On the residual inverse iteration for nonlinear eigenvalue problems admitting a Rayleigh functional*

Cedric Effenberger† Daniel Kressner‡

January 20, 2014

Abstract

The residual inverse iteration is a simple method for solving eigenvalue problems that are nonlinear in the eigenvalue parameter. In this paper, we establish a new expression and a simple bound for the asymptotic convergence factor of this iteration in the special case that the nonlinear eigenvalue problem is Hermitian and admits a so called Rayleigh functional. These results are then applied to discretized nonlinear PDE eigenvalue problems. For this purpose, we introduce an appropriate Hilbert space setting and show the convergence of the smallest eigenvalue of a Galerkin discretization. Under suitable conditions, we obtain a bound for the asymptotic convergence factor that is independent of the discretization. This also implies that the use of multigrid preconditioners yields mesh-independent convergence rates for finite element discretizations of nonlinear PDE eigenvalue problem. A simple numerical example illustrates our findings.

1 Introduction

We consider nonlinear eigenvalue problems of the form

$$T(\lambda)x = 0, \qquad x \neq 0, \tag{1.1}$$

where $T : \mathcal{D} \to \mathbb{C}^{n\times n}$ is a continuously differentiable matrix-valued function on some open interval $\mathcal{D} \subset \mathbb{R}$. A scalar $\lambda \in \mathcal{D}$ satisfying (1.1) is called an *eigenvalue* of T, and x is called an *eigenvector* belonging to λ. The numerical solution of nonlinear eigenvalue problems has attracted renewed attention during the last years; we refer to [3, 9, 14, 21] for recent surveys.

In the following, $T(\lambda)$ is supposed to be Hermitian for every $\lambda \in \mathcal{D}$. Moreover, we assume that the scalar nonlinear equation

$$x^*T(\lambda)x = 0 \tag{1.2}$$

*The work of C. Effenberger has been supported by the SNF research module *Robust numerical methods for solving nonlinear eigenvalue problems* within the SNF ProDoc *Efficient Numerical Methods for Partial Differential Equations*.

†ANCHP, MATHICSE, EPF Lausanne, Switzerland, cedric.effenberger@epfl.ch

‡ANCHP, MATHICSE, EPF Lausanne, Switzerland, daniel.kressner@epfl.ch

admits a unique solution $\lambda \in \mathcal{D}$ for every vector x in an open set $D_\rho \subset \mathbb{C}^n$. The resulting function $\rho : D_\rho \to \mathcal{D}$, which maps x to the solution λ of (1.2), is called *Rayleigh functional*, for which we additionally assume that

$$x^* T'(\rho(x))x > 0 \qquad \forall x \in D_\rho. \tag{1.3}$$

The existence of such a Rayleigh functional entails a number of important properties for the eigenvalue problem (1.1), see [21, Sec. 115.2] for an overview. In particular, the eigenvalues in D are characterized by a min-max principle and thus admit a natural ordering. Specifically, if

$$\lambda_1 := \inf_{x \in D_\rho} \rho(x) \in \mathcal{D}$$

then λ_1 is the first eigenvalue of T.

In this paper, we study the convergence of Neumaier's [10] residual inverse iteration (RESINVIT) for computing the eigenvalue λ_1 of T and an associated eigenvector x_1. In the Hermitian case, this iteration takes the form

$$v_{k+1} = \gamma_k(v_k + P^{-1}T(\rho(v_k))v_k), \quad k = 0, 1, \ldots, \tag{1.4}$$

for an initial guess $v_0 \in \mathbb{C}^n$ and a Hermitian preconditioner $P \in \mathbb{C}^{n \times n}$. Usually, $P = -T(\sigma)$ for some shift σ not too far away from λ_1 but the general formulation (1.4) allows for more flexibility, such as the use of multigrid preconditioners. Choosing a suitable normalization coefficient $\gamma_k \in \mathbb{C}$ in (1.4) is important in practice to avoid under- or overflow in the iterates. For the sake of our convergence analysis, however, the precise choice of γ_k is not important.

In [10, Sec. 3], it was shown that (1.4) with $P = -T(\sigma)$ converges linearly to an eigenvector belonging to a simple eigenvalue, provided that σ is sufficiently close to that eigenvalue. Jarlebring and Michiels [6] derived explicit expressions for the convergence rate by viewing (1.4) as a fixed point iteration and considering the spectral radius of the fixed point iteration matrix. These analyses yield that the convergence rate of RESINVIT is proportional to $|\lambda_1 - \sigma|$. Szyld and Xue [18] have shown that this property remains true when (1.4) is replaced by an inexact variant. A two-sided variant of (1.4) was derived and analyzed by Schreiber [14].

The convergence analysis presented in this paper is tailored to the particular situation under consideration and differs significantly from [6, 10]. Our major motivation for reconsidering this question was to establish mesh-independent convergence rates when applying (1.4) with a multigrid preconditioner to the finite element discretization of a nonlinear PDE eigenvalue problem. The results from [6, 10] do not seem to admit such a conclusion, at least it is not obvious to us. On the other hand, such results are well known for the linear case $T(\lambda) = \lambda I - A$, for which (1.4) comes down to the preconditioned inverse iteration (PINVIT). In particular, the seminal work by Neymeyr [11], see also [8, 13], establishes tight expressions for the convergence of the eigenvalue and eigenvector approximations produced by PINVIT. These results are established by performing a mini-dimensional analysis of the Rayleigh quotient; the convergence of the eigenvector then follows indirectly from the convergence of the eigenvalue. The elegance of this analysis seems to be strongly tied to linear eigenvalue problems; there seems little hope to carry it over to the general nonlinear case. The analysis in this paper is significantly more simplistic and leads to weaker bounds in the linear case, but it will still allow us to establish mesh-independent convergence rates.

For a particular class of nonlinear eigenvalue problems, Solov′ëv [17] proposed and analyzed several variants of RESINVIT. In particular, their mesh-independent convergence is established. We are not aware of any other results in this direction. However, such a result has been obtained by Hackbusch [4] when applying a very different approach to solving a certain class of nonlinear PDE eigenvalue problems. This approach consists of first linearizing the eigenvalue problems and then performing the discretization.

The rest of this paper is organized as follows. In Section 2, we establish a new expression and a simple bound for the asymptotic convergence factor. This result is used in Section 3 to study the convergence for Galerkin discretizations of nonlinear eigenvalue problems in a Hilbert space setting. Section 4 presents numerical results for a simple example, verifying the obtained mesh-independent convergence.

2 Asymptotic analysis of residual inverse iteration

In this section, we analyze one step of (1.4):

$$v^+ = v + P^{-1}T(\rho(v))v. \tag{2.1}$$

As the analysis is performed in terms of the angles between v, v^+ and x_1, we may drop the normalization coefficient in (1.4).

2.1 Geometry induced by P

For any $\sigma \in D$ with $\sigma < \lambda_1$, the eigenvalues of the Hermitian matrix $T(\sigma)$ are negative [21]. Consequently, any preconditioner P that is spectrally equivalent to $-T(\sigma)$ is also positive definite. The corresponding inner product is denoted by

$$\langle v, w\rangle_P := v^*Pw,$$

which induces the norm $\|v\|_P := \sqrt{v^*Pv}$. In this geometry, the angle $\phi_P(v, x_1)$ between v and x_1 is defined by

$$\cos\phi_P(v, x_1) := \frac{\mathsf{Re}\langle v, x_1\rangle_P}{\|v\|_P\|w\|_P}$$

Given a P-orthogonal decomposition

$$v = v_1 + v_\perp, \quad v_1 \in \operatorname{span}\{x_1\}, \quad \langle v_1, v_\perp\rangle_P = 0, \tag{2.2}$$

it follows that

$$\tan\phi_P(v, x_1) = \|v_\perp\|_P/\|v_1\|_P.$$

2.2 Main result

The following theorem represents the first main result of this paper.

Theorem 2.1 *Let $T : D \to \mathbb{C}^{n\times n}$ be Hermitian and twice continuously differentiable, admitting a Rayleigh functional $\rho : D_\rho \to D$. Let $\lambda_1 \in D$ be the first eigenvalue of T, with associated eigenvector $x_1 \in D_\rho$. Suppose that the preconditioner P is Hermitian positive definite and spectrally equivalent to $-T(\lambda_1)$ on the P-orthogonal complement of the eigenvector x_1, that is, there is $\gamma < 1$ such that*

$$(1-\gamma)z^*Pz \le -z^*T(\lambda_1)z \le (1+\gamma)z^*Pz \quad \forall z \in \mathbb{C}^n : \ \langle z, x_1\rangle_P = 0. \tag{2.3}$$

Then one step (2.1) *of the residual inverse iteration satisfies*

$$\tan\phi_P(v^+, x_1) \le \gamma \cdot \varepsilon + O(\varepsilon^2),$$

provided that $\varepsilon := \tan\phi_P(v, x_1)$ is sufficiently small.

Clearly, Theorem 2.1 implies local linear convergence with asymptotic convergence rate γ. To prove this theorem, we require the following auxiliary result.

Lemma 2.2 *Under the assumptions of Theorem 2.1, we have*

$$T(\rho(v)) = T(\lambda_1) + O(\varepsilon^2).$$

Proof By [6, Proposition 2.1], the gradient of the Rayleigh functional is given by

$$\nabla\rho(v) = -\frac{1}{v^*T'(\rho(v))v} v^*T(\rho(v)),$$

where the denominator is nonzero due to (1.3). In particular,

$$\nabla\rho(x_1) = -\frac{1}{x_1^*T'(\lambda_1)x_1} x_1^*T(\lambda_1) = 0,$$

where we used that T is Hermitian. Therefore, the Taylor expansion of $T(\rho(v))$ at x_1 is given by

$$T(\rho(v)) = T(\lambda_1) + T'(\lambda_1)\nabla\rho(x_1)\cdot(v-x_1) + O(\|v-x_1\|_P^2) = T(\lambda_1) + O(\|v-x_1\|_P^2).$$

By a suitable scaling, we can always choose x_1 such that $\|v - x_1\|_P = O(\varepsilon)$. ∎

Proof of Theorem 2.1 The result of Lemma 2.2 allows us to rewrite (2.1) as

$$v^+ = v + P^{-1}T(\lambda_1)v + O(\varepsilon^2). \tag{2.4}$$

Note that the second term in (2.4) is P-orthogonal to x_1:

$$\langle x_1, P^{-1}T(\lambda_1)v\rangle_P = x_1^*PP^{-1}T(\lambda_1)v = 0.$$

Together with $P^{-1}T(\lambda_1)v = P^{-1}T(\lambda_1)v_\perp$ for the P-orthogonal decomposition (2.2) of v, this gives the approximate P-orthogonal decomposition

$$v^+ = v_1 + \big(v_\perp + P^{-1}T(\lambda_1)v_\perp\big) + O(\varepsilon^2).$$

Hence,

$$\tan\phi_P(v^+, x_1) = \frac{\|v_\perp + P^{-1}T(\lambda_1)v_\perp\|_P}{\|v_1\|_P} + O(\varepsilon^2) \le \gamma \cdot \tan\phi_P(v, x_1) + O(\varepsilon^2),$$

with the convergence factor

$$\gamma := \max_{\substack{z\neq 0 \\ \langle z, x_1\rangle_P = 0}} \frac{\|(I + P^{-1}T(\lambda_1))z\|_P}{\|z\|_P} = \max_{\substack{y\neq 0 \\ y^*P^{1/2}x_1 = 0}} \frac{\|(I + P^{-1/2}T(\lambda_1)P^{-1/2})y\|_2}{\|y\|_2}.$$

Since $I + P^{-1/2}T(\lambda_1)P^{-1/2}$ is Hermitian, it follows that

$$\gamma = \max_{\substack{z\neq 0 \\ y^*P^{1/2}x_1 = 0}} \left|1 + \frac{y^*P^{-1/2}T(\lambda_1)P^{-1/2}y}{y^*y}\right| = \max_{\substack{z\neq 0 \\ \langle z, x_1\rangle_P = 0}} \left|1 + \frac{z^*T(\lambda_1)z}{z^*Pz}\right|. \tag{2.5}$$

This completes the proof. ∎

2.3 Convergence rate for $P = -T(\sigma)$

In the following, we assume that λ_1 is a simple eigenvalue, which implies that the eigenvalues of the Hermitian matrix $T(\lambda_1)$ satisfy

$$0 = \mu_1 > \mu_2 \ge \mu_3 \ge \cdots \ge \mu_n.$$

This assumptions allows us to derive bounds for the convergence rate γ when using the Hermitian positive definite preconditioner $P = -T(\sigma)$. For this purpose, we will use that the continuous differentiability of $T(\lambda)$ and (1.2) imply the existence of $\sigma_0 < \lambda_1$, $L > 0$, $\delta > 0$ such that

$$\|T'(\xi)\|_2 \le L, \quad x_1^*T'(\xi)x_1 \ge \delta, \qquad \forall \xi \in [\sigma_0, \lambda_1], \tag{2.6}$$

where we assumed w.l.o.g. that $\|x_1\|_2 = 1$.

Proposition 2.3 *With the quantities introduced above, suppose that σ is chosen such that $\sigma_0 \le \sigma < \lambda_1$ and*

$$\alpha := \frac{L(L+\delta)}{|\mu_2|\delta}(\lambda_1 - \sigma) < \frac{1}{2}. \tag{2.7}$$

Then the preconditioner $P = -T(\sigma)$ satisfies the condition (2.3) *of Theorem 2.1 with a convergence rate $\gamma > 0$ satisfying*

$$\gamma \le \frac{\alpha}{1-\alpha} < 1. \tag{2.8}$$

Proof We let $E := T(\sigma) - T(\lambda_1)$ such that $P = -T(\sigma) = -T(\lambda_1) - E$. Note that (2.6) implies

$$\|E\|_2 \le L(\lambda_1 - \sigma), \quad |x_1^*Ex_1| \ge \delta(\lambda_1 - \sigma).$$

Given an arbitrary vector $z \in \mathbb{C}^n$ with $\langle z, x_1 \rangle_P = 0$, we decompose $z = z_1 + z_\perp$ such that $z_1 \in \operatorname{span}\{x_1\}$ and $\langle z_1, z_\perp \rangle = 0$. From

$$0 = z^* P x_1 = -z^* E x_1 \quad \Rightarrow \quad z_1^* E x_1 = -z_\perp^* E x_1, \tag{2.9}$$

it follows that

$$\|z_1\|_2 = \frac{|z_1^* E x_1|}{|x_1^* E x_1|} = \frac{|z_\perp^* E x_1|}{|x_1^* E x_1|} \leq \frac{L}{\delta}(\lambda_1 - \sigma)\|z_\perp\|_2.$$

Using $z_1^* P z = 0$, the identity

$$\begin{aligned} z^* P z &= z_1^* P z + z_\perp^* P z_\perp + z_\perp^* P z_1 = z_\perp^* P z_\perp + z_\perp^* P z_1 \\ &= -z_\perp^* T(\lambda_1) z_\perp - z_\perp^* E z_\perp - z_\perp^* E z_1 \end{aligned} \tag{2.10}$$

is established. Hence, we obtain

$$\begin{aligned} \left| 1 + \frac{z^* T(\lambda_1) z}{z^* P z} \right| &= \left| 1 - \left[1 + \frac{z_\perp^* E z_\perp}{z_\perp^* T(\lambda_1) z_\perp} + \frac{z_\perp^* E z_1}{z_\perp^* T(\lambda_1) z_\perp} \right]^{-1} \right| \\ &\leq \left| 1 - \left[1 - L(\lambda_1 - \sigma) \frac{\|z_\perp\|_2^2 + \|z_1\|_2 \|z_\perp\|_2}{|\mu_2| \|z_\perp\|_2^2} \right]^{-1} \right| \\ &\leq \left| 1 - [1 - \alpha]^{-1} \right| = \frac{\alpha}{1 - \alpha}. \end{aligned}$$

Together with the characterization (2.5) of γ, this completes the proof. ∎

Proposition 2.3 implies a convergence rate proportional to $\lambda_1 - \sigma$. This fact was already noted in [10], see also [6, Thm. 4.2]. The novel aspect of Proposition 2.3 is that it gives a simple upper bound on the convergence rate.

2.4 Convergence rate for P spectrally equivalent to $-T(\sigma)$

Any symmetric positive definite matrix P is spectrally equivalent to $-T(\sigma)$, that is, there are constants $0 < C_L \leq C_U$ such that

$$C_L z^* P z \leq -z^* T(\sigma) z \leq C_U z^* P z \quad \forall z \in \mathbb{C}^n. \tag{2.11}$$

By rescaling P, we can always assume that $C_U = 1$ and hence $C_L \leq 1$, see [11] for a related discussion. When using RESINVIT to drive a subspace method, such as the Arnoldi method presented in [19], then the scaling of P becomes irrelevant.

The spectral equivalence (2.11) implies that condition (2.3) of Theorem 2.1 holds with the convergence rate γ from Proposition 2.3 replaced by

$$\tilde{\gamma} = 1 - C_L(1 - \gamma).$$

We thus still obtain asymptotically linear convergence, but the convergence rate may deteriorate as the preconditioner becomes worse, that is, C_L approaches 0. For standard preconditioners, this may happen when σ approaches λ_1 and thus $T(\sigma)$ becomes nearly singular. Multigrid preconditioners that take the near singularity of $T(\sigma)$ into account have been discussed in, e.g., [2].

3 Galerkin discretization of nonlinear PDE eigenvalue problem

In this section, we aim to give some insight into the convergence of RESINVIT for Galerkin discretizations of nonlinear PDE eigenvalue problems, using a variation of Proposition 2.3. This requires to set up a suitable mathematical framework for performing this analysis. To the best of our knowledge, the analyses of discretized nonlinear PDEs available in the literature only cover particular classes of nonlinear eigenvalue problems, see, e.g., [4, 16]. In the following, we will sketch a framework covering general self-adjoint nonlinear PDE eigenvalue problems that admit a Rayleigh functional. We omit details not needed for the purpose of this paper; this topic certainly merits a separate investigation.

3.1 Hilbert space setting

To allow for infinite-dimensional nonlinear eigenvalue problems, Voss and Werner [22, 20] have considered the setting of a selfadjoint and bounded operator $T(\lambda)$ on a Hilbert space. We consider a different setting, which is related to the developments by Solov'ëv [16].

Let V and H be real or complex Hilbert spaces with the corresponding scalar products denoted by $\langle\cdot,\cdot\rangle_V$ and $\langle\cdot,\cdot\rangle_H$, respectively. The embedding $V \hookrightarrow H$ is assumed to be dense and compact. In particular, we may consider $V = H^1(\Omega)$ and $H = L^2(\Omega)$ for some open, bounded Lipschitz domain Ω. We then consider a nonlinear eigenvalue problem in the variational formulation

$$\begin{aligned} &\text{Find } (u,\lambda) \in V \times \mathcal{D} \text{ such that} \\ &a(u,v,\lambda) = 0 \quad \forall v \in V. \end{aligned} \tag{3.1}$$

For each fixed $\lambda \in \mathcal{D}$, $a(\cdot,\cdot,\lambda)$ is assumed to be a symmetric continuous sesquilinear form on $V \times V$. Moreover, we assume continuous differentiability with respect to λ. In particular, this means that the derivative $a_\lambda(\cdot,\cdot,\lambda)$ is again a symmetric continuous sesquilinear form on $V \times V$.

When considering λ fixed, the linear eigenvalue problem associated with $\bar{a}(\cdot,\cdot,\lambda) := -a(\cdot,\cdot,\lambda)$ takes the form

$$\begin{aligned} &\text{Find } (u,\mu) \in V \times \mathcal{D} \text{ such that} \\ &\bar{a}(u,v,\lambda) = \mu\langle u,v\rangle_H \quad \forall v \in V. \end{aligned} \tag{3.2}$$

The following common assumption is made:

$$\exists c(\lambda), s(\lambda) \quad \text{such that} \quad \bar{a}(v,v,\lambda) \geq c(\lambda)\|v\|_V^2 - s(\lambda)\|v\|_H^2 \quad \forall v \in V. \tag{3.3}$$

Under the above assumptions, it is well known [12] that (3.2) has discrete eigenvalues $\mu_j(\lambda) \in \mathbb{R}$, $j = 1, 2, \ldots$, satisfying

$$s(\lambda) \leq \mu_1(\lambda) \leq \mu_2(\lambda) \leq \cdots .$$

Moreover, the variational characterization

$$\mu_j(\lambda) = \inf_{\substack{U \subset V \\ \dim U = j}} \max_{\substack{u \in U \\ u \neq 0}} \frac{\bar{a}(u,u,\lambda)}{\langle u,u\rangle_H} \tag{3.4}$$

holds.

Example 3.1 ([16]) *The nonlinear eigenvalue problem*

$$\begin{aligned} &-u''(\xi) = \lambda u(\xi), \quad \xi \in \Omega = (0,1), \\ &u(0) = 0, \quad -u'(1) = \varphi(\lambda) u(1), \end{aligned} \tag{3.5}$$

where $\varphi(\lambda) = \frac{\lambda\eta m}{\lambda - \eta}$ *with constants* $\eta, m > 0$*, models a string with an elastically attached mass. The existence of a Rayleigh functional for this problem is discussed in [3]. Let* $V = \{v \in H^1(\Omega) : v(0) = 0\}$ *and* $H = L^2(\Omega)$*, with*

$$\langle u, v \rangle_V = \int_\Omega u'v' \, d\xi, \quad \langle u, v \rangle_H = \int_\Omega uv \, d\xi.$$

Then the bilinear form in the variational formulation (3.1) *of* (3.5) *is given by*

$$a(u, v, \lambda) = -\langle u, v \rangle_V - \varphi(\lambda) u(1) v(1) + \lambda \langle u, v \rangle_H,$$

which is clearly bounded on V*. Assuming that* $\varphi(\lambda)$ *is positive in the region of interest, it follows that the ellipticity condition* (3.3) *is satisfied with* $s(\lambda) = \lambda$ *and* $c(\lambda) = 1$*.* ◇

3.2 Rayleigh functionals

To derive a variational characterization for the eigenvalues (3.1), we need to make two assumptions.

(A1) The equation $a(u, u, \lambda) = 0$ has at most one solution $\lambda \in \mathcal{D}$ for each $u \in V \setminus \{0\}$.

The set of all u admitting a solution λ to $a(u, u, \lambda) = 0$ is denoted by $D(\rho) \subset V$, which can be shown to be open, following the arguments in [20]. By (A1), there exists a *Rayleigh functional* $\rho : D(\rho) \mapsto \mathcal{D}$ mapping $u \in D(\rho)$ to the solution $\lambda \in \mathcal{D}$.

(A2) For all $u \in D(\rho)$, it holds that

$$a_\lambda(u, u, \rho(u)) > 0.$$

Similar to the developments in [22, 20], one may conclude from (A1) and (A2) a min-max principle for the nonlinear eigenvalue problem (3.1). In particular, if

$$\lambda_1 := \inf_{x \in D_\rho} \rho(x) \in \mathcal{D}$$

then λ_1 is the smallest eigenvalue of (3.1).

3.3 Convergence of Galerkin discretization

Let $V_1 \subset V_2 \subset \cdots \subset V$ be a nested sequence of finite-dimensional subspaces of V. We assume that this sequence is asymptotically dense in the sense that $\Pi_k v \to v$ as $k \to \infty$ for all $v \in V$, where Π_k denotes the orthogonal projection from V onto V_k in the inner product $\langle \cdot, \cdot \rangle_V$. The Galerkin projection of the nonlinear eigenvalue problem (3.1) with respect to V_k reads

$$\begin{aligned} &\text{Find } (u, \lambda) \in V_k \times \mathcal{D} \text{ such that} \\ &a(u, v, \lambda) = 0 \quad \forall v \in V_k. \end{aligned} \tag{3.6}$$

Letting $v_1, \ldots, v_{N_k}$ denote a basis of V_k, this is equivalent to the $N_k \times N_k$ nonlinear matrix eigenvalue problem

$$T_k(\lambda)x = 0, \qquad \big[T_k(\lambda)\big]_{ij} = a(v_i, v_j, \lambda), \quad \forall i, j = 1, \ldots, N_k.$$

Clearly, the restriction of ρ to $D(\rho) \cap V_k$, which we will again denote by ρ for convenience, constitutes a Rayleigh functional for the nonlinear eigenvalue problem (3.6). Hence, if $\lambda_1^{(k)} := \inf_{v \in D(\rho) \cap V_k} \rho(v) \in \mathcal{D}$, then $\lambda_1^{(k)}$ is the smallest eigenvalue of T_k in $\mathcal{D}$. As $k \to \infty$, the eigenvalue $\lambda_1^{(k)}$ of the nonlinear matrix eigenvalue problem (3.6) represents increasingly accurate approximations to the eigenvalue λ_1.

Proposition 3.2 *Let λ_1 and $\lambda_1^{(k)}$ be as above for $k = 1, 2, \ldots$. Then $\lambda_1^{(k)}$ converges to λ_1 as $k \to \infty$.*

Our proof of Proposition 3.2 will be based on a perturbation analysis of the Rayleigh functional ρ. Such an analysis has been given by Schwetlick and Schreiber in [15, Corollary 18]. Proposition 3.3 below is a refinement of their result. It relies on assumptions (A1) and (A2), of which the latter implies

$$\delta_a := a_\lambda(u_1, u_1, \lambda_1) > 0.$$

Moreover, the bilinear forms a and a_λ are bounded on $V \times V$, implying that there are constants $C_a, C_{a'}$ such that

$$|a(v, w, \lambda_1)| \le C_a \|v\|_V \|w\|_V, \quad |a_\lambda(v, w, \lambda_1)| \le C_{a'} \|v\|_V \|w\|_V, \quad \forall v, w \in V.$$

Additionally, we need to assume the Lipschitz continuity of $a_\lambda(\cdot, \cdot, \lambda)$ with respect to λ in a neighborhood of λ_1. More specifically, there is a constant $L_{a'}$ such that

$$|a_\lambda(v, w, \mu) - a_\lambda(v, w, \lambda_1)| \le L_{a'} |\mu - \lambda_1| \|v\|_V \|w\|_V$$

holds for all $\mu \in \big[\lambda_1 - 2\frac{\delta_a}{L_{a'}}, \lambda_1 + 2\frac{\delta_a}{L_{a'}}\big]$.

Proposition 3.3 *With the assumptions and quantities described above, let (u_1, λ_1) with $\|u_1\|_V = 1$ be a simple eigenpair of T. Choose $0 < \theta < 1$ and define*

$$\beta = 1 + \sqrt{1 + \Phi}, \qquad \Phi = \frac{\theta}{1 - \theta} \cdot \frac{L_{a'} C_a}{2 C_{a'}^2}.$$

Then every $v \neq 0$ with $\sin \angle(v, u_1) \le \frac{\theta \delta_a}{\beta C_{a'}}$ belongs to $D(\rho)$ and

$$|\rho(v) - \lambda_1| \le \frac{1}{1 - \theta} \cdot \frac{C_a}{\delta_a} \cdot \sin^2 \angle(y, x).$$

Proof Because neither $\angle(v, u_1)$ nor $\rho(v)$ depend on the scaling of v, we may rescale v such that $v = u_1 - e$ with e orthogonal to v in the inner product on V. In particular, this implies $\|v\|_V^2 + \|e\|_V^2 = \|u_1\|_V^2 = 1$ and $\|e\|_V = \sin \angle(v, u_1)$.

Since the function $f(\mu) := a(v, v, \mu)$ is continuously differentiable, it holds that

$$f(\mu) = f(\lambda_1) + (\mu - \lambda_1) \cdot R(\mu), \tag{3.7}$$

with

$$R(\mu) = \int_0^1 f'\big(\lambda_1 + \tau(\mu - \lambda_1)\big)\,\mathrm{d}\tau = f'(\lambda_1) + \int_0^1 \big[f'\big(\lambda_1 + \tau(\mu - \lambda_1)\big) - f'(\lambda_1)\big]\,\mathrm{d}\tau. \tag{3.8}$$

Exploiting that u is an eigenvector, we find that $f(\lambda_1) = a(u-e, u-e, \lambda_1) = a(e, e, \lambda_1)$ and conclude

$$|f(\lambda_1)| \le C_a \|e\|_V^2 \le C_a \frac{\theta^2 \delta_a^2}{\beta^2 C_{a'}^2}. \tag{3.9}$$

To bound $R(\mu)$ from below, we use

$$f'(\lambda_1) = \delta_a - a_\lambda(u_1, e, \lambda_1) - a_\lambda(e, v, \lambda_1) \ge \delta_a - 2C_{a'}\|e\|_V \ge (1 - \tfrac{2}{\beta}\theta)\delta_a$$

and the Lipschitz continuity of f' inherited from a_λ:

$$\int_0^1 \big[f'\big(\lambda_1 + \tau(\mu - \lambda_1)\big) - f'(\lambda_1)\big]\,\mathrm{d}\tau \le \int_0^1 L_{a'}\tau|\mu - \lambda_1|\,\mathrm{d}\tau = \tfrac{L_{a'}}{2}|\mu - \lambda_1|.$$

For $\mu \in [\mu^-, \mu^+]$ with $\mu^\pm := \lambda_1 \pm \frac{2}{L_{a'}}(1 - \frac{2}{\beta})\theta\delta_a$, combining the above estimates yields

$$R(\mu) \ge (1 - \theta)\delta_a > 0.$$

By construction, β satisfies $\beta(\beta - 2) = \Phi$, which is equivalent to $1 - \frac{2}{\beta} = \frac{\Phi}{\beta^2} > 0$. Therefore, using the definition of Φ and the estimate for $|f(\lambda_1)|$ in (3.9), we obtain

$$f(\mu^+) \ge (\mu^+ - \lambda_1) \cdot R(\mu^+) - |f(\lambda_1)| \ge \frac{\theta\delta_a^2}{\beta^2}\left[\frac{2\Phi}{L_{a'}}(1 - \theta) - \theta \cdot \frac{C_a}{C_{a'}^2}\right] = 0.$$

Analogously, one shows that $f(\mu^-) \le 0$. Consequently, f has a zero inside the interval $[\mu^-, \mu^+]$, proving that $v \in D(\rho)$ and $\rho(v) \in [\mu^-, \mu^+]$, i.e.,

$$|\rho(v) - \lambda_1| \le \tfrac{2}{L_{a'}}\big(1 - \tfrac{2}{\beta}\big)\theta\delta_a.$$

Inserting $\mu = \rho(v)$ into the expansion (3.7) and rearranging leads to the bound

$$|\rho(v) - \lambda_1| = \frac{|f(\lambda_1)|}{|R(\rho(v))|} \le \frac{C_a\|e\|_V^2}{(1 - \theta)\delta_a}$$

as claimed. ∎

Although we will utilize it only for the smallest eigenvalue λ_1, it is worth noting that Proposition 3.3 holds for any simple eigenvalue of (3.1).

We are now in the position to prove Proposition 3.2.

Proof of Proposition 3.2 Since λ_1 is an eigenvalue of (3.1), there exists a corresponding eigenvector $u_1 \neq 0$ such that $\lambda_1 = \rho(u_1)$. Since the spaces V_k are asymptotically dense in V, we have $\Pi_k u_1 \to u_1$ as $k \to \infty$, where Π_k denotes again the orthogonal projector onto V_k with respect to $\langle \cdot, \cdot \rangle_V$. Therefore,

$$\sin \angle(\Pi_k u_1, u_1) = \frac{\|u_1 - \Pi_k u_1\|_V}{\|u_1\|_V}$$

converges to zero as $k \to \infty$. Consequently, by Proposition 3.3, $\Pi_k u_1 \in D(\rho)$ for sufficiently large k and $|\rho(\Pi_k u_1) - \lambda_1| \to 0$ as $k \to \infty$. The proof is now finished by concluding from the definitions of λ_1 and $\lambda_1^{(k)}$ that

$$\lambda_1 \leq \lambda_1^{(k)} \leq \rho(\Pi_k u_1) \to \lambda_1.$$

∎

3.4 Convergence of RESINVIT

In the following, we use the results from Section 2 to gain insight into the asymptotic convergence of RESINVIT for computing the smallest eigenpair $(u_1^{(k)}, \lambda_1^{(k)})$ of the discretized eigenvalue problem (3.6). This requires us to reconsider and generalize the various requirements of Proposition 2.3.

1. Let us first consider the discretization of the linear eigenvalue problem (3.2):

$$\begin{aligned} &\text{Find } (u, \mu) \in V_k \times \mathcal{D} \text{ such that} \\ &\bar{a}(u, v, \lambda) = \mu \langle u, v \rangle_H \quad \forall v \in V_k. \end{aligned} \tag{3.10}$$

   Letting $\mu_1^{(k)}(\lambda) \leq \mu_2^{(k)}(\lambda) \leq \cdots$ denote the eigenvalues of (3.10), the variational characterization (3.4) implies $\mu_2^{(k)}(\lambda) \geq \mu_2(\lambda)$. Moreover, $\lambda_1^{(k)} \overset{k\to\infty}{\to} \lambda$ by Proposition 3.2; therefore the continuity of eigenvalues yields $\mu_2^{(k)}\big(\lambda_1^{(k)}\big) \overset{k\to\infty}{\to} \mu_2^{(k)}(\lambda_1)$. In particular, $\mu_2(\lambda_1) > 0$ implies that there is $\delta_\mu > 0$ such that $\mu_2^{(k)}\big(\lambda_1^{(k)}\big) > \delta_\mu$ for sufficiently large k. Thus,

$$a(v, v, \lambda) \geq \delta_\mu \|v\|_H^2 \geq \frac{\delta_\mu}{C_{V\hookrightarrow H}^2} \|v\|_V^2 \qquad \forall v \in V_k : \langle v, u_1^{(k)} \rangle_H = 0, \tag{3.11}$$

   where $u_1^{(k)}$ denotes an eigenvector of (3.6) belonging to λ_1. By the compact embedding $V \hookrightarrow H$, there is a constant $C_{V\hookrightarrow H}$ such that $\|v\|_V \leq C_{V\hookrightarrow H} \|z\|_H$ for all $v \in V$.

2. Since $u_1^{(k)}$ is also an eigenvector belonging to the simple eigenvalue zero of the linear eigenvalue problem (3.10), it follows [7] that the suitably normalized sequence $u_1^{(k)}$ converges in V to u_1. Hence, by Assumption (A2) there is $\delta_a > 0$ such that $a_\lambda\big(u_1^{(k)}, u_1^{(k)}, \lambda_1^{(k)}\big) \geq \delta_a$ for sufficiently large k. Together with the boundedness of a_λ, this implies the existence of constants $\sigma_0 < \lambda_1$, $L > 0$, $\delta > 0$ such that for all $v, w \in V$:

$$|a_\lambda(v, w, \xi)| \leq L \|v\|_V \|w\|_V, \quad a_\lambda(u_1^{(k)}, u_1^{(k)}, \xi) \geq \delta, \quad \forall \xi \in \big[\sigma_0, \lambda_1\big]. \tag{3.12}$$

To connect the conditions above to the findings of Section 2, let us rewrite the properties (3.11) and (3.12) in terms of matrices. Let M_H and M_V denote the mass matrices associated with the inner products $\langle\cdot,\cdot\rangle_H$ and $\langle\cdot,\cdot\rangle_V$, respectively:

$$\big[M_H(\lambda)\big]_{ij} = \langle v_i, v_j\rangle_M, \quad \big[M_V(\lambda)\big]_{ij} = \langle v_i, v_j\rangle_V, \quad \forall i,j = 1,\dots,N_k.$$

For vectors $y, z \in \mathbb{C}^{N_k}$, we set $\langle y,z\rangle_H = y^* M_H z$, $\langle y,z\rangle_V = y^* M_V z$, which induce the norms $\|\cdot\|_H$ and $\|\cdot\|_V$ on $\mathbb{C}^{N_k}$.

1. The eigenvalue problem (3.10) is equivalent to the $N_k \times N_k$ linear matrix eigenvalue problem $T_k(\lambda) + \mu M_H$. Property (3.11) states that its second eigenvalue is bounded from below by δ_μ. More specifically,

$$|z^* T_k(\lambda) z| \geq \delta_\mu \|z\|_H^2 \geq \frac{\delta_\mu}{C_{V\hookrightarrow H}^2}\|z\|_V^2 \qquad \forall z \in \mathbb{C}^{N_k} : \langle z, x_1^{(k)}\rangle_H = 0, \tag{3.13}$$

   where $x_1^{(k)}$ is an eigenvector of $T_k(\lambda)$ belonging to $\lambda_1^{(k)}$.

2. Property (3.12) states that

$$\|T_k'(\xi)\|_V \leq L, \quad \big(x_1^{(k)}\big)^* T'(\xi) x_1^{(k)} \geq \delta, \quad \forall \xi \in \big[\sigma_0, \lambda_1\big], \tag{3.14}$$

   where $\|\cdot\|_V$ denotes the matrix norm induced $\langle\cdot,\cdot\rangle_V$.

If $M_H = M_V = I$ and $\lambda_1 = \lambda_1^{(k)}$ then the properties (3.13) and (3.14) correspond exactly to what is needed in Proposition (2.3). The following theorem, which represents the second main result of this paper, gives a more general version of this proposition.

Theorem 3.4 *Let $\varepsilon > 0$. Suppose that k is sufficiently large such that* (3.13) *and* (3.14) *hold, and $\lambda_1^{(k)} - \lambda_1 \leq \varepsilon$. Moreover, suppose that σ is chosen such that $\sigma_0 \leq \sigma < \lambda_1$ and*

$$\alpha := \frac{C_{V\hookrightarrow H}^2 L(L+\delta)}{\delta_\mu \delta}(\lambda_1 + \varepsilon - \sigma) < \frac{1}{2}.$$

Then the preconditioner $P = -T_k(\sigma)$ satisifies the condition (2.3) *of Theorem 2.1 with a convergence rate $\gamma > 0$ satisfying*

$$\gamma \leq \frac{\alpha}{1-\alpha} < 1.$$

Proof The proof is along the lines of the proof of Proposition 2.3. For convenience, we write x_1 instead of $x_1^{(k)}$ and normalize such that $\|x_1\|_V = 1$. Letting $E := T(\sigma) - T(\lambda_1^{(k)})$, (3.14) implies

$$\|E\|_V \leq L(\lambda_1 + \varepsilon - \sigma), \quad \big|x_1^* E x_1\big| \geq \delta(\lambda_1 + \varepsilon - \sigma).$$

Given an arbitrary vector $z \in \mathbb{C}^n$ with $\langle z, x_1\rangle_P = 0$, we decompose $z = z_1 + z_\perp$ such that $z_1 \in \operatorname{span}\{x_1\}$ and $\langle z_1, z_\perp\rangle_H = 0$. Then, noting that $z_1 = \beta\|z_1\|_V x_1$ for some $\beta \in \mathbb{C}$ with $|\beta| = 1$,

$$\|z_1\|_V = \frac{|z_1^* E x_1|}{|x_1^* E x_1|} = \frac{|z_\perp^* E x_1|}{|x_1^* E x_1|} \leq \frac{L}{\delta}(\lambda_1 + \varepsilon - \sigma)\|z_\perp\|_V.$$

Using the identity (2.10), we obtain

$$\begin{aligned}\left|1+\frac{z^*T_k(\lambda_1^{(k)})z}{z^*Pz}\right| &= \left|1-\left[1+\frac{z_\perp^*Ez_\perp}{z_\perp^*T_k(\lambda_1^{(k)})z_\perp}+\frac{z_\perp^*Ez_1}{z_\perp^*T(\lambda_1^{(k)})z_\perp}\right]^{-1}\right| \\ &\le \left|1-\left[1-C_{V\hookrightarrow H}^2L(\lambda_1+\varepsilon-\sigma)\frac{\|z_\perp\|_V^2+\|z_1\|_V\|z_\perp\|_V}{\delta_\mu\|z_\perp\|_V^2}\right]^{-1}\right| \\ &\le \left|1-[1-\alpha]^{-1}\right| = \frac{\alpha}{1-\alpha},\end{aligned}$$

completing the proof. ∎

Theorem 3.4 establishes a convergence rate that remains bounded away from 1 as k grows. It is important to note that the shift σ also remains constant and does not need to be chosen progressively closer to λ_1 as k grows, which would lead to practical difficulties. By the discussion in Section 2.4, these statements still hold when using a preconditioner P satisfying the spectral equivalence (2.11) with constants C_L, C_U independent of k.

4 Numerical experiments

We consider Example 3.1 with $\eta = m = 1$, which corresponds to the parameters used in [16, Sec. 8]. The aim is to compute the smallest eigenvalue within the interval $(1, \infty)$, which, for the continuous problem, amounts to $\lambda_1 = 4.482024295$.

Discretizing the problem by linear finite elements on a uniform grid of size $h = 2^{-k}$ leads to a nonlinear eigenvalue problem (1.1) for a matrix $T_k(\lambda)$. As the preconditioner in RESINVIT, we employ one W-cycle of the symmetric multigrid algorithm with Jacobi smoother described in [5], applied to $-T_k(\sigma)$ for a suitably chosen $\sigma < \lambda_1$. This preconditioner is spectrally equivalent to $-T_k(\sigma)$, uniformly with respect to the mesh size h [1].

Table 1 reports the numbers of iterations required to reach an accuracy of 10^{-6} in the eigenvalue for different levels of mesh refinement and different choices of the parameter σ. Clearly, the iteration numbers do not increase as the mesh is further refined; in fact, they even slightly decrease. This confirms our theoretical findings from Section 3. As expected, the number of iterations reduces as σ approaches λ_1. All computations have been performed under MATLAB 7.13 (R2011b).

5 Conclusions

We have derived a new expression for the asymptotic convergence rate of the residual inverse iteration when computing the smallest eigenvalue of a Hermitian nonlinear eigenvalue problems admitting a Rayleigh functional. This expression allowed us to conclude mesh-independent convergence rates for Galerkin discretizations of certain infinite-dimensional problems.

	#iterations		
h	$\sigma = 0$	$\sigma = 2$	$\sigma = 4$
2^{-5}	12	8	5
2^{-6}	12	8	5
2^{-7}	12	8	5
2^{-8}	12	7	4
2^{-9}	12	7	4
2^{-10}	12	7	4
2^{-11}	11	7	4
2^{-12}	11	7	4
2^{-13}	11	6	4
2^{-14}	10	6	4
2^{-15}	10	6	4
2^{-16}	10	6	4
2^{-17}	9	6	3
2^{-18}	9	5	3

Table 1: Iteration numbers of RESINVIT applied to a finite element discretization of Example 3.1 for different grid sizes h. The employed preconditioners are spectrally equivalent to $-T_h(\sigma)$ for different values of σ.

References

[1] J. H. Bramble and J. E. Pasciak. New convergence estimates for multigrid algorithms. *Math. Comp.*, 49(180):311–329, 1987.

[2] Z. Cai, J. Mandel, and S. McCormick. Multigrid methods for nearly singular linear equations and eigenvalue problems. *SIAM J. Numer. Anal.*, 34(1):178–200, 1997.

[3] C. Effenberger. *Robust solution methods for nonlinear eigenvalue problems*. PhD thesis, EPF Lausanne, Switzerland, 2013. Available at `http://anchp.epfl.ch/students`.

[4] W. Hackbusch. On the computation of approximate eigenvalues and eigenfunctions of elliptic operators by means of a multi-grid method. *SIAM J. Numer. Anal.*, 16(2):201–215, 1979.

[5] W. Hackbusch. *Multi-Grid Methods and Applications*. Number 4 in Springer Series in Computational Mathematics. Springer, Berlin, 1985.

[6] E. Jarlebring and W. Michiels. Analyzing the convergence factor of residual inverse iteration. *BIT*, 51(4):937–957, 2011.

[7] T. Kato. *Perturbation theory for linear operators*. Classics in Mathematics. Springer-Verlag, Berlin, 1995. Reprint of the 1980 edition.

[8] A. V. Knyazev and K. Neymeyr. Gradient flow approach to geometric convergence analysis of preconditioned eigensolvers. *SIAM J. Matrix Anal. Appl.*, 31(2):621–628, 2009.

[9] V. Mehrmann and H. Voss. Nonlinear eigenvalue problems: A challenge for modern eigenvalue methods. *GAMM Mitteilungen*, 27, 2004.

[10] A. Neumaier. Residual inverse iteration for the nonlinear eigenvalue problem. *SIAM J. Numer. Anal.*, 22(5):914–923, 1985.

[11] K. Neymeyr. A geometric theory for preconditioned inverse iteration. I. Extrema of the Rayleigh quotient. *Linear Algebra Appl.*, 322(1-3):61–85, 2001.

[12] M. Reed and B. Simon. *Methods of modern mathematical physics. IV. Analysis of operators.* Academic Press [Harcourt Brace Jovanovich Publishers], New York, 1978.

[13] T. Rohwedder, R. Schneider, and A. Zeiser. Perturbed preconditioned inverse iteration for operator eigenvalue problems with applications to adaptive wavelet discretization. *Adv. Comput. Math.*, 34(1):43–66, 2011.

[14] K. Schreiber. *Nonlinear Eigenvalue Problems: Newton-type Methods and Nonlinear Rayleigh Functionals.* PhD thesis, Institut für Mathematik, TU Berlin, 2008.

[15] H. Schwetlick and K. Schreiber. Nonlinear Rayleigh functionals. *Linear Algebra Appl.*, 436(10):3991–4016, 2012.

[16] S. I. Solov′ëv. The finite element method for symmetric eigenvalue problems with nonlinear occurrence of the spectral parameter. *Zh. Vychisl. Mat. Mat. Fiz.*, 37(11):1311–1318, 1997.

[17] S. I. Solov′ëv. Preconditioned iterative methods for a class of nonlinear eigenvalue problems. *Linear Algebra Appl.*, 415(1):210–229, 2006.

[18] D. B. Szyld and F. Xue. Local convergence analysis of several inexact Newton-type algorithms for general nonlinear eigenvalue problems. *Numer. Math.*, 123(2):333–362, 2013.

[19] H. Voss. An Arnoldi method for nonlinear eigenvalue problems. *BIT*, 44(2):387–401, 2004.

[20] H. Voss. A minmax principle for nonlinear eigenproblems depending continuously on the eigenparameter. *Numer. Linear Algebra Appl.*, 16(11-12):899–913, 2009.

[21] H. Voss. Nonlinear eigenvalue problems. In L. Hogben, editor, *Handbook of Linear Algebra.* Chapman & Hall/CRC, FL, 2013. To appear.

[22] H. Voss and B. Werner. A minimax principle for nonlinear eigenvalue problems with applications to nonoverdamped systems. *Math. Methods Appl. Sci.*, 4(3):415–424, 1982.