathtt{dtol} \cdot \operatorname{nest}(\mathbf{A}) & \text{if } \deg \hat{\boldsymbol{\psi}}_n > p. \end{cases} \end{aligned} \tag{50}$$

Here, $\operatorname{nest}(\mathbf{A})$ is either $\|\mathbf{A}\|$ or an estimate of $\|\mathbf{A}\|$, and $\mathtt{dtol}$ is an absolute deflation tolerance. Based on our extensive numerical experiences for the applications outlined in Sections 8.1–8.4 below, we recommend $\mathtt{dtol} = \sqrt{\mathtt{eps}}$, where $\mathtt{eps}$ is the machine precision. In practical applications, the matrix $\mathbf{A}$ is often not available directly, and then the ideal choice $\operatorname{nest}(\mathbf{A}) = \|\mathbf{A}\|$ in (50) is not feasible in general. However, matrix-vector products with $\mathbf{A}$ and $\mathbf{A}^{\mathrm{T}}$ can usually be computed efficiently. In this case, one evaluates quotients of the form

$$\frac{\|\mathbf{A}\,\mathbf{v}\|}{\|\mathbf{v}\|} \quad \text{or} \quad \frac{\left\|\mathbf{A}^{\mathrm{T}}\,\mathbf{w}\right\|}{\|\mathbf{w}\|}$$

for a small number of vectors $\mathbf{v} \in \mathbb{C}^N$, $\mathbf{v} \neq \mathbf{0}$, or $\mathbf{w} \in \mathbb{C}^N$, $\mathbf{w} \neq \mathbf{0}$, and then takes the largest of these quotients as an estimate $\operatorname{nest}(\mathbf{A})$ of $\|\mathbf{A}\|$.

In the case of Hankel matrices $\mathbf{H}$ not given given in terms of a realization (12), the deflation tolerances $\mathtt{dtol}_n^{\mathtt{r}}$, respectively $\mathtt{dtol}_n^{\mathtt{l}}$, in (49) are chosen similar to (50) as products of an absolute deflation tolerance $\mathtt{dtol}$ and factors that take the actual scaling of the columns, respectively rows, of $\mathbf{H}$ into account.

5 Recurrence relations

In this section, we describe the recurrence relations that are used in our computational procedure for generating the polynomials (26). From now on, we always consider the general case where look-ahead may be needed, and we will point out simplifications that occur in the generic case.

5.1 Recurrences in matrix form

Recall that the matrix polynomials $\mathbf{\Phi}_n$ and $\mathbf{\Psi}_n$ introduced in (35) contain the first $n+1$ pairs,

$$\boldsymbol{\phi}_0, \boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_n \quad \text{and} \quad \boldsymbol{\psi}_0, \boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_n, \tag{51}$$

of the polynomials (26). Using the notation $\mathbf{\Phi}_n$ and $\mathbf{\Psi}_n$, the recurrences for generating all the polynomials (51) can be summarized compactly in matrix form as follows:

$$\begin{aligned} \left[\, \mathbf{I}_m \quad \lambda\, \mathbf{\Phi}_n \,\right] &= \mathbf{\Phi}_n\, \mathbf{T}_{n,m+n} + [\, \underbrace{\mathbf{0} \quad \cdots \quad \mathbf{0}}_{m+n+1-m_c} \quad \underbrace{\hat{\boldsymbol{\phi}}_{n+1} \quad \cdots \quad \hat{\boldsymbol{\phi}}_{n+m_c}}_{m_c} \,] + \hat{\mathbf{\Phi}}^{\text{defl}}_{m+n}, \\ \left[\, \mathbf{I}_p \quad \lambda\, \mathbf{\Psi}_n \,\right] &= \mathbf{\Psi}_n\, \widetilde{\mathbf{T}}_{n,p+n} + [\, \underbrace{\mathbf{0} \quad \cdots \quad \mathbf{0}}_{p+n+1-p_c} \quad \underbrace{\hat{\boldsymbol{\psi}}_{n+1} \quad \cdots \quad \hat{\boldsymbol{\psi}}_{n+p_c}}_{p_c} \,] + \hat{\mathbf{\Psi}}^{\text{defl}}_{p+n}. \end{aligned} \tag{52}$$

Relations (52) hold true for all $n = -1, 0, 1, 2, \ldots$, where $n \leq n_{\max}$. Here, we use the convention that $n = -1$ corresponds to the initialization of the first m right and p left auxiliary polynomials,

$$\hat{\boldsymbol{\phi}}_0, \hat{\boldsymbol{\phi}}_1, \ldots, \hat{\boldsymbol{\phi}}_{m-1} \quad \text{and} \quad \hat{\boldsymbol{\psi}}_0, \hat{\boldsymbol{\psi}}_1, \ldots, \hat{\boldsymbol{\psi}}_{p-1},$$

respectively, and we set $\mathbf{\Phi}_{-1} := \mathbf{\Psi}_{-1} := \emptyset$ and $\mathbf{T}_{-1,m-1} := \widetilde{\mathbf{T}}_{-1,p-1} := \emptyset$ in (52). For $n \geq 0$, the matrices

$$\begin{aligned} \mathbf{T}_{n,m+n} &= [\, t_{j,k} \,]_{0 \leq j \leq n,\, -m \leq k \leq n} \in \mathbb{C}^{(n+1) \times (m+n+1)} \\ \text{and} \quad \widetilde{\mathbf{T}}_{n,p+n} &= [\, \tilde{t}_{j,k} \,]_{0 \leq j \leq n,\, -p \leq k \leq n} \in \mathbb{C}^{(n+1) \times (p+n+1)} \end{aligned} \tag{53}$$

contain the recurrence coefficients used for the right and left polynomials, respectively. Corresponding to the partitioning of the matrices on the left-hand sides of (52), the matrices (53) can be written in the form

$$\mathbf{T}_{n,m+n} = [\, \boldsymbol{\rho}_n \quad \mathbf{T}_n \,] \quad \text{and} \quad \widetilde{\mathbf{T}}_{n,p+n} = [\, \boldsymbol{\eta}_n \quad \widetilde{\mathbf{T}}_n \,], \tag{54}$$

where

$$\begin{aligned} \boldsymbol{\rho}_n &:= [\, t_{j,k} \,]_{0 \leq j \leq n,\, -m \leq k < 0} \in \mathbb{C}^{(n+1) \times m}, \quad \mathbf{T}_n = [\, t_{j,k} \,]_{0 \leq j,k \leq n} \in \mathbb{C}^{(n+1) \times (n+1)}, \\ \boldsymbol{\eta}_n &:= [\, \tilde{t}_{j,k} \,]_{0 \leq j \leq n,\, -p \leq k < 0} \in \mathbb{C}^{(n+1) \times p}, \quad \widetilde{\mathbf{T}}_n := [\, \tilde{t}_{j,k} \,]_{0 \leq j,k \leq n} \in \mathbb{C}^{(n+1) \times (n+1)}. \end{aligned}$$

Finally, the matrix polynomials $\hat{\mathbf{\Phi}}_{m+n}^{\text{defl}}$ and $\hat{\mathbf{\Psi}}_{p+n}^{\text{defl}}$ in (52) contain mostly zero columns, together with the polynomials that have been deflated. Recall from Section 4.4 that in our algorithm the polynomials $\hat{\boldsymbol{\phi}}_n$ and $\hat{\boldsymbol{\psi}}_n$ are used to check for necessary deflation. If it is decided that $\hat{\boldsymbol{\phi}}_n$ needs to be deflated, then $\hat{\boldsymbol{\phi}}_n$ is moved into $\hat{\mathbf{\Phi}}_{m+n}^{\text{defl}}$ and becomes its column with index $m+n-m_{\text{c}}$. Otherwise, $\hat{\boldsymbol{\phi}}_n$ is accepted as the next right polynomial $\boldsymbol{\phi}_n$. Similarly, each deflated left polynomial $\hat{\boldsymbol{\psi}}_n$ becomes the $(p+n-p_{\text{c}})$-th column of $\hat{\mathbf{\Psi}}_{p+n}^{\text{defl}}$.

5.2 Structure of the recurrence matrices

Recall that the recurrence matrices for generating scalar formally orthogonal polynomials in the special case $m = p = 1$ are tridiagonal in the generic case and block-tridiagonal in the general case; see, e.g., [14,16,17]. Similarly, the recurrence matrices (53), $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$, exhibit certain structures, although, in the case of deflation, these structures are somewhat more complicated than those for $m = p = 1$. In this subsection, we describe the structures of the matrices (53).

First, consider the simplest case that neither deflation nor look-ahead occur during the construction of the polynomials (51). In this case, $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$ have a banded structure. More precisely, the entries of (53) satisfy

$$\begin{aligned} t_{j,k} &= 0 \quad \text{if} \quad j > k + m \quad \text{or} \quad k > j + p, \\ \tilde{t}_{j,k} &= 0 \quad \text{if} \quad j > k + p \quad \text{or} \quad k > j + m. \end{aligned}$$

In terms of the partitionings (54), this means that $\boldsymbol{\rho}_n$ and $\boldsymbol{\eta}_n$ are upper triangular matrices, $\mathbf{T}_n$ is a banded matrix with lower bandwidth m and upper bandwidth p, and $\widetilde{\mathbf{T}}_n$ is a banded matrix with lower bandwidth p and upper bandwidth m.

Next, consider the case that deflation occurs. Recall that we use the integers m_{c} and p_{c} to count deflations. More precisely, initially $m_{\text{c}} = m$ and $p_{\text{c}} = p$, and then m_{c}, respectively p_{c}, is reduced by one every time a right polynomial $\hat{\boldsymbol{\phi}}_n$, respectively a left polynomial $\hat{\boldsymbol{\psi}}_n$, is deflated. It turns out that m_{c} and p_{c} are also the "current" bandwidths of $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$. This means that the deflation of a right polynomial $\hat{\boldsymbol{\phi}}_n$ reduces m_{c} and thus both the lower bandwidth of $\mathbf{T}_{n,m+n}$ and the upper bandwidth of $\widetilde{\mathbf{T}}_{n,p+n}$ by one. Similarly, the deflation of a left polynomial $\hat{\boldsymbol{\psi}}_n$ reduces p_{c} and thus both the upper bandwidth of $\mathbf{T}_{n,m+n}$ and the lower bandwidth of $\widetilde{\mathbf{T}}_{n,p+n}$ by one. In addition, deflation has a second effect. A deflation of $\hat{\boldsymbol{\phi}}_n$ implies that from now on, the matrix $\widetilde{\mathbf{T}}_{n,p+n}$ will have additional potentially nonzero entries $\tilde{t}_{n-m_{\text{c}},k}$ in row

$n - m_c$ and to the right of its banded part. These additional entries mean that from now on, all left polynomials need to be explicitly orthogonalized against the right polynomial $\boldsymbol{\phi}_{n-m_c}$. Similarly, a deflation of $\hat{\boldsymbol{\psi}}_n$ implies that from now on, the matrix $\mathbf{T}_{n,m+n}$ will have additional potentially nonzero entries $t_{n-p_c,k}$ in row $n - p_c$ and to the right of its banded part. These additional entries mean that from now on, all right polynomials need to be explicitly orthogonalized against the left polynomial $\boldsymbol{\psi}_{n-p_c}$. The size of these additional entries in $\widetilde{\mathbf{T}}_{n,p+n}$ and $\mathbf{T}_{n,m+n}$ can be shown to be bounded by

$$\max_{j=0,1,\ldots,n} \mathtt{dtol}_j^{\mathtt{r}} \quad \text{and} \quad \max_{j=0,1,\ldots,n} \mathtt{dtol}_j^{\mathtt{l}},$$

respectively, where $\mathtt{dtol}_j^{\mathtt{r}}$ and $\mathtt{dtol}_j^{\mathtt{l}}$ are the tolerances used to check for deflation; see Section 4.4. In particular, these additional entries in $\widetilde{\mathbf{T}}_{n,p+n}$ and $\mathbf{T}_{n,m+n}$ all reduce to zero if only exact deflation is performed, i.e., if $\mathtt{dtol}_j^{\mathtt{r}} = \mathtt{dtol}_j^{\mathtt{l}} = 0$ for all j.

Finally, in the general case where both deflation and look-ahead occur, non-trivial look-ahead clusters, i.e., those with $n_{k+1} - n_k > 1$, result in "bulges" just above the banded parts of $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$. In Algorithm 1 below, the cluster indices $\ell_{\boldsymbol{\psi}}$ and $\ell_{\boldsymbol{\phi}}$ defined in (58) and (57) mark the first potentially nonzero elements of the banded parts, including the bulges due to look-ahead, in the n-th column of $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$, respectively. More precisely, $t_{n_{\ell_{\boldsymbol{\psi}}},n}$ is the first potentially nonzero element in the n-th column of $\mathbf{T}_{n,m+n}$, and $\tilde{t}_{n_{\ell_{\boldsymbol{\phi}}},n}$ is the first potentially nonzero element in the n-th column of $\widetilde{\mathbf{T}}_{n,p+n}$. Furthermore, if the row indices $n - m_c$, respectively $n - p_c$, of the additional potentially nonzero elements due to deflation are part of a non-trivial look-ahead cluster, then these additional potentially nonzero elements are spread out over all rows corresponding to that look-ahead cluster. In Algorithm 1 below, the sets $\mathcal{D}_{\boldsymbol{\psi}}$ and $\mathcal{D}_{\boldsymbol{\phi}}$ of cluster indices are used to record these additional nonzero rows above the banded parts of $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$, respectively.

At each n-th pass through the main loop of Algorithm 1, based on $\ell_{\boldsymbol{\psi}}$ and $\mathcal{D}_{\boldsymbol{\psi}}$, we form the set $\mathcal{I}_{\boldsymbol{\psi}}$ in Step 6b) and, based on $\ell_{\boldsymbol{\phi}}$ and $\mathcal{D}_{\boldsymbol{\phi}}$, the set $\mathcal{I}_{\boldsymbol{\phi}}$ in Step 7b). The set $\mathcal{I}_{\boldsymbol{\psi}}$ in (59) contains the indices k of those clusters for which the associated entries $t_{j,n}$, $n_k \leq j < n_{k+1}$ of column n of the matrix $\mathbf{T}_{n,m+n}$ are potentially nonzero. These are just the indices of the clusters $\boldsymbol{\Psi}^{(k)}$ of left polynomials against which the next right auxiliary polynomial $\hat{\boldsymbol{\phi}}_{n+m_c}$ has to be explicitly orthogonalized. Note that the set $\mathcal{I}_{\boldsymbol{\psi}}$ in (59) has two parts. The first part in (59) contains the cluster indices corresponding to spread-out rows due to deflation of earlier left polynomials, while the second part in (59) contains the cluster indices corresponding to the banded part of $\mathbf{T}_{n,m+n}$. Similarly, the set $\mathcal{I}_{\boldsymbol{\phi}}$ in (61) contains the indices k of those clusters for which the associated entries $\tilde{t}_{j,n}$, $n_k \leq j < n_{k+1}$ of column n of the matrix $\widetilde{\mathbf{T}}_{n,p+n}$ are potentially nonzero. These are just the indices of the clusters $\boldsymbol{\Phi}^{(k)}$ of right polynomials

$$
\mathbf{T}_{14,18} = \left[\begin{array}{cccc|ccccccccccccccc}
\bullet & * & * & * & * & * & * & * & * & * & \mathtt{1} & & & & & & & & \\
 & \bullet & * & * & * & * & * & * & * & * & * & & & & & & & & \\
 & & & \bullet & * & * & * & * & * & * & * & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\
 & & & & \bullet & * & * & * & * & * & * & * & \mathtt{1} & & & & & & \\
 & & & & & \bullet & * & * & * & * & * & * & * & \mathtt{1} & \mathtt{1} & \mathtt{1} & \mathtt{1} & \mathtt{1} & \mathtt{1} \\
 & & & & & & \bullet & * & * & * & * & * & * & * & \mathtt{1} & \mathtt{1} & \mathtt{1} & \mathtt{1} & \mathtt{1} \\
 & & & & & & & & \bullet & * & * & * & * & * & \diamond & \diamond & \diamond & \diamond & \diamond \\
 & & & & & & & & & \bullet & * & * & * & * & * & \mathtt{1} & & & \\
 & & & & & & & & & & \bullet & * & * & * & * & * & & & \\
 & & & & & & & & & & & \bullet & * & * & * & * & \diamond & \diamond & \diamond \\
 & & & & & & & & & & & & \bullet & * & * & * & * & & \\
 & & & & & & & & & & & & & & \bullet & * & * & * & \\
 & & & & & & & & & & & & & & & \bullet & * & * & * \\
 & & & & & & & & & & & & & & & & \bullet & * & * \\
 & & & & & & & & & & & & & & & & & \bullet & *
\end{array}\right]
$$

Fig. 1. Structure of $\mathbf{T}_{14,18}$

against which the next left auxiliary polynomial $\hat{\boldsymbol{\psi}}_{n+p_c}$ has to be explicitly orthogonalized. The first parts of the set $\mathcal{I}_{\boldsymbol{\phi}}$ in (61) contains the cluster indices corresponding to spread-out rows due to deflation of earlier right polynomials, while the second part in (61) contains the cluster indices corresponding to the banded part of $\widetilde{\mathbf{T}}_{n,m+n}$.

5.3 An example

In this subsection, we illustrate the structure of the recurrence matrices with an example.

Consider the case that $m = 4$ and $p = 5$, and assume that in the associated block Hankel matrix, the columns with index $k = 2+4i, 8+4i, 19+4i$, $i \in \mathbb{N}_0$, and the rows with index $j = 7+5i, 11+5i, 15+5i$, $i \in \mathbb{N}_0$, need to be deflated. The associated indices (18) of the columns and rows left after deflation are as follows:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	$\cdots$
μ_n	0	1	3	4	5	7	9	11	13	15	17	21	25	29	33	$\cdots$
ν_n	0	1	2	3	4	5	6	8	9	10	13	14	18	19	23	$\cdots$

In terms of the auxiliary polynomials (45), these deflations of columns and rows translate into deflations of the right auxiliary polynomials $\hat{\boldsymbol{\phi}}_2$ (when $m_c = 4$), $\hat{\boldsymbol{\phi}}_6$ (when $m_c = 3$), and $\hat{\boldsymbol{\phi}}_{11}$ (when $m_c = 2$), and the left auxiliary polynomials $\hat{\boldsymbol{\psi}}_2$ (when $p_c = 4$), $\hat{\boldsymbol{\psi}}_6$ (when $p_c = 3$), and $\hat{\boldsymbol{\psi}}_{11}$ (when $p_c = 2$).

$$
\widetilde{\mathbf{T}}_{14,19} = \left[\begin{array}{ccccc|ccccccccccccccc}
\bullet & * & * & * & * & * & * & * & * & & & & & & & & & & & \\
 & \bullet & * & * & * & * & * & * & * & * & 1 & 1 & & & & & & & & \\
 & & \bullet & * & * & * & * & * & * & * & * & 1 & & & & & & & & \\
 & & & \bullet & * & * & * & * & * & * & * & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond & \diamond \\
 & & & & \bullet & * & * & * & * & * & * & * & 1 & 1 & & & & & & \\
 & & & & & \bullet & * & * & * & * & * & * & * & 1 & & & & & & \\
 & & & & & & \bullet & * & * & * & * & * & * & * & & & & & & \\
 & & & & & & & & \bullet & * & * & * & * & * & * & 1 & & & & \\
 & & & & & & & & & \bullet & * & * & * & * & * & * & & & & \\
 & & & & & & & & & & \bullet & * & * & * & * & * & \diamond & \diamond & \diamond & \diamond \\
 & & & & & & & & & & & & \bullet & * & * & * & * & & & \\
 & & & & & & & & & & & & & \bullet & * & * & * & * & & \\
 & & & & & & & & & & & & & & & \bullet & * & * & * & \\
 & & & & & & & & & & & & & & & & \bullet & * & * & * \\
 & & & & & & & & & & & & & & & & & \bullet & * & *
\end{array}\right]
$$

Fig. 2. Structure of $\widetilde{\mathbf{T}}_{14,19}$

We now determine the structure of the recurrence matrices $\mathbf{T}_{14,18}$ and $\widetilde{\mathbf{T}}_{14,19}$ at $n = 14$. First, we assume that no look-ahead occurs. Recall from (44) that then $\gamma(i) = i$ for all i. Algorithm 1 produces the sets $\mathcal{D}_{\boldsymbol{\phi}} = \{3, 9\}$ and $\mathcal{D}_{\boldsymbol{\psi}} = \{6, 9\}$. Moreover, the values of the row indices $\ell_{\boldsymbol{\phi}}$ and $\ell_{\boldsymbol{\psi}}$ given by (57) and (58) are as follows:

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\ell_{\boldsymbol{\phi}}$	0	0	0	0	1	2	4	5	6	7	8	10	11	12	13
$\ell_{\boldsymbol{\psi}}$	0	0	0	0	0	0	1	3	4	5	7	8	10	11	12

The resulting structure of $\mathbf{T}_{14,18}$ and $\widetilde{\mathbf{T}}_{14,19}$ is shown in Figure 1 and 2, respectively. Here, the following convention is used: guaranteed positive elements are marked by "$\bullet$", other potentially nonzero elements within the banded parts are marked by "$*$", and potentially nonzero elements outside the banded parts due to deflation are marked by "$\diamond$". Moreover, the vertical lines in Figures 1 and 2 indicate the partitioning (54) of $\mathbf{T}_{14,18}$ into $\boldsymbol{\rho}_{14}$ and $\mathbf{T}_{14}$, and of $\widetilde{\mathbf{T}}_{14,19}$ into $\boldsymbol{\eta}_{14}$ and $\widetilde{\mathbf{T}}_{14}$, respectively.

Next, we assume that two non-trivial look-ahead clusters occur: one cluster of length $n_5 - n_4 = 3$ starting at $n_4 = 4$, followed by a cluster of length $n_6 - n_5 = 2$ starting at $n_5 = 7$. Algorithm 1 now produces the sets of cluster indices $\mathcal{D}_{\boldsymbol{\phi}} = \{3, 6\}$ and $\mathcal{D}_{\boldsymbol{\psi}} = \{2, 4, 6\}$, and the values of the cluster indices

$\ell_{\boldsymbol{\phi}}$ and $\ell_{\boldsymbol{\psi}}$ given by (57) and (58) are as follows:

k	0	1	2	3	4	5	6	7	8	9	10	11
n_k	0	1	2	3	4	7	9	10	11	12	13	14
$\ell_{\boldsymbol{\phi}}$	0	0	0	0	1	4	5	5	7	8	9	10
$\ell_{\boldsymbol{\psi}}$	0	0	0	0	0	3	4	5	5	7	8	9

The resulting structure of $\mathbf{T}_{14,18}$ and $\widetilde{\mathbf{T}}_{14,19}$ is again shown in Figure 1 and 2, respectively, where we have used "1" to mark the additional potentially nonzero entries caused by the two non-trivial look-ahead clusters.

6 The algorithm

In this section, we present a detailed statement of the complete computational procedure for constructing formally orthogonal polynomials associated with a given bilinear Hankel form.

At pass n through the main loop of Algorithm 1 below, we construct the n-th pair of polynomials, $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$. We use the counter ℓ to denote the index of the clusters to which $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$ are added. This means that, at pass n, the currently constructed look-ahead clusters are

$$\boldsymbol{\Phi}^{(\ell)} := [\, \boldsymbol{\phi}_{n_\ell} \quad \cdots \quad \boldsymbol{\phi}_n \,] \quad \text{and} \quad \boldsymbol{\Psi}^{(\ell)} := [\, \boldsymbol{\psi}_{n_\ell} \quad \cdots \quad \boldsymbol{\psi}_n \,].$$

We also check if these look-ahead clusters are complete. If they are, then the polynomials $\boldsymbol{\phi}_{n+1}$ and $\boldsymbol{\psi}_{n+1}$ constructed during the next, $(n+1)$-st, pass start new clusters. All the other notation used in the following statement of Algorithm 1 has already been introduced.

Algorithm 1 (Construction of polynomials associated with $\langle\cdot,\cdot\rangle$.)
INPUT: *A bilinear Hankel form* $\langle\cdot,\cdot\rangle\colon\ \mathcal{P}^{(p)} \times \mathcal{P}^{(m)} \mapsto \mathbb{C}$.

0) *Set* $\hat{\boldsymbol{\phi}}_i = \mathbf{e}_{i+1}^{(m)}$ *and* $\hat{\mu}_i = i$ *for* $i = 0, 1, \ldots, m-1$.
Set $\hat{\boldsymbol{\psi}}_i = \mathbf{e}_{i+1}^{(p)}$ *and* $\hat{\nu}_i = i$ *for* $i = 0, 1, \ldots, p-1$.
Set $m_{\mathrm{c}} = m$ *and* $p_{\mathrm{c}} = p$.
Set $\mathcal{D}_{\boldsymbol{\phi}} = \mathcal{D}_{\boldsymbol{\psi}} = \emptyset$ *and* $\ell_{\boldsymbol{\phi}} = \ell_{\boldsymbol{\psi}} = 0$.
Set $\ell = 0$, $n_0 = 0$, *and* $\boldsymbol{\Phi}^{(0)} = \boldsymbol{\Psi}^{(0)} = \emptyset$.

For $n = 0, 1, 2, \ldots,$ *do*:

1) (If necessary, deflate $\hat{\boldsymbol{\phi}}_n$.)
   Decide if $\hat{\boldsymbol{\phi}}_n$ *needs to be deflated.*
   If no, continue with Step 2).
   If yes, deflate $\hat{\boldsymbol{\phi}}_n$ *by doing the following*:
   1a) *If* $m_{\mathrm{c}} = 1$, *then stop.*
   (There are no more right orthogonal polynomials.)
   1b) *If* $n \geq m_{\mathrm{c}}$, *set* $\mathcal{D}_{\boldsymbol{\phi}} = \mathcal{D}_{\boldsymbol{\phi}} \cup \{\, \gamma(n - m_{\mathrm{c}}) \,\}$.
   ($\mathcal{D}_{\boldsymbol{\phi}}$ records the cluster indices of right polynomials against which we have to explicitly orthogonalize from now on.)
   1c) *Set* $\hat{\boldsymbol{\phi}}_i = \hat{\boldsymbol{\phi}}_{i+1}$ *and* $\hat{\mu}_i = \hat{\mu}_{i+1}$ *for* $i = n, n+1, \ldots, n + m_{\mathrm{c}} - 2$.
   (The polynomial $\hat{\boldsymbol{\phi}}_n$ is deflated and becomes the $(m + n - m_{\mathrm{c}})$-th column of the matrix $\hat{\boldsymbol{\Phi}}^{\mathrm{defl}}_{m+n}$ in (52); the indices of the remaining right auxiliary polynomials are reduced by one.)
   1d) *Set* $m_{\mathrm{c}} = m_{\mathrm{c}} - 1$.
   (The current right block size is reduced by one.)
   1e) *Repeat all of Step* 1).
2) (If necessary, deflate $\hat{\boldsymbol{\psi}}_n$.)
   Decide if $\hat{\boldsymbol{\psi}}_n$ *needs to be deflated.*
   If no, continue with Step 3).
   If yes, deflate $\hat{\boldsymbol{\psi}}_n$ *by doing the following*:
   2a) *If* $p_{\mathrm{c}} = 1$, *then stop.*
   (There are no more left orthogonal polynomials.)
   2b) *If* $n \geq p_{\mathrm{c}}$, *set* $\mathcal{D}_{\boldsymbol{\psi}} = \mathcal{D}_{\boldsymbol{\psi}} \cup \{\, \gamma(n - p_{\mathrm{c}}) \,\}$.
   ($\mathcal{D}_{\boldsymbol{\psi}}$ records the cluster indices of left polynomials against which we have to explicitly orthogonalize from now on.)
   2c) *Set* $\hat{\boldsymbol{\psi}}_i = \hat{\boldsymbol{\psi}}_{i+1}$ *and* $\hat{\nu}_i = \hat{\nu}_{i+1}$ *for* $i = n, n+1, \ldots, n + p_{\mathrm{c}} - 2$.
   (The polynomial $\hat{\boldsymbol{\psi}}_n$ is deflated and becomes the $(p + n - p_{\mathrm{c}})$-th column of the matrix $\hat{\boldsymbol{\Psi}}^{\mathrm{defl}}_{p+n}$ in (52); the indices of the remaining left auxiliary polynomials are reduced by one.)
   2d) *Set* $p_{\mathrm{c}} = p_{\mathrm{c}} - 1$.
   (The current right block size is reduced by one.)
   2e) *Repeat all of Step* 2).
3) (Normalize $\hat{\boldsymbol{\phi}}_n$ and $\hat{\boldsymbol{\psi}}_n$ to obtain $\boldsymbol{\phi}_n$ and $\boldsymbol{\psi}_n$, and add them to the current clusters $\boldsymbol{\Phi}^{(\ell)}$ and $\boldsymbol{\Psi}^{(\ell)}$.)
   Set

$$\boldsymbol{\phi}_n = \frac{\hat{\boldsymbol{\phi}}_n}{t_{n,n-m_{\mathrm{c}}}}, \quad \mu_n = \hat{\mu}_n, \quad \textit{and} \quad \boldsymbol{\psi}_n = \frac{\hat{\boldsymbol{\psi}}_n}{\tilde{t}_{n,n-p_{\mathrm{c}}}}, \quad \nu_n = \hat{\nu}_n,$$

   where $t_{n,n-m_{\mathrm{c}}} > 0$ *and* $\tilde{t}_{n,n-p_{\mathrm{c}}} > 0$ *are suitable scaling factors.*
   Set $\boldsymbol{\Phi}^{(\ell)} = [\, \boldsymbol{\Phi}^{(\ell)} \quad \boldsymbol{\phi}_n \,]$ *and* $\boldsymbol{\Psi}^{(\ell)} = [\, \boldsymbol{\Psi}^{(\ell)} \quad \boldsymbol{\psi}_n \,]$.
4) (Compute $\boldsymbol{\Delta}^{(\ell)}$ and check for end of look-ahead cluster.)

Form the matrix $\mathbf{\Delta}^{(\ell)} = \langle \mathbf{\Psi}^{(\ell)}, \mathbf{\Phi}^{(\ell)} \rangle$.
If the matrix $\mathbf{\Delta}^{(\ell)}$ *is singular or in some sense "close" to singular, continue with Step* 6).

5) (The ℓ-th look-ahead clusters $\mathbf{\Phi}^{(\ell)}$ and $\mathbf{\Psi}^{(\ell)}$ are complete and the following "end-of-clusters" updates are performed.)

5a) (Orthogonalize the polynomials $\hat{\boldsymbol{\phi}}_{n+1}, \hat{\boldsymbol{\phi}}_{n+2}, \ldots, \hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}-1}$ against $\mathbf{\Psi}^{(\ell)}$.)
For $i = n+1, n+2, \ldots, n+m_{\mathrm{c}}-1$, *set*

$$\begin{aligned} [\, t_{j,i-m_{\mathrm{c}}} \,]_{n_\ell \le j \le n} &= \left(\mathbf{\Delta}^{(\ell)}\right)^{-1} \langle \mathbf{\Psi}^{(\ell)}, \hat{\boldsymbol{\phi}}_i \rangle, \\ \hat{\boldsymbol{\phi}}_i &= \hat{\boldsymbol{\phi}}_i - \mathbf{\Phi}^{(\ell)} \, [\, t_{j,i-m_{\mathrm{c}}} \,]_{n_\ell \le j \le n} \, . \end{aligned} \tag{55}$$

5b) (Orthogonalize the polynomials $\hat{\boldsymbol{\psi}}_{n+1}, \hat{\boldsymbol{\psi}}_{n+2}, \ldots, \hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}-1}$ against $\mathbf{\Phi}^{(\ell)}$.)
For $i = n+1, n+2, \ldots, n+p_{\mathrm{c}}-1$, *set*

$$\begin{aligned} [\, \tilde{t}_{j,i-p_{\mathrm{c}}} \,]_{n_\ell \le j \le n} &= \left(\mathbf{\Delta}^{(\ell)}\right)^{-\mathrm{T}} \langle \hat{\boldsymbol{\psi}}_i, \mathbf{\Phi}^{(\ell)} \rangle^{\mathrm{T}}, \\ \hat{\boldsymbol{\psi}}_i &= \hat{\boldsymbol{\psi}}_i - \mathbf{\Psi}^{(\ell)} \, [\, \tilde{t}_{j,i-p_{\mathrm{c}}} \,]_{n_\ell \le j \le n} \, . \end{aligned} \tag{56}$$

5c) *If* $\mu_{n_\ell} \ge m$, *set*

$$\ell_{\boldsymbol{\phi}} = \gamma\Big(\mu^{-1}(\mu_{n_\ell} - m)\Big). \tag{57}$$

If $\nu_{n_\ell} \ge p$, *set*

$$\ell_{\boldsymbol{\psi}} = \gamma\Big(\nu^{-1}(\nu_{n_\ell} - p)\Big). \tag{58}$$

5d) *Set* $\ell = \ell + 1$, $n_\ell = n + 1$, *and* $\mathbf{\Phi}^{(\ell)} = \mathbf{\Psi}^{(\ell)} = \emptyset$.
(The polynomials $\boldsymbol{\phi}_{n+1}$ and $\boldsymbol{\psi}_{n+1}$ constructed in the next iteration start new look-ahead clusters.)

6) (Obtain new right polynomial $\hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}}$ and make it orthogonal to complete right clusters.)

6a) *Set* $\hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}} = \lambda \, \boldsymbol{\phi}_n$ *and* $\hat{\mu}_{n+m_{\mathrm{c}}} = \mu_n + m$.

6b) (Determine the indices of the left clusters $\mathbf{\Psi}^{(k)}$ against which $\hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}}$ needs to be orthogonalized.)
Set

$$\mathcal{I}_{\boldsymbol{\psi}} = \{\, k \mid k \in \mathcal{D}_{\boldsymbol{\psi}} \text{ and } k < \ell_{\boldsymbol{\psi}} \,\} \cup \{\, \ell_{\boldsymbol{\psi}}, \ell_{\boldsymbol{\psi}} + 1, \ldots, \ell - 1 \,\}. \tag{59}$$

6c) (Orthogonalize $\hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}}$ against these clusters.)
For all $k \in \mathcal{I}_{\boldsymbol{\psi}}$ *(in ascending order), set*

$$\begin{aligned} [\, t_{j,n} \,]_{n_k \le j < n_{k+1}} &= \left(\mathbf{\Delta}^{(k)}\right)^{-1} \langle \mathbf{\Psi}^{(k)}, \hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}} \rangle, \\ \hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}} &= \hat{\boldsymbol{\phi}}_{n+m_{\mathrm{c}}} - \mathbf{\Phi}^{(k)} \, [\, t_{j,n} \,]_{n_k \le j < n_{k+1}} \, . \end{aligned} \tag{60}$$

7) (Obtain new left polynomial $\hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}}$ and make it orthogonal to complete right clusters.)
   7a) *Set* $\hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}} = \lambda\, \boldsymbol{\psi}_n$ *and* $\hat{\nu}_{n+p_{\mathrm{c}}} = \nu_n + p$.
   7b) (Determine the indices of the right clusters $\boldsymbol{\Phi}^{(k)}$ against which $\hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}}$ needs to be orthogonalized.)
   Set

$$\mathcal{I}_{\boldsymbol{\phi}} = \{\, k \mid k \in \mathcal{D}_{\boldsymbol{\phi}} \text{ and } k < \ell_{\boldsymbol{\phi}} \,\} \cup \{\, \ell_{\boldsymbol{\phi}}, \ell_{\boldsymbol{\phi}}+1, \ldots, \ell-1 \,\}. \tag{61}$$

   7c) (Orthogonalize $\hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}}$ against these clusters.)
   For all $k \in \mathcal{I}_{\boldsymbol{\phi}}$ *(in ascending order), set*

$$\begin{aligned} [\, \tilde{t}_{j,n} \,]_{n_k \le j < n_{k+1}} &= \left(\boldsymbol{\Delta}^{(k)}\right)^{-\mathrm{T}} \langle \hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}}, \boldsymbol{\Phi}^{(k)} \rangle^{\mathrm{T}}, \\ \hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}} &= \hat{\boldsymbol{\psi}}_{n+p_{\mathrm{c}}} - \boldsymbol{\Psi}^{(k)}\, [\, \tilde{t}_{j,n} \,]_{n_k \le j < n_{k+1}}. \end{aligned} \tag{62}$$

In the following, let $\mathbf{T}_{n,m+n}$ and $\widetilde{\mathbf{T}}_{n,p+n}$ be the matrices (53) with entries $t_{j,k}$ and $\tilde{t}_{j,k}$ generated by Algorithm 1. Here we use the convention that entries that are not explicitly generated in Algorithm 1 are set to be zero.

7 Properties

In the following theorem, we summarize the key properties of Algorithm 1.

Theorem 4 (Properties of Algorithm 1.)

a) *For each* $n = -1, 0, 1, 2, \ldots$, *the polynomials and matrices that have been generated after the* n*-th pass through the main loop of Algorithm* 1*, satisfy the recurrence relations* (52)*.*
b) *For each* $n = 0, 1, 2, \ldots$, *the polynomials that have been generated after the* n*-th pass through the main loop of Algorithm* 1*, satisfy*

$$\begin{aligned} \boldsymbol{\Phi}_n &:= [\, \boldsymbol{\phi}_0 \quad \boldsymbol{\phi}_1 \quad \cdots \quad \boldsymbol{\phi}_n \,] = [\, \boldsymbol{\Phi}^{(0)} \quad \boldsymbol{\Phi}^{(1)} \quad \cdots \quad \boldsymbol{\Phi}^{(\ell)} \,], \\ \boldsymbol{\Psi}_n &:= [\, \boldsymbol{\psi}_0 \quad \boldsymbol{\psi}_1 \quad \cdots \quad \boldsymbol{\psi}_n \,] = [\, \boldsymbol{\Psi}^{(0)} \quad \boldsymbol{\Psi}^{(1)} \quad \cdots \quad \boldsymbol{\Psi}^{(\ell)} \,], \end{aligned} \tag{63}$$

and the cluster-wise orthogonality conditions

$$\begin{aligned} \langle \boldsymbol{\Psi}_n, \boldsymbol{\Phi}_n \rangle &= \boldsymbol{\Delta}_n := \operatorname{diag}\left(\boldsymbol{\Delta}^{(0)}, \boldsymbol{\Delta}^{(1)}, \ldots, \boldsymbol{\Delta}^{(\ell)}\right), \\ \langle \boldsymbol{\Psi}_{n_\ell - 1}, \hat{\boldsymbol{\phi}}_{n+i} \rangle &= \mathbf{0}, \quad i = 1, 2, \ldots, m_{\mathrm{c}}, \\ \langle \hat{\boldsymbol{\psi}}_{n+i}, \boldsymbol{\Phi}_{n_\ell - 1} \rangle &= \mathbf{0}, \quad i = 1, 2, \ldots, p_{\mathrm{c}}, \end{aligned} \tag{64}$$

where $\ell = \gamma(n+1)$.

SKETCH OF PROOF Part a), as well as the partitioning property (63), can be directly verified.

The cluster-wise orthogonality conditions (64) are proved using induction on n. By the induction hypothesis, before Step 5a) in Algorithm 1 is performed, the right auxiliary polynomials $\hat{\boldsymbol{\phi}}_{n+1}, \hat{\boldsymbol{\phi}}_{n+2}, \ldots, \hat{\boldsymbol{\phi}}_{n+m_c-1}$ are already orthogonal to all left clusters $\boldsymbol{\Psi}^{(k)}$ with $0 \leq k < \ell$. Thus it only remains to orthogonalize these polynomials against $\boldsymbol{\Psi}^{(\ell)}$, and this is obviously achieved by the update (55). Similarly, the update (56) is sufficient to ensure that the left auxiliary polynomials $\hat{\boldsymbol{\psi}}_{n+1}, \hat{\boldsymbol{\psi}}_{n+2}, \ldots, \hat{\boldsymbol{\psi}}_{n+p_c-1}$ are orthogonal against all right clusters $\boldsymbol{\Phi}^{(k)}$ with $0 \leq k \leq \ell$. Next, consider the update (60) of $\hat{\boldsymbol{\phi}}_{n+m_c}$. Here, we need to show that for the "omitted" clusters $\boldsymbol{\Psi}^{(k)}$, $k \notin \mathcal{I}_{\boldsymbol{\psi}}$, we have

$$\langle \boldsymbol{\Psi}^{(k)}, \hat{\boldsymbol{\phi}}_{n+m_c} \rangle = [\, \langle \boldsymbol{\psi}_j, \hat{\boldsymbol{\phi}}_{n+m_c} \rangle \,]_{n_k \leq j < n_{k+1}} = \mathbf{0}.$$

To this end, assume that $k \notin \mathcal{I}_{\boldsymbol{\psi}}$ and let $n_k \leq j < n_{k+1}$. Since $\hat{\boldsymbol{\phi}}_{n+m_c} = \lambda\, \boldsymbol{\phi}_n$ and using the shift property (6) of the bilinear Hankel form, it follows that

$$\langle \boldsymbol{\psi}_j, \hat{\boldsymbol{\phi}}_{n+m_c} \rangle = \langle \lambda\, \boldsymbol{\psi}_j, \boldsymbol{\phi}_n \rangle. \tag{65}$$

The polynomial $\lambda\, \boldsymbol{\psi}_j$ itself was used at the earlier pass j through the main loop of Algorithm 1 to obtain the auxiliary polynomial $\hat{\boldsymbol{\psi}}_{j+p_c(j)}$. Now, there are two cases. The first one is that $\hat{\boldsymbol{\psi}}_{j+p_c(j)}$ was deflated later on; in this case, however, $k \in \mathcal{D}_{\boldsymbol{\psi}}$, and thus, by (59), $k \in \mathcal{I}_{\boldsymbol{\psi}}$, which contradicts our assumption. This leaves the second case that $\hat{\boldsymbol{\psi}}_{j+p_c(j)}$ was not deflated. In this case, $\hat{\boldsymbol{\psi}}_{j+p_c(j)}$ was orthogonalized to become $\boldsymbol{\psi}_{j'}$ and was added to the cluster with index $k' = \gamma(j')$ at pass j'. One can show that $\ell_{\boldsymbol{\psi}}$ in (58) is just chosen such that $k < \ell_{\boldsymbol{\psi}}$ implies $k' < \ell$. Thus the polynomials $\boldsymbol{\psi}_{j'}$ and $\boldsymbol{\phi}_n$ are part of clusters with different indices, and together with (65), it follows that

$$\langle \boldsymbol{\psi}_j, \hat{\boldsymbol{\phi}}_{n+m_c} \rangle = \langle \lambda\, \boldsymbol{\psi}_j, \boldsymbol{\phi}_n \rangle = \langle \boldsymbol{\psi}_{j'}, \boldsymbol{\phi}_n \rangle = 0$$

for all $n_k \leq j < n_{k+1}$ with $k \notin \mathcal{I}_{\boldsymbol{\psi}}$.

Similarly, one shows that the clusters with $k \notin \mathcal{I}_{\boldsymbol{\phi}}$ can indeed be omitted in (62). $\square$

8 Applications

In this section, we sketch some applications of Algorithm 1. In the following, we assume that $\mathbf{A}$, $\mathbf{R}$, $\mathbf{L}$ is a given triplet of matrices of the form (12).

8.1 A Lanczos-type algorithm for multiple starting vectors

The first application is a Lanczos-type method that extends the classical Lanczos process [19] to multiple right and left starting vectors. We denote by $\mathcal{K}_n(\mathbf{A}, \mathbf{R})$ the *n-th right block Krylov subspace* spanned by the first $n+1$ columns of the deflated right block Krylov matrix (16), and by $\mathcal{K}_n(\mathbf{A}^{\mathrm{T}}, \mathbf{L})$ the *n-th left block Krylov subspace* spanned by the first $n+1$ columns of the deflated left block Krylov matrix (17). The goal of the Lanczos-type method is to generate bi-orthogonal basis vectors for $\mathcal{K}_n(\mathbf{A}, \mathbf{R})$ and $\mathcal{K}_n(\mathbf{A}^{\mathrm{T}}, \mathbf{L})$. To this end, using (14), we associate with the polynomials generated by Algorithm 1 the so-called right and left Lanczos vectors,

$$\mathbf{v}_n = \boldsymbol{\phi}_n(\mathbf{A}) \circ \mathbf{R} \quad \text{and} \quad \mathbf{w}_n = \boldsymbol{\psi}_n(\mathbf{A}^{\mathrm{T}}) \circ \mathbf{L}, \quad n = 0, 1, \ldots, \tag{66}$$

respectively. Then, by re-stating Algorithm 1 in terms of the vectors (66), instead of polynomials, we obtain the desired Lanczos-type method for multiple right and left starting vectors. In fact, the resulting algorithm is a look-ahead version of the Lanczos-type method stated in [10, Algorithm 9.2], and it can also be viewed as a variant of the Lanczos-type method proposed in [1].

It is easy to verify that the right and left Lanczos vectors (66) indeed span the right and left block Krylov subspaces, i.e.,

$$\begin{aligned} \operatorname{span}\{\, \mathbf{v}_0, \mathbf{v}_1, \ldots, \mathbf{v}_n \,\} &= \mathcal{K}_n(\mathbf{A}, \mathbf{R}) \\ \text{and} \quad \operatorname{span}\{\, \mathbf{w}_0, \mathbf{w}_1, \ldots, \mathbf{w}_n \,\} &= \mathcal{K}_n(\mathbf{A}^{\mathrm{T}}, \mathbf{L}). \end{aligned} \tag{67}$$

Now let $\mathbf{V}_n := [\, \mathbf{v}_0 \quad \mathbf{v}_1 \quad \cdots \quad \mathbf{v}_n \,]$ and $\mathbf{W}_n := [\, \mathbf{w}_0 \quad \mathbf{w}_1 \quad \cdots \quad \mathbf{w}_n \,]$ denote the matrices whose columns are the first $n+1$ right and left Lanczos vectors. The polynomial recurrences (52) then translate into the following compact formulation of the recurrences used to generate the Lanczos vectors:

$$
\begin{aligned}
[\,\mathbf{R} \quad \mathbf{A}\,\mathbf{V}_n\,] &= \mathbf{V}_n\,\mathbf{T}_{n,m+n} + [\,\underbrace{\mathbf{0} \ \cdots \ \mathbf{0}}_{m+n+1-m_c} \ \underbrace{\hat{\mathbf{v}}_{n+1} \ \cdots \ \hat{\mathbf{v}}_{n+m_c}}_{m_c}\,] + \hat{\mathbf{V}}^{\mathrm{defl}}_{m+n}, \\
[\,\mathbf{L} \quad \mathbf{A}^{\mathrm{T}}\,\mathbf{W}_n\,] &= \mathbf{W}_n\,\widetilde{\mathbf{T}}_{n,p+n} + [\,\underbrace{\mathbf{0} \ \cdots \ \mathbf{0}}_{p+n+1-p_c} \ \underbrace{\hat{\mathbf{w}}_{n+1} \ \cdots \ \hat{\mathbf{w}}_{n+p_c}}_{p_c}\,] + \hat{\mathbf{W}}^{\mathrm{defl}}_{p+n}.
\end{aligned}
\tag{68}
$$

Here, the only possible nonzero columns of $\hat{\mathbf{V}}^{\mathrm{defl}}_{m+n}$ and $\hat{\mathbf{W}}^{\mathrm{defl}}_{p+n}$ are the columns containing deflated vectors. Moreover, in view of (47), even these columns are zero and thus

$$
\hat{\mathbf{V}}^{\mathrm{defl}}_{m+n} = \mathbf{0} \quad \text{and} \quad \hat{\mathbf{W}}^{\mathrm{defl}}_{p+n} = \mathbf{0} \tag{69}
$$

if only exact deflation is performed. Finally, the orthogonality conditions (64) translate into the following cluster-wise bi-orthogonality relations of the Lanczos vectors:

$$
\begin{aligned}
\mathbf{W}_n^{\mathrm{T}}\,\mathbf{V}_n &= \boldsymbol{\Delta}_n := \operatorname{diag}\left(\boldsymbol{\Delta}^{(0)}, \boldsymbol{\Delta}^{(1)}, \ldots, \boldsymbol{\Delta}^{(\ell)}\right), \\
\mathbf{W}_{n_\ell-1}^{\mathrm{T}}\,\hat{\mathbf{v}}_{n+i} &= \mathbf{0}, \quad i = 1, 2, \ldots, m_{\mathrm{c}}, \\
\mathbf{V}_{n_\ell-1}^{\mathrm{T}}\,\hat{\mathbf{w}}_{n+i} &= \mathbf{0}, \quad i = 1, 2, \ldots, p_{\mathrm{c}}.
\end{aligned}
\tag{70}
$$

Here, $\ell = \gamma(n+1)$.

In the remainder of this paper, we always assume that n corresponds to the end of the ℓ-th look-ahead cluster, i.e., $n = n_\ell - 1$. This condition guarantees that, in (70), all blocks $\boldsymbol{\Delta}^{(k)}$ of the block-diagonal matrix $\boldsymbol{\Delta}_n$ and thus $\boldsymbol{\Delta}_n$ itself are nonsingular.

By multiplying the first relation in (68) from the left by $\boldsymbol{\Delta}_n^{-1}\,\mathbf{W}_n^{\mathrm{T}}$ and using the bi-orthogonality relations (70), as well as the partitioning of $\mathbf{T}_{n,m+n}$ in (54), we obtain

$$
\begin{aligned}
[\,\boldsymbol{\Delta}_n^{-1}\,\mathbf{W}_n^{\mathrm{T}}\,\mathbf{R} \quad \boldsymbol{\Delta}_n^{-1}\,\mathbf{W}_n^{\mathrm{T}}\,\mathbf{A}\,\mathbf{V}_n\,] &= [\,\boldsymbol{\rho}_n \quad \mathbf{T}_n\,] + \boldsymbol{\Delta}_n^{-1}\,\mathbf{W}_n^{\mathrm{T}}\,\hat{\mathbf{V}}^{\mathrm{defl}}_{m+n} \\
&=: [\,\boldsymbol{\rho}_n^{\mathrm{proj}} \quad \mathbf{T}_n^{\mathrm{proj}}\,].
\end{aligned}
\tag{71}
$$

Similarly, by multiplying the second relation in (68) from the left by $\mathbf{\Delta}_n^{-\mathrm{T}} \mathbf{V}_n^{\mathrm{T}}$, we get

$$\begin{aligned}\left[\, \mathbf{\Delta}_n^{-\mathrm{T}} \mathbf{V}_n^{\mathrm{T}} \mathbf{L} \quad \mathbf{\Delta}_n^{-\mathrm{T}} \mathbf{V}_n^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{W}_n \,\right] &= \left[\, \boldsymbol{\eta}_n \quad \widetilde{\mathbf{T}}_n \,\right] + \mathbf{\Delta}_n^{-\mathrm{T}} \mathbf{V}_n^{\mathrm{T}} \hat{\mathbf{W}}_{p+n}^{\mathrm{defl}} \\ &=: \left[\, \boldsymbol{\eta}_n^{\mathrm{proj}} \quad \widetilde{\mathbf{T}}_n^{\mathrm{proj}} \,\right].\end{aligned} \tag{72}$$

Recall that $\hat{\mathbf{V}}_{m+n}^{\mathrm{defl}}$ and $\hat{\mathbf{W}}_{p+n}^{\mathrm{defl}}$ contain mostly zero columns, together with the deflated vectors. This can be used to show that the matrices $\mathbf{T}_n$ and $\widetilde{\mathbf{T}}_n$ and the *projected* versions $\mathbf{T}_n^{\mathrm{proj}}$ and $\widetilde{\mathbf{T}}_n^{\mathrm{proj}}$ differ only in a few entries in their lower triangular parts, respectively. Furthermore, from (71) and (72), it follows that

$$\mathbf{W}_n^{\mathrm{T}} \mathbf{A} \mathbf{V}_n = \mathbf{\Delta}_n \mathbf{T}_n^{\mathrm{proj}} = \left(\widetilde{\mathbf{T}}_n^{\mathrm{proj}}\right)^{\mathrm{T}} \mathbf{\Delta}_n. \tag{73}$$

This relation implies that $\mathbf{T}_n^{\mathrm{proj}}$ can be generated directly from only $\mathbf{T}_n$, $\widetilde{\mathbf{T}}_n$, and $\mathbf{\Delta}_n$, without using the term $\mathbf{\Delta}_n^{-1} \mathbf{W}_n^{\mathrm{T}} \hat{\mathbf{V}}_{m+n}^{\mathrm{defl}}$ in (71).

For later use, we note that in the case of exact deflation, by (69), (71), and (72),

$$\boldsymbol{\rho}_n = \mathbf{\Delta}_n^{-1} \mathbf{W}_n^{\mathrm{T}} \mathbf{R}, \quad \boldsymbol{\eta}_n = \mathbf{\Delta}_n^{-\mathrm{T}} \mathbf{V}_n^{\mathrm{T}} \mathbf{L}, \quad \mathbf{T}_n = \mathbf{\Delta}_n^{-1} \mathbf{W}_n^{\mathrm{T}} \mathbf{A} \mathbf{V}_n. \tag{74}$$

Next, we describe three applications of the Lanczos-type method sketched in this subsection.

8.2 Padé approximation of matrix-valued transfer functions

The matrix triplet (12) induces the $p \times m$-matrix-valued transfer function

$$\mathbf{Z}(s) \equiv \mathbf{L}^{\mathrm{T}} \left(\mathbf{I} - s\, \mathbf{A}\right)^{-1} \mathbf{R}. \tag{75}$$

The matrix size N in (12) is called the *state-space dimension* of (75). Let $n < N$, and consider $p \times m$-matrix-valued transfer functions

$$\mathbf{Z}_{n+1}(s) \equiv \mathbf{L}_n^{\mathrm{T}} \left(\mathbf{I} - s\, \mathbf{G}_n\right)^{-1} \mathbf{R}_n, \tag{76}$$

where $\mathbf{G}_n \in \mathbb{C}^{(n+1)\times(n+1)}$, $\mathbf{R}_n \in \mathbb{C}^{(n+1)\times m}$, and $\mathbf{L}_n \in \mathbb{C}^{(n+1)\times p}$. Note that (76) is a transfer function of the same form as (75), but with smaller state-space dimension $n+1$, instead of N. A function of the form (76) is said to be an

$(n+1)$*-st matrix-Padé approximant* of $\mathbf{Z}$ (about the expansion point $s_0 = 0$) if the matrices $\mathbf{G}_n$, $\mathbf{R}_n$, and $\mathbf{L}_n$ are such that

$$\mathbf{Z}_{n+1}(s) = \mathbf{Z}(s) + \mathcal{O}\left(s^{q(n)}\right),$$

where $q(n)$ is as large as possible.

It turns out that, for the case of exact deflation, an $(n+1)$-st matrix-Padé approximant can be obtained by a suitable two-sided projection of $\mathbf{Z}$ onto the n-th block Krylov subspaces $\mathcal{K}_n(\mathbf{A}, \mathbf{R})$ and $\mathcal{K}_n(\mathbf{A}^{\mathrm{T}}, \mathbf{L})$. Recall from (67) that these subspaces are spanned by the columns of the matrices $\mathbf{V}_n$ and $\mathbf{W}_n$. In terms of $\mathbf{V}_n$ and $\mathbf{W}_n$, the two-sided projection of (75) is as follows:

$$\mathbf{Z}_{n+1}(s) \equiv \left(\mathbf{V}_n^{\mathrm{T}}\,\mathbf{L}\right)^{\mathrm{T}} \left(\mathbf{W}_n^{\mathrm{T}}\,\mathbf{V}_n - s\,\mathbf{W}_n^{\mathrm{T}}\,\mathbf{A}\,\mathbf{V}_n\right)^{-1} \left(\mathbf{W}_n^{\mathrm{T}}\,\mathbf{R}\right). \tag{77}$$

Using the first relation in (70), as well as (74), we can re-write (77) in the following form:

$$\mathbf{Z}_{n+1}(s) \equiv \left(\boldsymbol{\Delta}_n^{\mathrm{T}}\,\boldsymbol{\eta}_n\right)^{\mathrm{T}} \left(\mathbf{I} - s\,\mathbf{T}_n\right)^{-1}\,\boldsymbol{\rho}_n. \tag{78}$$

Hence $\mathbf{Z}_{n+1}$ is a function of the type (76). Furthermore, in [9, Theorem 1], it is shown that (78) is indeed an $(n+1)$-st matrix-Padé approximant of $\mathbf{Z}$.

8.3 Approximate eigenvalues

In this subsection, we consider the eigenvalue problem,

$$\mathbf{A}\,\mathbf{x} = \tau\,\mathbf{x}, \tag{79}$$

for $\mathbf{A}$. Using the same two-sided projection as in Section 8.2, we can derive from (79) a smaller eigenvalue problem whose eigenvalues can then be used as approximate eigenvalues of $\mathbf{A}$. More precisely, setting $\mathbf{x} = \mathbf{V}_n\,\mathbf{z}$ and multiplying (79) from the left by $\mathbf{W}_n^{\mathrm{T}}$, we get

$$\mathbf{W}_n^{\mathrm{T}}\,\mathbf{A}\,\mathbf{V}_n\,\mathbf{z} = \tau\,\mathbf{W}_n^{\mathrm{T}}\,\mathbf{V}_n\,\mathbf{z}. \tag{80}$$

By (73), the generalized eigenvalue problem (80) is equivalent to the standard eigenvalue problem,

$$\mathbf{T}_n^{\mathrm{proj}}\,\mathbf{z} = \tau\,\mathbf{z},$$

for the matrix $\mathbf{T}_n^{\mathrm{proj}}$ generated from the Lanczos-type method sketched in Section 8.1. The eigenvalues of $\mathbf{T}_n^{\mathrm{proj}}$ are then used as approximate eigenvalues of $\mathbf{A}$. For further details of this approach, we refer the reader to [11].

8.4 Linear systems with multiple right-hand sides

Next, consider systems of linear equations with coefficient matrix $\mathbf{A}$ and multiple, say m, given right-hand sides. Such linear systems can be written in compact matrix form,

$$\mathbf{A}\,\mathbf{X} = \mathbf{B}, \tag{81}$$

where $\mathbf{B} \in \mathbb{C}^{N\times m}$. Let $\mathbf{X}_0 \in \mathbb{C}^{N\times m}$ be any guess for the solution of (81), and let $\mathbf{R}_0 := \mathbf{B} - \mathbf{A}\,\mathbf{X}_0$ be the associated residual matrix. By running the Lanczos-type method sketched in Section 8.1 applied to $\mathbf{A}$, $\mathbf{R} = \mathbf{R}_0$, and any (for example, random) matrix $\mathbf{L} \in \mathbb{C}^{N\times p}$ until pass n, we obtain the matrices $\boldsymbol{\rho}_n^{\mathrm{proj}}$ and $\mathbf{T}_n^{\mathrm{proj}}$, which can be used to generate a Galerkin-type iterate for (81). More precisely, we set

$$\mathbf{X}_{n+1} = \mathbf{X}_0 + \mathbf{V}_n\,\mathbf{Z}_n, \quad \text{where} \quad \mathbf{Z}_n \in \mathbb{C}^{(n+1)\times m}, \tag{82}$$

and require that the free parameter matrix $\mathbf{Z}_n$ in (82) is chosen such that the Galerkin condition

$$\mathbf{W}_n^{\mathrm{T}}\,(\mathbf{B} - \mathbf{A}\,\mathbf{X}_{n+1}) = \mathbf{0} \tag{83}$$

is satisfied. By inserting (82) into (83) and using the definitions of $\boldsymbol{\rho}_n^{\mathrm{proj}}$ and $\mathbf{T}_n^{\mathrm{proj}}$ in (71), it follows that (83) is equivalent to the linear system

$$\mathbf{T}_n^{\mathrm{proj}}\,\mathbf{Z}_n = \boldsymbol{\rho}_n^{\mathrm{proj}}. \tag{84}$$

Provided that $\mathbf{T}_n^{\mathrm{proj}}$ is nonsingular, the solution of (84) defines a unique iterate (82). In the special case that $m = p$ and that no deflation occurs, the resulting iterative method for solving (81) is mathematically equivalent to block-biconjugate gradients [21].

We remark that the condition on the nonsingularity of $\mathbf{T}_n^{\mathrm{proj}}$ can be avoided by replacing (84) by a least-squares problem with a rectangular extension of $\mathbf{T}_n^{\mathrm{proj}}$, which always has full column rank. The resulting iterative method for solving (81) is the block-QMR algorithm [13,20].

8.5 A fast block Hankel solver

The last application is an extension of the fast solver for scalar Hankel matrices in [14] to general block Hankel matrices.

Let $\mathbf{A}_n$ and $\mathbf{B}_n$ the matrices given by (28) and (38). Recall that these matrices contain the potentially nonzero coefficients of the first $n+1$ pairs of polynomials (51) produced by Algorithm 1. By construction, these matrices are upper triangular and they satisfy $\langle \mathbf{\Psi}_n, \mathbf{\Phi}_n \rangle = \mathbf{B}_n^{\mathrm{T}} \mathbf{H}_n^{\mathrm{defl}} \mathbf{A}_n$. Hence, the first relation in (64) is equivalent to the following matrix factorization:

$$\mathbf{B}_n^{\mathrm{T}} \mathbf{H}_n^{\mathrm{defl}} \mathbf{A}_n = \mathbf{\Delta}_n = \operatorname{diag}\left(\mathbf{\Delta}^{(0)}, \mathbf{\Delta}^{(1)}, \ldots, \mathbf{\Delta}^{(\ell)}\right). \tag{85}$$

Note that (85) represents an inverse triangular factorization of $\mathbf{H}_n^{\mathrm{defl}}$. By rewriting the recurrences used to generate the polynomials in Algorithm 1 in terms of the columns of $\mathbf{A}_n$ and $\mathbf{B}_n$, one obtains a fast, i.e., $\mathcal{O}(n^2)$, algorithm for computing the factorization (85) of the deflated block Hankel matrix $\mathbf{H}_n^{\mathrm{defl}}$. Details of this fast block Hankel solver will be given in a future publication.

9 Concluding remarks

We have presented a computational procedure for generating vector-valued polynomials that are formally orthogonal with respect to a matrix-valued bilinear form induced by a general block Hankel matrix $\mathbf{H}$ with arbitrary, not necessarily square blocks. Existing algorithms for this problem require the assumption that $\mathbf{H}$ is strongly regular; unfortunately, this assumption is not satisfied in one of the most important special cases, namely bilinear forms given by a realization. In contrast, our approach can handle the most general case and does not require any assumptions on the block Hankel matrix $\mathbf{H}$.

We have briefly discussed some applications of the proposed computational procedure to problems in linear algebra. There are other potential applications, for example Gauss quadrature for matrix-valued bilinear forms, and these will be reported elsewhere.

References

[1] J.I. Aliaga, D.L. Boley, R.W. Freund, and V. Hernández, A Lanczos-type method for multiple starting vectors, *Math. Comp.* (2000), to appear. Also available online from `http://cm.bell-labs.com/cs/doc/98`.

[2] B. Beckermann, Nonsymmetric difference operators and polynomials being orthogonal with respect to rectangular matrix valued measures, Technical Report ANO–335, Laboratoire d'Analyse Numérique et d'Optimisation, Université des Sciences et Technologies de Lille, 1995.

[3] C. Brezinski, *Padé-Type Approximation and General Orthogonal Polynomials* (Birkhäuser, Basel, 1980).

[4] A. Bultheel and M. Van Barel, *Linear Algebra, Rational Approximation and Orthogonal Polynomials* (North-Holland, Amsterdam, 1997).

[5] A.J. Durán and W. Van Assche, Orthogonal matrix polynomials and higher-order recurrence relations, *Linear Algebra Appl.* **219** (1995) 261–280.

[6] P. Feldmann and R.W. Freund, Efficient linear circuit analysis by Padé approximation via the Lanczos process, *IEEE Trans. Computer-Aided Design* **14** (1995), 639–649.

[7] P. Feldmann and R.W. Freund, Reduced-order modeling of large linear subcircuits via a block Lanczos algorithm, in: Proc. 32nd Design Automation Conference (ACM, New York, 1995) 474–479.

[8] R.W. Freund, The look-ahead Lanczos process for nonsymmetric matrices and its applications, in: J.D. Brown, M.T. Chu, D.C. Ellison, and R.J. Plemmons, eds., *Proceedings of the Cornelius Lanczos International Centenary Conference* (SIAM, Philadelphia, 1994) 33–47.

[9] R.W. Freund, *Computation of matrix Padé approximations of transfer functions via a Lanczos-type process*, in: C.K. Chui and L.L. Schumaker, eds., *Approximation Theory VIII, Vol. 1: Approximation and Interpolation* (World Scientific Publishing Co., Inc., Singapore, 1995) 215–222.

[10] R.W. Freund, Reduced-order modeling techniques based on Krylov subspaces and their use in circuit simulation, in: B.N. Datta, ed., *Applied and Computational Control, Signals, and Circuits, Vol. 1* (Birkhäuser, Boston, 1999) 435–498.

[11] R.W. Freund, *Band Lanczos method*, in: Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: a Practical Guide* (SIAM, Philadelphia, 2000) to appear. Also available online from `http://cm.bell-labs.com/cs/doc/99`.

[12] R.W. Freund, M.H. Gutknecht, and N.M. Nachtigal, An implementation of the look-ahead Lanczos algorithm for non-Hermitian matrices, *SIAM J. Sci. Comput.* **14** (1993) 137–158.

[13] R.W. Freund and M. Malhotra, A block QMR algorithm for non-Hermitian linear systems with multiple right-hand sides, *Linear Algebra Appl.* **254** (1997) 119–157.

[14] R.W. Freund and H. Zha, A look-ahead algorithm for the solution of general Hankel systems, *Numer. Math.* **64** (1993) 295–321.

[15] W.B. Gragg, Matrix interpretations and applications of the continued fraction algorithm, *Rocky Mountain J. Math.* **4** (1974) 213–225.

[16] W.B. Gragg and A. Lindquist, On the partial realization problem, *Linear Algebra Appl.* **50** (1983) 277–319.

[17] M.H. Gutknecht, A completed theory of the unsymmetric Lanczos process and related algorithms, part I, *SIAM J. Matrix Anal. Appl.* **13** (1992) 594–639.

[18] R.E. Kalman, P.L. Falb, and M.A. Arbib, *Topics in Mathematical System Theory* (McGraw-Hill, New York, 1969).

[19] C. Lanczos, An iteration method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Nat. Bur. Standards* **45** (1950) 255–282.

[20] M. Malhotra, R.W. Freund, and P.M. Pinsky, Iterative solution of multiple radiation and scattering problems in structural acoustics using a block quasi-minimal residual algorithm, *Comput. Methods Appl. Mech. Engrg.* **146** (1997) 173–196.

[21] D.P.O'Leary, The block conjugate gradient algorithm and related methods, *Linear Algebra Appl.* **29** (1980) 293–322.

[22] B.N. Parlett, D.R. Taylor, and Z.A. Liu, A look-ahead Lanczos algorithm for unsymmetric matrices, *Math. Comp.* **44** (1985) 105–124.

[23] J. Rissanen, Basis of invariants and canonical forms for linear dynamical systems, *Automatica J. IFAC* **10** (1974) 175–182.

[24] A. Salam, Formal vector orthogonal polynomials, *Adv. Comput. Math.* **8** (1998) 267–289.

[25] A. Sinap and W. Van Assche, Orthogonal matrix polynomials and applications, *J. Comput. Appl. Math.* **66** (1996) 27–52.

[26] E.D. Sontag, *Mathematical Control Theory* (Springer-Verlag, New York, 1990).

[27] V.N. Sorokin and J. Van Iseghem, Algebraic aspects of matrix orthogonality for vector polynomials, *J. Approx. Theory* **90** (1997) 97–116.

[28] D.R. Taylor, *Analysis of the look ahead Lanczos algorithm*, Ph.D. Thesis, Department of Mathematics, University of California, Berkeley, 1982.