SIAM J. Matrix Anal. Appl.
Vol. 26, No. 3, pp. 640–659

SOAR: A SECOND-ORDER ARNOLDI METHOD FOR THE SOLUTION OF THE QUADRATIC EIGENVALUE PROBLEM[*]

ZHAOJUN BAI[†] AND YANGFENG SU[‡]

Abstract. We first introduce a *second-order Krylov subspace* $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ based on a pair of square matrices $\mathbf{A}$ and $\mathbf{B}$ and a vector $\mathbf{u}$. The subspace is spanned by a sequence of vectors defined via a second-order linear homogeneous recurrence relation with coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ and an initial vector $\mathbf{u}$. It generalizes the well-known Krylov subspace $\mathcal{K}_n(\mathbf{A}; \mathbf{v})$, which is spanned by a sequence of vectors defined via a first-order linear homogeneous recurrence relation with a single coefficient matrix $\mathbf{A}$ and an initial vector $\mathbf{v}$. Then we present a second-order Arnoldi (SOAR) procedure for generating an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$. By applying the standard Rayleigh–Ritz orthogonal projection technique, we derive an SOAR method for solving a large-scale quadratic eigenvalue problem (QEP). This method is applied to the QEP directly. Hence it preserves essential structures and properties of the QEP. Numerical examples demonstrate that the SOAR method outperforms convergence behaviors of the Krylov subspace–based Arnoldi method applied to the linearized QEP.

Key words. quadratic eigenvalue problem, second-order Krylov subspace, second-order Arnoldi procedure, Rayleigh–Ritz orthogonal projection

AMS subject classifications. 65F15, 65F30

DOI. 10.1137/S0895479803438523

1. Introduction.

The Krylov subspace

$$\mathcal{K}_n(\mathbf{A}; \mathbf{v}) = \operatorname{span}\{\mathbf{v}, \mathbf{A}\mathbf{v}, \mathbf{A}^2\mathbf{v}, \ldots, \mathbf{A}^{n-1}\mathbf{v}\} \tag{1.1}$$

based on a square matrix $\mathbf{A}$ and a vector $\mathbf{v}$ plays an indispensable role in modern numerical techniques for solving large-scale matrix computation problems, such as the linear eigenvalue problem of the form $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. A Krylov subspace–based method is often the method of choice due to its simplicity, its availability of reliable and efficient processes for generating its orthonormal basis, and the superiority of convergence behaviors [5, 6, 12, 15, 16]. Many state-of-the-art Krylov subspace methods for solving large-scale eigenvalue problems are presented in [3].

The generalized eigenvalue problem of the form $\mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$ must be reduced, explicitly or implicitly, to the linear eigenvalue problem in a form such as $(\mathbf{B}^{-1}\mathbf{A})\mathbf{x} = \lambda\mathbf{x}$, and then a Krylov subspace–based method can be applied. The quadratic eigenvalue problem (QEP) of the form

$$(\lambda^2\mathbf{M} + \lambda\mathbf{D} + \mathbf{K})\mathbf{x} = \mathbf{0} \tag{1.2}$$

is usually processed in two stages, as recommended in most literature, public domain packages, and proprietary software today. At the first stage, it transforms the QEP into an equivalent generalized eigenvalue problem:

$$\mathbf{C}\mathbf{y} = \lambda\mathbf{G}\mathbf{y}, \tag{1.3}$$

[*] Received by the editors December 9, 2003; accepted for publication (in revised form) by H. A. van der Vorst July 2, 2004; published electronically March 3, 2005.
http://www.siam.org/journals/simax/26-3/43852.html

[†] Department of Computer Science and Department of Mathematics, University of California, Davis, CA 95616 (bai@cs.ucdavis.edu). The research of this author was supported in part by the National Science Foundation under grant 0220104.

[‡] Department of Mathematics, Fudan University, Shanghai 200433, China (yfsu@fudan.edu.cn). The research of this author was supported in part by NSFC research project 10001009 and NSFC research key project 90307017.

where $\mathbf{y}^{\mathrm{T}} = \begin{bmatrix} \lambda\mathbf{x}^{\mathrm{T}} & \mathbf{x}^{\mathrm{T}} \end{bmatrix}$, and $\mathbf{C}$ and $\mathbf{G}$ are in forms such as

$$\mathbf{C} = \begin{bmatrix} -\mathbf{D} & -\mathbf{K} \\ \mathbf{I} & \mathbf{0} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix},$$

where we assume throughout the report that $\mathbf{M}$ is nonsingular. At the second stage, it reduces the generalized eigenvalue problem (1.3) to a linear eigenvalue problem "$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$" and then applies a Krylov subspace–based method. Such an approach takes advantages of Krylov subspace–based methods, such as the fast convergence rate and the simultaneous convergence of a group of eigenvalues. However, it also suffers some disadvantages, such as having to solve the generalized eigenvalue problem (1.3) of twice the dimension of the original QEP and, more importantly, the loss of the original structures of the QEP in the process of linearization. For example, when coefficient matrices $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ are symmetric positive definite, the transformed generalized eigenvalue problem (1.3) has to be either intrinsically nonsymmetric, where one of $\mathbf{C}$ and $\mathbf{G}$ has to be nonsymmetric, or symmetric indefinite, where both $\mathbf{C}$ and $\mathbf{G}$ are symmetric but neither will be positive definite. Subsequently, essential spectral properties of the QEP are not guaranteed to be preserved. The reader is referred to [24] for a recent survey on theory, applications, and algorithms of the QEP.

For years, researchers have been studying numerical methods which can be applied to the large-scale QEP directly. In these methods, they do not transform the QEP into an equivalent linear form; instead, they project the QEP onto a properly chosen low-dimensional subspace to reduce to a QEP directly with matrix dimension of lower order. The reduced QEP problem can then be solved by a standard dense matrix technique. The Jacobi–Davidson method [17, 18] is one such method. The method targets one eigenvalue at a time with local convergence versus Krylov subspace methods in which a group of eigenvalues is approximated with global convergence. A direct Krylov-type subspace method with a generalized Arnoldi procedure is briefly described in [13]. However, the procedure presented in [13] in fact does not compute an orthonormal basis of the desired Krylov-type subspace. In [7], Arnoldi- and Lanczos-type processes are developed to construct projections of the QEP. The convergence of these methods is usually slower than a Krylov subspace method applied to the mathematically equivalent linear eigenvalue problem. Finally, a subspace approximation method that uses perturbation theory of the QEP was recently presented in [8]. The success of the method is strongly dependent on the initial approximation, although Rayleigh quotient iteration can be used for acceleration.

Motivated by striking an ideal method which not only can be applied to the QEP directly to preserve the essential structures of the QEP but also achieves the superior global convergence behaviors of Krylov subspace methods via linearization, in this paper, we first introduce a second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ based on a pair of square matrices $\mathbf{A}$ and $\mathbf{B}$ and a vector $\mathbf{u}$. The basis vectors of the subspace are defined via a linear homogeneous recurrence of degree 2 with coefficient matrices $\mathbf{A}$ and $\mathbf{B}$. Consequently, a second-order Arnoldi (SOAR) procedure is presented for generating an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$. As an application of the SOAR procedure, a Rayleigh–Ritz orthogonal projection technique based on $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ is discussed for finding a few of the largest magnitude eigenvalues and the corresponding eigenvectors of the large-scale QEP (1.2). This method is applied to the QEP directly. Hence it preserves essential structures and properties of the QEP. Numerical examples presented in section 5 demonstrate that the new QEP solver outperforms convergence behaviors of the Krylov subspace–based Arnoldi method when applied to the linearized QEP.

In order to solve the large-scale QEP and, more generally, the matrix polynomial eigenvalue problem efficiently, the necessity for the extension of the standard Krylov subspace to explicitly involve more than one matrix has been recognized. In section 2, we will see that the definition of the subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ is a natural extension in the context of solving the QEP by a projection technique. It has been an interesting problem to find a scheme which can efficiently construct an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ that is comparable to the Arnoldi process for generating an orthonormal basis of the standard Krylov subspace $\mathcal{K}_n(\mathbf{A}; \mathbf{u})$. The first procedure presented in this paper is inspired by the work of Su and Craig [22], to which we are gratefully indebted.

The rest of this report is organized as follows. In section 2, we introduce the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ and a simple SOAR procedure for generating an orthonormal basis of the subspace. In section 3, we discuss the possible deflation and breakdown situations of the SOAR procedure, and we present a revised version of the SOAR procedure with deflation and memory saving. In section 4, we present a Rayleigh–Ritz procedure for solving the QEP (1.2). For completeness, we also present the basic Arnoldi method for solving the equivalent generalized eigenvalue problem (1.3). Numerical examples are presented in section 5. Discussion and future work are in section 6.

Throughout the paper, we follow the notational convention commonly used in matrix computation literature. Specifically, we use boldface letters to denote vectors (lower cases) and matrices (upper cases), $\mathbf{I}$ for the identity matrix, $\mathbf{e}_j$ for the jth column of the identity matrix $\mathbf{I}$, and $\mathbf{0}$ for zero vectors and matrices. The dimensions of these vectors and matrices are conformed with dimensions used in the context. We use $\cdot^{\mathrm{T}}$ to denote the transpose. N denotes the order of the original matrix triplet $(\mathbf{M}, \mathbf{D}, \mathbf{K})$ and associated QEP (1.2). $\mathrm{span}\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ and $\mathrm{span}\{\mathbf{Q}\}$ denote the space spanned by the vector sequence $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n$ and the columns of the matrix $\mathbf{Q}$, respectively. $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the 1-norm and 2-norm, respectively, for vector or matrix. $\mathbf{x}(i:j)$, as used in MATLAB, denotes the ith to jth entries of the vector $\mathbf{x}$. $\mathbf{A}(i:j, k:\ell)$ denotes the submatrix of $\mathbf{A}$ by the intersection of rows i to j and columns k to ℓ.

2. A second-order Krylov subspace.

In this section, we first define a generalized Krylov subspace induced by a pair of matrices $\mathbf{A}$ and $\mathbf{B}$ and a vector $\mathbf{u}$. Then we discuss the motivation for such a generalization and present an Arnoldi-like procedure for generating an orthonormal basis of the generalized Krylov subspace.

DEFINITION 2.1. *Let $\mathbf{A}$ and $\mathbf{B}$ be square matrices of order N, and let $\mathbf{u} \neq \mathbf{0}$ be an N vector. Then the sequence*

$$\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n-1}, \tag{2.1}$$

where

$$\begin{aligned} &\mathbf{r}_0 = \mathbf{u}, \\ &\mathbf{r}_1 = \mathbf{A}\mathbf{r}_0, \\ &\mathbf{r}_j = \mathbf{A}\mathbf{r}_{j-1} + \mathbf{B}\mathbf{r}_{j-2} \quad \textit{for } j \geq 2, \end{aligned}$$

is called a second-order Krylov sequence *based on $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{u}$. The space*

$$\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u}) = \mathrm{span}\{\mathbf{r}_0, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{n-1}\}$$

*is called an n*th second-order Krylov subspace.

First, we note that the subspace $\mathcal{G}_n(\mathbf{A},\mathbf{B};\mathbf{u})$ generalizes the standard Krylov subspace $\mathcal{K}_n(\mathbf{A};\mathbf{u})$ in the way that when $\mathbf{B}$ is a zero matrix, the second-order Krylov subspace is the standard Krylov subspace, namely,

$$\mathcal{G}_n(\mathbf{A},\mathbf{0};\mathbf{u}) = \mathcal{K}_n(\mathbf{A};\mathbf{u}).$$

Second, we know that the Krylov subspace $\mathcal{K}_n(\mathbf{A};\mathbf{u})$ has an important characterization in terms of matrix polynomials, which forms a foundation for convergence analysis of a Krylov subspace–based method. There is a similar one for the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A},\mathbf{B};\mathbf{u})$. With the starting vector $\mathbf{u}$, the first few vectors in the second-order Krylov sequence can be written as

$$\begin{aligned}
\mathbf{r}_0 &= \mathbf{u},\\
\mathbf{r}_1 &= \mathbf{A}\mathbf{u},\\
\mathbf{r}_2 &= (\mathbf{A}^2+\mathbf{B})\mathbf{u},\\
\mathbf{r}_3 &= (\mathbf{A}^3+\mathbf{A}\mathbf{B}+\mathbf{B}\mathbf{A})\mathbf{u},\\
\mathbf{r}_4 &= (\mathbf{A}^4+\mathbf{A}^2\mathbf{B}+\mathbf{A}\mathbf{B}\mathbf{A}+\mathbf{B}\mathbf{A}^2+\mathbf{B}^2)\mathbf{u}.
\end{aligned}$$

In general, the jth vector $\mathbf{r}_j$ in the second-order Krylov sequence defined by a linear homogeneous recurrence relation of degree 2 with coefficient matrices $\mathbf{A}$ and $\mathbf{B}$ can be written as

$$\mathbf{r}_j = p_j(\mathbf{A},\mathbf{B})\mathbf{u},$$

where $p_j(\alpha,\beta)$ are polynomials in α and β, defined by the recurrence

$$p_j(\alpha,\beta) = \alpha\cdot p_{j-1}(\alpha,\beta) + \beta\cdot p_{j-2}(\alpha,\beta)$$

with $p_0(\alpha,\beta)\equiv 1$ and $p_1(\alpha,\beta)=\alpha$.

We now discuss the motivation for the definition of the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A},\mathbf{B};\mathbf{u})$ in the context of solving the QEP (1.2). Recall that the QEP (1.2) can be transformed into an equivalent generalized eigenvalue problem (1.3). If one applies a Krylov subspace technique to (1.3), then an associated Krylov subspace would naturally be

$$\mathcal{K}_n(\mathbf{H};\mathbf{v}) = \operatorname{span}\{\mathbf{v},\mathbf{H}\mathbf{v},\mathbf{H}^2\mathbf{v},\ldots,\mathbf{H}^{n-1}\mathbf{v}\}, \tag{2.2}$$

where $\mathbf{v}$ is a starting vector of length $2N$, and

$$\mathbf{H} = \mathbf{G}^{-1}\mathbf{C} = \begin{bmatrix} -\mathbf{M}^{-1}\mathbf{D} & -\mathbf{M}^{-1}\mathbf{K}\\ \mathbf{I} & \mathbf{0}\end{bmatrix}. \tag{2.3}$$

Let $\mathbf{A}=-\mathbf{M}^{-1}\mathbf{D}$, $\mathbf{B}=-\mathbf{M}^{-1}\mathbf{K}$, and $\mathbf{v}=[\mathbf{u}^{\mathrm{T}}\ \ \mathbf{0}]^{\mathrm{T}}$; then we immediately derive that the second-order Krylov vectors $\{\mathbf{r}_j\}$ of length N defined in (2.1) and the standard Krylov vectors $\{\mathbf{H}^j\mathbf{v}\}$ of length $2N$ defined in (2.2) are related as the following form:

$$\begin{bmatrix}\mathbf{r}_j\\ \mathbf{r}_{j-1}\end{bmatrix} = \mathbf{H}^j\mathbf{v} \quad \text{for } j\geq 1. \tag{2.4}$$

In other words, the generalized Krylov sequence $\{\mathbf{r}_j\}$ *defines* the entire standard Krylov sequence based on $\mathbf{H}$ and $\mathbf{v}$. Equation (2.4) indicates that the subspace $\mathcal{G}_j(\mathbf{A},\mathbf{B};\mathbf{u})$ of $\mathcal{R}^N$ should be able to provide sufficient information to let us directly work with the QEP, instead of using the subspace $\mathcal{K}_n(\mathbf{H};\mathbf{v})$ of $\mathcal{R}^{2N}$ for the linearized eigenvalue problem (1.3). We will discuss this further in section 4.

We now turn to the question of how to construct an orthonormal basis $\{\mathbf{q}_i\}$ of $\mathcal{G}_j(\mathbf{A},\mathbf{B};\mathbf{u})$. Namely,

$$\mathrm{span}\{\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_j\} = \mathcal{G}_j(\mathbf{A},\mathbf{B};\mathbf{u}) \quad \text{for } j \geq 1.$$

The following is a procedure to implicitly apply to the sequence of the second-order Krylov vectors $\{\mathbf{r}_j\}$ to generate an orthonormal basis $\{\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_j\}$. Later we will see that it is an Arnoldi-like procedure. We call it an SOAR (second-order Arnoldi) procedure.

ALGORITHM 1. *SOAR procedure.*

```
 1. $\mathbf{q}_1 = \mathbf{u}/\|\mathbf{u}\|_2$
 2. $\mathbf{p}_1 = \mathbf{0}$
 3. for $j = 1, 2, \ldots, n$ do
 4.     $\mathbf{r} = \mathbf{A}\mathbf{q}_j + \mathbf{B}\mathbf{p}_j$
 5.     $\mathbf{s} = \mathbf{q}_j$
 6.     for $i = 1, 2, \ldots, j$ do
 7.         $t_{ij} = \mathbf{q}_i^{\mathrm{T}}\mathbf{r}$
 8.         $\mathbf{r} := \mathbf{r} - \mathbf{q}_i t_{ij}$
 9.         $\mathbf{s} := \mathbf{s} - \mathbf{p}_i t_{ij}$
10.     end for
11.     $t_{j+1\,j} = \|\mathbf{r}\|_2$
12.     if $t_{j+1\,j} = 0$, stop
13.     $\mathbf{q}_{j+1} = \mathbf{r}/t_{j+1\,j}$
14.     $\mathbf{p}_{j+1} = \mathbf{s}/t_{j+1\,j}$
15. end for
```

We note that matrices $\mathbf{A}$ and $\mathbf{B}$ are referenced only via the matrix-vector multiplications in line 4 of the algorithm above. Therefore, it is ideal for large and sparse matrices $\mathbf{A}$ and $\mathbf{B}$. Sparsity or structures of $\mathbf{A}$ and $\mathbf{B}$ can be exploited in the matrix-vector multiplications. This enjoys the same feature as the Arnoldi process for generating an orthonormal basis of the standard Krylov subspace $\mathcal{K}_n$.

The **for**-loop in lines 6–10 is an orthogonalization procedure with respect to the $\{\mathbf{q}_i\}$ vectors. The vector sequence $\{\mathbf{p}_j\}$ is an auxiliary sequence. In section 3, we will present a modified version of the algorithm to remove the requirement of explicit reference of the sequence $\{\mathbf{p}_j\}$. This will reduce the memory requirements by almost half.

Algorithm 1 stops prematurely when the norm of $\mathbf{r}$ computed at line 12 vanishes at a certain step j. In this case, we encounter either *deflation* or *breakdown.* We delay the discussion of deflation and breakdown till the next section.

We now consider basic relations between quantities generated by the algorithm. If $\mathbf{Q}_n$ denotes the $N \times n$ matrix with column vectors $\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_n$, $\mathbf{P}_n$ denotes the $N \times n$ matrix with column vectors $\mathbf{p}_1,\mathbf{p}_2,\ldots,\mathbf{p}_n$, and $\mathbf{T}_n$ denotes the $n \times n$ upper Hessenberg matrix with nonzero entries t_{ij} as defined in the algorithm, then the following relations hold:

$$\mathbf{A}\mathbf{Q}_n + \mathbf{B}\mathbf{P}_n = \mathbf{Q}_n\mathbf{T}_n + \mathbf{q}_{n+1}\mathbf{e}_n^{\mathrm{T}} t_{n+1\,n}, \tag{2.5}$$

$$\mathbf{Q}_n = \mathbf{P}_n\mathbf{T}_n + \mathbf{p}_{n+1}\mathbf{e}_n^{\mathrm{T}} t_{n+1\,n} \tag{2.6}$$

with the orthonormality of the vector sequence $\{\mathbf{q}_1,\mathbf{q}_2,\ldots,\mathbf{q}_n,\mathbf{q}_{n+1}\}$. Let $\widehat{\mathbf{T}}_n$ be an $(n+1)\times n$ upper Hessenberg matrix of the form $\widehat{\mathbf{T}}_n = \left[\begin{smallmatrix}\mathbf{T}_n \\ \mathbf{e}_n^{\mathrm{T}} t_{n+1\,n}\end{smallmatrix}\right]$. Then equations

(2.5) and (2.6) can be rewritten in the compact form

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_n \\ \mathbf{P}_n \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_{n+1} \\ \mathbf{P}_{n+1} \end{bmatrix} \widehat{\mathbf{T}}_n. \tag{2.7}$$

This relation assembles the similarity between the SOAR procedure and the well-known Arnoldi procedure [1]. Let us recall the following Arnoldi procedure for generating an orthonormal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of the Krylov subspace $\mathcal{K}_n(\mathbf{H}; \mathbf{v})$, where $\mathbf{H}$ and $\mathbf{v}$ are defined in (2.4).

Algorithm 2. *Arnoldi procedure.*

```
$\mathbf{v}_1 = \mathbf{v}/\|\mathbf{v}\|_2$
for $j = 1, 2, \ldots, n$ do
    $\mathbf{r} = \mathbf{H}\mathbf{v}_j$
    for $i = 1, 2, \ldots, j$ do
        $h_{ij} = \mathbf{v}_i^{\mathrm{T}}\mathbf{r}$
        $\mathbf{r} := \mathbf{r} - \mathbf{v}_i h_{ij}$
    end for
    $h_{j+1\,j} = \|\mathbf{r}\|_2$
    if $h_{j+1,j} = 0$, breakdown
    $\mathbf{v}_{j+1} = \mathbf{r}/h_{j+1\,j}$
end for
```

If $\mathbf{V}_n$ denotes the $2N \times n$ matrix with column vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ and $\mathbf{H}_n$ denotes the $n \times n$ Hessenberg matrix with nonzero entries h_{ij} as defined in the algorithm, then the Arnoldi procedure can be compactly expressed by the equation

$$\mathbf{H}\,\mathbf{V}_n = \mathbf{V}_n \mathbf{H}_n + \mathbf{v}_{n+1}\mathbf{e}_n^{\mathrm{T}} h_{n+1\,n}$$

or be cast in the form similar to (2.7),

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{V}_n = \mathbf{V}_{n+1}\widehat{\mathbf{H}}_n, \tag{2.8}$$

where $\mathbf{V}_{n+1} = \begin{bmatrix} \mathbf{V}_n & \mathbf{v}_{n+1} \end{bmatrix}$ is a $(2N) \times (n+1)$ orthonormal matrix, and $\widehat{\mathbf{H}}_n = \left[\begin{smallmatrix} \mathbf{H}_n \\ \mathbf{e}_n^{\mathrm{T}} h_{n+1\,n} \end{smallmatrix}\right]$ is a $(n+1) \times n$ upper Hessenberg matrix. By comparing (2.7) and (2.8), we see that the essential difference between the SOAR procedure and the Arnoldi procedure is that in SOAR, the nonzero elements t_{ij} of the $(n+1) \times n$ upper Hessenberg matrix $\widehat{\mathbf{T}}_n$ are chosen to enforce the orthonormality of the vectors $\{\mathbf{q}_j\}$ of dimension N, whereas in Arnoldi, the nonzero elements h_{ij} of $(n+1) \times n$ upper Hessenberg matrix $\widehat{\mathbf{H}}_n$ are determined to ensure the orthonormality of the vectors $\{\mathbf{v}_j\}$ of dimension $2N$. In the next section, we will further exploit the relationship between SOAR and Arnoldi to derive a revised version of the SOAR procedure, which remedies the deflation and saves half the memory requirement and floating point operations.

For the rest of this section, we prove that the vector sequence $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_n\}$ indeed is an orthonormal basis of the second-order Krylov subspace $\mathcal{G}_j(\mathbf{A}, \mathbf{B}; \mathbf{u})$. First, we have the following lemma, which reveals the connection between decomposition characteristics in (2.7) and (2.8) and a related Krylov subspace.

Lemma 2.2. *Let $\mathbf{A}$ be an arbitrary $n \times n$ matrix. Let $\mathbf{V}_{m+1} = \begin{bmatrix} \mathbf{V}_m & \mathbf{v}_{m+1} \end{bmatrix}$ be an $n \times (m+1)$ rectangular matrix that satisfies*

$$\mathbf{A}\mathbf{V}_m = \mathbf{V}_{m+1}\widehat{\mathbf{H}}_m$$

for an $(m+1)\times m$ upper Hessenberg matrix $\widehat{\mathbf{H}}_m$. Then there is an upper triangular matrix $\mathbf{R}_m$ such that

$$\mathbf{V}_m\mathbf{R}_m = \begin{bmatrix} \mathbf{v}_1 & \mathbf{A}\mathbf{v}_1 & \cdots & \mathbf{A}^{m-1}\mathbf{v}_1 \end{bmatrix}. \tag{2.9}$$

Furthermore, if the first $m-1$ subdiagonal elements of $\widehat{\mathbf{H}}_m$ are nonzero, then $\mathbf{R}_m$ is nonsingular and

$$\operatorname{span}\{\mathbf{V}_m\} = \mathcal{K}_m(\mathbf{A}, \mathbf{v}_1). \tag{2.10}$$

Proof. We first prove (2.9) by induction on m. When $m=1$, (2.9) holds with $\mathbf{R}_1 = 1$. Assume that (2.9) holds for $m-1$. Then for m,

$$\begin{aligned}
\begin{bmatrix} \mathbf{v}_1 & \mathbf{A}\mathbf{v}_1 & \cdots & \mathbf{A}^{m-1}\mathbf{v}_1 \end{bmatrix} &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{A}\begin{bmatrix} \mathbf{v}_1 & \mathbf{A}\mathbf{v}_1 & \cdots & \mathbf{A}^{m-2}\mathbf{v}_1 \end{bmatrix} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{v}_1 & \mathbf{A}\mathbf{V}_{m-1}\mathbf{R}_{m-1} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{V}_m\mathbf{e}_1 & \mathbf{V}_m\widehat{\mathbf{H}}_{m-1}\mathbf{R}_{m-1} \end{bmatrix} \\
&= \mathbf{V}_m\begin{bmatrix} \mathbf{e}_1 & \widehat{\mathbf{H}}_{m-1}\mathbf{R}_{m-1} \end{bmatrix} \equiv \mathbf{V}_m\mathbf{R}_m.
\end{aligned}$$

The fact of the upper triangularity of $\mathbf{R}_m$ is immediately followed by its definition. Furthermore, note that the diagonal elements of $\mathbf{R}_m$ are 1 and the products of the first $m-1$ subdiagonal elements of $\widehat{\mathbf{H}}_m$. Therefore, if these subdiagonal elements are nonzero, then $\mathbf{R}_m$ is nonsingular. Finally, (2.10) is established by (2.9) and the nonsingularity of $\mathbf{R}_m$. □

We note that in Lemma 2.2, the column vectors of $\mathbf{V}_n$ span the Krylov subspace $\mathcal{K}_n(\mathbf{A}, \mathbf{v}_1)$ as long as (2.9) is satisfied and $\mathbf{R}_m$ is nonsingular. It is still true even when some of the columns of $\mathbf{V}_n$ are zero vectors. Lemma 2.2 can be viewed as a generalization of the second part of Theorem 1.1 in [21, p. 298]. We will apply this fact when we discuss the deflation in the SOAR procedure. We now prove that Algorithm 1 generates an orthonormal basis of the second-order Krylov subspace $\mathcal{G}_j(\mathbf{A}, \mathbf{B}; \mathbf{u})$.

THEOREM 2.3. *If $t_{j+1,j} \neq 0$ for $j \geq 1$ in Algorithm 1, then the vector sequence $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j\}$ forms an orthonormal basis of the second-order Krylov subspace $\mathcal{G}_j(\mathbf{A}, \mathbf{B}; \mathbf{u})$, i.e.,*

$$\operatorname{span}\{\mathbf{Q}_j\} = \mathcal{G}_j(\mathbf{A}, \mathbf{B}; \mathbf{u}) \quad \textit{for } j \geq 1 \tag{2.11}$$

and $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_k = 0$ if $i \neq k$ and $\mathbf{q}_i^{\mathrm{T}}\mathbf{q}_i = 1$ for $i, k = 1, 2, \ldots, j$.

Proof. Equation (2.11) is established by the following sequence of equalities:

$$\begin{aligned}
\mathcal{G}_j(\mathbf{A}, \mathbf{B}; \mathbf{r}_0) &= \operatorname{span}\{\mathbf{r}_0, \mathbf{r}_1, \ldots, \mathbf{r}_{j-1}\} \\
&= \operatorname{span}\left\{\begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \cdots & \mathbf{r}_{j-1} \\ \mathbf{0} & \mathbf{r}_0 & \cdots & \mathbf{r}_{j-2} \end{bmatrix}\right\} \\
&= \operatorname{span}\left\{\begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{v}_1 & \mathbf{H}\mathbf{v}_1 & \ldots & \mathbf{H}^{j-1}\mathbf{v}_1 \end{bmatrix}\right\} \quad \text{by (2.4)} \\
&= \operatorname{span}\left\{\begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{bmatrix}\mathbf{R}_j\right\} \quad \text{by (2.7) and Lemma 2.2} \\
&= \operatorname{span}\left\{\begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{bmatrix}\right\} \quad \text{by the assumption that } \mathbf{R}_j \text{ is nonsingular} \\
&= \operatorname{span}\{\mathbf{Q}_j\}.
\end{aligned}$$

Finally, the orthogonality of the basis vectors $\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j\}$ is directly obtained from the orthogonalization inner **for**-loop (lines 6–10) and the normalization step at line 13 of the SOAR procedure. □

3. An SOAR procedure.

As we pointed out in the previous section, Algorithm 1 stops prematurely when the norm of $\mathbf{r}$ computed at line 12 vanishes at a certain step j. There are two possible explanations for this. One is that the vector sequence $\{\mathbf{r}_i\}$ for $i = 0, 1, \ldots, j-1$ is linearly dependent, but the double length vector sequence $\{[\,\mathbf{r}_i^{\mathrm{T}}\ \mathbf{r}_{i-1}^{\mathrm{T}}\,]^{\mathrm{T}}\}$ is linearly independent. We call this situation *deflation.* We will show that with a proper treatment, the SOAR procedure can continue. Another possible explanation is that both vector sequences $\{\mathbf{r}_i\}$ and $\{[\,\mathbf{r}_i^{\mathrm{T}}\ \mathbf{r}_{i-1}^{\mathrm{T}}\,]^{\mathrm{T}}\}$ are linearly dependent at a certain step j. In this situation, the SOAR procedure terminates. We call this *breakdown.*

The Arnoldi procedure (Algorithm 2) terminates when the norm of the vector $\mathbf{r}$ computed at line 9 vanishes at a certain step j. It happens when the vector sequence $\{\mathbf{H}^i\mathbf{v}\} = \{[\,\mathbf{r}_i^{\mathrm{T}}\ \mathbf{r}_{i-1}^{\mathrm{T}}\,]^{\mathrm{T}}\}$ for $i = 0, 1, \ldots, j-1$ is linearly dependent. This is known as the breakdown of the Arnoldi procedure.

In this section, we first discuss the deflation and then the breakdown. We will show the connection of breakdowns between the SOAR and Arnoldi procedures.

3.1. Deflation.

We now present the following modified version of Algorithm 1, which remedies the deflation.

ALGORITHM 3. *SOAR procedure with deflation.*

```
 1.  $\mathbf{q}_1 = \mathbf{u}/\|\mathbf{u}\|_2$
 2.  $\mathbf{p}_1 = \mathbf{0}$
 3.  for $j = 1, 2, \ldots, n$ do
 4.      $\mathbf{r} = \mathbf{A}\mathbf{q}_j + \mathbf{B}\mathbf{p}_j$
 5.      $\mathbf{s} = \mathbf{q}_j$
 6.      for $i = 1, 2, \ldots, j$ do
 7.          $t_{ij} = \mathbf{q}_i^{\mathrm{T}}\mathbf{r}$
 8.          $\mathbf{r} := \mathbf{r} - \mathbf{q}_i t_{ij}$
 9.          $\mathbf{s} := \mathbf{s} - \mathbf{p}_i t_{ij}$
10.      end for
11.      $t_{j+1\,j} = \|\mathbf{r}\|_2$
12.      if $t_{j+1\,j} = 0$
13.          if $\mathbf{s} \in \mathrm{span}\{\mathbf{p}_i \mid i : \mathbf{q}_i = 0, 1 \le i \le j\}$
14.              breakdown
15           else    % deflation
16.              reset $t_{j+1\,j} = 1$
17.              $\mathbf{q}_{j+1} = \mathbf{0}$
18.              $\mathbf{p}_{j+1} = \mathbf{s}$
19.          end if
20.      else    % normal case
21.          $\mathbf{q}_{j+1} = \mathbf{r}/t_{j+1\,j}$
22.          $\mathbf{p}_{j+1} = \mathbf{s}/t_{j+1\,j}$
23.      end if
24.  end for
```

We note that in the modified SOAR procedure above, when deflation is detected (line 15), it simply takes $\mathbf{q}_{j+1} = \mathbf{0}$ and sets the scaling element $t_{j+1\,j}$ to a nonzero value (line 16). Then the procedure continues.

Without repeating the discussion in section 2, we state that quantities generated by Algorithm 3 hold the same relations as Algorithm 1, e.g., (2.7) is still true and the vector sequence $\{\mathbf{q}_0, \mathbf{q}_1, \ldots, \mathbf{q}_{n-1}\}$ still spans the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$, except that some of the $\mathbf{q}$ vectors are zero vectors when deflations occur at the corresponding steps. The set of nonzero $\mathbf{q}$ vectors forms an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$.

3.2. Breakdown. Let us discuss the situation where breakdown occurs. We have the following theorem.

THEOREM 3.1. *The SOAR procedure* (*Algorithm* 3) *with matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *and starting vector* $\mathbf{u}$ *breaks down at a certain step* j *if and only if the Arnoldi procedure with matrix* $\mathbf{H}$ *and starting vector* $\mathbf{v}$ *breaks down at the same step* j.

To prove Theorem 3.1, we need the following lemma.

LEMMA 3.2. *For a sequence of linearly independent vectors* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *with partition* $\mathbf{v}_i = \{[\,\mathbf{q}_i^{\mathrm{T}}\ \mathbf{p}_i^{\mathrm{T}}\,]^{\mathrm{T}}\}$, *if there exists a subsequence* $\{\mathbf{q}_{i_1}, \mathbf{q}_{i_2}, \ldots, \mathbf{q}_{i_k}\}$ *of the* $\mathbf{q}$ *vectors that are linearly independent and the remaining vectors are zeros,* $\mathbf{q}_{i_{k+1}} = \mathbf{q}_{i_{k+2}} = \cdots = \mathbf{q}_{i_n} = \mathbf{0}$, *then a vector* $\mathbf{v} = \{[\,\mathbf{0}\ \mathbf{p}^{\mathrm{T}}\,]^{\mathrm{T}}\} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *if and only if* $\mathbf{p} \in \operatorname{span}\{\mathbf{p}_{i_{k+1}}, \mathbf{p}_{i_{k+2}}, \ldots, \mathbf{p}_{i_n}\}$.

Proof. If $\mathbf{v} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$, then there exist scalars α_i, such that $\mathbf{v} = \sum_{i=1}^{n} \alpha_i \mathbf{v}_i$. By the assumption that $\mathbf{v} = \{[\,\mathbf{0}\ \mathbf{p}^{\mathrm{T}}\,]^{\mathrm{T}}\} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ and some zero vectors in the $\mathbf{q}$ vector sequence, we have $\mathbf{0} = \sum_{j=1}^{n} \alpha_j \mathbf{q}_j = \sum_{j=1}^{k} \alpha_{i_j} \mathbf{q}_{i_j}$. Since vectors $\mathbf{q}_{i_1}, \mathbf{q}_{i_2}, \ldots, \mathbf{q}_{i_k}$ are linearly independent, it yields that $\alpha_{i_j} = 0$ for $j = 1, 2, \ldots, k$. Hence $\mathbf{v} = \sum_{j=k+1}^{n} \alpha_{i_j} \mathbf{v}_{i_j}$, which means that $\mathbf{p} = \sum_{j=k+1}^{n} \alpha_{i_j} \mathbf{p}_{i_j}$ or, equivalently, $\mathbf{p} \in \operatorname{span}\{\mathbf{p}_{i_{k+1}}, \mathbf{p}_{i_{k+2}}, \ldots, \mathbf{p}_{i_n}\}$. □

Proof of Theorem 3.1. Let us first consider that the Arnoldi procedure breaks down at a certain step j. This implies that

$$\dim(\mathcal{K}_n(\mathbf{H}, \mathbf{v})) = j \quad \text{and} \quad \mathbf{H}^n \mathbf{v} \in \mathcal{K}_j(\mathbf{H}, \mathbf{v}) \quad \text{for } n \geq j \tag{3.1}$$

From (2.7) and Lemma 2.2, we have

$$\operatorname{span}\left\{ \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{bmatrix} \right\} = \mathcal{K}_j(\mathbf{H}, \mathbf{v}).$$

Since $\dim(\mathcal{K}_j(\mathbf{H}, \mathbf{v})) = j$, $\left[\begin{smallmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{smallmatrix}\right]$ is full column rank. By Lemma 2.2 again and (3.1), we have

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \in \operatorname{span}\left\{ \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{bmatrix} \right\}. \tag{3.2}$$

We now show that $\mathbf{r} = \mathbf{0}$ (at line 11 of Algorithm 3). Suppose $\mathbf{r} \neq \mathbf{0}$. Since $\mathbf{r}^{\mathrm{T}} \mathbf{q}_i = 0$ for $i = 1, 2, \ldots, j$, it implies that

$$\mathbf{r} \notin \operatorname{span}\{\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j\},$$

which indicates that

$$\begin{bmatrix} \mathbf{r} \\ \mathbf{s} \end{bmatrix} \notin \operatorname{span}\left\{ \begin{bmatrix} \mathbf{Q}_j \\ \mathbf{P}_j \end{bmatrix} \right\}.$$

This contradicts (3.2). Therefore $\mathbf{r} = \mathbf{0}$. Thus Algorithm 3 proceeds to execute line 13. By (3.2) and Lemma 3.2, we have

$$\mathbf{s} \in \operatorname{span}\{\mathbf{p}_i \mid i : \mathbf{q}_i = \mathbf{0}, 1 \leq i \leq j\}.$$

Therefore, Algorithm 3 also breaks down (line 14 of Algorithm 3).

Conversely, if Algorithm 3 breaks down at a certain step j, then

$$\begin{bmatrix}\mathbf{r}\\ \mathbf{s}\end{bmatrix} = \begin{bmatrix}\mathbf{0}\\ \mathbf{s}\end{bmatrix} \in \operatorname{span}\left\{\begin{bmatrix}\mathbf{Q}_j\\ \mathbf{P}_j\end{bmatrix}\right\}. \tag{3.3}$$

Note that (2.7) still holds with the choice of $t_{j+1\,j} = 1$. Thus by Lemma 2.2, we have

$$\operatorname{span}\left\{\begin{bmatrix}\mathbf{Q}_j\\ \mathbf{P}_j\end{bmatrix}\right\} = \mathcal{K}_j(\mathbf{H},\mathbf{v}) \quad \text{and} \quad \operatorname{span}\left\{\begin{bmatrix}\mathbf{Q}_{j+1}\\ \mathbf{P}_{j+1}\end{bmatrix}\right\} = \mathcal{K}_{j+1}(\mathbf{H},\mathbf{v}).$$

On the other hand, by induction, we can show that after $j-1$ steps in Algorithm 3, we have

$$\operatorname{rank}\left(\begin{bmatrix}\mathbf{Q}_j\\ \mathbf{P}_j\end{bmatrix}\right) = j. \tag{3.4}$$

Combining (3.3) and (3.4), it yields that $\dim(\mathcal{K}_j(\mathbf{H},\mathbf{v})) = j$ and $\mathcal{K}_j(\mathbf{H},\mathbf{v}) = \mathcal{K}_{j+1}$ $(\mathbf{H},\mathbf{v})$. These two conditions ensure that the Arnoldi procedure breaks down at the same step j. □

In the Arnoldi procedure, when breakdown occurs, it indicates that the Krylov subspace $\mathcal{K}_j(\mathbf{H},\mathbf{v})$ is an invariant subspace of $\mathbf{H}$, and the vector sequence $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j\}$ is an orthonormal basis of the subspace. It is regarded as a lucky breakdown. For the SOAR procedure (Algorithm 3), by (2.7) we know that the column vectors of the $2N \times j$ matrix $[\begin{smallmatrix}\mathbf{Q}_j\\ \mathbf{P}_j\end{smallmatrix}]$ also span an invariant subspace of $\mathbf{H}$, but it is not an orthonormal basis.

3.3. An SOAR procedure. Now we further exploit the relations in Algorithm 3 to derive a new version, which avoids the explicit references and updates of the $\mathbf{p}$ vectors at lines 9 and 22. The resulting algorithm reduces memory requirement by almost half.

First, by (2.6) and noting that $\mathbf{p}_1 = \mathbf{0}$, we have

$$\mathbf{Q}_n = \mathbf{P}_{n+1}\widehat{\mathbf{T}}_n = \mathbf{P}_{n+1}(:, 2:n+1) \cdot \widehat{\mathbf{T}}_n(2:n+1, 1:n).$$

Then (2.5) can be rewritten as

$$\mathbf{A}\mathbf{Q}_n + \mathbf{B}\mathbf{Q}_n\mathbf{S}_n = \mathbf{Q}_n\mathbf{T}_n + \mathbf{q}_{n+1}\mathbf{e}_n^{\mathrm{T}} t_{n+1\,n}, \tag{3.5}$$

where $\mathbf{S}_n$ is an $n \times n$ strictly upper triangular matrix of the form

$$\mathbf{S}_n = \begin{bmatrix}\mathbf{0} & \widehat{\mathbf{T}}_n(2:n, 1:n-1)^{-1}\\ 0 & \mathbf{0}\end{bmatrix}.$$

Equation (3.5) suggests a method for computing vector $\mathbf{q}_{j+1}$ from $\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_j$. This leads to the following algorithm, which needs only about a half of the memory and floating point operations of Algorithm 3.

Algorithm 4. *SOAR procedure with deflation and memory saving.*

```
 1. $\mathbf{q}_1 = \mathbf{u}/\|\mathbf{u}\|_2$
 2. $\mathbf{f} = \mathbf{0}$
 3. for $j = 1, 2, \ldots, n$ do
 4.     $\mathbf{r} = \mathbf{A}\mathbf{q}_j + \mathbf{B}\mathbf{f}$
 5.     for $i = 1, 2, \ldots, j$ do
 6.         $t_{ij} = \mathbf{q}_i^{\mathrm{T}}\mathbf{r}$
 7.         $\mathbf{r} := \mathbf{r} - \mathbf{q}_i t_{ij}$
 8.     end for
 9.     $t_{j+1\,j} = \|\mathbf{r}\|_2$
10.     if $t_{j+1\,j} \neq 0$,
11.         $\mathbf{q}_{j+1} := \mathbf{r}/t_{j+1\,j}$
12.         $\mathbf{f} = \mathbf{Q}_j \widehat{\mathbf{T}}(2 : j+1, 1 : j)^{-1}\mathbf{e}_j$
13.     else
14.         reset $t_{j+1\,j} = 1$
15.         $\mathbf{q}_{j+1} = \mathbf{0}$
16.         $\mathbf{f} = \mathbf{Q}_j \widehat{\mathbf{T}}(2 : j+1, 1 : j)^{-1}\mathbf{e}_j$
17.         save $\mathbf{f}$ and check deflation and breakdown
18.     end if
19. end for
```

Note that at line 17 of the algorithm above, if $\mathbf{f}$ belongs to the subspace spanned by previously saved $\mathbf{f}$ vectors, then the algorithm encounters breakdown and terminates. Otherwise, there is a deflation at step j; after setting $t_{j+1\,j}$ to 1 or any nonzero constant, the algorithm continues. Those saved $\mathbf{f}$ vectors are the $\mathbf{p}_i$ vectors corresponding to the vector $\mathbf{q}_i = \mathbf{0}$ in Algorithm 3. To check whether $\mathbf{f}$ is in the subspace spanned by the previously saved $\mathbf{f}$, we can use a modified Gram–Schmidt procedure [21]. It is not necessary to use extra storage to save those $\mathbf{f}$ vectors. They can be stored at the columns of $\mathbf{Q}_n$ where the corresponding $\mathbf{q}_i = \mathbf{0}$.

4. A projection method applied directly to the QEP.

In this section, we apply the concept of the second-order Krylov subspace and its orthonormal basis generated by the SOAR procedure to develop a projection technique to solve the QEP (1.2). We follow the orthogonal Rayleigh–Ritz approximation procedure to derive a method which approximates a large-scale QEP by a small-scale QEP.

Following the standard derivation, to apply the Rayleigh–Ritz approximation technique based on the subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ with $\mathbf{A} = -\mathbf{M}^{-1}\mathbf{D}$ and $\mathbf{B} = -\mathbf{M}^{-1}\mathbf{K}$, we seek an approximate eigenpair $(\theta, \mathbf{z})$, where $\theta \in \mathcal{C}$ and $\mathbf{z} \in \mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$, by imposing the following orthogonal condition, also called the Galerkin condition:

$$(\theta^2\mathbf{M} + \theta\mathbf{D} + \mathbf{K})\mathbf{z} \;\perp\; \mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$$

or, equivalently,

$$\mathbf{v}^{\mathrm{T}}(\theta^2\mathbf{M} + \theta\mathbf{D} + \mathbf{K})\mathbf{z} = 0 \quad \text{for all } \mathbf{v} \in \mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u}). \tag{4.1}$$

Since $\mathbf{z} \in \mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$, it can be written as

$$\mathbf{z} = \mathbf{Q}_m\mathbf{g}, \tag{4.2}$$

where the $N \times m$ matrix $\mathbf{Q}_m$ is an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ generated by the SOAR procedure (Algorithm 4), and $\mathbf{g}$ is an m vector and $m \leq n$. When there are deflations, $m < n$. By (4.1) and (4.2), it yields that θ and $\mathbf{g}$ must satisfy the reduced QEP:

$$(\theta^2 \mathbf{M}_m + \theta \mathbf{D}_m + \mathbf{K}_m)\mathbf{g} = \mathbf{0} \tag{4.3}$$

with

$$\mathbf{M}_m = \mathbf{Q}_m^{\mathrm{T}} \mathbf{M} \mathbf{Q}_m, \quad \mathbf{D}_m = \mathbf{Q}_m^{\mathrm{T}} \mathbf{D} \mathbf{Q}_m, \quad \mathbf{K}_m = \mathbf{Q}_m^{\mathrm{T}} \mathbf{K} \mathbf{Q}_m. \tag{4.4}$$

The eigenpairs $(\theta, \mathbf{g})$ of (4.3) define the *Ritz pairs* $(\theta,\ \mathbf{z})$. The Ritz pairs are approximate eigenpairs of the QEP (1.2). The accuracy of the approximate eigenpairs $(\theta, \mathbf{z})$ can be assessed by the norms of the residual vectors $(\theta^2 \mathbf{M} + \theta \mathbf{D} + \mathbf{K})\mathbf{z}$.

We note that by explicitly formulating the matrices $\mathbf{M}_m$, $\mathbf{D}_m$, and $\mathbf{K}_m$, essential structures of $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ are preserved. For example, if $\mathbf{M}$ is symmetric positive definite, so is $\mathbf{M}_m$. As a result, essential spectral properties of the QEP will be preserved. For example, if the QEP is a gyroscopic dynamical system in which $\mathbf{M}$ and $\mathbf{K}$ are symmetric, one of them is positive definite, and $\mathbf{D}$ is skew-symmetric, then the reduced QEP is also a gyroscopic system. It is known that in this case, the eigenvalues λ are symmetrically placed with respect to both the real and imaginary axes [10]. Such a spectral property will be preserved in the reduced QEP.

The following algorithm is a high-level description of the Rayleigh–Ritz projection procedure based on $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ for solving the QEP (1.2) directly.

Algorithm 5. *SOAR method for solving the QEP directly.*

1. *Run the SOAR procedure* (*Algorithm* 4) *with* $\mathbf{A} = -\mathbf{M}^{-1}\mathbf{D}$ *and* $\mathbf{B} = -\mathbf{M}^{-1}\mathbf{K}$ *and a starting vector* $\mathbf{u}$ *to generate an* $N \times m$ *orthogonal matrix* $\mathbf{Q}_m$ *whose columns span an orthonormal basis of* $\mathcal{G}_n\,(\mathbf{A}, \mathbf{B}; \mathbf{u})$.
2. *Compute* $\mathbf{M}_m$, $\mathbf{D}_m$, *and* $\mathbf{K}_m$ *as defined in* (4.4).
3. *Solve the reduced QEP* (4.3) *for* $(\theta, \mathbf{g})$ *and obtain the Ritz pairs* $(\theta, \mathbf{z})$, *where* $\mathbf{z} = \mathbf{Q}_m \mathbf{g} / \|\mathbf{Q}_m \mathbf{g}\|_2$.
4. *Test the accuracy of Ritz pairs* $(\theta,\ \mathbf{z})$ *as approximate eigenvalues and eigenvectors of the QEP* (1.2) *by the relative norms of residual vectors:*

$$\frac{\|(\theta^2 \mathbf{M} + \theta \mathbf{D} + \mathbf{K})\mathbf{z}\|_2}{|\theta|^2 \|\mathbf{M}\|_1 + |\theta| \|\mathbf{D}\|_1 + \|\mathbf{K}\|_1}. \tag{4.5}$$

A few remarks are in order:

- At step 1, the matrix-vector product operations $-\mathbf{M}^{-1}\mathbf{D}\mathbf{u}$ and $-\mathbf{M}^{-1}\mathbf{K}\mathbf{u}$ for an arbitrary vector $\mathbf{u}$ must be provided to run the SOAR procedure (Algorithm 4). A factorized form of $\mathbf{M}$, such as the LU factorization, should be made available outside of the first **for**-loop of Algorithm 4 for computational efficiency.
- At step 2, the orthonormal basis matrix $\mathbf{Q}_m$ computed in step 1 is used to explicitly compute the projection matrices $\mathbf{M}_n$, $\mathbf{D}_n$, and $\mathbf{K}_n$. This can be done by using matrix-vector product operations $\mathbf{M}\mathbf{q}$, $\mathbf{D}\mathbf{q}$, and $\mathbf{K}\mathbf{q}$ for an arbitrary vector $\mathbf{q}$. This is an overhead comparison of the method based on the Arnoldi procedure, in which the projection of the matrix is obtained as a by-product without any additional cost (see Algorithm 6 below). This overhead could be significant in some applications. However, this is a numerically better way to use the computed orthonormal basis $\mathbf{Q}_m$ since we can preserve the structures of coefficient matrices as we discussed early. Structure preservation often outweighs the extra cost of floating point operations in the modern

computing environment. For the numerical examples, presented in the next section, we observed that this step takes a small fraction of the total work, due to extreme sparsity of the matrices $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ in practical problems we encountered. The bottleneck of computational costs is often associated with the matrix-vector multiplication operations involving $\mathbf{M}^{-1}$ at step 1.

- At step 3, to solve the small QEP (4.3), we transform it to a generalized eigenvalue problem in the form of (1.3) and then use a dense matrix method, such as the QZ algorithm [5, 6], to find all eigenvalues and eigenvectors $(\theta, \mathbf{g})$ of the small QEP.
- At step 4, we use the relative residual norms (4.5) as the accuracy assessment to indicate the backward errors of the approximate eigenpairs $(\theta, \mathbf{z})$. The discussion of forward errors of approximate eigenvalues and eigenvectors is beyond the scope of this report; the interested reader is referred to [11, 23, 24].

Let us review the basic Arnoldi method for solving the QEP (1.2) based on linearization (1.3). At this stage of our study, we are concerned only with the fundamental properties and behaviors of the SOAR method. It is implemented in a straightforward way as outlined in Algorithm 5. Therefore, we will compare the SOAR method with the following simple implementation of the Arnoldi method for solving the QEP via linearization.

Algorithm 6. *Basic Arnoldi method for linearized QEP.*

1. *Transform the QEP* (1.2) *into the equivalent generalized eigenvalue problem* (1.3).
2. *Run the Arnoldi procedure* (*Algorithm* 2) *with the matrix* $\mathbf{H} = \mathbf{G}^{-1}\mathbf{C}$ *and the vector* $\mathbf{v} = [\,\mathbf{u}^{\mathrm{T}}\ \mathbf{0}\,]^{\mathrm{T}}$ *to generate an orthonormal basis* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ *of the Krylov subspace* $\mathcal{K}_n(\mathbf{H}; \mathbf{v})$. *Let* $\mathbf{V}_n = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$.
3. *Solve the reduced eigenvalue problem*
$$(\mathbf{V}_n^{\mathrm{T}}\mathbf{H}\mathbf{V}_n)\mathbf{t} = \theta\mathbf{t}$$
and obtain the Ritz pairs $(\theta, \mathbf{y})$ *of the eigenvalue problem of the single matrix* $\mathbf{H}$, *where* $\mathbf{y} = \mathbf{V}_n\mathbf{t}$. *Note that by* (2.8), $\mathbf{V}_n^{\mathrm{T}}\mathbf{H}\mathbf{V}_n = \mathbf{H}_n(1:n, 1:n)$ *is an* $n \times n$ *upper Hessenberg matrix returned directly from the Arnoldi procedure without additional cost.*
4. *Extract the approximate eigenpairs* $(\theta,\ \mathbf{z})$ *of the QEP* (1.2) *and test their accuracy by the residual norms as described in* (4.5), *where* $\mathbf{z} = \mathbf{y}(N+1 : 2N)/\|\mathbf{y}(N+1:2N)\|_2$.

Finally, we discuss a hybrid method of the SOAR method (Algorithm 5) and the Arnoldi method (Algorithm 6) to solve the QEP directly. This method provides a good verification for the SOAR method. Let $\mathbf{K}_n$ denote the matrix of the explicit Krylov basis of $\mathcal{K}_n(\mathbf{H}, \mathbf{v})$:

$$\mathbf{K}_n = [\ \mathbf{v}\ \ \mathbf{H}\mathbf{v}\ \ \mathbf{H}^2\mathbf{v}\ \ \cdots\ \ \mathbf{H}^{n-1}\mathbf{v}\ ].$$

Then it is well known (for example, see [21, section 5.1]) that $\mathbf{V}_n$, generated by the Arnoldi procedure with $\mathbf{H}$ and $\mathbf{v}$, is the Q-factor of the QR factorization of $\mathbf{K}_n$:

$$\mathbf{K}_n = \mathbf{V}_n\mathbf{R}_n.$$

By (2.4), the equation above can be written in the form

$$\begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \cdots & \mathbf{r}_{n-1} \\ \mathbf{0} & \mathbf{r}_0 & \cdots & \mathbf{r}_{n-2} \end{bmatrix} = \begin{bmatrix} \mathbf{V}_n^{(1)} \\ \mathbf{V}_n^{(2)} \end{bmatrix} \mathbf{R}_n,$$

where $\mathbf{V}_n$ is partitioned into a 2×1 block matrix with $N \times n$ subblocks $\mathbf{V}_n^{(1)}$ and $\mathbf{V}_n^{(2)}$. From the first N rows of the previous equation, we have

$$\begin{bmatrix} \mathbf{r}_0 & \mathbf{r}_1 & \cdots & \mathbf{r}_{n-1} \end{bmatrix} = \mathbf{V}_n^{(1)}\mathbf{R}_n. \tag{4.6}$$

Hence, we have

$$\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u}) = \operatorname{span}\{\mathbf{V}_n^{(1)}\}.$$

Therefore, an alternative way to generate an orthonormal basis of $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ is to first run the Arnoldi procedure with $2N \times 2N$ matrix $\mathbf{H}$ and starting vector $\mathbf{v} = [\,\mathbf{u}^{\mathrm{T}} \;\; \mathbf{0}]^{\mathrm{T}}$, then orthonormalize the first block $\mathbf{V}_n^{(1)}$ of $\mathbf{V}_n$ to obtain an orthonormal basis of the projection subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$. This method provides a good verification for the SOAR method, although it is expensive in terms of memory and computational requirements. For numerical results presented in the next section, we observed that the convergence rate and behaviors of this method and the SOAR method are essentially the same.

5. Numerical examples.

In this section, we present numerical examples to demonstrate the promises of the SOAR method (Algorithm 5) for solving the QEP (1.2). Following the discussion presented in the previous sections, we focus on the illustration of the fundamental properties of the SOAR method in terms of the following two aspects:

1. The convergence behaviors of the SOAR method applied directly to the QEP are generally comparable to the Arnoldi method applied to the linearized QEP. Specifically,
   (a) eigenvalues with the largest magnitude converge first;
   (b) the convergence rate of the SOAR method is at least as fast as the Arnoldi method.
2. The SOAR method preserves the essential structures of the QEP, such as symmetry and positive definiteness in coefficient matrices $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$. As a result, we should expect the preservation of spectral properties of the large QEP (1.2) in the reduced QEP (4.3).

In the following examples, the starting vector $\mathbf{u}$ of the SOAR method is chosen as a vector with all components equal to 1. $\mathbf{v} = [\mathbf{u}^{\mathrm{T}} \;\; \mathbf{0}]^{\mathrm{T}}$ is used as the starting vector of the Arnoldi-based methods (Algorithms 6 and 7). The so-called *exact* eigenvalues of the QEP are computed by the dense method, namely, the QZ method for computing all eigenvalues and eigenvectors of the generalized eigenvalue problem (1.3). The deflation and breakdown thresholds are set to be the same, namely, 10^{-10}. In fact, with this threshold, deflation and breakdown were detected only in Example 1.

Example 1. This example shows the deflation and breakdown phenomena in the SOAR procedure (Algorithm 4). The matrices $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ are from the modeling of a simple vibrating spring-mass system with damping in linear connection [5, 9]. $\mathbf{M}$ and $\mathbf{D}$ are diagonal matrices, and $\mathbf{K}$ is tridiagonal. For this particular run, we choose 50×50 matrices, where $\mathbf{M} = 0.1 \times \mathbf{I}$, $\mathbf{D} = \mathbf{I}$, and

$$\mathbf{K} = \begin{bmatrix} 0.2 & -0.1 & & & \\ -0.1 & 0.2 & -0.1 & & \\ & \ddots & \ddots & \ddots & \\ & & -0.1 & 0.2 & -0.1 \\ & & & -0.1 & 0.1 \end{bmatrix}.$$

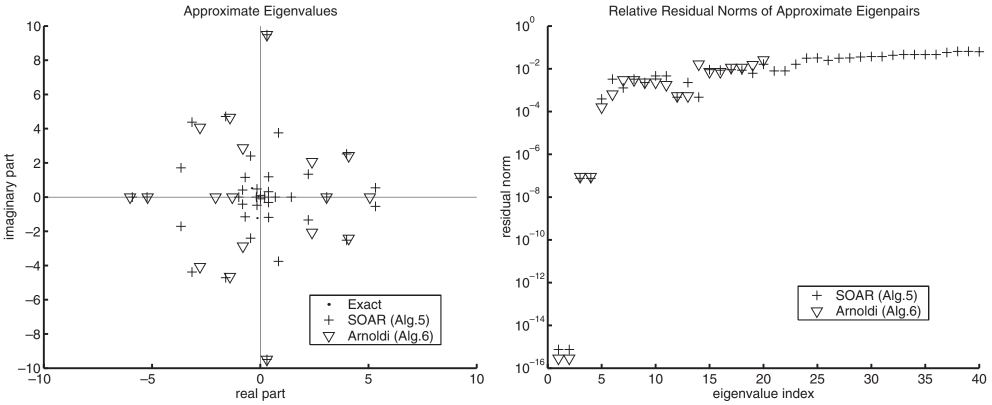

FIG. 5.1. *Random nonsymmetric QEP; exact and approximate eigenvalues* (*left*), *and relative residual norms* (*right*) (*Example* 2).

This example illustrates the following two main issues:

1. Deflation occurs at every even step of the SOAR procedure, i.e., $\mathbf{q}_j = 0$ for all even number j.
2. Suppose the starting vector $\mathbf{u}$ is chosen as a linear combination of κ eigenvectors of the matrix $\mathbf{K}$ corresponding to the κ eigenvalues closest to 0. For $\kappa = 1, 2, 3$, both the SOAR procedure (Algorithm 4) and the Arnoldi procedure (Algorithm 2) break down at steps $j = 2\kappa$. However, for large κ, breakdown has not been detected due to numerical noises.

Example 2. The purpose of this example is to show that the convergence behaviors of the SOAR and Arnoldi methods are generally the same for a "general" QEP. Let $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ be 200×200 random nonsymmetric matrices. Elements of these matrices are chosen from a normal distribution with mean zero, variance one, and standard deviation one. The left plot of Figure 5.1 shows the partial approximate eigenvalues computed by two methods with the reduced dimension $n = 20$. The right plot of Figure 5.1 shows the relative residual norms. This example shows that the convergence behaviors of the two methods are essentially the same, as we expected.

Example 3. As in Example 2, this example is to show that the convergence rates of the SOAR and Arnoldi methods are comparable. However, only the SOAR method preserves the essential properties of the QEP. Specifically, $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ are chosen as 200×200 random matrices with the elements chosen from a normal distribution with mean zero, variance one, and standard deviation one. Furthermore, $\mathbf{M}$ is symmetric positive definite, $\mathbf{D}$ is skew-symmetric, and $\mathbf{K}$ is symmetric negative definite, as one encounters in a gyroscopic dynamical system. The gyroscopic system is a widely studied system. There are many interesting properties associated with such a system. For example, it is known that the distribution of the eigenvalues of the system in the complex plane is symmetric with respect to both the real and imaginary axes. The left plot of Figure 5.2 shows the approximate eigenvalues computed by two algorithms with $n = 20$. The right plot of Figure 5.2 shows the relative residual norms. This example shows that the SOAR method (Algorithm 5) preserves the gyroscopic spectral property. Furthermore, the residual norms indicate that the SOAR method has a slightly better convergence rate.

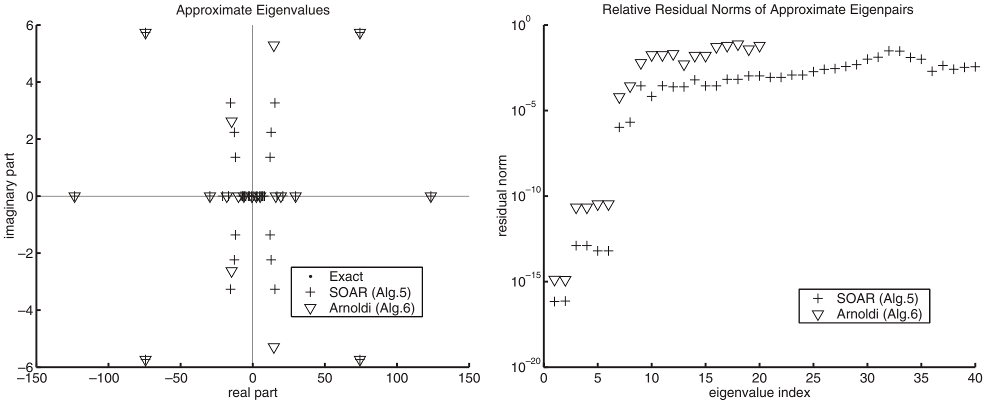

FIG. 5.2. *Random gyroscopic QEP; exact and approximate eigenvalues* (*left*) *and relative residual norms* (*right*) (*Example* 3).

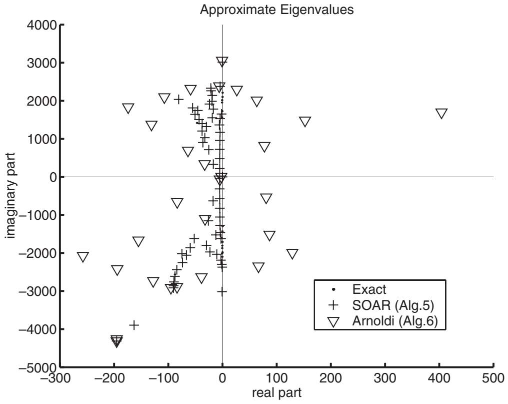

FIG. 5.3. *Acoustic QEP; exact and approximate eigenvalues* (*Example* 4).

Example 4. This is a QEP encountered in modeling the propagation of sound waves in a room in which one wall was made of a sound-absorbing material. This is a scaled-down version of the test problem as presented in [18]. The matrices $\mathbf{M}$, $\mathbf{D}$, and $\mathbf{K}$ are of order 1331. Furthermore, $\mathbf{M}$ and $\mathbf{K}$ are real symmetric positive definite, and $\mathbf{D}$ is complex non-Hermitian. The largest magnitude eigenvalue computed by the standard dense matrix method (for all eigenvalues) and by the SOAR and Arnoldi methods with $n = 30$ are

$$\lambda_{\max} = -1.952652244810165 \times 10^2 - 4.314162072894026 \times 10^3\mathtt{i} \text{ (“exact”)},$$
$$\lambda_{\max}^{\mathrm{S}} = -\underline{1.9526522448}09287 \times 10^2 - \underline{4.314162072894}454 \times 10^3\mathtt{i} \text{ (SOAR)},$$
$$\lambda_{\max}^{\mathrm{A}} = -\underline{1.9526522}50694968 \times 10^2 - \underline{4.314162072}541710 \times 10^3\mathtt{i} \text{ (Arnoldi)}.$$

We observed that both the SOAR and Arnoldi methods converge to the largest magnitude eigenvalue first. The relative errors are $|\lambda_{\max}^{\mathrm{S}} - \lambda_{\max}|/|\lambda_{\max}| = 2.64 \times 10^{-12}$ and $|\lambda_{\max}^{\mathrm{A}} - \lambda_{\max}|/|\lambda_{\max}| = 1.95 \times 10^{-8}$, respectively. The largest magnitude eigenvalues produced by the SOAR method (Algorithm 5) are more accurate than the Arnoldi method (Algorithm 6). Furthermore, Figure 5.3 shows that all eigenvalues of the

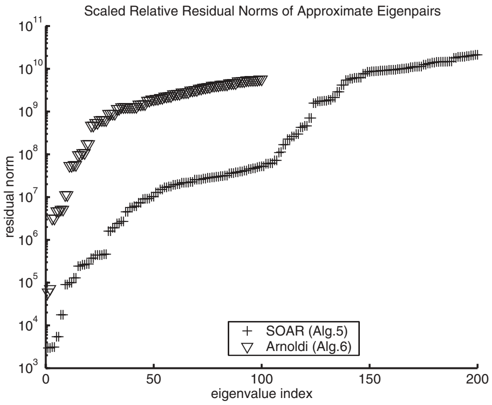

FIG. 5.4. *Scaled relative residual norms of Example* 5.

original QEP are distributed in the left half of the complex plane, known as stable eigenvalues. The reduced QEP by the SOAR method inherits such a property in the process of approximation. On the other hand, the linearized QEP used in the Arnoldi method loses this important property.

Example 5. This is a QEP problem from the NASTRAN simulation of a fluid-structure coupling cylinder model with both acoustic elements and structure elements. The order of the matrices **M**, **D**, and **K** is $N = 3600$. The following table is a profile of other properties of the matrix triplet. The last column is an estimated lower bound for the 1-norm condition number using MATLAB's `condest` function.

	Nonzeros	Symmetry	Pos.def.	1-norm	Cond.est
M	5521	yes	no	36.00	Inf
D	19570	yes	no	1.025	Inf
K	59062	yes	no	2.19×10^{12}	8.42×10^{16}

We solved the shift-and-invert QEP

$$(\mu^2 \widehat{\mathbf{M}} + \mu \widehat{\mathbf{D}} + \widehat{\mathbf{K}})\mathbf{x} = \mathbf{0}, \tag{5.1}$$

where $\mu = 1/(\lambda - \sigma)$, $\widehat{\mathbf{M}} = \sigma^2 \mathbf{M} + \sigma \mathbf{D} + \mathbf{K}$, $\widehat{\mathbf{D}} = \mathbf{D} + 2\sigma \mathbf{M}$, and $\widehat{\mathbf{K}} = \mathbf{M}$. The largest (in modulus) eigenvalue μ approximates the eigenvalues λ of the original QEP closest to the shift σ. These eigenvalues are given by $\sigma + 1/\mu$. With the shift $\sigma = 10^4$, a lower bound for the 1-norm condition number of the matrix $\widehat{\mathbf{M}}$ is 4.09×10^{13}. Figure 5.4 reports the scaled relative residual norms of the two methods with the subspace dimension $n = 100$. The scaled relative residual norm for an approximate eigenpair $(\theta, \mathbf{z})$ is defined by

$$\frac{\|(\theta^2 \mathbf{M} + \theta \mathbf{D} + \mathbf{K})\mathbf{z}\|_2}{\epsilon\,(|\theta|^2 \|\mathbf{M}\|_1 + |\theta| \|\mathbf{D}\|_1 + \|\mathbf{K}\|_1)},$$

where ϵ is the machine precision, which is at the order of 10^{-16} in double precision arithmetic. Since the norm of the matrix **M** is at the order of 10^{12}, it is better to show the scaled relative residual norm. To machine precision backward accuracy, the scaled relative residual norm should be about one.

Example 6. This final example arises from a finite element analysis of dissipative acoustics [4, 24]. Our matrix data for the associated algebraic quadratic eigenvalue

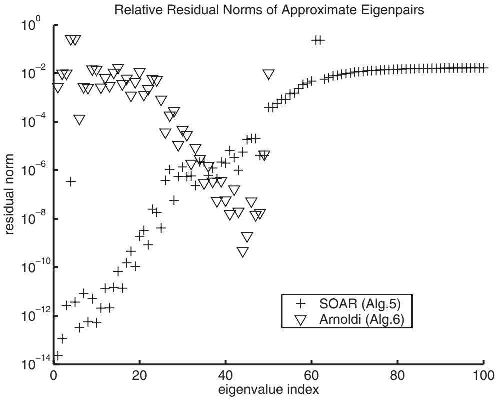

FIG. 5.5. *Relative residual norms of Example* 6.

problem are from [7]. The dimension of the QEP is $N = 9168$. Matrix $\mathbf{M}$ is symmetric positive definite, and matrices $\mathbf{D}$ and $\mathbf{K}$ are symmetric positive semidefinite. As described in [7], to find the eigenvalues of interest, we solve the shift-and-invert QEP (5.1) with the shift $\sigma = -253$. Figure 5.5 shows the relative residual norms for the approximated eigenpairs computed by the SOAR and Arnoldi methods with $n = 50$. We observe that SOAR converges faster than Arnoldi. By the Krylov-type subspace method proposed in [7], it is reported that with the number of iterations $n = 250$ to 300, three approximated eigenpairs converge with relative residual norms less than 10^{-12}. By contrast, the SOAR method delivers twice as many approximated eigenpairs with the same accuracy but only uses one-fifth of the number of iterations.

6. Discussion and future work. The primary purpose of this paper is to present the basic concept of the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ and its straightforward application for solving a large-scale QEP. There are many issues to examine. Foremost, one can ask whether the subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ is a better projection subspace to work with for an iterative solution of the QEP. A partial answer is based on the following observation. Let $\mathbf{A} = -\mathbf{M}^{-1}\mathbf{D}$ and $\mathbf{B} = -\mathbf{M}^{-1}\mathbf{K}$; then the QEP (1.2) is equivalent to the QEP

$$(\lambda^2\mathbf{I} - \lambda\mathbf{A} - \mathbf{B})\mathbf{x} = \mathbf{0}, \tag{6.1}$$

which can be written as the linear eigenvalue problem

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \lambda\mathbf{x} \\ \mathbf{x} \end{bmatrix} = \lambda \begin{bmatrix} \lambda\mathbf{x} \\ \mathbf{x} \end{bmatrix}. \tag{6.2}$$

In the Arnoldi basis $\mathbf{V}_n$ of the Krylov subspace $\mathcal{K}_n$, the coefficient matrix of (6.2) is represented by an upper Hessenberg matrix of order n,

$$\mathbf{V}_n^{\mathrm{T}} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{V}_n = \mathbf{H}_n. \tag{6.3}$$

On the other hand, using an orthonormal basis $\mathbf{Q}_n$ of the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$, the coefficient matrix of (6.2) is represented by a 2×2 block matrix of order $2n$,

$$\begin{bmatrix} \mathbf{Q}_n^{\mathrm{T}} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_n^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_n & \mathbf{B}_n \\ \mathbf{I}_n & \mathbf{0} \end{bmatrix}. \tag{6.4}$$

It can be shown that the subspace spanned by the columns of $\mathbf{V}_n$ can be embedded into the subspace spanned by the columns of the 2×2 block diagonal matrix $\mathrm{diag}(\mathbf{Q}_n, \mathbf{Q}_n)$, namely,

$$\mathrm{span}\{\mathbf{V}_n\} \subset \mathrm{span}\left\{ \begin{bmatrix} \mathbf{Q}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{Q}_n \end{bmatrix} \right\}.$$

Therefore, the $2n \times 2n$ block matrix in (6.4) should deliver at least as many good approximations of eigenpairs as the $n \times n$ Hessenberg matrix $\mathbf{H}_n$ does.

We note that the explicit triangular inversion in the SOAR procedure (Algorithm 4) brings the potential numerical instability. Many elaborate and proven techniques for robust and efficient implementation of Krylov subspace techniques developed over the years could be considered for the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$. The other subjects of further study include maintaining the orthogonality in the presence of finite precision arithmetic and a restarting strategy for solving the QEP by the SOAR method.

Krylov subspaces have an important characterization in terms of univariate matrix polynomials. Convergence theory of a Krylov subspace–based method has been established based on the theory of univariate polynomials and the distribution of eigenvalues of the underlying matrix. In section 2, we showed the connection between the second-order Krylov subspace $\mathcal{G}_n(\mathbf{A}, \mathbf{B}; \mathbf{u})$ and the bivariate polynomials $p_j(\alpha, \beta)$. It is unclear whether it can be used to develop a convergence theory which is directly based on the distribution of the matrices $\mathbf{A}$ and $\mathbf{B}$.

A closely related problem to the central theme of this paper is that of model-order reduction of a second-order dynamical system. The problem is about how to produce a reduced-order system of the same second-order form. One pioneering work is due to Su and Craig [22] back to 1991. In recent years, this approach has been repeatedly applied, studied, and improved; for example, see [2, 14, 19, 20]. In particular, the dissertation work of Slone [19] has essentially extended Su and Craig's approach to the model reduction of high-order dynamical systems but is based the popular AWE (asymptotic waveform evaluation) approach as widely known in interconnect analysis of integrated circuits and computational electromagnetics. In a forthcoming work, we will examine the application of the SOAR method for the model reduction of a second-order dynamical system and its connections to those previous works.

Acknowledgments. We are grateful to Gerard Sleijpen, Tom Kowalski, and Leonard Hoffnung for providing the matrix data used in Examples 4, 5, and 6, respectively. We would like to thank Ming Gu, Beresford Parlett, and Rodney Slone for helpful discussions during the course of this work. We thank the referees for valuable comments and suggestions to improve the presentation of the paper. One of the referees provided us the reference [13].

REFERENCES

[1] W. E. Arnoldi, *The principle of minimized iterations in the solution of the matrix eigenvalue problem*, Quart. Appl. Math., 9 (1951), pp. 17–29.

[2] Z. Bai, *Krylov subspace techniques for reduced-order modeling of large-scale dynamical systems*, Appl. Numer. Math., 43 (2002), pp. 9–44.

[3] 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.

[4] A. Bermúdez, R. G. Durán, R. Rodríguez, and J. Solomin, *Finite element analysis of a quadratic eigenvalue problem arising in dissipative acoustics*, SIAM J. Numer. Anal., 38 (2000), pp. 267–291.

[5] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[6] G. Golub and C. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[7] L. Hoffnung, R. C. Li, and Q. Ye, *Krylov type subspace methods for matrix polynomials*, Linear Algebra Appl., to appear.

[8] U. B. Holz, G. Golub, and K. H. Law, *A Subspace Approximation Method for the Quadratic Eigenvalue Problem*, Technical report SCCM-03-01, Stanford University, Stanford, CA, 2003.

[9] T. Kowalski, *Extracting a Few Eigenpairs of Symmetric Indefinite Matrix Pencils*, Ph.D. thesis, University of Kentucky, Lexington, KY, 2000.

[10] P. Lancaster, *Lambda-Matrices and Vibrating Systems*, Pergamon Press, Oxford, UK, 1966.

[11] N. K. Nichols and J. Kautsky, *Robust eigenstructure assignment in quadratic matrix polynomials: Nonsingular case*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 77–102.

[12] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice–Hall, Englewood Cliffs, NJ, 1980; revised reprint, Classics Appl. Math. 20, SIAM, Philadelphia, 1997.

[13] F. A. Raeven, *A new Arnoldi approach for polynomial eigenproblems*, in Proceedings of the Copper Mountain Conference on Iterative Methods, http://www.mgnet.org/mgnet/Conferences/CMCIM96/Psfiles/raeven.ps.gz, 1996.

[14] D. Ramaswamy and J. White, *Automatic generation of small-signal dynamic macromodels from* 3-*D simulation*, in Technical Proceedings of the Fourth International Conference on Modeling and Simulation of Microsystems, Nano Science and Technology Institute (NSTI), 2000, pp. 27–30.

[15] Y. Saad, *Numerical Methods for Large Eigenvalue Problems*, Halsted Press, New York, 1992.

[16] Y. Saad, *Iterative Methods for Linear Systems*, PWS, Boston, 1996.

[17] G. L. G. Sleijpen, A. G. L. Booten, D. R. Fokkema, and H. A. van der Vorst, *Jacobi–Davidson type methods for generalized eigenproblems and polynomial eigenproblems*, BIT, 36 (1996), pp. 595–633.

[18] G. L. G. Sleijpen, H. A. van der Vorst, and M. B. van Gijzen, *Quadratic eigenproblems are no problem*, SIAM News, 29 (1996), pp. 8–9.

[19] R. D. Slone, *Fast Frequency Sweep Model Order Reduction of Polynomial Matrix Equations Resulting from Finite Element Discretization*, Ph.D. thesis, Ohio State University, Columbus, OH, 2002.

[20] R. D. Slone, R. Lee, and J.-F. Lee, *Broadband model order reduction of polynomial matrix equations using single-point well-conditioned asymptotic waveform evaluation: Derivations and theory*, Internat. J. Numer. Methods Engrg., 58 (2003), pp. 2325–2342.

[21] G. W. Stewart, *Matrix Algorithms, Volume* II: *Eigensystems*, SIAM, Philadelphia, 2001.

[22] T.-J. Su and R. R. Craig, Jr., *Model reduction and control of flexible structures using Krylov vectors*, J. Guidance Control Dynam., 14 (1991), pp. 260–267.

[23] F. Tisseur, *Backward error and condition of polynomial eigenvalue problems*, Linear Algebra Appl., 309 (2000), pp. 339–361.

[24] F. Tisseur and K. Meerbergen, *The quadratic eigenvalue problem*, SIAM Rev., 43 (2001), pp. 235–286.