Two-sided Grassmann-Rayleigh quotient iteration[a]

P.-A. Absil[bd] and P. Van Dooren[c]

Revised version, submitted on April 23, 2003
(First version submitted to SIMAX on October 11, 2002)

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 symmetric eigenproblem, the Hamiltonian eigenproblem and the skew-Hamiltonian eigenproblem.

Keywords. Block Rayleigh quotient iteration, two-sided iteration, Grassmann manifold, generalized eigenproblem, Hamiltonian eigenproblem.

AMS subject classification. 65F15

1 Introduction

The Rayleigh quotient iteration (RQI) is a classical method for computing eigenvectors of a symmetric matrix $A = A^T$ [Par74, Par98]. It is a special kind of inverse iteration [Ips97] where the shift is the Rayleigh quotient evaluated at the current iterate. The Rayleigh quotient is an efficient shift because in a neighbourhood of any eigenvector of A it yields a quadratic approximation of the corresponding eigenvalue. This remarkable property endows the RQI with cubic rate of convergence to the eigenvectors of A (see [Par74, Par98] or the sketch of proof in [AMSV02]). Thanks to its high rate of convergence, the RQI is particularly efficient for refining estimates of eigenvectors.

In some cases, one has to refine an estimate of a p-dimensional invariant subspace (also called *eigenspace*) of A. A reason for considering an eigenspace instead of individual eigenvectors may be that the eigenvectors themselves are ill-conditioned while the subspace is not (see e.g. [Ste73]) or just because they are not relevant for the application. Several methods have been proposed for refining invariant subspace estimates. A quadratically convergent iteration for refining eigenspaces of arbitrary matrices was proposed by Chatelin [Cha84, Dem87]. It

[a]This paper presents research results of the Belgian Programme on Interuniversity Attraction Poles, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with its authors.

[b]Department of Electrical Engineering and Computer Science, Université de Liège, Bât. B28 Systèmes, Grande Traverse 10, B-4000 Liège, Belgium (www.montefiore.ulg.ac.be/systems).

[c]Department of Mathematical Engineering, Université Catholique de Louvain, Bâtiment Euler (A.119), Avenue Georges Lemaître 4, B-1348 Louvain-la-Neuve, Belgium (vdooren@anma.ucl.ac.be).

[d]Aspirant du FNRS (Research Fellow with the Belgian National Fund for Scientific Research).

uses Newton's method for solving the Riccati equation obtained by expressing the eigenproblem in inhomogeneous coordinates. Edelman, Arias and Smith [EAS98] derived a coordinate-free Newton method on the set of the p-dimensional subspaces of $\mathbb{R}^n$ with a generalized Rayleigh quotient as cost function. This Newton method, closely related to Chatelin's iteration, converges cubically to the eigenspaces of a symmetric matrix A (see also [LE02, AMS02]). The method can also be used to compute eigenspaces of arbitrary matrices with a rate of convergence that is (at least) quadratic [AMS02]. In [AMSV02], it was shown that the RQI, originally defined on the set of one-dimensional subspaces of $\mathbb{R}^n$, can be generalized to operate on the set of p-dimensional subspaces of $\mathbb{R}^n$. The generalized iteration, called Grassmann-RQI (because the set of the p-dimensional subspaces of $\mathbb{R}^n$ is termed a *Grassmann manifold*), or GRQI for short, converges locally cubically to the p-dimensional eigenspaces of $A = A^T$.

It is natural to ask whether the Grassmann-RQI method can be adapted to deal with nonsymmetric matrices. This is the topic of the present paper.

The key idea comes from Ostrowski's series of papers dedicated to the RQI [Ost59a]. Let C be a nonsymmetric matrix (more precisely, let C be nonnormal). Then the quadratic approximation property of the Rayleigh quotient is lost and the RQI behaves badly [BS90]. This drawback was avoided by Ostrowski [Ost59b, Par74] by considering the bilateral Rayleigh quotient $\rho(y_L, y_R) := y_L^T C y_R / y_L^T y_R$ which displays the quadratic property in the neighbourhood of the pairs of left-right nondefective eigenvectors of C. Using this Rayleigh quotient as a shift, he derived a two-sided iteration (see Algorithm 2.5 below) that operates on pairs of vectors (or pairs of one-dimensional subspaces, because the norm is irrelevant) and aims at converging to pairs of left-right eigenvectors of C. The rate of convergence is in general cubic.

In the present paper, we generalize Ostrowski's two-sided RQI to operate on pairs of p-dimensional subspaces (instead of one-dimensional subspaces in the original iteration). The new iteration, called *Two-Sided Grassmann-RQI* (*2sGRQI*), converges locally cubically to the pairs of left-right p-dimensional eigenspaces of C (see Section 4).

The iteration involves solving a pair of Sylvester equations of the form $AXB - CXD = E$ in the $n \times p$ unknown X. Such linear equations in X can be efficiently solved using orthogonal transformations (Section 5). We report on numerical experiments in Section 8.

Comparison between Chatelin's iteration and the 2sGRQI (Section 6) shows that each method has its advantages and drawbacks. Main advantages of the 2sGRQI over Chatelin's iteration are the higher rate of convergence, the simultaneous computation of left and right eigenspaces, and the simpler structure of the Sylvester equations. On the other hand, the 2sGRQI behaves badly when C is defective, involves two Sylvester equations instead of one, and little is known about its global behaviour.

We also show that in some structured eigenproblems, namely E-(skew-)symmetric matrices with $E = \pm E^T$, a relation $\mathcal{Y}_L = E\mathcal{Y}_R$ between left and right subspaces is invariant by the 2sGRQI mapping (Section 7). This e.g. yields a modified one-sided Grassmann-RQI for the Hamiltonian eigenproblem. For normal nonsymmetric eigenproblems, we show that the unmodified one-sided GRQI converges also cubically.

2 Preliminaries

The goal of this section is to introduce some definitions, concepts and notations.

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 [Wil65]. Notions of subspaces and distance between them can be found in [Ste73].

Unless otherwise stated, all scalars, vectors and matrices are real (the analysis of the complex case is simpler). The superscript ^{T} denotes the transpose. In accordance with Parlett's conventions [Par74, Par98], the letter A always stands for a symmetric matrix while C denotes an arbitrary matrix. Following the conventions in [HM94], we use $\mathrm{Grass}(p,n)$ to denote the Grassmann manifold of the p-dimensional subspaces of $\mathbb{R}^n$, $\mathbb{RP}^{n-1}$ to denote the real projective space (i.e. the set of all one-dimensional subspaces of $\mathbb{R}^n$), and $\mathrm{ST}(p,n)$ 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 $Y \in \mathrm{ST}(p,n)$ is denoted by $\lfloor Y \rfloor$ and called the *span* of Y. The notation GL_p stands for the set of invertible $p \times p$ matrices. We now briefly recall some well-known facts about invariant subspaces.

Definition 2.1 *Let $\mathcal{X}$ be a p-dimensional subspace of $\mathbb{R}^n$ and let $X = [X_1, X_2]$ be an orthogonal $n \times n$ matrix such that X_1 spans $\mathcal{X}$. Then X^TCX may be partitioned in the form $X^TCX = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$ where $C_{11} \in \mathbb{R}^{p\times p}$. The subspace $\mathcal{X}$ is an* invariant subspace *(or* eigenspace*) 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}. We say that $\mathcal{X}$ is a* nondefective *invariant subspace of C if C_{11} is nondefective. The invariant subspace $\mathcal{X}$ will be termed* spectral *[RR02] if C_{11} and C_{22} have no eigenvalue in common. The eigenspaces of C^T are called* left eigenspaces *of C. We say that $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a* pair of spectral left-right eigenspaces *of C if $\mathcal{Y}_L$ and $\mathcal{Y}_R$ are spectral left and right eigenspaces of C with the same spectrum.*

The span of $Y \in \mathrm{ST}(p,n)$ is an eigenspace of C if and only if there exists a matrix M such that $CY = YM$. Each spectral invariant subspace is *isolated*, i.e. there exists a ball in $\mathrm{Grass}(p,n)$ centered on $\mathcal{V}$ that does not contain any eigenspace of C other than $\mathcal{V}$. We will also need the following result.

Proposition 2.2 *If $(\mathcal{Y}_L, \mathcal{Y}_R)$ is a pair of spectral 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^{-T} span $\mathcal{Y}_L$, and $S^{-1}CS = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$ with $D_1 \in \mathbb{R}^{p\times p}$.*

Proof. Let $(\mathcal{Y}_L, \mathcal{Y}_R)$ be a pair of spectral left-right eigenspaces of C. Then there exists X orthogonal such that $X^TCX = \begin{bmatrix} C_{11} & C_{12} \\ 0 & C_{22} \end{bmatrix}$, the first p columns of X span $\mathcal{Y}_R$, and C_{11} and C_{22} have no eigenvalue in common. Therefore, there exists a matrix L such that $C_{11}L - LC_{22} = -C_{12}$ [Ste73]. Let $S = X \begin{bmatrix} I & L \\ 0 & I \end{bmatrix}$. Then the first p columns of S span $\mathcal{Y}_R$. One easily checks that $S^{-1}CS = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix}$. Moreover $C^TS^{-T} = S^{-T} \begin{bmatrix} C_{11}^T & 0 \\ 0 & C_{22}^T \end{bmatrix}$. Therefore the first p columns of S^{-T} span an eigenspace of C^T whose spectrum is the same as the one of $\mathcal{Y}_R$. That is, the first p columns of S^{-T} span $\mathcal{Y}_L$. □

The Rayleigh quotient iteration (RQI) is a classical method for computing a single eigenvector of a symmetric matrix A [Par74, Par98]. It induces an iteration on the real projective space $\mathbb{RP}^{n-1}$ that can be written as follows.

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

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

It is shown in [BS89] that around each (isolated) eigenvector of A, there is a ball on 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 any initial point outside a set of measure zero (the result is actually stronger, see [BS89]).

The *Grassmann-Rayleigh Quotient Iteration* or GRQI [AMSV02] is a generalization of the RQI that operates on $\mathrm{Grass}(p, n)$.

Algorithm 2.4 (GRQI) *Let $A = A^T$ 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 $\mathbb{R}^n$ 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 $Y \in \mathrm{ST}(p, n)$ such that $\lfloor Y \rfloor = \mathcal{Y}$.*
2. *Solve the Sylvester equation*
$$\boldsymbol{T}_Y Z := AZ - Z(Y^T Y)^{-1} Y^T AY = Y \tag{2}$$
for $Z \in \mathbb{R}^{n \times p}$.
3. *Define $\mathcal{Y}_+ := \lfloor Z \rfloor$.*

It is shown in [AMSV02] 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 (attractive) fixed points.

When the matrix A is not normal (for the normal nonsymmetric case, see Section 7.3), the stationary property of the Rayleigh quotient fails. Consequently, the convergence rate of the RQI can be at best quadratic. In order to recover cubic convergence, Ostrowski [Ost59b] proposed a two-sided version of the RQI, formulated as follows in [Par74].

Algorithm 2.5 (Two-Sided RQI) *Let C be an n-by-n matrix. Pick initial vectors v_0^T and u_0 satisfying $v_0^T u_0 \neq 0$, $\|v_0\| = \|u_0\| = 1$. For $k = 0, 1, 2, \ldots$,*
1. Compute $\rho_k = v_k^T C u_k / v_k^T u_k$.
2. If $C - \rho_k I$ is singular solve $y^T(C - \rho_k I) = 0$ and $(C - \rho_k I)x = 0$ for $y^T, x \neq 0$ and stop, otherwise
3. Solve both $v_{k+1}^T(C - \rho_k I) = v_k^T \nu_k$, $(C - \rho_k I)u_{k+1} = u_k \tau_k$, where ν_k and τ_k are normalizing factors.
4. If $v_{k+1}^T u_{k+1} = 0$, 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.

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-by-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 $\mathbb{R}^n$ 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 $\mathrm{ST}(p,n)$ such that $\lfloor Y_L \rfloor = \mathcal{Y}_L$ and $\lfloor Y_R \rfloor = \mathcal{Y}_R$.
2. Solve the Sylvester equations

$$CZ_R - Z_R \underbrace{(Y_L^T Y_R)^{-1} Y_L^T C Y_R}_{R_R} = Y_R \tag{3a}$$

$$Z_L^T C - \underbrace{Y_L^T C Y_R (Y_L^T Y_R)^{-1}}_{R_L} Z_L^T = Y_L^T \tag{3b}$$

for Z_L and Z_R in $\mathbb{R}^{n \times p}$.
3. Define $\mathcal{Y}_{L+} := \lfloor Z_L \rfloor$ and $\mathcal{Y}_{R+} := \lfloor Z_R \rfloor$.

Before studying the convergence properties of the algorithm, a few comments are in order.

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 $\{YM : M \in \mathrm{GL}_p\}$ is the (infinite) set of all other bases of 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 [AMSV02] for the GRQI algorithm, if (Y_L, Y_R, Z_L, Z_R) solve (3) then $(Y_L M, Y_R N, Z_L M, Z_R N)$ also solve (3) for all M, N in GL_p. Therefore, 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 $Y_L^T Y_R$ 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 spectral left-right eigenspaces of C, then $Y_L^T Y_R$ is invertible (as a consequence of Proposition 2.2), and by continuity invertibility holds on a neighbourhood of the pair of eigenspaces.

In point 2, equations (3) –uncoupled Sylvester equations– do not always admit a unique solution. They do not if and only if (Y_R, Y_L) belongs to the set

$$\begin{aligned}\mathcal{S} &:= \{(Y_L, Y_R) \in \mathrm{ST}(p,n) \times \mathrm{ST}(p,n) : R_R \text{ exists and } \sigma(C) \cap \sigma(R_R) \neq \emptyset\} \\ &= \bigcup_{\lambda \in \sigma(C)} \{(Y_L, Y_R) \in \mathrm{ST}(p,n) \times \mathrm{ST}(p,n) : R_R \text{ exists and } \det(R_R - \lambda I) = 0\}.\end{aligned}$$

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 neighbourhood of this pair on which the solution (Z_L, Z_R) of (3) exists and is unique. We did not detail this in the above Algorithms 2.4 and 3.1, but when such a singularity occurs, we slightly perturb Y_R and Y_L. A justification for this is given in [AMSV02] and the numerical tests performed in Section 8 illustrate that this works well in practice. In most cases, singularities are naturally prevented by numerical errors.

In Section 5 we propose a numerical algorithm for solving the Sylvester equations (3). Until then, we assume that computations are performed in exact arithmetic.

In point 3, if Z_L or Z_R is not full rank, then $(\mathcal{Y}_{L+}, \mathcal{Y}_{R+})$ does not belong to $\mathrm{Grass}(p,n) \times \mathrm{Grass}(p,n)$. 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

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

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 do not improve any more after a few (typically two or three) iteration steps because of numerical errors (see numerical experiments in Section 8). 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 $A^T\mathcal{Y}_L$.

4 Local convergence

This section is dedicated to proving that the cubic convergence property of Ostrowski's two-sided iteration (Algorithm 2.5) extends to its Grassmannian generalization (Algorithm 3.1).

Theorem 4.1 (cubic convergence of 2sGRQI) *Let C be an n-by-n matrix. Let $(\mathcal{V}_L, \mathcal{V}_R)$ be a pair of p-dimensional spectral nondefective left-right eigenspaces of C (see Definition 2.1). Then the 2sGRQI mapping (Algorithm 3.1) admits a continuous extension on a neighbourhood of $(\mathcal{V}_L, \mathcal{V}_R)$. The point $(\mathcal{V}_L, \mathcal{V}_R)$ is an attractive fixed point of the extended mapping, and the rate of convergence is cubic.*

Cubic convergence of Algorithm 3.1 means, among others, cubic convergence to zero of the scalar $\mathrm{dist}(\mathcal{Y}_L, \mathcal{V}_L) + \mathrm{dist}(\mathcal{Y}_R, \mathcal{V}_R)$, where "dist" denotes the largest principal angle between its arguments. In other words, given C and $(\mathcal{V}_L, \mathcal{V}_R)$, there exist $\epsilon > 0$ and $c > 0$ such that, for all $(\mathcal{Y}_L, \mathcal{Y}_R)$ with $\mathrm{dist}(\mathcal{Y}_{L+}, \mathcal{V}_{L+}) + \mathrm{dist}(\mathcal{Y}_{R+}, \mathcal{V}_{R+}) < \epsilon$, one has $\mathrm{dist}(\mathcal{Y}_{L+}, \mathcal{V}_{L+}) + \mathrm{dist}(\mathcal{Y}_{R+}, \mathcal{V}_{R+}) < c\,(\mathrm{dist}(\mathcal{Y}_L, \mathcal{V}_L) + \mathrm{dist}(\mathcal{Y}_R, \mathcal{V}_R))^3$. Other compatible definitions of dist can be found in [EAS98].

In order to prove Theorem 4.1, we will need the following modification of a technical lemma in [AMSV02].

Lemma 4.2 *Let A be a nondefective $p \times p$ matrix. Let $B_\epsilon = \{E \in \mathbb{R}^{p\times p} : \|E\| \leq \epsilon\}$. Let $\Gamma = \{E \in \mathbb{R}^{p\times p} : \sigma(A) \cap \sigma(A+E) = \emptyset\}$. Then*

1. *Γ is an open dense subset of $\mathbb{R}^{p\times p}$.*
2. *The equation*
$$AX - X(A+E) = I \qquad (4)$$
admits a unique solution if and only if $E \in \Gamma$.
3. *There exists $\epsilon > 0$ such that, for all $E \in \Gamma \cap B_\epsilon$, the solution X of (4) is invertible. The function $\Gamma \cap B_\epsilon \ni E \mapsto X^{-1}$ is continuous and admits a continuous extension on B_ϵ. We will also use the notation X^{-1} to denote the continuous extension.*

4. There exist $\epsilon > 0$ and $c > 0$ such that, for all $E \in B_\epsilon$,

$$\|X^{-1}\| \leq c\|E\|. \tag{5}$$

5. If A is normal, then (5) holds with $c = 1 + \epsilon$.

Here are a few comments before we proceed with the proof of Lemma 4.2.

In [AMSV02], the corresponding technical lemma only deals with A symmetric. The continuous extension of X^{-1} is not mentioned in the statement, although it is implied in the proof.

The condition A nondefective cannot be removed. Moreover, there is no universal value for the constant c in (5). We show these two claims on an example. Take

$$A = \begin{bmatrix} 0 & 1 \\ 0 & \delta \end{bmatrix}, \; E = \begin{bmatrix} \eta & 0 \\ 0 & \frac{\eta}{2} \end{bmatrix}.$$

Then (4) yields

$$X = \begin{bmatrix} -\frac{1}{\eta} & \frac{1}{\eta(\eta+\delta)} \\ 0 & -\frac{2}{\eta} \end{bmatrix} \text{ and } X^{-1} = \begin{bmatrix} -\eta & -\frac{\eta}{2(\delta+\eta)} \\ 0 & -\frac{\eta}{2} \end{bmatrix}. \tag{6}$$

Let $\delta = 0$. Then A is defective, and one observes that X^{-1} is bounded away from zero. For δ fixed and nonzero, one has $\|X^{-1}\| \simeq \frac{1}{\delta}O(\eta)$, so c becomes arbitrarily large for δ small.

Proof. (Lemma 4.2) Point 1. The complement of Γ is $\bigcup_{\lambda\in\sigma(A)}\{E : \det(A+E-\lambda I) = 0\}$, i.e. a finite union of algebraic sets. Therefore Γ is open and dense.
Point 2. See [Ste73, §4.3].
Points 3 and 4. The proof of Lemma 5.3 in [AMSV02] implies that points 3 and 4 are valid for A (complex) diagonal. Here A is any nondefective matrix, so there exists a complex invertible matrix S and a complex diagonal matrix D such that $A = SDS^{-1}$. Rewrite (4) as $D\tilde{X} - \tilde{X}(D+\tilde{E}) = I$, where $\tilde{E} := S^{-1}ES$ and $\tilde{X} := S^{-1}XS$ whence $\tilde{X}^{-1} := S^{-1}X^{-1}S$. The relation between the tilded and original quantities are linear invertible maps. Therefore, the results 3 and 4 also hold for any A nondefective. In particular, one has $\|X^{-1}\| \leq c_1\|\tilde{X}^{-1}\|$, $\|\tilde{X}^{-1}\| \leq c_2\|\tilde{E}\|$ and $\|\tilde{E}\| \leq c_3\|E\|$, so $\|X^{-1}\| \leq c\|E\|$.
Point 5. If A is normal, then S can be chosen unitary. Therefore the inequalities above hold with $c_1 = c_3 = 1$. Moreover, Lemma 5.3 in [AMSV02] shows that $\|\tilde{X}^{-1}\| \leq (1+\epsilon)\|\tilde{E}\|$ holds for all $\tilde{E}$ in B_ϵ, i.e. E in B_ϵ. □

Proof. (Theorem 4.1) The proof is inspired by the proof of cubic convergence of the Two-Sided RQI in [Par74] and the proof of cubic convergence of the GRQI in [AMSV02]. By Proposition 2.2, since $(\mathcal{V}_L, \mathcal{V}_R)$ is a pair of spectral nondefective left-right eigenspaces of C, there is an invertible (real) matrix S such that $\mathcal{V}_R$ is spanned by the first p columns of S, $\mathcal{V}_L$ is spanned by the first p columns of S^{-T} and $S^{-1}CS =: D = \begin{bmatrix} D_1 & 0 \\ 0 & D_2 \end{bmatrix}$ where D_1 is nondefective (so one can even choose it quasi-diagonal) and has no common eigenvalue with D_2.

Let

$$Y_R = S\begin{pmatrix} I_p \\ K_R \end{pmatrix}, Y_L = S^{-T}\begin{pmatrix} I_p \\ K_L \end{pmatrix}$$

so that $\lfloor Y_R \rfloor$, $\lfloor Y_L \rfloor$ are arbitrary perturbations of $\mathcal{V}_R$, $\mathcal{V}_L$. The 2sGRQI mapping sends $(\lfloor Y_L \rfloor, \lfloor Y_R \rfloor)$ to $(\lfloor Z_L \rfloor, \lfloor Z_R \rfloor)$ where Z_R and Z_L satisfy the Sylvester equations (3) repeated here

$$CZ_R - Z_R(Y_L^TY_R)^{-1}Y_L^TCY_R = Y_R \tag{7a}$$
$$Z_L^TC - Y_L^TCY_R(Y_L^TY_R)^{-1}Z_L^T = Y_L^T. \tag{7b}$$

Put

$$Z_R = S\begin{pmatrix} Z_{R1} \\ Z_{R2} \end{pmatrix},\ Z_L = S^{-T}\begin{pmatrix} Z_{L1} \\ Z_{L2} \end{pmatrix},\ K_{R+} = Z_{R2}Z_{R1}^{-1},\ K_{L+} = Z_{L2}Z_{L1}^{-1},$$

so that

$$\lfloor Z_R \rfloor = \left\lfloor S\begin{pmatrix} I_p \\ K_{R+} \end{pmatrix} \right\rfloor, \lfloor Z_L \rfloor = \left\lfloor S^{-T}\begin{pmatrix} I_p \\ K_{L+} \end{pmatrix} \right\rfloor.$$

Premultiplying (7a) by S^{-1} yields the two uncoupled equations

$$D_1Z_{R1} - Z_{R1}(I + K_L^TK_R)^{-1}(D_1 + K_L^TD_2K_R) = I_p \tag{8}$$
$$D_2Z_{R2} - Z_{R2}(I + K_L^TK_R)^{-1}(D_1 + K_L^TD_2K_R) = K_R. \tag{9}$$

From (9), one has $\|Z_{R2}\| \le \|\boldsymbol{T}_{K_L,K_R}^{-1}\|\,\|K_R\|$ where

$$\boldsymbol{T}_{K_L,K_R} : Z \mapsto D_2Z - Z(I + K_L^TK_R)^{-1}(D_1 + K_L^TD_2K_R)$$

is a linear operator depending continuously on K_L, K_R. $\boldsymbol{T}_{0,0}$ is invertible because D_1 and D_2 have no eigenvalue in common. Therefore, there exists $c_2 > 0$ such that $\|\boldsymbol{T}_{K_L,K_R}^{-1}\| \le c_2$ for all K_L, K_R sufficiently small, whence

$$\|Z_{R2}\| \le c_2\,\|K_R\| \tag{10}$$

for all K_L, K_R sufficiently small.

From (8), by Lemma 4.2 there exists $c_1 > 0$ such that

$$\|Z_{R1}^{-1}\| \le c_1\,\|(I + K_L^TK_R)^{-1}(D_1 + K_L^TD_2K_R) - D_1\| \le c_1'\|K_L\|\,\|K_R\| \tag{11}$$

for all K_L, K_R sufficiently small, where Z_{R1}^{-1} is defined by continuous extension as stated in Lemma 4.2.

So (8)-(9) define a continuous function $(K_L, K_R) \mapsto K_{R+} = Z_{R2}Z_{R1}^{-1}$ everywhere in a neighbourhood of $(0, 0)$, and from (10) and (11), one has

$$\|K_{R+}\| \le c_3\,\|K_L\|\,\|K_R\|^2. \tag{12}$$

Applying a similar reasoning to (7b), we obtain

$$\|K_{L+}\| \le c_4\,\|K_L\|^2\,\|K_R\|. \tag{13}$$

From (12)-(13), one has

$$\|K_{L+}\| + \|K_{R+}\| \le c\,\|K_L\|\,\|K_R\|(\|K_L\| + \|K_R\|) \le c(\|K_L\| + \|K_R\|)^3 \tag{14}$$

and this is cubic convergence. □

Intuitively, the right-hand sides in the 2sGRQI equations (3) are responsible via equation (9) for the linear convergence that would be observed with a fixed shift. Cubic convergence is reached via equation (8) thanks to the particular adaptive shifts R_R and R_L utilized in the 2sGRQI equations (3).

5 Practical implementation

The major computational task in both GRQI (Algorithm 2.4) and 2sGRQI (Algorithm 3.1) is to solve the Sylvester equations. For GRQI, it is recommended to choose Y orthonormal (i.e. $Y^TY = I_p$) such that Y^TAY is diagonal. This involves a p-dimensional eigenproblem, which is cheap when p is small. Then (2) decouples into p linear systems for the p columns of Z. See [AMSV02] for details.

The case of 2sGRQI (Algorithm 3.1) is quite different. The matrices R_R and R_L in the 2sGRQI equations (3) are not symmetric and they might not be diagonalizable. A possible approach to solving an equation like (3a) is to reduce it to a special triangular structure by means of orthogonal transformations, as we now explain. Firstly, (3a) can be rewritten as

$$C\tilde{Z}_R(Y_L^TY_R) - \tilde{Z}_RY_L^TCY_R = Y_R \tag{15}$$

where $\tilde{Z}_R := Z_R(Y_L^TY_R)$. Since $\tilde{Z}_R$ and Z_R have the same span, equation (15) can be solved instead of (3a). The same reasoning can be applied to the left equation (3b). Therefore Algorithm 3.1 is left unmodified if (3) is replaced by

$$CZ_R\,(Y_L^TY_R) - Z_R\,(Y_L^TCY_R) = Y_R \tag{16a}$$

$$C^TZ_L\,(Y_R^TY_L) - Z_L\,(Y_R^TC^TY_L) = Y_L. \tag{16b}$$

Let us focus for the moment on the first equation, (16a), rewritten as

$$CZB^T - ZD^T = E \tag{17}$$

for simplicity. Rewrite (17) as

$$\underbrace{(Q_1CU_1)}_{P}\,\underbrace{(U_1^TZU_2)}_{Y}\,\underbrace{(U_2^TB^TQ_2^T)}_{R^T} - \underbrace{(Q_1U_1)}_{S}\,\underbrace{(U_1^TZU_2)}_{Y}\,\underbrace{(U_2^TD^TQ_2^T)}_{T^T} = \underbrace{Q_1EQ_2^T}_{F}$$

in which Q_1, U_1, Q_2, U_2 are orthogonal matrices. The QZ algorithm of Moler and Stewart [MS73] can be applied on the pair (D, B) so that T is quasi-upper-triangular and R is upper triangular. This is a cheap problem here because both B and D are small p-by-p matrices. Due to its particular triangular structure, the system

$$PYR^T - SYT^T = F$$

can be solved by back substitution techniques. Denoting by a_k the kth column of A, one has to solve $2n \times 2n$ block structured systems like

$$\begin{pmatrix} r_{k-1,k-1}P - t_{k-1,k-1}S & r_{k-1,k}P - t_{k-1,k}S \\ -t_{k,k-1}S & r_{kk}P - t_{kk}S \end{pmatrix} \begin{pmatrix} y_{k-1} \\ y_k \end{pmatrix} = \begin{pmatrix} f_{k-1}^{(k)} \\ f_k^{(k)} \end{pmatrix}$$

where $f_k^{(j)}$ is defined by the recursion $f_k^{(p)} = f_k$, $f_k^{(j-1)} = f_k^{(j)} - r_{kj}(Py_j) + t_{kj}(Sy_j)$. Moreover, one can take advantage of the degrees of freedom in Q_1 and U_1 to make these systems easier to solve. For more details, see [GLAM92].

The same kind of argument applies to the left equation (16b). Note that the QZ transformations need not be computed a second time.

The generalized eigenvalue problem

Let $A, B \in \mathbb{R}^{n\times n}$. 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. [Ste73, GV96]. 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

$$A\mathcal{X}, B\mathcal{X} \ \subset \ \mathcal{Y}. \tag{18}$$

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 (18) 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 (16), one obtains after some manipulations

$$\hat{A}Z_R\hat{Y}_L^T\hat{B}Y_R - \hat{B}Z_R\hat{Y}_L^T\hat{A}Y_R = \hat{B}Y_R \tag{19a}$$

$$\hat{A}^T\hat{Z}_LY_R^T\hat{B}^T\hat{Y}_L - \hat{B}^T\hat{Z}_LY_R^T\hat{A}^T\hat{Y}_L = \hat{B}^T\hat{Y}_L \tag{19b}$$

where $\hat{Y}_L := \hat{B}^{-T}Y_L$ and $\hat{Z}_L := \hat{B}^{-T}Z_L$. 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 Comparisons with other methods

In [Cha84], Chatelin proposes a Newton method for refining eigenspace estimates. The reasoning can be explained as follows. An $n \times p$ matrix Y spans an eigenspace of C if and only if there exists a $p \times p$ matrix M such that

$$CY = YM. \tag{20}$$

However, any subspace admits infinitely many bases, and the solutions Y of (20) are not isolated. A way to remove the freedom in the choice of basis is to impose a normalization condition $W^TY = I$ where W is a given full-rank $n \times p$ matrix. Then (20) becomes

$$F(Y) := CY - Y(W^TCY) = 0 \tag{21}$$

where the unknown Y is normalized by $W^TY = I$. The Newton iteration for solving (21) is given by

$$(I - YW^T)C\Delta - \Delta(W^TCY) = -F(Y),\ W^T\Delta = 0 \tag{22}$$

$$Y_+ := Y + \Delta. \tag{23}$$

Choosing the basis Y orthonormal and $W := Y$, (22) becomes

$$\Pi C\Pi\Delta - \Delta(Y^TCY) = -\Pi CY,\ Y^T\Delta = 0 \tag{24}$$

where $\Pi := I - YY^T$. Equation (24), with update performed along the geodesic in direction Δ instead of (23), admits an interpretation as a Riemannian Newton method [Smi94] on the Grassmann manifold [EAS98, AMS02]. The rate of convergence is quadratic in general (cubic when C is symmetric).

In [Dem87], Demmel makes $\Delta = Y_\perp K$, where $Y_\perp$ is an orthonormal matrix with $Y^T Y_\perp = 0$ and K is an $(n-p) \times p$ matrix. Then $Y^T \Delta = 0$ is trivially satisfied and equation (24) becomes

$$(Y_\perp^T C Y_\perp) K - K(Y^T C Y) = -Y_\perp^T C Y, \tag{25}$$

i.e. a Sylvester equation without constraints on the unknown K. As was pointed out in [AMSV02], solving (25) takes $O(n^3)$ operations even when C is condensed (e.g. tridiagonal) because $Y_\perp^T C Y_\perp$ is a full big $(n-p) \times (n-p)$ matrix. However, Lundström and Eldén recently proposed an algorithm for solving (24) that does not require the computation of $Y_\perp^T C Y_\perp$ [LE02, alg. 2]. It takes $O(np^2)$ operations to solve (24) when C is block diagonal of sufficiently moderate block size and $O(n^2 p)$ when C is Hessenberg. The complexity of the Two-Sided GRQI (Algorithm 3.1) is of the same order.

Main advantages of 2sGRQI (Algorithm 3.1) over Chatelin's method:

- The convergence is cubic instead of quadratic.
- A pair of left-right eigenspaces is computed, and not only a right-eigenspace.
- Solving the 2sGRQI equations (16) is more straightforward than solving Chatelin's equation (22) where projectors appear and the solution is constrained. However, when (22) is solved using the algorithm described in [LE02], the numerical cost is of the same order for both methods.

Advantages of Chatelin's method:

- Only one Sylvester equation –instead of two in 2sGRQI– has to be solved to compute a new iterate.
- The convergence analysis also applies when the targeted eigenspace is defective [Dem87, AMS02].

7 Structured eigenproblems

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

7.1 E-symmetric eigenproblem

Let C be an $n \times n$ matrix. Since C and C^T have the same spectrum, they are similar, i.e. there exists an invertible matrix E such that

$$EC = C^T E, \tag{26}$$

and we say that C is *E-symmetric.* If C is E-symmetric, then its left and right eigenspaces are related by a multiplication by 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 (26), one obtains $C^T ES = ESD$, to be compared with $CS = SD$.

The case where E is symmetric or skew-symmetric, i.e. $E^T = \pm E$, is of particular interest because, as we show in the next proposition, the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant under the 2sGRQI (Algorithm 3.1). Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, it is sufficient to solve (3a) only, and then compute $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, since the pairs of left-right eigenspaces of C also satisfy $\mathcal{V}_L = E\mathcal{V}_R$, Theorem 4.1 also applies.

Proposition 7.1 *Let E be invertible with $E^T = \pm E$ and let C be E-symmetric, i.e. $EC = C^T E$. If $Y_L = EY$ and $Y_R = Y$, then $Z_L = EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations* (16) *with*

$$\boxed{CZ\,(Y^T EY) - Z\,(Y^T ECY) = Y.} \tag{27}$$

Therefore, 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 (27) *converges locally cubically to the spectral nondefective 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 (16a) and (16b) yields (27) 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 4.1. □

Efficient numerical methods are available for solving (27). See Section 5 and [GLAM92, GV96].

Generalized symmetric eigenproblem

Using Proposition 7.1, we show that the 2-sided GRQI yields a Grassmannian RQI for the symmetric 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 symmetric 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 $V \in \mathbb{R}^{n\times p}$ 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 7.1 are satisfied with $E := B$. The modified GRQI equation (27) becomes

$$\boxed{AZ\,(Y^T BY) - BZ\,(Y^T AY) = BY} \tag{28}$$

and local cubic convergence holds to the spectral nondefective eigenspaces of $B^{-1}A$. In particular, $B^{-1}A$ is nondefective when A or B is positive definite.

Skew-Hamiltonian eigenproblem

Let T be a skew-Hamiltonian matrix, i.e. $(TJ)^T = -TJ$, where $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$, see e.g. [BBMX02]. Equivalently, $JT = T^T J$, i.e. T is J-symmetric. Conditions in Proposition 7.1 are satisfied with $C := T$ and $E := J$. The modified GRQI equation (27) becomes

$$\boxed{TZ\,(Y^T JY) - Z\,(Y^T JTY) = Y} \tag{29}$$

and local cubic convergence holds to the spectral nondefective right-eigenspaces of T.

7.2 E-skew-symmetric eigenproblem

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

$$EC = -C^T E. \tag{30}$$

We saw in the previous section that the corresponding left and right eigenspaces of E-symmetric matrices are related by a multiplication by E. The case of E-skew-symmetric matrices is slightly different. The eigenvalues appear in pairs of opposite numbers, and the left and right eigenspaces with opposite eigenvalues are related by multiplication by E, as we now show. Let S be an invertible matrix of principal vectors of C, i.e.

$$CS = SD \tag{31}$$

where D is a Jordan matrix. Then from (30), one obtains

$$C^T ES = ES(-D). \tag{32}$$

The matrix $-D$ is therefore a Jordan matrix of C^T. Therefore, if λ is an eigenvalue of C then $-\lambda$ is an eigenvalue of C^T, thus of C. Since C is real, $\bar{\lambda}$ and $-\bar{\lambda}$ are eigenvalues of C, too. Moreover, equations (31) and (32) 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 $-\lambda_{i_1}, \ldots, -\lambda_{i_p}$. So, $(E\mathcal{V}, \mathcal{V})$ forms a pair of spectral left-right eigenspaces of C if and only if $\mathcal{V}$ is a spectral right-eigenspace and the spectrum of $\mathcal{V}$ is symmetric with respect to zero. We call such a $\mathcal{V}$ a *full eigenspace*.

As was the case for the E-symmetric matrices, if E is symmetric or skew-symmetric, i.e. $E^T = \pm E$ then the relation $\mathcal{Y}_L = E\mathcal{Y}_R$ is invariant by the 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 (3a) only, and then compute $\mathcal{Y}_{L+} := E\mathcal{Y}_{R+}$. Moreover, the 2sGRQI iteration restricted to the pairs $(\mathcal{Y}_L, \mathcal{Y}_R) = (E\mathcal{Y}, \mathcal{Y})$ converges locally cubically to the full nondefective eigenspaces of C.

Proposition 7.2 *Let E be invertible with $E^T = \pm E$ and let C be E-skew-symmetric, i.e. $EC = -C^T E$. If $Y_L = EY$ and $Y_R = Y$, then $Z_L = -EZ$ and $Z_R = Z$ satisfy the 2sGRQI equations* (16) *with*

$$\boxed{CZ\,(Y^T EY) - Z\,(Y^T ECY) = Y.} \tag{33}$$

Therefore, if $\mathcal{Y}_L = E\mathcal{Y}_R$, then $\mathcal{Y}_{L+} = E\mathcal{Y}_{R+}$.
Moreover, let $\mathcal{V}$ be a full *nondefective 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 invariant by multiplication by -1 and $C|_{\mathcal{V}}$ is nondefective). Then the subspace iteration $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ defined by (27) *converges locally cubically to $\mathcal{V}$.*

This proposition differs from Proposition 7.1 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 (16a) and (16b) yields (33) 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 nondefective 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 4.1. □

Skew-symmetric eigenproblem

Let Ω be skew-symmetric. Then one has $EC = -C^T E$ with $C := \Omega$ and $E := I$. The modified GRQI equation (33) becomes

$$\boxed{\Omega Z\,(Y^T Y) - Z\,(Y^T \Omega Y) = Y.} \tag{34}$$

This is simply the classical GRQI equation (2). This is not surprising, since skew-symmetric matrices are normal (i.e. they commute with their transpose).

Hamiltonian eigenproblem

Let H be Hamiltonian, i.e. $(HJ)^T = HJ$, where $J = \left(\begin{smallmatrix} 0 & I \\ -I & 0 \end{smallmatrix}\right)$. This is equivalent to $JH = -H^T J$. Thus $EC = -C^T E$ with $C := H$ and $E := J$, and the modified GRQI equation (33) reads

$$\boxed{HZ\,(Y^T JY) - Z\,(Y^T JHY) = Y.} \tag{35}$$

Proposition 7.2 implies that the subspace iteration with iteration mapping $\lfloor Y \rfloor \mapsto \lfloor Z \rfloor$ defined by (35) converges locally cubically to the full nondefective right-eigenspaces of H.

7.3 Normal eigenproblem

The (real) matrix C is *normal* if $CC^T = C^T C$. A real matrix is normal if and only if it is quasi-diagonalizable by an orthogonal matrix. In particular, symmetric, skew-symmetric and orthogonal matrices are normal.

When C is symmetric, the relation $\mathcal{Y}_L = \mathcal{Y}_R$ is evidently invariant by the 2sGRQI mapping, since the two equations (3) are then identical. This property does not extend to normal nonsymmetric matrices. Nevertheless, if C is normal and $\mathcal{Y}_L = \mathcal{Y}_R$, then $\mathrm{dist}(\mathcal{Y}_{R+}, \mathcal{V}) = O(\mathrm{dist}^3(\mathcal{Y}_R, \mathcal{V}))$ for all spectral nondefective eigenspace $\mathcal{V}$ of C. To show this, note that in the proof of cubic convergence (Theorem 4.1), the matrix S can be taken orthogonal since C is normal. Therefore, $S^{-T} = S$, so $Y_L = Y_R$ implies $K_R = K_L$. The result follows from (12). Now making $Y_L = Y_R =: Y$ in (7) yields the GRQI equation (2). This proves that the cubic local convergence property of the GRQI (Algorithm 2.4) to the spectral (trivially nondefective) eigenspaces of A is also valid when A is normal.

8 Numerical experiments

The aim of this section is to illustrate via extensive numerical experiments that 2sGRQI (Algorithm 3.1) has solid potential as a numerical method. The Two-Sided GRQI has been implemented as follows.

Algorithm 8.1 (implementation of 2sGRQI) *Let C be an $n \times n$ matrix. Given two $n \times p$ matrices Y_{L_0} and Y_{R_0} satisfying $Y_{L_0}^T Y_{L_0} = I = Y_{R_0}^T Y_{R_0}$, the algorithm produces a sequence of matrices (Y_{L_k}, Y_{R_k}) as follows. For $k = 1, 2, \ldots$,*
1. Apply the real QZ algorithm on the pair $(Y_{R_k}^T C^T Y_{L_k}, Y_{R_k}^T Y_{L_k})$ to obtain the pair (T, R) where T is quasi-upper-triangular and R is upper-triangular. In Matlab, `[T,R,Q,U] = qz(YL'*C'*YL,YR'*YL,'real')`.
2. Solve the equations $CZ_R R^T - Z_R T^T = Y_R Q^T$ and $C^T Z_L R - Z_L T = Y_L U$ by successive substitution as explained in Section 5. The $n \times n$ and $2n \times 2n$ linear systems are solved in Matlab using the operator \.
3. Orthonormalize Z_L and Z_R to obtain $Y_{L_{k+1}}$ and $Y_{R_{k+1}}$. In Matlab, `[YL,dummy] = qr(ZL,0)` *and* `[YR,dummy] = qr(ZR,0)`.

It is well known [BS89] 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 [ASVM03] and obviously extends to 2sGRQI (Algorithm 3.1). Moreover, in 2sGRQI the matrix C is not necessarily symmetric; its eigenspaces can thus be arbitrarily close to each other. In the experiments reported below, we have ruled out clustered eigenvalues/vectors by taking $C = SDS^{-1}$ where D is a diagonal matrix whose diagonal elements are random permutations of $1, \ldots, n$ and $S = I + E$ where E is taken from normally distributed random variables and scaled such that $\|E\|_2$ is uniformly distributed on the interval $(0, 0.1)$. The initial matrices Y_{L_0} and Y_{R_0} are randomly chosen such that $\mathrm{dist}(\lfloor Y_{R_0} \rfloor, \lfloor S(:, 1:p) \rfloor) < 0.1$ and $\mathrm{dist}(\lfloor Y_{L_0} \rfloor, \lfloor S^{-T}(:, 1:p) \rfloor) < 0.1$, where "dist" is the largest principal angle. The behaviour of the 2sGRQI algorithm in case of ill-separated eigenvectors/values would need further investigation. The symmetric case is studied in [ASVM03] where improvements of GRQI and the Riemannian Newton algorithm are proposed.

Algorithm 8.1 was run 10^6 times with $n = 20$, $p = 5$, and C, Y_{L_0}, Y_{R_0} randomly chosen as explained above. Experiments were run using Matlab 6.0 with floating point relative accuracy approximately equal to $2 \cdot 10^{-16}$. Results are summarized in Table 1. The error e is defined as the largest principal angle between $\lfloor Y_R \rfloor$ and $\lfloor S(:, 1:p) \rfloor$ plus the largest principal angle between $\lfloor Y_L \rfloor$ and $\lfloor S^{-T}(:, 1:p) \rfloor$. Mean and max are computed over the 10^6 experiments.

Iterate number	mean(log10(e))	max(log10(e))
0	-1.0482	-0.6996
1	-4.2584	-2.4413
2	-13.5275	-7.9426
3	-16.2589	-11.0974
4	-16.2572	-11.5222
5	-16.2567	-11.3846

Table 1: Numerical experiments for Algorithm 8.1. See details in the text.

9 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 spectral nondefective left-right eigenspaces of arbitrary square matrices. The iteration is competitive with Chatelin's Newton method and it yields one-sided formulations adapted to some structured eigenproblems, including the Hamiltonian and generalized symmetric eigenproblems.

In this paper, all matrices used in the iterations were real even though we used complex matrices in some of the proofs. When going to complex arithmetic everything becomes much simpler. The QZ algorithm gives triangular matrices, the solution of the Sylvester equations are simpler recurrences but implying complex vectors and systems of equations.

Acknowledgements

Part of this work was performed while 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 PhD research and the impact this had on the results of this paper.

References

[AMS02] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, submitted to Acta Appl. Math., November 2002.

[AMSV02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Review **44** (2002), no. 1, 57–73.

[ASVM03] P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, submitted to SIAM J. Matrix Anal. Appl., January 2003.

[BBMX02] Peter Benner, Ralph Byers, Volker Mehrmann, and Hongguo Xu, *Numerical computation of deflating subspaces of skew-Hamiltonian/Hamiltonian pencils*, SIAM J. Matrix Anal. Appl. **24** (2002), no. 1, 165–190.

[BS89] S. Batterson and J. Smillie, *The dynamics of Rayleigh quotient iteration*, SIAM J. Numer. Anal. **26** (1989), no. 3, 624–636.

[BS90] ——, *Rayleigh quotient iteration for nonsymmetric matrices*, Math. Comp. **55** (1990), no. 191, 169–178.

[Cha84] F. Chatelin, *Simultaneous Newton's iteration for the eigenproblem*, Computing, Suppl. **5** (1984), 67–74.

[Dem87] J. W. Demmel, *Three methods for refining estimates of invariant subspaces*, Computing **38** (1987), 43–57.

[EAS98] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.

[GLAM92] J. D. Gardiner, A. J. Laub, J. J. Amato, and C. B. Moler, *Solution of the Sylvester matrix equation $AXB^T + CXD^T = E$*, ACM Trans. Math. Software **18** (1992), no. 2, 223–231.

[GV96] G. H. Golub and C. F. Van Loan, *Matrix computations, third edition*, The John Hopkins University Press, 1996.

[HM94] U. Helmke and J. B. Moore, *Optimization and dynamical systems*, Springer, 1994.

[Ips97] I. C. F. Ipsen, *Computing an eigenvector with inverse iteration*, SIAM Rev. **39** (1997), no. 2, 254–291.

[LE02] E. Lundström and L. Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2002), no. 3, 819–839.

[MS73] C. B. Moler and G. W. Stewart, *An algorithm for generalized matrix eigenvalue problems*, SIAM J. Numer. Anal. **10** (1973), 241–256.

[Ost59a] A. M. Ostrowski, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. I-VI*, Arch. Rational Mech. Anal. (1959), 1, 233–241, 2, 423–428, 3, 325–340, 3, 341–347, 3, 472–481, and 4, 153–165.

[Ost59b] ———, *On the convergence of the Rayleigh quotient iteration for the computation of the characteristic roots and vectors. III*, Arch. Rational Mech. Anal. **3** (1959), 325–340.

[Par74] B. N. Parlett, *The Rayleigh quotient iteration and some generalizations for non-normal matrices*, Mathematics of Computation **28** (1974), no. 127, 679–693.

[Par98] ———, *The symmetric eigenvalue problem*, Classics in applied mathematics, vol. 20, SIAM, Philadelphia, 1998.

[PW79] G. Peters and J. H. Wilkinson, *Inverse iteration, ill-conditioned equations and Newton's method*, SIAM Review **21** (1979), no. 3, 339–360.

[RR02] A. C. M. Ran and L. Rodman, *A class of robustness problems in matrix analysis*, Interpolation Theory, Systems Theory and Related Topics, The Harry Dym Anniversary Volume (D. Alpay, I. Gohberg, and V. Vinnikov, eds.), Operator Theory: Advances and Applications, vol. 134, Brikhäuser, 2002, pp. 337–383.

[Smi94] S. T. Smith, *Optimization techniques on Riemannian manifolds*, Hamiltonian and gradient flows, algorithms and control (Anthony Bloch, ed.), Fields Institute Communications, vol. 3, American Mathematical Society, 1994, pp. 113–136.

[Ste73] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review **15** (1973), no. 4, 727–764.

[Wil65] J. H. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.