SIAM REVIEW
Vol. 44, No. 1, pp. 57–73

© 2002 Society for Industrial and Applied Mathematics

A Grassmann–Rayleigh Quotient Iteration for Computing Invariant Subspaces*

P.-A. Absil†
R. Mahony‡
R. Sepulchre†
P. Van Dooren§

Abstract. The classical Rayleigh quotient iteration (RQI) allows one to compute a one-dimensional invariant subspace of a symmetric matrix A. Here we propose a generalization of the RQI which computes a p-dimensional invariant subspace of A. Cubic convergence is preserved and the cost per iteration is low compared to other methods proposed in the literature.

Key words. Rayleigh quotient iteration, invariant subspace, Grassmann manifold

AMS subject classification. 65F15

PII. S0036144500378648

1. Introduction.

Many classical methods exist for computing a single eigenpair (eigenvector and eigenvalue) of a symmetric matrix A. Among them are the power method and (shifted) inverse iteration, with the Rayleigh quotient iteration (RQI) as a particular case [14, 15]. The RQI is of particular interest because of its cubic convergence and its potential use in the shifted QR algorithm [22, 15].

In some cases, especially for multiple or clustered eigenvalues, it is advisable to compute the whole invariant subspace spanned by the corresponding eigenvectors. This leads one to consider algorithms whose iterates are general p-dimensional subspaces of $\mathbb{R}^n$. Algorithms for refining estimates of invariant subspaces were first proposed by Stewart [20], Dongarra, Moler, and Wilkinson [7], and Chatelin [4]. Demmel [5] unified these methods by showing that they amount to solving the same Riccati equation. Recently, other methods have been proposed, notably by Helmke and Moore [13] and Edelman, Arias, and Smith [9], based on the fact that the invariant subspaces are the stationary points of a well-chosen cost function. The fundamental

*Received by the editors September 26, 2000; accepted for publication (in revised form) August 22, 2001; published electronically February 1, 2002.
http://www.siam.org/journals/sirev/44-1/37864.html

†Institut d'Electricité Montefiore, Université de Liège, Bât. B28 Systèmes, Grande Traverse 10, 4000 Liège, Belgium (absil@montefiore.ulg.ac.be, R.Sepulchre@ulg.ac.be). The first author's work was supported by FNRS.

‡Department of Engineering, Australian National University, ACT, 0200, Australia (Robert.Mahony@anu.edu.au).

§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).

difference between the former methods [20, 7, 4, 5] and the latter [13, 9] is that the former represent subspaces by means of a fixed projective space representation, while the latter rely on the quotient geometry of the Grassmann manifold (see section 4.8 in [9] and section 7 in the present paper). Methods of the first category [5] all display linear convergence (as in the classical subspace iteration of Rutishauser [17, 18]) or quadratic convergence [4] but at a high numerical cost of $O(n^3)$ per iteration. In the second category, Newton's method on the Grassmann manifold displays cubic convergence for symmetric A [19, 9] but at a high cost of $O(n^3)$ per iteration because it does not take advantage of a condensed form of A.

In the present paper, we propose a generalization of the RQI dealing with p-dimensional subspaces of $\mathbb{R}^n$. The property of cubic convergence of the classical RQI extends to the generalized algorithm. Moreover, the numerical cost of each iteration is shown to be as low as $O(np^2)$ after a preliminary reduction to condensed form. The generalized RQI we propose for invariant subspace computation is defined as follows (practical implementation is discussed in section 6).

ALGORITHM 1.1. *Pick an orthonormal basis $X_{(0)}$ in $\mathbb{R}^{n\times p}$, i.e., $X_{(0)}^T X_{(0)} = I_p$. Then, for $k = 0, 1, 2, \ldots$, repeat the following:*

1. *Compute the solution $Z \in \mathbb{R}^{n\times p}$ of the Sylvester equation*

$$AZ - ZX_{(k)}^T AX_{(k)} = X_{(k)}. \tag{1.1}$$

2. *Compute $X_{(k+1)} := \mathrm{qf}(Z)$, where $\mathrm{qf}(Z)$ denotes the Q-factor of the QR decomposition of Z.*

The iterates in Algorithm 1.1 are orthogonal matrices $X_{(k)}$, but our interest is in fact directed towards a p-dimensional invariant *subspace* of A. In section 4, we show that Algorithm 1.1 indeed defines an algorithm on the set of p-dimensional (linear) subspaces of $\mathbb{R}^n$. This set can be endowed with a manifold structure, called the *Grassmann manifold* $\mathrm{Gr}(p, n)$, that provides a suitable mathematical framework in which to analyze (1.1). Section 2 motivates the approach taken in the analysis of (1.1) via a geometrical interpretation of the classical Rayleigh quotient algorithm, i.e., Algorithm 1.1 with $p = 1$.

The organization of the paper is as follows. Section 2 reviews the classical RQI in the projective space setting, preparing the generalization to the Grassmann manifold. This generalization is carried out in section 4 after an overview of the essential features of the Grassmann manifold in section 3. The local and global convergence properties of the new algorithm are analyzed in section 5, and its practical implementation is considered in section 6. Section 7 is dedicated to a comparison with other Grassmannian methods for eigenspace computation. Conclusions are drawn in section 8.

2. A Geometrical View of the Classical RQI. In this section, we briefly review the classical RQI. We show that it defines an iteration on the set of one-dimensional (linear) subspaces of $\mathbb{R}^n$, and we provide coordinate charts for this set. These concepts will be generalized to higher dimensional subspaces in the following sections.

Let A be a real symmetric $n \times n$ matrix. The classical RQI is as follows [14, 15].

ALGORITHM 2.1 (RQI). *Pick a unit vector $x_{(0)}$; then, for $k = 0, 1, 2, \ldots$, repeat the following:*

1. *Compute the Rayleigh quotient $\rho_{(k)} = x_{(k)}^T A x_{(k)}$.*
2. *If $A - \rho_{(k)} I$ is singular, then solve $(A - \rho_{(k)} I)x_{(k+1)} = 0$ for unit vector $x_{(k+1)}$ and stop. Otherwise, solve equation*

$$(A - \rho_{(k)} I) z_{(k+1)} = x_{(k)} \tag{2.1}$$

for $z_{(k+1)}$.

3. *Normalize, i.e.,* $x_{(k+1)} = z_{(k+1)}/\|z_{(k+1)}\|$.
4. *If* $\|z_{(k+1)}\|$ *is big enough, then stop.*

The iterates $x_{(k)}$ of Algorithm 2.1 belong to the unit sphere S^{n-1}. But the unitary norm of the iterates is irrelevant: the required information is contained in the direction of $x_{(k)}$. In other words, an iterate $x \in S^{n-1}$ is a *representation* of the one-dimensional subspace $\{x\alpha : \alpha \in \mathbb{R}\}$, which we shall denote as $\lfloor x \rfloor$, anticipating a broader definition appearing in section 3. The set of all one-dimensional subspaces of $\mathbb{R}^n$ is called the *real projective space* $\mathbb{RP}^{n-1}$.

A subspace $\lfloor x \rfloor$ can be represented by two elements of S^{n-1}: x or $-x$. It is straightforward to show that if $x_{(k)}$ maps to $x_{(k+1)}$ through an iteration of Algorithm 2.1, then $-x_{(k)}$ maps to $-x_{(k+1)}$. Thus, Algorithm 2.1 induces an algorithm on $\mathbb{RP}^{n-1}$ with step $\lfloor x_{(k)} \rfloor \mapsto \lfloor x_{(k+1)} \rfloor$, which can be written as follows.

ALGORITHM 2.2 (RQI on real projective space). *Pick an element* $\mathcal{S}_0$ *in* $\mathbb{RP}^{n-1}$*; then, for* $k = 0, 1, 2, \ldots$*, repeat the following:*

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

$$(A - \rho_k I)z = y \tag{2.2}$$

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

It is easy to see that $\mathcal{S}_{k+1}$ does not depend on the representation y chosen in step 1 and that Algorithm 2.1 is a realization of Algorithm 2.2 using unitary vectors for representing elements of $\mathbb{RP}^{n-1}$.

The representation of one-dimensional subspaces by elements of S^{n-1} is convenient for Algorithm 2.1 but awkward from the analysis viewpoint for at least two reasons: the representation is not unique (two candidates for the same element); and a point of S^{n-1} is represented in $\mathbb{R}^n$ by n components, whereas the dimension of $\mathbb{RP}^{n-1}$ is $n-1$. These drawbacks also hold for a representation by elements of $\mathbb{R}^n \setminus \{0\}$ as in Algorithm 2.2.

For analysis purposes, a convenient representation of one-dimensional subspaces is as follows. Pick W on S^{n-1} and complete it with $W_\perp$ so that $(W|W_\perp)$ is an orthonormal basis of $\mathbb{R}^n$. Denote by Γ_W the elements of $\mathbb{RP}^{n-1}$ which are not orthogonal to W. For any element $\lfloor y \rfloor$ of Γ_W, there exists one and only one $K \in \mathbb{R}^{n-1}$ such that $\lfloor y \rfloor = \lfloor W + W_\perp K \rfloor$. This defines a bijection between Γ_W and $\mathbb{R}^{n-1}$, i.e., a *chart*. A single chart cannot represent $\mathbb{RP}^{n-1}$ ($\mathbb{RP}^{n-1}$ is not diffeomorphic to $\mathbb{R}^{n-1}$) but different charts are obtained by changing the reference point W.

Figure 2.1 illustrates this concept of chart in combination with the RQI in the case $n = 2$. Let $A = \mathrm{diag}(\lambda_1, \lambda_2)$, with λ_1 and λ_2 real and distinct. Take $W = (1, 0)^T$ (note that W is one of the two distinct eigenvectors of A) and $W_\perp = (0, 1)^T$. The bijection between one-dimensional subspaces and K's is best viewed on Figure 2.1. We now show that if $x_{(k)}$ and $x_{(k+1)}$ are successive iterates of Algorithm 2.1 with $\lfloor x_{(k)} \rfloor = \lfloor W + W_\perp K_{(k)} \rfloor$ and $\lfloor x_{(k+1)} \rfloor = \lfloor W + W_\perp K_{(k+1)} \rfloor$, then $K_{(k+1)} = K_{(k)}^3$. This demonstrates the property that iterates of RQI (Algorithm 2.1) cubically converge to eigenvectors of A. Indeed, if $\lfloor x_{(k)} \rfloor = \lfloor W + W_\perp K_{(k)} \rfloor$, then one has [3]

$$x_{(k)} = \pm(1, K_{(k)})^T / \sqrt{1 + K_{(k)}^2},$$
$$\rho_k = x_{(k)}^T A x_{(k)} = (\lambda_1 + \lambda_2 K_{(k)}^2)/(1 + K_{(k)}^2),$$

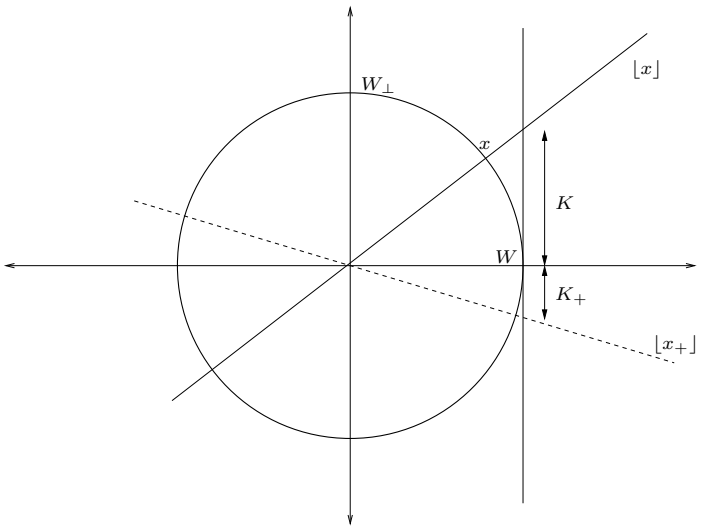

Fig. 2.1 *RQI for $n = 2$.*

$$z_{(k+1)} = \pm \frac{\sqrt{1 + K_{(k)}^2}}{(\lambda_1 - \lambda_2)} \left(\frac{1}{K_{(k)}^2}, -K_{(k)} \right)^T,$$
$$\lfloor x_{(k+1)} \rfloor = \lfloor z_{(k+1)} \rfloor = \lfloor (1, -K_{(k)}^3)^T \rfloor.$$

3. Grassmann Manifold and Local Coordinates. In view of generalizing the RQI and proving its cubic convergence, we need a representation of p-dimensional subspaces by means of matrices and an associated measure of distance between subspaces.

The set of the p-dimensional linear subspaces of $\mathbb{R}^n$ ("linear" will be omitted in what follows) is the *Grassmann manifold*, denoted here by $\mathrm{Gr}(p, n)$, reducing to $\mathbb{RP}^{n-1}$ when $p = 1$. Referring the reader to the literature (e.g., [8, 9]) for a comprehensive treatment of this manifold, let us simply view $\mathrm{Gr}(p, n)$ as the set of the p-dimensional subspaces of $\mathbb{R}^n$. A p-dimensional subspace $\mathcal{Y}$ of $\mathbb{R}^n$ will be represented by an $n \times p$ full column rank matrix Y whose columns span this space $\mathcal{Y}$. For ease of reference, we denote by $\mathrm{ST}(p, n)$, termed the *noncompact Stiefel manifold*, the set of the $p \times n$ real matrices with full column rank. Since the column space concept is widely used in this article, we introduce the notation $\lfloor Y \rfloor$ for the column space of Y:

$$\lfloor Y \rfloor := \{Y\alpha : \alpha \in \mathbb{R}^p\}. \tag{3.1}$$

Note that $\lfloor Y \rfloor = \lfloor \hat{Y} \rfloor$ if and only if there exists a nonsingular p-by-p matrix M such that $\hat{Y} = YM$.

The charts we used in section 2 in order to represent $\mathbb{RP}^{n-1}$ are adapted to $\mathrm{Gr}(p, n)$ with little effort, as we now show. Consider the *Stiefel manifold* $\mathrm{St}(p, n)$ (see [9]), defined as the set of all n-by-p orthogonal real matrices,

$$\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} : X^T X = I_p\} \tag{3.2}$$

(observe that $\mathrm{St}(1,n) = S^{n-1}$). Consider a point $W \in \mathrm{St}(p,n)$ and pick $W_\perp \in \mathrm{St}(n-p,n)$ such that $\lfloor W \rfloor$ and $\lfloor W_\perp \rfloor$ are orthogonal, that is, $W_\perp^T W_\perp = I_{n-p}$ and $W_\perp^T W = 0$. In other words, $(W|W_\perp)$ belongs to the orthogonal group $O_n = \{Q \in \mathbb{R}^{n\times n} : Q^T Q = I_n\}$. Since the columns of $(W|W_\perp)$ form an orthonormal basis of $\mathbb{R}^n$, any Y in $\mathrm{ST}(p,n)$ admits the decomposition $Y = WY_1 + W_\perp Y_2$, where $Y_1 = W^T Y$ and $Y_2 = W_\perp^T Y$. Note that Y_1 and Y_2 depend on the choice of the reference $(W|W_\perp)$ in O_n. If Y_1 is invertible, then there is a unique $K \in \mathbb{R}^{(n-p)\times p}$ such that

$$\lfloor Y \rfloor = \lfloor W + W_\perp K \rfloor. \tag{3.3}$$

Obviously, $K = Y_2 Y_1^{-1} = (W_\perp^T Y)(W^T Y)^{-1}$. This K is called the *local expression* or *local coordinates* of $\lfloor Y \rfloor$, with reference $(W|W_\perp)$. The mapping $\lfloor Y \rfloor \mapsto K$ defines a *coordinate chart*. Any other reference point $(W|W_\perp)$ in O_n defines another coordinate chart, and these charts make $\mathrm{Gr}(p,n)$ a *smooth manifold*, the so-called *Grassmann manifold* (of p-dimensional linear subspaces of $\mathbb{R}^n$).

Note that in the case $p = 1$, the operation $Y \to (W + W_\perp K)$ is the central projection of Y (a point of $\mathbb{R}^n$) onto the tangent plane to the unit sphere of $\mathbb{R}^n$ at W (see Figure 2.1). The $n \times (n-1)$ matrix $W_\perp$ sets a basis in the tangent plane, and K is the decomposition of the aforementioned central projection in this basis.

A classical definition for the distance between two linear subspaces $\lfloor X \rfloor$ and $\lfloor Y \rfloor$ is given by (see [20])

$$\mathrm{dist}(\lfloor X \rfloor, \lfloor Y \rfloor) = \|P_X - P_Y\|_2, \tag{3.4}$$

where P_X and P_Y are the orthogonal projectors on $\lfloor X \rfloor$ and $\lfloor Y \rfloor$, respectively. Let $X \in \mathrm{St}(p,n)$ and $Y = X + X_\perp K$ with $(X|X_\perp)$ in O_n. The following theorem establishes that $\|K\|_2$ is a valid measure of the distance (3.4) for assessing the rate of convergence of $\lfloor X + X_\perp K \rfloor$ to $\lfloor X \rfloor$.

THEOREM 3.1. *Let $X \in \mathrm{St}(p,n)$ and $(X|X_\perp) \in O_n$. The tangents of the canonical angles between the subspaces $\lfloor X \rfloor$ and $\lfloor X + X_\perp K \rfloor$ are the singular values of K. In particular, the tangent of the largest canonical angle is equal to $\|K\|_2$. Moreover,*

$$\mathrm{dist}(\lfloor X \rfloor, \lfloor X + X_\perp K \rfloor) = \sin\ \mathrm{atan}\ \|K\|_2 = \frac{\|K\|_2}{\sqrt{1 + \|K\|_2^2}} = \|K\|_2(1 + O(\|K\|_2^2)). \tag{3.5}$$

Proof. The proof is a direct consequence of the material in [20, section 2]. The distance $\mathrm{dist}(\lfloor X \rfloor, \lfloor X + X_\perp K \rfloor)$ is the sine of the largest canonical angle between $\lfloor X \rfloor$ and $\lfloor X + X_\perp K \rfloor$. The cosines of the canonical angles θ_i between $\lfloor X \rfloor$ and $\lfloor X + X_\perp K \rfloor$ are the singular values of $\hat{Y}^T X = (I_p + K^T K)^{-1/2}$, where $\hat{Y} = (X + X_\perp K)(I_p + K^T K)^{-1/2}$, so that $\hat{Y} \in \mathrm{St}(p,n)$ and $\lfloor \hat{Y} \rfloor = \lfloor X + X_\perp K \rfloor$. Denoting by τ_i $(i = 1, \ldots, p)$ the singular values of K, one has $\cos\theta_i = (1 + \tau_i^2)^{-1/2}$, whence $\theta_i = \mathrm{atan}\ \tau_i$. Finally, the largest singular value of K is equal to the spectral norm of K, defined as $\|K\|_2 := \max_{\|u\|_2 = 1} \|Ku\|_2$. ∎

4. Extending the RQI from $\mathbb{RP}^{n-1}$ to $\mathrm{Gr}(p,n)$.

4.1. Rayleigh Quotient. The quest for a generalization of the RQI (Algorithm 2.1) leads us to generalize the Rayleigh quotient itself.

When $Y \in \mathbb{R}^n \setminus \{0\}$, the Rayleigh quotient [14] is defined as $\rho_A(Y) = \frac{Y^T A Y}{Y^T Y}$. A generalization for $Y \in \mathrm{ST}(p,n)$ uses the *matrix Rayleigh quotient*

$$R_A(Y) = (Y^T Y)^{-1} Y^T A Y \tag{4.1}$$

to define a *generalized (scalar) Rayleigh quotient* [13] as

$$\rho_A(Y) = \text{tr } R_A(Y) = \text{tr } Y^T AY(Y^TY)^{-1} = \text{tr } (Y^TY)^{-1/2}Y^TAY(Y^TY)^{-1/2}. \tag{4.2}$$

Important properties of the classical Rayleigh quotient [14] are preserved, as summarized in the following proposition.

PROPOSITION 4.1. *Definitions* (4.1)–(4.2) *satisfy the following properties:*

(i) *(Homogeneity)* $\rho_A(YM) = \rho_A(Y)$ *for all invertible p-by-p matrices* M. *This means that if* $\lfloor Y_1 \rfloor = \lfloor Y_2 \rfloor$, *then* $\rho_A(Y_1) = \rho_A(Y_2)$. *In other words,* ρ_A *defines a scalar field on* $\text{Gr}(p,n)$.

(ii) *(Stationarity)* $D\rho_A(Y) = 0$ *if and only if* $\lfloor Y \rfloor$ *is an invariant subspace of* A.

(iii) *(Minimal residual)* $\|AY - YB\|_F^2 \geq \|AY\|_F^2 - \|YR_A(Y)\|_F^2$, *and the equality holds if and only if* $B = R_A(Y)$. *Hence* $B = R_A(Y)$ *is the unique minimizer of* $\|AY - YB\|_F^2$.

Proof.

(i) $\rho_A(YM) = \text{tr } M^{-1}(Y^TY)^{-1}Y^TAYM = \rho_A(Y)$.

(ii)

$$\begin{aligned} D\rho_A(Y)[Z] &= \text{tr } (Z^TAY(Y^TY)^{-1} + Y^TAZ(Y^TY)^{-1} \\ &\qquad -Y^TAY(Y^TY)^{-1}(Z^TY + Y^TZ)(Y^TY)^{-1}) \\ &= \text{tr } (2(Y^TY)^{-1}Y^TA - 2(Y^TY)^{-1}Y^TAY(Y^TY)^{-1}Y^T)Z \\ &= \text{tr } 2(Y^TY)^{-1}Z^T(AY - Y(Y^TY)^{-1}Y^TAY), \end{aligned}$$

$$D\rho_A(Y) = 0 \Leftrightarrow D\rho_A(Y)[Z] = 0 \ \forall Z \in \mathbb{R}^{n\times p} \tag{4.3}$$

$$\Leftrightarrow AY - Y(Y^TY)^{-1}Y^TAY = 0 \tag{4.4}$$

$$\Leftrightarrow \exists\, R_A = (Y^TY)^{-1}Y^TAY \in \mathbb{R}^{p\times p} : AY - YR_A(Y) = 0 \tag{4.5}$$

$$\Leftrightarrow \lfloor AY \rfloor \subset \lfloor Y \rfloor; \tag{4.6}$$

i.e., $\lfloor Y \rfloor$ is an invariant subspace of A.

(iii)

$$\begin{aligned} \|AY - YB\|_F^2 &= \text{tr}(Y^TAAY + B^TY^TYB - B^TY^TAY - Y^TAYB) \\ &= \text{tr}(Y^TAAY + (B - R_A)^TY^TY(B - R_A) - R_A^TY^TYR_A) \\ &= \|AY\|_F^2 - \|YR_A\|_F^2 + \|Y(B - R_A)\|_F^2 \end{aligned}$$

and, because $Y \in \text{ST}(p,n)$, the last term vanishes if and only if $B = R_A(Y)$. ∎

Property (iii) establishes that $YR_A(Y)$ is the point in $\lfloor Y \rfloor$ that is closest (in the sense of the Frobenius norm) to AY. We shall see in section 5.2 that this is also true for the 2-norm.

The development for point (ii) can be recast in local coordinates (see section 3). In local coordinates with reference $(W|W_\perp)$, the generalized Rayleigh quotient reads

$$\rho_A(\lfloor Y \rfloor) = \text{tr } (A_{11} + A_{12}K + K^TA_{21} + K^TA_{22}K)(I_p + K^TK)^{-1}, \tag{4.7}$$

where $A_{11} := W^TAW$, $A_{12} := W^TAW_\perp$, $A_{21} := W_\perp^TAW$, $A_{22} := W_\perp^TAW_\perp$, $\lfloor Y \rfloor = \lfloor W + W_\perp K \rfloor$. Computing the derivative of (4.7) shows that ρ_A is stationary at $K = 0$ (local coordinates of $\lfloor W \rfloor$) if and only if $A_{12} = 0 = A_{21}$, i.e., $\lfloor W \rfloor$ is an invariant subspace of A.

Moreover, (4.7) shows that if $\lfloor W \rfloor$ is an invariant subspace of A, then ρ_A is even in K. This feature, which of course is also valid when $p = 1$, plays a key role in the performance of the Rayleigh quotient shifts for accelerating the convergence of the inverse iteration.

4.2. Grassmann–Rayleigh Quotient Iteration. When $p = 1$, (4.5) reduces to $(A - \rho_A(Y)I)Y = 0$. By analogy with the classical RQI (Algorithm 2.2), (4.5) leads us to consider the following *Grassmann–Rayleigh Quotient Iteration* (GRQI).

ALGORITHM 4.2 (GRQI). *Pick a p-dimensional subspace $\mathcal{Y}_0$ of $\mathbb{R}^n$ (that is, $\mathcal{Y}_0$ belongs to the Grassmann manifold $\mathrm{Gr}(p,n)$). Then, for $k = 0, 1, 2, \ldots$, repeat the following:*

1. *Pick $Y \in \mathrm{ST}(p,n)$ such that $\lfloor Y \rfloor = \mathcal{Y}_k$.*
2. *Solve*

$$AZ - Z\underbrace{(Y^TY)^{-1}Y^TAY}_{R_A(Y)} = Y \tag{4.8}$$

   for $Z \in \mathbb{R}^{n\times p}$.
3. *Define $\mathcal{Y}_{k+1} := \lfloor Z \rfloor$.*

Notice that the mapping

$$\boldsymbol{T}_Y : Z \mapsto AZ - Z(Y^TY)^{-1}Y^TAY \tag{4.9}$$

can be singular. What to do for this nongeneric case will be discussed in section 6.

Algorithm 4.2 indeed defines an algorithm on $\mathrm{Gr}(p,n)$.

PROPOSITION 4.3. *The sequence $\{\mathcal{Y}_{(k)}\}_{k\geq 0}$ generated by Algorithm 4.2 is independent of the choice of Y used to represent $\mathcal{Y}_k$. That is, if Y_a and Z_a (Y_b and Z_b, respectively) verify (4.8) and if $\lfloor Y_a \rfloor = \lfloor Y_b \rfloor$, then $\lfloor Z_a \rfloor = \lfloor Z_b \rfloor$.*

Proof. If $AZ - Z(Y^TY)^{-1}Y^TAY = Y$, then a right-multiplication by an invertible matrix M leads to $AZM - ZM((YM)^T(YM))^{-1}(YM)^TAYM = YM$. ∎

Moreover, Algorithm 4.2 and our original Algorithm 1.1 are equivalent in the sense of the p-subspaces.

PROPOSITION 4.4. *Algorithms 4.2 and 1.1 are equivalent on the manifold $\mathrm{Gr}(p,n)$; i.e., they define the same subspace iteration.*

Proof. Algorithm 1.1 is a realization of Algorithm 4.2 with the particular choice $Y_+ = \mathrm{qf}(Z)$. Note that the qf operation does not alter the column space. ∎

Clearly, for computational matters, it is preferable to represent elements of $\mathrm{Gr}(p,n)$ by orthogonal bases in order to prevent loss of numerical rank, which would deteriorate the quality of the subspace representation. Thus, Algorithm 1.1 must be interpreted as a numerically reliable realization of Algorithm 4.2.

5. Convergence of the GRQI.

5.1. Cubic Local Convergence. This section is dedicated to the proof of the following theorem, which generalizes the result of cubic local convergence of the classical RQI [14, 15].

THEOREM 5.1 (cubic convergence of GRQI). *Suppose A is symmetric real. Let $\mathcal{V}$ be an invariant subspace of A and $\mathcal{V}_\perp$ be its orthogonal complement. If $A|_{\mathcal{V}}$ and $A|_{\mathcal{V}_\perp}$ have no eigenvalue in common, then the iterates $\mathcal{Y}_k$ of the GRQI (Algorithm 4.2) converge cubically to $\mathcal{V}$ for all $\mathcal{Y}_0$ in a neighborhood of $\mathcal{V}$.*

Consider the iteration (4.8):

$$\boldsymbol{T}_Y Z \equiv AZ - Z(Y^TY)^{-1}Y^TAY = Y.$$

Take $V \in \mathrm{St}(p,n)$ such that $\lfloor V \rfloor = \mathcal{V}$ and that $\Lambda_1 := V^TAV$ is diagonal. Also take $V_\perp \in \mathrm{St}(n-p,n)$ such that $\lfloor V_\perp \rfloor = \mathcal{V}_\perp$ and that $\Lambda_2 := V_\perp^TAV_\perp$ is diagonal. Note that Λ_1 and Λ_2 have no element in common.

The columns of $(V|V_\perp)$ form a basis of $\mathbb{R}^n$. The matrix Y decomposes as $Y = VY_1 + V_\perp Y_2$, and $K := Y_2 Y_1^{-1}$ is the local representation of $\lfloor Y \rfloor$ in the coordinate chart with reference $(V|V_\perp)$; see section 3. Define $\tilde{Z} := ZY_1^{-1}$ and note that $\lfloor \tilde{Z} \rfloor = \lfloor Z \rfloor$. Decompose $\tilde{Z}$ according to $\tilde{Z} = VZ_1 + V_\perp Z_2$, where the tilde is omitted in the decomposition for ease of notation. Then $K_+ = Z_2 Z_1^{-1}$ is the local representation of the next iterate $\lfloor Y \rfloor_+$ of Algorithm 4.2. Using $V + V_\perp K$ to represent $\lfloor Y \rfloor$, (4.8) reads

$$\Lambda_1 Z_1 - Z_1(I_p + K^T K)^{-1}(\Lambda_1 + K^T \Lambda_2 K) = I_p, \tag{5.1}$$

$$\Lambda_2 Z_2 - Z_2(I_p + K^T K)^{-1}(\Lambda_1 + K^T \Lambda_2 K) = K, \tag{5.2}$$

$$K_+ = Z_2 Z_1^{-1}. \tag{5.3}$$

We shall now prove that

$$\|Z_1^{-1}\| \leq c_1 \|K\|^2, \tag{5.4}$$

$$\|Z_2\| \leq c_2 \|K\| \tag{5.5}$$

for all K sufficiently small, from which it follows that $K_+ = O(K^3)$, i.e., that the convergence is cubic.

While (5.4)–(5.5) is valid for any matrix norm, the spectral norm $\|\cdot\|_2$ is of particular interest because of its interpretation in terms of distance and canonical angles between $\lfloor Y \rfloor$ and $\lfloor V \rfloor$; see section 3. We shall also use the Frobenius norm, denoted by $\|\cdot\|_F$.

Intuitively, (5.5) is due to the right-hand side of (4.8). It accounts for the linear convergence of unshifted or constant-shifted inverse iteration. Inequality (5.4) follows from the stationary property of $R_A(Y)$ around $\lfloor V \rfloor$ (see Proposition 4.1).

LEMMA 5.2. *Let A, B be $p \times p$ symmetric matrices. Let C be a $p \times p$ matrix. Define* $\operatorname{gap}[A, B] = \min |\lambda(A) - \lambda(B)|$. *Suppose* $\operatorname{gap}[A, B] > 0$. *Then, for all E, F in $\mathbb{R}^{p \times p}$ with $\|E\|_F$, $\|F\|_F < \operatorname{gap}[A, B]/2$, the equation*

$$(A + E)Z - Z(B + F) = C$$

admits one and only one solution $Z \in \mathbb{R}^{p \times p}$, and

$$\|Z\|_2 \leq \frac{\sqrt{p}}{\operatorname{gap}[A, B] - \|E\|_F - \|F\|_F} \|C\|_2.$$

Proof. One has

$$\begin{aligned} \|Z\|_2 \leq \frac{\|C\|_2}{\operatorname{sep}_2(A + E, B + F)} &\leq \frac{\|C\|_2}{\operatorname{sep}_F(A + E, B + F)/\sqrt{p}} \\ &\leq \frac{\sqrt{p}\|C\|_2}{\operatorname{gap}[A, B] - \|E\|_F - \|F\|_F} \end{aligned}$$

as a consequence of Definition 4.5 and Theorems 4.6, 4.7, and 4.10 in [20]. ∎

LEMMA 5.3. *Let D be diagonal and let $\|\cdot\|$ denote any consistent matrix norm. Then for every $\epsilon > 0$ there exists a $\delta > 0$ with the following property: for every E such that $\|E\| < \delta$ and such that D and $D - E$ have no eigenvalues in common, the unique solution Z to*

$$DZ - Z(D - E) = I \tag{5.6}$$

is invertible, and its inverse satisfies $\|Z^{-1}\| \leq (1 + \epsilon)\|E\|$.

Proof. The existence and uniqueness of Z are a consequence of the fact that the operator $Z \mapsto AZ - ZB$ is invertible if and only if A and B have no eigenvalues in common [11].

Denote by n the size of D, Z, E, and I. Suppose, without loss of generality, that $D = \operatorname{diag}(D_1, \ldots, D_q)$, where $D_j = d_j I_{m(j)}$, with $d_j \neq d_l$ whenever $j \neq l$. Partition E, I, and Z in the same manner.

Let $i \in \{1, \ldots, q\}$. Define $\mathcal{I} = (1, \ldots, i-1, i+1, \ldots, q)$. After a permutation bringing block i in the upper left position, (5.6) decomposes as

$$\begin{aligned}(5.7)\quad &\begin{pmatrix} D_i & 0 \\ 0 & D_{\mathcal{I}} \end{pmatrix} \begin{pmatrix} Z_{ii} & Z_{i\mathcal{I}} \\ Z_{\mathcal{I}i} & Z_{\mathcal{I}\mathcal{I}} \end{pmatrix} \\ &- \begin{pmatrix} Z_{ii} & Z_{i\mathcal{I}} \\ Z_{\mathcal{I}i} & Z_{\mathcal{I}\mathcal{I}} \end{pmatrix} \left[\begin{pmatrix} D_i & 0 \\ 0 & D_{\mathcal{I}} \end{pmatrix} - \begin{pmatrix} E_{ii} & E_{i\mathcal{I}} \\ E_{\mathcal{I}i} & E_{\mathcal{I}\mathcal{I}} \end{pmatrix} \right] = \begin{pmatrix} I_i & 0 \\ 0 & I_{\mathcal{I}} \end{pmatrix}.\end{aligned}$$

Blocks ii and $i\mathcal{I}$ of this equation give, respectively,

$$(5.8)\qquad Z_{ii}E_{ii} + Z_{i\mathcal{I}}E_{\mathcal{I}i} = I_i$$

and

$$(5.9)\qquad d_i Z_{i\mathcal{I}} - Z_{i\mathcal{I}}D_{\mathcal{I}} + Z_{ii}E_{i\mathcal{I}} + Z_{i\mathcal{I}}E_{\mathcal{I}\mathcal{I}} = 0.$$

Solving (5.9) for $Z_{i\mathcal{I}}$ leads to

$$(5.10)\qquad Z_{i\mathcal{I}} = -Z_{ii}E_{i\mathcal{I}}(d_i I_{\mathcal{I}} - D_{\mathcal{I}} + E_{\mathcal{I}\mathcal{I}})^{-1} = -Z_{ii}\, O(E),$$

where $(d_i I_i - D_{\mathcal{I}} + E_{\mathcal{I}\mathcal{I}})$ is invertible whenever E is sufficiently small. Substituting (5.10) in (5.8) yields

$$Z_{ii}[E_{ii} - E_{i\mathcal{I}}(d_i I_{\mathcal{I}} - D_{\mathcal{I}} + E_{\mathcal{I}\mathcal{I}})^{-1}E_{\mathcal{I}i}] = I_i,$$

whence

$$(5.11)\qquad Z_{ii}^{-1} = E_{ii} - E_{i\mathcal{I}}(d_i I_{\mathcal{I}} - D_{\mathcal{I}} + E_{\mathcal{I}\mathcal{I}})^{-1}E_{\mathcal{I}i} = E_{ii} + O(E^2).$$

From (5.10), we have $Z = \operatorname{diag}(Z_{11}, \ldots, Z_{qq})(I_n + O(E))$. Then, from (5.11), we conclude that Z is invertible whenever E is sufficiently small (provided that Z exists, which is guaranteed when D and $D + E$ have no common eigenvalue), and

$$\begin{aligned}(5.12)\qquad Z^{-1} &= (I + O(E))^{-1} \operatorname{diag}(Z_{ii}^{-1}) \\ (5.13)\qquad &= (I + O(E)) \operatorname{diag}(Z_{ii}^{-1}) \\ (5.14)\qquad &= (I + O(E)) \left(\operatorname{diag}(E_{11}, \ldots, E_{qq}) + O(E^2)\right) \\ (5.15)\qquad &= \operatorname{diag}(E_{11}, \ldots, E_{qq}) + O(E^2),\end{aligned}$$

whence

$$(5.16)\qquad \|Z^{-1}\| \leq (1 + O(\|E\|))\ \|E\|,$$

which completes the proof. ∎

We are now able to prove (5.5) and (5.4). First note that

$$(I + K^T K)^{-1}(\Lambda_1 + K^T \Lambda_2 K) = \Lambda_1 - (K^T K \Lambda_1 - K^T \Lambda_2 K + O(K^4)).$$

Lemma 5.2 applied to (5.2) directly leads to

$$\|Z_2\|_2 \leq \frac{\sqrt{p}}{\mathrm{gap}[\Lambda_1, \Lambda_2] + O(\|K\|^2)} \|K\|_2, \tag{5.17}$$

which proves (5.5). For (5.4), defining $E = K^T K \Lambda_1 - K^T \Lambda_2 K + O(K^4)$ and applying Lemma 5.3 to the Sylvester equation (5.1), we obtain

$$\begin{aligned} \|Z_1^{-1}\| &\leq (1 + O(\|E\|))\|E\| \\ &\leq (\|\Lambda_1\| + \|\Lambda_2\|)\|K\|^2 + O(K^4). \end{aligned} \tag{5.18}$$

In conclusion,

$$\|K_+\|_2 \leq \sqrt{p}\, \frac{\|A|_\nu\|_2 + \|A|_{\nu_\perp}\|_2}{\mathrm{gap}[A|_\nu, A|_{\nu_\perp}]} \|K\|_2^3 + O(K^5). \tag{5.19}$$

This concludes the proof of Theorem 5.1.

REMARK 5.4. *Note that Lemma* 5.3 *cannot be used in the nongeneric case where* Λ_1 *and* $(I+K^TK)^{-1}(\Lambda_1+K^T\Lambda_2K)$ *have a common eigenvalue. This case corresponds to a singular* $\boldsymbol{T}_Y$ (4.9). *An easy remedy when a singular* $\boldsymbol{T}_Y$ *shows up during the iteration is to slightly perturb the offending iterate so as to dismiss the singularity while staying in the neighborhood where cubic convergence holds.*

Theorem 5.1 guarantees local cubic convergence but provides no information about the size of the basin of attraction of a given invariant subspace $\mathcal{V}$. Obtaining such an estimate from the above proof amounts to quantifying the region around $\mathcal{V}$ where the bounds (5.4) and (5.5) hold. This estimate will thus shrink if Z is poorly conditioned. The term $(d_i I_{\mathcal{I}} - D_{\mathcal{I}} + E_{\mathcal{I}\mathcal{I}})^{-1}$ in (5.10) and (5.11) suggests that this will happen when the distinct eigenvalues of $A|_{\mathcal{V}}$ are not well separated. The following example illustrates that in such situations, some subspaces that are "close" to $\mathcal{V}$ may not belong to its basin of attraction.

EXAMPLE 5.5. *Let* $A = \mathrm{diag}(1.01, 1, 2)$ *and*

$$Y = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0.1232 & 0.1593 \end{bmatrix}.$$

The initial subspace is thus "close" to the invariant subspace $\mathcal{V}$ *spanned by the columns of*

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Nevertheless, one iterate of Algorithm 4.2 *yields*

$$Z = \begin{bmatrix} -20.9669 & -162.8163 \\ -16.0128 & -124.8679 \\ 0.2046 & 0.0253 \end{bmatrix} \quad \text{and} \quad X_+ = \begin{bmatrix} 0.7947 & -0.1495 \\ 0.6069 & 0.2081 \\ -0.0078 & 0.9666 \end{bmatrix},$$

where X_+ *nearly contains the eigenvector* $[0, 0, 1]^T$. *This is a situation where the algorithm initialized "close" to* $\mathcal{V}$ *converges to a different invariant subspace. This behavior is a consequence of the two eigenvalues of* $A|_{\mathcal{V}}$ *being close to each other, which results in an ill-conditioned* Z *for some initial conditions close to* $\mathcal{V}$.

Note that there exists a set (of measure zero) of matrices Y such that the solution Z of the GRQI equation (4.8) is not full rank. As a corollary of the proof of cubic convergence, this set is bounded away from the invariant subspaces of A; indeed it was shown that Z_1^{-1} exists for all K sufficiently small.

5.2. Global Convergence. Algorithm 1.1 (implemented with slight perturbation as explained above) was extensively tested on randomly selected matrices, as well as on some ill-conditioned matrices such as the Laplacian

$$A = \begin{pmatrix} -2 & 1 & \cdots & & \\ 1 & -2 & 1 & \cdots & \\ 0 & 1 & -2 & 1 & \cdots \\ \cdots & & & \ddots & \end{pmatrix}$$

and matrices with repeated eigenvalues. The particular case where there is no invariant subspace $\mathcal{V}$ such that $A|_{\mathcal{V}}$ and $A|_{\mathcal{V}_\perp}$ have no eigenvalue in common has also been investigated. In all these situations, Algorithm 1.1 invariably showed numerical convergence to an invariant subspace of A. This suggests that the generalized RQI inherits the good global convergence properties of the classical RQI. Unfortunately, the developments available for the RQI do not generalize, as we now show.

The global convergence properties of the classical RQI are well understood [14, 15, 3]. In [15], it was shown that the iterates $x_{(k)}$ of Algorithm 2.1 either converge to an eigenvector of A or converge to the bisectors of a pair of eigenvectors of A, the latter situation being unstable under perturbations of $x_{(k)}$. Batterson and Smillie [3] showed later that the set of points for which the RQI does not converge to an eigenvector is a set of measure zero.

The proof in [15], due to Kahan, relies on the following observation. Given $X \in \mathrm{St}(1,n) = S^{n-1}$, one defines a residual as

$$r = (A - \rho_A(X)I)X. \tag{5.20}$$

The Euclidean norm of the residual r can be interpreted as the distance between AX and $\lfloor X \rfloor$, i.e., $\|r\| = \min_{Y \in \lfloor X \rfloor} \|AX - Y\|$. The Euclidean norm of the residual is shown to be monotonically decreasing under the (classical) RQI and is used as a Lyapunov function in the proof of [15].

A straightforward generalization of (5.20) in the case where $X \in \mathrm{St}(p,n)$ would be

$$Res = AX - XR_A(X), \tag{5.21}$$

where $R_A(X) = X^T AX$ is the *matrix Rayleigh quotient*. Indeed,

$$\begin{aligned} \|AX - XB\|_{2,F} &= \|(X|X_\perp)^T(AX - XB)\|_{2,F} \\ &= \left\| \begin{pmatrix} X^TAX - B \\ X_\perp^T AX \end{pmatrix} \right\|_{2,F} = \left\| \begin{pmatrix} A_{11} - B \\ A_{21} \end{pmatrix} \right\|_{2,F} \end{aligned}$$

is minimal for $B = R_A(X)$. We see that $\|Res\|_{2,F} = \|A_{21}\|_{2,F}$ expresses the departure of

$$(X|X_\perp)^T A(X|X_\perp) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

from being a block upper triangular matrix.

Another way to measure the "eigenness" of a space $\lfloor X \rfloor$ is the principal angle between $\lfloor AX \rfloor$ and $\lfloor X \rfloor$. In the case $p = 1$, this principal angle would be

$$\theta = \arccos \frac{x^T Ax}{\|x\|\|Ax\|}.$$

This measure seems more appropriate because it is invariant under a multiplication of A by a constant.

Unfortunately, neither of the two distances proposed above shows a systematic decrease in numerical tests. This compromises the possibility of establishing a global convergence analysis along the lines of Kahan's proof.

6. Practical Implementation of the Algorithm. The practical relevance of the new method depends on two important issues: (i) keeping the computational cost of each iteration step as low as possible and (ii) detecting stagnation or convergence of the method.

6.1. Computational Efficiency. The key equation of Algorithm 1.1 is the Sylvester equation

$$AZ - Z\hat{A} = X, \quad X^T X = I_p, \tag{6.1}$$

where A and $\hat{A} = R_A(X) = X^T AX$ are both symmetric. It is easy to see that orthogonal similarity transformations on A and $\hat{A}$ do not alter the nature or solvability of the equation. Define indeed

$$A_u := U^T AU, \quad \hat{A}_v := V^T \hat{A} V, \quad X_{uv} := U^T XV \tag{6.2}$$

for $U \in O(n)$ and $V \in O(p)$. Then $Z_{uv} := U^T ZV$ solves the transformed equation

$$A_u Z_{uv} - Z_{uv} \hat{A}_v = X_{uv} \tag{6.3}$$

if and only if Z solves (6.1). One shows, moreover, that both systems of equations have the same conditioning [21]. The matrix $\hat{A} = X^T AX$ changes at each step of the iteration but is small compared to A. Transforming the n-by-n matrix A to a more convenient form A_u thus needs to be done *only once*, whereas computing a special form of the p-by-p matrix $\hat{A}_v$ can be done at each step of the algorithm. For this reason, we assume A to be already in an "appropriate" form $A_u = A$ and dismiss the index u in what follows. An "appropriate" form can be a banded matrix (see below). Efficient methods to solve (6.1) are typically based on an eigenvalue decomposition of the smallest matrix $\hat{A}$ [2, 12]. Then $\hat{A}_v$ is diagonal:

$$\hat{A}_v = \operatorname{diag}\{\rho_1, \ldots, \rho_p\} = V^T \hat{A} V. \tag{6.4}$$

An implementation of Algorithm 1.1 based on such reductions then requires the following basic steps at each iteration:

$$\hat{A} := X^T AX. \tag{6.5}$$

$$\text{Find } V \text{ such that } V^T \hat{A} V = \hat{A}_v = \operatorname{diag}\{\rho_1, \ldots, \rho_p\}. \tag{6.6}$$

$$X_v := XV. \tag{6.7}$$

$$\text{Find } Z_v \text{ such that } AZ_v - Z_v \hat{A}_v = X_v. \tag{6.8}$$

$$X_+ := \operatorname{qf}(Z_v V^T). \tag{6.9}$$

The eigenvalues ρ_i of $\hat{A} = R_A(X)$ are called the Ritz values of A with respect to $\lfloor X \rfloor$, and the corresponding columns of X_v are called the Ritz vectors.

Steps (6.6), (6.7), and (6.9) involve dense matrices and require $O(p^3)$, $O(np^2)$, and $O(np^2)$ floating point operations (flops), respectively. One would like the other two steps (6.5) and (6.8) to require at most $O(np^2)$ flops as well. This is possible by putting

A in a special "condensed" form. An obvious one is to choose A tridiagonal. Step (6.5) then requires $O(np^2)$ flops and step (6.8) can be decoupled into the individual systems

$$(A - \rho_i I)z_i = x_i \tag{6.10}$$

for each column x_i of X_v and z_i of Z_v, respectively. Since these systems are tridiagonal, they each require $O(n)$ flops.

We should point out that the preliminary reduction of A to tridiagonal form requires $O(n^3)$ flops and becomes therefore the most time-consuming step of the procedure. A more economical idea would be to reduce A to a banded matrix with bandwidth $2k + 1$. Such a matrix can also be viewed as a block tridiagonal matrix with $k \times k$ blocks. The complexity of obtaining such a form is comparable to that of obtaining a "scalar" tridiagonal form, but it can be computed more efficiently on parallel architectures [6]. Moreover, if $k^2 \simeq p$, then steps (6.5) and (6.8) both still require $O(np^2)$ flops.

In many applications (e.g., PDEs), the matrix A has a special sparsity pattern that can also be exploited. For example, one often encounters matrices A that already have a banded form and therefore do not need a preliminary reduction. Other forms of A allow cheap solutions for the systems (6.10) although A does not have a banded form. These include, e.g., low-rank modifications of banded matrices or matrices with special structure such as Toeplitz or Hankel matrices. Finally, we point out that in many applications A is sparse and should therefore no longer be transformed to a condensed form. Instead, one can use iterative solvers to find a solution z_i to (6.10) for each right-hand side x_i. It turns out [10] that a high relative accuracy of (6.10) is only needed in the last few steps of the iteration and hence that a lot of flexibility can be built into the iterative procedure.

Our recommendation is thus to use this algorithm on a block tridiagonal form of A when A is dense and not to reduce A at all when A is sparse, but rather to use flexible iterative solvers for the equations (6.10).

6.2. Handling Singularities. The Sylvester equation (6.1) has a unique solution except when the operator

$$\boldsymbol{T}_X : Z \mapsto AZ - ZX^T AX$$

is singular, which happens when A and $R_A(X)$ have a common eigenvalue, that is, when some Ritz value ρ_i is also an eigenvalue of A. In this nongeneric situation, one could take advantage of the knowledge of the eigenvalue ρ_i of A and solve for the kernel of $(A - \rho_i I)$. Another possibility is to dismiss the singularity by slightly perturbing ρ_i (see Remark 5.4). In any case, the conditioning of systems (6.10) is doomed to deteriorate during the iterations. This characteristic feature of shifted inverse iterations is discussed in [16].

6.3. Stopping Criterion. For the stopping criterion of the algorithm, one has to detect whether the solution Z of (6.1) is such that $\lfloor Z \rfloor$ is nearly equal to $\lfloor X \rfloor$. A possibility is to rely on $\operatorname{dist}(\lfloor X \rfloor, \lfloor Z \rfloor)$ defined in section 3. From Theorem 3.1,

$$\operatorname{dist}(\lfloor X \rfloor, \lfloor Z \rfloor) = \sin \operatorname{atan} \|K_Z\|_2,$$

and

$$\|K\|_2 = \|X_\perp K\|_2 = \|X_\perp X_\perp^T Z(X^T Z)^{-1}\|_2 = \|(I - XX^T)Z(X^T Z)^{-1}\|_2,$$

the computation of which requires $O(np^2)$ flops. One can also use a criterion based on the "block-diagonality" of $(X|X_\perp)^T A(X|X_\perp)$ by checking the value of $\|X_\perp^T AX\|$, i.e., the norm of the residual *Res* (see section 5.2). When A is suitably condensed, the computation requires $O(np^2)$ flops.

6.4. Comparison with Independent RQIs. The practical implementation (6.5)–(6.9) remains a realization of Algorithm 4.2 if we redefine (6.9) as

$$X_+ = \mathrm{qf}(Z_v). \tag{6.11}$$

In fact, if we are not interested in the X's generated by Algorithm 1.1 but only in the $\lfloor X \rfloor$'s, we can content ourselves with an implementation of Algorithm 4.2, which can be as follows.

ALGORITHM 6.1 (implementation of GRQI). *Pick* $Y_{(0)} \in \mathrm{ST}(p,n)$. *Then, for* $k = 0, 1, 2, \ldots$:

1. *Orthonormalize the columns of* $Y_{(k)}$ *to get* $X \in \mathrm{St}(p,n)$.
2. *Compute* $\hat{A} := X^T AX$.
3. *Find* V *such that* $V^T \hat{A} V = \mathrm{diag}\{\rho_1, \ldots, \rho_p\}$.
4. $X_v := XV$.
5. *Find* Z *such that* $AZ - Z\mathrm{diag}\{\rho_1, \ldots, \rho_p\} = X_v$. *If this system is singular, dismiss the singularity by slightly perturbing the* ρ_i*'s.*
6. $Y_{(k+1)} := Z$.

Algorithm 6.1 shows that the GRQI step can be interpreted as p classical RQI steps in parallel preceded by an orthonormalization step. Because this orthonormalization is expensive ($O(np^2)$), one may prefer to skip this step and to define $x_{i+} = z_i$ after (6.10), which would result in p independent RQIs. This simplification of the GRQI has two disadvantages. First, the local convergence property is lost. Indeed, there exist starting points arbitrarily close to certain invariant subspaces for which parallel evolution of p RQIs results in convergence to a different invariant subspace [3]. Second, parallel evolution of p RQIs without orthogonalization is very prone to column merging in the course of the iterations.

One can imagine implementing only a few steps of Algorithm 4.2 and then switching to p independent RQIs. It is difficult, however, to evaluate the right time to switch from one strategy to the other, because there is no bound on Ritz vectors [15]. If Algorithm 4.2 is abandoned too early, then the independent RQIs are likely to merge or to behave oddly as described above. Moreover, the convergence of GRQI is so fast in the neighborhood of an invariant subspace that one hardly has the time to make a decision.

7. Comparison with Other Grassmannian Methods. In [5], Demmel compared three methods for refining estimates of invariant subspaces. These methods are based on the fact that $\lfloor W + W_\perp K \rfloor$ is an invariant subspace of A if and only if the Riccati equation

$$A_{22}K - KA_{11} = -A_{21} + KA_{12}K \tag{7.1}$$

is satisfied, where $(W|W_\perp)$ is an orthogonal basis and

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} W^T \\ W_\perp^T \end{pmatrix} A \, (W|W_\perp). \tag{7.2}$$

Equation (7.1) can be solved by the iteration

$$A_{22}K_{i+1} - K_{i+1}A_{11} = -A_{21} + K_i A_{12} K_i, \quad K_0 = 0, \tag{7.3}$$

Table 7.1 *Comparison of methods for invariant subspace computation.*

Method	Order of convergence	Numerical cost of one iteration
GRQI	Cubic	$O(np^2)$ (A condensed)
Newton on $\mathrm{Gr}(p,n)$	Cubic	$O(n^3)$ (even for A condensed)
Iteration (7.4)	Quadratic	$O(n^3)$ (even for A condensed)
Iteration (7.3)	Linear	$O(np^2)$ (A condensed)

used in [20, 7], or by the Newton–Raphson method used in [4], which yields

$$(A_{22} - K_i A_{12})K_{i+1} - K_{i+1}(A_{11} + A_{12}K_i) = -A_{21} - K_i A_{12} K_i. \tag{7.4}$$

Note that the A_{ij}'s remain constant in the course of the iteration, since W is fixed. Iteration (7.3) converges linearly and costs $O(np^2)$ per iteration after reduction of A to condensed form. The Newton–Raphson method (7.4) converges quadratically, even for symmetric A's. It cannot reach cubic convergence for symmetric A because working on the fixed affine space $\{W + W_\perp K : K \in \mathbb{R}^{(n-p)\times p}\}$ destroys the symmetry of the problem. Each step of (7.4) costs $O(n^3)$ because matrices $(A_{22} - K_i A_{12})$ and $(A_{11} + A_{12}K_i)$ vary from step to step.

In [9], Edelman, Arias, and Smith proposed a Newton method on the Grassmann manifold for the cost $\rho_A(\mathcal{Y})$ defined in (4.2). If $\lfloor X \rfloor$ is the current iterate with X orthonormal, then the next iterate is $\lfloor X + N \rfloor$, where N is orthogonal to X (i.e., $X^T N = 0$) and solves

$$(I - XX^T)(-NX^T AX + AN) = (I - XX^T)AX. \tag{7.5}$$

Using $N = X_\perp H$ and the notation

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} X^T \\ X_\perp^T \end{pmatrix} A (X | X_\perp), \tag{7.6}$$

equation (7.5) reads

$$A_{22}H - HA_{11} = -A_{21}. \tag{7.7}$$

The method converges cubically thanks to the symmetry of ρ_A: if $\lfloor V \rfloor$ is an invariant subspace of A, then $\rho_A(\lfloor V + V_\perp K \rfloor) = \rho_A(\lfloor V - V_\perp K \rfloor)$. On the other hand, since A_{22} is modified in each iteration, one cannot take advantage of a preliminary reduction of A, and the cost per iteration is thus $O(n^3)$.

In comparison, the GRQI (Algorithm 1.1) achieves cubic convergence with a cost of $O(np^2)$ per iteration when A is in condensed form (see section 6). Algorithm 1.1 is faster and cheaper than the methods of [5] and [9].

It is worthwhile to note that Algorithm 1.1 is explicitly mentioned in section 4.8 of Edelman, Arias, and Smith [9] in relation to Newton's method. However, the GRQI and Newton's method are not equivalent, as we now show (this fact is also observed in numerical simulations). Let us use the same notations as above for A_{ij} along with

$$Z = XZ_1 + X_\perp Z_2, \tag{7.8}$$

$$K = Z_2 Z_1^{-1}. \tag{7.9}$$

The GRQI maps $\lfloor X \rfloor$ to

$$\lfloor Z \rfloor = \lfloor XZ_1 + X_\perp Z_2 \rfloor = \lfloor X + X_\perp K \rfloor, \tag{7.10}$$

to be compared with $\lfloor X + X_\perp H \rfloor$ in the case of the Newton step. The equation for K is obtained from (1.1); i.e.,

$$\begin{cases} A_{11}Z_1 + A_{12}Z_2 - Z_1A_{11} = I_p, \\ A_{21}Z_1 + A_{22}Z_2 - Z_2A_{11} = 0, \end{cases} \tag{7.11}$$

which implies

$$A_{21} + A_{22}K - KZ_1A_{11}Z_1^{-1} = 0 \tag{7.12}$$

or equivalently

$$A_{21} + A_{22}K - KA_{11} = -K(A_{11}Z_1 - Z_1A_{11})Z_1^{-1}. \tag{7.13}$$

Comparing the definitions of H in (7.7) and K in (7.13) shows that Newton and GRQI are not equivalent unless $A_{11} = X^TAX$ and $Z_1 = X^TZ$ commute, which is not the case in general. Note that when $p = 1$, A_{11} and Z_1 commute, since they are scalars. In this case, Newton (with tangent update) is equivalent to GRQI. This fact was pointed out in a number of earlier works [16, 19].

8. Conclusion. We have shown how the RQI can be extended to the Grassmann manifold $\mathrm{Gr}(p, n)$ in view of computing a p-dimensional invariant subspace of a symmetric matrix A. Taking advantage of the geometry of $\mathrm{Gr}(p, n)$ and of some properties of Sylvester equations, we have shown cubic convergence of the algorithm. Comparison with other methods for computing invariant subspaces [20, 7, 4, 9] leads to the conclusion that the GRQI achieves a higher rate of convergence at a lower numerical cost.

Even though the algorithm presented in this paper has a number of appealing features, we do not claim that it should replace other eigenvalue solvers. For large sparse matrices, there are other well-known methods such as the Jacobi–Davidson method or other Krylov-based methods which do not require a preliminary reduction to a condensed form. Even when a condensed form is available, one can use other eigenvalue solvers such as inverse iteration or divide and conquer techniques. Each of these methods are well suited to particular types of problems, i.e., the type of matrix that is given or the number and location of the requested eigenvalues (see [1] for a survey).

The purpose of our paper is not to present a robust and efficient eigenvalue solver, but rather to show the existence of a cubically convergent subspace method and to show that its implementation can be quite efficient. In the future, such a technique may become a valuable component in a flexible eigenvalue solver.

REFERENCES

[1] Z. Bai, J. Demmel, J. Dongarra, A. Ruhe, and H. van der Vorst, eds., *Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide*, SIAM, Philadelphia, 2000.

[2] R. H. Bartels and G. W. Stewart, *Algorithm* 432: *Solution of the matrix equation* $AX + XB = C$, Comm. ACM, 15 (1972), pp. 820–826.

[3] S. Batterson and J. Smillie, *The dynamics of Rayleigh quotient iteration*, SIAM J. Numer. Anal., 26 (1989), pp. 624–636.

[4] F. Chatelin, *Simultaneous Newton's iteration for the eigenproblem*, in Defect Correction Methods, Comput. Suppl. 5, Springer-Verlag, Vienna, 1984, pp. 67–74.

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

[6] J. J. Dongarra, I. S. Duff, D. C. Sorensen, and H. A. van der Vorst, *Solving Linear Systems on Vector and Shared Memory Computers*, SIAM, Philadelphia, 1991.
[7] J. J. Dongarra, C. B. Moler, and J. H. Wilkinson, *Improving the accuracy of computed eigenvalues and eigenvectors*, SIAM J. Numer. Anal., 20 (1983), pp. 23–45.
[8] B. F. Doolin and C. F. Martin, *Introduction to Differential Geometry for Engineers*, Monogr. Textbooks Pure Appl. Math. 136, Marcel Dekker, New York, 1990.
[9] A. Edelman, T. A. Arias, and S. T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.
[10] S. Eisenstat and H. F. Walker, *Choosing the forcing terms in an inexact Newton method*, SIAM J. Sci. Comput., 17 (1996), pp. 16–32.
[11] F. R. Gantmacher, *The Theory of Matrices* I, II, Chelsea, New York, 1959.
[12] G. H. Golub, S. Nash, and C. F. V. Loan, *A Hessenberg-Schur method for the problem* $AX + XB = C$, IEEE Trans. Automat. Control, 24 (1979), pp. 909–913.
[13] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*, Springer-Verlag, New York, 1994.
[14] B. N. Parlett, *The Rayleigh quotient iteration and some generalizations for nonnormal matrices*, Math. Comp., 28 (1974), pp. 679–693.
[15] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Classics Appl. Math. 20, SIAM, Philadelphia, 1998.
[16] G. Peters and J. H. Wilkinson, *Inverse iteration, ill-conditioned equations and Newton's method*, SIAM Rev., 21 (1979), pp. 339–360.
[17] H. R. Rutishauser, *Computational aspects of F. L. Bauer's simultaneous iteration method*, Numer. Math., 13 (1969), pp. 4–13.
[18] H. R. Rutishauser, *Simultaneous iteration method for symmetric matrices*, Numer. Math., 16 (1970), pp. 205–223.
[19] S. T. Smith, *Optimization techniques on Riemannian manifolds*, in Hamiltonian and Gradient Flows, Algorithms and Control, A. Bloch, ed., Fields Institute Communication 3, AMS, Providence, RI, 1994, pp. 113–136.
[20] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Rev., 15 (1973), pp. 727–764.
[21] P. Van Dooren and M. Verhaegen, *On the use of unitary state-space transformations*, in Linear Algebra and Its Role in Linear Systems Theory, B. N. Datta, ed., Contemp. Math. 47, AMS, Providence, RI, 1985, pp. 447–463.
[22] D. S. Watkins, *Understanding the QR algorithm*, SIAM Rev., 24 (1982), pp. 427–440.