Numer. Math. (2010) 114:549–571
DOI 10.1007/s00211-009-0266-y

Numerische Mathematik

Two-sided Grassmann–Rayleigh quotient iteration

P.-A. Absil · P. Van Dooren

Received: 19 April 2007 / Revised: 29 July 2009 / Published online: 10 October 2009

Abstract The two-sided Rayleigh quotient iteration proposed by Ostrowski computes a pair of corresponding left–right eigenvectors of a matrix C. We propose a Grassmannian version of this iteration, i.e., its iterates are pairs of p-dimensional subspaces instead of one-dimensional subspaces in the classical case. The new iteration generically converges locally cubically to the pairs of left–right p-dimensional invariant subspaces of C. Moreover, Grassmannian versions of the Rayleigh quotient iteration are given for the generalized Hermitian eigenproblem, the Hamiltonian eigenproblem and the skew-Hamiltonian eigenproblem.

Mathematics Subject Classification (2000) 65F15

1 Introduction

The Rayleigh quotient iteration (RQI) is a classical single-vector iteration for computing eigenvectors of a Hermitian matrix $A = A^H$; see, e.g., [19]. It generates a

This paper presents research results of the Belgian Network Dynamical systems, control, and optimization (DYSCO), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. This work was supported in part by the US National Science Foundation under Grant OCI-0324944 and by the School of Computational Science of Florida State University through a postdoctoral fellowship.

P.-A. Absil (✉) · P. Van Dooren
Département d'ingénierie mathématique, Université catholique de Louvain,
1348 Louvain-la-Neuve, Belgium
e-mail: absil@inma.ucl.ac.be
URL: http://www.inma.ucl.ac.be/∼absil

P. Van Dooren
URL: http://www.inma.ucl.ac.be/∼vdooren

sequence of vectors $\{x_k\}_{k=0,1,\ldots}$ from a given starting vector x_0 by solving the linear systems

$$(A - \rho(x_k)I)x_{k+1} = s_k x_k, \quad k \geq 0, \tag{1}$$

where $\rho(x_k) = x_k^H A x_k / (x_k^H x_k)$ is the Rayleigh quotient of A at x_k, I denotes the identity matrix, s_k is a scaling factor, and the superscript ^{H} stands for the conjugate transpose. The RQI can be viewed as combination of Lord Rayleigh's technique for improving an approximate eigenvector x by solving $(A - \rho(x)I)y = e_1$ (i.e., 1 with fixed right-hand side), and Wielandt's inverse iteration with fixed shift; see [31,40], and [21] and references therein for a historical perspective on inverse iteration. The RQI per se is attributed to Crandall [12], and major contributions to its analysis were made by Ostrowski [24], Parlett and Kahan [26–28], and Batterson and Smillie [9,10]. A key element behind the popularity of the Rayleigh shift in inverse iteration is that, around an eigenvector, it produces a quadratic approximation of the corresponding eigenvalue; see [27, (4.20)]. This property endows the RQI with a cubic rate of convergence to the eigenvectors of A; see [26,27] or the sketch of proof in [4]. Another reason to find the RQI of particular interest is that it lives in hiding within the Rayleigh-shifted QR algorithm [39].

The purpose of this paper is to propose and analyze a method that combines two existing generalizations of the RQI: a two-sided version, and a block version.

The underlying idea for the two-sided version is that when the matrix A is nonnormal—we will call it C to follow Parlett's convention of letting letters that are symmetric about a vertical axis stand for Hermitian matrices—, the quadratic approximation property of the Rayleigh quotient is lost. This drawback was avoided by Ostrowski [25] (or see [26]) by considering the generalized Rayleigh quotient $\rho(y_L, y_R) := y_L^H C y_R / y_L^H y_R$ which displays the quadratic property in the neighborhood of the pairs of left–right eigenvectors that belong to a nondefective eigenvalue of C. Using this Rayleigh quotient as a shift, Ostrowski derived a two-sided iteration that operates on pairs of vectors and aims at converging to pairs of left–right eigenvectors of C. The rate of convergence is cubic in nondegenerate cases. The two-sided RQI generalizes the RQI in the sense that, if C is Hermitian and the initial left and right vectors are chosen identical, then the two-sided RQI reduces to two copies of the RQI.

The other existing generalization of the RQI is the block version proposed by Smit [34] and rediscovered in [4] under the name *Grassmann–Rayleigh quotient iteration* (GRQI). The choice of the name Grassmann was prompted by the fact that the block RQI induces an iteration on the set of p dimensional subspaces of $\mathbb{R}^n$, where p is the block size and n is the order of A; the set of all p-dimensional subspaces of $\mathbb{R}^n$ is a *Grassmann manifold*, whose differentiable structure underpins the convergence analysis proposed in [4]. The GRQI converges locally cubically to the p-dimensional eigenspaces of the Hermitian matrix A. The method is similar, but not identical, to several other block methods for eigenspace computation chiefly stemming from a Newton approach [11,13,14,23,36], as well as to Newton-like iterations for eigenspace refinement derived from a differential-geometric approach [2,15,22] (or see [3] for an overview).

Thus, in the present paper, we merge Ostrowski's two-sided RQI with the block (or Grassmann) RQI to obtain an two-sided iteration that operates on pairs of p-dimensional subspaces. While the definition of the new iteration, which we dub *Two-Sided Grassmann-RQI* (*2sGRQI*), is straightforward, its convergence analysis is more challenging. We show that the iteration converges locally cubically to the pairs of left–right p-dimensional eigenspaces of C. On the way to the result, we prove several lemmas that may be of interest in their own right.

The choice of the practical implementation of 2sGRQI is also a nontrivial issue, because of the need to solve Sylvester equations that become increasingly ill-conditioned as the iteration progresses, in such a way that the column space of the solution is accurate. It should be emphasized that solving increasingly ill-conditioned systems is *inherent* to RQI-like methods: it is indeed this ill-conditioning that confers to the method its superlinear convergence. In the classical RQI, it has been shown in [29] that, barring particular situations with clustered eigenvalues, the column space (namely, here, the direction) of the solution is only mildly affected by numerical errors. In the classical two-sided RQI and in the GRQI, numerical experience points to similar conclusions. In the case of 2sGRQI, we report on numerical experiments showing that certain ways of solving the Sylvester equations are preferrable to others. We found that, when the Sylvester equations are solved by diagonalizing the small matrix, the accuracy is similar with the one obtained with the previously-known versions of RQI.

We also show that in some structured eigenproblems, namely E-(skew-)Hermitian matrices with $E = \pm E^H$, a relation $\mathcal{Y}_L = E\mathcal{Y}_R$ between left and right subspaces is invariant by the 2sGRQI mapping (Sect. 5). In particular, this observation yields a modified one-sided Grassmann-RQI for the Hamiltonian eigenproblem.

2 Preliminaries

This paper uses a few elementary concepts related to the algebraic eigenvalue problem, such as principal vectors, Jordan blocks and nonlinear elementary divisors. A classical reference is [41].

The superscript ^{H} denotes the conjugate transpose. In accordance with Parlett's conventions [26,27], we try to reserve the letter A for Hermitian matrices while C may denote any matrix. We use $\mathrm{Grass}(p, n)$ to denote the Grassmann manifold of the p-dimensional subspaces of $\mathbb{C}^n$, $\mathbb{P}^{n-1}$ to denote the projective space (i.e., the set of all one-dimensional subspaces of $\mathbb{C}^n$), and $\mathbb{C}_*^{n\times p}$ to denote the noncompact Stiefel manifold, i.e., the set of n-by-p matrices with full rank. The space spanned by the columns of $Y \in \mathbb{C}_*^{n\times p}$ is denoted by $\lfloor Y \rfloor$ and called the *span* of Y. The norm of a vector x is $\|x\| = \sqrt{x^H x}$. The spectral norm of a matrix F, denoted by $\|F\|$, is the largest singular value of F. The Hermitian angle $\angle(x, y)$ in $[0, \frac{\pi}{2}]$ between two vectors x and y in $\mathbb{C}^n$ is given by $\cos\angle(x, y) = \frac{|x^H y|}{\|x\|\|y\|}$; see, e.g., [32]. The angle between a vector $y \in \mathbb{C}^n$ and a subspace $\mathcal{X}$ spanned by $X \in \mathbb{C}_*^{n\times p}$ is $\angle(X, y) = \min_{x\in\mathcal{X}} \angle(x, y)$. The angle $\angle(X, Y)$ between two subspaces spanned by $X \in \mathbb{C}_*^{n\times p}$ and $Y \in \mathbb{C}_*^{n\times p}$ is defined as the largest principal angle between the two subspaces, given by $\cos\angle(X, Y) = \sigma_{\min}(\tilde{X}^H \tilde{Y})$, where $\tilde{X}$ and $\tilde{Y}$ are orthonormal bases for $\lfloor X \rfloor$ and $\lfloor Y \rfloor$ respectively, and $\sigma_{\min}$ denotes

the smallest singular value; see, e.g., [38, Sect. 4.2.1]. The angle between subspaces defines a metric (i.e., a distance function) on the Grassmann manifold [30].

We now briefly recall some basic facts about invariant subspaces.

Definition 2.1 [eigenspaces] Let $\mathcal{X}$ be a p-dimensional subspace of $\mathbb{C}^n$ and let $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ be a unitary $n \times n$ matrix such that X_1 spans $\mathcal{X}$. Partition $X^H C X$ in the form $X^H C X = \left[\begin{smallmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{smallmatrix}\right]$ where $C_{11} \in \mathbb{C}^{p\times p}$. The subspace $\mathcal{X}$ is an *eigenspace* (or *invariant subspace*) of C if $C_{21} = 0$, i.e., $C\mathcal{X} \subset \mathcal{X}$. By *spectrum* of $\mathcal{X}$, we mean the set of eigenvalues of C_{11}. The invariant subspace $\mathcal{X}$ is termed *simple* [35,38], or *spectral* [17], if C_{11} and C_{22} have no eigenvalue in common. We say that $\mathcal{X}$ is a *nondegenerate* invariant subspace of C if all the eigenvalues of C_{11} are simple, i.e., distinct. The eigenspaces of C^H are called *left eigenspaces* of C. We say that $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a *pair of simple left–right eigenspaces* of C if $\mathcal{Y}_L$ and $\mathcal{Y}_R$ are simple left and right eigenspaces of C with the same spectrum.

The span of $Y \in \mathbb{C}_*^{n\times p}$ is an eigenspace of C if and only if there exists a matrix M such that $CY = YM$. Every simple eigenspace $\mathcal{V}$ is *isolated*, i.e., there exists a neighborhood of $\mathcal{V}$ in $\mathrm{Grass}(p, n)$ that does not contain any eigenspace of C other than $\mathcal{V}$. We will also need the following result [20, Sect. 7.6.3], which is a reformulation of [35, Theorem V.1.5].

Proposition 2.2 *If $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of simple left–right eigenspaces of C, then there exists an invertible matrix S such that the first p columns of S span $\mathcal{Y}_R$, the first p columns of S^{-H} span $\mathcal{Y}_L$, and $S^{-1}CS = \left[\begin{smallmatrix} D_1 & 0 \\ 0 & D_2 \end{smallmatrix}\right]$ with $D_1 \in \mathbb{C}^{p\times p}$.*

The Rayleigh quotient iteration is a classical method for computing a single eigenvector of a Hermitian matrix A. It induces an iteration on the projective space $\mathbb{P}^{n-1}$ that can be written as follows.

Algorithm 2.3 [RQI on projective space] *Let $A = A^H$ be an $n \times n$ matrix. Given $\mathcal{S}_0$ in the projective space $\mathbb{P}^{n-1}$, the RQI algorithm produces a sequence of elements of $\mathbb{P}^{n-1}$ as follows. For $k = 0, 1, 2, \ldots$,*

1. *Pick y in $\mathbb{C}^n \setminus \{0\}$ such that $\lfloor y \rfloor = \mathcal{S}_k$.*
2. *Compute the Rayleigh quotient $\rho_k = (y^H Ay)/(y^H y)$.*
3. *If $A - \rho_k I$ is singular, then solve for its kernel and stop. Otherwise, solve the system*

$$(A - \rho_k I)z = y \tag{2}$$

   for z.
4. $\mathcal{S}_{k+1} := \lfloor z \rfloor$.

It is shown in [9] that around each (isolated) eigenvector of A, there is a ball in which cubic convergence to the eigenvector is uniform. The size of the ball depends on the spacing between the eigenvalues. Globally, the RQI converges to an eigenvector for all initial points outside a certain set of measure zero described in [9].

The *Grassmann–Rayleigh Quotient Iteration* (GRQI) is a generalization of the RQI that operates on $\mathrm{Grass}(p, n)$, the set of all p-dimensional subspaces of $\mathbb{C}^n$ [4].

Algorithm 2.4 [GRQI] *Let $A = A^H$ be an $n \times n$ matrix. Given $\mathcal{Y}_0 \in \mathrm{Grass}(p, n)$, the GRQI algorithm produces a sequence of p-dimensional subspaces of $\mathbb{C}^n$ by iterating from $\mathcal{Y}_0$ the mapping $\mathrm{Grass}(p, n) \to \mathrm{Grass}(p, n) : \mathcal{Y} \mapsto \mathcal{Y}_+$ defined as follows.*

1. *Pick $Y \in \mathbb{C}_*^{n\times p}$ such that $\lfloor Y \rfloor = \mathcal{Y}$.*
2. *Solve the Sylvester equation*

$$AZ - Z(Y^H Y)^{-1} Y^H AY = Y \tag{3}$$

   for $Z \in \mathbb{C}^{n\times p}$.
3. *Define $\mathcal{Y}_+ := \lfloor Z \rfloor$.*

It is shown in [4] that the subspace $\mathcal{Y}_+$ does not depend on the choice of basis Y for $\mathcal{Y}$ in the first step. This iteration converges cubically to the p-dimensional eigenspaces of A, which are the only fixed points.

When the matrix A is not normal, the Rayleigh quotient no longer produces a quadratic approximation of the eigenvalues. Consequently, the convergence rate of the RQI can be at best quadratic. In order to recover cubic convergence, Ostrowski [25] proposed a two-sided version of the RQI, formulated as follows in [26].

Algorithm 2.5 [Two-Sided RQI] *Let C be an $n \times n$ matrix. Pick initial vectors v_0 and u_0 satisfying $v_0^H u_0 \neq 0$, $\|v_0\| = \|u_0\| = 1$. For $k = 0, 1, 2, \ldots$,*

1. *Compute $\rho_k = v_k^H C u_k / v_k^H u_k$.*
2. *If $C - \rho_k I$ is singular solve $y^H (C - \rho_k I) = 0$ and $(C - \rho_k I)x = 0$ for $y, x \neq 0$ and stop, otherwise*
3. *Solve both $v_{k+1}^H (C - \rho_k I) = v_k^H \nu_k$, $(C - \rho_k I) u_{k+1} = u_k \tau_k$, where ν_k and τ_k are normalizing factors.*
4. *If $v_{k+1}^H u_{k+1} = 0$, then stop and admit failure.*

The Two-Sided RQI converges with cubic rate to the pairs of left–right eigenvectors of C with linear elementary divisor [26].

3 Two-sided GRQI

We propose the following generalization of the two-sided RQI, which we call the *Two-Sided Grassmann–Rayleigh Quotient Iteration* (*2sGRQI*).

Algorithm 3.1 [2sGRQI] *Let C be an $n \times n$ matrix. Given $(\mathcal{Y}_{L_0}, \mathcal{Y}_{R_0}) \in \mathrm{Grass}(p, n) \times \mathrm{Grass}(p, n)$, the 2sGRQI algorithm produces a sequence of pairs of p-dimensional subspaces of $\mathbb{C}^n$ by iterating from $(\mathcal{Y}_{L_0}, \mathcal{Y}_{R_0})$ the mapping $(\mathcal{Y}_L, \mathcal{Y}_R) \mapsto (\mathcal{Y}_{L+}, \mathcal{Y}_{R+})$ defined as follows.*

1. *Pick Y_L and Y_R in $\mathbb{C}_*^{n\times p}$ such that $\lfloor Y_L \rfloor = \mathcal{Y}_L$ and $\lfloor Y_R \rfloor = \mathcal{Y}_R$.*
2. *Solve the Sylvester equations*

$$CZ_R - Z_R \underbrace{\left(Y_L^H Y_R\right)^{-1} Y_L^H C Y_R}_{R_R} = Y_R \tag{4a}$$

and

$$Z_L^H C - \underbrace{Y_L^H C Y_R \left(Y_L^H Y_R\right)^{-1}}_{R_L} Z_L^H = Y_L^H \tag{4b}$$

for Z_L and Z_R in $\mathbb{C}^{n\times p}$.

3. *Define $\mathcal{Y}_{L+} := \lfloor Z_L \rfloor$ and $\mathcal{Y}_{R+} := \lfloor Z_R \rfloor$.*

In point 1, one has to choose bases for $\mathcal{Y}_L$ and $\mathcal{Y}_R$. There are infinitely many possibilities. Indeed, if Y is a basis of $\mathcal{Y}$, then $\{YM : M \in \mathbb{C}_*^{p\times p}\}$ is the (infinite) set of all bases of $\mathcal{Y}$. Therefore, one has to make sure that $\mathcal{Y}_{L+}$ and $\mathcal{Y}_{R+}$ do not depend on the choice of basis. By a straightforward adaptation of the development carried out in [4] for the GRQI algorithm, if (Y_L, Y_R, Z_L, Z_R) solve (4) then $(Y_L M, Y_R N, Z_L M, Z_R N)$ also solve (4) for all M, N in $\mathbb{C}_*^{p\times p}$. Hence, the spans of Z_L and Z_R only depend on $\mathcal{Y}_L$ and $\mathcal{Y}_R$, and not on the choice of the bases Y_L and Y_R.

In point 2, the matrix $Y_L^H Y_R$ may not be invertible. This corresponds to point 4 in the Two-Sided RQI (Algorithm 2.5). However, if $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of simple left–right eigenspaces of C, then $Y_L^H Y_R$ is invertible as a consequence of Proposition 2.2 and, by continuity, invertibility holds on a neighborhood of the pair of eigenspaces.

In point 2, Eqs. 4 are two uncoupled Sylvester equations. Numerical methods for solving these equations are discussed in Sect. 6.

The Sylvester equations (4) may fail to admit one and only one solution. This situation happens if and only if (Y_R, Y_L) belongs to the set

$$\begin{aligned}\mathcal{S} &:= \left\{(Y_L, Y_R) \in \mathbb{C}_*^{n\times p} \times \mathbb{C}_*^{n\times p} : R_R \text{ exists and } \operatorname{spec}(C) \cap \operatorname{spec}(R_R) \neq \emptyset\right\} \\ &= \bigcup_{\lambda\in\operatorname{spec}(C)} \{(Y_L, Y_R) \in \mathbb{C}_*^{n\times p} \times \mathbb{C}_*^{n\times p} : R_R \text{ exists and } \det(R_R - \lambda I) = 0\},\end{aligned}$$

where $\operatorname{spec}(C)$ denotes the spectrum of C; this follows directly from the characterization of the eigenvalues of Sylvester operators [36, Theorem 4.4]. Since $\mathcal{S}$ is the finite union of algebraic sets, it has measure zero and the interior of its closure is empty. This means that if $(\hat{Y}_L, \hat{Y}_R)$ does not yield a unique solution, then there exists, arbitrarily close to $(\hat{Y}_L, \hat{Y}_R)$, a pair (Y_L, Y_R) and a neighborhood of this pair on which the solution (Z_L, Z_R) of (4) exists and is unique. Hence, when such a singularity occurs (in Matlab, when the computed solution of the Sylvester equations contains `Inf`'s or `NaN`'s), a simple remedy is to slightly perturb the system. We have used this remedy in our numerical experiments described in Sect. 6.

In point 3, if Z_L or Z_R is not full rank, then $(\mathcal{Y}_{L+}, \mathcal{Y}_{R+})$ does not belong to $\operatorname{Grass}(p, n) \times \operatorname{Grass}(p, n)$. A tall $n \times p$ matrix Z is rank deficient if and only if all its $p \times p$ minors are zero. Therefore, the set

$$\mathcal{D} := \{(Y_L, Y_R) : \operatorname{rank}(Z_L) < p \text{ or } \operatorname{rank}(Z_R) < p\}$$

is a subset of a finite union of algebraic sets. So here again, Z_L and Z_R are full rank for a generic choice of Y_L, Y_R.

In practice, only a few iterates will be computed. In finite precision arithmetic, the iterates no longer improve after a few (typically two or three) iterations because of numerical errors; see numerical experiments in Sect. 6. Stopping criteria can rely on the principal angles between two successive iterates and on the principal angles between $\mathcal{Y}_R$ and $A\mathcal{Y}_R$ or $\mathcal{Y}_L$ and $A^H\mathcal{Y}_L$.

4 Local convergence

This section is dedicated to showing local cubic convergence of 2sGRQI. The analysis can be thought of as a two-sided generalization of the proof of cubic convergence of the block RQI (equivalent to the Grassmann-RQI of [4]) given in [34]. The development is rather long, but the rationale is quite simple. It is presented in the next two paragraphs.

Let $(\mathcal{V}_L, \mathcal{V}_R)$ be a pair of simple left–right eigenspaces of C, and let V_L and V_R be corresponding eigenbases. We assume that the eigenspaces are nondegenerate, that is, all the eigenvalues of the matrix $(V_L^H V_R)^{-1} V_L^H C V_R$ are simple. The set of matrices with all simple eigenvalues is open (this follows, e.g., from [35, Theorem IV.1.1]), therefore, for all $\mathcal{Y}_L$ and $\mathcal{Y}_R$ sufficiently close to $\mathcal{V}_L$ and $\mathcal{V}_R$, the block Rayleigh quotients R_R and R_L have all simple eigenvalues, and are thus diagonalizable by similarity transformations W_R and W_L. Equations 4 can then be solved in a two-step procedure, which relates to the Bartels–Stewart approach [8]: (i) Diagonalize the small block Rayleigh quotients, hence decoupling the 2sGRQI equations (4) and reducing them to classical two-sided RQI equations:

$$C\tilde{Z}_R - \tilde{Z}_R \mathrm{diag}(\rho_1, \ldots, \rho_p) = \tilde{Y}_R \tag{5a}$$

$$C^H\tilde{Z}_L - \tilde{Z}_L \mathrm{diag}(\bar{\rho}_1, \ldots, \bar{\rho}_p) = \tilde{Y}_L, \tag{5b}$$

where $R_R = W_R \,\mathrm{diag}(\rho_1, \ldots, \rho_p)\, W_R^{-1}$, $R_L = W_L \,\mathrm{diag}(\rho_1, \ldots, \rho_p)\, W_L^{-1}$, $\tilde{Y}_R = Y_R W_R$, $\tilde{Y}_L = Y_L W_L^{-H}$, $\tilde{Z}_R = Z_R W_R$, and $\tilde{Z}_L = Z_L W_L^{-H}$. Observe that R_R and R_L have the same spectrum since they are related by a similarity transformation, and W_L can be chosen as $(Y_L^H Y_R) W_R$. (ii) Solve the decoupled two-sided RQI equations, yielding matrices $\tilde{Z}_L$ and $\tilde{Z}_R$ that span $\mathcal{Y}_{L+}$ and $\mathcal{Y}_{R+}$.

In view of this two-step procedure for solving the 2sGRQI equations (4), the local convergence analysis of 2sGRQI can be carried out in three steps: (a) Show that the column-wise angles between $\tilde{Y}_R$ and V_R is $O(\epsilon)$, where $\epsilon = \angle(Y_L, V_L) + \angle(Y_R, V_R)$, and likewise for the left-side objects. (b) Invoke the cubic local convergence of the classical two-sided RQI to show that the column-wise angles between $\tilde{Z}_R$ and V_R is $O(\epsilon^3)$, and likewise for the left-side objects. (c) Show that $\angle(\tilde{Z}_R, V_R)$ is $O(\epsilon^3)$, and likewise for $\angle(\tilde{Z}_L, V_L)$.

We find that obtaining an explicit bound, rather than simply an $O(\epsilon^3)$, is informative. To reach this goal, we need a succession of technical lemmas that produce explicit bounds for the three steps.

As in [35], we let

$$\operatorname{sep}(L_1, L_2) = \inf_{\|P\|=1} \|PL_1 - L_2P\|.$$

Observe that sep is unitarily invariant, i.e., for all unitary matrices Q_1 and Q_2,

$$\operatorname{sep}\left(Q_1^H L_1 Q_1, Q_2^H L_2 Q_2\right) = \operatorname{sep}(L_1, L_2).$$

For any diagonalizable matrix L of order n, we let

$$\kappa(L) = \inf\{\operatorname{cond}(S) : S \in \mathbb{C}^{n\times n}, S^{-1}LS \text{ diagonal}\},$$

where $\operatorname{cond}(S)$ denotes the condition number of S. We refer to [35, Sect. IV.3.2] for more information on the concept of condition number of matrices of eigenvectors. Finally, we let

$$B_{X_L,X_R} := \left(X_L^H X_R\right)^{-1} X_L^H C X_R$$

denote the block two-sided Rayleigh quotient; the matrix C is always clear from the context, and the subscripts will sometimes be omitted, too.

The first lemma below is a preparation for the lemma that addresses step (a).

Lemma 4.1 *Let V_L, V_R be orthonormal bases of a pair of left–right simple nondegenerate eigenspaces of an $n \times n$ matrix C, and let $\lambda_1, \ldots, \lambda_p$ be the eigenvalues of the block two-sided Rayleigh quotient B_{V_L,V_R}. Observe that, if $\angle(Y_L, V_L)$ and $\angle(Y_R, V_R)$ are sufficiently small, then the eigenvalues of B_{Y_L,Y_R} are sufficiently close to the eigenvalues $\lambda_1, \ldots, \lambda_p$, and, therefore, they can be assigned to them in a unique manner. Then for all $\theta > 0$, there exists $\epsilon_{\text{sep}} > 0$ such that, for all indices $1 \le i \le p$ and for all $n \times p$ orthonormal matrices Y_R and Y_L with $\angle(Y_L, V_L) < \epsilon_{\text{sep}}$ and $\angle(Y_R, V_R) < \epsilon_{\text{sep}}$, we have*

$$\begin{aligned}\min_i \operatorname{sep}&\left(w_{i,Y_L,Y_R}^H B_{Y_L,Y_R} w_{i,Y_L,Y_T}, (w_{i,Y_L,Y_R})_\perp^H B_{Y_L,Y_R}(w_{i,Y_L,Y_R})_\perp\right)\\ &\ge (1-\theta)\min_i \frac{\min_{j\neq i}|\lambda_j - \lambda_i|}{\kappa((w_{i,V_L,V_R})_\perp^H B_{V_L,V_R}(w_{i,V_L,V_R})_\perp)} > 0,\end{aligned}$$

where w_{i,Y_L,Y_R} is the eigenvector of B_{Y_L,Y_R} associated to its eigenvalue that converges to λ_i, $(w_{i,V_L,V_R})_\perp$ denotes an orthonormal basis of the orthogonal complement of w_{i,Y_L,Y_R} in $\mathbb{C}^p$.

Proof $F := (w_{i,V_L,V_R})_\perp^H B_{V_L,V_R}(w_{i,V_L,V_R})_\perp$ is diagonalizable to $D = \operatorname{diag}(\lambda_j)_{j\neq i}$ by a matrix S. For any scalar μ, we have $\operatorname{sep}(F,\mu) = \min_{\|x\|=1}\|(F-\mu I)x\|$. We also have $\|(F-\mu I)x\| = \|S(D-\mu I)S^{-1}x\| \ge \sigma_{\min}(S)\|(D-\mu I)S^{-1}x\| \ge$

$\sigma_{\min}(S)\min_{j\neq i}|\lambda_j-\mu|\|S^{-1}x\| \geq \frac{\sigma_{\min}(S)}{\sigma_{\max}(S)}\min_{j\neq i}|\lambda_j-\mu|$. For the choice $\mu = w^H_{i,V_L,V_R}B_{V_L,V_R}w_{i,V_L,V_T} = \lambda_i$, this shows that

$$\begin{aligned}&\operatorname{sep}\left(w^H_{i,V_L,V_R}B_{V_L,V_R}w_{i,V_L,V_T},(w_{i,V_L,V_R})^H_\perp B_{V_L,V_R}(w_{i,V_L,V_R})_\perp\right)\\ &\quad\geq\frac{\min_{j\neq i}|\lambda_j-\lambda_i|}{\kappa((w_{i,V_L,V_R})^H_\perp B_{V_L,V_R}(w_{i,V_L,V_R})_\perp)}.\end{aligned}$$

Observe now that $\operatorname{sep}(w^H_{i,Y_L,Y_R}B_{Y_L,Y_R}w_{i,Y_L,Y_T},(w_{i,Y_L,Y_R})^H_\perp B_{Y_L,Y_R}(w_{i,Y_L,Y_R})_\perp)$ is a continuous (and even smooth) function of (Y_L, Y_R) around (V_L, V_R). Moreover, since sep is unitarily invariant, the function only depends on (Y_L, Y_R) through $(\mathcal{Y}_L, \mathcal{Y}_R)$, thus it projects to a function on $\operatorname{Grass}(p,n)\times\operatorname{Grass}(p,n)$ (see, e.g., [3, Sect. 3.4.2]). This function is as smooth as the original function (see, e.g., [1, Sect. 3.5.21(i)]); in particular it is continuous. The result follows, since $\angle(Y_L,V_L)+\angle(Y_R,V_R)$ is a measure of distance that induces the manifold topology on $\operatorname{Grass}(p,n)\times\operatorname{Grass}(p,n)$. □

The next lemma gives a bound for step (a). It can be viewed as an "oblique" generalization of [37, Theorem 2], showing that the angles between the right Ritz vectors (the columns of Y_RW_R) and the "corresponding" right eigenvectors of C are of the order of the largest principal angle between $\mathcal{Y}_R$ and $\mathcal{V}_R$, and likewise for the left Ritz vectors and eigenvectors. We give an explicit asymptotic bound, but we do not attempt to make it tight.

Lemma 4.2 *Let $(\mathcal{V}_L,\mathcal{V}_R)$ be a pair of simple nondegenerate left–right eigenspaces of an $n\times n$ matrix C, let V_L, V_R be orthogonal bases for $\mathcal{V}_L$, $\mathcal{V}_R$, and let (λ,x) be an eigenpair of C with $x\in\mathcal{V}_R$. Given Y_L and Y_R $n\times p$ orthonormal with $Y_L^HY_R$ invertible, let $\epsilon_L=\angle(Y_L,V_L)$ and $\epsilon_R=\angle(Y_R,V_R)$. Then there exists $\epsilon_0>0$ such that, whenever $\epsilon_L<\epsilon_0$ and $\epsilon_R<\epsilon_0$, there exists an eigenvector w_R of the block Rayleigh quotient B_{Y_L,Y_R} such that*

$$\boxed{\sin\angle(Y_Rw_R,x)\leq\left[1+\frac{2(\cos\delta)^{-1}r_L\alpha_\delta(\epsilon_x)}{\operatorname{sep}\left(w_R^HBw_R,w_{R\perp}^HBw_{R_\perp}\right)-r_L\gamma_\delta(\epsilon_x)}\right](1+\tan\delta)\epsilon_x}\quad(6)$$

where $\epsilon_x:=\sin\angle(Y_R,x)$ is the angle between the direction of x and the span of Y_R, $\delta:=\angle(Y_R,Y_L)$ is the largest principal angle between the spans of Y_R and Y_L, $\alpha_\delta(\epsilon_x):=\frac{1}{\sqrt{1-\epsilon_x^2}-\epsilon_x\tan\delta}$ satisfies $\lim_{\epsilon_x\to0}\alpha_\delta(\epsilon_x)=1$, $\gamma_\delta(\epsilon_x):=(\cos\delta(\sqrt{1-\epsilon_x^2}-\epsilon_x\tan\delta))^{-1}(1+\tan\delta)\epsilon_x$ satisfies $\lim_{\epsilon_x\to0}\gamma_\delta(\epsilon_x)=0$, and $r_L:=\|Y_{L\perp}^HA^HY_L\|$ where $Y_{L\perp}\in\mathbb{C}^{n\times(n-p)}$ is an orthonormal basis of the orthogonal complement of the span of Y_L. In particular, for all $\theta>0$, by taking ϵ_0 sufficiently small, we have that

$$\angle(Y_Rw_R,x)\leq(1+\theta)\,(1+\tan\angle(V_L,V_R))\,\angle(Y_R,V_R).\quad(7)$$

Proof It is readily checked that the statement is not affected by a unitary change of coordinates in $\mathbb{C}^n$. Therefore, without loss of generality, we work in a unitary

coordinate system such that $Y_R = \begin{bmatrix} I_p \\ 0_{(n-p)\times p} \end{bmatrix}$. Let $Y_{L\perp} \in \mathbb{C}^{n\times(n-p)}$ and $Y_{R\perp} \in \mathbb{C}^{n\times(n-p)}$ be orthonormal bases of the orthogonal complements of the spans of Y_L and Y_R, respectively. Assume without loss of generality that the eigenvector x has unit norm. Consider the block decompositions $x = \begin{bmatrix} x_a \\ x_b \end{bmatrix}$ and $Y_L = \begin{bmatrix} Y_{La} \\ Y_{Lb} \end{bmatrix}$. Consider also the decomposition $x = Y_R x_R + Y_{L\perp} x_{L\perp}$, which yields

$$x_R := \left(Y_L^H Y_R\right)^{-1} Y_L^H x, \quad x_{L\perp} := \left(Y_{R\perp}^H Y_{L\perp}\right)^{-1} Y_{R\perp}^H x.$$

Since $\epsilon_x = \sin\angle(Y_R, x)$, we have $\|x_a\|^2 = 1 - \epsilon_x^2$ and $\|x_b\| = \epsilon_x$. We also have $(Y_L^H Y_R)^{-1} Y_L^H = \begin{bmatrix} I & T \end{bmatrix}$ where $T = (Y_{La})^{-1} Y_{Lb}$. It follows as a direct consequence of [38, Theorem 4.2.4] that $\|T\| = \tan\delta$. We also obtain

$$Y_R x_R = \begin{bmatrix} I \\ 0 \end{bmatrix} \begin{bmatrix} I & T \end{bmatrix} x = \begin{bmatrix} x_a + T x_b \\ 0 \end{bmatrix}.$$

Acceptable choices for $Y_{L\perp}$ and $Y_{R\perp}$ are $Y_{L\perp} = \begin{bmatrix} -T \\ I_{n-p} \end{bmatrix} (I_{n-p} + T^H T)^{-1/2}$ and $Y_{R\perp} = \begin{bmatrix} 0_{p\times(n-p)} \\ I_{n-p} \end{bmatrix}$. This yields $x_{L\perp} = (I_{n-p} + T^H T)^{1/2} x_b$ and thus $\|x_{L\perp}\| \leq \sqrt{1+\tan^2\delta}\,\epsilon_x$.

Since $\sin\angle(u, v) \leq \sin\angle(u, w) + \sin\angle(w, v)$ for all $u, v, w \in \mathbb{C}_0^n$, we have

$$\angle(Y_R w_R, x) \leq \angle(Y_R w_R, Y_R x_R) + \angle(Y_R x_R, x). \tag{8}$$

Let us first consider the second term in (8). Since

$$\sin\angle(Y_R x_R, x) \leq \|Y_R x_R - x\| \leq \left\| Y_R x_R - \begin{bmatrix} x_a \\ 0 \end{bmatrix} \right\| + \left\| \begin{bmatrix} x_a \\ 0 \end{bmatrix} - x \right\|,$$

it follows that

$$\sin\angle(Y_R x_R, x) \leq \|T x_b\| + \|x_b\| \leq \tan\delta\,\epsilon_x + \epsilon_x = (1+\tan\delta)\epsilon_x. \tag{9}$$

Note also for later use that, for all small ϵ_x such that $\sqrt{1-\epsilon_x^2} > \epsilon_x \tan\delta$, we also obtain that $\|x_R\| \geq |\|x_a\| - \|T x_b\|| \geq \sqrt{1-\epsilon_x^2} - \epsilon_x \tan\delta$.

We now tackle the first term in (8). Since Y_R is orthonormal, it follows that $\angle(Y_R w_R, Y_R x_R) = \angle(w_R, x_R)$. Pre-multiplying the equation $Cx = \lambda x$ by $(Y_L^H Y_R)^{-1} Y_L^H$ yields

$$(Y_L^H Y_R)^{-1} Y_L^H C (Y_R x_R + Y_{L\perp} x_{L\perp}) = x_R \lambda,$$

which can be rewritten as

$$(B + E)\hat{x}_R = \lambda \hat{x}_R,$$

where $\hat{x}_R := x_R \|x_R\|^{-1}$ and

$$E := \left(Y_L^H Y_R\right)^{-1} Y_L^H A Y_{L\perp} x_{L\perp} \|x_R\|^{-1} \hat{x}_R^H.$$

Then, by [35, Theorem V.2.7] and using [38, Theorem 4.2.4], there exists an eigenvector w_R of B such that

$$\sin \angle(w_R, \hat{x}_R) \le \tan \angle(w_R, \hat{x}_R) \le \frac{2\|E\|}{\operatorname{sep}\left(w_R^H B w_R, (w_R)_\perp^H B (w_R)_\perp\right) - 2\|E\|}$$

holds under some conditions that, in view of Lemma 4.1, hold whenever $\|E\|$, ϵ_R, and ϵ_L are sufficiently small. Choosing $\|E\|$ sufficiently small is achieved by taking ϵ_x sufficiently small, since

$$\begin{aligned}\|E\| = \left\| \left(Y_L^H Y_R\right)^{-1} Y_L^H C Y_{L\perp} x_{L\perp} \right\| \|x_R\|^{-1} &\le \left\| \left(Y_L^H Y_R\right)^{-1} \right\| \|Y_{L\perp}^H C^H Y_L\| \|x_{L_\perp}\| \|x_R\|^{-1} \\ &\le \frac{1}{\cos\delta} r_L (1 + \tan\delta) \epsilon_x \frac{1}{\sqrt{1-\epsilon_x^2} - \epsilon_x \tan\delta},\end{aligned}$$

where we have used the bound $\sqrt{1 + \tan^2\delta} \le (1 + \tan\delta)$ that holds for all $\delta \in [0, \frac{\pi}{2})$. Replacing all these results in (8) yields the desired bound.

The bound (7) follows from (6) and from the fact that $\angle(Y_R, V_R) \ge \angle(Y_R, x)$ since $x \in \mathcal{V}_R$. □

The next lemma is about how well the two-sided Rayleigh quotient approximates the eigenvalues.

Lemma 4.3 *Let C be an $n \times n$ matrix, let v_L and v_R be left and right eigenvectors of C with eigenvalue λ and $v_L^H v_R \neq 0$. Let w_L and w_R be unit vectors orthogonal to v_L and v_R, respectively. Then, for all $\theta > 0$ and for all ϵ, η sufficiently small,*

$$\left| \frac{(v_L + \epsilon w_L)^H C (v_R + \eta w_R)}{(v_L + \epsilon w_L)^H (v_R + \eta w_R)} - \lambda \right| \le (1 + \theta)\epsilon\eta \frac{\|P_{v_L}^\perp (C - \lambda I) P_{v_R}^\perp\|}{|v_L^H v_R|},$$

where $P_v^\perp$ denotes the orthogonal projector onto the orthogonal complement of v.

Proof Routine manipulations show that $\left| \frac{(v_L+\epsilon w_L)^H C (v_R+\eta w_R)}{(v_L+\epsilon w_L)^H (v_R+\eta w_R)} - \lambda \right| = \left| \frac{\epsilon\eta w_L^H (C-\lambda I) w_R}{v_L^H v_R + \epsilon w_L^H v_R + \eta v_L^H w_R + \epsilon\eta w_L^H w_R} \right|$, and the result directly follows. □

We now obtain a bound for step (b). In view of the preceding lemma, a careful inspection of the proof of local convergence in [26, Sect. 11] yields the following local convergence result for the classical two-sided RQI.

Lemma 4.4 *Let C be an $n \times n$ matrix, let v_L and v_R be left and right eigenvectors of C with eigenvalue λ and $v_L^H v_R \neq 0$. Let w_L and w_R be unit vectors orthogonal to v_L and v_R, respectively. Then, for all $\theta > 0$ and for all ϵ, η sufficiently small and such that*

$$\rho := \frac{(v_L + \epsilon w_L)^H C (v_R + \eta w_R)}{(v_L + \epsilon w_L)^H (v_R + \eta w_R)}$$

is not an eigenvalue of C, if we let $(v_L + \epsilon_+ w_{L+})(C - \rho I) = (v_L + \epsilon w_L)^H v$ and $(C - \rho I)(v_R + \eta_+ w_{R+}) = (v_R + \eta w_R)\tau$ with w_{L+} and w_{R+} unit vectors orthogonal to v_L and v_R respectively, then we have

$$\epsilon_+ \leq \epsilon^2 \eta (1+\theta) \frac{\|P_{v_L}^\perp (C - \lambda I) P_{v_R}^\perp\| \|P_{v_L}^\perp (C - \lambda I)^{-1} P_{v_R}^\perp\|}{|v_L^H v_R|}$$
$$\eta_+ \leq \eta^2 \epsilon (1+\theta) \frac{\|P_{v_L}^\perp (C - \lambda I) P_{v_R}^\perp\| \|P_{v_L}^\perp (C - \lambda I)^{-1} P_{v_R}^\perp\|}{|v_L^H v_R|}.$$

It remains to address step (c). The final lemma gives a bound for the angle between two subspaces in terms of the column-wise angles between bases.

Lemma 4.5 *Let X be a full-rank $n \times p$ matrix. Then for all $\theta > 0$, there exists $\epsilon > 0$ such that, for all $Y \in \mathbb{C}^{n \times p}$ with $\angle(Xe_i, Ye_i) \leq \epsilon$, $i = 1, \ldots, p$, one has*

$$\angle(X, Y) \leq (1+\theta) \|(\hat{X}^H \hat{X})^{-1/2}\| p \max_i \angle(Xe_i, Ye_i),$$

where $\hat{X} e_i = Xe_i / \|Xe_i\|$, $i = 1, \ldots, p$.

Proof Let D_X and D_Y be such that $\hat{X} = X D_X$ and $\hat{Y} = Y D_Y$ have unit columns with $(\hat{Y} e_i)^H \hat{X} e_i > 0$, $i = 1, \ldots, p$, which can be achieved by an appropriate choice of D_Y, i.e., by an additional scaling of $\hat{Y} e_i$ by a phase factor $\exp(i\alpha_i)$. Let E be such that $\hat{Y} = \hat{X} + E$. Observe that $\lfloor \hat{X} \rfloor = \lfloor X \rfloor$ and likewise for Y. We have $\|E\| \leq \|E\|_F \leq \|Ee_1\| + \cdots + \|Ee_p\| \leq \sum_i 2 \sin(\angle(Xe_i, Ye_i)/2) \leq \sum_i \angle(Xe_i, Ye_i) \leq p \max_i \angle(Xe_i, Ye_i)$, where $\|E\|_F$ denotes the Frobenius norm of E. Since by definition $\cos \angle(X, Y) = \sigma_{\min}(\tilde{X}^H \hat{Y} (\hat{Y}^H \hat{Y})^{-1/2})$, where $\tilde{X}$ is any orthonormal basis of $\lfloor X \rfloor$, it is easily deduced (e.g., from the CS decomposition) that $\sin \angle(X, Y) = \sigma_{\max}(X_\perp^H \hat{Y} (\hat{Y}^H \hat{Y})^{-1/2})$, where $X_\perp$ is $n \times (n-p)$ orthonormal with $X^H X_\perp = 0$. We have

$$\begin{aligned}
\sin \angle(X, Y) &= \sigma_{\max} \left(X_\perp^H E (\hat{X}^H \hat{X} + \hat{X}^H E + E^H \hat{X} + E^H E)^{-1/2} \right) \\
&= \|X_\perp^H E (\hat{X}^H \hat{X} + \hat{X}^H E + E^H \hat{X} + E^H E)^{-1/2}\| \\
&\leq (1+\theta) \|E\| \|(\hat{X}^H \hat{X})^{-1/2}\| \\
&\leq (1+\theta) p \max_i \angle(Xe_i, Ye_i) \|(\hat{X}^H \hat{X})^{-1/2}\|
\end{aligned}$$

whenever $\max_i \angle(Xe_i, Ye_i)$ is sufficiently small. □

We are now ready to prove the main theorem, showing local cubic convergence of 2sGRQI.

Theorem 4.6 *Let $(\mathcal{V}_L, \mathcal{V}_R)$ be a pair of p-dimensional simple nondegenerate left–right eigenspaces of an $n \times n$ matrix C (Definition 2.1). Then there is a neighborhood $\mathcal{N}$ of $(\mathcal{V}_L, \mathcal{V}_R)$ in $\mathrm{Grass}(p,n) \times \mathrm{Grass}(p,n)$ and a $c > 0$ such that, for all $(\mathcal{Y}_L, \mathcal{Y}_R) \in \mathcal{N}$ for which the Sylvester equations (4) admit one and only one solution, the subspaces $\mathcal{Y}_{L+}$ and $\mathcal{Y}_{R+}$ produced by the 2sGRQI mapping (Algorithm 3.1) satisfy*

$$\angle(\mathcal{Y}_{L_+}, \mathcal{V}_L) + \angle(\mathcal{Y}_{R_+}, \mathcal{V}_R) \leq c\,(\angle(\mathcal{Y}_L, \mathcal{V}_L) + \angle(\mathcal{Y}_R, \mathcal{V}_R))^3. \tag{10}$$

Letting V_L and V_R denote matrices of unit left and right eigenvectors that span $\mathcal{V}_L$ and $\mathcal{V}_R$, the bound (10) is satisfied with

$$c = (1+\theta)p\left(\left\|\left(V_R^H V_R\right)^{-1/2}\right\| + \left\|\left(V_L^H V_L\right)^{-1/2}\right\|\right) \max_{v_L = V_L e_i, v_R = V_R e_i} \times \frac{\left\|P_{v_L}^{\perp}(C-\lambda I)P_{v_R}^{\perp}\right\| \left\|P_{v_L}^{\perp}(C-\lambda I)^{-1}P_{v_R}^{\perp}\right\|}{\left|v_L^H v_R\right|} \left(1+\tan\angle(V_L, V_R)\right)^3,$$

for any $\theta > 0$, by taking $\mathcal{N}$ sufficiently small.

Proof Since the pair of eigenspaces is assumed to be simple, it follows that $\angle(V_L, V_R) < \pi/2$. Therefore, taking the neighborhood $\mathcal{N}$ sufficiently small, one has $\angle(Y_R, Y_L) \leq \delta' < \pi/2$. The result follows from (5) by exploiting Lemmas 4.2, 4.4, and 4.5 successively. □

5 Structured eigenproblems

In this section, we show that the 2sGRQI algorithm induces particular one-sided formulations for some structured eigenproblems.

5.1 E-Hermitian eigenproblem

Let C be an $n \times n$ matrix. If there exists an invertible matrix E such that

$$EC = C^H E, \tag{11}$$

then we say that C is *E-Hermitian*. If C is E-Hermitian, then its left and right eigenspaces are related by the action of E. Indeed, let S be a (complex) matrix of principal vectors of C, i.e.,

$$CS = SD$$

where D is a (complex) Jordan matrix; then, from (11), one obtains $C^H(ES) = (ES)D$.

The case where E is Hermitian or skew-Hermitian, i.e., $E^H = \pm E$, is of particular interest because, as we show in the next proposition, the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant under 2sGRQI (Algorithm 3.1). Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, it is not necessary to solve both (4a) and (4b): just solve (4a) to get $\mathcal{Y}_{R+}$, and obtain $\mathcal{Y}_{L+}$ as $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, since the pairs of left–right eigenspaces of C also satisfy $\mathcal{V}_L = E\mathcal{V}_R$, Theorem 4.6 also applies.

Proposition 5.1 *Let E be invertible with $E^H = \pm E$ and let C be E-Hermitian, i.e., $EC = C^H E$. If $Y_L = EY$, $Y_R = Y$, and Z satisfies*

$$\boxed{CZ - Z(Y^H EY)^{-1}(Y^H ECY) = Y,} \tag{12}$$

then $Z_L = EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations (4). Hence, if $\mathcal{Y}_L = E\mathcal{Y}_R$, then $\mathcal{Y}_{L+} = E\mathcal{Y}_{R+}$. Moreover, the subspace iteration $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ defined by (12) converges locally cubically to the simple nondegenerate right-eigenspaces of C.

Proof It is easy to check that replacing $Y_R := Y$, $Z_R := Z$, $Y_L := EY_R$, $Z_L := EZ_R$ in (4a) and (4b) yields (12) in both cases. In order to prove cubic convergence, it is sufficient to notice that the pairs $(\mathcal{V}_L, \mathcal{V}_R)$ of eigenspaces satisfy $\mathcal{V}_L = E\mathcal{V}_R$, as was shown above. Therefore, if $\mathcal{Y}$ is close to $\mathcal{V}_R$, then the pair $(\mathcal{Y}_L, \mathcal{Y}_R) := (E\mathcal{Y}, \mathcal{Y})$ is close to $(\mathcal{V}_L, \mathcal{V}_R)$ and local cubic convergence to $\mathcal{V}_R$ follows from Theorem 4.6. □

The discussion in Sect. 6 on solving Sylvester equations applies likewise to (12).

5.1.1 Generalized Hermitian eigenproblem

Using Proposition 5.1, we show that 2sGRQI yields a Grassmannian RQI for the Hermitian generalized eigenproblem $A\mathcal{V} \subset B\mathcal{V}$ which does not involve an explicit computation of $B^{-1}A$. Let A and B be two Hermitian n-by-n matrices with B invertible. Consider the problem of finding a p-dimensional subspace $\mathcal{V}$ such that $A\mathcal{V} \subset B\mathcal{V}$. Let $V \in \mathbb{C}^{n\times p}$ be a basis for $\mathcal{V}$, then $A\mathcal{V} \subset B\mathcal{V}$ if and only if there is a matrix M such that $AV = BVM$. Equivalently, V spans a right-eigenspace of $B^{-1}A$, i.e.,

$$B^{-1}AV = VM.$$

The problem is thus to find a right-eigenspace of $C := B^{-1}A$. The conditions in Proposition 5.1 are satisfied with $E := B$. The modified GRQI equation (12) becomes

$$\boxed{AZ - BZ(Y^H BY)^{-1}(Y^H AY) = BY} \tag{13}$$

and the subspace iteration $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ converges locally cubically to the simple nondegenerate eigenspaces of $B^{-1}A$.

5.1.2 Skew-Hamiltonian eigenproblem

Let T be a skew-Hamiltonian matrix, i.e., $(TJ)^H = -TJ$, where $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$, see e.g., [6]. Equivalently, $JT = T^H J$, i.e., T is J-Hermitian. Conditions in Proposition 5.1 are satisfied with $C := T$ and $E := J$. The modified GRQI equation (12) becomes

$$\boxed{TZ - Z(Y^H JY)^{-1}(Y^H JTY) = Y} \tag{14}$$

and the subspace iteration $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ converges locally cubically to the simple nondegenerate right-eigenspaces of T.

5.2 E-skew-Hermitian eigenproblem

Let E be an invertible $n \times n$ matrix and let C be an *E-skew-Hermitian* $n \times n$ matrix, namely

$$EC = -C^H E. \tag{15}$$

We saw in the previous section that the corresponding left and right eigenspaces of E-Hermitian matrices are related by a multiplication by E. The case of E-skew-Hermitian matrices is slightly different.

Proposition 5.2 *Let C be an E-skew-Hermitian matrix. Then the spectrum of C is symmetric with respect to the imaginary axis. In other words, if λ is an eigenvalue of C, then so is $-\overline{\lambda}$. Moreover, if $\mathcal{V}_L$ and $\mathcal{V}_R$ are left and right eigenspaces of C whose spectra are the symmetric image one of the other with respect to the imaginary axis, then $\mathcal{V}_L = E\mathcal{V}_R$.*

Proof Letting S be an invertible matrix of principal vectors of C, i.e.,

$$CS = SD \tag{16}$$

where D is a Jordan matrix, (15) yields

$$C^H ES = ES(-D). \tag{17}$$

Hence, the matrix $-D$ is a Jordan matrix of C^H. Therefore, if λ is an eigenvalue of C, then $-\lambda$ is an eigenvalue of C^H, and thus $-\overline{\lambda}$ is an eigenvalue of C. Moreover, equations (16) and (17) show that if $\mathcal{V}$ is a right-eigenspace of C with eigenvalues $\lambda_{i_1}, \ldots, \lambda_{i_p}$, then $E\mathcal{V}$ is a left-eigenspace of C with eigenvalues $-\overline{\lambda}_{i_1}, \ldots, -\overline{\lambda}_{i_p}$. □

Consequently, letting $\mathcal{V}$ be a simple right-eigenspace of C, we have that $(E\mathcal{V}, \mathcal{V})$ forms a pair of simple left–right eigenspaces of C if and only if the spectrum of $\mathcal{V}$ is symmetric with respect to the imaginary axis. We call such an invariant subspace $\mathcal{V}$ a *full eigenspace* of the E-skew-Hermitian matrix C.

If E is Hermitian or skew-Hermitian, then the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant by 2sGRQI (Algorithm 3.1), as we show in the forthcoming proposition. Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, it is sufficient to solve (4a) only, and then compute $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, the 2sGRQI method restricted to the pairs $(\mathcal{Y}_L, \mathcal{Y}_R) = (E\mathcal{Y}, \mathcal{Y})$ converges locally cubically to the full nondegenerate eigenspaces of C.

Proposition 5.3 *Let E be invertible with $E^H = \pm E$ and let C be E-skew-Hermitian, i.e., $EC = -C^H E$. If $Y_L = EY$ and $Y_R = Y$, then $Z_L = -EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations (4) with*

$$\boxed{CZ - Z\,(Y^H EY)^{-1}(Y^H ECY) = Y.} \tag{18}$$

Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, then $\mathcal{Y}_{L+} = E\mathcal{Y}_{R+}$.

Moreover, let $\mathcal{V}$ be a full *nondegenerate right-eigenspace of C (which means that the eigenvalues of $C|_{\mathcal{V}}$ have the same multiplicity as in C, the spectrum of $C|_{\mathcal{V}}$ is symmetric with respect to the imaginary axis, and $C|_{\mathcal{V}}$ has all simple eigenvalues). Then the subspace iteration $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ defined by (12) converges locally cubically to $\mathcal{V}$.*

Note that this proposition differs from Proposition 5.1 in two points: $Z_L = -EZ$ and the specification that $\mathcal{V}$ must be full.

Proof It is easy to check that replacing $Y_R := Y, Z_R := Z, Y_L := EY_R, Z_L := -EZ_R$ in (4a) and (4b) yields (18) in both cases. In order to prove cubic convergence, it is sufficient to notice that the pairs $(\mathcal{V}_L, \mathcal{V}_R)$ of full nondegenerate left–right eigenspaces satisfy $\mathcal{V}_L = E\mathcal{V}_R$, as was shown above. Therefore, if $\mathcal{Y}$ is close to $\mathcal{V}_R$, then the pair $(\mathcal{Y}_L, \mathcal{Y}_R) := (E\mathcal{Y}, \mathcal{Y})$ is close to $(\mathcal{V}_L, \mathcal{V}_R)$ and local cubic convergence to $\mathcal{V}$ follows from Theorem 4.6. □

5.2.1 Skew-Hermitian eigenproblem

Let Ω be skew-Hermitian. Then we have $EC = -C^H E$ with $C := \Omega$ and $E := I$. The modified GRQI equation (18) becomes

$$\boxed{\Omega Z - Z\,(Y^H Y)^{-1}(Y^H \Omega Y) = Y.} \tag{19}$$

This is simply the classical GRQI equation (3). This is not surprising as skew-Hermitian matrices are normal matrices.

5.2.2 Hamiltonian eigenproblem

Let H be Hamiltonian, i.e. $(HJ)^H = HJ$, where $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. This is equivalent to $JH = -H^H J$. Thus we have $EC = -C^H E$ with $C := H$ and $E := J$, and the modified GRQI equation (18) reads

$$\boxed{HZ - Z\,(Y^H JY)^{-1}(Y^H JHY) = Y.} \tag{20}$$

Proposition 5.3 implies that the subspace iteration with iteration mapping $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ defined by (20) converges locally cubically to the full nondegenerate right-eigenspaces of H.

5.3 The generalized eigenvalue problem

We briefly discuss the application of the 2sGRQI concept to the generalized eigenvalue problem. Let $A, B \in \mathbb{C}^{n\times n}$. The generalized eigenvalue problem consists in finding the nontrivial solutions of the equation $Ax = \lambda Bx$. Corresponding to the notion of invariant subspace for a single matrix, we have the notion of a *deflating subspace*, see e.g., [20,36]. The p-dimensional subspace $\mathcal{X}$ is deflating for the pencil $A - \lambda B$ if there exists a p-dimensional subspace $\mathcal{Y}$ such that

$$A\mathcal{X}, B\mathcal{X} \subset \mathcal{Y}. \tag{21}$$

Here we suppose that the pencil $A - \lambda B$ is nondegenerate, i.e., $\det(A - \lambda B)$ is not trivially zero. Then there exists α and β such that $\hat{B} := \alpha B - \beta A$ is invertible. Now take γ, δ such that $\alpha\delta - \gamma\beta \neq 0$ and let $\hat{A} := \gamma B - \delta A$. Then (21) is equivalent to

$$\begin{aligned} \hat{B}^{-1}\hat{A}\mathcal{X} &\subset \mathcal{X} \\ \hat{B}\mathcal{X} &= \mathcal{Y}, \end{aligned}$$

i.e., $\mathcal{X}$ is an invariant subspace of $\hat{B}^{-1}\hat{A}$. Replacing this expression for C in (4), one obtains after some manipulations

$$\hat{A}Z_R\hat{Y}_L^H\hat{B}Y_R - \hat{B}Z_R\hat{Y}_L^H\hat{A}Y_R = \hat{B}Y_R \tag{22a}$$

$$\hat{A}^H\hat{Z}_LY_R^H\hat{B}^H\hat{Y}_L - \hat{B}^H\hat{Z}_LY_R^H\hat{A}^H\hat{Y}_L = \hat{B}^H\hat{Y}_L \tag{22b}$$

where $\hat{Y}_L := \hat{B}^{-H}Y_L$ and $\hat{Z}_L := \hat{B}^{-H}Z_L$. It yields an iteration for which Y_R and $\hat{Y}_L$ locally cubically converge to pairs of left–right deflating subspaces of the pencil $A - \lambda B$. Note that if B is invertible then we can choose $\hat{B} := B$ and $\hat{A} := A$.

6 Numerical experiments

In our numerical tests, 2sGRQI (Algorithm 3.1) has been implemented in Matlab as follows. The initial spaces $\mathcal{Y}_{L_0}$ and $\mathcal{Y}_{R_0}$ are represented by orthonormal bases Y_{L_0} and Y_{R_0}. Step 1 is irrelevant. In Step 3, $\mathcal{Y}_{R+}$ is represented by an orthonormal basis obtained as the Q factor of the thin QR decomposition of Z_R; likewise for the left-hand objects. Remains Step 2, which is the topic of the next section.

6.1 Solving the Sylvester equations

The crucial point in the practical implementation is to choose a method for solving the decoupled Sylvester equations (4) in Step 2. We discuss the case of (4a), namely,

$$CZ_R - Z_R \underbrace{\left(Y_L^H Y_R\right)^{-1} Y_L^H C Y_R}_{R_R} = Y_R;$$

the discussion is easily adapted to the case of (4b). There is a vast literature on the solution of Sylvester equations; we refer to [7,8,16,18,33] and references therein. We performed a set of experiments to investigate whether the choice of the Sylvester solver has an impact on the accuracy with which 2sGRQI is able to compute eigenspaces. We compared two approaches.

In the first approach, the Sylvester equation (4a) is decoupled into p classical two-sided RQI equations as explained in (5). This involves a diagonalization of the block Rayleigh quotient R_R, which can always be performed if $(\mathcal{Y}_L, \mathcal{Y}_R)$ are sufficiently close to a pair of simple nondegenerate left–right eigenspaces of C.

It should be observed, however, that even though R_R is diagonalizable for almost all $(\mathcal{Y}_L, \mathcal{Y}_R)$, its eigenstructure may be very ill-conditioned, in the sense that the eigensystem condition number $\kappa(R_R)$ may be very large. A remedy is to reduce R_R to a certain triangular structure (Schur form) by means of unitary transformations and solve the new system of equations using back substitution, as described in [16]. This is the second approach.

It is known [9] that the basins of attraction of RQI (Algorithm 2.3) may collapse around attractors when the eigenvalues of A are not well separated. This property also holds for GRQI [5] and obviously extends to 2sGRQI (Algorithm 3.1). Moreover, in 2sGRQI the matrix C is not necessarily Hermitian; its eigenspaces can thus be arbitrarily close to each other. In this set of experiments, in order to ensure a reasonably large basin of attraction around the left–right eigenspaces, we ruled out clustered eigenvalues and ill-separated eigenvectors by choosing C as follows: $C = SDS^{-1}$, where D is a diagonal matrix whose diagonal elements are random permutations of $1, \ldots, n$ and

$$S = I + \tfrac{\alpha}{\|E\|_2} E, \tag{23}$$

where the elements of E are observations of independent random variables with standard normal distribution and α is chosen from the uniform distribution on the interval $(0, 0.1)$. The initial matrices Y_{L_0} and Y_{R_0} are randomly chosen such that $\angle(Y_{R_0}, V_R) < 0.1$ and $\angle(Y_{L_0}, V_L) < 0.1$, where V_R denotes the p first columns of S, V_L denotes the first p columns of S^{-H}, and where $\angle$ still denotes the largest principal angle.

Algorithm 3.1 was run 10^6 times with $n = 4$, $p = 2$. We chose these very low dimensions because we found that they reveal the same kind of behavior as the higher-dimensinal cases, while making it possible to run many experiments and to investigate more easily the pathological runs. The matrices C, Y_{L_0}, and Y_{R_0} were randomly chosen in each experiment as explained above. Experiments were run using Matlab 7.2 with

Table 1 Numerical experiments for Algorithm 3.1 with diagonal form

Iterate number	Mean [log10(e)]	Max [log10(e)]
0	−1.43	−1.00
1	−4.70	−2.81
2	−13.93	−7.48
3	−17.17	−12.70
4	−17.17	−11.96
5	−17.17	−12.82

Table 2 Numerical experiments for Algorithm 3.1 with Schur form

Iterate number	Mean [log10(e)])	Max [log10(e)]
0	−1.43	−1.00
1	−4.70	−2.81
2	−13.92	−6.95
3	−17.17	−10.98
4	−17.17	−10.39
5	−17.16	−11.20

floating point relative accuracy approximately equal to 2×10^{-16}. Results are summarized in Tables 1 (diagonal form) and 2 (Schur form), where the error e is defined as

$$e := \angle(Y_R, V_R) + \angle(Y_L, V_L). \tag{24}$$

These results show that convergence to the target eigenspace occurred in each of the 10^6 runs. The evolution of the error is compatible with cubic order of convergence.

We also observe that the worst error observed over the 10^6 runs is slightly favorable to the diagonal version. We tried to understand the reason by investigating a case where the Schur version performs much more poorly than the diagonal version. For a certain $C = SDS^{-1}$, X_L, X_R, we obtained

$$S^{-1}\tilde{Z}_{R(\text{diag})} = \begin{bmatrix} 8.1250e-01 & -1.8226e+15 \\ 5.5567e+18 & 1.0681e-03 \\ -9.5000e+00 & -5.2490e-03 \\ 0 & 7.8125e-03 \end{bmatrix},$$

$$S^{-1}\tilde{Z}_{R(\text{Schur})} = \begin{bmatrix} 8.1250e-01 & -9.1049e+13 \\ 5.5567e+18 & -1.2981e+17 \\ -9.5000e+00 & -8.0141e+01 \\ 0 & -6.9000e+01 \end{bmatrix}.$$

We have $\angle(\tilde{Z}_{R(\text{diag})}, V_R) = 5.8849e-16$ and $\angle(\tilde{Z}_{R(\text{Schur})}, V_R) = 1.2525e-12$. The matrix $\tilde{Z}_{R(\text{Schur})}$ is a source of worry because the angle between its columns is very small (of the order of the machine precision), and hence the subspace that it spans is very ill-conditioned.

6.2 Clustered eigenvalues

We now report on experiments where the matrices C are chosen as in Sect. 6.1, except that the matrix D of eigenvalues is chosen as $\mathrm{diag}(\begin{bmatrix}1 & 1+10^{-3} & 2 & 3\end{bmatrix})$, i.e., with a small internal gap. Convergence is declared when the error (24) is below 10^{-12} at the 10th iterate. For 2sGRQI (in diagonal form), convergence was declared 100% of the time over the 10^4 runs that we have performed.

We compared with a simple "independent 2sRQI" method that works as follows: first compute the Ritz vectors, then run a certain number of steps of the classical two-sided RQI starting from these p vectors, and *without* any coupling. This contrasts with 2sGRQI, where the Ritz vectors are extracted after each step. For the "independent 2sRQI" method, convergence was declared 74% of the time. This shows that 2sGRQI improves on independent 2sRQI in case of clustered eigenvalues.

6.3 Ill-conditioned eigenspaces

In these experiments, the matrices C are chosen as in Sect. 6.1, except that the matrix S of eigenvectors is chosen as in (23) with I replaced by $\begin{bmatrix}e_1 & e_1 & e_3 & e_4\end{bmatrix}$. Convergence is declared when the error (24) is below 10^{-12} at the 10th iterate. For 2sGRQI (in diagonal form), convergence was declared 83% of the time. With the "independent 2sRQI" method, convergence was declared 78% of the time. In this case, 2sGRQI does not improve much on independent 2sRQI.

6.4 Hamiltonian eigenproblem

In another set of experiments, real Hamiltonian matrices C were selected randomly as

$$C = \begin{bmatrix} F & \tilde{G} + \tilde{G}^H \\ \tilde{H} + \tilde{H}^H & -F^H \end{bmatrix}$$

where F, $\tilde{G}$ and $\tilde{H}$ are matrices of dimension $\frac{n}{2} \times \frac{n}{2}$ whose elements are independent observations of the standard normally distributed random variable. A new matrix C was selected for each experiment. For testing purposes, an eigenvalue decomposition $C = SDS^{-1}$ was computed using the Matlab `eig` function, and the *full* left and right real eigenspaces corresponding to the eigenvalues with largest real part in magnitude were chosen as the target left and right eigenspaces. (The notion of *full* eigenspace is defined in Sect. 5.2. The *real* eigenspace associated to a pair $(\lambda, \overline{\lambda})$ of complex conjugate eigenvalues with eigenvectors $v_r + i v_i$ and $v_r - i v_i$ is the span of v_r and v_i.) The eigenvalue decomposition was ordered in such a way that $\lfloor V_L \rfloor$ is the target left-eigspace and $\lfloor V_R \rfloor$ is the target right-eigenspace, where V_R denotes the first p columns of S and V_L the first p columns of S^{-H}. Note that we have $p = 2$ when the target eigenvalues are real (λ and $-\lambda$), or $p = 4$ when the target eigenvalues have a nonzero imaginary part (λ, $\overline{\lambda}$, $-\lambda$, and $-\overline{\lambda}$). The initial matrix Y_{R_0} was randomly chosen such that $\angle(Y_{R_0}, V_R) < 0.1$, and Y_{L_0} was chosen as JY_{R_0} in accordance with the material

of Sect. 5.2. Convergence to the target left and right eigenspaces was declared when the error e as defined above was smaller than 10^{-12} at the 10th iterate. Algorithm 3.1 was run 10^6 times with $n = 20$ with the matrices C, Y_{L_0} and Y_{R_0} randomly chosen in each experiment as described above. Note that, in accordance with the material in Sect. 5.2, only Z_R was computed at each iteration; Z_L was chosen as JZ_R. We observed that convergence to the target eigenspaces was declared for 99.95% of the 10^6 experiments. Next, the experiment was run 10^6 times with the distance bound on the initial condition set to 0.001 instead of 0.1. Convergence to the target eigenspaces was declared for all but seven of the 10^6 randomly generated experiments. This confirms the potential of Algorithm 3.1 for refining initial estimates of *full eigenspaces* of Hamiltonian matrices.

7 Conclusion

We have shown that Ostrowski's two-sided iteration generalizes to an iteration on $\mathrm{Grass}(p, n) \times \mathrm{Grass}(p, n)$ that converges locally cubically to the pairs of simple non-degenerate left–right eigenspaces of arbitrary square matrices. The cubic order of convergence has been corroborrated by numerical experiments. The behavior of the 2sGRQI algorithm in case of ill-separated eigenvectors or eigenvalues would deserve further investigation, akin to the one performed for the Hermitian case in [5], where improvements of GRQI and of the Riemannian Newton algorithm were also proposed.

Acknowledgment This work was initiated when the first author was a guest in the Mathematics Department of the University of Würzburg under a grant from the European Nonlinear Control Network. The hospitality of the members of the Department is gratefully acknowledged. The first author would also like to thank especially Rodolphe Sepulchre for his careful guidance throughout his Ph.D. research and the impact this had on the results of this paper. The authors would like to thank the two referees for their careful reading of the manuscript and for their many helpful comments.

References

1. Abraham, R., Marsden, J.E., Ratiu, T.: Manifolds, Tensor Analysis, and Applications. Applied Mathematical Sciences, vol. 75, 2nd edn. Springer, New York (1988)
2. Absil, P.-A., Mahony, R., Sepulchre, R.: Riemannian geometry of Grassmann manifolds with a view on algorithmic computation. Acta Appl. Math. **80**(2), 199–220 (2004)
3. Absil, P.-A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton (2008)
4. Absil, P.-A., Mahony, R., Sepulchre, R., Van Dooren, P.: A Grassmann–Rayleigh quotient iteration for computing invariant subspaces. SIAM Rev. **44**(1), 57–73 (2002)
5. Absil, P.-A., Sepulchre, R., Van Dooren, P., Mahony, R.: Cubically convergent iterations for invariant subspace computation. SIAM J. Matrix Anal. Appl. **26**(1), 70–96 (2004)
6. Benner, P., Byers, R., Mehrmann, V., Xu, H.: Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils. SIAM J. Matrix Anal. Appl. **24**(1), 165–190 (2002)
7. Brandts, J.: Computing tall skinny solutions of $AX - XB = C$. Math. Comput. Simul. **61**(3–6), 385–397 MODELLING 2001 (Pilsen) (2003)
8. Bartels, R.H., Stewart, G.W.: Algorithm 432: solution of the matrix equation $AX + XB = C$. Commun. ACM **15**, 820–826 (1972)
9. Batterson, S., Smillie, J.: The dynamics of Rayleigh quotient iteration. SIAM J. Numer. Anal. **26**(3), 624–636 (1989)

10. Batterson, S., Smillie, J.: Rayleigh quotient iteration for nonsymmetric matrices. Math. Comp. **55**(191), 169–178 (1990)
11. Chatelin, F.: Simultaneous Newton's iteration for the eigenproblem. In: Defect Correction Methods (Oberwolfach, 1983), vol. 5 of Comput. Suppl., pp. 67–74. Springer, Vienna (1984)
12. Crandall, S.H.: Iterative procedures related to relaxation methods for eigenvalue problems. Proc. R. Soc. Lond. **207**(1090), 416–423 (1951)
13. Demmel, J.W.: Three methods for refining estimates of invariant subspaces. Computing **38**(1), 43–57 (1987)
14. Dongarra, J.J., Moler, C.B., Wilkinson, J.H.: Improving the accuracy of computed eigenvalues and eigenvectors. SIAM J. Numer. Anal. **20**(1), 23–45 (1983)
15. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998)
16. Gardiner, J.D., Laub, A.J., Amato, J.J., Moler, C.B.: Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$. ACM Trans. Math. Softw. **18**(2), 223–231 (1992)
17. Gohberg, I., Lancaster, P., Rodman, L.: Invariant subspaces of matrices with applications. Canadian Mathematical Society Series of Monographs and Advanced Texts. Wiley, New York, Republished 2006 as Classics in Applied Mathematics 51 by SIAM, Philadelphia (1986)
18. Golub, G.H., Nash, S., Van Loan, C.F.: A Hessenberg–Schur method for the problem $AX + XB = C$. IEEE Trans. Automat. Control **24**, 909–913 (1979)
19. Gu, M.: Single- and multiple-vector iterations. In: Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H. (eds.) Templates for the Solution of Algebraic Eigenvalue Problems, pp. 51–56. SIAM, Philadelphia (2000)
20. Golub, G.H., Van Loan, C.F.: Matrix Computations. Johns Hopkins Studies in the Mathematical Sciences, 3rd edn. Johns Hopkins University Press, Baltimore (1996)
21. Ipsen, I.C.F.: Computing an eigenvector with inverse iteration. SIAM Rev. **39**(2), 254–291 (1997)
22. Lundström, E., Eldén, L.: Adaptive eigenvalue computations using Newton's method on the Grassmann manifold. SIAM J. Matrix Anal. Appl., **23**(3), 819–839 (2001/2002)
23. Lösche, R., Schwetlick, H., Timmermann, G.: A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix. Linear Algebra Appl. **275/276**, 381–400 (1998)
24. Ostrowski, A.M.: On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. I–VI. Arch. Ration. Mech. Anal., **1**, 233–241, **2**, 423–428, **3**, 325–340, **3**, 341–347, **3**, 472–481, and **4**, 153–165 (1959)
25. Ostrowski, A.M.: On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III. Arch. Ration. Mech. Anal. **3**, 325–340 (1959)
26. Parlett, B.N.: The Rayleigh quotient iteration and some generalizations for nonnormal matrices. Math. Comp. **28**, 679–693 (1974)
27. Parlett, B.N.: The Symmetric Eigenvalue Problem. Prentice-Hall, Englewood Cliffs. Republished 1998 as Classics in Applied Mathematics 20 by SIAM, Philadelphia (1980)
28. Parlett, B.N., Kahan, W.: On the convergence of a practical QR algorithm. In: Information Processing 68 (Proc. IFIP Congress, Edinburgh, 1968), vol. 1. Mathematics, Software, pp 114–118, North-Holland, Amsterdam (1969)
29. Peters, G., Wilkinson, J.H.: Inverse iteration, ill-conditioned equations and Newton's method. SIAM Rev. **21**(3), 339–360 (1979)
30. Qiu, L., Zhang, Y., Li, C.-K.: Unitarily invariant metrics on the Grassmann space. SIAM J. Matrix Anal. Appl. **27**(2), 507–531 (electronic) (2005)
31. Rayleigh, J.W.: The Theory of Sound. Macmillan, London (1877)
32. Scharnhorst, K.: Angles in complex vector spaces. Acta Appl. Math. **69**(1), 95–103 (2001)
33. Simoncini, V.: On the numerical solution of $AX - XB = C$. BIT **36**(4), 814–830 (1996)
34. Smit, P.: Numerical Analysis of Eigenvalue Algorithms Based on Subspace Iterations. Ph.D. Thesis, CentER, Tilburg University, Tilburg, The Netherlands (1997)
35. Stewart, G.W., Sun, J.G.: Matrix Perturbation Theory. Science And Scientific Computing. Academic Press, Boston (1990)
36. Stewart, G.W.: Error and perturbation bounds for subspaces associated with certain eigenvalue problems. SIAM Rev. **15**, 727–764 (1973)
37. Stewart, G.W.: A generalization of Saad's theorem on Rayleigh–Ritz approximations. Linear Algebra Appl. **327**(1–3), 115–119 (2001)
38. Stewart, G.W.: Matrix Algorithms, vol. II. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (Eigensystems) (2001)

39. Watkins, D.S.: Understanding the QR algorithm. SIAM Rev. **24**(4), 427–440 (1982)
40. Wielandt, H.: Beiträge zur mathematischen behandlung komplexer eigenwertprobleme, Teil V: Bestimmung höherer eigenwerte durch gebrochene iteration. Bericht B 44/J/37, Aerodynamische Versuchsanstalt, Göttingen, Germany (1944)
41. Wilkinson, J.H.: The Algebraic Eigenvalue Problem. Clarendon Press, Oxford (1965)