Report

On harmonic Ritz values

Author(s):
Vömel, Christof

Publication Date:
2011

Permanent Link:
https://doi.org/10.3929/ethz-a-006810333

Rights / License:
In Copyright - Non-Commercial Use Permitted
ON HARMONIC RITZ VALUES

CHRISTOF VÖMEL

Abstract. One application of harmonic Ritz values is to approximate, with a projection method, the interior eigenvalues of a matrix $A$ while avoiding the explicit use of the inverse $A^{-1}$. In this context, harmonic Ritz values are commonly derived from a Petrov-Galerkin condition for the residual of a vector from the test space.

In this paper, we investigate harmonic Ritz values from a slightly different perspective. We consider a bounded functional $\psi$ that yields the reciprocals of the harmonic Ritz values of a symmetric matrix $A$. The crucial observation is that with an appropriate residual $s$ that differs from the standard one, many results from Rayleigh quotient and Rayleigh-Ritz theory can be extended. The same is true for the generalization to matrix pencils $(A, B)$ when $B$ is symmetric positive definite.

AMS subject classifications. 65F15, 65Y15.

Key words. Rayleigh quotient, Ritz value, harmonic Ritz value, symmetric eigenproblem, generalized symmetric positive definite eigenproblem.

1. Introduction. Prompted by the seminal works of Morgan [15], of Paige, Parlett, and van der Vorst [20], and of Sleijpen and van der Vorst [25], there has been a lot of recent research dedicated to the topic of harmonic Ritz values, see for example [2, 8, 9, 10, 12, 17, 24, 31] and also [1, Chapter 3] and [30, Chapter 4.4].

Harmonic Ritz values can rival the standard ones as approximation to the interior eigenvalues of a matrix. In this role, they have a natural place in iterative eigensolvers based on matrix projection on a subspace, including LOBPCG [13], Lanczos [21, 27], (generalized) Davidson [4, 5, 16, 18], and Jacobi-Davidson [11, 14, 23, 25, 26, 28, 29]. We are interested in their application in a very general projection method where the subspace in question might not be of Krylov or block-Krylov type.

For a real symmetric matrix $A$, the standard Ritz values can be viewed as a matrix extension to the Rayleigh quotient

$$\rho(v) := \frac{v^t A v}{v^t v}, \quad v \neq 0.$$  

Likewise, one can see harmonic Ritz values as matrix extension of the functional

$$\phi(v) := \frac{v^t A^2 v}{v^t A v}, \quad v \neq 0.$$  

What started this current work is the observation that $\phi$ is unbounded when $A$ is indefinite, even in the case where it is nonsingular. From this point of view, it seems natural to instead consider the reversed version

$$\psi(v) := \frac{v^t A v}{v^t A^2 v}, \quad v \neq 0.$$  

This is the topic of Section 2. There, we also introduce the key definition of the associated residual

$$s(v) := v - A v \psi(v)$$

\footnote{Institute of Computational Science, ETH Zürich, CAB, Universitätstrasse 6, 8092 Zürich, Switzerland, cvoemel@inf.ethz.ch}
which features prominently in subsequent sections.

In addition to having a greater domain of definition, we found that $\psi$ has some interesting features in common with $\rho$ some of which are not shared by $\phi$. In Section 3, we thus make the case that for both practical and theoretical reasons, $\psi$ is a worthy subject of study itself.

In Section 4, we investigate extension of the functionals $\rho$, $\phi$ and $\psi$ to subspaces. This leads to Ritz values and vectors. For the harmonic case, the Petrov-Galerkin condition on the residual that commonly is the fundamental definition follows here from the matrix functional. Attractive features of the matrix functional $\Psi(V)$ and the subspace residual $S(V)$ are illuminated.

Last but not least, Section 5 exhibits the generalization to matrix pencils $(A,B)$ where $B$ is symmetric positive definite. In this case, we study

$$\psi(v) := \frac{v^t A v}{v^t B^{-1} A v}$$

with the associated generalized residual

$$s(v) := Bv - Av\psi(v).$$

Section 6 summarizes and concludes.

2. Some rational quadratic forms. The main point of this section is to motivate the study of $\psi$ which could be called the ‘reversed harmonic Rayleigh quotient.’

2.1. Review of the Rayleigh quotient. Let $A$ denote a real symmetric matrix. The Rayleigh quotient $\rho$ of a vector $v \neq 0$ is defined as

$$\rho_A(v) := \frac{v^t A v}{v^t v}. \quad (2.1)$$

When there is no risk of ambiguity, the subscript mention of the matrix $A$ is avoided.

**Theorem 2.1.** The Rayleigh quotient enjoys three fundamental properties [21]:

1. Invariance under scaling: for $\mathbb{R} \ni \alpha \neq 0$: $\rho(\alpha v) = \rho(v)$.

2. Boundedness:

$$\lambda_{\min}(A) = \min_{v \neq 0} \rho(v), \quad \lambda_{\max}(A) = \max_{v \neq 0} \rho(v), \quad (2.2)$$

where $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the smallest and largest eigenvalue of $A$, respectively.

3. Stationarity: $\rho$ is stationary in its argument $v$ if and only if $v$ is an eigenvector of $A$.

Because of scaling-invariance, it is sufficient to consider $\rho$ for vectors of unit length. For any such vector, $\rho$ is a weighted average of the eigenvalues of $A$, this shows boundedness. Stationarity follows from inspection of the gradient

$$\nabla \rho(v) = 2\frac{Av - v\rho(v)}{v^t v} =: \frac{2}{v^t v} r(v). \quad (2.3)$$

Here,

$$r(v) := Av - v\rho(v) \quad (2.4)$$

denotes the residual which satisfies the Galerkin condition

$$r(v) \perp v. \quad (2.5)$$
By (2.2), one can compute the largest (or smallest) eigenvalues of $A$ by an optimization method such as Preconditioned Conjugate Gradients [3, 19, 22]. In order to compute eigenvalues in the interior of the spectrum of $A$, one can replace $A$ by a shifted-and-inverted version, that is apply the method to

$$\frac{v^t(A - \sigma I)^{-1}v}{v^t v}.$$ (2.6)

Here, the shift $\sigma$ is chosen close but not exactly equal to the eigenvalues of interest.

Why does one usually not consider the reversed Rayleigh quotient

$$\frac{v^t v}{v^t Av}?$$ (2.7)

The answer is: when $A$ is indefinite, then $v \neq 0$ does not imply that $v^t Av \neq 0$. Thus the functional is unbounded, even if $A$ is nonsingular. For example:

$$\begin{pmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} = 0.$$ (2.8)

In this example, the problem does not seem to be very severe. The bisectors of the true eigenvectors constitute the critical points and are thus far away, in terms of the angle, from the points of interest. However, by considering a matrix $\text{diag}(10^{-10})$, instead say, one sees that depending on the spectrum, the critical points can be very close to an eigenvector and make the computation challenging.

2.2. A harmonic functional and a variation. In his pioneering paper [15], Morgan considers the inverse-free functional

$$\phi(v) := \frac{v^t(A - \sigma I)^2 v}{v^t(A - \sigma I)v}$$ (2.9)

for computing an interior eigenvalue close to $\sigma$. He proves that the eigenvector associated to the closest eigenvalue minimizes the magnitude of (2.9). Implicitly, this functional has been revisited by Paige, Parlett, and van der Vorst as it arises in an investigation of MINRES [20]. The authors point out among others the further occurrence in [7]. In [25], Sleijpen and van der Vorst generalize it to a Petrov-Galerkin extraction approach for general subspaces. Further extensions are for example given in [2, 9, 10].

We now return to the functional (2.9), where for simplicity of notation we henceforth assume $\sigma = 0$. As the denominator generally is indefinite, the discussion at the end of Section 2.1 is relevant in the context of $\phi$ as well. From the computational point of view, it thus seems worthwhile to also consider the following functional

$$\psi_A(v) := \frac{v^t Av}{v^t A^2 v}. $$ (2.10)

Without the risk of ambiguity, the subscript mention of the matrix $A$ is again avoided. Of course, this is just the reversed version of (2.9), and for ease of notation, we do not mention $\sigma$. Nevertheless, as long as $A$ is nonsingular, the denominator of (2.10) is definite and thus the functional is defined for any $v \neq 0$. Furthermore, we have the fundamental relation

$$\psi_A(v) = \rho_{A^{-1}}(w), \ w := Av.$$ (2.11)
We now investigate the similarities between $\rho$, $\phi$, and $\psi$.

**Theorem 2.2.** Let $A$ be nonsingular, and consider $\phi$ from (2.9) with $\sigma = 0$, and $\psi = \psi_A$ as in (2.10).

1. For all $v$ where $\phi$ is defined, one has $\phi(v) \cdot \psi(v) = 1$.
2. Both $\phi$ and $\psi$ are invariant under scaling of the argument $v$.
3. $\psi$ is bounded for $v \neq 0$.

$$\lambda_{\min}(A^{-1}) = \min_{v \neq 0} \psi(v), \quad \lambda_{\max}(A^{-1}) = \max_{v \neq 0} \psi(v).$$  
(2.12)

$\phi$ is only bounded if $A$ is definite.

4. $\phi$ and $\psi$ are stationary if and only if $v$ is an eigenvector of $A$.

**Proof.** Scaling invariance is immediate. Without loss of generality, let $\|w\|_2 = 1$. Then $\alpha_i := w^t v_i, \sum \alpha_i^2 = 1$, and $v = \sum \alpha_i v_i$ imply that

$$\psi_A(v) = \sum \frac{\alpha_i^2}{\lambda_i}.$$  
(2.13)

Thus $\psi$ is a weighted average of the eigenvalues of $A^{-1}$. As $A$ is nonsingular, the mapping $v \rightarrow w$ is one-to-one and (2.12) follows from (2.11) and (2.2). $\phi$ is unbounded when $A$ is indefinite, compare the example at the end of Section 2.1. To prove stationarity, we note that

$$\nabla \phi(v) = 2A v^t (Av - v \phi(v)).$$  
(2.14)

The last factor vanishes if and only if $v$ is an eigenvector. Similarly,

$$\nabla \psi(v) = 2A v^t (v - Av \psi(v)).$$  
(2.15)

When $v$ is an eigenvector, then the last factor vanishes. Conversely, assume that the factor vanishes. In this case $\psi$ cannot be zero: cancellation is needed as $v \neq 0$ and $Av \neq 0$. Thus $\phi$ is defined, and using the first part of the theorem shows that also (2.14) must be zero, that is $v$ is an eigenvector. □

The preceding results show that where defined, one has for $v, \|Av\|_2 = 1$,

$$\frac{1}{\phi_A(v)} = \sum \frac{\alpha_i^2}{\lambda_i}.$$  
(2.16)

Hence, $\phi$ is a harmonic weighted average of the eigenvalues of $A$, motivating the name [20]. The last factor from (2.14) could be called the harmonic residual

$$r^{(h)}(v) := Av - v \phi(v).$$  
(2.17)

It is defined wherever $\phi$ is defined and differs from the standard residual (2.4) only in the choice of the scaling factor for $v$. The Petrov-Galerkin condition

$$r^{(h)}(v) \perp Av$$  
(2.18)

holds which, when $A$ is definite, can be written as $r^{(h)}(v) \perp_A v$. (2.15)'s last factor,

$$s(v) := v - Av \psi(v)$$  
(2.19)

is the residual with respect to $\psi$ and defined for any $v \neq 0$. It also satisfies a Petrov-Galerkin condition

$$s(v) \perp Av.$$  
(2.20)
3. Similarities between ψ and ρ.

3.1. Minimal residual properties. This short section points out a similarity between ρ, r and ψ, s that is not shared by φ, r^(h).

We first recall the following minimal residual property.

**Theorem 3.1.** [21, Fact 1.8] For a given vector v and any scalar σ holds

\[ \|r(v)\|_2 = \|Av - \rho(v)v\|_2 \leq \|Av - \sigma v\|_2. \]  

(3.1)

**Corollary 3.2.** Let v ≠ 0 and assume that its harmonic residual r^(h)(v) from (2.17) is defined. Then

\[ \|r(v)\|_2 \leq \|r^(h)(v)\|_2. \]  

(3.2)

Remarkably, ψ and s enjoy a minimal residual property similar to the relation for ρ and r in Theorem 3.1.

**Theorem 3.3.** Let A be nonsingular. For a given vector v and any scalar σ, it holds for s from (2.19) that

\[ \|s(v)\|_2 = \|v - Av\psi(v)\|_2 \leq \|v - Av\sigma\|_2. \]  

(3.3)

**Proof.** At the end of Section 2.2 was noted that s(v) ⊥ Av. Hence

\[ v - Av\sigma = [v - Av\psi(v)] - (\sigma - \psi(v)) Av \]

is an orthogonal decomposition. This in turn yields

\[ \|v - Av\sigma\|_2^2 = \|s(v)\|_2^2 + |\sigma - \psi(v)|^2 \|Av\|_2^2, \]

which proves the result. \(\blacksquare\)

3.2. Backward error. It is remarkable that for a vector v of unit length, the norm of the standard residual r(v) from (2.4) equals the norm of a rank-two perturbation of A that has (ρ_A(v), v) as an eigenpair. Precisely:

**Theorem 3.4.** [21, Theorem 4.5.2] Let v ≠ 0, then (ρ_A(v), v) is an eigenpair of A - M, where

\[ M = \frac{vr(v)^t + r(v)v^t}{\|v\|_2^2}, \quad \|M\|_2 = \frac{\|r(v)\|_2}{\|v\|_2}. \]  

(3.4)

**Proof.** Without loss of generality, assume that v is not an eigenvector, otherwise nothing needs to be shown. The eigenpair property follows from the Galerkin condition r(v) ⊥ v and definition (3.4). M has rank two and the eigenvectors belonging to its nonzero eigenvalues must be linear combinations of v and r. The Ansatz \(x = \alpha v + \beta r\) yields

\[ Mx = \frac{1}{\|v\|_2^2} (\beta r^t r)v + \alpha (v^t v)r. \]

For the right-hand side to be collinear with x, one needs to have

\[ \frac{\alpha^2}{\beta^2} = \frac{r^t r}{v^t v}. \]
This yields the expression for $\|M\|_2$. \(\square\)

Relation (2.11) between standard and reversed harmonic Rayleigh quotient
\[
\psi_A(v) = \rho_{A^{-1}}(w), \ w := Av,
\]
indicates that one can hope to find a similar result involving $\psi$.

**Theorem 3.5.** Let $A$ be nonsingular, $v \neq 0$, and $s(v)$ as in (2.19). Then $(\psi_A(v), Av)$ is an eigenpair of $A^{-1} - N$, where
\[
N = \frac{Avs(v)^t + s(v)v' A}{\|Av\|_2^2}, \quad \|N\|_2 = \frac{\|s(v)\|_2}{\|Av\|_2}
\] (3.6)

**Proof.** If $v$ is an eigenvector of $A$ then $s(v) = 0$ and nothing is to show.

\((A^{-1} - N)Av = v - NAv = v - s(v) = \psi_A(v)Av,
\)

by the Petrov-Galerkin condition $s(v) \perp Av$. Comparison of (3.6) with (3.4) shows that $\|N\|_2$ can be computed in the same way as $\|M\|_2$ in the proof of Theorem 3.4. \(\square\)

It seems difficult to find a similar relation for $\phi$.

### 3.3. Distance bounds for eigenvalues.

We now compare how the residual vectors introduced in Section 2.2 can be used in bounding errors in eigenvalues. A starting point is Weinstein’s

**Theorem 3.6.** [21, Theorem 4.5.1] For any scalar $\sigma$ and any vector $v \neq 0$, there is an eigenvalue $\lambda$ of $A$ such that
\[
|\lambda - \sigma| \leq \frac{\|Av - v\sigma\|_2}{\|v\|_2},
\] (3.7)

Two different proofs can be found in [21]. According to Theorem 3.1, the right hand side is minimized by $\|r(v)\|_2/\|v\|_2$.

**Corollary 3.7.** Let $v \neq 0$, then

1. there is an eigenvalue $\lambda$ of $A$ such that
\[
|\lambda - \rho(v)| \leq \frac{\|r(v)\|_2}{\|v\|_2},
\] (3.8)

2. if $v' Av \neq 0$, there is an eigenvalue $\lambda$ of $A$ such that
\[
|\lambda - \phi(v)| \leq \frac{\|r(v)\|_2}{\|v\|_2}.
\] (3.9)

Now, a Weinstein-type bound for $s$ is established.

**Theorem 3.8.** Let $A$ be nonsingular. For any scalar $\sigma$ and any vector $v \neq 0$, there is an eigenvalue $\lambda$ of $A^{-1}$ such that
\[
|\lambda - \sigma| \leq \frac{\|v - Av\sigma\|_2}{\|Av\|_2},
\] (3.10)

**Proof.**
\[
\frac{\|v - Av\sigma\|_2^2}{\|Av\|_2^2} = \frac{\|(A^{-1} - \sigma I)Av\|_2^2}{\|Av\|_2^2} = \rho_{(A^{-1} - \sigma I)^2}(Av) \geq \min \left(\lambda_i(A^{-1} - \sigma)^2\right),
\]
where the last line exploits the bound from (2.2) on the standard Rayleigh-quotient.

As discussed in Section 3.1, the right-hand side of the bound (3.10) is minimized by \( \psi \).

### 3.4. Angle with the eigenvector.

This Section gives an extension to the famous Davis-Kahan gap theorem [6, 21] for an inexact eigenpair.

For a given matrix \( A \) and \( v \neq 0 \), denote by \( \bar{\lambda} = \lambda(A) \) the eigenvalue closest to \( \rho_A(v) \), and by \( \bar{v} \) an associated eigenvector of unit length. With this notation, the result is given as

**Theorem 3.9.** For a unit vector \( v \) with Rayleigh quotient \( \rho(v) \) and residual \( r(v) \), one has

\[
|\sin \angle(v, \bar{v})| \leq \frac{\|r(v)\|_2}{\text{gap}(\rho(v), A)},
\]

where \( \text{gap}(\rho, A) = \min \{|\rho - \lambda| : \lambda \neq \bar{\lambda}, \lambda \in \text{spectrum}(A)\} \).

Key to the proof is an orthogonal decomposition of \( v \) into the direction of the ‘closest’ eigenvector \( \bar{v} \) and perpendicular to that. In the following proof, a similar decomposition

\[
Av = \cos \vartheta \bar{A} \bar{v} + \sin \vartheta \bar{A} w, \quad \bar{A} \bar{v} \perp \bar{A} w,
\]

is used (\( \vartheta := \angle(Av, \bar{A} \bar{v}) \)).

**Theorem 3.10.** Let \( A \) be nonsingular. Let \( v \neq 0 \) such that \( \|Av\|_2 = 1 \), with \( \psi(v) \) as in (2.10) and residual \( s(v) \) as in (2.19). Denote by \( \bar{\lambda} = \lambda(A^{-1}) \) the eigenvalue of \( A^{-1} \) closest to \( \psi_A(v) \), and by \( \bar{v} \) a corresponding eigenvector normalized such that \( \|A\bar{v}\|_2 = 1 \). Then one has

\[
|\sin \angle(Av, \bar{A} \bar{v})| \leq \frac{\|s(v)\|_2}{\text{gap}(\psi(v), A^{-1})}.
\]

**Proof.** First note that

\[
s(v) = v - A \psi(v) = (A^{-1} - \psi(v)I)Av.
\]

Inserting here-in the orthogonal decomposition (3.12) one finds that

\[
s(v) = \cos \vartheta (A^{-1} - \psi(v)I)A \bar{v} + \sin \vartheta (A^{-1} - \psi(v)I)Aw
\]

is an orthogonal decomposition of \( s \). This implies

\[
\|s\|_2^2 = \cos^2 \vartheta |\bar{\lambda} - \psi(v)|^2 + \sin^2 \vartheta \|(A^{-1} - \psi(v)I)Aw\|_2^2.
\]

Now, since

\[
\|(A^{-1} - \psi(v)I)Aw\|_2^2 \geq \min_{\rho(A^{-1} - \psi(v)I)^2} \text{gap}^2(\psi(v), A^{-1}),
\]

the result follows.

Another useful direct consequence of (3.12) is that the error in \( \psi \) is proportional to the square of the error in \( v \), analogous to the Rayleigh quotient.

**Theorem 3.11.** Let \( A \) be nonsingular. Let \( v \neq 0 \) such that \( \|Av\|_2 = 1 \), with \( \psi(v) \) as in (2.10). Denote by \( \bar{\lambda} = \lambda(A^{-1}) \) the eigenvalue of \( A^{-1} \) closest to \( \psi_A(v) \) and let (3.12) be in force. Then

\[
\psi(v) = \bar{\lambda} - \sin^2 \angle(Av, \bar{A} \bar{v}) (\bar{\lambda} - \psi(w))
\]

(3.14)
3.5. Hesse matrices. Newton’s method can be used to compute eigenvalues as stationary points of the Rayleigh quotient. In this approach, the Hessian of $\rho$ plays a central role.

**Theorem 3.12.** [33] Let $v \neq 0$, then

$$\nabla^2 \rho(v) = \frac{2}{v^tv} \left\{ \left( I - \frac{2}{v^tv} vv^t \right) (A - \rho(v)I) \left( I - \frac{2}{v^tv} vv^t \right) \right\}. \quad (3.15)$$

**Proof.** By (2.3)

$$\nabla^2 \rho(v) = \nabla \left( \frac{2}{v^tv} r(v) \right)$$

$$= \nabla \left( \frac{2}{v^tv} \right) r(v)^t + \frac{2}{v^tv} \{ (A - \rho(v)I) - \nabla (\rho(v))^t \}$$

$$= \frac{2A - \rho(v)I}{v^tv} - \frac{4r(v)v^t + vr(v)^t}{(v^tv)^2}.$$

The last line equals (3.15) as $r(v) = (A - \rho(v)I)v \perp v$. □

It is interesting to note that the Hessian of $\psi$ can be written using Householder reflectors as well.

**Theorem 3.13.** Let $A$ be nonsingular and $v \neq 0$. If $w := Av$, then

$$\nabla^2 \psi(v) = \frac{2}{w^tw} A \left\{ \left( I - \frac{2}{w^tw} ww^t \right) (A^{-1} - \psi(v)I) \left( I - \frac{2}{w^tw} ww^t \right) \right\} A. \quad (3.16)$$

**Proof.** By (2.15)

$$\nabla^2 \psi(v) = \nabla \left( \frac{2A}{v^tA^2v} [v - Av\psi(v)] \right) = \nabla \left( \frac{2A}{v^tA^2v} s \right)$$

$$= \frac{2A - \psi(v)A^2}{v^tA^2v} - \frac{4As(v)v^tA^2 + A^2vs(v)^tA}{(v^tA^2v)^2}.$$

Substitution of $s = (A^{-1} - \psi(v)I)Av$ yields

$$\nabla^2 \psi(v) =$$

$$\frac{2}{v^tA^2v} A \left\{ (A^{-1} - \psi(v)I) - \frac{2(A^{-1} - \psi(v)I) Av v^t A + Avv^t A (A^{-1} - \psi(v)I)}{(v^tA^2v)} \right\} A. \quad (3.17)$$

This equals (3.16) as $v^tA (A^{-1} - \psi(v)I) Av = v^tAs(v) = 0$. □


4.1. The matrix Rayleigh quotient. This section briefly reviews standard material that can be found for example in [21, Chapter 11], in [1, Chapter 3], or in [30, Chapter 4.4].

Let the columns of $V \in \mathbb{R}^{n \times k}$ be linearly independent. Define the matrix Rayleigh quotient

$$\varrho(V) := (V^tV)^{-1} V^tAV \quad (4.1)$$
As $V^TV$ is nonsingular, it is even positive definite with a Cholesky factorization $V^TV = LL^t$. Thus $g(V)$ is similar to the symmetric matrix $L^{-1}g(V)L^{-t}$. Consequently, it has $k$ real eigenvalues $\mu_i$, the so-called Ritz values. These also are exactly the eigenvalues of the (definite) generalized symmetric eigenvalue problem

$$V^tAV y = \mu V^TV y, \ 0 \neq y \in \mathcal{R}^k.$$  \hspace{1cm} (4.2)

Let $y_i$ denote the eigenvector of (4.2) associated to $\mu_i$. Then $V y_i \in \mathcal{R}^n$ is called a Ritz vector.

In practice, one commonly chooses $V$ to have orthonormal columns whenever possible. As this also allows for theoretical simplification, we henceforth assume that $V^TV = I$ and refer to [21, Section 11.10] for the more general setting. When the columns of $V$ are orthonormal, one has

$$g(V) = \begin{bmatrix}
\rho(v_1) & v_1^tAv_2 & \ldots & v_1^tAv_k \\
\vdots & \ddots & \vdots \\
v_k^tAv_1 & \ldots & v_k^tAv_{k-1} & \rho(v_k)
\end{bmatrix}.$$  \hspace{1cm} (4.3)

We note that $\mu_i = \rho(V y_i)$ so that (4.2) can be rephrased as

$$0 = V^t [(A - \mu_i I)V y_i] = V^tr(V y_i)$$  \hspace{1cm} (4.4)

to yield a Ritz-Galerkin condition stronger than (2.5).

As an extension of (2.4), one can also define the subspace residual

$$R(V) := AV - Vg(V).$$  \hspace{1cm} (4.5)

It satisfies the subspace Galerkin condition

$$R(V) \perp V.$$  \hspace{1cm} (4.6)

The subspace span$(V)$ is invariant under $A$ if and only if $R(V) = 0$.

**4.2. $\Phi$ and $\Psi$: matrix versions of $\phi$ and $\psi$.** Let $A$ be nonsingular and the columns of $V \in \mathcal{R}^{n \times k}$ be linearly independent. The matrix versions of (2.9) and (2.10) are

$$\Phi(V) := (V^tAV)^{-1}V^tA^2V,$$  \hspace{1cm} (4.7)

and

$$\Psi(V) := (V^tA^2V)^{-1}V^tAV,$$  \hspace{1cm} (4.8)

respectively. When both $\Phi$ and $\Psi$ are defined, they have real eigenvalues as they are similar to symmetric matrices. Since one has $\Phi(V)\Psi(V) = \Psi(V)\Phi(V) = I$, the eigenvalues of one are just the reciprocals of the other.

The eigenvalues $\mu_i$ of $\Phi$ are called the harmonic Ritz values. They are solutions of the (generally indefinite) generalized symmetric eigenvalue problem

$$V^tA^2V y = \mu V^tAV y, \ 0 \neq y \in \mathcal{R}^k.$$  \hspace{1cm} (4.9)

Let the harmonic subspace residual

$$R^{(h)}(V) := AV - V\Phi(V),$$  \hspace{1cm} (4.10)
then one has as the generalization of (2.18) that
\[ R^{(h)}(V) \perp AV. \quad (4.11) \]

This Petrov-Galerkin condition is normally used to define the harmonic Ritz values and vectors, see for example [25] or [1, Chapter 3].

The blemish of \( \Phi \) is that \( V^tAV \) need not be invertible for indefinite \( A \), recall the example at the end of Section 2.1. This is not a negligible detail: for example, in [12], convergence of the harmonic Ritz values to an eigenvalue of \( A \) is shown under the condition that \( \|(V^tAV)^{-1}\| \) is uniformly bounded. However, for interior eigenvalues of an indefinite (shifted) \( A \), this condition might never be satisfied!

On the other hand, \( V^tA^2V \) is positive definite when \( A \) is nonsingular. This makes \( \Psi \) attractive. The associated definite generalized symmetric eigenvalue problem is
\[ V^tAVy = \mu V^tA^2Vy, \quad 0 \neq y \in \mathbb{R}^k. \quad (4.12) \]

With
\[ S(V) := V - AV\Psi(V) \quad (4.13) \]

one has the alternative Petrov-Galerkin condition
\[ S(V) \perp AV. \quad (4.14) \]

4.3. Similarities between \( \Psi \) and \( \rho \). Section 3 pointed out interesting parallels between \( \rho \) and \( \psi \). We investigate here how some of them generalize to the matrix versions of the respective functionals. The focus is on results that involve \( \Psi \) and \( S(V) \). In [24, Section 3], a nice summary of other similar features is given, including a translation of the Courant-Fischer Minmax/Maxmin principle.

First, we consider the generalization of the minimal residual properties from Section 3.1. For the matrix Rayleigh quotient, we have as extension of Theorem 3.1 that

**Theorem 4.1.** [21, Theorem 11.4.2] Let \( V \in \mathbb{R}^{n \times k} \) with \( V^tV = I \). Let \( R(V) \) as in (4.5), then
\[ \|R(V)\|_2 = \|AV - V\rho(V)\|_2 \leq \|AV - VB\|_2 \quad (4.15) \]

for any matrix \( B \in \mathbb{R}^{k \times k} \).

As a generalization of Theorem 3.3, we show

**Theorem 4.2.** Let \( A \) be nonsingular and \( V \in \mathbb{R}^{n \times k} \) such that \( V^tA^2V = I \). Then
\[ \|S(V)\|_2 = \|V - AV\Psi(V)\|_2 \leq \|V - AVB\|_2. \quad (4.16) \]

for any matrix \( B \in \mathbb{R}^{k \times k} \).

**Proof.** We first note the auxiliary result
\[ \|S(V)\|_2 = \|V^tS(V)\|_2 \quad (4.17) \]

which is a consequence of \( S(V)^tS(V) = S(V)^tV + S(V)^tAV\Psi(V) \) and the Petrov-Galerkin condition (4.14). Now we have
\[
(V - AVB)^t(V - AVB) = V^tV - \Psi(V)^tB - B^t\Psi(V) + B^t(V^tA^2V)B
= V^tS(V) + (\Psi(V) - B)^t(\Psi(V) - B)
\]
The matrix \((\Psi(V) - B)^t(\Psi(V) - B)\) is positive semidefinite. Weyl’s Monotonicity Theorem \([21, \text{Theorem 10.3.1}]\) yields that every eigenvalue, in particular the largest one, of \((V - AVB)^t(V - AVB)\) is greater or equal than the corresponding one of \(V^tS(V)\). By (4.17), this completes the proof. \footnote{\[21, \text{Theorem 10.3.1}\]}

As a generalization of Theorem 3.6, one has

**Theorem 4.3.** \([21, \text{Theorem 11.5.1}]\) Let \(V \in \mathbb{R}^{n \times k}\) with \(V^tV = I\). Let \(R(V)\) as in (4.5), then there are \(k\) eigenvalues \(\lambda(A)\) such that for \(i = 1, \ldots, k:\)

\[
|\lambda_i(A) - \mu_i(\Psi)| \leq \|R(V)\|_2. \tag{4.18}
\]

For \(\Psi(V)\) and \(S(V)\), Theorem 3.8 generalizes to

**Theorem 4.4.** Let \(A\) be nonsingular and \(V \in \mathbb{R}^{n \times k}\) such that \(V^tA^2V = I\). Then there are \(k\) eigenvalues \(\lambda(A^{-1})\) such that for \(i = 1, \ldots, k:\)

\[
|\lambda_i(A^{-1}) - \mu_i(\Psi)| \leq \|S(V)\|_2. \tag{4.19}
\]

*Proof.* Let \(P := (AV, AW)\) such that \(P^tP = PP^t = I\). The matrix \(P^tA^{-1}P\) is orthogonally similar to \(A^{-1}\) and has the shape

\[
P^tA^{-1}P = \begin{bmatrix}
\Psi(V) & \\
\Psi(W) & V^tAW
\end{bmatrix} + \begin{bmatrix}
W^tAV & V^tAW
\end{bmatrix}.
\]

To prove the result using again Weyl’s Monotonicity Theorem \([21, \text{Theorem 10.3.1}]\), one thus has to find \(\|W^tAV\|_2\) which equals in magnitude the largest and smallest eigenvalue of the second term. For this, note that

\[
P^tS(V) = \begin{bmatrix}
V^tAV - V^tA^2V\Psi(V) & 0 \\
W^tAV - W^tA^2V\Psi(V) & W^tAV
\end{bmatrix} = \begin{bmatrix}
0 & W^tAV
\end{bmatrix}
\]

using the definition of \(\Psi\) and \(AV \perp AW\). Now, since \(\|S(V)\|_2 = \|P^tS(V)\|_2 = \|W^tAV\|_2\), everything is shown. \footnote{\[21, \text{Theorem 11.5.1}\]}

5. **Matrix pencils.** We consider here functionals for the matrix pencil \((A, B)\) where \(A\) and \(B\) are real symmetric matrices, and \(B\) is also positive definite.

5.1. **Generalized \(\rho\).** This review is intentionally brief, for background material see for example \([21, \text{Chapter 15}]\) or \([1, \text{Chapter 5.7}]\). The (generalized) Rayleigh quotient \(\rho\) is defined as

\[
\rho(v) := \rho_{(A,B)}(v) := \frac{v^tAv}{v^tBv}. \tag{5.1}
\]

The gradient yields the residual

\[
\nabla \rho(v) = 2\frac{Av - Bvp(v)}{v^tBv} =: \frac{2}{v^tBv}r(v). \tag{5.2}
\]

Theorem 2.1 also holds in the generalized setting. The positive-definiteness of \(B\) implies existence of the Cholesky factorization \(B = LL^T\). Thus, the pencils \((A, B)\) and \((L^{-1}AL^{-t}, I)\) as well as \((AB^{-1}, I)\) and \((B^{-1}A, I)\) have the same eigenvalues. Further, one can define for both \(M := B\) and \(M := B^{-1}\) an induced scalar product that yields the associated norm and angle:

\[
(v, w)_M := w^tMv, \quad \|v\|_M := \sqrt{(v, v)_M}, \quad \angle_M(v, w) := \arccos \frac{|(v, w)_M|}{\|v\|_M\|w\|_M}.
\]
One important application is the fact that the eigenvectors of \((A, B)\) can always be chosen \(B\)-orthonormal. Another one is the Galerkin condition \(r(v) \perp v\) which can be written as
\[
r(v) \perp_{B^{-1}} Bv.
\] (5.3)

The following Theorem 5.1 summarizes the beautiful extensions of the results in Section 3 to the generalized Rayleigh quotient.

**Theorem 5.1.** Let \(v \neq 0\) be a vector and \(\sigma \in \mathbb{R}\) a scalar.

1. **Minimal residual property** [21, Theorem 15.9.2]:
\[
\|r(v)\|_{B^{-1}} = \|Av - \rho(v)Bv\|_{B^{-1}} \leq \|Av - \sigma Bv\|_{B^{-1}}.
\] (5.4)

2. **Backward error** [1, Section 5.7.1]: \((\rho(v), v)\) is an eigenpair of \((A - M, B)\), where
\[
M = \frac{v \rho(v) + r(v)v^t}{\|v\|^2}, \quad \|M\|_2 = \|r\|_2 \|v\|_2
\] (5.5)

3. **Weinstein’s bound** [21, Theorem 15.9.1]: there is an eigenvalue \(\lambda\) of \((A, B)\) such that
\[
|\lambda - \sigma| \leq \frac{\|Av - Bv\|_{B^{-1}}}{\|Bv\|_{B^{-1}}^2}.
\] (5.6)

4. **Angle with the closest eigenvector** [1, Section 5.7.1]:
\[
|\sin \theta(v, \tilde{v})| = |\sin \theta_{B^{-1}}(Bv, B\tilde{v})| \leq \frac{\|r(v)\|_{B^{-1}}}{\text{gap}(\rho(v), (A, B)) \cdot \|Bv\|_{B^{-1}}},
\] (5.7)
with the straight-forward definition of the spectral gap as in Theorem 3.9.

5. **Hessian:**
\[
\nabla^2 \rho(v) = \frac{2}{v^t Bv} \left\{ \left( I - \frac{2}{v^t Bv} Bvv^t \right) (A - \rho(v)B) \left( I - \frac{2}{v^t Bv} v^t B \right) \right\}.
\] (5.8)

Some results in Theorem 5.1 can be partly simplified using the normalization \(\|v\|_B = \|Bv\|_{B^{-1}} = 1\).

### 5.2. Generalized \(\psi\)

In order to define \(\psi\) for pencils, we will again assume that \(A\) is nonsingular. Then, one way to extend \(\psi\) to the generalized case is via the fundamental relation (2.11) for the standard case, \(\psi_A(v) = \rho_{A^{-1}}(Av)\). It can be applied to the pencil \((L^{-1}AL^{-t}, I)\) which is equivalent to \((A, B)\) as noted in Section 5.1. This yields \((L^{-1}AL^{-t}, L^{-1}AB^{-1}AL^{-t})\) and the simpler, equivalent
\[
\psi(v) := \psi_{(A,B)}(v) := \frac{v^t Av}{v^t AB^{-1}Av}.
\] (5.9)

Incidentally, it satisfies \(\psi_{(A,B)}(v) = \rho_{(A^{-1}, B^{-1})}(Av)\).

The gradient of (5.9) guides the search for the definition of an appropriate residual:
\[
\nabla \psi(v) = 2\frac{AB^{-1}}{v^t AB^{-1}Av} [Bv - Av\psi(v)].
\] (5.10)
We let (5.10)'s last factor,
\[ s(v) := Bv - Av\psi(v) \] (5.11)
denote the (generalized) residual with respect to \( \psi \). It is defined for any \( v \neq 0 \) and satisfies the Petrov-Galerkin condition
\[ s(v) \perp B^{-1}Av. \] (5.12)

We are now ready to state the analogue to Theorem 5.1 for \( \psi \).

**Theorem 5.2.** Let \( A \) be nonsingular, \( v \neq 0 \) be a vector and \( \sigma \in \mathbb{R} \) be a scalar.

1. Minimal residual property:
\[ \| s(v) \|_{B^{-1}} = \| Bv - \psi(v)Av \|_{B^{-1}} \leq \| Bv - \sigma Av \|_{B^{-1}}. \] (5.13)

2. Backward error: \( (\psi(v), Av) \) is an eigenpair of the pencil \( (A^{-1} - N, B^{-1}) \), where
\[ N = \frac{Avs(v)^tB^{-1} + B^{-1}s(v)s^tA}{\| Av \|_2^2}, \| N \|_2 = \frac{\| B^{-1}s(v) \|_2}{\| Av \|_2}. \] (5.14)

3. Weinstein-type bound: there is an eigenvalue \( \lambda \) of \( (A^{-1}, B^{-1}) \) such that
\[ |\lambda - \sigma| \leq \frac{\| Bv - \sigma Av \|_{B^{-1}}}{\| Av \|_{B^{-1}}}. \] (5.15)

4. Angle with the closest eigenvector \( \bar{v} \): let \( \| Av \|_{B^{-1}} = \| A\bar{v} \|_{B^{-1}} = 1 \), then
\[ |\sin \angle_{B^{-1}}(Av, A\bar{v})| \leq \frac{\| s(v) \|_{B^{-1}}}{\text{gap}(\psi(v), (A^{-1}, B^{-1}))}. \] (5.16)

5. Hessian: let \( w := Av \), then
\[ \nabla^2 \psi(v) = \frac{2}{w^tB^{-1}w} \left( I - \frac{2}{w^tB^{-1}w}B^{-1}ww^t \right) \left( A^{-1} - \psi(v)B^{-1} \right) \left( I - \frac{2}{w^tB^{-1}w}w(B^{-1}w)^t \right) A. \] (5.17)

**Proof.** The Petrov-Galerkin condition (5.12) shows that
\[ Bv - \sigma Av = s(v) + (\psi - \sigma)Av \]
is a \( B^{-1} \)-orthogonal decomposition. This is sufficient to show (5.13), see also the proof of Theorem 3.3.

The difficulty of the backward error result is to find the shape of \( N \). With (5.14) at hand, one sees that
\[ (A^{-1} - N)Av = v - B^{-1}s = \psi B^{-1}Av. \]

The rest of the proof proceeds as the ones in Section 3.2.

For proving (5.15), we first note that
\[ Bv - \sigma Av = B(A^{-1} - \sigma B^{-1})Av. \]
Thus
\[
\frac{\|Bv - Av\sigma\|_{B^{-1}}}{\|Av\|_{B^{-1}}} = \frac{(Av)^t(A^{-1} - \sigma B^{-1})B(A^{-1} - \sigma B^{-1})Av}{(Av)^tB^{-1}Av}.
\] (5.18)

One now needs the key insight that the right-hand side of (5.18) actually is the Rayleigh quotient of a ‘folded’ pencil. The eigenvalues are exactly the squares of the eigenvalues of the pencil \((A^{-1} - \sigma B^{-1}, B^{-1})\). This in turn shows that the right-hand side is an upper bound on the squared distance to the closest eigenvalue, analogously to the proof of Theorem 3.8.

The proof of (5.16) becomes analogous to the one of Theorem 3.10 when one substitutes orthogonality by \(B^{-1}\)-orthogonality and recognizes that the quantities involved are again from a ‘folded’ pencil as seen in (5.18).

Proceeding as in the proof of Theorem 3.13, one obtains as generalization of (3.17)
\[
\nabla^2 \psi(v) = 2\frac{\nu^t AB^{-1}Av}{\nu^tBv} \left\{ (A^{-1} - \psi(v)B^{-1}) \right. \\
\left. - 2(A^{-1} - \psi(v)B^{-1})Av^tAB^{-1}B^{-1}Av + B^{-1}Av^tA(A^{-1} - \psi(v)B^{-1}) \right\} A.
\]

As (5.12) says that \(\nu^tA(A^{-1} - \psi(v)B^{-1})Av = 0\) one has (5.17).

At last, we comment on the interesting observation that in both standard and generalized case, \(\psi\) is a quotient of Rayleigh and ‘folded’ ([15, ‘Method 2’][32]) Rayleigh quotient. This is clear in the standard case
\[
\psi(v) = \frac{\nu^tAv}{\nu^tBv},
\] (5.19)
but equally true in the generalized one:
\[
\psi(v) = \frac{\nu^tAv}{\nu^tBv}.
\] (5.20)

When \(v\) is an eigenvector, one has as expected \(\psi(v) = 1/\rho(v)\). (Note that \(A\) is assumed to be nonsingular.) Otherwise, the general relationship can be derived from (5.20) and \(Av = \rho Bv + r(v), v \perp r(v)\), to yield
\[
\psi(v) = \frac{\rho(v)\|Bv\|^2_{B^{-1}}}{\rho^2(v)\|Bv\|^2_{B^{-1}} + \|r(v)\|^2_{B^{-1}}} = \frac{\rho(v)}{\rho^2(v) + \|r(v)\|^2_{B^{-1}}}. \] (5.21)

Thus, \(\|r(v)\|^2_{B^{-1}}\) measures how much \(\psi\) deviates from \(\rho^{-1}\), in the direction towards zero. It is also this term that makes all the difference for indefinite \(A\), compared to the simply reversed Rayleigh quotient from (2.7).

6. Summary. This paper investigates harmonic Ritz values in their role as approximants to interior eigenvalues of a real symmetric matrix. Instead of the standard derivation from a Petrov Galerkin condition, we follow a functional-based approach.

It turns out that the reversed harmonic Rayleigh quotient \(\psi\) has attractive properties. In particular, it remains bounded even for an indefinite nonsingular matrix. This is important for the computation of interior eigenvalues.

Guided by the gradient \(\nabla \psi\), we derived a new type of residual \(s\). We showed that the theory for \(\psi\) and \(s\) parallels in remarkable ways the one of the standard Rayleigh quotient. Useful features including minimal residual and Weinstein-type bounds and expressions for backward error Hessian matrix have been established.
REFERENCES

2007.


