Extracting partial canonical structure for large scale eigenvalue problems

Bo Kågström and Petter Wiberg *

Department of Computing Science, Umeå University, SE-901 87 Umeå, Sweden
E-mail: bokg@cs.umu.se, wiberg@cs.toronto.edu

Received 24 November 1998; revised 13 April 2000
Communicated by P. van Dooren

We present methods for computing a nearby partial Jordan–Schur form of a given matrix and a nearby partial Weierstrass–Schur form of a matrix pencil. The focus is on the use and the interplay of the algorithmic building blocks – the implicitly restarted Arnoldi method with prescribed restarts for computing an invariant subspace associated with the dominant eigenvalue, the clustering method for grouping computed eigenvalues into numerically multiple eigenvalues and the staircase algorithm for computing the structure revealing form of the projected problem. For matrix pencils, we present generalizations of these methods. We introduce a new and more accurate clustering heuristic for both matrices and matrix pencils. Particular emphasis is placed on reliability of the partial Jordan–Schur and Weierstrass–Schur methods with respect to the choice of deflation parameters connecting the steps of the algorithm such that the errors are controlled. Finally, successful results from computational experiments conducted on problems with known canonical structure and varying ill-conditioning are presented.

Keywords: partial Jordan canonical stucture, eigenvalue problems, Jordan–Schur form, Weierstrass–Schur form, large scale, Krylov methods, implicitly restarted Arnoldi, eigenvalue clustering, staircase algorithms

AMS subject classification: 65F15, 65F10, 15A21, 15A22, 68N99

1. Introduction

In the literature about large scale eigenvalue computations, multiple eigenvalues are often treated in passing, although they appear in several applications, including dynamical systems in control theory. Most of the existing iterative methods do not tackle degenerate or defective eigenvalues satisfactory. This motivated us to consider the problem of computing partial canonical information for large scale eigenvalue problems. Inspired by methods for computing partial Schur forms (e.g., see [12,38,39]), we present algorithms for extracting partial information about the canonical structure

* Current address: Department of Computer Science, University of Toronto, 10 King's College Road, Toronto, Ontario, Canada M5S 3G4.

of large scale matrices and regular matrix pencils. For matrices, our method computes a partial Jordan–Schur form, which includes the Jordan structure of the dominant eigenvalue along with an invariant subspace of principal vectors. For regular matrix pencils, we compute a corresponding partial Weierstrass–Schur form associated with the dominant (finite) eigenvalue of $A - \lambda B$.

The primary objective of this paper is to give a unified presentation of such methods and to examine their behaviour in practice. Our methods are natural extensions that fit into the family of methods for computing partial canonical forms of various eigenproblems. We report in detail what accuracy may be expected, and we show that the methods have reliable behavior for not too ill-conditioned, defective as well as derogatory, problems. We also show how the tolerances connecting different steps in the algorithms should be chosen such that perturbations are controlled. In this process, we also indentify the limitations and difficulties encountered including open problems for future research. To the best of our knowledge this is the first attempt to consider the steps in a unified way, which has been necessary for several of our results and findings.

1.1. Canonical forms and staircase regularization

The Jordan canonical form (JCF) completely characterizes the algebraic properties of a matrix A. Similarly, the Weierstrass canonical form (WCF) describes the algebraic properties of a regular matrix pencil $A - \lambda B$. The JCF and the WCF are both unstable. Arbitrary small perturbations to any defective or derogatory problem may result in a completely different but less degenerate canonical structure. Given the problem dimensions and assuming that the eigenvalues are fixed (the nilpotent case), the relation between all possible canonical structures is given by the closure hierarchy of the similarity orbits for matrices, and the equivalence orbits for matrix pencils. In the general case, when we allow eigenvalues to vary and coalesce, it is the closures of the corresponding matrix (or matrix pencil) bundles that define the structure hierarchy of the possible Jordan (or Weierstrass) canonical forms. (For this type of qualitative results, see [10 and further references therein].)

Since computing the canonical forms are ill-posed problems, the perturbations introduced by the finite precision arithmetic are likely to corrupt the computed form. However, it is possible to regularize the problem by allowing a deflation criterion for range/null-space separations, and thereby make it possible to compute the canonical structure of a nearby matrix or pencil. Instead of computing the canonical form, we compute a special Schur form of a nearby matrix or matrix pencil. The canonical structure of an eigenvalue is revealed by imposing further conditions on the block structure of the triangular matrix in the Schur form of a matrix or the generalized Schur form of a matrix pencil. These Schur forms are the Jordan–Schur and the Weierstrass–Schur forms.

Kublanovskaya's staircase algorithm is the standard tool for computing the Jordan–Schur and the Weierstrass–Schur forms of matrices and regular matrix pen-

cils [25]. They are computed by a sequence of range/null-space separations and orthogonal deflations. The choice of method for computing a null space results in different variants of the staircase algorithm (e.g., see [2,14,17,18,21,22,25,26,42]). The singular value decomposition (SVD) is the most reliable choice. But staircase algorithms based on the SVD require a lot of floating point operations already for quite modest sized problems. For large scale problems, one approach would be to use some other method for finding a basis for the null space of a matrix. However, reliable methods for computing the null space of general large scale matrices do not exist today.

1.2. Extracting partial canonical structure information

An alternative approach is to use a projection method to find a restriction with respect to an invariant subspace or a pair of eigenspaces in the matrix pencil case (e.g., see [1,4,8,29,32,33,36,37]). Our approach is to combine the implicitly restarted Arnoldi (IRA) method [27,28,37] for computing a partial Schur form with staircase algorithms for computing the Jordan–Schur and the Weierstrass–Schur forms.

In the first step of the algorithm, we use a variant of the IRA method combined with explicit restarts to find an approximate invariant subspace of A containing the principal vectors of interest. The explicit restarts are used for extracting information about the Jordan blocks of the matrix that were not computed in previous applications of the IRA method. In theory, only one Jordan block can be identified per starting vector, but due to round-off errors we can accidentally get more information. This property may sometimes play into our hands, but it is not safe to rely on it to find the complete structure information (e.g., all Jordan blocks).

We focus on computing canonical information for the dominant eigenvalue, i.e., the eigenvalue of largest modulus (magnitude). It is necessary to be able to solve this simpler problem before taking on the general problem of computing partial canonical information for arbitrary eigenvalues. To find canonical information of interior eigenvalues some spectral enhancement technique such as shift-and-invert [33] must be incorporated, but we will not pursue this topic further in this paper. For matrix pencils we use a similar generalized IRA method [38] to compute approximate left and right eigenspaces containing the principal vectors associated with the dominant eigenvalue. For the generalized Arnoldi method to be constructible we have to assume that the matrix B in $A-\lambda B$ is nonsingular. In general, this should not pose any problem since any regular matrix pencil can be transformed into a problem with a nonsingular B. We then use a heuristic based on Gerschgorin circles to isolate a cluster of eigenvalues associated with the computed subspace(s). If the heuristic is successful, the mean of the isolated cluster is a good approximation to a multiple eigenvalue [31]. Finally, we apply Kublanovskaya's staircase algorithm to compute the Jordan–Schur form (or the Weierstrass–Schur form) of the clustered mean [17,21,22,25,30].

1.3. Outline of the paper

Before we go any further we outline the rest of the paper. In section 2, we give a brief introduction to the standard eigenvalue problem and the Jordan canonical form. We also define the partial Jordan canonical form. The background and the partial Weierstrass canonical form for matrix pencils are introduced in section 3. The structure revealing Jordan–Schur and Weierstrass–Schur forms together with the staircase algorithms used to compute these forms are introduced in section 4. In section 5, we present new heuristics for Gerschgorin clustering of eigenvalues of matrices and matrix pencils. Section 6 introduces the Arnoldi method and discusses the issues involved in computing an invariant subspace containing the principal vectors of interest. The generalized Arnoldi method for matrix pencils is presented in section 7. In sections 8 and 9, we present the overall methods for computing the partial Jordan–Schur and Weierstrass–Schur forms based on the building blocks from the previous sections. We express the nearby matrices corresponding to the computed Jordan–Schur or Weierstrass–Schur forms in terms of the deflation errors introduced during the Arnoldi and staircase algorithms, respectively. These nearby matrices represent the regularized problems that have a stable canonical structure with respect to the deflation criteria of the algorithms. We also present results from computational experiments to illustrate the reliability and robustness of our methods. Finally, we give some concluding remarks in section 10, focusing on applicability and reliability issues.

2. A partial Jordan canonical form

An $n \times n$ matrix A has n eigenvalues λ_i defined as the n zeros of the characteristic polynomial $p(\lambda) \equiv \det(A - \lambda I)$. The eigenvalues can take values anywhere in the complex plane $\mathbb{C}$. The set $\sigma(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is the *spectrum* of A. For each distinct λ_i there is at least one nonzero eigenvector x_i satisfying $Ax_i = \lambda_i x_i$. In the general case, we have $t \leqslant n$ distinct eigenvalues and if $t < n$, then A has at least one multiple eigenvalue.

The *algebraic multiplicity* of λ_i, $n_{\mathrm{a}}(\lambda_i)$, is the multiplicity of the zero of the characteristic polynomial. The *geometric multiplicity* of λ_i, $n_{\mathrm{g}}(\lambda_i)$, is the dimension of the null space of $\mathrm{null}(A - \lambda_i I)$. If $n_{\mathrm{g}}(\lambda_i) < n_{\mathrm{a}}(\lambda_i)$, then λ_i is a *defective* eigenvalue and A is said to be a defective matrix. A is *nondefective* if $n_{\mathrm{g}}(\lambda_i) = n_{\mathrm{a}}(\lambda_i)$ for all i. A is *derogatory* if $n_{\mathrm{g}}(\lambda_i) \geqslant 2$, i.e., there exist more than one eigenvector for at least one of the eigenvalues. For example, the $n \times n$ identity matrix is derogatory but nondefective ($n_{\mathrm{g}}(1) = n_{\mathrm{a}}(1) = n$).

By placing the eigenvectors as the columns of X, we see that every non-defective matrix is diagonalizable by a similarity transformation:

$$AX = X\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n).$$

However, this is not the case for a defective matrix. The simplest matrix similar to a general matrix is a direct sum of Jordan blocks

$$J_k(\lambda) = \begin{pmatrix} \lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda \end{pmatrix}.$$

A Jordan block $J_k(\lambda)$ of size $k \times k$ is a cyclic transformation. The range is generated by the kth unit vector e_k – the vectors of the Krylov sequence $e_k, J_k e_k, J_k^2 e_k, \ldots, J_k^{k-1} e_k$ span the range. From the cyclicity of the transformation we get the *vector chain relations* $(J_k - \lambda I)e_{j+1} = e_j$. The vector e_{j+1} is said to be a *principal vector of grade* $j+1$. The order k is often referred to as the *height* of the chain. The eigenvectors are principal vectors of grade 1.

Let the subset $\mathcal{L}(A) = \{\lambda_1, \ldots, \lambda_t\}$ of $\sigma(A)$ be the set of distinct eigenvalues. Then for any $A \in \mathbb{C}^{n\times n}$ there exits a nonsingular matrix $X \in \mathbb{C}^{n\times n}$ that transforms A to *Jordan canonical form* (JCF):

$$AX = XJ,$$

where

$$J = \operatorname{diag}\big(J(\lambda_1), J(\lambda_2), \ldots, J(\lambda_t)\big).$$

The *Jordan matrix* $J(\lambda_i)$ is a direct sum of the Jordan blocks associated with the eigenvalue λ_i [13,16,35]. The number of Jordan blocks for an eigenvalue is the same as its geometric multiplicity. To simplify the notation we let $g_i = n_{\mathrm{g}}(\lambda_i)$ and we let $s_k^{(i)}$ be the sizes of the Jordan blocks associated with λ_i where $s_1^{(i)} \geqslant s_2^{(i)} \geqslant \cdots \geqslant s_{g_i}^{(i)}$. Algebraically, the $s_k^{(i)}$s are the degrees of the elementary divisors of $A - \lambda I$ at $\lambda = \lambda_i$, also known as the *Segre characteristics*. Later we also need $h_i = \max_k s_k^{(i)}$, where h_i is the maximum height of the vector chains for the eigenvalue λ_i. The Jordan matrix $J(\lambda_i)$ can be expressed as

$$J(\lambda_i) = \operatorname{diag}\big(J_{s_1^{(i)}}(\lambda_i), J_{s_2^{(i)}}(\lambda_i), \ldots, J_{s_{g_i}^{(i)}}(\lambda_i)\big).$$

The Jordan matrix J is uniquely defined up to the order of the Jordan blocks.

The JCF induces a factorization of the vector space $\mathbb{C}^n$ into a direct sum of subspaces

$$\mathbb{C}^n = \mathcal{T}_1 \oplus \mathcal{T}_2 \oplus \cdots \oplus \mathcal{T}_t, \tag{1}$$

such that $\mathcal{T}_i \cap \mathcal{T}_j = \{0\}$ if $i \neq j$ and each subspace $\mathcal{T}_i$ is annihilated by the transformation $(A - \lambda_i I)^{h_i}$. The restriction of A with respect to any of the subspaces has Jordan blocks associated with only *one* eigenvalue.

Suppose that the column vectors of $V \in \mathbb{C}^{n\times k}$ span an invariant subspace. The restriction of A is represented by a smaller $k \times k$ matrix C so that $AV = VC$. If

$CY = YJ$ is the JCF of C, then we get the factorization $AX = XJ$ for $X = VY$. The vectors in a chain of principal vectors $y_1, y_2, \ldots, y_{s_j}$ for $\lambda \in \mathcal{L}(J)$ satisfy

$$(A - \lambda I)Vy_i = V(C - \lambda I)y_i = \begin{cases} 0 & \text{if } i = 1, \\ Vy_{i-1} & \text{if } i > 1. \end{cases}$$

The vectors $Vy_1, Vy_2, \ldots, Vy_{s_j}$ are a chain of principal vectors for the eigenvalue $\lambda \in \mathcal{L}(A)$. Each Jordan block of the restriction C corresponds to a Jordan block – possibly of lower order – of the larger matrix A. This leads to the following definition.

Definition 1. Let $X \in \mathbb{C}^{n\times k}$, where $k < n$, have full rank and satisfy

$$AX = XJ, \tag{2}$$

where J is a $k \times k$ Jordan matrix. We say that the factorization (2) is a *partial Jordan canonical form* of A with respect to the subspace spanned by X.

The problem of computing a partial JCF is twofold. First, we compute an invariant subspace corresponding to a subset of the spectrum. Second, the JCF of the restriction to this subspace is computed.

3. A partial Weierstrass canonical form

A *matrix pencil* is the set of matrices $A - \lambda B$ parameterized by the scalar λ. A pencil is *regular* if the characteristic polynomial $p(\lambda) \equiv \det(A-\lambda B)$ is not identically equal to zero. The n eigenvalues of a regular pencil are points in the extended complex plane $\mathbb{C}\cup\{\infty\}$. The eigenvalues λ_i are defined as the zeros of $p(\lambda)$ and $n-\text{degree}(p(\lambda))$ additional eigenvalues $\lambda_i = \infty$. The spectrum of $A - \lambda B$ is the set of eigenvalues $\sigma(A - \lambda B) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

An alternative representation of a matrix pencil is the *cross product form*: the set of matrices $\beta A - \alpha B$ where $(\alpha, \beta) \in \mathbb{C}^2$. The mapping $(\alpha, \beta) \mapsto \alpha/\beta$ shows the relation between the eigenvalues of $\beta A - \alpha B$ and $A - \lambda B$. For example, zero and infinite eigenvalues are represented as $(0, \beta)$ and $(\alpha, 0)$, and can be treated as any other points in $\mathbb{C}^2$. The cross product representation simplifies and unifies the perturbation theory for the generalized eigenvalue problem [41].

For each $\lambda_i \neq \infty$, there is at least one nonzero eigenvector x_i satisfying $Ax_i = \lambda_i Bx_i$, and, for each $\lambda_i = \infty$, there is at least one nonzero eigenvector x_i such that $Bx_i = 0$. A regular pencil $A - \lambda B$ has $t \leqslant n$ distinct eigenvalues. The *algebraic multiplicity* of λ_i, $n_{\text{a}}(\lambda_i)$, is the multiplicity of the zero of the characteristic polynomial. The *geometric multiplicity* of λ_i, $n_{\text{g}}(\lambda_i)$, is the dimension of the null spaces null$(A - \lambda_i B)$ and null(B), respectively, for $\lambda_i \neq \infty$ and $\lambda_i = \infty$.

The JCF is generalized to include the canonical structure of the infinite eigenvalue. For any regular pencil $A - \lambda B$, there are two nonsingular matrices $X \in \mathbb{C}^{n\times n}$ and $Y \in \mathbb{C}^{n\times n}$ that transform $A - \lambda B$ to *Weierstrass canonical form* (WCF):

$$(A - \lambda B)X = Y\left(\begin{pmatrix} J & 0 \\ 0 & I \end{pmatrix} - \lambda \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}\right),$$

where J is a Jordan matrix and N is a nilpotent Jordan matrix [13]. The Jordan blocks $J_k(0)$ of N describe the canonical structure of the infinite eigenvalue. Both J and N are unique up to the order of the blocks. We introduce a notation similar to the standard case. Let $g_i = n_{\mathrm{g}}(\lambda_i)$ and let $s_k^{(i)}$ be the sizes of the Jordan blocks associated with λ_i where $s_1^{(i)} \geqslant s_2^{(i)} \geqslant \cdots \geqslant s_{g_i}^{(i)}$. The height of λ_i is defined as $h_i = \max_k s_k^{(i)}$.

If the matrix B is nonsingular, the pencil $A - \lambda B$ is regular and the generalized eigenvalue problem $Ax = \lambda Bx$ is mathematically equivalent to the standard problems $B^{-1}Ax = \lambda x$ and $AB^{-1}y = \lambda y$ with $Bx = y$. Moreover, the pencil has no infinite eigenvalues and the nilpotent matrix N vanishes in the WCF. Later, in our discussion about Arnoldi methods for generalized eigenvalue problems, we enforce the assumption that B is nonsingular (see section 7). For any regular pencil $A - \lambda B$, we may without loss of generality consider a shifted problem $A - \lambda(B - \alpha A)$, where $B - \alpha A$ is nonsingular.

The columns of $V \in \mathbb{C}^{n\times k}$ span a (right) eigenspace of $A - \lambda B$ if there is a matrix $W \in \mathbb{C}^{n\times k}$ such that the restriction to V can be represented as a $k \times k$ pencil $C - \lambda D$ satisfying

$$(A - \lambda B)V = W(C - \lambda D).$$

Consider the WCF of the pencil $C - \lambda D$,

$$(C - \lambda D)\widehat{X} = \widehat{Y}\left(\begin{pmatrix} J & 0 \\ 0 & I \end{pmatrix} - \lambda \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}\right).$$

Then, we get the factorization

$$(A - \lambda B)X = Y\left(\begin{pmatrix} J & 0 \\ 0 & I \end{pmatrix} - \lambda \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}\right)$$

for $X = V\widehat{X}$ and $Y = W\widehat{Y}$. Each block of the pencil $C - \lambda D$ corresponds to a block in the WCF of $A - \lambda B$.

Definition 2. Let $X, Y \in \mathbb{C}^{n\times k}$, where $k < n$, have full rank and satisfy

$$(A - \lambda B)X = Y\left(\begin{pmatrix} J & 0 \\ 0 & I \end{pmatrix} - \lambda \begin{pmatrix} I & 0 \\ 0 & N \end{pmatrix}\right), \tag{3}$$

where J is a $k_J \times k_J$ Jordan matrix and N is a $k_N \times k_N$ Jordan matrix with eigenvalue zero ($k_J + k_N = k$). We say that the factorization (3) is a *partial Weierstrass canonical form* of $A - \lambda B$ with respect to the pair of subspaces spanned by X and Y.

The two major steps in computing a partial WCF are similar to the partial JCF case. First, a pair of eigenspaces associated with a subset of the spectrum is computed. Second, the staircase form of the restriction to the pair of eigenspaces is computed to extract information about the partial Weierstrass form of the pencil.

4. The Jordan–Schur and Weierstrass–Schur forms

The Jordan canonical form is unstable in the sense that it is not a continuous function of the matrix entries. The 2×2 matrix

$$B(\varepsilon) = \begin{pmatrix} 0 & 1 \\ \varepsilon & 0 \end{pmatrix},$$

illustrates the ill-conditioned nature of the JCF. The matrix has the JCF $J_2(0)$ for $\varepsilon = 0$ and $\mathrm{diag}(J_1(-\sqrt{\varepsilon}), J_1(\sqrt{\varepsilon}))$, otherwise. We see that the JCF of $B(\varepsilon)$ is not continuous at $\varepsilon = 0$. For small ε, the diagonalizing similarity transformation is very ill-conditioned. Extending the example to arbitrary matrices, it is obvious that computing the JCF is an ill-posed problem. However, by regularizing the problem in terms of a deflation criterion for range/null-space separations, we can compute exact canonical information of a nearby matrix $A + G$. The size of $\|G\|_F$ is an upper bound on the distance from A to the regularized problem with the computed structure. The standard tool used for computing canonical structure information of matrices and matrix pencils is the staircase algorithm (e.g., see [14,21,22,25,30] for matrices and [2,6,7,18,26,42] for matrix pencils).

The first staircase algorithm for computing the Jordan structure associated with a given eigenvalue λ_i of a non-Hermitian matrix A is due to Kublanovskaya [25]. In theory the Jordan structure of λ_i is given by a special Schur form with staircase structure – the *Jordan–Schur form* of A with respect to λ_i:

$$Q^H A Q = S \equiv \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}, \tag{4}$$

where Q is unitary and S is upper triangular. The block $S_{11} = \lambda_i I + N$, where N is a nilpotent matrix with block structure in *staircase form* that reveals the Jordan structure associated with λ_i:

$$N \equiv \begin{pmatrix} 0 & N_{12} & \dots & N_{1h_i} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & N_{h_i-1h_i} \\ 0 & 0 & \dots & 0 \end{pmatrix}. \tag{5}$$

The diagonal blocks of N are of size $m_k \times m_k$, with $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{h_i} \geqslant 1$. The m_ks are the set of nonzero successive differences in the nullities of the matrices $(A - \lambda_i I)^k$, $k = 1, 2, \dots$, also known as the *Weyr characteristics*. The number m_k is the dimension of the space spanned by the principal vectors of grade k and the

corresponding columns of Q form a basis for the space spanned by the principal vectors of grade k. As before, h_i is the height of the longest principal vector chain of λ_i.

4.1. Computing a Jordan–Schur form

Considering the shifted transformation $A - \lambda_i I$, the problem of computing the Jordan structure of λ_i is equivalent to computing the dimension of the null space of $(A-\lambda_i I)^j$ for increasing j. Without loss of generality, we assume that we are searching for the Jordan structure of A associated with the zero eigenvalue. The Jordan–Schur form of $\lambda_i = 0$ is constructed from A by applying a sequence of unitary similarity transformations and deflations that step by step compute the principal vectors of increasing grade. The similarity transformation is built from a rank revealing factorization used to find an orthonormal basis for the null space. We present the algorithm based on the singular value decomposition [14,21,22,30]. Other rank revealing factorizations and the issues involved for determining the numerical rank of a matrix is, for example, discussed in [40, chapter 5].

The SVD has the desirable property that it gives information about the distance from a matrix to the closest lower rank matrix. Let $A = U\Sigma V^H$ be the SVD of A, where U and V are unitary and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_n)$ with $0 \leqslant \sigma_1 \leqslant \cdots \leqslant \sigma_n$. Notice that the singular values are in increasing order. From the Eckart–Young theorem, $\|A - A_k\|_2$ with $\mathrm{rank}(A_k) \leqslant k$ is minimized by

$$A_k = U\Sigma^{(k)}V^H,$$

where

$$\Sigma^{(k)} = \begin{pmatrix} 0_{(n-k)\times(n-k)} & & & \\ & \sigma_{n-k+1} & & \\ & & \ddots & \\ & & & \sigma_n \end{pmatrix},$$

and $\|A - A_k\|_2 = \sigma_{n-k}$ [15,41].

Due to errors in the original problem and rounding errors introduced in the computations, the singular values of a rank deficient matrix will be small rather than zero. How to determine the *numerical rank* of a matrix is an important question. The result above on lower rank approximations suggests an answer. Given a tolerance *tol*, the numerical rank of the matrix is the largest k such that $\sigma_{n-k+1} > tol \cdot \sigma_n$. This definition does not ensure that the rank is stable in the sense that a small perturbation relative to *tol* could cause the matrix to change rank. Once again the theorem above gives us the means to recognize the danger. The rank is sensitive to perturbations if σ_{n-k+1} is only slightly larger than σ_{n-k}. Given a constant *gap*, we add the requirement that k should be chosen as the largest number such that $\sigma_{n-k+1} > tol \cdot \sigma_n$ and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$ [6,7]. In practice, we use $\sigma_{n-k+1} > tol\|A\|_F$, i.e., $\|A\|_2$ is replaced by $\|A\|_F$. If the gap criterion fails, the numerical rank is not defined for that tolerance.

This situation has traditionally been resolved by increasing the tolerance until all of the matrix is included in the null space or a tolerance with a well defined numerical rank is found. The down side to this approach is that it is difficult to know how changes in the tolerance affect earlier steps in an algorithm.

In each step of the staircase algorithm, we use the criterion above to determine the numerical rank of the matrix. The current matrix is then transformed and replaced by a nearby rank-deficient matrix with the exact rank equal to the numerical rank of A. Consider the SVD

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix},$$

where the matrices are partitioned conforming to the numerical rank of A. Hence, Σ_1 corresponds to the "zero" singular values. The null space is then deflated by a unitary similarity transformation,

$$A^{(1)} = \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix} A \begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} V_1^H U_1 \Sigma_1 & V_1^H U_2 \Sigma_2 \\ V_2^H U_1 \Sigma_1 & V_2^H U_2 \Sigma_2 \end{pmatrix}.$$

By the criterion for numerical rank we accept the leading columns of $A^{(1)}$ to be zero and replace the matrix by the exactly rank deficient matrix,

$$\widetilde{A} \equiv \begin{pmatrix} 0 & \widetilde{A}_{12} \\ 0 & \widetilde{A}_{22} \end{pmatrix} = \begin{pmatrix} 0 & V_1^H U_2 \Sigma_2 \\ 0 & V_2^H U_2 \Sigma_2 \end{pmatrix}.$$

Now, observe that if $z^H = (0, x^H)$, where x is in the null space of $\widetilde{A}_{22}$, then

$$\widetilde{A} \begin{pmatrix} 0 \\ x \end{pmatrix} = \begin{pmatrix} \widetilde{A}_{12} x \\ \widetilde{A}_{22} x \end{pmatrix} = \begin{pmatrix} \widetilde{A}_{12} x \\ 0 \end{pmatrix}.$$

Hence, $\widetilde{A}z$ is in the null space of $\widetilde{A}^2$, i.e., z is a second grade principal vector of $\widetilde{A}$. The higher grade principal vectors are computed by applying the procedure recursively to $\widetilde{A}_{22}$ to produce a Jordan–Schur matrix in staircase form (5). We remark that in each step the nullity is maximized with respect to the deflation criterion, which corresponds to finding the least generic Jordan structure with respect to the deflation criterion [9,10]. For a zero eigenvalue of multiplicity n, the most generic JCF is a $J_n(0)$ block (defective but nonderogatory) and the least generic JCF is n Jordan blocks $J_1(0)$ (derogatory but nondefective). All other possible Jordan structures under small perturbations are given by the closure-hierarchy of the nilpotent orbits (e.g., see [10]). This seemingly complicated result has a simple solution; the closure hierarchy in terms of the Segre characteristics is given by the dominance ordering of the integer partitions of n.

The deflations made in the staircase algorithm result in an exact Jordan–Schur form of a nearby matrix $A + G$:

$$Q^H(A + G)Q = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}, \tag{6}$$

Algorithm 1. Staircase algorithm for matrices

Input: $A \in \mathbb{C}^{n\times n}$. Deflation tolerance parameters *tol* and *gap*.
Output: $S \in \mathbb{C}^{n\times n}$ and a unitary matrix $Q \in \mathbb{C}^{n\times n}$ such that $(A+G)Q = QS$, where $\|G\|_F = \mathrm{O}(tol\|A\|_F)$. The zero eigenvalue of $A+G$ has the Weyr characteristics m_k listed in *str*.
function $[S, Q, str] = \mathrm{staircase}(A, tol, gap)$
$S = A,\ Q = I_n$
Let $U\Sigma V^H$ be the SVD of A.
Find the largest index k such that $\sigma_{n-k+1} > tol \cdot \|A\|_F$ and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$.
Partition U, Σ and V conformally.
$i = 1,\ m = n - k$
$str = m$
while $m > 0$ **do**
  $Q(:, i:n) = Q(:, i:n)V$
  if $i = 1$
    $S = (\,0 \quad V^H U_2 \Sigma_2\,)$
  else
    $S(:, i:n) = \begin{pmatrix} S(1:i-1, i:n)V \\ 0 \quad V^H U_2 \Sigma_2 \end{pmatrix}$
  $i = i + m$
  Let $U\Sigma V^H$ be the SVD of $S(i:n, i:n)$.
  Find the largest index k such that $\sigma_{n-k+1} > tol \cdot \|A\|_F$ and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$.
  Partition U, Σ and V conformally.
  $m = n - k - i + 1$
  $str = [str, m]$
end
Compute Schur form of remaining part $S(i:n, i:n)$ and update $Q(:, i:n)$.

Algorithm 1. Staircase algorithm for computing the Jordan–Schur form with respect to the zero eigenvalue.

where $\|G\|_F = \mathrm{O}(tol\|A\|_F)$. We say that the original A has a well-defined canonical structure at $\lambda_i = 0$, if all range/null-space separations in the staircase algorithm are well-defined with respect to the deflation tolerance parameters *tol* and *gap*. If A has zero as an exact multiple eigenvalue with well-conditioned principal vectors, then $\|G\|_F/\|A\|_F$ will be of size $\mathrm{O}(n \cdot macheps)$, where *macheps* is the relative machine accuracy. Algorithm 1 presents a high-level description of the staircase algorithm. For simplicity we have left out the case when the gap criterion in the numerical rank fails. We remark that $G = -(AQ - QS)$ can easily be computed from the output of the algorithm. Notice that we get the algebraic multiplicity $n_{\mathrm{a}}(\lambda_i)$ by summing the Weyr characteristics, i.e., the entries in *str*. The staircase form associated with the zero eigenvalue occupy the first $n_{\mathrm{a}}(\lambda_i)$ rows and columns of S, and the first $n_{\mathrm{a}}(\lambda_i)$ columns of Q is a base for the eigenspace spanned by the principal vectors of $\lambda_i = 0$.

4.2. Computing a Weierstrass–Schur form

The Weierstrass canonical form for regular matrix pencils suffers from the same type of instability as the JCF. Small perturbations to the original problem may change the canonical structure. But, the matrix pencil problem is more complicated since a regular pencil may be close to a singular pencil. The pencil

$$A - \lambda B(\varepsilon) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 \\ 0 & \varepsilon \end{pmatrix}$$

is regular for $\varepsilon \neq 0$ and $A - \lambda B(0)$ is singular since $\det(A - \lambda B(0)) \equiv 0$ for any λ. We shall assume that the regular pencil is stable in the sense that it requires a large perturbation relative to the tolerance for numerical rank *tol* to singularly perturb the problem. In the Arnoldi method for matrix pencils this requirement is reflected in the requirement on B to be well-conditioned with respect to inversion. Theory, algorithms and software for the general problem, including the cases when $\det(A - \lambda B) \equiv 0$ and A and B are rectangular matrices, are presented in [6,7].

In the spirit of the Jordan–Schur form, we define the *Weierstrass–Schur form* of a regular pencil $A - \lambda B$ with respect to the eigenvalue λ_i:

$$P^H(A - \lambda B)Q = S - \lambda T \equiv \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} - \lambda \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix},$$

where P and Q are unitary matrices and S and T are upper triangular. If $\lambda_i \neq \infty$, the block $S_{11} = \lambda_i T_{11} + N$, where N is a nilpotent block matrix that exposes the Weierstrass structure of λ_i. The block T_{11} is nonsingular since $A - \lambda B$ is regular. If $\lambda_i = \infty$, then $T_{11} = N$ is nilpotent and reveals the structure at ∞. Moreover, the regularity of $A - \lambda B$ implies that S_{11} is nonsingular. As for the matrix case, N is an upper triangular blocked matrix in staircase form,

$$N \equiv \begin{pmatrix} 0 & N_{12} & \cdots & N_{1h_i} \\ 0 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & N_{h_i-1h_i} \\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

The diagonal blocks of N (with zero entries) are of size $m_k \times m_k$, with $m_1 \geqslant m_2 \geqslant \cdots \geqslant m_{h_i} \geqslant 1$, where m_k is the number of principal vectors of grade k ($m_1 =$ the number of eigenvectors). Moreover, the corresponding columns of Q and P span the right and left spaces of grade k principal vectors.

The staircase algorithm for regular matrix pencils resembles the one for matrices. Suppose that λ_i is an eigenvalue of $A - \lambda B$. If $\lambda_i \neq \infty$, the zero eigenvalue of the shifted matrix pencil $(A - \lambda_i B) - \lambda B$ has the same canonical structure as λ_i. Similarly, if $\lambda_i = \infty$, the structure at infinity is the same as the structure of the zero eigenvalue of the pencil $B - \mu A$ with $\mu = 1/\lambda$. The Weierstrass–Schur form of the zero eigenvalue of a matrix pencil $A - \lambda B$ is computed by a sequence of orthogonal

deflations in terms of unitary equivalence transformations. Consider the SVD of A partitioned conforming to the numerical rank of the matrix:

$$A = \begin{pmatrix} U_1 & U_2 \end{pmatrix} \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix} \begin{pmatrix} V_1^H \\ V_2^H \end{pmatrix}.$$

The null space of A is deflated by postmultiplication:

$$(A - \lambda B)\begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} U_1\Sigma_1 & U_2\Sigma_2 \end{pmatrix} - \lambda \begin{pmatrix} BV_1 & BV_2 \end{pmatrix}.$$

The submatrix $U_1\Sigma_1$ has tiny entries and by setting these to zero we get an exactly rank deficient nearby matrix pencil

$$\left(\widetilde{A} - \lambda B\right)\begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} 0 & U_2\Sigma_2 \end{pmatrix} - \lambda \begin{pmatrix} BV_1 & BV_2 \end{pmatrix}.$$

Since the pencil $A-\lambda B$ is regular, A and B have non-intersecting null spaces, implying that the submatrix BV_1 has full rank. Let

$$BV_1 = \begin{pmatrix} W_1 & W_2 \end{pmatrix} \begin{pmatrix} R \\ 0 \end{pmatrix},$$

be the QR factorization of BV_1. Then

$$\begin{pmatrix} W_1 & W_2 \end{pmatrix}^H \left(\widetilde{A} - \lambda B\right)\begin{pmatrix} V_1 & V_2 \end{pmatrix} = \begin{pmatrix} 0 & W_1^H U_2\Sigma_2 \\ 0 & W_2^H U_2\Sigma_2 \end{pmatrix} - \lambda \begin{pmatrix} R & W_1^H BV_2 \\ 0 & W_2^H BV_2 \end{pmatrix}.$$

We introduce some new notation and represent the last equivalence transformation as

$$W^H\left(\widetilde{A} - \lambda B\right)V = \begin{pmatrix} 0 & \widetilde{A}_{12} \\ 0 & \widetilde{A}_{22} \end{pmatrix} - \lambda \begin{pmatrix} \widetilde{B}_{11} & \widetilde{B}_{12} \\ 0 & \widetilde{B}_{22} \end{pmatrix},$$

which corresponds to the first step in the staircase algorithm for $A - \lambda B$ with B non-singular. The complete Weierstrass–Schur form is computed by recursively applying the procedure to the pencil $\widetilde{A}_{22} - \lambda\widetilde{B}_{22}$.

The deflations made in the staircase algorithm result in an exact Weierstrass–Schur form of a nearby matrix pencil $A + G - \lambda B$:

$$P^H(A + G - \lambda B)Q = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} - \lambda \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix}, \tag{7}$$

where $\|G\|_F = \mathrm{O}(\mathit{tol}\max(\|A\|_F, \|B\|_F))$. As for the matrix case, we say that the original pencil has a well-defined canonical structure at $\lambda_i = 0$, if all range/null-space separations in the staircase algorithm are well-defined with respect to the deflation tolerance parameters *tol* and *gap*. Notice that we only discard elements in the A-part of the pencil. If $A - \lambda B$ has zero as an exact multiple eigenvalue with well-conditioned principal vectors, then $\|G\|_F / \max(\|A\|_F, \|B\|_F)$ is of size $\mathrm{O}(n \cdot \mathit{macheps})$. Algorithm 2 presents a high-level description of the complete algorithm. For simplicity we have left out the case when the gap criterion in the numerical rank fails. As for the matrix case, we get the algebraic multiplicity $n_{\mathrm{a}}(\lambda_i)$ by summing the Weyr characteristics, i.e., the entries in *str*. The staircase form associated with the zero eigenvalue

Algorithm 2. Staircase algorithm for regular pencils

Input: $A, B \in \mathbb{C}^{n\times n}$ such that $A - \lambda B$ is regular. Deflation tolerance parameters *tol* and *gap*.
Output: $S, T \in \mathbb{C}^{n\times n}$ and two unitary matrices $P, Q \in \mathbb{C}^{n\times n}$ such that $S - \lambda T = P^H((A+G) - \lambda B)Q$, where $\|G\|_F = \mathrm{O}(tol \max(\|A\|_F, \|B\|_F))$. The zero eigenvalue of $A + G - \lambda B$ has the Weyr characteristics m_k listed in *str*.
function $[S, T, P, Q, str] = \mathrm{staircase}(A, B, tol, gap)$
$S = A,\ T = B,\ P = I_n,\ Q = I_n$
Let $U\Sigma V^H$ be the SVD of A.
Find the largest index k such that $\sigma_{n-k+1} > tol \cdot \max(\|A\|_F, \|B\|_F)$
and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$.
Partition U, Σ and V conformally.
$i = 1,\ m = n - k$
$str = m$
while $m > 0$ **do**
  $Q(:, i:n) = Q(:, i:n)V$
  if $i = 1$
    $S = \begin{pmatrix} 0 & U_2\Sigma_2 \end{pmatrix},\ T = TV$
  else
    $S(:, i:n) = \begin{pmatrix} S(1:i-1, i:n)V \\ 0 \quad U_2\Sigma_2 \end{pmatrix}, \quad T(:, i:n) = T(:, i:n)V$
  $j = i + m - 1$
  $[W, T(i:n, i:j)] = \mathrm{qr}(T(i:n, i:j))$
  $T(i:n, j+1:n) = W^H T(i:n, j+1:n)$
  $S(i:n, j+1:n) = W^H S(i:n, j+1:n)$
  $P(:, i:n) = P(:, i:n)W$
  $i = i + m$
  Let $U\Sigma V^H$ be the SVD of $S(i:n, i:n)$.
  Find the largest index k such that $\sigma_{n-k+1} > tol \cdot \max(\|A\|_F, \|B\|_F)$
  and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$.
  Partition U, Σ and V conformally.
  $m = n - k - i + 1$
  $str = [str, m]$
end
Compute generalized Schur form of $S(i:n, i:n)$ and $T(i:n, i:n)$ and
update $P(:, i:n)$ and $Q(:, i:n)$.

Algorithm 2. Staircase algorithm for computing a Weierstrass–Schur form with respect to the zero eigenvalue.

occupies the first $n_\mathrm{a}(\lambda_i)$ rows and columns of S and T. The corresponding left and right eigenspaces occupy the first $n_\mathrm{a}(\lambda_i)$ columns of P and Q.

We remark that the algorithm presented presumes that A and B have non-intersecting null spaces. To handle situations reliably when $A - \lambda B$ is close to singular and perhaps A and B have a common null space with respect to the deflation criterion, we have to use the more "powerful" staircase algorithms for singular pencils, e.g., see [6,7].

5. Clustering of computed eigenvalues

When computing the canonical structure of an arbitrary eigenvalue of a matrix or a matrix pencil, the staircase algorithm is applied to a shifted problem, e.g., $A-\lambda_{\text{shift}}I$, where λ_{shift} is the multiple eigenvalue. Since the eigenvalues are computed in finite precision, the approximations to an exact multiple eigenvalue will at best appear to be a group of nearby eigenvalues. For example, if A has a multiple eigenvalue λ_i, then the individual eigenvalues of $A+E$ with $\|E\|=\varepsilon$ may spread out inside and on the boundary of a circle with origin at λ_i and radius of size O($\varepsilon^{1/h}$), where h is the size of the largest Jordan block of λ_i, while the mean of the corresponding eigenvalues is only perturbed by a factor of size O(ε) [24]. To summarize, the mean of a set of nearby eigenvalues is less sensitive to perturbations than individual eigenvalues [31]. If we can identify a group of perturbed eigenvalues corresponding to a true multiple eigenvalue, the mean of the group is most likely a much better approximation than the individual eigenvalues. Therefore, enforcing the mean as a multiple eigenvalue makes sense. *Clustering* is the process of identifying and isolating a group of nearby eigenvalues associated with a multiple eigenvalue.

The theory for Gerschgorin circles provides us with a tool for clustering the eigenvalues of both matrices and matrix pencils. Let us first consider the matrix case.

5.1. A Gerschgorin heuristic for matrices

Given a diagonal matrix $D=\text{diag}(d_1,d_2,\ldots,d_n)$. The Gerschgorin theorem states that for $A\in\mathbb{C}^{n\times n}$, the region of the complex plane defined by the n disks

$$|z-a_{ii}|\leqslant\sum_{\substack{j\neq i\\ j=1}}^{n}\left|\frac{d_i}{d_j}\right||a_{ji}|$$

contains the eigenvalues of DAD^{-1}. Moreover, if k of the circles intersect, then the union of these disks contains k eigenvalues [15,41].

Considering the perturbed matrix $\widetilde{A}=A+E$, the disks are

$$\left|z-\tilde{a}_{ii}\right|\leqslant\sum_{j\neq i}\left|\frac{d_i}{d_j}\right|\left|\tilde{a}_{ji}\right|,$$

and, provided that $|d_i|>1$ for all i, we get the upper bound [30],

$$|z-a_{ii}|\leqslant|d_i|\left(\|E\|_1+\sum_{j\neq i}\frac{|a_{ji}|}{|d_j|}\right). \tag{8}$$

Given a size for $\|E\|_1$, the bound describes a set of disks centered at the eigenvalues. The trick in the clustering strategy devised by Ruhe [30] is to compute the Schur form and choose a diagonal matrix to isolate groups of intersecting disks. The mean of each group is then used as numerically multiple eigenvalues. When Ruhe's strategy

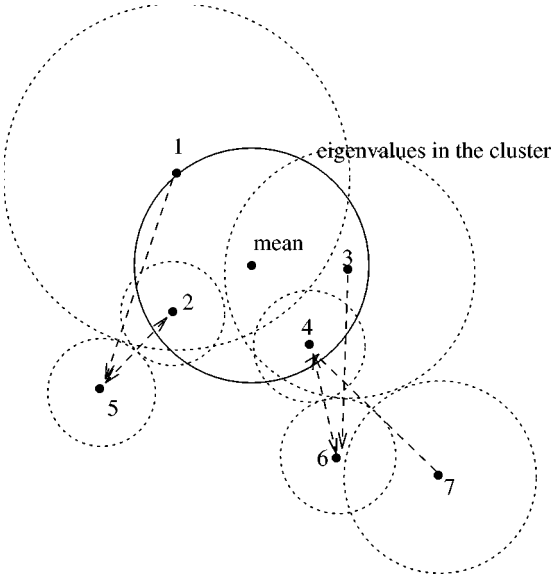

Figure 1. Example of circles centered at the eigenvalues constructed in the clustering algorithm. The dots inside the solid circle are the candidates for a cluster.

is implemented as in [21,22], it may be overly conservative in its construction of the circles. We extend the strategy to take previously computed circles into account and we get a sharper heuristic in the sense that it allows for larger radii for some of the clustered eigenvalues. Both algorithms are O(n^3), but the new algorithm requires some extra data structures, in particular, it uses an $n \times n$ matrix for temporary storage.

If we want to find a cluster containing the largest eigenvalue in magnitude, we compute the Schur form and reorder the eigenvalues in decreasing order of magnitude along the diagonal of the Schur triangle [15,21,22]. Eigenvalues are added to the candidate set for a cluster in the order given by the Schur triangle, and we consider step s with $S = \{\lambda_1, \ldots, \lambda_s\}$. The idea is to construct circles centered at each eigenvalue with as large radii as possible without intersecting any of the circles in the other set. For each eigenvalue, we require the radius to be greater than or equal to half the distance to the closest eigenvalue in the other subset, i.e., if the eigenvalue is in S then we take the distance to the closest eigenvalue in the complement of S and vice versa. See figure 1.

Define a matrix v by $v(i,j) = \frac{1}{2}|\lambda_i - \lambda_j|$, if one and only one of λ_i and λ_j is in S, and $v(i,j) = \infty$, otherwise. The matrix elements are the maximum radius allowed considering only λ_i and λ_j, and we use it to construct the n-vector *rad* with the radius for each circle. The first radius is given by $\mathit{rad}(i) = v(i,j) = \min_{\mathit{rad}(s)=\infty} v(s,t)$, and the matrix is updated accordingly, $v(i,s) = v(s,i) = 2v(i,s) - \mathit{rad}(i)$ for $s = 1, \ldots, n$. The step is repeated $n-1$ more times to determine all elements of *rad*.

Inserting the radii into (8) starting with $i = 1$, we may compute the largest value for d_i satisfying the expression with equality. If $d_i \geqslant 1$ for $i = 1, \ldots, n$, then the

Algorithm 3. Clustering for matrices

Input: $A \in \mathbb{C}^{n\times n}$. A point λ_{goal} that we want to compute the cluster closest to. The size of the perturbation *tol* $\approx \|E\|_1$ known to affect the problem.
Output: λ_{cluster}, the mean of the s clustered eigenvalues.
function $[\lambda_{\text{cluster}}, s] =$ cluster(A, λ_{goal}, *tol*)
Compute the Schur form $Q^H AQ = S$.
Reorder the Schur form such that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are sorted in increasing order with respect to the distance to λ_{goal}.
$s = 0$
repeat
  $s = s + 1$
  Let $v(i,j) = \frac{1}{2}|\lambda_i - \lambda_j|$ if $i \leqslant s < j$ or $j \leqslant s < i$
  and $v(i,j) = \infty$ otherwise, for $i, j = 1, \ldots, n$.
  Let *rad*$(i) = \infty$ for $i = 1, \ldots, n$.
  for $i = 1$ **to** n
    Let $v(i,j) = \min_{\textit{rad}(s)=\infty} v(s,t)$.
    rad$(i) = v(i,j)$
    Let $v(i,k) = v(k,i) = 2v(i,k) -$ *rad*(i) for $k = 1, \ldots, n$.
  end
  $i = 0$
  repeat
    $i = i + 1$
    $\rho =$ *tol* $+ \sum_{j<i} |s_{ji}/d_j|$
    $d_i =$ *rad*$(i)/\rho$
  until $d_i < 1$ **or** $i = n$
  if $s = n - 1$ **and** $d_i < 1$ **then** $s = n$
until $d_n \geqslant 1$ **or** $s = n$
$\lambda_{\text{cluster}} = \sum_{i=1}^{s} \lambda_i / s$

Algorithm 3. Clustering of eigenvalues for matrices.

circles belonging to the cluster do not intersect the rest of the circles. Hence, a cluster of eigenvalues has been isolated. If any $d_i < 1$, then we add λ_{s+1} to the cluster S and repeat the procedure. Algorithm 3 shows the algorithm for computing the cluster closest to a given point in the complex plane.

5.2. A Gerschgorin heuristic for regular matrix pencils

Clustering for matrix pencils is more difficult than clustering for matrices. Although the Gerschgorin theorem can be modified for generalized eigenvalue problems, another problem is to justify the use of the mean to approximate a multiple eigenvalue. The problem appears when $A - \lambda B$ has finite eigenvalues in the neighborhood of infinity or $A - \lambda B$ is close to a singular pencil. The generalized eigenvalue problem $Ax = \lambda Bx$ is equivalent to the standard problem $B^{-1}Ax = \lambda x$ provided that B is nonsingular. Our assumption that B is well-conditioned with respect to inversion in

the generalized Arnoldi method also warrants the use of the traditional mean to approximate a multiple eigenvalue. A perturbation bound for this situation is presented in [20].

To define the distance between generalized eigenvalues it is necessary to consider the pencil in cross product form $\beta A - \alpha B$. In this form, the eigenvalue pairs (α_i, β_i) can be normalized and mapped to points on the unit sphere $|\alpha_i|^2 + |\beta_i|^2 = 1$. Considering the eigenvalues as points on the sphere the *chordal metric* defines the distance between generalized eigenvalues:

$$\operatorname{chord}(\lambda_1, \lambda_2) = \frac{|\lambda_1 - \lambda_2|}{\sqrt{1 + |\lambda_1|^2}\sqrt{1 + |\lambda_2|^2}}.$$

The Gerschgorin theorem for regular pencils $A - \lambda B$ states that the eigenvalues are contained in the region defined by the disks

$$|a_{ii} - z b_{ii}| \leqslant \sum_{\substack{j \neq i \\ j=1}}^{n} |a_{ji} - z b_{ji}|.$$

If k of the disks are isolated from the other disks, the union of the regions contain k eigenvalues [41]. It is not possible to use the inequalities defining the disks in this form since the parameter z is included on both sides. A corollary giving a bound based on the chordal metric can be proven [41]. Consider the perturbed pencil $\widetilde{A} - \lambda\widetilde{B} = (A + E) - \lambda(B + F)$. Inserting this in the inequalities above it follows that

$$\begin{aligned} |a_{ii} - z b_{ii}| &\leqslant \|E\|_1 + |z|\|F\|_1 + \sum_{j \neq i} |a_{ji} - z b_{ji}| \\ &\leqslant \begin{pmatrix} 1 & |z| \end{pmatrix} \begin{pmatrix} \|E\|_1 + \sum_{j \neq i} |a_{ji}| \\ \|F\|_1 + \sum_{j \neq i} |b_{ji}| \end{pmatrix}. \end{aligned}$$

By the Cauchy–Schwarz inequality and the definition of the chordal metric we get the easily computed bound

$$\operatorname{chord}\left(z, \frac{a_{ii}}{b_{ii}}\right) \leqslant \sqrt{\frac{\|E\|_1^2 + \|F\|_1^2 + \sum_{j \neq i} |a_{ji}|^2 + \sum_{j \neq i} |b_{ji}|^2}{|a_{ii}|^2 + |b_{ii}|^2}}.$$

The clusters of eigenvalues are determined by computing the generalized Schur form of $A - \lambda B$ and grouping the eigenvalues by searching for intersecting regions. Algorithm 4 shows the pseudocode for computing the cluster closest to a given point in the complex plane.

Algorithm 4. Clustering for matrix pencils

Input: $A, B \in \mathbb{C}^{n\times n}$ such that $A - \lambda B$ is regular and B is well-conditioned with respect to inversion. A point λ_{goal} that we want to compute the cluster closest to. The size of the perturbations *tola* $\approx \|E\|_1$ and *tolb* $\approx \|F\|_1$ known to affect the problem.
Output: λ_{cluster}, the mean of the s clustered eigenvalues.
function$[\lambda_{\text{cluster}}, s] = \text{cluster}(A, B, \lambda_{\text{goal}}, \mathit{tola}, \mathit{tolb})$

Compute the generalized Schur form $P^H(A - \lambda B)Q = S - \lambda T$.
$\lambda_i = s_{ii}/t_{ii}$ for $i = 1, \ldots, n$.
Reorder the generalized Schur form such that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are sorted in increasing order with respect to the distance to λ_{goal}.
$s = 0$
repeat
  $s = s + 1$
  Let $v(i, j) = \frac{1}{2}\text{chord}(\lambda_i, \lambda_j)$ if $i \leqslant s < j$ or $j \leqslant s < i$
  and $v(i, j) = \infty$ otherwise, for $i, j = 1, \ldots, n$.
  Let $\mathit{rad}(i) = \infty$ for $i = 1, \ldots, n$.
  for $i = 1$ **to** n
    Let $v(i, j) = \min_{\mathit{rad}(s)=\infty} v(s, t)$.
    $\mathit{rad}(i) = v(i, j)$
    Let $v(i, k) = v(k, i) = 2v(i, k) - \mathit{rad}(i)$ for $k = 1, \ldots, n$.
  end
  $i = 0$
  repeat
    $i = i + 1$
    $$\rho = \sqrt{\frac{\mathit{tola}^2 + \mathit{tolb}^2 + \sum_{j<i} |s_{ji}|^2 + \sum_{j<i} |t_{ji}|^2}{|s_{ii}|^2 + |t_{ii}|^2}}$$
    $d_i = \mathit{rad}(i)/\rho$
  until $d_i < 1$ **or** $i = n$
  if $s = n - 1$ **and** $d_i < 1$ **then** $s = n$
until $d_n \geqslant 1$ **or** $s = n$
$\lambda_{\text{cluster}} = \sum_{i=1}^{s} \lambda_i/s$

Algorithm 4. Clustering of eigenvalues for matrix pencils.

6. The Arnoldi method

The Arnoldi method provides an efficient way to compute an invariant subspace of a large scale matrix A and the restriction of A with respect to this subspace [15,33]. The method is based on the Krylov sequence

$$\mathcal{K}_i(A, v_1) = \{v_1, Av_1, \ldots, A^{i-1}v_1\}.$$

The Krylov sequence has the important property that it converges to the eigenvector corresponding to the largest eigenvalue in magnitude, provided that the starting vector v_1 contains a component in this direction. In the rest of this section we discuss

the Arnoldi method and the important concept of implicit restarts used to reduce the number of floating point operations [27].

In the kth step of the Arnoldi method, Gram–Schmidt orthogonalization is applied to produce a basis for the space spanned by the first k vectors in the Krylov sequence. Let $v_1 \in \mathbb{C}^n$, $\|v_1\|_2 = 1$, be the starting vector. The method is defined by

$$\begin{aligned} h_{ij} &= v_i^H A v_j, \quad i \leqslant j, & \hat{v}_{j+1} &= A v_j - \sum_{i \leqslant j} h_{ij} v_i, \\ h_{j+1j} &= \left\|\hat{v}_{j+1}\right\|_2, & v_{j+1} &= \frac{\hat{v}_{j+1}}{\|\hat{v}_{j+1}\|_2}. \end{aligned} \tag{9}$$

The recursive procedure builds a factorization – the *Arnoldi factorization* – of the form

$$AV_i = V_i H_{ii} + f_i e_i^{\mathrm{T}},$$

where H_{ii} is in unreduced Hessenberg form and $f_i = h_{i+1i} v_{i+1}$. The residual $f_i e_i^{\mathrm{T}}$ is orthogonal to V_i. Algorithm 5 extends an Arnoldi factorization of dimension k to a new factorization of higher dimension. The vectors $v_1, v_2, \ldots, v_i$ generated by the Arnoldi method span the same space as the Krylov sequence $\mathcal{K}_i(A, v_1)$. The Krylov sequence converges to the dominant eigenvector and the residual $f_i e_i^{\mathrm{T}}$ converges to zero producing an invariant subspace spanned by V_i and a restriction to this subspace H_{ii}. The method typically converges to a restriction containing the largest eigenvalues in moduli of A. To compute eigenvalues of the interior spectrum, it is necessary to transform A such that the eigenvalues of interest are mapped to the dominant part of the spectrum. This is often achieved by applying the method to a shifted and inverted matrix $A - \sigma I$ [33]. In this paper, we assume that the problem somehow

Algorithm 5. Arnstep

Input: $A \in \mathbb{C}^{n\times n}$, $V \in \mathbb{C}^{n\times k+1}$ and $H \in \mathbb{C}^{k+1\times k}$ such that $AV(:, 1:k) = V(:, 1:k+1)H$ is an Arnoldi factorization. The number of steps p.
Output: $V \in \mathbb{C}^{n\times k+p+1}$ and $H \in \mathbb{C}^{k+p+1\times k+p}$ such that $AV(:, 1:k+p) = V(:, 1:k+p+1)H$ is an Arnoldi factorization.
function $[H, V] = \text{arnstep}(A, H, V, k, p)$
$j = k$
while $j \leqslant k + p$ **do**
  $V(:, j+1) = AV(:, j)$
  for $i = 1$ **to** j
    $H(i, j) = V(:, i)^H V(:, j+1)$
    $V(:, j+1) = V(:, j+1) - H(i, j)V(:, i)$
  end
  $H(j+1, j) = \|V(:, j+1)\|_2$
  $V(:, j+1) = V(:, j+1)/H(j+1, j)$
  $j = j + 1$
end

Algorithm 5. Extends an Arnoldi factorization of dimension k to a factorization of dimension $k + p$.

has been transformed such that the eigenvalues of interest correspond to the dominant ones. Convergence characteristics of the Arnoldi method and other Krylov methods are discussed in detail in [5,33].

If $\|f_i\|_2 = 0$, then the columns of V_i span an invariant subspace. An eigenpair (θ, y) of H_{ii} satisfies $H_{ii}y = \theta y$ and it follows that $(\theta, V_i y)$ is an eigenpair of A. When the method converges, the residual $f_i e_i^{\mathrm{T}}$ becomes small and when it is smaller than some given threshold, it is set to zero. Approximations for the eigenvalues are computed from the small matrix H_{ii}. This approximation procedure is called the Rayleigh–Ritz process and the approximations to the eigenvalues are referred to as Ritz values. The residual of the Ritz pair $(\theta, V_i y)$ is

$$\|AV_i y - \theta V_i y\| = \left\|f_i e_i^{\mathrm{T}} y\right\| = h_{j+1j}\left|e_i^{\mathrm{T}} y\right|.$$

The vector space $\mathbb{C}^n$ can be expressed as a direct sum of subspaces annihilated by the transformations $(A-\lambda_i I)^{h_i}$, (see equation (1)). The polynomial $p(A) = \prod_{i=1}^{t}(A-\lambda_i I)^{h_i}$ maps the vector space $\mathbb{C}^n$ to zero. The dimension of the Krylov space is therefore bounded by $\sum_{i=1}^{t} h_i$ and the Arnoldi method must terminate within this number of steps. The matrix H_{ii} is in unreduced Hessenberg form implying that all the Ritz values have geometrical multiplicity one – H_{ii} is nonderogatory. For a derogatory matrix A, the restriction H_{ii} can in theory only contain information about one Jordan block for each eigenvalue. If the starting vector contains a component in the direction of the principal vector of highest grade, then H_{ii} will include information about a Jordan block of maximum dimension. In practice, the computations are carried out in finite precision and we may benefit from this in the sense that more than one block may be revealed. Unfortunately, we have found through experiments that relying on this property to find information about the complete Jordan structure of the dominant eigenvalue is not a safe and reliable procedure and we do not advise this approach. Instead, we use *explicit restarts* to extract information about the Jordan blocks of the matrix that were not computed in the previous applications of the method (see section 6.3).

6.1. The implicitly restarted Arnoldi method

The number of operations of the Arnoldi method increases proportionally to the number of columns of V. The ith step of the Arnoldi method requires O($i^2 n$) floating point operations. The payoff from limiting the number of vectors involved in the procedure is therefore substantial. In [27,37], *implicit restarts* are presented as a strategy for transforming the Arnoldi factorization to reduce the number of vectors and filter out the unwanted part of the spectrum. Consider the $r = (k + p)$-dimensional Arnoldi factorization

$$AV_r = V_r H_{rr} + f_r e_r^{\mathrm{T}}.$$

Suppose the iteration is to be restarted with a k-dimensional factorization. First, the Ritz values of H_{rr} are computed. Then, p Ritz values $\rho_1, \rho_2, \ldots, \rho_p$ are chosen to be

removed from the current factorization. These Ritz values are moved to the bottom right part of the factorization by applying p explicit QR iterations

$$H_{rr}^{(i-1)} - \rho_i I = Q_i R_i,$$
$$H_{rr}^{(i)} = Q_i^H H_{rr}^{(i-1)} Q_i = R_i Q_i + \rho_i I.$$

Let $Q = Q_1 \cdots Q_p$ and $\widetilde{H}_{rr} = H_{rr}^{(p)}$. The Arnoldi factorization

$$AV_r = V_r Q \widetilde{H}_{rr} Q^H + f_r e_r^{\mathrm{T}}$$

is postmultiplied by Q and by defining $\widetilde{V}_r = V_r Q$ we get

$$A\widetilde{V}_r = \widetilde{V}_r \widetilde{H}_{rr} + f_r e_r^{\mathrm{T}} Q.$$

The matrix $\widetilde{H}_{rr}$ is also in unreduced Hessenberg form. Moreover, since the unitary matrix Q is in Hessenberg form it follows that

$$e_{k+p}^{\mathrm{T}} Q = (\underbrace{0, \ldots, 0}_{k-1}, q_{rk}, \ldots, q_{rr}).$$

The leading $k \times k$ part of the factorization is, therefore,

$$A\widetilde{V}_k = \widetilde{V}_k \widetilde{H}_{kk} + \big(\tilde{h}_{k+1k} \tilde{v}_{k+1} + f_r q_{rk}\big) e_k^{\mathrm{T}},$$

and by redefining $V_k = \widetilde{V}_k$, $H_k = \widetilde{H}_{kk}$ and $f_k = \tilde{h}_{k+1k} \tilde{v}_{k+1} + f_r q_{rk}$, we get an Arnoldi factorization of dimension k. The Arnoldi method can now be applied to the factorization $AV_k = V_k H_{kk} + f_k e_k^{\mathrm{T}}$ of dimension k and thereby extending it to a new Arnoldi factorization of dimension $k + p$. The iteration proceeds in this manner by restarting and extending the factorization until the implicitly restarted method has converged.

6.2. Stopping criterion for the invariant subspace

A concern of all iterative methods is to decide when to stop and declare the iteration to be converged. In the implicitly restarted Arnoldi method the subdiagonal elements of the Hessenberg matrix H_{k+1k} approaches zero when the method converges. If one of the subdiagonal elements is exactly zero, then the subspace spanned by the leading columns of V is invariant. In practice, we find a situation similar to the one for the QR algorithm where the subdiagonal elements become small but not exactly zero. For our experiments we use an approach similar to the criterion used in the QR method. The iteration is considered to have converged if

$$\frac{\|AV_i - V_i H_{ii}\|_2}{\|A\|_2} = \frac{|h_{i+1i}|}{\|A\|_2} \leqslant \varepsilon,$$

for any i smaller than the dimension of the Arnoldi factorization. Computing $\|A\|_2$ can be expensive and we instead use the bound

$$\frac{\|AV_i - V_iH_{ii}\|_2}{\|A\|_F} = \frac{\|AV_i - V_iH_{ii}\|_F}{\|A\|_F} = \frac{|h_{i+1i}|}{\|A\|_F} \leqslant \varepsilon. \tag{10}$$

The stopping criterion is backward stable since setting a subdiagonal element to zero causes a perturbation in A of size less or equal to the deleted element. This criterion and other similar criteria are discussed in [3,4,34].

6.3. *Arnoldi factorization associated with the dominant eigenvalue*

In exact arithmetic the IRA method converges to an invariant subspace containing information about one Jordan chain per starting vector. To extract information about the complete Jordan structure associated with the dominant eigenvalue, the method must be restarted with a new starting vector when the residual vanishes. Deflation is performed by orthogonalizing the vectors in the successive Arnoldi applications with respect to the invariant subspace computed during the previous steps [33]. Consider the converged Arnoldi factorization

$$AV_1 \approx V_1H_{11}.$$

We use the notation $\approx$ to show that equality is only satisfied with respect to the stopping criterion of the iterative process. A new starting vector u_1 is chosen orthogonal to the space spanned by V_1. The factorization is easily extended to a factorization of the form

$$A\begin{pmatrix} V_1 & u_1 \end{pmatrix} = \begin{pmatrix} V_1 & u_1 \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} \\ 0 & h_{22} \end{pmatrix} + re_i^{\mathrm{T}}. \tag{11}$$

The IRA method is then applied to the new factorization (11). The resulting restriction H is no longer an unreduced Hessenberg matrix. Instead, we get a Hessenberg matrix H with the number of zeros on the subdiagonal equal to the number of explicit restarts. After a sequence of s explicit restarts we get a converged factorization of the form

$$A\begin{pmatrix} V_1 & \dots & V_s \end{pmatrix} \approx \begin{pmatrix} V_1 & \dots & V_s \end{pmatrix} \begin{pmatrix} H_{11} & H_{12} & \dots & H_{1s} \\ 0 & H_{22} & \dots & H_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & H_{ss} \end{pmatrix}.$$

Let the dimension k used in the Arnoldi procedure be greater than the height h_i of the dominant eigenvalue and assume that all starting vectors contain components in the direction of the principal vector of highest grade not already included in V. Then, provided that the number of explicit restarts s is greater than the number of Jordan blocks – the geometric multiplicity $n_{\mathrm{g}}(\lambda_i)$ of the dominant eigenvalue – the factorization will contain all the canonical information associated with the dominant eigenvalue. A high-level description of the IRA method with prescribed number of explicit restarts is presented in algorithm 6.

Algorithm 6. IRA with prescribed number of explicit restarts

Input: $A \in \mathbb{C}^{n\times n}$, two numbers k and p describing the dimensions of the Arnoldi factorization to be used. The number s of explicit restarts – find at least s blocks. *tol* is the tolerance for the convergence criterion.
Output: $V \in \mathbb{C}^{n\times lock}$ and $H \in \mathbb{C}^{lock\times lock}$ such that $AV \approx VH$.
function $[H, V] =$ ira($A, k, p,$ *tol*, s)
lock $= 0$, *iter* $= 1$, $m = k + p$
while *iter* $\leqslant s$ **and** *lock* $\leqslant n$ **do**
  $V = V(:, 1 : lock)$
  Choose v_{lock+1} randomly s.t. $\|v_{lock+1}\|_2 = 1$ and $v_{lock+1} \perp V(:, 1 : lock)$
  $[H, V] =$ arnstep($A, H, V, lock + 1, k - 1$)
  dfl $=$ *lock*, $l = lock + 1$
  while *dfl* $=$ *lock* **do**
    $[H, V] =$ arnstep($A, H, V, k + lock, p$)
    Let the vector *ritz* be the eigenvalues of $H(l : m, l : m)$
    sorted in increasing order.
    $e = (0, \ldots, 0, 1)$, $\eta = H(m + 1, m)$
    for $i = 1$ **to** p
      $[Q, R] =$ qr($H(l : m, l : m) -$ *ritz*$(i)I$)
      $e = eQ$
      $V(:, l : m) = V(:, l : m)Q$
      $H(1 :$ *lock*$, l : m) = H(1 :$ *lock*$, l : m)Q$
      $H(l : m, l : m) = RQ +$ *ritz*$(i)I$
    end
    $e = H(l + k,$ *lock* $+ k)V(:, l + k) + \eta e(k)V(:, m + 1)$
    $h = (0, \ldots, 0, \|e\|_2)$
    $H = [H(1 : k +$ *lock*$, 1 : k +$ *lock*$); h]$
    $V = [V(:, 1 : k +$ *lock*$), e/\|e\|_2]$
    Let *dfl* be the largest index such that $H($*dfl* $+ 1,$ *dfl*$)$ satisfies the
    stopping criteria for *tol*.
  end
  lock $=$ *dfl*, $m =$ *lock* $+ k + p$, *iter* $=$ *iter* $+ 1$
end
$H = H(1 :$ *lock*$, 1 :$ *lock*$)$
$V = V(:, 1 :$ *lock*$)$

Algorithm 6. Implicitly restarted Arnoldi method with prescribed number of explicit restarts.

7. A generalized Arnoldi method

The Arnoldi method can be modified to compute a subset of the eigenvalues of a generalized eigenvalue problem $Ax = \lambda Bx$. Krylov methods for the generalized eigenvalue problem are discussed in for instance [28,32,33,38]. If the matrix B is non-singular then the eigenvalue problem $Ax = \lambda Bx$ is in theory equivalent to the standard problem $AB^{-1}y = \lambda y$ with $y = Bx$. This relation is the key to deriving an Arnoldi

method for the generalized eigenvalue problem. Consider the Arnoldi factorization

$$AB^{-1}W_i = W_i G_{ii} + r_i e_i^{\mathrm{T}}.$$

Suppose that

$$B^{-1}W_i = V_i U_{ii}^{-1}$$

is the QR factorization of $B^{-1}W_i$ where $V_i \in \mathbb{C}^{n\times i}$ has orthonormal columns and $U_{ii}^{-1} \in \mathbb{C}^{i\times i}$ is upper triangular. Define the Hessenberg matrix $H_{ii} = G_{ii}U_{ii}^{-1}$ and the vector $f_i = u_{ii}r_i$. By inserting the definitions into the Arnoldi factorization for AB^{-1}, we get a *generalized Arnoldi factorization* of the matrix pencil $A - \lambda B$:

$$\begin{aligned} AV_i &= W_i H_{ii} + f_i e_i^{\mathrm{T}}, \\ BV_i &= W_i U_{ii}. \end{aligned} \tag{12}$$

The residual $f_i e_i^{\mathrm{T}}$ is orthogonal to W_i. A recurrence similar to (9) can be derived by applying Gram–Schmidt orthogonalization to the vectors in the Krylov sequence $\mathcal{K}(AB^{-1}, w)$. Let $w_1 \in \mathbb{C}^n$, $\|w_1\|_2 = 1$, be the starting vector. Let $z = V_{j-1}^H B^{-1} w_j$. The vector v_j is then computed by orthogonalizing $B^{-1}w_j$ with respect to the columns of V_{j-1},

$$\hat{v}_j = B^{-1}w_j - V_{j-1}z, \qquad v_j = \frac{\hat{v}_j}{\|\hat{v}_j\|_2}.$$

From (12) it follows that the last (jth) column of U_{jj} can be computed as

$$u_j = \frac{1}{\|\hat{v}_j\|_2} \begin{pmatrix} -U_{j-1j-1}z \\ 1 \end{pmatrix}.$$

Inserting V_j and U_{jj} in (12), we may then derive the vector w_{j+1} and the jth column of H_{j+1j}:

$$\begin{aligned} h_{ij} &= w_i^H A v_j, \quad i \leqslant j & \widehat{w}_{j+1} &= Av_j - \sum_{i\leqslant j} h_{ij} w_i, \\ h_{j+1j} &= \|\widehat{w}_{j+1}\|_2, & w_{j+1} &= \frac{\widehat{w}_{j+1}}{\|\widehat{w}_{j+1}\|_2}. \end{aligned}$$

In algorithm 7, the recurrence is used to extend a generalized Arnoldi factorization to a factorization of higher dimension.

The Krylov sequence $\mathcal{K}(AB^{-1}, w_1)$ converges to the eigenvector associated with the dominant eigenvalue provided that w_1 has a component in this direction. The vectors $w_1, w_2, \ldots, w_i$ in the generalized Arnoldi method span the same subspace as the first i terms in the Krylov sequence $\mathcal{K}(AB^{-1}, w_1)$. If the Krylov sequence converges, then the residual $f_i e_i^{\mathrm{T}}$ in (12) goes to zero. For a converged generalized Arnoldi factorization with a zero residual, W_i and V_i span the left and right eigenspaces of $A - \lambda B$ and $H_{ii} - \lambda U_{ii}$ is the restriction with respect to the pair of eigenspaces. Moreover, if (θ, y) is an eigenpair of $H_{ii} - \lambda U_{ii}$, then $(\theta, V_i y)$ is an eigenpair of

Algorithm 7. Garnstep

Input: $A, B \in \mathbb{C}^{n\times n}$ such that $A - \lambda B$ is a regular pencil with a nonsingular B.
$V \in \mathbb{C}^{n\times k}$, $W \in \mathbb{C}^{n\times k+1}$, $H \in \mathbb{C}^{k+1\times k}$ and $U \in \mathbb{C}^{k\times k}$ describing a generalized Arnoldi factorization (12) of dimension k. The number of steps p.
Output: $V \in \mathbb{C}^{n\times k+p}$, $W \in \mathbb{C}^{n\times k+p+1}$, $U \in \mathbb{C}^{k+p\times k+p}$ and $H \in \mathbb{C}^{k+p+1\times k+p}$ describing a generalized Arnoldi factorization (12) of dimension $k + p$.
function $[H, U, V, W]$ = garnstep(A, B, H, U, V, W, k, p)
$j = k$
while $j \leqslant k + p - 1$ **do**
  $j = j + 1$
  Solve $BV(:, j) = W(:, j)$
  for $i = 1$ **to** $j - 1$
    $z(i) = V(:, i)^H V(:, j)$
    $V(:, j) = V(:, j) - z(i)V(:, i)$
  end
  $\rho = \|V(:, j)\|_2$
  $V(:, j) = V(:, j)/\rho$
  if $j > 1$ **then**
    $U(:, j) = -Uz/\rho$
  end
  $U(j, j) = 1/\rho$
  $W(:, j + 1) = AV(:, j)$
  for $i = 1$ **to** j
    $H(i, j) = W(:, i)^H HW(:, j + 1)$
    $W(:, j + 1) = W(:, j + 1) - H(i, j)W(:, i)$
  end
  $H(j + 1, j) = \|W(:, j + 1)\|_2$
  $W(:, j + 1) = W(:, j + 1)/H(j + 1, j)$
end

Algorithm 7. Extends a k-dimensional generalized Arnoldi factorization to a factorization of dimension $k + p$.

$A - \lambda B$. When the residual is sufficiently small, the eigenvalues of $H_{ii} - \lambda U_{ii}$ can be used to approximate the eigenvalues of the original pencil $A - \lambda B$. The unreduced Hessenberg-triangular form $H_{ii} - \lambda U_{ii}$ implies that its WCF can only contain one Weierstrass block per eigenvalue. Therefore, in exact arithmetic we can only compute eigenspaces containing information about one Weierstrass block per starting vector.

We propose a stopping criterion which is a straight forward generalization of the criterion used in the standard case. Let ε be a small number. The iteration is considered to have converged if

$$\frac{\|AV_i - W_i H_{ii}\|_2}{\max(\|A\|_2, \|B\|_2)} = \frac{|h_{i+1i}|}{\max(\|A\|_2, \|B\|_2)} \leqslant \varepsilon,$$

for any i less than the dimension of the generalized Arnoldi factorization. Since it is expensive to compute $\max(\|A\|_2, \|B\|_2)$, we replace the 2-norm with the Frobenius norm and use the bound

$$\frac{\|AV_i - W_i H_{ii}\|_2}{\max(\|A\|_F, \|B\|_F)} = \frac{\|AV_i - W_i H_{ii}\|_F}{\max(\|A\|_F, \|B\|_F)} = \frac{|h_{i+1i}|}{\max(\|A\|_F, \|B\|_F)} \leqslant \varepsilon, \qquad (13)$$

which is easier to compute. The deflation is backward stable since the backward perturbation introduced by deflation is bounded by the size of the element in H that is set to zero.

7.1. An implicitly restarted Arnoldi method for the generalized eigenvalue problem

In analogy with the standard problem, the generalized Arnoldi factorization of dimension $r = k + p$

$$\begin{aligned} AV_r &= W_r H_{rr} + f_r e_r^{\mathrm{T}}, \\ BV_r &= W_r U_{rr}, \end{aligned}$$

can be transformed to filter out p unwanted eigenvalues [38]. Suppose that the Ritz values $\rho_1, \rho_2, \ldots, \rho_p$ of $H_{rr} - \lambda U_{rr}$ are to be removed from the current factorization. The Ritz values are moved to the bottom right part of the factorization by applying p explicit QZ iterations using the chosen Ritz values for shifts,

$$\begin{aligned} H_{rr}^{(i-1)} - \rho_i U_{rr}^{(i-1)} &= Q_i R_i, \\ Q_i^H U_{rr}^{(i-1)} &= S_i Z_i^H, \\ H_{rr}^{(i)} = Q_i^H H_{rr}^{(i-1)} Z_i &= R_i Z_i + \rho_i S_i, \\ U_{rr}^{(i)} &= S_i. \end{aligned}$$

Let $Q = Q_1 \cdots Q_p$, $\widetilde{H}_{rr} = H_{rr}^{(p)}$, $Z = Z_1 \cdots Z_p$ and $\widetilde{U}_{rr} = U_{rr}^{(p)}$. The transformed factorization is then

$$\begin{aligned} AV_r Z &= W_r Q \widetilde{H}_{rr} + f_r e_r^{\mathrm{T}} Z, \\ BV_r Z &= W_r Q \widetilde{U}_{rr}. \end{aligned}$$

The unitary transformation Z is in Hessenberg form and the leading $k \times k$ part of the factorization is a generalized Arnoldi factorization. We let $V_r = V_r Z$, $W_r = W_r Q$, $H_{rr} = \widetilde{H}_{rr}$ and $U_{rr} = \widetilde{U}_{rr}$ define a new generalized Arnoldi factorization of dimension k. The method is then restarted with the new $k \times k$ factorization discarding the trailing part of the $r \times r$ factorization. An algorithm with prescribed number of explicit restarts is displayed in algorithm 8.

8. The partial Jordan–Schur method

Our method for computing a partial Jordan–Schur form of a matrix, including complete canonical information for the dominant eigenvalue, is based on the build-

Algorithm 8. Generalized IRA with explicit restarts

Input: $A, B \in \mathbb{C}^{n\times n}$ such that $A - \lambda B$ is a regular pencil with a nonsingular B, the dimensions k and p of the generalized Arnoldi factorization to be used, the number of explicit restarts s and the tolerance *tol* for the convergence criterion.
Output: $V, W \in \mathbb{C}^{n\times lock}$ and $H, U \in \mathbb{C}^{lock\times lock}$ such that $(A - \lambda B)V \approx W(H - \lambda U)$.
function $[H, U, V, W] =$ gira($A, B, k, p,$ *tol*, s)
lock $= 0,$ *iter* $= 1,\ m = k + p$
while *iter* $\leqslant s$ **and** *lock* $\leqslant n$ **do**
  $V = V(:, 1 : lock)$
  Choose w_{lock+1} randomly s.t. $\|w_{lock+1}\|_2 = 1$ and $w_{lock+1} \perp W(:, 1 : lock)$.
  $[H, U, V, W] =$ garnstep($A, B, H, U, V, W,$ *lock*, k)
  dfl $=$ *lock*, $l =$ *lock* $+ 1$
  while *dfl* $=$ *lock* **do**
    $[H, U, V, W] =$ garnstep($A, B, H, U, V, W, k +$ *lock*, p)
    Let the vector *ritz* be the eigenvalues of the matrix pair
    $(H(l : m, l : m), U(l : m, l : m))$ sorted in increasing order.
    $e = (0, \ldots, 0, 1), \eta = H(m + 1, m)$
    for $i = 1$ **to** p
      $[Q, R] =$ qr($H(l : m, l : m) -$ *ritz*$(i)U(l : m, l : m)$)
      $[T, Z] =$ rq($Q^H U(l : m, l : m)$)
      $V(:, l : m) = V(:, l : m)Z^H,\ W(:, l : m) = W(:, l : m)Q$
      $H(1 :$ *lock*$, l : m) = H(1 :$ *lock*$, l : m)Z^H,\ H(l : m, l : m) = RZ^H +$ *ritz*$(i)T$
      $U(1 :$ *lock*$, l : m) = U(1 :$ *lock*$, l : m)Z^H,\ U(l : m, l : m) = T$
      $e = eZ^H$
    end
    $e = H(l + k,$ *lock* $+ k)W(:, l + k) + \eta e(k)W(:, m + 1)$
    $h = (0, \ldots, 0, \|e\|_2)$
    $H = [H(1 : k +$ *lock*$, 1 : k +$ *lock*$); h],\ U = U(1 : k +$ *lock*$, 1 : k +$ *lock*$)$
    $V = V(:, 1 : k +$ *lock*$),\ W = [W(:, 1 : k +$ *lock*$), e/\|e\|_2]$
    Let *dfl* be the largest index such that $H($*dfl* $+ 1,$ *dfl*$)$ satisfies the
    stopping criteria for *tol*.
  end
  lock $=$ *dfl*, $m =$ *lock* $+ k + p,$ *iter* $=$ *iter* $+ 1$
end
$H = H(1 :$ *lock*$, 1 :$ *lock*$),\ U = U(1 :$ *lock*$, 1 :$ *lock*$)$
$V = V(:, 1 :$ *lock*$),\ W = W(:, 1 :$ *lock*$)$

Algorithm 8. Implicitly restarted Arnoldi method for generalized eigenvalue problems with prescribed number of explicit restarts.

ing blocks presented in sections 4.1, 5.1 and 6. We give an overview of the partial Jordan–Schur method and discuss its reliability from the perspective of deflation errors introduced in the Arnoldi method and the staircase algorithm. Moreover, we present and discuss results from computational experiments conducted on problems with known canonical structure and varying ill-conditioning.

8.1. Algorithm and deflation errors

The first step in the partial Jordan–Schur form algorithm is to compute an approximate invariant subspace containing the subspace spanned by the principal vectors of the dominant eigenvalue. We use the IRA method with s explicit restarts to compute the invariant subspace of interest. Let k and p be the dimensions used in the implicit restarts. In order to find complete chains of principal vectors, the dimension k must be greater than the size of the largest Jordan block, h_i, and s must be greater than or equal to the number of Jordan blocks, $n_{\mathrm{g}}(\lambda_i)$. In the implicit restarts, we filter out the p smallest Ritz values keeping the k Ritz values of largest moduli. Moreover, for the method to converge to a subspace containing the space of principal vectors, it is necessary that the space spanned by the starting vectors has a component in the direction of all of the highest grade principal vectors. Provided that the implicit IRA method with s explicit restarts converges we get an Arnoldi factorization

$$AV = VH + E,$$

where the Arnoldi residual E is a matrix with small entries. The stopping criterion (10) ensures that $\|E\|_F = \mathrm{O}(tol\|A\|_F)$. The matrices V and H describe a converged Arnoldi factorization of the nearby matrix:

$$\left(A - EV^H\right)V = VH.$$

In [23], Kahan et al. examine the general problem for the left and right residuals and the relation between the residuals and the backward error.

In the next step, we use the clustering heuristic to isolate a group of eigenvalues of H containing the largest eigenvalue in magnitude. We use the average of the eigenvalues in the cluster as an approximate multiple eigenvalue. The size of the perturbation $\|E\|_1$ in the clustering heuristic is approximated by $tol\|A\|_F$, i.e., the deflation tolerance used in the Arnoldi step. If the clustering is successful, then the average λ_{cluster} is a good approximation to the dominant eigenvalue of A.

Finally, the staircase algorithm is applied to the shifted problem $H - \lambda_{\text{cluster}}I$ to find the Jordan–Schur form for λ_{cluster}. The deflation criterion used in the staircase algorithm ($\sigma_{n-k+1} > tol\|A\|_F$ and $\sigma_{n-k+1} > gap \cdot \sigma_{n-k}$, see section 4.1), implies that the dimension of the space of principal vectors in each step is taken as large as possible and short vector chains are preferred over longer ones in the computation. This property of the staircase algorithm corresponds to finding the least generic Jordan structure with respect to the deflation criterion [9,10] (see section 4.1).

As the result from the staircase algorithm, we get a decomposition

$$Q^H HQ \equiv \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} + F, \tag{14}$$

where $S_{11} = \lambda_{\text{cluster}}I + N$ with N in staircase form (5). F is the perturbation introduced by discarding small singular values in the orthogonal deflation process for computing N. In general, S_{11} reveals the Jordan structure of λ_{cluster} which is an exact

eigenvalue of $H - FQ^H$. If $\|F\|_F$ is small – normally of size O(*tol*$\|A\|_F$) – then we accept λ_{cluster} as an approximate multiple eigenvalue of H with the canonical structure given by the block structure of N.

By inserting the decomposition (14) in the Arnoldi factorization and multiplying by Q from the left, we get

$$\big(A - EV^H\big)VQ = VQS + QF.$$

The perturbation QF can be considered a perturbation in A,

$$\big(A - EV^H - QFV^H\big)VQ = VQS.$$

Let $X = VQ$ and $G = -EV^H - QFV^H$. We then get an exact partial Jordan–Schur form with respect to λ_{cluster},

$$(A+G)\begin{pmatrix} X_1 & X_2 \end{pmatrix} = \begin{pmatrix} X_1 & X_2 \end{pmatrix}\begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix}.$$

If the method is successful and the problem has well defined canonical structure with respect to the deflation criteria, then the perturbation is small and $\|G\|_F = \text{O}(tol\|A\|_F)$. When the perturbation is small, we consider the factorization to be a partial Jordan–Schur form of A with respect to the dominant eigenvalue λ_{cluster}. We remark that for a converged partial Jordan–Schur form, the orthogonality of the columns of X is granted through the Gram–Schmidt orthogonalization processes in the IRA method with explicit restarts and the unitary similarity transformations used in the staircase algorithm. The method is summarized in algorithm 9.

One conclusion from the discussion above is that the clustering method and staircase algorithm used in the partial Jordan–Schur algorithm must take the norm of

Algorithm 9. Partial Jordan–Schur form

Input: $A \in \mathbb{C}^{n\times n}$. Two numbers k and p defining the IRA dimensions and the number s of explicit restarts. A deflation tolerance *tol* and the *gap* parameter used in the staircase algorithm.

Output: A triangular matrix S revealing the Jordan structure of the dominant eigenvalueand a rectangular matrix X. S and X define a partial Jordan–Schur form of A and satisfy $(A+G)X = XS$, where $\|G\|_F = \text{O}(tol\|A\|_F)$. The Weyr characteristics of the dominant eigenvalue λ_{cluster} of $A+G$ are listed in *str*.

function $[S, X, \lambda_{\text{cluster}}, str] = \text{pjs}(A, k, p, tol, gap, s)$
$[H, V] = \text{ira}(A, k, p, tol, s)$
$\lambda_{\text{cluster}} = \text{cluster}(H, \infty, tol)$
$[S_{\text{shifted}}, Q, str] = \text{staircase}(H - \lambda_{\text{cluster}}I, tol, gap)$
$S = S_{\text{shifted}} + \lambda_{\text{cluster}}I$
$X = VQ$

Algorithm 9. Algorithm for computing a partial Jordan–Schur form associated with the dominant eigenvalue of a matrix A.

the deflation errors from the Arnoldi method into account. We do this by using the same value on *tol* in the three building blocks.

8.2. Sample problems and computational experiments

We have implemented the partial Jordan–Schur method with full reorthogonalization in Matlab. For the computational experiments presented in this section we have used $tol = 10^{-9}$. The tolerance can be chosen smaller for problems with mild departure from normality, but for the most ill-conditioned problems the method does not converge with a smaller value on *tol*. In general, the main obstacle for computing partial forms of higher accuracy is the computation of the invariant subspace. The SVD variant of the staircase algorithm can be used with much stricter criteria for numerical rank determination. In the staircase algorithm we have used $gap = 1000$ for our tests. Whenever the *gap*-criterion for the numerical rank fails, we increase the size of the tolerance used to determine the numerical rank and thereby increase the dimension of the null space until we find a gap in the singular values.

For the computational experiments we have used a family of 500×500 dense, complex matrices. We consider matrices for which the dominant eigenvalue λ_{mult} has algebraic multiplicity $n_{\text{a}} = 12$ and Jordan structure $J_6 \oplus J_3 \oplus J_2 \oplus J_1$. The other $500 - n_{\text{a}}$ eigenvalues are chosen randomly smaller than λ_{mult}. Let

$$\Lambda_2 = \text{diag}(\lambda_{n_{\text{a}}+1}, \ldots, \lambda_{500}).$$

Moreover, we choose λ_{mult} such that the gap in the moduli of the eigenvalues

$$\min_{n_{\text{a}} < i \leqslant n} \frac{|\lambda_{\text{mult}}| - |\lambda_i|}{|\lambda_{\text{mult}}|}$$

is equal to a given parameter. The convergence rate of the Arnoldi method depends on the gap in the moduli of the eigenvalues and decreases when the gap is small [33]. Figure 2 shows the spectrum of a lower order matrix generated in this manner. The small circle is the multiple eigenvalue.

We then build a triangular matrix of the form

$$S = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} = \begin{pmatrix} \lambda_{\text{mult}} I & 0 \\ 0 & \Lambda_2 \end{pmatrix} + \begin{pmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{pmatrix}.$$

Notice that the block-partitioning of S above is only for the construction of the test matrices (S_{11} is $n_{\text{a}} \times n_{\text{a}}$, S_{22} is $(n - n_{\text{a}}) \times (n - n_{\text{a}})$). The block-partitioning of S in (14) corresponds to the computed partial Jordan–Schur form. If the algorithm is successful, S_{11} in (14) is of size $n_{\text{a}} \times n_{\text{a}}$ as well.

The strictly upper triangular matrix N is generated with random numbers and we scale the blocks N_{ij} to control the ill-conditioning of the eigenvalue problem. The block structure of N_{11} is chosen to get the desired Jordan structure for λ_{mult}. The connection between the size of the elements in N and the ill-conditioning of the

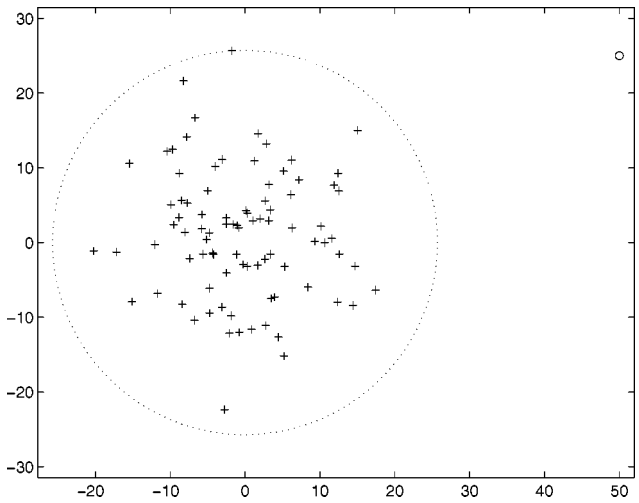

Figure 2. Spectrum of a smaller matrix from the same family as the test problems. The circle is the multiple eigenvalue.

multiple eigenvalue becomes clear when considering the block diagonalizing similarity transformation [41],

$$\begin{pmatrix} I & -X \\ 0 & I \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \begin{pmatrix} S_{11} & S_{11}X - XS_{22} - S_{12} \\ 0 & S_{22} \end{pmatrix}.$$

The invariant subspaces $(I, 0)^{\mathrm{T}}$ and $(X^{\mathrm{T}}, I)^{\mathrm{T}}$ are close to linearly dependent if $\|X\|_F$ is large. From the equation $S_{11}X - XS_{22} - S_{12} = 0$, it follows that $\|X\|_F$ is bounded by[1]

$$\|X\|_F \leqslant \frac{\|S_{12}\|_F}{\mathrm{sep}(S_{11}, S_{22})}.$$

Hence, $\|X\|_F$ can be large either if $\|N_{12}\|_F$ is large or if $\mathrm{sep}(S_{11}, S_{22})$ is small. Since $\mathrm{sep}(S_{11}, S_{22}) \leqslant \min_{n_a < i \leqslant n} |\lambda_{\mathrm{mult}} - \lambda_i|$, we see that decreasing the gap between λ_{mult} and the remaining eigenvalues also makes the separation between S_{11} and S_{22} smaller. The size of the elements in N_{11} controls the coupling among the spaces of different grade principal vectors. To get a dense matrix we generate a random unitary matrix Q and define

$$A = Q^H S Q.$$

To illustrate the performance of the partial Jordan–Schur method, we select eight matrices generated in this manner and apply the method to these with $k = 7$ and $p = 16$. We let four of the matrices have relative gap in the moduli of the eigenvalues

[1] Given two matrices A and B, the separation is defined as

$$\mathrm{sep}(A, B) = \inf_{\|X\|_F = 1} \|AX - XB\|_F.$$

equal to 1, and let the other four have gap equal to 0.25. Moreover, Dense A.1 and Dense A.25 have $\|N_{12}\|_F = \|N_{22}\|_F = 0$, i.e., except for the multiple eigenvalue all eigenvalues are normal. The departure from normality is very small and these are the most well-conditioned matrices of the eight matrices considered. For Dense B.1 and Dense B.25 the departure from normality is mild and equally distributed over the blocks of N. For Dense C.1 and Dense C.25 the blocks N_{12} have much larger entries than the other two blocks of N. Also, Dense D.1 and Dense D.25 have large relative departure from normality but the large elements are more evenly distributed over the blocks.

The results from our computational tests are summarized in tables 1–4. In table 1, some matrix characteristics are displayed, including the relative gap between eigenvalues and the distribution of the non-normality over the blocks of N. Table 2 shows that the number of matrix–vector multiplications and the number of implicit restarts for each explicit restart ($s = 4$) of the Arnoldi method both increase when the gap between the dominant (multiple) eigenvalue and the remaining spectrum de-

Table 1
The gap in the moduli of the eigenvalues and the relative departure from normality for the test matrices.

	$\frac{\min \lvert\lambda_{\text{mult}}\rvert - \lvert\lambda\rvert}{\lvert\lambda_{\text{mult}}\rvert}$	$\frac{\|N\|_F}{\|A\|_F}$	$\frac{\|N_{11}\|_F}{\|A\|_F}$	$\frac{\|N_{12}\|_F}{\|A\|_F}$	$\frac{\|N_{22}\|_F}{\|A\|_F}$
Dense A.1	1.00	0.00	0.00	0.00	0.00
Dense B.1	1.00	0.18	0.00	0.04	0.18
Dense C.1	1.00	0.60	0.03	0.59	0.13
Dense D.1	1.00	0.87	0.02	0.35	0.79
Dense A.25	0.25	0.00	0.00	0.00	0.00
Dense B.25	0.25	0.12	0.00	0.03	0.12
Dense C.25	0.25	0.50	0.02	0.48	0.11
Dense D.25	0.25	0.81	0.02	0.33	0.74

Table 2
The number of matrix–vector products and the number of implicit restarts for each explicit restart before the invariant subspaces converged. The relative Arnoldi residual.

	Matvec	Implicit restarts	$\frac{\|AV - VH\|_F}{\|A\|_F}$
Dense A.1	244	4, 3, 3, 2	7.6e−12
Dense B.1	208	3, 3, 2, 2	2.5e−10
Dense C.1	208	3, 3, 2, 2	2.1e−11
Dense D.1	226	3, 3, 3, 2	4.0e−10
Dense A.25	334	5, 4, 4, 4	7.9e−10
Dense B.25	478	7, 7, 6, 5	2.4e−10
Dense C.25	496	4, 6, 3, 13	1.4e−09
Dense D.25	496	5, 7, 3, 11	8.6e−10

Table 3
The number of eigenvalues isolated in the cluster and the relative error in the computed cluster mean.

	Cluster size	$\frac{\lvert\lambda_{\text{cluster}} - \lambda_{\text{mult}}\rvert}{\lvert\lambda_{\text{mult}}\rvert}$
Dense A.1	12	6.1e−16
Dense B.1	12	1.9e−13
Dense C.1	12	4.0e−13
Dense D.1	12	1.6e−11
Dense A.25	12	1.2e−13
Dense B.25	12	4.2e−11
Dense C.25	12	1.7e−11
Dense D.25	12	1.6e−10

Table 4
The Jordan structure (Segre characteristics) of the dominant eigenvalue in the computed partial Jordan–Schur form and the relative residual of the partial Jordan–Schur form.

	Computed structure	$\frac{\lVert AX - XS\rVert_F}{\lVert A\rVert_F}$
Dense A.1	6, 3, 2, 1	7.6e−12
Dense B.1	6, 3, 2, 1	2.5e−10
Dense C.1	6, 3, 2, 1	2.1e−11
Dense D.1	6, 3, 2, 1	4.1e−10
Dense A.25	6, 3, 2, 1	7.9e−10
Dense B.25	6, 3, 2, 1	2.4e−10
Dense C.25	6, 3, 2, 1	1.4e−09
Dense D.25	6, 3, 2, 1	8.6e−10

creases. The number of matrix–vector multiplications for the matrices Dense C.25 and Dense D.25 is large. This is the result of the large number of implicit restarts needed for convergence during the final step. A closer examination of the execution revealed that the method had captured all 12 dominant eigenvalues during the three previous steps, and the convergence is slow because the gap in moduli between the remaining eigenvalues is small. Compensating for the extra work, we find that only 262 multiplications were actually needed in these two cases. This is in accordance with the convergence properties of the implicit restarted Arnoldi method [27,33,37]. If we use $s \geqslant 5$, the convergence behavior deteriorates since the remaining spectra of our test problems do not have well-defined gaps between the individual eigenvalues. Table 3 shows the number of eigenvalues in the cluster and the relative error in the cluster mean, λ_{cluster}. We conclude that the clustering strategy works well and as expected the smallest error is for the matrices with large gap and small departure from normality. Finally, the computed Jordan structure and the residual for the computed partial Jordan–Schur form are reported in table 4. We see that for all test matrices the

computed structure is the correct one and the size of the final residuals are of the same size as the Arnoldi residuals in table 2.

If we continue to increase the departure from normality or decrease the gap between the dominant multiple eigenvalue and the remaining eigenvalues, it is easy to find matrices for which the method breaks down. We have found that for our implementation the method tends to break down already in the computation of the invariant subspace. Although it is easy to construct matrices with a very ill-conditioned Jordan–Schur form for which the staircase method fails, it seems that the main challenge lies in the computation of invariant subspaces for large matrices with a small gap in the eigenvalues or large departure from normality.

9. The partial Weierstrass–Schur method

The partial Weierstrass–Schur method is a straight forward generalization of the Jordan–Schur method to regular matrix pairs. Our algorithm for computing a partial Weierstrass–Schur form of a matrix pencil, including complete canonical information of the dominant eigenvalue, is based on the building blocks presented in sections 4.2, 5.2 and 7. As for the standard case, we give an overview of the method and discuss its reliability from the perspective of deflation errors introduced in the generalized Arnoldi and staircase algorithms. We also present and discuss results from computational experiments conducted on problems with known canonical structure and varying ill-conditioning.

9.1. Algorithm and deflation errors

We use the generalized IRA method with s explicit restarts to compute a pair of eigenspaces of interest. Let k and p be the dimensions used in the implicit restarts. In order for the method to converge successfully, the conditions on k, p and s and the space spanned by the starting vectors used in the iterative process are the same as for the matrix case ($k > h_i,\ s \geqslant n_{\mathrm{g}}(\lambda_i)$ and $k + p \ll n$).

Provided that the generalized IRA method converges, we get a factorization of the form

$$\begin{aligned} AV &= WH + E, \\ BV &= WU. \end{aligned}$$

The deflation criterion (13) ensures that the norm of the residual satisfies $\|E\|_F = \mathrm{O}(tol \max(\|A\|_F, \|B\|_F))$. The perturbation E may, like in the standard case, be considered a perturbation to the original pencil:

$$\begin{aligned} \left(A - EV^H\right)V &= WH, \\ BV &= WU. \end{aligned} \tag{15}$$

We then apply the clustering method for matrix pencils to isolate a cluster of eigenvalues containing the largest eigenvalue in magnitude. The mean over the isolated cluster λ_{cluster} is used to approximate the multiple eigenvalue. The size of the perturbations $\|E\|_1,\ \|F\|_1$ in the clustering heuristic is approximated by $tol \max(\|A\|_F, \|B\|_F)$, i.e., the deflation tolerance used in the generalized Arnoldi step.

The Weierstrass–Schur form of the dominant eigenvalue is then computed from the shifted pair $(H - \lambda_{\text{cluster}}U, U)$. In the criterion for numerical rank decisions, we use the condition $\sigma_{n-k+1} > tol \max(\|A\|_F, \|B\|_F)$. The result is then a decomposition of the form

$$P^H(H - \lambda U)Q \equiv \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} - \lambda \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} + F, \tag{16}$$

where $S_{11} = \lambda_{\text{cluster}} T_{11} + N$ with N in staircase form. If the deflation error $\|F\|_F$ is small – normally of size $\mathrm{O}(tol \max(\|A\|_F, \|B\|_F))$ – we accept λ_{cluster} as a multiple eigenvalue with the Weierstrass structure given by the block structure of the nilpotent matrix N. Observe that T_{11} is a general upper triangular matrix and the structure is, therefore, not accessible by inspecting the nonzero elements in S_{11}. This does not pose a problem since N is easily computed from S and T. Let $X = VQ$ and $Y = WP$. By inserting (16) into (15) and postmultiplying by Q^H we get the decomposition

$$\begin{aligned} (A + G)X &= YS, \\ BX &= YT, \end{aligned}$$

where $G = -EV^H - PFV^H$. The perturbation G is small if the method is successful and typically $\|G\|_F = \mathrm{O}(tol \max(\|A\|_F, \|B\|_F))$. In this case, we consider the factorization to be a partial Weierstrass–Schur form of $A - \lambda B$ with respect to the dominant eigenvalue λ_{cluster}. We remark that for a converged partial Weierstrass–Schur form, the orthogonality of the columns of X and Y is granted through the Gram–Schmidt orthogonalization processes in the generalized IRA method with explicit restarts and the unitary equivalence transformations used in the staircase algorithm. The method is summarized in algorithm 10.

9.2. Sample problems and computational experiments

We have implemented the partial Weierstrass–Schur method in Matlab. We have used the same deflation tolerance $tol = 10^{-9}$ and $gap = 1000$ as in the experiments with the Jordan–Schur method. The generalized IRA method requires a linear system to be solved in each step. The conditioning of these linear systems and the accuracy in their solutions also affect the convergence behavior of the generalized IRA method. So far, we have only used direct solvers, and even so our computational experiments show that the generalized Arnoldi method is less stable than the Arnoldi method.

For our computational experiments we have used a family of 200×200 complex regular matrix pencils. We let the dominant eigenvalue λ_{mult} have algebraic multiplicity

Algorithm 10. Partial Weierstrass–Schur form

Input: $A, B \in \mathbb{C}^{n\times n}$ such that $A - \lambda B$ is regular with B nonsingular. Two numbers k and p defining the IRA dimensions and the number s of explicit restarts. A deflation tolerance *tol* and the *gap* parameter.

Output: Two triangular matrices S and T and two rectangular matrices X and Y. The matrices describe a partial Weierstrass–Schur form and satisfy $(A + G - \lambda B)X = Y(S - \lambda T)$, where $\|G\|_F = \mathrm{O}(tol \max(\|A\|_F, \|B\|_F))$. The Weyr characteristics of the dominant eigenvalue λ_{cluster} of $A + G - \lambda B$ are listed in *str*.

function $[S, T, X, Y, \lambda_{\text{cluster}}, str] = \text{pws}(A, B, k, p, tol, gap, s)$
$[H, U, V, W] = \text{gira}(A, B, k, p, tol, s)$
$\lambda_{\text{cluster}} = \text{cluster}(H, U, \infty, tol)$
$[S_{\text{shifted}}, T, P, Q, str] = \text{staircase}(H - \lambda_{\text{cluster}}U, U, tol, gap)$
$S = S_{\text{shifted}} + \lambda_{\text{cluster}}T$
$X = VQ$
$Y = WP$

Algorithm 10. Algorithm for computing a partial Weierstrass–Schur form associated with the dominant eigenvalue of a matrix pencil.

$n_a = 7$ and the Weierstrass structure $J_4 \oplus J_2 \oplus J_1$. The remaining $200 - n_a$ eigenvalues are generated randomly smaller than λ_{mult}. Let

$$D_2 = \text{diag}(\lambda_{n_a+1}, \ldots, \lambda_{200}).$$

Moreover, we choose λ_{mult} such that the gap in the moduli of the eigenvalues,

$$\min_{n_a < i \leqslant 200} \frac{|\lambda_{\text{mult}}| - |\lambda_i|}{|\lambda_{\text{mult}}|},$$

is equal to a given parameter. The size of the gap in the moduli of the eigenvalues affects both the conditioning of the multiple eigenvalue and the convergence behavior of the Arnoldi method. We then construct a pair of upper triangular matrices (S, T),

$$T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} = \begin{pmatrix} I + M_{11} & M_{12} \\ 0 & I + M_{22} \end{pmatrix},$$

$$S = \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} = \begin{pmatrix} \lambda_{\text{mult}}T_{11} & \lambda_{\text{mult}}T_{12} \\ 0 & (D_2 - \lambda_{\text{mult}}I) + \lambda_{\text{mult}}T_{22} \end{pmatrix} + \begin{pmatrix} N_{11} & N_{12} \\ 0 & N_{22} \end{pmatrix},$$

where M and N are strictly upper triangular matrices. As for the matrix case, the block-partitioning above is only for the construction of the test matrix pencils.

The block N_{11} is chosen such that λ_{mult} gets the desired canonical structure. The elements of N and M are generated with random numbers and we scale the blocks N_{ij}

and M_{ij} to control the conditioning of the problem. The relation to the conditioning of the problem is explained by the block diagonalizing transformation [19,41],

$$\begin{pmatrix} I & -L \\ 0 & I \end{pmatrix} \begin{pmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{pmatrix} \begin{pmatrix} I & R \\ 0 & I \end{pmatrix} = \begin{pmatrix} S_{11} & S_{11}R - LS_{22} + S_{12} \\ 0 & S_{22} \end{pmatrix},$$

$$\begin{pmatrix} I & -L \\ 0 & I \end{pmatrix} \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix} \begin{pmatrix} I & R \\ 0 & I \end{pmatrix} = \begin{pmatrix} T_{11} & T_{11}R - LT_{22} + T_{12} \\ 0 & T_{22} \end{pmatrix}.$$

The left eigenspaces $(I, 0)^{\mathrm{T}}$ and $(L^{\mathrm{T}}, I)^{\mathrm{T}}$ are close to linearly dependent if $\|L\|_F$ is large. Similarly, the right eigenspaces are close to linearly dependent if $\|R\|_F$ is large. From the pair of matrix equations

$$S_{11}R - LS_{22} + S_{12} = 0,$$
$$T_{11}R - LT_{22} + T_{12} = 0,$$

it follows that $\|(L, R)\|_F$ is bounded by[2]

$$\big\|(L, R)\big\|_F \leqslant \frac{\|(S_{12}, T_{12})\|_F}{\mathrm{dif}((S_{11}, T_{11}), (S_{22}, T_{22}))}.$$

Similar to the standard case, $\|(L, R)\|_F$ can be large either if $\|(S_{12}, T_{12})\|_F$ is large or if $\mathrm{dif}((S_{11}, T_{11}), (S_{22}, T_{22}))$ is small [19]. Since

$$\mathrm{dif}\big((S_{11}, T_{11}), (S_{22}, T_{22})\big) \leqslant \min_{n_a < i \leqslant n} |\lambda_{\mathrm{mult}} - \lambda_i|,$$

we see that decreasing the gap between λ_{mult} and the remaining eigenvalues also makes the separation between (S_{11}, T_{11}) and (S_{22}, T_{22}) smaller. To get a dense pencil we generate two random unitary matrices P and Q and define the pencil $A - \lambda B = P(S - \lambda T)Q^H$.

For the numerical experiments we select four matrices and apply the partial Weierstrass–Schur method with three explicit restarts ($s = 3$). The dimensions describing the generalized implicitly restarted Arnoldi method are $k = 5$ and $p = 12$. The gap in the moduli of the eigenvalues is equal to 2. The blocks N_{ij} and M_{ij} are scaled with the same parameter for each individual pencil. For Pencil A, the blocks N_{12}, N_{22}, M_{12} and M_{22} are zero. Pencil A is the most well-conditioned pencil in the set. For pencils B–D we increase the size of the entries of M and N to make the problems increasingly ill-conditioned. The relative gap of the eigenvalues and the relative size of $\|(N_{ij}, M_{ij})\|_F$ are reported in table 5.

The results from the computations are reported in tables 6–8. In table 6 we see that the number of linear system solves and the number of implicit restarts slightly increase with the size of the elements in M and N. From table 7 it is obvious that the clustering is successful for all four matrices and that the computed mean

[2] Given two matrix pairs (A_1, A_2) and (B_1, B_2), the difference (or separation) is defined as

$$\mathrm{dif}((A_1, A_2), (B_1, B_2)) = \min_{\|(X,Y)\|_F = 1} \|(A_1X - YB_1, A_2X - YB_2)\|_F.$$

Table 5
The gap in the moduli of the eigenvalues and the relative size of the blocks of N and M for the test matrix pencils.

	$\frac{\min\lvert\lambda_{\text{mult}}\rvert - \lvert\lambda\rvert}{\lvert\lambda_{\text{mult}}\rvert}$	$\frac{\lVert(N,M)\rVert_F}{\lVert(A,B)\rVert_F}$	$\frac{\lVert(N_{11},M_{11})\rVert_F}{\lVert(A,B)\rVert_F}$	$\frac{\lVert(N_{12},M_{12})\rVert_F}{\lVert(A,B)\rVert_F}$	$\frac{\lVert(N_{22},M_{22})\rVert_F}{\lVert(A,B)\rVert_F}$
Pair A	2.00	0.06	0.06	0.00	0.00
Pair B	2.00	0.90	0.03	0.24	0.87
Pair C	2.00	0.94	0.07	0.59	0.73
Pair D	2.00	0.98	0.04	0.37	0.91

Table 6
The number of linear system solves, the condition of these systems and the number of implicit restarts for each explicit restart before the invariant subspaces converged. The relative generalized Arnoldi residual for A (the residual for B is much smaller).

	Linsolve	$\kappa_2(B)$	Implicit restarts	$\frac{\lVert AV - WH\rVert_F}{\max(\lVert A\rVert_F, \lVert B\rVert_F)}$
Pair A	123	1.2e+00	3, 3, 3	4.5e−11
Pair B	111	4.1e+00	3, 3, 2	9.3e−12
Pair C	111	1.5e+01	3, 3, 2	1.1e−12
Pair D	147	5.2e+81	5, 3, 3	3.3e−11

Table 7
The number of eigenvalues isolated in the cluster and the chordal distance from the exact multiple eigenvalue to the computed cluster mean.

	Cluster size	chord(λ_{cluster}, λ_{mult})
Pair A	7	1.6e−16
Pair B	7	2.1e−15
Pair C	7	8.7e−16
Pair D	7	3.7e−14

Table 8
The Weierstrass structure (Segre characteristics) of the dominant eigenvalue in the computed partial Weierstrass–Schur form $A-\lambda B$ and the relative residuals of S and T.

	Computed structure	$\frac{\lVert AX - YS\rVert_F}{\max(\lVert A\rVert_F, \lVert B\rVert_F)}$	$\frac{\lVert AX - YT\rVert_F}{\max(\lVert A\rVert_F, \lVert B\rVert_F)}$
Pair A	4, 2, 1	4.5e−11	8.4e−16
Pair B	4, 2, 1	9.3e−12	2.7e−16
Pair C	4, 2, 1	1.1e−12	2.2e−16
Pair D	4, 2, 1	3.3e−11	1.2e−16

λ_{cluster} is a good approximation to the multiple eigenvalue λ_{mult}. The second column of table 8 shows that the computed Weierstrass structure is correct for all of the pencils. Moreover, the residuals for the computed partial Weierstrass–Schur form are of the same size as the generalized IRA residuals in table 6. If we continue to increase the size of the elements in M and N, the pencils become very ill-conditioned and eventually the method starts to break down. Either the method breaks down by not converging at all or it converges to some strange answer with a large residual and the wrong structure.

10. Concluding remarks

10.1. Present approach and methods

The combination of the implicitly restarted Arnoldi method with explicit restarts, a new Gerschgorin clustering heuristic and the staircase algorithm provide a reliable method for computing a partial Jordan–Schur form associated with the dominant eigenvalue of a large scale matrix A. Indeed, we compute an exact partial Jordan–Schur form of a regularized nearby problem $A + G$. The size of $\|G\|_F$ is a function of the deflation tolerance parameters used. A small value on $\|G\|_F$ indicates that A has a dominant eigenvalue with the computed Jordan structure.

A similar generalized method for regular matrix pencils is presented and it is demonstrated to work reliably as well, provided the matrix B in $A - \lambda B$ is well-conditioned with respect to inversion. This requirement is necessary for the generalized IRA method to be constructible.

10.2. Applicability and reliability

Along with the description of the major building blocks of the partial Jordan–Schur and Weierstrass–Schur methods we have discussed what assumptions we make in order to get convergence and warrant meaningful and reliable results. These include the conditioning of the problems as well as the setting of dimension parameters (k, p and s) and deflation tolerances (*tol* and *gap*). We refer to earlier sections for how to choose values for these parameters.

The IRA methods require that there is a well-defined gap between the dominant eigenvalue and the remaining spectrum. Otherwise, the convergence will be too slow or the method breaks down. For the generalized IRA it is at least in theory possible to transform the problem so that the B-part becomes well-conditioned, but in general this requires some knowledge about the spectrum of the problem.

The clustering methods based on Gerschgorin circles are heuristics and therefore by definition they are not able to solve all types of problems. Without any knowledge about the true spectrum, the clustering problem in finite precision is in principle unsolvable. For example, there may be unrelated roots that lie within or near a true cluster. This situation may arise in dynamical systems.

Given an approximation of a multiple eigenvalue, the most stable part of partial Jordan–Schur and Weierstrass–Schur methods is the computation of the canonical structure associated with this eigenvalue. At first this may seem somewhat surprising, but all deflations are done using unitary similarity or equivalence transformations. So we will always produce a "nearby" matrix or matrix pencil which has a true dominant eigenvalue with the computed canonical structure. However, there may exist other closer nearby problems with different partial Jordan–Schur and Weierstrass–Schur forms [10,11]. It is the size of the final residuals ($\|G\|_F = \|AX - SX\|_F$ and $\|X^H X - I\|_F$ in the matrix case) with respect to the deflation tolerances that determine how well the methods have succeeded.

The results from computational experiments conducted on problems with known canonical structure and varying ill-conditioning show that the methods deliver partial Jordan–Schur and Weierstrass–Schur forms of nearby problems with the correct canonical structure of the dominant eigenvalue. The distance from a sample problem to the corresponding regularized problem is proportional to the size of the deflation tolerance used.

Due to the ill-posedness of the problem under consideration it is of course always possible to construct problems so that any of the three major building blocks produce erroneous result. Nonetheless, our study shows that for large scale problems with a well-defined staircase form it is possible to construct methods for extracting partial canonical structure.

The main challenge for future work lies in improving the first step of the methods, i.e., the computation of an invariant subspace or a pair of eigenspaces. If we could achieve better accuracy and more stable behavior in these steps for ill-conditioned problems and large scale problems, then the overall reliability would benefit since both the clustering heuristic and the staircase method can be used successfully with smaller tolerances.

Acknowledgements

The authors are grateful to Professor Dan Sorensen who visited the department during a week in the summer of 1997. He provided several valuable comments on Krylov methods and a manuscript (now published [38]) concerning the generalized implicitly restarted Arnoldi method which was helpful in our work. Special thanks to Pedher Johansson for his style file for the layout of algorithms and for his assistance in making our LaTeX files compatible with the class file of Baltzer Science.

Financial support has been received by the Swedish Research Council for Engineering Sciences under grant 222-97-112.

References

[1] Z. Bai and G.W. Stewart, Algorithm 776: SRRIT: A Fortran Subroutine to calculate the dominant invariant subspace of a nonsymmetric matrix, ACM Trans. Math. Software 23(4) (1997) 494–523.

[2] T. Beelen and P. Van Dooren, An improved algorithm for the computation of Kronecker's canonical form of a singular pencil, Linear Algebra Appl. 105 (1988) 9–65.
[3] M. Bennani and T. Braconnier, Stopping criteria for eigensolvers, Technical Report TR/PA/94/22, CERFACS (1994).
[4] T. Braconnier, V. Frayssé and J.-C. Rioual, ARNCHEB users' guide: Solution of large non symmetric or non Hermitian eigenvalue problems by the Arnoldi–Tchebycheff method, Technical Report TR/PA/97/50, CERFACS (1997).
[5] F. Chatelin, *Eigenvalues of Matrices* (Wiley, New York, 1993).
[6] J. Demmel and B. Kågström, The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part I: Theory and algorithms, ACM Trans. Math. Software 19(2) (1993) 160–174.
[7] J. Demmel and B. Kågström, The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: Robust software with error bounds and applications. Part II: Software and applications, ACM Trans. Math. Software 19(2) (1993) 175–201.
[8] I.S. Duff and J.A. Scott, Computing selected eigenvalues of sparse unsymmetric matrices using subspace iteration, ACM Trans. Math. Software 19(2) (1993) 137–159.
[9] A. Edelman, E. Elmroth and B. Kågström, A geometric approach to perturbation theory of matrices and matrix pencils. Part I: Versal deformations, SIAM J. Matrix Anal. Appl. 18(3) (1997) 653–692.
[10] A. Edelman, E. Elmroth and B. Kågström, A geometric approach to perturbation theory of matrices and matrix pencils. Part II: A stratification-enhanced staircase algorithm, SIAM J. Matrix Anal. Appl. 20(3) (1999) 667–699.
[11] E. Elmroth and B. Kågström, The set of 2-by-3 matrix pencils – Kronecker structures and their transitions under perturbations, SIAM J. Matrix Anal. Appl. 17(1) (1996) 1–34.
[12] G.L. Fokkema, G.L.G. Sleijpen and H.A. Van der Vorst, A Jacobi–Davidson style QR and QZ algorithm for partial reduction of matrix pencils, Technical Report 941, Department of Mathematics, University of Utrecht (1996).
[13] F.R. Gantmacher, *The Theory of Matrices*, Vol. II (Chelsea, New York, 1959).
[14] G. Golub and J.H. Wilkinson, Ill-conditioned eigensystems and the computation of the Jordan canonical form, SIAM Review 18(4) (1976) 578–619.
[15] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 2nd ed. (Johns Hopkins Univ. Press, Baltimore, MD, 1989).
[16] B. Hartley and T.O. Hawkes, *Rings, Modules and Linear Algebra* (Chapman & Hall, London, 1970).
[17] B. Kågström, How to compute the Jordan normal form – the choice between similarity transformations and methods using the chain relations, Technical Report UMINF-91.81, Department of Numerical Analysis, Institute of Information Processing, University of Umeå (1981).
[18] B. Kågström, RGSVD – an algorithm for computing the Kronecker canonical form and reducing subspaces of singular matrix pencils $A - \lambda B$, SIAM J. Sci. Statist. Comput. 7(1) (1986) 185–211.
[19] B. Kågström, A perturbation analysis of the generalized Sylvester equation $(AR-LB, DR-LE) = (C, F)$, SIAM J. Matrix Anal. Appl. 15(4) (1994) 1045–1060.
[20] B. Kågström and P. Poromaa, Computing eigenspaces with specified eigenvalues of a regular matrix pair (A, B) and condition estimation: Theory, algorithms and software, Numer. Algorithms 12 (1996) 369–407.
[21] B. Kågström and A. Ruhe, An algorithm for the numerical computation of the Jordan normal form of a complex matrix, ACM Trans. Math. Software 6(3) (1980) 389–419.
[22] B. Kågström and A. Ruhe, ALGORITHM 560: An algorithm for the numerical computation of the Jordan normal form of a complex matrix [F2], ACM Trans. Math. Software 6(3) (1980) 437–443.
[23] W. Kahan, B.N. Parlett and E. Jiang, Residual bounds on approximate eigensystems of nonnormal matrices, SIAM J. Num. Anal. 19(3) (1982) 470–484.
[24] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer, Berlin, 1980).
[25] V.N. Kublanovskaya, On a method of solving the complete eigenvalue problem for a degenerate matrix (in Russian), Zh. Vychisl. Mat. i Mat. Fiz. 6 (1966) 611–620. (USSR Comput. Math. Math. Phys. 6(4) (1968) 1–16.)

[26] V. Kublanovskaya, AB-algorithm and its modifications for the spectral problem of linear pencils of matrices, Numer. Math. 43 (1984) 329–342.
[27] R.B. Lehoucq and D.C. Sorensen, Deflation techniques for an implicitly restarted Arnoldi iteration, SIAM J. Matrix Anal. Appl. 17(2) (1996) 789–821.
[28] R.B. Lehoucq, D.C. Sorensen and C. Yang, *ARPACK Users' Guide, Solution of Large-Scale Eigenvalue Problems with Implicitly Restarted Arnoldi Methods* (SIAM, Philadelphia, PA, 1998).
[29] K. Meerbergen and D. Roose, The restarted Arnoldi method applied to iterative linear system solvers for the computation of rightmost eigenvalus, SIAM J. Matrix Anal. Appl. 18(1) (1997) 1–20.
[30] A. Ruhe, An algorithm for numerical determination of the structure of a general matrix, BIT 10 (1970) 196–216.
[31] A. Ruhe, Perturbation bound for means of eigenvalues and invariant subspaces, BIT (1970).
[32] A. Ruhe, Rational Krylov algorithms for nonsymmetric eigenvalue problems. II. Matrix pairs, Linear Algebra Appl. 197/198 (1994) 283–295.
[33] Y. Saad, *Numerical Methods for Large Eigenvalue Problems* (Halsted Press, New York, 1992).
[34] J.A. Scott, An Arnoldi code for computing selected eigenvalues of sparse, real, unsymmetric matrices, ACM Trans. Math. Software 21(4) (1995) 432–475.
[35] G.E. Shilov, *Linear Algebra* (Dover, New York, 1977).
[36] G.L.G. Sleijpen and H.A. van der Vorst, A Jacobi–Davidson iteration method for linear eigenvalue problems, SIAM J. Matrix Anal. Appl. 17(2) (1996) 401–425.
[37] D.C. Sorensen, Implicit application of polynomial filters in a k-step Arnoldi method, SIAM J. Matrix Anal. Appl. 13(2) (1992) 357–385.
[38] D.C. Sorensen, Truncated QZ methods for large scale generalized eigenvalue problems, Electron. Trans. Numer. Anal. 7 (1998) 141–168.
[39] D.C. Sorensen and C. Yang, A truncated RQ-iteration for large scale eigenvalue calculations, SIAM J. Matrix Anal. Appl. 19(4) (1998) 1045–1073.
[40] G.W. Stewart, *Matrix Algorithms: Basic Decompositions* (SIAM, Philadelphia, PA, 1998).
[41] G.W. Stewart and J.-G. Sun, *Matrix Perturbation Theory* (Academic Press, London, 1990).
[42] P. Van Dooren, The computation of Kronecker's canonical form of a singular pencil, Linear Algebra Appl. 27 (1979) 103–141.