Available online at www.sciencedirect.com
ScienceDirect
Linear Algebra and its Applications 423 (2007) 183–208
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

A generalized isometric Arnoldi algorithm

Michael Stewart

Department of Mathematics and Statistics, Georgia State University, Atlanta, GA 30303, United States

Received 30 September 2005; accepted 8 December 2006
Available online 19 January 2007
Submitted by V. Mehrmann

Abstract

This paper describes a generalization of the isometric Arnoldi algorithm and shows that it can be interpreted as a structured form of modified Gram–Schmidt. Given an isometry A, the algorithm efficiently orthogonalizes the columns of a sequence of matrices M_j for $j \geqslant 0$ (with $M_{-1} = 0$) for which the columns of $M_j - AM_{j-1}$ are in a fixed finite dimensional subspace for each $j \geqslant 0$. The dimension of the subspace is analogous to displacement rank in the generalized Schur algorithm. The algorithm is described in terms of projections and inner products. This is in contrast to orthogonalization methods based on the generalized Schur algorithm, for which Cholesky factorization is central to the computation. Numerical experiments suggest that, relative to a generalized Schur algorithm, the new algorithm improves the numerical orthogonality of the computed orthonormal sequence.

Keywords: Isometric Arnoldi algorithm; Orthogonalization; Structured matrices; Fast algorithms

1. Introduction

We assume throughout this paper that A is an isometry acting on a complex Hilbert space $\mathscr{H}$ with inner product $\langle \mathbf{x}, \mathbf{y} \rangle$ and norm $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. Given a vector $\mathbf{x} \in \mathscr{H}$ with $\|\mathbf{x}\| = 1$ the isometric Arnoldi algorithm [3,4] is an efficient procedure for orthogonalizing the Krylov sequence

$$\mathbf{x},\ A\mathbf{x},\ A^2\mathbf{x},\ \ldots.$$

It can be viewed as a generalization of the Szegö recurrence [9] for the orthogonalization of the polynomial power basis

E-mail address: mastewart@gsu.edu

0024-3795/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2006.12.021

$$1, z, z^2, z^3, \dots$$

with respect to an inner product on the unit circle or as a generalization of the lattice algorithm [1] for the orthogonalization of the columns of an $m \times l$ windowed Toeplitz matrix

$$T = [\mathbf{t} \quad Z\mathbf{t} \quad Z^2\mathbf{t} \quad \cdots \quad Z^{l-1}\mathbf{t}],$$

where Z is the circulant shift matrix and

$$\mathbf{t}^{\mathrm{T}} = [t_0 \quad t_1 \quad \cdots \quad t_{m-l} \quad \mathbf{0}_{l-1}^{\mathrm{T}}].$$

The orthogonalized sequence gives a basis with respect to which A reduces to a product of plane rotations. In the matrix case this corresponds to a unitary similarity that reduces A to unitary Hessenberg form, providing an efficient means to solve the unitary eigenvalue problem [2]. The procedure also provides efficient methods for solving systems involving shifts of unitary matrices, i.e. systems of the form $(\alpha A + \beta I)\mathbf{x} = \mathbf{b}$ [6].

The goal of this paper is to modify the isometric Arnoldi algorithm so as to orthogonalize a generalization of the class of Krylov sequences. We generalize in two ways. First, instead of sequences of vectors, we consider sequences of matrices of the form

$$M_j = [\mathbf{m}_{j,1} \quad \mathbf{m}_{j,2} \quad \cdots \quad \mathbf{m}_{j,p}] \tag{1}$$

for $j \geqslant 0$ and where $\mathbf{m}_{j,k} \in \mathscr{H}$. Throughout this paper we assume that $M_{-1} = 0$. Second, instead of requiring that $M_j = AM_{j-1}$, we require that the columns of $M_j - AM_{j-1}$ lie in some finite dimensional subspace $\mathscr{M} \subseteq \mathscr{H}$.

We make a few comments about notation. It is convenient to interpret a vector $\mathbf{x} \in \mathscr{H}$ as an operator mapping a complex number a to the product $a\mathbf{x} \in \mathscr{H}$. The vector $\mathbf{x}$ then has an adjoint $\mathbf{x}^* : \mathscr{H} \to \mathbb{C}$ defined by $\mathbf{x}^*\mathbf{y} = \langle \mathbf{y}, \mathbf{x} \rangle$. We similarly interpret a matrix M with p columns that are each in $\mathscr{H}$ as an operator from $\mathbb{C}^p$ to $\mathscr{H}$ acting through matrix vector multiplication in the obvious way. If $M : \mathbb{C}^p \to \mathscr{H}$ is a matrix with columns $\mathbf{m}_k \in \mathscr{H}$ the adjoint $M^* : \mathscr{H} \to \mathbb{C}^p$ is a matrix with rows $\mathbf{m}_k^*$. Given an arbitrary operator B we use the notation $\mathscr{P}(B)$ to represent the orthogonal projector onto $\mathrm{Im}(B)$.

Let

$$M^{(\infty)} = [M_0 \quad M_1 \quad M_2 \quad \cdots]. \tag{2}$$

Orthogonalizing the columns of $M^{(\infty)}$ against the columns of each M_j for $0 \leqslant j \leqslant k-1$ results in a matrix of the form

$$\begin{bmatrix} 0 & \cdots & 0 & M_k^{(k)} & M_{k+1}^{(k)} & \cdots \end{bmatrix}.$$

The sequence $M_j^{(k)}$ is M_j projected onto the orthogonal complement of the span of the columns of $M_0, M_1, \dots, M_{k-1}$. The k leading zero blocks are the columns of $M_0, M_1, \dots, M_{k-1}$ projected onto the orthogonal complement of their own span. The sequence $M_j^{(j)}$ is the orthogonalization of the sequence M_j in the sense that $\mathrm{Im}(M_j^{(j)}) \perp \mathrm{Im}(M_k^{(k)})$ for $j \neq k$ and

$$\mathrm{Im}[M_0 \quad M_1 \quad \cdots \quad M_j] = \mathrm{Im}\begin{bmatrix} M_0^{(0)} & M_1^{(1)} & \cdots & M_j^{(j)} \end{bmatrix}$$

for $j \geqslant 0$. The explicit computation of each of the sequences $M_j^{(k)}$ for $k = 0, 1, 2, \dots$ can be interpreted as a block form of modified Gram–Schmidt. The generalized isometric Arnoldi algorithm can be interpreted as a structured form of the above unstructured orthogonalization procedure. The algorithm exploits the fact that if M_j is Krylov-like then the partially orthogonalized sequences $M_j^{(k)}$ are also Krylov-like.

In order to describe the Krylov-like structure of the sequences $M_j^{(k)}$ we need a more detailed description of Krylov-like structure. For a variety of reasons it is convenient to work with projectors. If P_0 is the orthogonal projector onto $\mathscr{M}$ then a Krylov-like sequence could be defined as a sequence satisfying a relation of the form

$$M_j - AM_{j-1} = P_0 M_j - P_0 A M_{j-1} \tag{3}$$

for $j \geqslant 0$. Unfortunately, if we start with a sequence M_j satisfying a relation of the form (3) then the partially orthogonalized sequence $M_j^{(k)}$ satisfies a relation of the form (3) only if P_0 is replaced by a projector of greater rank. The following definition is based on a relation that is preserved during orthogonalization with no increase in the ranks of the projectors.

Definition 1. Any projectors P_0 and Q_0 satisfying

$$P_0[(I - Q_0)A]^j P_0 = 0 \tag{4}$$

for $j \geqslant 1$ are referred to as *displacement projectors*. A sequence $M_j : \mathbb{C}^p \to \mathscr{H}$ with $M_{-1} = 0$ is *Krylov-like* with displacement projectors P_0 and Q_0 if

$$M_j - AM_{j-1} = P_0 M_j - Q_0 A M_{j-1} \tag{5}$$

for $j \geqslant 0$.

Example 1. An ordinary Krylov sequence $\mathbf{m}_j = A^j \mathbf{x}$ for $\|\mathbf{x}\| = 1$ where $\mathbf{m}_{-1} = 0$ satisfies

$$\mathbf{m}_j - A\mathbf{m}_{j-1} = \delta_j \mathbf{x} \in \operatorname{Span}(\mathbf{x})$$

for $j \geqslant 0$ and where $\delta_j = 1$ for $j = 0$ and $\delta_j = 0$ for $j \neq 0$. Thus

$$\mathbf{m}_j - A\mathbf{m}_{j-1} = P_0 \mathbf{m}_j - Q_0 A \mathbf{m}_{j-1}$$

for $j \geqslant 0$ with $P_0 = Q_0 = \mathbf{x}\mathbf{x}^*$.

Example 2. An $m \times n$ real Toeplitz matrix

$$T = [\mathbf{t}_0 \quad \mathbf{t}_1 \quad \cdots \quad \mathbf{t}_{n-1}] \tag{6}$$

has columns

$$\mathbf{t}_j = [t_{-j} \quad t_{-j+1} \quad \cdots \quad t_{m-1-j}]^{\mathrm{T}}$$

that satisfy

$$\mathbf{t}_j - Z\mathbf{t}_{j-1} = \delta_j \mathbf{t}_0 + (t_{-j} - t_{m-j})\mathbf{e}_1 \in \operatorname{Span}(\mathbf{t}_0, \mathbf{e}_1)$$

for $j \geqslant 0$ where Z is the circulant downshift matrix and $\mathbf{t}_{-1} = 0$. Thus

$$\mathbf{t}_j - Z\mathbf{t}_{j-1} = P_0 \mathbf{t}_j - Q_0 Z \mathbf{t}_{j-1},$$

where

$$P_0 = Q_0 = \mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} + \frac{1}{\|\mathbf{t}_0 - t_0 \mathbf{e}_1\|^2} (\mathbf{t}_0 - t_0 \mathbf{e}_1)(\mathbf{t}_0 - t_0 \mathbf{e}_1)^{\mathrm{T}}. \tag{7}$$

In the above examples (4) is satisfied for the simple reason that $P_0 = Q_0$. As orthogonalization proceeds we generate projectors P_k and Q_k that are displacement projectors for $M_j^{(k)}$. In general P_k and Q_k are not equal. Nevertheless they satisfy

$$P_k[(I - Q_k)A]^j P_k = 0 \tag{8}$$

for $j \geqslant 1$. This relation is of fundamental importance to the proof that the algorithm correctly orthogonalizes a Krylov-like sequence.

An outline of this paper is as follows. In §2 we derive a form of the isometric Arnoldi algorithm that can also be applied to a Krylov-like sequence. The derivation assumes that the sequence is an ordinary Krylov sequence. In §3 we describe some simple properties of Krylov-like sequences, including a connection with Toeplitz-like matrices. We also show that Krylov-like structure is preserved by orthogonalization. In §4 we prove that the generalized isometric Arnoldi algorithm orthogonalizes a Krylov-like sequence M_j. In §5 we factor the projectors and describe the algorithm in terms of the bases for the images of the projectors. In §6 we show how to extend the procedure with recurrences to compute the factor R in a QR factorization. The recurrences reveal the connection between the generalized isometric Arnoldi algorithm and the generalized Schur algorithm. In §7 we present some numerical experiments. Finally in §8 we comment on some open problems and ongoing research.

2. A general form of the isometric Arnoldi algorithm

We now put the isometric Arnoldi algorithm in a form that is applicable to general Krylov-like sequences. The initial derivation assumes that the sequence to be orthogonalized is an ordinary Krylov sequence. It is only in §4 that we prove that the algorithm also correctly orthogonalizes Krylov-like sequences. In order to avoid worrying about the dimension of various subspaces and the choice of particular bases for the subspaces it is convenient to state the general form of the algorithm in terms of projectors. The projectors P_k and Q_k described in this section are in fact displacement projectors for a Krylov-like sequence, although the proof of this fact is also put off to §4.

Given a Krylov sequence $\mathbf{m}_j = A^j\mathbf{x}$ where $\|\mathbf{x}\| = 1$ and $A^*A = I$, the isometric Arnoldi algorithm of [3,4] is as follows.

Algorithm 1. Isometric Arnoldi

$\mathbf{x}_0 = \mathbf{x}, \mathbf{y}_0 = \mathbf{x}, k = 0$

$\gamma_0 = -\langle A\mathbf{x}, \mathbf{x}\rangle$

While $|\gamma_k| \neq 1$

$\quad \mathbf{x}_{k+1} = (A\mathbf{x}_k + \gamma_k\mathbf{y}_k)/\sqrt{1 - |\gamma_k|^2}$

$\quad \mathbf{y}_{k+1} = (\overline{\gamma_k}A\mathbf{x}_k + \mathbf{y}_k)/\sqrt{1 - |\gamma_k|^2}$

$\quad \gamma_{k+1} = -\langle A\mathbf{x}_{k+1}, \mathbf{y}_{k+1}\rangle$

$\quad k = k + 1$

End While

It can be shown that the quantity γ_k satisfies $|\gamma_k| \leqslant 1$. If $|\gamma_k| < 1$ for $0 \leqslant k \leqslant n - 1$ then Algorithm 1 generates an orthonormal sequence of vectors $\mathbf{x}_k$ for which

$$\mathrm{Span}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_k) = \mathrm{Span}(\mathbf{x}, A\mathbf{x}, \ldots, A^k\mathbf{x}) \tag{9}$$

for each $0 \leqslant k \leqslant n$. If $|\gamma_n| = 1$ then

$$\mathrm{Span}(\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n) = \mathrm{Span}(\mathbf{x}, A\mathbf{x}, \ldots, A^n\mathbf{x})$$

is an invariant subspace of A.

In describing the isometric Arnoldi algorithm, we differ from [3,4] in that we enforce the normalization $\|\mathbf{x}_k\| = \|\mathbf{y}_k\| = 1$. Starting with $\|\mathbf{x}_0\| = \|\mathbf{y}_0\| = 1$, it is easily verified that if $\|\mathbf{x}_k\| = \|\mathbf{y}_k\| = 1$ then

$$\|A\mathbf{x}_k + \gamma_k \mathbf{y}_k\|^2 = \|\overline{\gamma_k} A\mathbf{x}_k + \mathbf{y}_k\|^2 = 1 - |\gamma_k|^2$$

so that $\|\mathbf{x}_{k+1}\| = \|\mathbf{y}_{k+1}\| = 1$.

Let P_k be the orthogonal projector onto $\mathrm{Span}(\mathbf{x}_k)$ and let Q_k be the orthogonal projector onto $\mathrm{Span}(\mathbf{y}_k)$. Then $\|\mathbf{x}_k\| = \|\mathbf{y}_k\| = 1$ implies

$$\mathbf{x}_k^*\mathbf{x}_k = \mathbf{y}_k^*\mathbf{y}_k = 1, \quad P_k = \mathbf{x}_k\mathbf{x}_k^*, \quad \text{and} \quad Q_k = \mathbf{y}_k\mathbf{y}_k^*.$$

It follows that

$$\|(I - Q_k)A\mathbf{x}_k\|^2 = 1 - \|Q_k A\mathbf{x}_k\|^2 = 1 - \|\mathbf{y}_k\mathbf{y}_k^* A\mathbf{x}_k\|^2 = 1 - |\gamma_k|^2$$

so that

$$\mathbf{x}_{k+1} = \frac{1}{\sqrt{1 - |\gamma_k|^2}}(A\mathbf{x}_k - \mathbf{y}_k(\mathbf{y}_k^* A\mathbf{x}_k)) = \frac{1}{\|(I - Q_k)A\mathbf{x}_k\|}(I - Q_k)A\mathbf{x}_k$$

or

$$P_{k+1} = \mathbf{x}_{k+1}\mathbf{x}_{k+1}^* = \frac{1}{\|(I - Q_k)A\mathbf{x}_k\|^2}(I - Q_k)A\mathbf{x}_k\mathbf{x}_k^* A^*(I - Q_k). \tag{10}$$

Thus

$$P_{k+1} = \mathscr{P}((I - Q_k)AP_k).$$

Define $V_k = \mathscr{P}((I - Q_k)AP_k)$ and suppose for the moment that $U_k = P_k$. Then we can write $P_{k+1} = V_k$ as

$$P_{k+1} = P_k - U_k + V_k. \tag{11}$$

When considering the case of a general Krylov-like sequence we choose U_k to be the projector onto a particular subspace of $\mathrm{Im}(P_k)$. Hence in general we do not have $P_k = U_k$. Nevertheless, with an appropriate choice of U_k, (11) is applicable to the orthogonalization of general Krylov-like sequences.

We now consider the computation of Q_{k+1}. Since

$$[\mathbf{x}_{k+1} \quad \mathbf{y}_{k+1}] = \frac{1}{\sqrt{1 - |\gamma_k|^2}}[A\mathbf{x}_k \quad \mathbf{y}_k]\begin{bmatrix} 1 & \overline{\gamma_k} \\ \gamma_k & 1 \end{bmatrix}$$

and

$$\left(\frac{1}{\sqrt{1 - |\gamma_k|^2}}\begin{bmatrix} 1 & \overline{\gamma_k} \\ \gamma_k & 1 \end{bmatrix}\right)\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\left(\frac{1}{\sqrt{1 - |\gamma_k|^2}}\begin{bmatrix} 1 & \overline{\gamma_k} \\ \gamma_k & 1 \end{bmatrix}\right)^{\mathrm{H}} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

we have

$$[\mathbf{x}_{k+1} \quad \mathbf{y}_{k+1}]\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \mathbf{x}_{k+1}^* \\ \mathbf{y}_{k+1}^* \end{bmatrix} = \begin{bmatrix} A\mathbf{x}_k & \mathbf{y}_k \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} \mathbf{x}_k^* A^* \\ \mathbf{y}_k^* \end{bmatrix}$$

or

$$\mathbf{x}_{k+1}\mathbf{x}_{k+1}^* - \mathbf{y}_{k+1}\mathbf{y}_{k+1}^* = A\mathbf{x}_k\mathbf{x}_k^* A^* - \mathbf{y}_k\mathbf{y}_k^*.$$

Thus $Q_{k+1} = Q_k - AP_kA^* + P_{k+1}$. Using the relations $P_k = U_k$ and $P_{k+1} = V_k$ we can write

$$Q_{k+1} = Q_k - AU_kA^* + V_k. \tag{12}$$

As with the relation for P_{k+1}, (12) is applicable to a Krylov-like sequence if U_k is chosen to be the projector onto a suitable subspace of $\mathrm{Im}(P_k)$.

With regard to termination of the algorithm, we note that

$$\|Q_kAU_k\| = \|Q_kAP_k\| = \|\mathbf{y}_k\mathbf{y}_k^*A\mathbf{x}_k\mathbf{x}_k^*\| = |\gamma_k|$$

so that terminating when $|\gamma_k| = 1$ is the same as terminating when $\|Q_kAU_k\| = 1$.

Finally, since the relation $U_k = P_k$ is suitable only for the orthogonalization of an ordinary Krylov sequence and does not apply in the case of a Krylov-like sequence, we introduce a more generally applicable formula. If $\|Q_kAU_k\| < 1$ then $|\gamma_k| < 1$ so that $\mathbf{m}_0, \mathbf{m}_1, \ldots, \mathbf{m}_{k+1}$ are linearly independent. The vector $\mathbf{x}_{k+1}$ is $\mathbf{m}_{k+1}$ orthogonalized against $\mathrm{Span}(\mathbf{m}_0, \mathbf{m}_1, \ldots, \mathbf{m}_k)$ and then normalized. Linear independence thus ensures that $\mathbf{x}_{k+1}^*\mathbf{m}_{k+1} \neq 0$ so that

$$\mathscr{P}(P_{k+1}\mathbf{m}_{k+1}) = \mathscr{P}(\mathbf{x}_{k+1}\mathbf{x}_{k+1}^*\mathbf{m}_{k+1}) = P_{k+1} = U_{k+1}.$$

Combining this relation for U_k and the definition of V_k with the recurrences (11) and (12) gives the following form of the isometric Arnoldi algorithm.

Algorithm 2. Isometric Arnoldi in terms of projectors

$P_0 = \mathbf{x}\mathbf{x}^*$, $Q_0 = P_0$, $k = 0$
$U_0 = P_0$, $V_0 = \mathscr{P}((I - Q_0)AP_0)$
While $\|Q_kAU_k\| < 1$

$$P_{k+1} = P_k - U_k + V_k$$
$$Q_{k+1} = Q_k - AU_kA^* + V_k$$
$$U_{k+1} = \mathscr{P}(P_{k+1}\mathbf{m}_{k+1})$$
$$V_{k+1} = \mathscr{P}((I - Q_{k+1})AU_{k+1})$$
$$k = k + 1$$

End While

Recall that in the context of an ordinary Krylov sequence, $U_k = P_k = \mathbf{x}_k\mathbf{x}_k^*$ so that $\mathbf{x}_j^*\mathbf{x}_k = 0$ for $j \neq k$ implies that

$$U^{(n)} = U_0 + U_1 + \cdots + U_n$$

is the orthogonal projector onto $\mathrm{Span}(\mathbf{x}, A\mathbf{x}, \ldots, A^n\mathbf{x}) = \mathrm{Span}(\mathbf{x}_0, \ldots, \mathbf{x}_n)$.

We have claimed that Algorithm 2 can also be used to orthogonalize the broader class of Krylov-like sequences. The only modifications required to apply Algorithm 2 to the more general problem are replacing $U_{k+1} = \mathscr{P}(P_{k+1}\mathbf{m}_{k+1})$ with $U_{k+1} = \mathscr{P}(P_{k+1}M_{k+1})$ and setting P_0 to be the orthogonal projector onto the subspace $\mathscr{M}$. Although working with projectors represented as dense matrices is clearly inefficient, the images of the projectors can be represented by orthonormal bases, in which case the relations for P_{k+1} and Q_{k+1} can be implemented as two updating/downdating problems. Doing so reverses the steps of the preceding derivation, leading back to Algorithm 1.

3. Krylov-like sequences

Clearly the knowledge that a sequence M_j satisfies a relation of the form (5) for given P_0, Q_0, and A does not suffice to uniquely determine M_j. The additional information that is required to determine M_j is its projection onto the image of P_0. In fact the recurrence

$$M_j = B_j + (I - Q_0)AM_{j-1}, \quad M_{-1} = 0 \tag{13}$$

is a bijection mapping sequences of matrices B_j, $j \geqslant 0$ with columns in the image of P_0 (i.e. with $P_0B_j = B_j$) onto the set of all Krylov-like sequences satisfying (5). This bijection guarantees that $P_0M_j = B_j$ so that a Krylov-like sequence M_j is uniquely determined by its projection on the image of P_0.

Theorem 1. *Let Q_0 and P_0 be orthogonal projectors*:

1. *If* (4) *holds then the mapping* (13) *is a bijection from the set of sequences B_j satisfying $P_0B_j = B_j$ to the set of Krylov-like sequences M_j satisfying $M_j - AM_{j-1} = P_0M_j - Q_0AM_{j-1}$. In addition we have $P_0M_j = B_j$.*
2. *If for every B_j satisfying $P_0B_j = B_j$ there is a sequence M_j such that $M_j - AM_{j-1} = P_0M_j - Q_0AM_{j-1}$ and $P_0M_j = B_j$ then* (4) *holds.*

Proof. For M_j computed from (13) with B_j satisfying $P_0B_j = B_j$ we have

$$M_j = B_j + \sum_{m=1}^{j} [(I - Q_0)A]^m P_0 B_{j-m}$$

from which it follows that if (4) holds then $P_0M_j = B_j$. The relation (13) can then be rewritten

$$M_j = P_0M_j + (I - Q_0)AM_{j-1}.$$

Thus if (4) holds then (13) maps sequences B_j satisfying $P_0B_j = B_j$ into the set of Krylov-like sequences satisfying $M_j - AM_{j-1} = P_0M_j - Q_0AM_{j-1}$.

That (13) maps onto the set of all Krylov-like sequences follows from the fact that for an arbitrary sequence M_j satisfying $M_j - AM_{j-1} = P_0M_j - Q_0AM_{j-1}$, we can choose $B_j = P_0M_j$, in which case the recurrence (13) generates M_j. The map is one-to-one since if B_j and $\hat{B}_j$ map to the same sequence M_j then

$$M_j = B_j - (I - Q_0)AM_{j-1}, \quad \text{and} \quad M_j = \hat{B}_j - (I - Q_0)AM_{j-1}$$

immediately imply $B_j = \hat{B}_j$.

To prove the second part of the theorem, we set $B_j = 0$ for $j \neq 0$ and let B_0 be an arbitrary matrix with columns in $\mathrm{Im}(P_0)$. If there is a sequence M_j satisfying (5) and $P_0M_j = B_j$ then

$$M_j = B_j + \sum_{m=1}^{j} [(I - Q_0)A]^m B_{j-m} = [(I - Q_0)A]^j B_0.$$

Since $P_0M_j = B_j = 0$ for $j \geqslant 1$ we then have

$$P_0[(I - Q_0)A]^j B_0 = 0$$

for all $j \geqslant 1$. Since B_0 is an arbitrary matrix with columns in $\mathrm{Im}(P_0)$ this implies (4). $\square$

Krylov-like sequences are connected in a simple way with Toeplitz-like matrices. In particular, the relation (5) is closely related to the displacement equation [8] of the block Toeplitz-like matrix with blocks given by $T_{i,j} = M_i^* M_j$.

Theorem 2. *Suppose that the sequence M_j satisfies (5) with $M_{-1} = 0$. If*

$$T_{i,j} : \mathbb{C}^p \to \mathbb{C}^p = M_i^* M_j$$

then

$$T_{i,j} - T_{i-1,j-1} = M_i^* P_0 M_j - M_{i-1}^* A^* Q_0 A M_{j-1}.$$

Proof. Multiplying both sides of

$$M_j = P_0 M_j + (I - Q_0) A M_{j-1}$$

by P_0 gives $P_0(I - Q_0)AM_{j-1} = 0$. Thus M_j can be represented as the sum of two components: its own projection on $\mathrm{Im}(P_0)$ and a component that is orthogonal to both $\mathrm{Im}(P_0)$ and $\mathrm{Im}(Q_0)$.

Multiplying (5) by M_i^* gives

$$M_i^* M_j - M_i^* A M_{j-1} = M_i^* P_0 M_j - M_i^* Q_0 A M_{j-1}.$$

Multiplying (5) by A^* gives

$$A^* M_i = M_{i-1} + A^* P_0 M_i - A^* Q_0 A M_{i-1}$$

so that

$$M_i^* M_j - (M_{i-1} + A^* P_0 M_i - A^* Q_0 A M_{i-1})^* M_{j-1} = M_i^* P_0 M_j - M_i^* Q_0 A M_{j-1}$$

or

$$M_i^* M_j - M_{i-1}^* M_{j-1} = M_i^* P_0 M_j - M_{i-1}^* A^* Q_0 A M_{j-1} + M_i^* (P_0 - Q_0) A M_{j-1}.$$

Since M_i can be represented as a $P_0 M_i$ plus a component orthogonal to both $\mathrm{Im}(Q_0)$ and $\mathrm{Im}(P_0)$ we have

$$M_i^*(P_0 - Q_0) A M_{j-1} = M_i^* P_0 (P_0 - Q_0) A M_{j-1} = M_i^* P_0 (I - Q_0) A M_{j-1} = 0. \qquad \square$$

Example 3. For a Krylov sequence $\mathbf{m}_j = A^j \mathbf{x}$ with $\|\mathbf{x}\| = 1$ and $\mathbf{m}_{-1} = 0$ we have $P_0 = Q_0 = \mathbf{x}\mathbf{x}^*$. If $t_{i,j} = \mathbf{m}_i^* \mathbf{m}_j$ then

$$t_{i,j} - t_{i-1,j-1} = \mathbf{m}_i^* (\mathbf{x}\mathbf{x}^*) \mathbf{m}_j - \mathbf{m}_{i-1}^* A^* (\mathbf{x}\mathbf{x}^*) A \mathbf{m}_{j-1}.$$

Let Z_0 be the $n \times n$ shift matrix $[Z_0]_{ij} = 1$ for $i = j + 1$ and $[Z_0]_{ij} = 0$ otherwise. If

$$K = [\mathbf{m}_0 \quad \mathbf{m}_1 \quad \cdots \quad \mathbf{m}_{n-1}]$$

then $T = K^* K$ satisfies

$$T - Z_0 T Z_0^{\mathrm{T}} = K^* \mathbf{x}\mathbf{x}^* K - Z_0 K^* A^* \mathbf{x}\mathbf{x}^* A K Z_0^{\mathrm{T}}.$$

Thus T is a displacement rank 2 Toeplitz-like matrix. It is well known and trivial to verify that T is in fact Toeplitz.

Example 4. If $S = T^{\mathrm{T}} T$ where T is the real Toeplitz matrix (6) then

$$\begin{aligned} s_{i,j} - s_{i-1,j-1} = {}& \frac{1}{\|\mathbf{t}_0 - t_0 \mathbf{e}_1\|_2^2} (\mathbf{t}_i^{\mathrm{T}} (\mathbf{t}_0 - t_0 \mathbf{e}_1)(\mathbf{t}_0 - t_0 \mathbf{e}_1)^{\mathrm{T}} \mathbf{t}_j - \mathbf{t}_{i-1}^{\mathrm{T}} Z^{\mathrm{T}} (\mathbf{t}_0 - t_0 \mathbf{e}_1) \\ & \times (\mathbf{t}_0 - t_0 \mathbf{e}_1)^{\mathrm{T}} Z \mathbf{t}_{j-1}) + \mathbf{t}_i^{\mathrm{T}} \mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} \mathbf{t}_j - \mathbf{t}_{i-1}^{\mathrm{T}} Z^{\mathrm{T}} \mathbf{e}_1 \mathbf{e}_1^{\mathrm{T}} Z \mathbf{t}_{j-1}. \end{aligned}$$

Equivalently

$$S - Z_0 S Z_0^{\mathrm{T}} = \frac{1}{\|\mathbf{t}_0 - t_0\mathbf{e}_1\|_2^2}(T^{\mathrm{T}}(\mathbf{t}_0 - t_0\mathbf{e}_1)(\mathbf{t}_0 - t_0\mathbf{e}_1)^{\mathrm{T}}T - Z_0T^{\mathrm{T}}Z^{\mathrm{T}}(\mathbf{t}_0 - t_0\mathbf{e}_1) \times (\mathbf{t}_0 - t_0\mathbf{e}_1)^{\mathrm{T}}ZTZ_0^{\mathrm{T}}) + T^{\mathrm{T}}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}T - Z_0T^{\mathrm{T}}Z^{\mathrm{T}}\mathbf{e}_1\mathbf{e}_1^{\mathrm{T}}ZTZ_0^{\mathrm{T}}.$$

Thus S is a displacement rank 4 Toeplitz-like matrix.

Just as Krylov-like sequences are closely related to Toeplitz-like matrices, the generalized isometric Arnoldi algorithm is closely related to the generalized Schur algorithm [8]. However instead of exploiting the fact that Toeplitz-like structure is preserved by Schur complementation, we exploit the fact that Krylov-like structure is preserved by orthogonalization. Given a Krylov-like sequence $M_j^{(k)}$, $j \geqslant 0$ with displacement projectors P_k and Q_k and a projector U_k with $\mathrm{Im}(U_k) \subseteq \mathrm{Im}(P_k)$, the sequence

$$M_j^{(k+1)} = (I - U_k)M_j^{(k)}$$

for $j \geqslant 0$ is a Krylov-like sequence with displacement projectors P_{k+1} and Q_{k+1} with ranks less than or equal to the ranks of P_k and Q_k. The following theorem justifies these claims, with the notable exception that we put off the proof that P_{k+1} and Q_{k+1} satisfy (8).

Theorem 3. *Suppose that a Krylov-like sequence $M_j^{(k)}$ satisfies*

$$M_j^{(k)} - AM_{j-1}^{(k)} = P_kM_j^{(k)} - Q_kAM_{j-1}^{(k)}$$

for each $j \geqslant 0$ and for displacement projectors P_k and Q_k (i.e. for projectors satisfying (8)). Let U_k be the orthogonal projector for an arbitrary subspace of $\mathrm{Im}(P_k)$. *Let*

$$V_k = \mathscr{P}((I - Q_k)AU_k),$$
$$P_{k+1} = P_k - U_k + V_k, \quad \text{and} \quad Q_{k+1} = Q_k - AU_kA^* + V_k.$$

Then P_{k+1} and Q_{k+1} are orthogonal projectors with ranks less than or equal to those of P_k and Q_k respectively. If

$$M_j^{(k+1)} = (I - U_k)M_j^{(k)}$$

is the sequence $M_j^{(k)}$ orthogonalized against $\mathrm{Im}(U_k)$ *then*

$$M_j^{(k+1)} - AM_{j-1}^{(k+1)} = P_{k+1}M_j^{(k+1)} - Q_{k+1}AM_{j-1}^{(k+1)}$$

for $j \geqslant 0$.

Proof. It follows from (8) and the fact that $\mathrm{Im}(U_k) \subseteq \mathrm{Im}(P_k)$ that

$$P_kV_k = P_k\mathscr{P}((I - Q_k)AU_k) = P_k\mathscr{P}((I - Q_k)AP_kU_k) = 0.$$

This also implies $U_kV_k = 0$. It is obvious from the definition of V_k that $Q_kV_k = 0$. These observations imply that $P_k + V_k$ and $Q_k + V_k$ are orthogonal projectors with ranks equal to $\mathrm{rank}(P_k) + \mathrm{rank}(V_k)$ and $\mathrm{rank}(Q_k) + \mathrm{rank}(V_k)$.

Since $P_kU_k = U_k$ and $P_kV_k = 0$, it is trivial to verify that P_{k+1} is self-adjoint and idempotent so that it is an orthogonal projector onto its own image. In fact P_{k+1} is the orthogonal projector onto the orthogonal complement of $\mathrm{Im}(U_k)$ in $\mathrm{Im}(P_k + V_k) = \mathrm{Im}(P_k) \oplus \mathrm{Im}(V_k)$. The claim for the rank of P_{k+1} follows from the fact that the rank of V_k is no larger than the rank of U_k.

Since

$$V_k(I - Q_k)AU_k = (I - Q_k)AU_k$$

we have

$$AU_k = V_k(I - Q_k)AU_k + Q_k AU_k$$

so that $\mathrm{Im}(AU_k) \subseteq \mathrm{Im}(Q_k + V_k)$ Since $AU_k A^*$ is the orthogonal projector onto $\mathrm{Im}(AU_k)$ it follows that Q_{k+1} is the projector onto the orthogonal complement of $\mathrm{Im}(AU_k A^*)$ in $\mathrm{Im}(Q_k + V_k)$. The claim for the rank of Q_{k+1} follows from the fact that the rank of V_k is no larger than the rank of $AU_k A^*$.

The Krylov-like structure of the sequence $M_j^{(k)}$ gives

$$\begin{aligned}
&(I - U_k)M_j^{(k)} - A(I - U_k)M_{j-1}^{(k)} \\
&\quad = (P_k - U_k)M_j^{(k)} - (Q_k - AU_k A^*)AM_{j-1}^{(k)} \\
&\quad = (P_k - U_k)M_j^{(k+1)} - (Q_k - AU_k A^*)AM_{j-1}^{(k+1)} - (Q_k - AU_k A^*)AU_k M_{j-1}^{(k)} \\
&\quad = (P_k - U_k)M_j^{(k+1)} - (Q_k - AU_k A^*)AM_{j-1}^{(k+1)} + (I - Q_k)AU_k M_{j-1}^{(k)}.
\end{aligned} \tag{14}$$

In the second line we have used the fact that $(P_k - U_k)(I - U_k) = (P_k - U_k)$ so that $(P_k - U_k)M_j^{(k)} = (P_k - U_k)M_j^{(k+1)}$. Since V_k is the projector onto $\mathrm{Im}((I - Q_k)AU_k)$ we have

$$(I - Q_k)AU_k = V_k(I - Q_k)AU_k = V_k AU_k.$$

Using $V_k P_k = V_k Q_k = V_k U_k = 0$ we get

$$\begin{aligned}
(I - Q_k)AU_k M_{j-1}^{(k)} &= V_k AU_k M_{j-1}^{(k)} \\
&= V_k\big(P_k M_j^{(k)} - Q_k AM_{j-1}^{(k)}\big) + V_k AU_k M_{j-1}^{(k)} \\
&= V_k\big(M_j^{(k)} - AM_{j-1}^{(k)}\big) + V_k AU_k M_{j-1}^{(k)} \\
&= V_k(I - U_k)M_j^{(k)} - V_k A(I - U_k)M_{j-1}^{(k)} \\
&= V_k M_j^{(k+1)} - V_k AM_{j-1}^{(k+1)}.
\end{aligned}$$

Substituting the final expression into (14) gives the desired result. □

The theorem gives recurrences for computing the displacement projectors P_k and Q_k for the Krylov-like sequence $M_j^{(k)}$. Given initial displacement projectors P_0 and Q_0 for a Krylov-like sequence M_j the recurrences define two sequences of subspaces $\mathrm{Im}(P_k)$ and $\mathrm{Im}(Q_k)$ of nonincreasing dimension.

Theorem 3 suggests a structured orthogonalization algorithm that looks very much like Algorithm 2. Given a Krylov-like sequence $M_j^{(0)} = M_j$, the sequence can be orthogonalized against the columns of $M_0^{(0)}$ to obtain the sequence $M_j^{(1)} = (I - U_0)M_j^{(0)}$ where U_0 is the projector onto the span of the columns of $M_0^{(0)}$. The $j = 0$ case of (5) with $M_{-1} = 0$ implies that $\mathrm{Im}(U_0) = \mathrm{Im}(M_0^{(0)}) \subseteq \mathrm{Im}(P_0)$ so that Theorem 3 applies to show that $M_j^{(1)}$ is Krylov-like. The theorem also gives explicit relations for the displacement projectors of $M_j^{(1)}$. Since $M_0^{(1)} = 0$ we have from

(5) that $\mathrm{Im}(M_1^{(1)}) \subseteq \mathrm{Im}(P_1)$. Thus the process can be repeated to orthogonalize the sequence $M_j^{(1)}$ against the columns of $M_1^{(1)}$ to get $M_j^{(2)} = (I - U_1)M_j^{(1)}$ where U_1 is the projector onto $\mathrm{Im}(U_1) = \mathrm{Im}(M_1^{(1)}) \subseteq \mathrm{Im}(P_1)$. In general $\mathrm{Im}(M_k^{(k)}) \subseteq \mathrm{Im}(P_k)$ so that this procedure can be used to compute displacement projectors for each of the partially orthogonalized sequences $M_j^{(k)}$.

Unfortunately this outline of the algorithm is incomplete for two reasons. First, while the proof of Theorem 3 depends on (8) holding for P_k and Q_k, we have not shown that P_{k+1} and Q_{k+1} satisfy (8). Second, although we have defined U_k, we have not given a computationally useful formula for computing it. We have suggested that U_k should be the projector onto $\mathrm{Im}(M_k^{(k)})$. Since we do not expect to have an explicit representation of the partially orthogonalized sequence $M_j^{(k)}$, this definition is not computationally useful.

Both gaps are filled in the next section. It can be shown that Krylov-like structure, including the relation (8), is indeed preserved during orthogonalization. We can also show that $P_k M_j = 0$ for $j < k$. Since $M_k^{(0)} - M_k^{(k)}$ has columns that are in the span of the columns of $M_0, M_1, \ldots, M_{k-1}$, it follows that $P_k(M_k^{(0)} - M_k^{(k)}) = 0$ so that

$$U_k = \mathscr{P}\big(M_k^{(k)}\big) = \mathscr{P}\big(P_k M_k^{(k)}\big) = \mathscr{P}\big(P_k M_k^{(0)}\big). \tag{15}$$

Thus U_k can be obtained from P_k and the original Krylov-like sequence $M_k^{(0)}$. This results in the following algorithm.

Algorithm 3. Generalized isometric Arnoldi in terms of projectors

Given: Q_0, P_0, and M_j for $j \geqslant 0$.
$k = 0$
$U_0 = \mathscr{P}(M_0)$
$V_0 = \mathscr{P}((I - Q_0)AU_0)$
For $k = 0, 1, 2, \ldots$

$$P_{k+1} = P_k - U_k + V_k$$
$$Q_{k+1} = Q_k - AU_kA^* + V_k$$
$$U_{k+1} = \mathscr{P}(P_{k+1}M_{k+1})$$
$$V_{k+1} = \mathscr{P}((I - Q_{k+1})AU_{k+1})$$

End For

4. Orthogonality relations

Theorem 3 is almost a proof of the correctness of Algorithm 3. As noted what remains to be proven is that the desired $U_k = \mathscr{P}(M_k^{(k)})$ can be computed through the relation $U_k = \mathscr{P}(P_k M_k)$ and that the relation (8) is satisfied for the sequences P_k and Q_k. The two issues are closely related. We have justified (15) by the claim that $\mathrm{Im}(P_k)$ is orthogonal to the columns of M_j for $j < k$. If U_l is chosen to be the projector onto $\mathrm{Im}(M_l^{(l)})$ for each $0 \leqslant l < k$ then this follows from the $j = 0$ case of

$$U_l[(I - Q_k)A]^j P_k = 0$$

for $l < k$ and $j \geqslant 0$. This property of U_l is clearly similar to (8).

Relations of this type can be proven by arguments that depend on properties of suitably defined invariant subspaces of $(I - Q_k)A$. We define $W = (I - Q_k)A$ so that both relations are of the form $YW^jX = 0$ for $j \geqslant 0$ for suitably chosen Y and X. Let $\mathscr{H}_0$ be an invariant subspace of W. If we decompose $\mathscr{H}$ as $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_0^\perp$ then W can be written as

$$W = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix} : \mathscr{H}_0 \oplus \mathscr{H}_0^\perp \to \mathscr{H}_0 \oplus \mathscr{H}_0^\perp, \tag{16}$$

where

$$W_{11} = P_{\mathscr{H}_0} W|_{\mathscr{H}_0}, \quad W_{12} = P_{\mathscr{H}_0} W|_{\mathscr{H}_0^\perp}, \quad \text{and} \quad W_{22} = P_{\mathscr{H}_0^\perp} W|_{\mathscr{H}_0^\perp}.$$

The operators X and Y can be similarly written as

$$Y = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}, \quad \text{and} \quad X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}. \tag{17}$$

The relation $YW^jX = 0$ for $j \geqslant 0$ has an interpretation in terms of systems theory. In particular, it shows that the controllability subspace of the pair (W, X) is orthogonal to the observability subspace of the pair (Y, W) [7]. If we let $\mathscr{H}_0$ be the controllability subspace of the pair (W, X) then we obtain the following decomposition.

Lemma 1. *Suppose that Y, W, and X are bounded operators on $\mathscr{H}$ satisfying $YW^jX = 0$ for $j \geqslant 0$. Let*

$$\mathscr{H}_1 = \{\mathbf{x} : \mathbf{x} = X\mathbf{x}_0 + WX\mathbf{x}_1 + \cdots + W^lX\mathbf{x}_l \textit{ for } \mathbf{x}_k \in \mathscr{H} \textit{ and } l \geqslant 0\}$$

and let $\mathscr{H}_0$ be the closure of $\mathscr{H}_1$. Then $\mathscr{H}_0$ is an invariant subspace of W and, with respect to the decomposition $\mathscr{H} = \mathscr{H}_0 \oplus \mathscr{H}_0^\perp$, we have

$$W = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix}, \quad X = \begin{bmatrix} X_{11} & X_{12} \\ 0 & 0 \end{bmatrix}, \quad \textit{and} \quad Y = \begin{bmatrix} 0 & Y_{12} \\ 0 & Y_{22} \end{bmatrix}. \tag{18}$$

Proof. Clearly $\mathscr{H}_0$ has been defined to be an invariant subspace of W. Thus $P_{\mathscr{H}_0^\perp} W|_{\mathscr{H}_0} = 0$ and W has the form (18) with respect to $\mathscr{H}_0 \oplus \mathscr{H}_0^\perp$. Let X and Y be partitioned as (17). By construction $\operatorname{Im}(X) \subseteq \mathscr{H}_0$ so that $P_{\mathscr{H}_0^\perp} X = 0$ which gives the desired form for X. Since $YW^jX = 0$ for $j \geqslant 0$ and $\operatorname{Im}(W^jX) \subseteq \mathscr{H}_0$ we have

$$Y_{11}W^jX = P_{\mathscr{H}_0} Y|_{\mathscr{H}_0} W^jX = P_{\mathscr{H}_0} YW^jX = 0$$

for $j \geqslant 0$. Similarly $Y_{21} = P_{\mathscr{H}_0^\perp} YW^jX = 0$ for $j \geqslant 0$. Thus for any $\mathbf{x} \in \mathscr{H}_1$

$$Y_{11}\mathbf{x} = Y_{11}X\mathbf{x}_0 + Y_{11}WX\mathbf{x}_1 + \cdots + Y_{11}W^lX\mathbf{x}_l = 0$$

This implies $Y_{11}\mathbf{x} = 0$ for any $\mathbf{x} \in \mathscr{H}_0$. Since $Y_{11} = P_{\mathscr{H}_0}.Y|_{\mathscr{H}_0}$, this is equivalent to $Y_{11} = 0$. That $Y_{21} = 0$ follows using the obvious variation of this argument. $\square$

The following theorem establishes that P_k and Q_k as generated by Algorithm 3 are displacement projectors. Note that the properties of these projectors depend only on the U_k being chosen to be the projector onto a subspace of P_k and not on U_k being chosen to be the projector onto $\operatorname{Im}(M_k^{(k)})$.

Theorem 4. *Let P_0 and Q_0 be orthogonal projectors satisfying*

$$P_0[(I - Q_0)A]^j P_0 = 0$$

for $j \geqslant 1$. For $k \geqslant 0$ let the sequences P_k, Q_k, U_k, and V_k be generated by the following procedure: *U_k is chosen to be the orthogonal projector onto an arbitrary subspace of* $\operatorname{Im}(P_k)$ *and*

$$V_k = \mathscr{P}((I - Q_k)AU_k), \quad P_{k+1} = P_k - U_k + V_k, \quad \textit{and} \quad Q_{k+1} = Q_k - AU_kA^* + V_k.$$

(*This is just Algorithm 3 except that we do not require that $U_k = \mathscr{P}(P_kM_k)$.*) *Then the sequences P_k and Q_k are sequences of orthogonal projectors satisfying*

$$P_k[(I - Q_k)A]^j P_k = 0 \tag{19}$$

for $j \geqslant 1$. We also have the relations $U_kV_k = Q_kV_k = P_kV_k = P_{k+1}U_k = 0$ and $P_{k+1}V_k = V_k$.

Proof. The proof is inductive. We assume that P_l and Q_l are orthogonal projectors satisfying $P_l[(I - Q_l)A]^j P_l = 0$ for $0 \leqslant l \leqslant k$ and $j \geqslant 1$. This assumption is sufficient to prove all the orthogonality relations in addition to showing that P_{k+1} and Q_{k+1} are projectors satisfying $P_{k+1}[(I - Q_{k+1})A]^j P_{k+1} = 0$ for $j \geqslant 1$ which completes the induction.

We start with the orthogonality relations between U_k, V_k, Q_k, and P_k. In the proof of Theorem 3 we have shown that $P_kV_k = U_kV_k = Q_kV_k = 0$ and that P_{k+1} and Q_{k+1} are orthogonal projectors. Since $V_kU_k = 0$ and $(P_k - U_k)U_k = 0$ we have $P_{k+1}U_k = U_kP_{k+1} = 0$. Since $(P_k - U_k)V_k = 0$ we have $P_{k+1}V_k = V_k$.

To prove (19) we use Lemma 1 with

$$W = (I - Q_k)A, \quad Y = P_k, \quad \text{and} \quad X = WY = (I - Q_k)AP_k. \tag{20}$$

The induction hypothesis gives $YW^jX = 0$ for $j \geqslant 0$. If we define $\mathscr{H}_0$ and $\mathscr{H}_0^\perp$ as in Lemma 1 then with respect to the decomposition $\mathscr{H}_0 \oplus \mathscr{H}_0^\perp$ we have

$$W = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix}, \quad X = \begin{bmatrix} 0 & X_{12} \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad Y = \begin{bmatrix} 0 & 0 \\ 0 & Y_{22} \end{bmatrix}. \tag{21}$$

The additional zero blocks in X and Y that are not present in (17) arise as follows. Since Y is an orthogonal projector it is self-adjoint and we therefore have $Y_{12} = Y_{21}^* = 0$. Since $X = WY$, block multiplication gives $X_{11} = 0$. Note that the relation $X = WY$ also implies $W_{22}Y_{22} = 0$.

Let

$$\hat{Y} = P_{k+1}, \quad \hat{W} = (I - Q_{k+1})A, \quad \text{and} \quad \hat{X} = \hat{W}\hat{Y} = (I - Q_{k+1})AP_{k+1}.$$

Proving (19) is then equivalent to proving that $\hat{Y}\hat{W}^j\hat{X} = 0$ for $j \geqslant 0$. We do this by considering the block structure of $\hat{X}$, $\hat{Y}$, and $\hat{W}$ with respect to the decomposition $\mathscr{H}_0 \oplus \mathscr{H}_0^\perp$.

The projector U_k satisfies $U_kP_k = U_k$ so that

$$U_kW^jX = U_kP_k[(I - Q_k)A]^{j+1}P_k = 0$$

for $j \geqslant 0$ by the induction hypothesis. From the definition of $\mathscr{H}_0$ in Lemma 1, it follows that $P_{\mathscr{H}_0}U_k = 0$ so that

$$U_k = \begin{bmatrix} 0 & 0 \\ 0 & U_{22} \end{bmatrix}.$$

Since V_k is the projector onto $\operatorname{Im}((I - Q_k)AU_k) = \operatorname{Im}((I - Q_k)AP_kU_k) = \operatorname{Im}(XU_k)$ we clearly have $\operatorname{Im}(V_k) \subseteq \mathscr{H}_0$ so that $P_{\mathscr{H}_0^\perp}V_k = 0$ and

$$V_k = \begin{bmatrix} V_{11} & 0 \\ 0 & 0 \end{bmatrix}.$$

If we similarly partition A with respect to $\mathscr{H}_0 \oplus \mathscr{H}_0^\perp$ we get

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}. \tag{22}$$

Combining the partitionings of W, U_k, and V_k with the definitions of $\hat{W}$, $\hat{Y}$, and $\hat{X}$ gives

$$\hat{W} = W + AU_k - V_k A = \begin{bmatrix} W_{11} - V_{11}A_{11} & W_{12} + A_{12}U_{22} - V_{11}A_{12} \\ 0 & W_{22} + A_{22}U_{22} \end{bmatrix},$$

$$\hat{Y} = Y - U_k + V_k = \begin{bmatrix} V_{11} & 0 \\ 0 & Y_{22} - U_{22} \end{bmatrix},$$

and

$$\hat{X} = \begin{bmatrix} W_{11} - V_{11}A_{11} & W_{12} + A_{12}U_{22} - V_{11}A_{12} \\ 0 & W_{22} + A_{22}U_{22} \end{bmatrix} \begin{bmatrix} V_{11} & 0 \\ 0 & Y_{22} - U_{22} \end{bmatrix} = \begin{bmatrix} \hat{X}_{11} & \hat{X}_{12} \\ 0 & 0 \end{bmatrix}.$$

In the equation for $\hat{X}$ we have used $W_{22}Y_{22} = 0$ together with the fact that $\mathrm{Im}(U_k) \subseteq \mathrm{Im}(P_k) = \mathrm{Im}(Y)$ so that $Y_{22} - U_{22}$ is the projector onto the orthogonal complement of $\mathrm{Im}(U_{22})$ in $\mathrm{Im}(Y_{22})$. Thus

$$(W_{22} + A_{22}U_{22})(Y_{22} - U_{22}) = W_{22}Y_{22}(Y_{22} - U_{22}) = 0.$$

The block structures of Y, $\hat{W}$, and $\hat{X}$ imply that

$$P_k[(I - Q_{k+1})A]^{j+1}P_{k+1} = Y\hat{W}^j\hat{X} = 0 \tag{23}$$

for $j \geqslant 0$. To complete the proof we recall that $V_k(I - Q_k) = V_k$ so that

$$V_k\hat{W} = V_k[(I - Q_k)A + AU_k - V_kA] = V_kA + V_kAU_k - V_kA = V_kAU_k.$$

Therefore

$$V_k\hat{W}^j\hat{X} = V_k\hat{W}\hat{W}^j\hat{Y} = V_kAU_k\hat{W}^j\hat{Y} = 0 \tag{24}$$

for $j \geqslant 0$. For $j \geqslant 1$ have used $Y\hat{W}^j\hat{X} = 0$ and $U_kP_k = U_kY = U_k$ which imply that $U_k\hat{W}^j\hat{Y} = U_kY\hat{W}^{j-1}\hat{X} = 0$ for $j \geqslant 1$. For $j = 0$ we note that $U_k\hat{Y} = U_k(P_k - U_k + V_k) = 0$ since $U_k(P_k - U_k) = 0$ and $U_kV_k = 0$. Since $V_k\hat{W}^j\hat{X} = 0$ for $j \geqslant 0$, it follows that in $\hat{Y}\hat{W}^j\hat{X}$ the V_{11} block of $\hat{Y}$ can be ignored to get

$$\hat{Y}\hat{W}^j\hat{X} = \begin{bmatrix} 0 & 0 \\ 0 & Y_{22} - U_{22} \end{bmatrix} \begin{bmatrix} W_{11} - V_{11}A_{11} & W_{12} + A_{12}U_{22} - V_{11}A_{12} \\ 0 & W_{22} + A_{22}U_{22} \end{bmatrix}^j \begin{bmatrix} \hat{X}_{11} & \hat{X}_{12} \\ 0 & 0 \end{bmatrix}$$
$$= 0. \qquad \square$$

Theorem 5. *With P_k, Q_k, and U_k generated as in Theorem 4*

$$P_l[(I - Q_k)A]^jP_k = 0 \quad \text{for } j \geqslant 1 \text{ and } l \leqslant k, \tag{25}$$

$$Q_l[(I - Q_k)A]^jP_k = 0 \quad \text{for } j \geqslant 1 \text{ and } l \leqslant k, \tag{26}$$

and

$$U_l[(I - Q_k)A]^jP_k = 0 \quad \text{for } j \geqslant 0 \text{ and } l < k. \tag{27}$$

Proof. We start with (25). The proof is by induction on k. If $k = l$ the result follows immediately from Theorem 4. We assume that it is true for some $k \geqslant l$ and prove it for $k + 1$. Define

$$Y = P_l, \quad W = (I - Q_k)A, \quad \text{and} \quad X = (I - Q_k)AP_k$$

and let $\mathscr{H}_0$ be as in Lemma 1. The induction hypothesis gives $YW^jX = 0$ for $j \geqslant 0$. As in the proof of Theorem 4 this implies

$$Y = \begin{bmatrix} 0 & 0 \\ 0 & Y_{22} \end{bmatrix}, \quad W = \begin{bmatrix} W_{11} & W_{12} \\ 0 & W_{22} \end{bmatrix}, \quad \text{and} \quad X = \begin{bmatrix} 0 & X_{12} \\ 0 & 0 \end{bmatrix}.$$

If we define

$$\hat{W} = (I - Q_{k+1})A, \quad \text{and} \quad \hat{X} = (I - Q_{k+1})AP_{k+1}$$

then the induction step is equivalent to proving $Y\hat{W}^j\hat{X} = 0$ for $j \geqslant 0$. Since $\mathscr{H}_0$, $\hat{X}$, and $\hat{W}$ are exactly the same as in the proof of Theorem 4 and since Y has the same form with respect to $\mathscr{H}_0 \oplus \mathscr{H}_0^\perp$, the same proof establishes $Y\hat{W}^j\hat{X}_0 = 0$.

For (26) we note that the $l = k$ case follows from the fact that $Q_k(I - Q_k) = 0$. Assuming that the relation holds for some $k \geqslant l$ we can prove it for $k + 1$ in the same way as was done for (25). The only difference is that $Y = Q_l$ instead of $Y = P_l$.

Since $U_l = U_lP_l$, (25) implies the $j \geqslant 1$ case of (27). For $j = 0$ we need to prove that $U_lP_k = 0$ for $k > l$. This is done by induction on k. For $k = l + 1$ we have

$$U_lP_{l+1} = U_l(P_l - U_l + V_l) = 0,$$

where we have used $U_lP_l = U_l$ and $U_lV_l = 0$ which were established in Theorem 4. We assume that $U_lP_k = 0$ for some $k \geqslant l + 1$. It follows that

$$U_lP_{k+1} = U_l(P_k - U_k + V_k) = U_lV_k$$

since $U_l(P_k - U_k) = U_lP_k(P_k - U_k) = 0$ by the induction assumption. However V_k is the projector onto $\mathrm{Im}((I - Q_k)AU_k)$ and

$$U_l(I - Q_k)AU_k = U_l[(I - Q_k)A]^1P_kU_k = 0$$

since we have already proven the $j = 1$ case of (27). Thus $U_lP_{k+1} = U_lV_k = 0$. □

Corollary 1. *For a sequence U_k computed by Algorithm 3*

$$U_iU_j = 0 \quad \text{for } i \neq j.$$

We are ready to prove the correctness of Algorithm 3. Instead of letting U_k be the projector onto an arbitrary subspace of $\mathrm{Im}(P_k)$ we choose $U_k = \mathscr{P}(P_kM_k)$ and combine the results of this section with Theorem 3 to prove the following.

Theorem 6. *Let M_j be Krylov-like with displacement projectors P_0 and Q_0 and let P_k, Q_k, and U_k be the sequences of projectors computed by Algorithm 3. Define $M_j^{(0)} = M_j$ and let $M_j^{(k)}$ be the result of orthogonalizing the sequence $M_j^{(0)}$ for $j \geqslant 0$ against the subspace spanned by the columns of $M_l^{(0)}$ for $0 \leqslant l \leqslant k - 1$. The orthogonalization corresponds to a block form of modified Gram–Schmidt that can be written in terms of projections as*

$$M_j^{(k)} = \left(I - \mathscr{P}\left(M_{k-1}^{(k-1)}\right)\right)\left(I - \mathscr{P}\left(M_{k-2}^{(k-2)}\right)\right)\cdots\left(I - \mathscr{P}\left(M_0^{(0)}\right)\right)M_j^{(0)}. \tag{28}$$

The sequence $M_j^{(k)}$ is a Krylov-like sequence satisfying

$$M_j^{(k)} - AM_{j-1}^{(k)} = P_kM_j^{(k)} - Q_kAM_{j-1}^{(k)}. \tag{29}$$

The projectors U_k satisfy $U_k = \mathscr{P}\left(M_k^{(k)}\right)$ and $U_j U_i = 0$ for $j \neq i$. The projector

$$U^{(k)} = U_0 + U_1 + \cdots + U_{k-1}$$

is the orthogonal projector onto

$$\operatorname{Im}([M_0 \quad M_1 \quad \cdots \quad M_{k-1}]).$$

Proof. We first note that U_k, as computed in Algorithm 3 is an orthogonal projector for a subspace of $\operatorname{Im}(P_0)$. Consequently we have all the properties of P_k, Q_k, U_k and V_k given in Theorem 4 and Theorem 5.

The proofs of (29) and $U_k = \mathscr{P}\left(M_k^{(k)}\right)$ are by induction. For $k = 0$, (29) is just a restatement of the assumption that M_j is Krylov-like. Since $M_{-1}^{(0)} = 0$ we have $P_0 M_0^{(0)} = M_0^{(0)}$ so that

$$U_0 = \mathscr{P}\left(P_0 M_0^{(0)}\right) = \mathscr{P}\left(M_0^{(0)}\right).$$

We assume that

$$U_l = \mathscr{P}\left(P_l M_l^{(0)}\right) = \mathscr{P}\left(M_l^{(l)}\right), \quad \text{and} \quad M_j^{(l)} - A M_{j-1}^{(l)} = P_l M_j^{(l)} - Q_l A M_{j-1}^{(l)}$$

for $0 \leqslant l \leqslant k$ and prove these relations for $l = k+1$. Since $U_k = \mathscr{P}(P_k M_k^{(0)}) = \mathscr{P}(M_k^{(k)})$ and $M_j^{(k+1)}$ is defined to be $M_j^{(0)}$ orthogonalized against $M_l^{(0)}$ for $0 \leqslant l \leqslant k$ we have

$$M_j^{(k+1)} = (I - U_k) M_j^{(k)}.$$

The Krylov-like structure of $M_j^{(k+1)}$ then follows from Theorem 3. The fact that P_{k+1} and Q_{k+1} are displacement projectors follows from Theorem 4.

Clearly $M_k^{(k+1)} = 0$. Thus $M_j^{(k+1)} - A M_{j-1}^{(k+1)} = P_{k+1} M_j^{(k+1)} - Q_{k+1} A M_{j-1}^{(k+1)}$ implies $M_{k+1}^{(k+1)} = P_{k+1} M_{k+1}^{(k+1)}$. By Theorem 5 $U_l P_{k+1} = P_{k+1} U_l = 0$ for $l \leqslant k$. Combining this with

$$M_j^{(k+1)} = (I - U_k) \cdots (I - U_0) M_j^{(0)}$$

gives $P_{k+1} M_j^{(k+1)} = P_{k+1} M_j^{(0)}$ so that

$$\mathscr{P}\left(M_{k+1}^{(k+1)}\right) = \mathscr{P}\left(P_{k+1} M_{k+1}^{(k+1)}\right) = \mathscr{P}\left(P_{k+1} M_{k+1}^{(0)}\right) = U_{k+1}.$$

The claim for $U^{(k)}$ is obvious. □

This completes the proof that Algorithm 3 orthogonalizes a Krylov-like sequence. Unfortunately, the algorithm is inefficient: If $\mathscr{H} = \mathbb{C}^n$, then storing the projectors P_k and Q_k requires $O(n^2)$ memory and each update of the projectors requires $O(n^2)$ operations. If multiplication by A is fast and if $\operatorname{Im}(P_0)$ and $\operatorname{Im}(Q_0)$ are subspaces of low dimension, it is possible to obtain a fast algorithm by factoring the projectors P_k and Q_k into products of the form $P_k = X_k X_k^*$ and $Q_k = Y_k Y_k^*$ where the columns of X_k form an orthonormal basis for $\operatorname{Im}(P_k)$ and the columns of Y_k form an orthonormal basis for $\operatorname{Im}(Q_k)$. The resulting bases can be computed more efficiently than the corresponding projectors. This modification requires knowledge of the dimensions of the subspaces and that special measures to be taken when the dimensions change. In preparation for describing a factored form of the algorithm in the next section, we give further details relating the dimensions of the images of P_k, Q_k, U_k, and V_k.

Let

$$x_k = \operatorname{rank}(P_k), \quad y_k = \operatorname{rank}(Q_k), \quad t_k = \operatorname{rank}(V_k), \quad p_k = \operatorname{rank}(U_k), \tag{30}$$

and

$$r_k = \dim(\mathrm{Im}(Q_k) \cap \mathrm{Im}(AU_k)). \tag{31}$$

The relation between x_k, y_k, t_k and r_k is given in the following theorem. The theorem also identifies the criterion $\|Q_k AU_k\| = 1$ as signaling a drop in dimension and describes a deflation step in which the intersection of $\mathrm{Im}(Q_k)$ and $\mathrm{Im}(AU_k)$ is removed from P_k, Q_k, and U_k without changing the computed V_k, P_{k+1} and Q_{k+1}.

Theorem 7. *Let U_k, V_k, Q_k, and P_k be computed as in Algorithm 3. Define*

$$\mathscr{X}_k = \{\mathbf{x} : \|Q_k AU_k \mathbf{x}\| = \|\mathbf{x}\|\}.$$

Then

$$\mathscr{X}_k = \mathrm{Ker}((I - Q_k)AU_k) \cap \mathrm{Im}(U_k) \tag{32}$$

and

$$A\mathscr{X}_k = AU_k\mathscr{X}_k = \mathrm{Im}(Q_k) \cap \mathrm{Im}(AU_k).$$

If r_k is the dimension of $\mathscr{X}_k$ then

$$t_k = p_k - r_k, \quad y_{k+1} = y_k - r_k, \quad \textit{and} \quad x_{k+1} = x_k - r_k,$$

where t_k, y_k, and x_k are as in (30) and (31).

If R_k is the orthogonal projector onto $\mathscr{X}_k$ and

$$\widetilde{P}_k = P_k - R_k, \quad \widetilde{Q}_k = Q_k - AR_kA^*, \quad \textit{and} \quad \widetilde{U}_k = U_k - R_k$$

then

$$V_k = \mathscr{P}((I - Q_k)AU_k) = \mathscr{P}((I - \widetilde{Q}_k)A\widetilde{U}_k),$$
$$\|\widetilde{Q}_k A\widetilde{U}_k\| < 1,$$

and

$$P_{k+1} = \widetilde{P}_k - \widetilde{U}_k + V_k, \quad \textit{and} \quad Q_{k+1} = \widetilde{Q}_k - A\widetilde{U}_kA^* + V_k.$$

Proof. If $\mathbf{x} \in \mathscr{X}_k$ then $\mathbf{x} \in \mathrm{Im}(U_k)$ since otherwise $\|Q_kAU_k\mathbf{x}\| \leqslant \|Q_kA\|\|U_k\mathbf{x}\| < \|Q_kA\|\|\mathbf{x}\| \leqslant \|\mathbf{x}\|$. Similarly, we must also have $A\mathbf{x} \in \mathrm{Im}(Q_k)$ since otherwise $\|Q_kAU_k\mathbf{x}\| = \|Q_kA\mathbf{x}\| < \|A\mathbf{x}\| = \|\mathbf{x}\|$. Thus $A\mathscr{X}_k \subseteq \mathrm{Im}(AU_k) \cap \mathrm{Im}(Q_k)$.

For any $\mathbf{y} \in \mathrm{Im}(AU_k) \cap \mathrm{Im}(Q_k)$ there exists $\mathbf{x} \in \mathrm{Im}(U_k)$ such that $\mathbf{y} = AU_k\mathbf{x} = A\mathbf{x}$. Since $\mathbf{y} \in \mathrm{Im}(Q_k)$,

$$\|Q_kAU_k\mathbf{x}\| = \|AU_k\mathbf{x}\| = \|A\mathbf{x}\| = \|\mathbf{x}\|$$

so that $\mathbf{x} \in \mathscr{X}_k$ and $\mathbf{y} \in A\mathscr{X}_k$. Thus $A\mathscr{X}_k = \mathrm{Im}(AU_k) \cap \mathrm{Im}(Q_k)$.

We have shown that for $\mathbf{x} \in \mathscr{X}_k$, $Q_kAU_k\mathbf{x} = Q_kA\mathbf{x} = A\mathbf{x}$. Conversely, if $Q_kAU_k\mathbf{x} = A\mathbf{x}$ then the fact that A is an isometry implies $\|Q_kAU_k\mathbf{x}\| = \|\mathbf{x}\|$ so that $\mathbf{x} \in \mathscr{X}_k$. Thus $\mathscr{X}_k$ can be characterized as the set of $\mathbf{x} \in \mathscr{H}$ for which $(A - Q_kAU_k)\mathbf{x} = 0$. Since $\mathscr{X}_k \subseteq \mathrm{Im}(U_k)$, this implies that if $\mathbf{x} \in \mathscr{X}_k$ then $(I - Q_k)AU_k\mathbf{x} = 0$. Thus $\mathscr{X}_k \subseteq \mathrm{Ker}((I - Q_k)AU_k) \cap \mathrm{Im}(U_k)$. Conversely if $\mathbf{x} \in \mathrm{Ker}((I - Q_k)AU_k) \cap \mathrm{Im}(U_k)$ then $Q_kAU_k\mathbf{x} = AU_k\mathbf{x} = A\mathbf{x}$ from which it follows that $\|Q_kAU_k\mathbf{x}\| = \|\mathbf{x}\|$. Thus $\mathbf{x} \in \mathscr{X}_k$. This establishes (32).

Since $\mathscr{X}_k \subseteq \mathrm{Im}(U_k)$, we can define $\mathscr{X}_k^{\perp,U} \subseteq \mathrm{Im}(U_k)$ to be the orthogonal complement of $\mathscr{X}_k$ in $\mathrm{Im}(U_k)$. Then

$$\mathrm{Im}(V_k) = \mathrm{Im}((I - Q_k)AU_k) = (I - Q_k)AU_k\mathscr{X}_k^{\perp,U}.$$

The kernel of $(I - Q_k)AU_k|_{\mathscr{X}_k^{\perp,U}}$ is trivial since anything in the kernel would also have to be in $\mathscr{X}_k$ by (32). It follows that

$$\operatorname{rank}(V_k) = \dim\left((I - Q_k)AU_k\mathscr{X}_k^{\perp,U}\right) = \dim\left(\mathscr{X}_k^{\perp,U}\right) = \operatorname{rank}(U_k) - \dim(\mathscr{X}_k).$$

Since, as noted in the proof of Theorem 3, $Q_kV_k = 0$ and $\operatorname{Im}(AU_kA^*) \subseteq \operatorname{Im}(Q_k) + \operatorname{Im}(V_k)$ the relation $Q_{k+1} = Q_k - AU_kA^* + V_k$ implies

$$\operatorname{rank}(Q_{k+1}) = \operatorname{rank}(Q_k) + \operatorname{rank}(V_k) - \operatorname{rank}(AU_kA^*) = \operatorname{rank}(Q_k) - \dim(\mathscr{X}_k).$$

In a similar manner $\operatorname{Im}(U_k) \subseteq \operatorname{Im}(P_k)$, $P_kV_k = 0$, and $P_{k+1} = P_k - U_k + V_k$ imply the expression for x_{k+1}.

That $\widetilde{P}_k$, $\widetilde{Q}_k$, and $\widetilde{U}_k$ are orthogonal projectors with $\operatorname{Im}(\widetilde{P}_k) \subseteq \operatorname{Im}(P_k)$, $\operatorname{Im}(\widetilde{Q}_k) \subseteq \operatorname{Im}(Q_k)$, and $\operatorname{Im}(\widetilde{U}_k) \subseteq \operatorname{Im}(U_k)$ follows from $\mathscr{X}_k \subseteq \operatorname{Im}(U_k) \subseteq \operatorname{Im}(P_k)$ and $A\mathscr{X}_k \subseteq \operatorname{Im}(Q_k)$. From this it is clear that $\|\widetilde{Q}_kA\widetilde{U}_k\| \leqslant 1$. If $\|\widetilde{Q}_kA\widetilde{U}_k\| = 1$ then there is an $\mathbf{x} \neq 0$ such that $\|\widetilde{Q}_kA\widetilde{U}_k\mathbf{x}\| = \|\mathbf{x}\|$. As before this implies $\mathbf{x} \in \operatorname{Im}(\widetilde{U}_k) \subseteq \operatorname{Im}(U_k)$ and $A\mathbf{x} \in \operatorname{Im}(\widetilde{Q}_k) \subseteq \operatorname{Im}(Q_k)$ which implies that $\mathbf{x} \in \mathscr{X}_k$. However $\widetilde{U}_k\mathbf{x} = (U_k - R_k)\mathbf{x} = 0$ for all $\mathbf{x} \in \mathscr{X}_k$ so that $\|\widetilde{Q}_kA\widetilde{U}_k\mathbf{x}\| = 0$. From this contradiction we conclude that $\|\widetilde{Q}_kA\widetilde{U}_k\mathbf{x}\| < 1$.

The relations for P_{k+1} and Q_{k+1} are obvious from the definition of $\widetilde{P}_k$, $\widetilde{Q}_k$, and $\widetilde{U}_k$. The relation for V_k follows from

$$\begin{aligned}\mathscr{P}((I - \widetilde{Q}_k)A\widetilde{U}_k) &= \mathscr{P}((I - (Q_k - AR_kA^*))A(U_k - R_k))\\ &= \mathscr{P}((I - Q_k)AU_k + AR_k(U_k - R_k) - (I - Q_k)AR_k)\\ &= \mathscr{P}((I - Q_k)AU_k)\\ &= V_k. \qquad \square\end{aligned}$$

There are several special cases that are covered by the theorem but merit further description.

1. If $r_n = x_n$ then $P_{n+1} = 0$ so that $U_k = 0$ for $k \geqslant n + 1$. The algorithm can stop at this point. The columns of M_k for $k \geqslant n + 1$ are in the span of the columns of M_l for $0 \leqslant l \leqslant n$. If, as in the case of the isometric Arnoldi algorithm, $\operatorname{rank}(P_0) = 1$ then $\|Q_nAU_n\| = 1$ implies $r_n > 0$ so that $P_l = 0$ for $l \geqslant n + 1$ and $U_l = 0$ for $l \geqslant n + 1$. Thus the isometric Arnoldi algorithm terminates whenever $|\mathbf{y}_n^*A\mathbf{x}_n| = \|Q_nAU_n\| = 1$.
2. If $r_n < x_n$ then $r_n > 0$ does not necessarily indicate linear dependence. While this situation does not occur in the case of the isometric Arnoldi algorithm, in general $\|Q_nAU_n\| = 1$ indicates only a decrease in x_{n+1} and y_{n+1} and not linear dependence in the columns of the Krylov-like sequence. The correct general criterion for identifying linear dependence is $\operatorname{rank}(U_n) < p$.
3. If $r_n = y_n$ then $y_{n+1} = 0$, $Q_{n+1} = 0$, and $(I - Q_{n+1})AU_{n+1} = AU_{n+1}$ so that $V_{n+1} = AU_{n+1}A^*$. From this it follows that $Q_{n+2} = Q_{n+1} = 0$ and that $\operatorname{rank}(P_{n+2}) = \operatorname{rank}(P_{n+1})$. This pattern continues: for $k \geqslant n + 1$ we have $Q_k = 0$ and $\operatorname{rank}(P_k) = \operatorname{rank}(P_{n+1})$.

5. The factored algorithm

Let

$$P_k = X_kX_k^*, \quad Q_k = Y_kY_k^*, \quad U_k = S_kS_k^*, \quad \text{and} \quad V_k = T_kT_k^*,$$

where

$$X_k : \mathbb{C}^{x_k} \to \mathscr{H}, \quad Y_k : \mathbb{C}^{y_k} \to \mathscr{H}, \quad S_k : \mathbb{C}^{p_k} \to \mathscr{H}, \quad \text{and} \quad T_k : \mathbb{C}^{p_k - r_k} \to \mathscr{H}$$

have columns that form orthonormal bases for $\mathrm{Im}(P_k)$, $\mathrm{Im}(Q_k)$, $\mathrm{Im}(U_k)$, and $\mathrm{Im}(V_k)$ respectively. Thus

$$X_k^* X_k = I_{x_k}, \quad Y_k^* Y_k = I_{y_k}, \quad S_k^* S_k = I_{p_k}, \quad \text{and} \quad T_k^* T_k = I_{p_k - r_k}.$$

For the deflated projectors $\widetilde{P}_k$, $\widetilde{Q}_k$, and $\widetilde{U}_k$ described in Theorem 7 we define in a similar manner $\widetilde{X}_k$, $\widetilde{Y}_k$, and $\widetilde{S}_k$ with $x_k - r_k$, $y_k - r_k$, and $p_k - r_k$ columns, respectively. If $r_k = 0$ so that no deflation is necessary then $\widetilde{X}_k = X_k$, $\widetilde{Y}_k = Y_k$, and $\widetilde{S}_k = S_k$.

We require a deflation procedure for computing $\widetilde{X}_k$, $\widetilde{Y}_k$, and $\widetilde{S}_k$ from X_k, Y_k, and S_k. In the following we assume that X_k is of the form $X_k = [S_k \quad X_{k,2}]$. Since $\mathrm{Im}(U_k) \subseteq \mathrm{Im}(P_k)$ implies $\mathrm{Im}(S_k) \subseteq \mathrm{Im}(X_k)$, X_k can be put in this form by a suitable choice of orthonormal basis for $\mathrm{Im}(P_k)$.

Algorithm 4. Deflation of X_k, Y_k, and S_k

Given X_k, Y_k, and S_k with $\|Y_k^* A S_k\|_2 = 1$
and X_k partitioned as $X_k = [S_k \quad X_{k,2}]$:
Compute unitary $E_k = [E_{k,1} \quad E_{k,2}]$ and $F_k = [F_{k,1} \quad F_{k,2}]$ such that

$$E_k^{\mathrm{H}} S_k^* A^* Y_k F_k = \begin{bmatrix} I_{r_k} & 0 \\ 0 & \Sigma_{k,2} \end{bmatrix}$$

where $\Sigma_{k,2}$ is $(p_k - r_k) \times (y_k - r_k)$ and $\|\Sigma_{k,2}\|_2 < 1$.

Define $Y_{k,j}$ and $S_{k,j}$ by

$$Y_k F_k = [Y_{k,1} \quad Y_{k,2}]$$
$$S_k E_k = [S_{k,1} \quad S_{k,2}]$$

Let

$$\widetilde{X}_k = [S_{k,2} \quad X_{k,2}]$$
$$\widetilde{Y}_k = Y_{k,2} \quad \text{and} \quad \widetilde{S}_k = S_{k,2}.$$

We now show that the above algorithm computes $\widetilde{X}_k$, $\widetilde{Y}_k$, and $\widetilde{S}_k$ such that $\widetilde{P}_k = \widetilde{X}_k \widetilde{X}_k^*$, $\widetilde{Q}_k = \widetilde{Y}_k \widetilde{Y}_k^*$, and $\widetilde{U}_k = \widetilde{S}_k \widetilde{S}_k^*$. Let $\hat{\mathscr{X}}_k$ be the right singular subspace of $Y_k^* A S_k$ associated with the singular value 1. We claim that the subspace $\mathscr{X}_k$ defined in Theorem 7 is given by $\mathscr{X}_k = S_k \hat{\mathscr{X}}_k$. If $\hat{\mathbf{x}} \in \hat{\mathscr{X}}_k$ then

$$\|Q_k A U_k S_k \hat{\mathbf{x}}\| = \|Y_k Y_k^* A S_k \hat{\mathbf{x}}\| = \|Y_k^* A S_k \hat{\mathbf{x}}\|_2 = \|\hat{\mathbf{x}}\|_2 = \|S_k \hat{\mathbf{x}}\|$$

so that $S_k \hat{\mathscr{X}}_k \subseteq \mathscr{X}_k$. If $\mathbf{x} \in \mathscr{X}_k$ then $\mathbf{x} \in \mathrm{Im}(U_k) = \mathrm{Im}(S_k)$ and if we define $\hat{\mathbf{x}} = S_k^* \mathbf{x}$ so that $\mathbf{x} = S_k \hat{\mathbf{x}}$ then

$$\|Y_k^* A S_k \hat{\mathbf{x}}\|_2 = \|Y_k^* A S_k S_k^* \mathbf{x}\|_2 = \|Q_k A U_k \mathbf{x}\| = \|\mathbf{x}\| = \|S_k^* \mathbf{x}\|_2 = \|\hat{\mathbf{x}}\|_2.$$

Thus every $\mathbf{x} \in \mathscr{X}_k$ is of the form $\mathbf{x} = S_k \hat{\mathbf{x}}$ for some $\hat{\mathbf{x}} \in \hat{\mathscr{X}}_k$ and therefore $\mathscr{X}_k \subseteq S_k \hat{\mathscr{X}}_k$. It follows that $\mathscr{X}_k = S_k \hat{\mathscr{X}}_k$ so that $\dim(\mathscr{X}_k) = \dim(S_k \hat{\mathscr{X}}_k) = \dim(\hat{\mathscr{X}}_k)$ and the r_k computed by Algorithm 4 satisfies $r_k = \dim(\mathscr{X}_k)$.

From Theorem 7 we have

$$\mathrm{Im}(A S_k) \cap \mathrm{Im}(Y_k) = A \mathscr{X}_k = A S_k \hat{\mathscr{X}}_k = \mathrm{Im}(A S_{k,1}) = \mathrm{Im}(Y_{k,1}),$$

where in the final equality we have used the fact that $Y_{k,1}^* A S_{k,1} = I_{r_k}$ and the fact that $Y_{k,1}$ and $S_{k,1}$ are isometries imply that $Y_{k,1} = A S_{k,1}$. Note also that $\|\widetilde{Y}_k^* A \widetilde{S}_k\|_2 < 1$ implies that $\mathrm{Im}(A \widetilde{S}_k) \cap \mathrm{Im}(\widetilde{Y}_k) = \{\mathbf{0}\}$.

We then have

$$\widetilde{Y}_k\widetilde{Y}_k^* = Y_kY_k^* - Y_{k,1}Y_{k,1}^* = Y_kY_k^* - AS_{k,1}S_{k,1}^*A^* = Q_k - AR_kA^* = \widetilde{Q}_k,$$
$$\widetilde{X}_k\widetilde{X}_k^* = X_kX_k^* - S_{k,1}S_{k,1}^* = P_k - R_k = \widetilde{P}_k,$$

and

$$\widetilde{S}_k\widetilde{S}_k^* = S_kS_k^* - S_{k,1}S_{k,1}^* = U_k - R_k = \widetilde{U}_k.$$

Since Theorem 7 shows that $\widetilde{U}_k$, $\widetilde{P}_k$, and $\widetilde{Q}_k$ can be used to compute V_k, P_{k+1}, and Q_{k+1}, we can use $\widetilde{S}_k$, $\widetilde{X}_k$, and $\widetilde{Y}_k$ in a factored algorithm to compute T_k, X_{k+1}, and Y_{k+1}.

In stating the factored algorithm we assume that

$$\Sigma_k = I_{p_k-r_k} \oplus -I_{y_k-r_k}.$$

The condition $\|Y_k^*AS_k\|_2 = 1$ signals the need for a deflation. Otherwise, if $\|Y_k^*AS_k\|_2 < 1$, we take $\widetilde{X}_k = X_k$, $\widetilde{Y}_k = Y_k$, and $\widetilde{S}_k = S_k$.

Algorithm 5. Generalized Isometric Arnoldi

Given: X_0, Y_0, and M_j for $j \geqslant 0$

Let $S^{(-1)} = [\,]$ and $\hat{X}_0 = X_0$

For $k = 0, 1, 2, \ldots$

- Let $p_k = \operatorname{rank}(M_k^*\hat{X}_k)$
- Compute W_k such that
  - $M_k^*\hat{X}_kW_k = [B_k \quad 0]$
  - where B_k is $p \times p_k$
- Let $X_k = \hat{X}_kW_k$
- Partition $X_k = [S_k \quad X_{k,2}]$
  - where S_k has p_k columns
- Let $S^{(k)} = [S^{(k-1)} \quad S_k]$
- If $\|Y_k^*AS_k\|_2 = 1$
  - Use Algorithm 4 to compute $\widetilde{X}_k$, $\widetilde{Y}_k$, and $\widetilde{S}_k$.
- Else
  - Let $\widetilde{X}_k = X_k$, $\widetilde{Y}_k = Y_k$, $\widetilde{S}_k = S_k$
- End if
- Compute H_k satisfying $H_k^{\mathrm{H}}\Sigma_kH_k = \Sigma_k$ and
  - $[I_{p_k-r_k} \quad \widetilde{S}_k^*A^*\widetilde{Y}_k]H_k = [C_k \quad 0]$
  - for some $(p_k - r_k) \times (p_k - r_k)$ matrix C_k.
- Let $[T_k \quad Y_{k+1}] = [A\widetilde{S}_k \quad \widetilde{Y}_k]H_k$
- Let $\hat{X}_{k+1} = [T_k \quad X_{k,2}]$

End for

In the particular case in which $y_k = r_k$ it is not necessary to compute H_k or Y_{k+1}. See the note on this case at the end of §4. It is easily verified that in this case we may simply set $T_k = A\widetilde{S}_k$ and continue with the computation of X_{k+1}.

The following theorem states that Algorithm 5 is a factored form of Algorithm 3 in which the columns of X_k, Y_k, S_k, and T_k are orthonormal bases for the images of the projectors P_k, Q_k, U_k, and V_k.

Theorem 8. *Let M_j be a Krylov-like sequence with displacement projectors P_0 and Q_0. Let the projectors be factored as*

$$P_0 = X_0 X_0^* \quad \textit{and} \quad Q_0 = Y_0 Y_0^*,$$

where $X_0^* X_0 = I_{x_k}$ and $Y_0^* Y_0 = I_{y_k}$. Then Algorithm 5 computes sequences X_k and Y_k such that

$$X_k^* X_k = I_{x_k}, \quad Y_k^* Y_k = I_{y_k}, \quad S_k^* S_k = I_{p_k}, \quad \textit{and} \quad T_k^* T_k = I_{p_k - r_k} \tag{33}$$

and

$$P_k = X_k X_k^*, \quad Q_k = Y_k Y_k^*, \quad U_k = S_k S_k^*, \quad \textit{and} \quad V_k = T_k T_k^*, \tag{34}$$

where P_k, Q_k, U_k, and V_k are the projectors computed by Algorithm 3. For each $n \geqslant 0$ the columns of

$$S^{(n)} = [S_0 \quad S_1 \quad S_2 \quad \cdots \quad S_n]$$

have the same span as the columns of

$$M^{(n)} = [M_0 \quad M_1 \quad M_2 \quad \cdots \quad M_n].$$

Proof. To prove the theorem we show that given $\hat{X}_k$ and Y_k satisfying (33) and (34) the algorithm computes S_k, T_k, X_{k+1}, and Y_{k+1} satisfying (33) and (34). The claim for the image of $S^{(n)}$ will then follow as a consequence of Theorem 6.

Recall that $\hat{X}_k$ and X_k have columns that are simply different orthonormal bases for the same subspace. For S_k it is immediate from $\hat{X}_k^* \hat{X}_k = I_{x_k}$ and the fact that W_k is unitary that $S_k^* S_k = I_{p_k}$. To show that $S_k S_k^* = U_k$ we observe that

$$p_k = \operatorname{rank}(U_k) = \operatorname{rank}(\hat{X}_k \hat{X}_k^* M_k) = \operatorname{rank}(M_k^* \hat{X}_k) = \operatorname{rank}(M_k^* S_k)$$

where we have used the fact that $X_{k,2}^* M_k = 0$. Thus the $p \times p_k$ matrix $B_k = M_k^* S_k$ has linearly independent columns and

$$\operatorname{Im}(S_k) = \operatorname{Im}(S_k S_k^* M_k) = \operatorname{Im}(\hat{X}_k \hat{X}_k^* M_k) = \operatorname{Im}(P_k M_k) = \operatorname{Im}(U_k).$$

Together with $S_k^* S_k = I_{p_k}$ this implies that $S_k S_k^*$ is the projector onto $\operatorname{Im}(U_k)$ or equivalently $U_k = S_k S_k^*$.

We have already shown that given X_k, Y_k, and S_k, Algorithm 4 correctly computes $\widetilde{X}_k$, $\widetilde{Y}_k$, and $\widetilde{S}_k$. We now verify the calculation of T_k, X_{k+1}, and Y_{k+1} by describing the structure of H_k. The matrix H_k is invertible with its inverse given by $H_k^{-1} = \Sigma_k H_k^{\mathrm{H}} \Sigma_k$. It follows that C_k is nonsingular. Multiplying the relation that defines C_k by $H_k^{-1} \Sigma_k$ on the right gives

$$[I_{p_k - r_k} \quad -\widetilde{S}_k^* A^* \widetilde{Y}_k] = [C_k \quad 0] H_k^{\mathrm{H}}$$

which implies that H_k has the form

$$H_k = \begin{bmatrix} C_k^{-\mathrm{H}} & H_{k,12} \\ -\widetilde{Y}_k^* A \widetilde{S}_k C_k^{-\mathrm{H}} & H_{k,22} \end{bmatrix}.$$

For T_k we note that the form of H_k together with $[T_k \quad Y_{k+1}] = [A\widetilde{S}_k \quad \widetilde{Y}_k]H_k$ implies that

$$T_k = [A\widetilde{S}_k \quad \widetilde{Y}_k]\begin{bmatrix} I_{p_k-r_k} \\ -\widetilde{Y}_k^* A\widetilde{S}_k \end{bmatrix} C_k^{-\mathrm{H}} = (I - \widetilde{Y}_k\widetilde{Y}_k^*)A\widetilde{S}_k C_k^{-\mathrm{H}}.$$

Thus

$$\mathrm{Im}(T_k) = \mathrm{Im}((I - \widetilde{Y}_k\widetilde{Y}_k^*)A\widetilde{S}_k C_k^{-\mathrm{H}}) = \mathrm{Im}((I - \widetilde{Q}_k)A\widetilde{U}_k) = \mathrm{Im}(V_k).$$

In addition

$$T_k^* T_k = C_k^{-1}\widetilde{S}_k^* A^*(I - \widetilde{Y}_k\widetilde{Y}_k^*)A\widetilde{S}_k C_k^{-\mathrm{H}} = C_k^{-1}(I_{p_k-r_k} - \widetilde{S}_k^* A^*\widetilde{Y}_k\widetilde{Y}_k^* A\widetilde{S}_k)C_k^{-\mathrm{H}}.$$

However the (1, 1) block of the relation $H_k^{\mathrm{H}}\Sigma_k H_k = \Sigma_k$ and the form of the matrix H_k give

$$I_{p_k-r_k} = C_k^{-1}C_k^{-\mathrm{H}} - C_k^{-1}\widetilde{S}_k^* A^*\widetilde{Y}_k\widetilde{Y}_k^* A\widetilde{S}_k C_k^{-\mathrm{H}} = C_k^{-1}(I_{p_k-r_k} - \widetilde{S}_k^* A^*\widetilde{Y}_k\widetilde{Y}_k^* A\widetilde{S}_k)C_k^{-\mathrm{H}}$$

so that $T_k^* T_k = I_{p_k-r_k}$. Thus $T_k T_k^* = V_k$.

For X_{k+1} we note that $\widetilde{X}_k = [\widetilde{S}_k \quad X_{k,2}]$ and $\hat{X}_{k+1} = [T_k \quad X_{k,2}]$ imply that

$$X_{k+1}X_{k+1}^* = \hat{X}_{k+1}\hat{X}_{k+1}^* = \widetilde{X}_k\widetilde{X}_k^* - \widetilde{S}_k\widetilde{S}_k^* + T_k T_k^* = \widetilde{P}_k - \widetilde{U}_k + V_k = P_{k+1}.$$

Thus $X_{k+1}X_{k+1}^* = P_{k+1}$. Since P_{k+1} has rank $x_{k+1} = x_k - r_k$ and X_{k+1} has x_{k+1} columns, the columns of X_{k+1} are linearly independent and X_{k+1} has a left inverse. Since $X_{k+1}X_{k+1}^*$ is a projector we have $P_{k+1} = X_{k+1}X_{k+1}^* = X_{k+1}(X_{k+1}^* X_{k+1})X_{k+1}^*$ which implies $X_{k+1}^* X_{k+1} = I_{x_{k+1}}$.

For Y_{k+1} we note that the relation $H_k\Sigma_k H_k^{\mathrm{H}} = \Sigma_k$ and the relation used to compute Y_{k+1} imply

$$Y_{k+1}Y_{k+1}^* = \widetilde{Y}_k\widetilde{Y}_k^* - A\widetilde{S}_k\widetilde{S}_k^* A^* + T_k T_k^* = \widetilde{Q}_k - A\widetilde{U}_k A^* + V_k = Q_{k+1}.$$

The relation $Y_{k+1}^* Y_{k+1} = I_{y_{k+1}}$ follows in the same way as for X_{k+1}. □

6. Computing a QR factorization

The generalized isometric Arnoldi orthogonalizes the columns of the matrix $M^{(n)}$ to compute $S^{(n)}$. In this section we consider computation of

$$R^{(n)} = S^{(n)^*}M^{(n)} = \begin{bmatrix} R_0^{(n)} \\ R_1^{(n)} \\ \vdots \\ R_n^{(n)} \end{bmatrix}.$$

The blocks of rows $R_k^{(n)}$ are $p_k \times p(n+1)$ for $0 \leqslant k \leqslant n$. If we define

$$p^{(n)} = \sum_{k=0}^{n} p_k = \mathrm{rank}(M^{(n)})$$

then $R^{(n)}$ is an $p^{(n)} \times p(n+1)$. If all the columns of $M^{(n)}$ are linearly independent then $p^{(n)} = p(n+1)$ and $M^{(n)} = S^{(n)}R^{(n)}$ is a QR factorization of $M^{(n)}$.

To compute $R^{(n)}$ we extend Algorithm 5 with recurrences to compute $S_k^* M_j$.

We start by defining

$$E_{k,j} = S_k^* M_j, \quad F_{k,j} = X_{k,2}^* M_j, \quad \text{and} \quad G_{k,j} = Y_k^* A M_{j-1},$$

so that

$$X_k^* M_j = \begin{bmatrix} E_{k,j} \\ F_{k,j} \end{bmatrix}$$

and

$$R_k^{(n)} = S_k^* M^{(n)} = [E_{k,0} \quad E_{k,1} \quad \cdots \quad E_{k,n}].$$

To compute the rows of $R^{(n)}$ we will give a recurrence for computing $E_{k+1,j}$, $F_{k+1,j}$, and $G_{k+1,j}$ from $E_{k,j}$, $F_{k,j}$, and $G_{k,j}$. Note that $E_{k,j}$, $F_{k,j}$, and $G_{k,j}$ are $p_k \times p$, $(x_k - p_k) \times p$, and $y_k \times p$ respectively. Thus the number of rows in these matrices varies with k.

Algorithm 4 deflates X_k and Y_k by applying unitary transformations on the right and removing columns to get $\widetilde{X}_k$ and $\widetilde{Y}_k$. This corresponds in an obvious way to applying unitary transformations on the left and removing rows from $E_{k,j}$ and $G_{k,j}$ Thus

$$\widetilde{E}_{k,j} = \widetilde{S}_k^* M_j, \quad \widetilde{F}_{k,j} = F_{k,j} = X_{k,2}^* M_j, \quad \text{and} \quad \widetilde{G}_{k,j} = \widetilde{Y}_k^* A M_{j-1}.$$

We partition H_k as

$$H_k = \begin{bmatrix} H_{k,11} & H_{k,12} \\ H_{k,21} & H_{k,22} \end{bmatrix},$$

where $H_{k,11}$ is $(p_k - r_k) \times (p_k - r_k)$. Since

$$[A\widetilde{S}_k \quad \widetilde{Y}_k]H_k = [T_k \quad Y_{k+1}]$$

we have

$$\begin{bmatrix} H_{k,11}^{\mathrm{H}} & 0 & H_{k,21}^{\mathrm{H}} \\ 0 & I_{x_k - p_k} & 0 \\ H_{k,12}^{\mathrm{H}} & 0 & H_{k,22}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \widetilde{S}_k^* M_{j-1} \\ X_{k,2}^* M_j \\ \widetilde{Y}_k^* A M_{j-1} \end{bmatrix} = \begin{bmatrix} T_k^* A M_{j-1} \\ X_{k,2}^* M_j \\ Y_{k+1}^* A M_{j-1} \end{bmatrix}.$$

The Krylov-like structure of M_j gives

$$A M_{j-1} = M_j - P_0 M_j + Q_0 A M_{j-1}.$$

By (25) and $P_k U_k = U_k$ we have $P_0(I - Q_k) A U_k = 0$ which implies that $P_0 T_k T_k^* = P_0 V_k = 0$. By (26) we have $Q_0(I - Q_k) A U_k = 0$ so that $Q_0 T_k T_k^* = 0$. Thus $T_k^* A M_{j-1} = T_k^* M_j$ and

$$\begin{bmatrix} H_{k,11}^{\mathrm{H}} & 0 & H_{k,21}^{\mathrm{H}} \\ 0 & I_{x_k - p_k} & 0 \\ H_{k,12}^{\mathrm{H}} & 0 & H_{k,22}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \widetilde{S}_k^* M_{j-1} \\ X_{k,2}^* M_j \\ \widetilde{Y}_k^* A M_{j-1} \end{bmatrix} = \begin{bmatrix} T_k^* M_j \\ X_{k,2}^* M_j \\ Y_{k+1}^* A M_{j-1} \end{bmatrix}.$$

Since $X_{k+1} = [S_{k+1} \quad X_{k+1,2}] = [T_k \quad X_{k,2}]W_{k+1}$ we get

$$\begin{bmatrix} W_{k+1}^{\mathrm{H}} & 0 \\ 0 & I_{y_k - r_k} \end{bmatrix} \begin{bmatrix} H_{k,11}^{\mathrm{H}} & 0 & H_{k,21}^{\mathrm{H}} \\ 0 & I_{x_k - p_k} & 0 \\ H_{k,12}^{\mathrm{H}} & 0 & H_{k,22}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \widetilde{S}_k^* M_{j-1} \\ X_{k,2}^* M_j \\ \widetilde{Y}_k^* A M_{j-1} \end{bmatrix} = \begin{bmatrix} S_{k+1}^* M_j \\ X_{k+1,2}^* M_j \\ Y_{k+1}^* A M_{j-1} \end{bmatrix}$$

or

$$\begin{bmatrix} W_{k+1}^{\mathrm{H}} & 0 \\ 0 & I_{y_k - r_k} \end{bmatrix} \begin{bmatrix} H_{k,11}^{\mathrm{H}} & 0 & H_{k,21}^{\mathrm{H}} \\ 0 & I_{x_k - p_k} & 0 \\ H_{k,12}^{\mathrm{H}} & 0 & H_{k,22}^{\mathrm{H}} \end{bmatrix} \begin{bmatrix} \widetilde{E}_{k,j-1} \\ \widetilde{F}_{k,j} \\ \widetilde{G}_{k,j} \end{bmatrix} = \begin{bmatrix} E_{k+1,j} \\ F_{k+1,j} \\ G_{k+1,j} \end{bmatrix}. \tag{35}$$

This is the desired recurrence for the sequences $F_{k,j}$ and $G_{k,j}$.

The recurrence can be related to the generalized Schur algorithm as follows. Theorem 2 states that

$$M^{(n)*}M^{(n)} - ZM^{(n)^*}M^{(n)}Z^{\mathrm{T}} = M^{(n)*}X_0X_0^*M^{(n)} - ZM^{(n)*}A^*Y_0Y_0^*AM^{(n)}Z^{\mathrm{T}},$$

where Z is the $(n+1)p \times (n+1)p$ block downshift matrix with $p \times p$ blocks. If

$$E_k^{(n)} = [E_{k,0} \quad E_{k,1} \quad \cdots \quad E_{k,n}],$$
$$F_k^{(n)} = [F_{k,0} \quad F_{k,1} \quad \cdots \quad F_{k,n}],$$

and

$$G_k^{(n)} = [G_{k,0} \quad F_{k,1} \quad \cdots \quad F_{k,n}]$$

then

$$M^{(n)^*}M^{(n)} - ZM^{(n)^*}M^{(n)}Z^{\mathrm{T}} = E_0^{(n)H}E_0^{(n)} + F_0^{(n)H}F_0^{(n)} - G_0^{(n)H}G_0^{(n)}.$$

Thus the matrices $E_0^{(n)}$, $F_0^{(n)}$, and $G_0^{(n)}$ are generators, in the sense of [8], for the block Toeplitz-like matrix $M^{(n)*}M^{(n)}$. It can be shown that (35) is the generalized Schur algorithm with the transformations H_k and W_k computed from X_k, Y_k, and the Krylov-like sequence M_j instead of from the generators of $M^{(n)*}M^{(n)}$. Of course there is the important difference that if the matrix $M^{(n)*}M^{(n)}$ is singular then the generalized Schur algorithm fails while the generalized isometric Arnoldi algorithm can continue after a deflation.

7. Preliminary numerical experiments

In order to observe the effect of ill-conditioning on the procedure we compare three methods of orthogonalizing three 20×10 Toeplitz matrices. All numerical experiments were run using Matlab code written by the author on a PC with a Pentium 4 processor. The first matrix T_1 has first column and first row

$$\begin{bmatrix} 1 \\ 1.001 \\ \vdots \\ 1.019 \end{bmatrix}, \quad \text{and} \quad [1 \quad 1.001 \quad 1.002 \quad \cdots \quad 1.009].$$

The matrix has condition number $\kappa_2(T_1) \approx 2.8 \times 10^4$. The second matrix T_2 has elements

$$t_{ij} = e^{-(i-j)^2/25}$$

for $1 \leqslant i \leqslant 20$ and $1 \leqslant j \leqslant 10$ and condition number $\kappa_2(T_2) = 3.1 \times 10^7$. The third matrix is similar to the second but with elements $t_{ij} = e^{-(i-j)^2/50}$ and condition number $\kappa_2(T_3) = 4.0 \times 10^9$.

The first method is Algorithm 5. The second method is Algorithm 5 with the following reorthogonalization step. In the absence of numerical error the matrices X_k and Y_k satisfy $X_k^*X_k = I_{x_k}$ and $Y_k^*Y_k = I_{y_k}$. However, in finite precision, the columns of X_k and Y_k do not remain exactly orthogonal. The reorthogonalization step involves computing QR factorizations $X_k = Q_XR_X$ and $Y_k = Q_YR_Y$ and setting $X_k = Q_X$ and $Y_k = Q_Y$ each time through the main loop of the algorithm.

The last method is based on displacement structure and is described in [8]. The generalized Schur algorithm is applied to generators of the matrix

Table 1
Loss of orthogonality

Matrix	$\kappa_2(T)$	Isometric Arnoldi 1	Isometric Arnoldi 2	Generalized Schur
T_1	2.8×10^4	3.9×10^{-9}	3.9×10^{-10}	1.5×10^{-8}
T_2	3.1×10^7	1.9×10^{-4}	1.6×10^{-11}	6.8×10^{-3}
T_3	4.0×10^9	1.9×10^{-1}	4.7×10^{-10}	1.0×10^0

Table 2
Backward errors

Matrix	$\kappa_2(T)$	Isometric Arnoldi 1	Isometric Arnoldi 2	Generalized Schur
T_1	2.8×10^4	2.7×10^{-11}	3.2×10^{-10}	3.3×10^{-14}
T_2	3.1×10^7	9.1×10^{-9}	2.1×10^{-11}	3.4×10^{-16}
T_3	4.0×10^9	6.8×10^{-6}	2.3×10^{-8}	8.8×10^{-16}

$$\begin{bmatrix} T^{\mathrm{T}}T & T^{\mathrm{T}} \\ T & 0 \end{bmatrix}.$$

The Σ-unitary transformations used in the generalized Schur approach are computed from a fast Cholesky factorization of $T^{\mathrm{T}}T$ while in Algorithm 5 the transformations are computed using inner products. The computational complexity of the algorithms is comparable, in each case $O(mn)$ for an $m \times n$ Toeplitz matrix.

Each algorithm was applied to T_1, T_2, and T_3 to compute a matrix Q with orthonormal columns. Table 1 gives $\|Q^{\mathrm{T}}Q - I\|_2$, the loss of orthogonality of the computed Q, for each of the three algorithms. In each case the factor R in the QR factorization was also computed. We used the generalized Schur algorithm [8] without any modification to compute R. The generalized isometric Arnoldi algorithm was augmented with the recurrences from §6. The relative backward errors $\|QR - T_k\|_2/\|T_k\|_2$ are given in Table 2.

For the second algorithm, the orthonormality of the columns of the computed Q is comparable to what might be expected from modified Gram–Schmidt, which satisfies an error bound $\|Q^{\mathrm{T}}Q - I\|_2 \leqslant cu\kappa_2(A)$ [5] where u is the unit roundoff. The results for the other two algorithms are dramatically worse. In contrast, the generalized Schur algorithm achieves the best backward error as is shown in Table 2. This is not surprising; the generalized Schur algorithm is known to compute a factorization for which the backward error is of the order of the machine precision.

8. Additional topics

We now comment on a few problems that have not been addressed and have been only partially solved. The isometric Arnoldi algorithm can be used to reduce a unitary matrix A, by unitary similarity, to a product of plane rotations The generalized isometric Arnoldi algorithm can be used to reduce A to a slightly more complicated form. In particular under the assumption that the columns of $M^{(n)}$ are linearly independent the matrix $S^{(n)^*}AS^{(n)}$ can be shown to have a structure of the form

$$H^{(n)} = S^{(n)^*}AS^{(n)} = \hat{J}_{n+1}^{\mathrm{H}}\hat{J}_n^{\mathrm{H}}\hat{J}_{n-1}^{\mathrm{H}} \cdots \hat{J}_0^{\mathrm{H}}\hat{G}_0\hat{G}_1 \cdots \hat{G}_n,$$

where

$$\hat{G}_k = I_{pk} \oplus G_k \oplus I_{(n-k)p+x_{n+1}-x_{k+1}}, \quad \text{and} \quad \hat{J}_k = I_{pk} \oplus J_k \oplus I_{p(n+1-k)+x_{n+1}-x_{k+1}}.$$

The transformations G_k and J_k are unitary and are defined in terms of H_k and W_k. The definition of H_k is involved and it does not lead directly to a stable method for computing H_k. Further research is needed into how to represent H_k in terms of plane rotations or Householder transformations.

This suggests an alternative method for computing $R^{(n)}$. If we consider the relation (5) and multiply by the unitary matrix $S^{(n)^*}$ on both sides then we get

$$S^{(n)^*}M_j = H^{(n)}S^{(n)^*}M_{j-1} + S^{(n)^*}(P_0M_j - Q_0AM_{j-1}). \tag{36}$$

The matrix $S^{(n)^*}M_j$ is a block of columns of $R^{(n)} = S^{(n)^*}M^{(n)}$. Multiplication of a vector by $H^{(n)}$ is $O(n)$ if it is implemented as a product of rotations. Except for $H^{(n)}S^{(n)^*}M_{j-1}$ everything on the right hand side of (36) is in $\mathrm{Im}(P_0) \cup \mathrm{Im}(Q_0)$. If this subspace is of low dimension and the Krylov-like sequence M_j is available, then (36) is a fast recurrence for computing the columns of $R^{(n)}$. The backward errors from Table 2 suggest that the recurrences of §6 are not a satisfactory way to compute $R^{(n)}$. Straightforward implementation of (36) have not given better results. However there are numerous variations on the basic recurrences that have not yet been tried. It is also possible that a direct recurrence for least squares solutions would be a better option. This is the subject of ongoing research.

Finally there are a variety of issues surrounding linear dependence in Krylov-like sequences. The generalized isometric Arnoldi algorithm is able to detect and effectively skip over vectors that can be expressed as linear combination of previous vectors in the sequence. This is in striking contrast to the generalized Schur algorithm for which fast Cholesky fails when the columns of $M^{(n)}$ are not linearly independent. The numerical properties of deflation are not clear and merit further investigation.

Acknowledgment

The author would like to thank two anonymous referees for their careful reading and for suggestions that substantially improved this paper.

References

[1] G. Cybenko, The numerical stability of the lattice algorithm for least squares linear prediction problems, BIT 24 (1984) 441–455.
[2] W.B. Gragg, The QR algorithm for unitary Hessenberg matrices, J. Comput. Appl. Math. 16 (1986) 1–8.
[3] W.B. Gragg, Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle (in Russian), in: E.S. Nikolaev (Ed.), Numerical Methods in Linear Algebra, Moscow University Press, 1982, pp. 16–32.
[4] W.B. Gragg, Positive definite Toeplitz matrices, the Arnoldi process for isometric operators, and Gaussian quadrature on the unit circle, J. Comput. Appl. Math. 46 (1993) 183–198 (English translation of [3]).
[5] N.J. Higham, Accuracy and Stability of Numerical Algorithms, second ed., SIAM, Philadelphia, 2002.
[6] C.F. Jagels, L. Reichel, A fast minimal residual algorithm for shifted unitary matrices, Numer. Linear Algebra Appl. 1 (1994) 31–46.
[7] T. Kailath, Linear Systems Theory, Prentice Hall, 1980.
[8] T. Kailath, J. Chun, Generalized displacement structure for block-Toeplitz, Toeplitz-block, and Toeplitz-derived matrices, SIAM J. Matrix Anal. Appl. 15 (1994) 114–128.
[9] G. Szegö, Orthogonal Polynomials, third ed., American Mathematical Society, Providence, Rhode Island, 1967.