ALTERNATIVES TO THE RAYLEIGH QUOTIENT FOR THE QUADRATIC EIGENVALUE PROBLEM

MICHIEL E. HOCHSTENBACH* AND HENK A. VAN DER VORST*

Abstract. We consider the quadratic eigenvalue problem $\lambda^2 Ax + \lambda Bx + Cx = 0$. Suppose that u is an approximation to an eigenvector x (for instance obtained by a subspace method), and that we want to determine an approximation to the corresponding eigenvalue λ. The usual approach is to impose the Galerkin condition $r(\theta, u) = (\theta^2 A + \theta B + C)u \perp u$ from which it follows that θ must be one of the two solutions to the quadratic equation $(u^*Au)\theta^2 + (u^*Bu)\theta + (u^*Cu) = 0$. An unnatural aspect is that if $u = x$, the second solution has in general no meaning. When u is not very accurate, it may not be clear which solution is the best. Moreover, when the discriminant of the equation is small, the solutions may be very sensitive to perturbations in u.

In this paper we therefore examine alternative approximations to λ. We compare the approaches theoretically and by numerical experiments. The methods are extended to approximations from subspaces and to the polynomial eigenvalue problem.

Key words. Quadratic eigenvalue problem, Rayleigh quotient, Galerkin, minimum residual, subspace methods, polynomial eigenvalue problem, backward error, refined Ritz vector.

AMS subject classifications. 65F15.

1. Introduction. First consider the eigenvalue problem $Ax = \lambda x$, with A a real symmetric $n \times n$ matrix. Suppose that we have an approximate eigenvector u. The usual approximation to the corresponding eigenvalue is given by the *Rayleigh quotient* of u

$$\rho = \rho(u) := \frac{u^*Au}{u^*u}. \tag{1.1}$$

This Rayleigh quotient has the following attractive properties:

1. ρ satisfies the *Ritz-Galerkin condition* on the *residual* $r(\theta, u)$:

$$r(\rho, u) := Au - \rho u \perp u. \tag{1.2}$$

2. ρ satisfies the *minimum residual condition* on the residual

$$\rho = \mathrm{argmin}_{\theta \in \mathbb{R}} \|Au - \theta u\|. \tag{1.3}$$

(Here and elsewhere in the paper, $\|\cdot\|$ stands for $\|\cdot\|_2$).

3. The function $\rho(u)$ has as its *stationary points* exactly the n eigenvectors x_i, so

$$\frac{d\rho}{du}(x_i) = 0. \tag{1.4}$$

(Recall that stationary means that all directional derivatives are zero.) This implies that a first order perturbation of the eigenvector only gives a second order perturbation of the Rayleigh quotient: $\rho(x_i + h) = \rho(x_i) + \mathcal{O}(\|h\|^2)$.

REMARK 1.1. *When A is nonsymmetric, (1.2) and (1.3) still hold, but (1.4) fails to hold. One can show that instead of this $\rho(u, v) := \frac{v^*Au}{v^*u}$ has as its stationary points exactly the right/left eigenvectors combinations (x_i, y_i). This suggests to replace the Ritz-Galerkin condition (1.2) by the* Petrov-Galerkin condition

$$r(\theta, u) = Au - \theta u \perp v,$$

which is used in two-sided methods such as two-sided Lanczos [4] and two-sided Jacobi–Davidson [2]. However, in this paper we assume that we have no information about the left eigenvector.

Now consider the quadratic eigenvalue problem

$$Q(\lambda)x := \lambda^2 Ax + \lambda Bx + Cx = 0, \tag{1.5}$$

*Mathematical Institute, Utrecht University, P.O. Box 80.010, NL-3508 TA Utrecht, The Netherlands, `www.math.uu.nl/people/[hochstenbach, vorst]`, `[hochstenbach, vorst]@math.uu.nl`. November 2001 where A, B, and C are (complex) $n \times n$ matrices. In this paper, we examine generalizations of the properties (1.2)–(1.4) for the quadratic eigenvalue problem, to derive different eigenvalue approximations. See [5] for a nice overview of the quadratic eigenvalue problem. For an eigenvector x we have either one of the following properties:

- Ax and Bx are dependent, then Cx is also dependent, and there are two eigenvalues (counting multiplicities) corresponding to x,
- Ax and Bx are independent, Cx lies in the span of Ax and Bx, and the corresponding eigenvalue λ is unique.

We will assume in the remainder of the paper that x has the second property. For a motivation see Remark 2.2 at the end of Section 2.4.

Now let u be an approximation to an eigenvector x, for instance one obtained by a subspace method. We will also assume that Au and Bu are independent, which is not unnatural in view of the assumptions that Ax and Bx are independent, and $u \approx x$; see also Remark 2.2. We study ways to determine an approximation θ to the eigenvalue λ, from the information of u. In Section 2.1 we discuss the "classical" one-dimensional Galerkin method, while in Sections 2.2, 2.3, and 2.4 we introduce new approaches. The methods are compared in Section 3 and extended to subspaces of dimension larger than one and to the polynomial eigenvalue problem in Section 4. Numerical experiments and a conclusion can be found in Section 5 and 6.

2. Approximations for the quadratic eigenvalue problem.

2.1. One-dimensional Galerkin. For an approximate eigenpair $(\theta, u) \approx (\lambda, x)$ we define the *residual* $r(\theta, u)$ by

$$r(\theta, u) := Q(\theta)u = (\theta^2 A + \theta B + C)u.$$

The usual approach to derive an approximate eigenvalue θ from the approximate eigenvector u is to impose the Galerkin condition $r(\theta, u) \perp u$. Then it follows that $\theta = \theta(u)$ must be one of the two solutions to the quadratic equation

$$\alpha\theta^2 + \beta\theta + \gamma = 0, \tag{2.1}$$

where $\alpha = \alpha(u) = u^*Au$, $\beta = \beta(u) = u^*Bu$, and $\gamma = \gamma(u) = u^*Cu$. An unnatural aspect is that if $u = x$, the second solution of (2.1) has in general no meaning. If u is close to x, we will be able to decide which one is best by looking at the norms of the residuals. But if u is not very accurate, it may not be clear which solution is the best. This may for instance happen when we try to solve (1.5) by a subspace method; in the beginning of the process, the search space may not contain good approximations to an eigenvector. This problem is also mentioned in [1, p. 282].

Moreover, when the discriminant

$$\delta = \delta(u) := \beta^2 - 4\alpha\gamma \tag{2.2}$$

is small, then the solutions of (2.1) may be very sensitive to perturbations in u (see also Section 3). Thus the second solution of (2.1) is not only useless, but it may also hinder the accuracy of the solution that is of interest!

We therefore examine alternative ways to approximate λ. We generalize the Galerkin property (1.2) and minimum residual property (1.3) for the quadratic eigenvalue problem in the following three subsections. In Section 3 the approaches are compared using a generalization of (1.4).

2.2. Two-dimensional Galerkin. In the standard eigenvalue problem, we deal with two vectors u and Au, which are asymptotically (by which we mean when $u \to x$) dependent. Therefore it is natural to take the length of the projection of Au onto the span of u as an approximation to the eigenvalue, which is exactly what the Rayleigh quotient $\rho(u)$ does. For the generalized eigenvalue problem we have a similar situation.

In the quadratic eigenvalue problem, however, we deal with three vectors Au, Bu, and Cu, which asymptotically lie in a plane. Therefore it is natural to consider the projection of these three vectors onto a certain plane, spanned by two independent vectors p and q. To generalize the approach of (1.2), define the *generalized residual* $r(\mu, \nu, u)$ by

$$r(\mu, \nu, u) := (\mu A + \nu B + C)u. \tag{2.3}$$

The idea behind this is that we want to impose conditions on r such that μ forms an approximation to λ^2, and ν an approximation to λ. Then both μ/ν and ν may be good approximations to the eigenvalue λ. A generalization of (1.2) is obtained by imposing two Galerkin conditions $r(\mu, \nu, u) \perp p$ and $r(\mu, \nu, u) \perp q$ for specific independent vectors p, q. This leads to the system

$$W^* Z \begin{bmatrix} \mu \\ \nu \end{bmatrix} = -W^* Cu, \quad \text{where} \quad W = \begin{bmatrix} p & q \end{bmatrix}, \quad Z = \begin{bmatrix} Au & Bu \end{bmatrix}. \tag{2.4}$$

When W^*Z is nonsingular, (2.4) defines a unique μ and ν. A logical choice for p and q is any linear combination of Au, Bu, and Cu. Specifically, one could take the "least-squares" plane such that

$$\|(I - \Pi)Au\|^2 + \|(I - \Pi)Bu\|^2 + \|(I - \Pi)Cu\|^2$$

is minimal, where Π is the orthogonal projection onto the plane. Let z be the normal of the sought plane, then one may verify that $\|(I - \Pi)Au\|^2 = \|(z^*Au)z\|^2 = |z^*Au|^2$. If D denotes the $n \times 3$ matrix with Au, Bu, and Cu as its columns, then z is the vector of unit length such that $\|z^*D\|^2$ is minimal. So we conclude that z is the minimal left singular vector of D, and for p and q we can take the two "largest" left singular vectors. Another choice for p and q as well as its meaning are discussed in Section 2.4.

2.3. One-dimensional minimum residual. Two other approaches, discussed in this and the following subsection, generalize the minimum residual approach (1.3). First, we can minimize the norm of the residual with respect to θ:

$$\min_{\theta \in \mathbb{C}} \|(\theta^2 A + \theta B + C)u\|. \tag{2.5}$$

For complex θ, differentiating the square of (2.5) with respect to $\mathrm{Re}(\theta)$ and $\mathrm{Im}(\theta)$ gives two mixed equations of degree three in $\mathrm{Re}(\theta)$ and $\mathrm{Im}(\theta)$, or an equation (the so-called *resultant*) of degree nine in only $\mathrm{Re}(\theta)$ or $\mathrm{Im}(\theta)$ (see Section 5). Of course, only the real solutions of these equations are of interest. We may solve the equations numerically (see the numerical experiments in Section 5). In the special case that we know that λ is real, we would like to have a real approximation θ. Then differentiating the square of (2.5) with respect to θ gives the cubic equation with real coefficients

$$4\,\|Au\|^2\theta^3 + 6\,\mathrm{Re}((Au)^*Bu)\theta^2 + 2\,(\|Bu\|^2 + 2\,\mathrm{Re}((Cu)^*Au)\theta + 2\,\mathrm{Re}((Cu)^*Bu) = 0, \tag{2.6}$$

which can be solved analytically. This is for instance the case for the important class of *quasi-hyperbolic quadratic eigenvalue problems*:

DEFINITION 2.1. *(Cf. [5, p. 257]) A quadratic eigenvalue problem $Q(\lambda)x = 0$ is called* quasi-hyperbolic *if A is Hermitian positive definite, B and C are Hermitian, and for all eigenvectors of $Q(\lambda)$ we have*

$$(x^*Bx)^2 > 4(x^*Ax)(x^*Cx).$$

It is easy to see that all eigenvalues of quasi-hyperbolic quadratic eigenvalue problems are real.

In the next subsection we will also discuss a suboptimal solution of (2.5) that involves the solution of a resultant equation of degree five instead of nine.

2.4. Two-dimensional minimum residual. Another idea is to minimize the norm of the generalized residual (2.3) with respect to μ, ν:

$$(\mu_*, \nu_*) = \mathrm{argmin}_{(\mu,\nu)\in\mathbb{C}^2} \|(\mu A + \nu B + C)u\|. \tag{2.7}$$

To solve this, consider the corresponding overdetermined $n \times 2$ linear system

$$Z \begin{bmatrix} \mu \\ \nu \end{bmatrix} = -Cu,$$

with Z as in (2.4). By assumption Au and Bu are independent, so μ_* and ν_* are uniquely determined by

$$\begin{bmatrix} \mu_* \\ \nu_* \end{bmatrix} := -Z^+ Cu = -\left(Z^* Z\right)^{-1} Z^* Cu,$$

where Z^+ denotes the pseudoinverse of Z. We see that (2.7) is a special case of (2.4), namely the case where we choose $p = Au$ and $q = Bu$, so $W = Z$.

Returning to (2.5), we can define a suboptimal solution by solving for $\theta \in \mathbb{C}$ such that

$$\left\| \begin{bmatrix} \theta^2 \\ \theta \end{bmatrix} - \begin{bmatrix} \mu_* \\ \nu_* \end{bmatrix} \right\| \tag{2.8}$$

is minimal. Differentiating the square of (2.8) with respect to $\mathrm{Re}(\theta)$ and $\mathrm{Im}(\theta)$ gives two mixed equations of degree three in $\mathrm{Re}(\theta)$ and $\mathrm{Im}(\theta)$, or a resultant equation of degree five in only $\mathrm{Re}(\theta)$ or $\mathrm{Im}(\theta)$ (see Section 5); compare this with the resultant of degree nine for the optimal solution.

The following remark explains why we assumed in Section 2 that both of the pairs Ax and Bx, and Au and Bu are independent.

REMARK 2.2. *When Au and Bu are dependent, then the one-dimensional minimum residual approach reduces to the one-dimensional Galerkin approach, while the two-dimensional methods are not uniquely determined. When Ax and Bx are dependent, then, though the approaches may be uniquely determined, the results may be bad. For example, the matrix Z in the two-dimensional methods is ill-conditioned if u is a good approximation to x.*

3. Comparison of the methods. Concerning the cost, all methods require three matrix-vector multiplications (Au, Bu, and Cu) and additionally $\mathcal{O}(n)$ time. In this section, we compare the quality of the methods by two different means. First, we investigate the influence of perturbations of u to θ, and then we examine backward errors.

A nice property that an approximate eigenvalue can (or should) have is that it is close to the eigenvalue if the corresponding approximate eigenvector is close to the eigenvector. In other words, we like the situation where

$$|\theta(x+h) - \lambda| = |\theta(x+h) - \theta(x)| \quad \text{is small}$$

for small $\|h\|$. When θ is differentiable with respect to u in the point x this is equivalent to the condition

$$\left\| \frac{\partial \theta}{\partial u}(x) \right\| \quad \text{is small.} \tag{3.1}$$

We now examine the four approaches from the previous section with this criterion, starting with the one-dimensional Galerkin approach. Equation (2.1) defines θ implicitly as a function of α, β, and γ, say $f(\theta, \alpha, \beta, \gamma) = 0$, with $f(\lambda, \alpha(x), \beta(x), \gamma(x)) = 0$. When $\delta(x) \neq 0$, the Implicit Function Theorem states that locally θ is a function of α, β, and γ, say $\theta = \varphi(\alpha, \beta, \gamma)$, and that

$$\begin{aligned} D\varphi(\alpha(x), \beta(x), \gamma(x)) &= -((D_\theta f)^{-1} D_{(\alpha,\beta,\gamma)} f)(\lambda, \alpha(x), \beta(x), \gamma(x)) \\ &= \pm \tfrac{1}{\sqrt{\delta(x)}} \cdot (\lambda^2, \lambda, 1). \end{aligned}$$

So when δ is small, which means that (2.1) has two roots that are close, we may expect that $|\theta - \lambda|$ is large for small perturbations of x, see the numerical experiments.

REMARK 3.1. *As in the standard eigenvalue problem, $\theta = \theta(u, v)$ as solution of*

$$(v^* Au)\theta^2 + (v^* Bu)\theta + (v^* Cu) = 0$$

is stationary in the right/left eigenvector combinations (x_i, y_i). *We assume, however, that we have no information about the (approximate) left eigenvector.*

Now consider the two-dimensional Galerkin method (2.4), and the two-dimensional minimum residual method (2.7). In both cases, ν and μ/ν can be taken as approximation to λ. By differentiating (2.4), it can be seen that $\frac{\partial\mu}{\partial u}(x) = w_1^* Q(\lambda)$ and $\frac{\partial\nu}{\partial u}(x) = w_2^* Q(\lambda)$, where w_1, w_2 are certain linear combinations of the vectors p and q that span the plane of projection. From Section 2 it is clear that the plane for the two-dimensional Galerkin method is contained in $\mathrm{span}\{Au, Bu, Cu\}$, while the plane for the two-dimensional minimum residual method is $\mathrm{span}\{Au, Bu\}$. Since $\mathrm{span}\{Ax, Bx, Cx\} = \mathrm{span}\{Ax, Bx\}$, we conclude that $\frac{\partial\mu}{\partial u}(x)$ and $\frac{\partial\nu}{\partial u}(x)$ are the same for both two-dimensional methods.

For the second approximation, μ/ν, we have that

$$\frac{\partial(\mu/\nu)}{\partial u}(x) = \frac{1}{\lambda} \cdot \frac{\partial\mu}{\partial u}(x) - \frac{\partial\nu}{\partial u}(x). \tag{3.2}$$

This suggests that μ/ν might give inaccurate approximations for small λ, which is confirmed by numerical experiments, see Experiment 5.1.

The effects of perturbations of u for the results of the one-dimensional minimum residual approach is hard to analyze: amongst other things it depends upon the position of the zeros of the polynomials (see (5.1)).

A second interesting tool to compare the methods of Section 2 is the notion of the *backward error.*

DEFINITION 3.2. *(Cf. [7]) The backward error of an approximate eigenpair* (θ, u) *of* Q *is defined as*

$$\begin{aligned}\eta(\theta, u) := \min\{\varepsilon :\ & (\theta^2(A + \Delta A) + \theta(B + \Delta B) + (C + \Delta C))u = 0,\\ & \|\Delta A\| \le \varepsilon\zeta_1, \|\Delta B\| \le \varepsilon\zeta_2, \|\Delta C\| \le \varepsilon\zeta_3\ \}.\end{aligned}$$

The backward error of an approximate eigenvalue θ *of* Q *is defined as*

$$\eta(\theta) := \min_{\|u\|=1} \eta(\theta, u).$$

In [7, Theorems 1 and 2], the following results are proven:

$$\eta(\theta, u) = \frac{\|r\|}{\zeta_1|\theta|^2 + \zeta_2 \cdot |\theta| + \zeta_3}, \qquad \eta(\theta) = \frac{\sigma_{\min}(Q(\theta))}{\zeta_1|\theta|^2 + \zeta_2 \cdot |\theta| + \zeta_3}. \tag{3.3}$$

In the numerical experiments we therefore examine the quality of the computed θ by examining $\|r\|$ and $\sigma_{\min}(Q(\theta))$, which, for convenience, are also called backward errors. Note that the backward errors are related: $\sigma_{\min}(Q(\theta)) \le \|r\|$.

4. Extensions.

4.1. Approximations from subspaces. We can also use the techniques described in Section 2 for approximations to eigenpairs from subspaces of dimension larger than one. Let $\mathcal{U}$ be a k-dimensional subspace, where for subspace methods one typically has $k \ll n$, and let the columns of U form a basis for $\mathcal{U}$. The Ritz-Galerkin condition

$$\theta^2 Au + \theta Bu + Cu \perp \mathcal{U}, \qquad u \in \mathcal{U},$$

leads, with the substitution $u = Us$, to the projected quadratic eigenvalue problem

$$(\theta^2 U^*AU + \theta U^*BU + U^*CU)s = 0, \tag{4.1}$$

which in general yields $2k$ Ritz pairs (θ, u). For a specific pair, one can "refine" the value θ by the methods of Section 2. Although it is not guaranteed that the new $\widetilde{\theta}$ is better, it seems to be often the case, see the numerical experiments. Moreover, we have knowledge of the backward error, which we will discuss in a moment.

Then, as a second step, one can "refine" the vector u by taking $\widetilde{u} = U\widetilde{s}$, where

$$\widetilde{s} = \text{the "smallest" right singular vector of } \quad \widetilde{\theta}^2 AU + \widetilde{\theta} BU + CU$$

(For the Arnoldi method for the standard eigenvalue problem, a similar refinement of a Ritz vector has been proposed in [3].) This step is relatively cheap, because all matrices are "skinny". Given $\widetilde{\theta}$, the vector $\widetilde{u}$ minimizes the backward error $\eta(\widetilde{\theta}, u)$, see (3.3). It is also possible to repeat these two steps to get better and better approximations, leading to Algorithm 4.1.

Input: a subspace $\mathcal{U}$
Output: an approximate eigenpair (θ, u) with $u \in \mathcal{U}$
1. Compute an approximate eigenpair (θ, u) according to the standard Ritz-Galerkin method
   for $k = 1, 2, \ldots$
2. Compute a new θ_k choosing one of the methods of Section 2
3. Compute the "smallest" singular vector s_k of $\theta_k^2 AU + \theta_k BU + CU$
4. $u_k = U s_k$

ALG. 4.1. *Refinement of an approximate eigenpair* (θ, u).

During this algorithm, we do not know the (forward) error $|\theta_k - \lambda|$, but the backward errors $\|r\|$ and $\sigma_{\min}(\theta_k^2 AU + \theta_k BU + CU)$ are cheaply available; they can be used to decide whether or not to continue the algorithm. When we take the optimal one-dimensional minimum residual method in each step, we are certain that the backward error $\|r\|$ decreases monotonically. In Experiment 5.3 we use the two-dimensional Galerkin approach in every step.

REMARK 4.1. *For the symmetric eigenvalue problem, the possibility of an iterative procedure to minimize $\|Au - \rho(u)u\|$ over the subspace $\mathcal{U}$ is mentioned in [6], in the context of finding inclusion intervals for eigenvalues. Moreover, a relation between the minimalization of $\|Au - \rho(u)u\|$ and the smallest possible Lehmann interval is given.*

4.2. The polynomial eigenvalue problem. Consider the polynomial eigenvalue problem

$$(\lambda^l A_l + \lambda^{l-1} A_{l-1} + \cdots + \lambda A_1 + A_0)x = 0.$$

Define the generalized residual as

$$r(\mu_1, \ldots, \mu_l, u) := (\mu_l A_l + \mu_{l-1} A_{l-1} + \cdots + \mu_1 A_1 + A_0)u.$$

Both the l-dimensional Galerkin method

$$r(\mu_1, \ldots, \mu_l, u) \perp \{p_1, \ldots, p_l\}$$

and the l-dimensional minimum residual method

$$\min_{\mu_1, \ldots, \mu_l} \|r(\mu_1, \ldots, \mu_l, u)\|$$

lead to a system of the form

$$W^* Z \begin{bmatrix} \mu_l \\ \vdots \\ \mu_1 \end{bmatrix} = -W^* A_0 u, \tag{4.2}$$

where $Z = [\ A_l u \quad \cdots \quad A_1 u\ ]$. For the l-dimensional minimum residual method we have $W = Z$; for the l-dimensional Galerkin approach with "least-squares" l-dimensional plane, W consists of the l largest left singular vectors of $[\ Z \quad A_0 u\ ]$. Assuming that the vectors $A_1 u, \ldots, A_l u$ are independent, (4.2) has a unique solution. In principle we can try every quotient μ_l/μ_{l-1}, μ_{l-1}/μ_{l-2}, $\ldots$, μ_2/μ_1, μ_1, and also some other combinations like $\mu_l/(\mu_{l-2}\mu_1)$, as an approximation to λ. When λ is small, μ_1 will probably be the best. The one-dimensional minimum residual approach is less attractive, as the degree of the associated polynomials (cf. (2.6) and (5.1)) increases fast.

5. Numerical experiments. The experiments are carried out in Matlab and Maple. First a word on solving (2.5) for the optimal, and (2.8) for the suboptimal one-dimensional minimum residual approach. Write $\theta = \theta_1 + i\theta_2$, $\mu_* = \mu_1 + i\mu_2$, and $\nu_* = \nu_1 + i\nu_2$. Differentiating the square of (2.5) with respect to θ_1 and θ_2 leads to two mixed equations (in θ_1 and θ_2) of degree three. With Maple the equations are manipulated so that we have two equations of degree nine in θ_1 or θ_2 only, which are called the resultants. When we know that λ is real, then we get the cubic equation (2.6).

Differentiation of (2.8) with respect to θ_1 and θ_2, leads to

$$\left\{ \begin{array}{rcl} \theta_1^3 + (\theta_2^2 - \mu_1 + \frac{1}{2})\theta_1 - \mu_2\theta_2 - \frac{1}{2}\nu_1 & = & 0, \\ \theta_2^3 + (\theta_1^2 + \mu_1 + \frac{1}{2})\theta_2 - \mu_2\theta_1 - \frac{1}{2}\nu_2 & = & 0. \end{array} \right. \tag{5.1}$$

Because of the missing θ_1^2 and θ_2^2 terms, in the first and second equation respectively, the corresponding resultants have degree only five. All equations were solved numerically by a Maple command of the form

`solve(resultant(` $equation_1$ `(x, y),` $equation_2$ `(x, y) , y), x).`

Of course, we only have to solve one resultant, say for $\mathrm{Re}(\theta)$, then $\mathrm{Im}(\theta)$ can be solved from a cubic equation. In our experiments, many equations have a unique real solution, making it unnecessary to choose. When there is more than one real solution, we take the one that minimizes the norm of the residual.

EXPERIMENT 5.1. Our first example is taken from [5, p. 250]:

$$A = \begin{bmatrix} 0 & 6 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & -6 & 0 \\ 2 & -7 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad C = I_3.$$

This problem has two eigenvectors for each of which there exist two eigenvalues: $\begin{bmatrix} 1 & 1 & 0 \end{bmatrix}^T$ corresponds to $\lambda = 1/2$ and $\lambda = 1/3$, while $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$ corresponds to $\lambda = \pm i$. In line with our assumptions, we do not consider these. Instead, we focus on the other eigenpairs $(\lambda, x) = (1, \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T)$ and $(\lambda, x) = (\infty, \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T)$. For the last pair we consider the problem for $\lambda^{-1} = 0$. We simulate the situation of having a good approximation $u \approx x$ by adding a random (complex) perturbation to x:

$$u := x + \varepsilon \cdot w \,/\, \|x + \varepsilon \cdot w\|,$$

where w is a normalized vector of the form `rand(3,1)`$+i\cdot$`rand(3,1)`. (For all experiments, we take "`seed=0`" so that our results are reproducible.) Table 5.2 gives the results of the four approaches for $\varepsilon = 0.01$. The first row of the two-dimensional Galerkin (Gal-2) and two-dimensional minimum residual (MR-2) approaches represents μ/ν, while the second gives ν as approximation to λ. The first row of the one-dimensional minimum residual method (MR-1) represents the optimal solution, while the second is the suboptimal solution. For clarity, the meaning of the different rows is first summarized in Table 5.1.

TABLE 5.1: The rows of Tables 5.2 to 5.4, with their meaning.

row nr.	label	meaning
1	Gal-1	best approximation (of the two) of the one-dimensional Galerkin method
2	Gal-2	μ/ν approximation of the two-dimensional Galerkin method
3		ν approximation of the two-dimensional Galerkin method
4	MR-1	optimal approximation of the one-dimensional minimum residual method
5		suboptimal approximation of the one-dimensional minimum residual method
6	MR-2	μ/ν approximation of the one-dimensional minimum residual method
7		ν approximation of the one-dimensional minimum residual method

For $\lambda = 1$, all other approaches (Gal-2, MR-1, and MR-2) give a smaller (forward) error than the classical one-dimensional Galerkin method (Gal-1). The "ν" approximation of the two-dimensional approaches Gal-2 (row 3) and MR-2 (row 7) is particularly good. The sensitivities

TABLE 5.2: The approximations of the one-dimensional Galerkin (Gal-1), two-dimensional Galerkin (Gal-2, μ/ν and ν), one-dimensional minimum residual (MR-1, optimal and suboptimal), and two-dimensional minimum residual (MR-2, μ/ν and ν) approaches for $\lambda = 1$ and $\lambda^{-1} = 0$. The other columns give the (forward) error $|\theta - \lambda|$, and $\|r\|$ and $\sigma_{\min}(Q(\theta))$ for the backward errors.

Method	appr. for $\lambda = 1$	error	$\|r\|$	$\sigma_{\min}$	appr. for $\lambda^{-1} = 0$	error	$\|r\|$	$\sigma_{\min}$
Gal-1	0.99935-0.00192i	0.00202	0.0112	0.00142	0.00117-0.03956i	0.03958	0.0399	0.0279
Gal-2	0.99947-0.00159i	0.00168	0.0111	0.00118	1.00009-0.00229i	1.00009	2.8285	0.0016
	1.00004	0.00004	0.0181	0.00002	-0.00069-0.01986i	0.01987	0.0206	0.0140
MR-1	0.99942-0.00173i	0.00182	0.0111	0.00128	-0.00036-0.02384i	0.02384	0.0186	0.0168
	0.99970-0.00064i	0.00070	0.0217	0.00049	0.00009-0.01987i	0.01990	0.0206	0.0141
MR-2	0.99946-0.00159i	0.00168	0.0111	0.00119	1.00005-0.00229i	1.00006	2.8284	0.0016
	0.99986	0.00014	0.0178	0.00009	-0.00069-0.01986i	0.01987	0.0206	0.0140

for the two-dimensional approaches $\|\partial\nu/\partial u\| = 0$ and $\|\partial(\mu/\nu)/\partial u\| \approx 0.33$ also indicate this. The suboptimal solution of MR-1 has a larger backward error $\|r\|$, but a smaller forward error than the optimal solution. For the discriminant (2.2) we have $\delta = 25$.

For $\lambda^{-1} = 0$, the "μ/ν" approximations (rows 2 and 6) are bad, which was already predicted by (3.2). The sensitivities are $\|\partial\nu/\partial u\| \approx 3.0$ and $\|\partial(\mu/\nu)/\partial u\| = \infty$, and for the discriminant we have $\delta = 1$.

EXPERIMENT 5.2. For the second example we construct matrices such that the discriminant δ is small and hence the zeros of (2.1) almost coincide. For small $\zeta > 0$ define

$$A = I_3, \qquad B = \begin{bmatrix} 1 & 1 & 0 \\ 0 & -2 & 2 \\ 0 & 0 & 1 \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & -1-\sqrt{\zeta} & 0 \\ 0 & 1-\zeta & 2 \\ 0 & 0 & 1 \end{bmatrix}.$$

One may check that $x = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T$ is an eigenvector with corresponding eigenvalue $1 + \sqrt{\zeta}$. (The second solution $1 - \sqrt{\zeta}$ to (2.1) is close to the eigenvalue, but has no meaning.) The discriminant is equal to 4ζ. We take $\zeta = 10^{-4}$, so $\lambda = 1.01$. We test the approaches for $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-3}$, see Table 5.3.

TABLE 5.3: The approximations of the one-dimensional Galerkin (Gal-1), two-dimensional Galerkin (Gal-2, μ/ν and ν), one-dimensional minimum residual (MR-1, optimal and suboptimal), and two-dimensional minimum residual (MR-2, μ/ν and ν) approaches for $\lambda = 1.01$, for $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-3}$, respectively. The other columns give the (forward) error $|\theta - \lambda|$, and $\|r\|$ and $\sigma_{\min}(Q(\theta))$ for the backward errors.

Method	appr. ($\varepsilon = 10^{-2}$)	error	$\|r\|$	$\sigma_{\min}$	appr. ($\varepsilon = 10^{-3}$)	error	$\|r\|$	$\sigma_{\min}$
Gal-1	1.0317-0.1442i	0.1458	0.1343	0.01326	1.0117-0.0444i	0.0444	0.0431	0.001325
Gal-2	0.9861-0.0249i	0.0345	0.0348	0.00052	1.0076-0.0025i	0.0034	0.0034	0.000036
	1.0050-0.0142i	0.0151	0.0274	0.00018	1.0095-0.0014i	0.0015	0.0027	0.000018
MR-1	1.0046-0.0150i	0.0159	0.0274	0.00020	1.0094-0.0014i	0.0015	0.0027	0.000018
	0.9971-0.0186i	0.0226	0.0287	0.00027	1.0087-0.0018i	0.0022	0.0029	0.000025
MR-2	0.9857-0.0249i	0.0348	0.0350	0.00052	1.0076-0.0025i	0.0034	0.0034	0.000037
	1.0047-0.0142i	0.0152	0.0274	0.00018	1.0095-0.0014i	0.0015	0.0027	0.000018

The sensitivities for the two-dimensional methods Gal-2 and MR-2 are $\|\partial\nu/\partial u\| \approx 3.0$ and $\|\partial(\mu/\nu)/\partial u\| \approx 5.0$, and $|\delta| \approx 4.0 \cdot 10^{-4}$. Because the discriminant is small, and the sensitivities are very modest, it is no surprise that all other approximations are much better (measured in forward or backward error) than Gal-1.

EXPERIMENT 5.3. For the last example we take A, B, and C random symmetric matrices of size 100×100. We try to approximate the eigenvalue $\lambda \approx 7.2288 + 2.7803i$, for $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-4}$, see Table 5.4.

The sensitivities for Gal-2 and MR-2 are $\|\partial\nu/\partial u\| \approx 3.9 \cdot 10^2$ and $\|\partial(\mu/\nu)/\partial u\| \approx 2.0 \cdot 10^2$, and $|\delta| \approx 2.4 \cdot 10^{-5}$. Indeed, we see that the two "μ/ν" approximations (row 2 and 6) are the best, together with the optimal MR-1 solution (row 4). Note that for larger matrices, the computation of $\sigma_{\min}(Q(\theta))$ is expensive. In practice, one does not compute it, but it is shown here to compare the methods.

TABLE 5.4: The approximations of the one-dimensional Galerkin (Gal-1), two-dimensional Galerkin (Gal-2, μ/ν and ν), one-dimensional minimum residual (MR-1, optimal and suboptimal), and two-dimensional minimum residual (MR-2, μ/ν and ν) approaches for $\lambda \approx 7.2288 + 2.7803i$, and $\varepsilon = 10^{-3}$ and $\varepsilon = 10^{-4}$, respectively. The other columns give the (forward) error $|\theta - \lambda|$, and $\|r\|$ and $\sigma_{\min}(Q(\theta))$ for the backward errors.

Method	appr. ($\varepsilon = 10^{-3}$)	error	$\|r\|$	$\sigma_{\min}$	appr. ($\varepsilon = 10^{-4}$)	error	$\|r\|$	$\sigma_{\min}$
Gal-1	6.86+2.71i	0.37	2.88	0.185	7.218+2.738i	0.0428	0.307	0.0221
Gal-2	7.25+2.67i	0.10	2.90	0.053	7.230+2.769i	0.0110	0.290	0.0057
	6.87+3.04i	0.44	3.15	0.233	7.189+2.800i	0.0445	0.329	0.0232
MR-1	7.04+2.61i	0.24	2.81	0.123	7.227+2.769i	0.0107	0.290	0.0055
	5.13+2.08i	2.20	5.23	0.822	7.176+2.772i	0.0529	0.332	0.0247
MR-2	7.23+2.65i	0.12	2.88	0.063	7.230+2.769i	0.0112	0.290	0.0058
	3.66+1.62i	3.74	6.33	0.709	7.123+2.775i	0.1057	0.436	0.0545

Next, we test Algorithm 4.1. We start with a three-dimensional subspace $\mathcal{U}$, consisting of the same vector as above ($\varepsilon = 10^{-3}$), completed by two random (independent) vectors. We determine six Ritz pairs according to (4.1), and refine the one with θ approximating the eigenvalue $\lambda \approx 7.2288 + 2.7803i$ by Algorithm 4.1, where in every step we choose the μ/ν- approximation of the two-dimensional Galerkin method. The results, shown in Table 5.5, reveal that both u and θ are improved four times, after which they keep fixed in the decimals shown. Note that the smallest possible angle of a vector in $\mathcal{U}$ with x is

$$\angle(\mathcal{U}, x) = \angle((I - UU^*)x, x) \approx 6.2809 \cdot 10^{-4}.$$

TABLE 5.5: Refinement of an approximate eigenvalue by Algorithm 4.1 for $\lambda \approx 7.2288+2.7803i$. The columns give the iteration number, angle between u and x, (forward) error $|\theta - \lambda|$, and $\|r\|$, $\sigma_1 := \sigma_{\min}(\theta^2 AU + \theta BU + CU)$ and $\sigma_2 := \sigma_{\min}(Q(\theta)) = \sigma_{\min}(\theta^2 A + \theta B + C)$ for the backward errors.

iteration	$\angle(u, x)$ $(\cdot 10^{-4})$	θ	error $(\cdot 10^{-3})$	$\|r\|$ $(\cdot 10^{-1})$	σ_1 $(\cdot 10^{-3})$	σ_2 $(\cdot 10^{-3})$
0	7.1924	7.2228+2.7788i	6.1127	1.2348	1.1667	3.1788
1	6.5423	7.2312+2.7836i	4.1139	1.1960	1.1370	2.1425
2	6.5295	7.2312+2.7812i	2.6279	1.1374	1.1373	1.3681
3	6.5289	7.2312+2.7811i	2.5972	1.1374	1.1374	1.3520
≥ 4	6.5289	7.2312+2.7811i	2.5961	1.1374	1.1374	1.3515

We see that in particular the first step of the algorithm considerably improves the approximate eigenpair. After four steps, the angle of the refined approximate eigenvector with the optimal vector in $\mathcal{U}$ is less than 30% of the angle that the Ritz vector makes with the optimal vector. The error in θ is more than halved. Note again that $\sigma_2 := \sigma_{\min}(\theta^2 A + \theta B + C)$ is expensive, but $\sigma_1 := \sigma_{\min}(\theta^2 AU + \theta BU + CU)$ is readily available in the algorithm.

6. Conclusions. The usual one-dimensional Galerkin approach for the determination of an approximate eigenvalue corresponding to an approximate eigenvector may give inaccurate results, especially when the discriminant of equation (2.1) is small. We have proposed several alternative ways that all require the same order of time and that often give better results. Based on our analysis and the numerical experiments, we recommend the approximations of the two-dimensional approaches Gal-2 and MR-2, because they are cheap to compute and give good results. For small eigenvalues, one should take the "ν" approximations. The MR-1 method ensures a minimal residual (backward error).

The approaches are also useful for approximations from a subspace and for polynomial eigenvalue problems of higher degree.

Acknowledgments We thank Hans Duistermaat for discussions about (5.1) and Jasper van den Eshof for helpful comments.

REFERENCES

[1] Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., and van der Vorst, H., Eds. *Templates for the solution of algebraic eigenvalue problems.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. A practical guide.

[2] Hochstenbach, M., and Sleijpen, G. Two-sided and alternating Jacobi-Davidson. Preprint 1196, Dep. Math., University Utrecht, Utrecht, the Netherlands, July 2001. To appear in LAA.

[3] Jia, Z. Refined iterative algorithms based on Arnoldi's process for large unsymmetric eigenproblems. *Linear Algebra Appl. 259* (1997), 1–23.

[4] Lanczos, C. An iteration method for the solution of the eigenvalue problem of linear differential and integral operators. *J. Res. Nat. Bur. Stand. 45*, 4 (1950), 255–282.

[5] Meerbergen, K., and Tisseur, F. The quadratic eigenvalue problem. *SIAM Rev. 43(2)* (2001), 235–286.

[6] Sleijpen, G. L. G., and van den Eshof, J. On the use of harmonic ritz pairs in approximating internal eigenpairs. Preprint 1184 (revised version), Dep. Math., University Utrecht, Utrecht, the Netherlands, August 2001. To appear in LAA.

[7] Tisseur, F. Backward error and condition of polynomial eigenvalue problems. *Linear Algebra Appl. 309(1-3)* (2000), 339–361.