ON A CERTAIN ALGORITHM OF EIGENVALUE LOCALIZATION FOR NORMAL OPERATORS

KRZYSZTOF MOSZYŃSKI and TERESA REGIŃSKA

Institute of Mathematics, Polish Academy of Sciences, Warsaw, Poland

1. Introduction

The paper is concerned with the problem of localization of eigenvalues of a normal operator $A$ in a Hilbert space $H$, $A: D(A) \to H$. An iterative method for computing bounds for eigenvalues, based on iterative algorithm of minimization of functionals of two variables is considered.

For illustration, let us describe the simplest variant of this method for a selfadjoint operator $A$. For any $x \in D(A)$, $x \neq 0$, and $\lambda \in \mathbb{R}$ we have

$$\text{dist}(\lambda, \sigma(A)) = \frac{1}{\| (\lambda - A)^{-1} \|} \leq \frac{\| (\lambda - A)x \|}{\| x \|}$$

where $\sigma(A)$ denotes the spectrum of $A$. Computing the norm of residuum, we get a certain rough bound for the distance between given $\lambda \in \mathbb{R}$ and $\sigma(A)$. In order to improve this bound, the right-hand side can be minimized with respect to $x$ over some fixed finite dimensional subspace $X_N \subset D(A)$. Next, minimizing this term over $\lambda \in \mathbb{R}$ we get a new value $\lambda_1$ at which this infimum is attained. The bound

$$\mu_1 = \frac{\| (\lambda_1 - A)x \|}{\| x \|}$$

for $\text{dist}(\lambda_1, \sigma(A))$ is less than the previous one for $\lambda$. This process can be continued. In such a way we get a sequence of approximate eigenvalues $\{\lambda_i\}$ as well as the computable bounds $\mu_i$ for the distances between $\lambda_i$ and the spectrum $\sigma(A)$.

Such an iterative method in a more general case when $X_N \notin D(A)$ was proposed in [5] and its convergence was analysed in [6]. In this work we present some more detailed analysis of such a method.
The general procedure consists in minimizing a certain functional $\Phi$ on the right-hand side of the following inequality:

\[
\inf_{s \in \sigma(A)} \left| \frac{\lambda - s}{s} \right| \leq \Phi(\lambda, x).
\]

This process is discussed in Section 4 for a certain class of two variable functionals $\Phi$. The idea of convergence proof is taken from [6].

In Section 5 we generalize the Kuttler and Sigillito \textit{a posteriori} inequality [5] (which implies inequality (1.1)) for a wider class of operators. It allows us to apply the considered method not only in selfadjoint case but also for a wider class of operators including nonselfadjoint ones.

Section 6 contains more detailed study of the case where $A$ is selfadjoint and $\Phi(\lambda, x) = \| (A - \lambda)x \|^2$. In this case, a characterization of a limit of the sequence $\{\lambda_n\}$ is given. This process is compared with the Galerkin (finite element) method described in Section 2.

2. Galerkin method

For simplicity let us assume that $A$ is a bounded linear operator in a Hilbert space $H$ with a scalar product $(,)$ and norm $\| \|$. Let $\{X_N\}$ be a sequence of finite dimensional subspaces of $H$ such that

$$ \forall v \in H \quad \inf_{x \in X_N} \| v - x \| \xrightarrow{N \to \infty} 0. $$

Let $\{\hat{X}_N, p_N, r_N\}$ be corresponding approximation $H$ defined as follows:

\( \hat{X}_N = C^N \) (the complex $N$-dimensional space) with the Euclidian norm denoted by $\| \|_N$;

\[
p_N : \hat{X}_N \to X_N, \quad \forall \alpha \in C^N, \quad \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N) \quad p_N \alpha = \sum_{j=1}^{N} \alpha_j \xi_j,
\]

\[
r_N : H \to \hat{X}_N, \quad \forall x \in H \quad r_N x = [(x, \xi_1), \ldots, (x, \xi_N)]^T,
\]

where $\xi_1, \ldots, \xi_N$ is an orthonormal basis of $X_N$. Then

$$ \Pi_N \overset{df}{=} p_N r_N : H \to X_N $$

is the orthogonal projection onto $X_N$ and

\[
\hat{A}_N \overset{df}{=} r_N A p_N = \{ (A \xi_j, \xi_j) \}_{j=1}^{N}
\]

is the standard Galerkin approximation of $A$ with respect to the basis $\xi_1, \xi_2, \ldots, \xi_N$. If $A$ is selfadjoint then the matrix $\hat{A}_N$ is hermitian. Put

\[
A_N \overset{df}{=} \Pi_N A \Pi_N.
\]
In this case the following relations can easily be proved

\[(2.4) \quad \forall N > 0 \forall x \in \mathbb{X}_N \| p_N x \| = \| x \|;\]
\[(2.5) \quad \forall N > 0 \forall x \in H |r_N x|_N \leq \| x \|;\]
\[(2.6) \quad \exists K < \infty \forall N > 0 \| A_N \| = \| \hat{A} \| \leq K;\]
\[(2.7) \quad \forall x \in H \|(I - \Pi_N)x\| \to 0 \quad \text{and} \quad \|(A - A_N)x\| \to 0 \quad \text{when} \ N \to \infty.\]

Thus general theorems on spectral approximation given in [2] imply the following one.

**Theorem 1.** If \( s \) is an isolated eigenvalue of \( A \) then there exists a sequence \( \{s_N\} \) of eigenvalues of \( \hat{A}_N \) converging to \( s \) as \( N \to \infty \). If the algebraic multiplicity of \( s \) is one, then \( s_N \) is also single for sufficiently large \( N \). Moreover, there exist an integer \( N_0 \) and a positive constant \( \gamma \) such that for \( N > N_0 \) the interval \( (s - \gamma, s + \gamma) \) contains \( s_N \) and no other eigenvalue of \( \hat{A}_N \).

3. An implicit function theorem

Let \( X, Y, Z \) be Banach spaces. Let \( B_X(w, \delta), B_Y(w, \delta) \) denote the ball with the center \( w \) and the radius \( \delta \) in the space \( X \) and \( Y \), respectively. Let us consider the equation

\[(3.1) \quad G(x, y) = 0\]

where \( G: \Omega \overset{def}{=} B_X(x_0, \delta) \times B_Y(y_0, \delta) \to Z \). \( x_0, y_0 \) are given elements and \( \delta \) is a fixed positive number.

In Section 6 we will use the following two versions of an implicit function theorem (cf. [3]):

**Theorem 2.** Let us assume that

1° \( G: \Omega \to Z \) is a function of the class \( C^p \), \( p \geq 2 \);

2° the Fréchet derivative \( \frac{\partial}{\partial y} G(x_0, y_0) \) is an isomorphism of \( Y \) onto \( Z \);

3° there exist positive constants \( c_k \), \( k = 0, \ldots, p \) such that

\[\left\| \left( \frac{\partial}{\partial y} G(x_0, y_0) \right)^{-1} \right\| \leq c_0 \quad \forall x, y \in \Omega \quad \left\| G^{(k)}(x, y) \right\| \leq c_k, \quad k = 1, \ldots, p\]

where \( G^{(k)} \) is the \( k \)-th Fréchet derivative of \( G \). Then there exist positive constants \( a, b, \delta \) depending uniquely on \( \delta \) and \( c_i \), \( i = 0, 1, 2 \), and the constants \( M_k \), \( k = 1, \ldots, p \), depending uniquely on \( c_i \), \( i = 0, \ldots, k \) such that

(i) \( \forall (x, y) \in B_X(x_0, a) \times B_Y(y_0, b) \quad \frac{\partial}{\partial y} G(x, y): Y \to Z \) is an isomorphism of
Y onto Z and
\[ \left\| \left( \frac{\partial}{\partial y} G(x, y) \right)^{-1} \right\| \leq 2c_0; \]

(ii) if \( \| G(x_0, y_0) \| \leq d \), then there exists a function \( g: B_x(x_0, a) \to B_y(y_0, b) \) of the class \( C^1 \) such that for any \( x \in B_x(x_0, a) \) \( y = g(x) \) is the unique solution of the equation (3.1) in \( B_x(x_0, a) \times B_y(y_0, b) \). Moreover,
\[ \forall x \in B_x(x_0, a) \quad \| g^{(k)}(x) \| \leq M_k \quad \text{for } k = 1, 2, \ldots, p. \]

THEOREM 3. Let the assumptions of Theorem 2 be satisfied with \( \| G(x_0, y_0) \| \leq d \). Let \( q: B_x(x_0, a) \to B_y(y_0, b) \) be a function of class \( C^1 \). Then there exists a constant \( K \) depending uniquely on \( c_0 \) and \( \delta \) such that
\[ \forall x \in B(x, a) \quad \| g(x) - q(x) \| \leq K \| G(x, q(x)) \|. \]

4. Minimization of functionals of two variables

Let us consider a functional \( \Phi: C \times H \to C \) of the form
\[ \forall \lambda \in C \quad \forall x \in D \subset H \quad \Phi(\lambda, x) = (B_2 x, x) \]
where \( B_2: D \to H \) is a positive selfadjoint linear operator and \( B_\lambda = B_0 + \lambda B_1 + \lambda^2 B^*_1 + |\lambda|^2 B_2 \). The domain \( D \) is supposed to be dense and independent of \( \lambda \). We will assume that \( B_2 \) is positive definite, i.e.
\[ \exists \alpha, 0 < \alpha < \infty \quad \forall x \in D \quad (B_2 x, x) \geq \alpha \| x \|^2. \]

Let \( X_N \) be a fixed finite dimensional subspace contained in \( D \) and let \( \Pi_N \) be the orthogonal projection of \( H \) onto \( X_N \). Then, for \( y \in X_N \), \( \Phi(\lambda, y) = (B_\lambda y, y) = (B_\lambda y, \Pi_N y) = (\Pi_N B_\lambda y, y) \), i.e. the functional \( \Phi|_{C \times X_N} \) is related to the operator \( \Pi_N B_\lambda|_{X_N} \) which is the orthogonal Galerkin approximation of \( B_\lambda \).

Now, we are going to discuss the convergence of two sequences \( \{ x_j \}_{j=1}^\infty \) and \( \{ \lambda_j \}_{j=0}^\infty \) defined as follows:
\[ x_j \in X_N \] is such that \( \Phi(\lambda_j, x_j) = \inf \{ \Phi(\lambda, x) \mid x \in X_N, \| x \| = 1 \} \),
\[ \lambda_{j+1} \in C \] is such that \( \Phi(\lambda_{j+1}, x) = \inf \{ \Phi(\lambda, x) \mid \lambda \in A \} \),
where \( \lambda_0 \) is a given starting point. We will consider two cases
(i) \( A \) is the whole complex plane \( C \),
(ii) \( A \) is a straight line in \( C \),
\[ A = \{ \lambda(t) = u(t) + iv(t) \in C \mid u(t) = a + tb, \ v(t) = a' + tb', \ t \in \mathbb{R} \} \quad \text{(e.g. } A = \mathbb{R} \).

Since \( \Phi(\lambda_j, x_j)/(x, x) \) is the Rayleigh quotient for \( B_\lambda \), it follows that in the both cases (i) and (ii) its infimum over finite dimensional space \( X_N \) is attained at an arbitrary normed eigenvector corresponding to the minimal eigenvalue
\( \mu_n(\lambda_j) \) of the operator \( \Pi_N B_{\lambda_j} |_{x_N} \). If the minimal eigenvalue of \( \Pi_N B_{\lambda_j} |_{x_N} \) is of multiplicity greater than 1, \( x_j \) is not uniquely determined by (4.2). In this case we take as \( x_j \) one of elements satisfying the minimum condition (4.2). We have

\[
(4.4) \quad \mu_n(\lambda_j) = (B_{\lambda_j} x_j, x_j)
\]

and evidently \( \mu_n(\lambda_0) \geq \mu_n(\lambda_1) \geq \ldots \geq \mu_n(\lambda_i) \geq 0 \), since \( B_2 \) is positive definite by the assumption. Thus the sequence \( \{\mu_n(\lambda_j)\}_{j=0}^{\infty} \) being decreasing and bounded from below is convergent, but not necessarily convergent to zero. For \( \lambda = u + iv \) (\( u, v \in \mathbb{R} \))

\[
\Phi(\lambda, x_j, x_j) = (B_0 x_j, x_j) + 2u(\text{Re} B_1 x_j, x_j) - 2v(\text{Im} B_1 x_j, x_j) + (u^2 + v^2)(B_2 x_j, x_j),
\]

where \( \text{Re} B_1 = (B_1 - B_1^*)/2 \), \( \text{Im} B_1 = (B_1 + B_1^*)/2i \). From the conditions

\[
\frac{\partial \Phi}{\partial u} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial v} = 0,
\]

in the case (i) \( (A = C) \) we get

\[
(4.5) \quad \lambda_{j+1} = u_{j+1} + iv_{j+1},
\]

where

\[
u_{j+1} = -(\text{Re} B_1 x_j, x_j)/(B_2 x_j, x_j), \quad v_{j+1} = -(\text{Im} B_1 x_j, x_j)/(B_2 x_j, x_j).
\]

In case (ii), finding \( \frac{d}{dt} \Phi(\lambda(t), x_j) \) for \( \lambda(t) = u(t) + iv(t), u(t) = a + tb, u(t) = a' + tb' \), (where \( a, a', b, b' \) are constants defining \( A \)) we easily see that

\[
\lambda_{j+1} = \lambda(t_{j+1}), \quad \text{where} \quad t_{j+1} \text{ satisfies the equation}
\]

\[
(4.6) \quad b\left(\frac{A_j}{C_j} + u(t_{j+1})\right) + b'\left(-\frac{D_j}{C_j} + v(t_{j+1})\right) = 0
\]

with

\[
(4.7) \quad A_j = (\text{Re} B_1 x_j, x_j), \quad D_j = (\text{Im} B_1 x_j, x_j), \quad C_j = (B_2 x_j, x_j).
\]

Observe that

\[
\frac{d^2}{dt^2} \Phi(\lambda(t), x_j) = 2C_j(b^2 + b'^2) > 0.
\]

**Lemma 1.** Let \( \Phi(\lambda, x) \) satisfy the assumptions (4.1) and let

\[3 \alpha > 0, \quad \forall x \in X_N, \quad (B_2 x, x) \geq \alpha(x, x).\]

Then for the sequence \( \{\lambda_j\} \) defined by (4.2) and (4.3) with \( A = C \) or \( A \) being a straight line in \( C \), the following condition holds:

\[
|\lambda_{j+1} - \lambda_j| \to 0 \quad \text{when} \quad j \to \infty.
\]
Proof. In both cases (i) and (ii), we have

$$\mu_N(\lambda_j) - \mu_N(\lambda_j+1) \geq \Phi(j, x_j) - \Phi(j+1, x_j).$$

Using the notation (4.7) and formulae (4.5), we get

$$\Phi(j, x_j) - \Phi(j+1, x_j) = C_j|\lambda_j - \lambda_j+1|^2$$

in case (i). In case (ii), after simple transformations we obtain

$$\Phi(j, x_j) - \Phi(j+1, x_j) = C_j|\lambda_j - \lambda_j+1|^2$$

$$+ 2C_j\left((u(t_j) - u(t_{j+1}))\left(A_j + u(t_{j+1})\right) + (v(t_j) - v(t_{j+1}))\left(-\frac{D_j}{C_j} + v(t_{j+1})\right)\right).$$

Since $u(t_j) - u(t_{j+1}) = (t_j - t_{j+1})b$ and $v(t_j) - v(t_{j+1}) = (t_j - t_{j+1})b'$, from (4.6) it follows that the last term vanishes. So, the convergence of the sequence \( \{\mu_N(\lambda_j)\}_{j=0}^{\infty} \) implies the lemma.

Theorem 4. Let the assumptions of Lemma 1 be satisfied and let \( \{\lambda_j\} \) be defined by iterative formulae (4.2), (4.3) for a fixed straight line \( \Lambda \) (case (ii)). Then the sequence \( \{\lambda_j\} \) is convergent.

Proof. Since \( \dim X_N < \infty \) and \( C_j > \alpha(x_j, x_j) \), \( A_j/C_j \) and \( B_j/C_j \) are uniformly bounded with respect to \( j \). From (4.6) it follows that the sequence \( \{\lambda_j\} \) is bounded, and hence, it contains a converging subsequence. If every converging subsequence converges to the same limit then, clearly, \( \{\lambda_j\} \) converges to this limit too.

Suppose that there are two different accumulation points \( r_1 \) and \( r_2 \) of \( \{\lambda_j\} \) \( (r_1, r_2) \in \Lambda \) since \( \lambda_j \in \Lambda \) for every \( j \). Let \( \{\lambda_{k_0}\} \) and \( \{\lambda_{l_0}\} \) be subsequences converging to \( r_1 \) and \( r_2 \), respectively. Clearly, we can choose them in such a way that

$$\ldots < k_j < l_j < k_{j+1} < l_{j+1} < k_{j+2} < \ldots$$

Let \( r \) be arbitrary point of interval \([r_1, r_2]\). Let \( \lambda_{m_j} \) be the nearest element to \( r \) among \( \lambda_j, k_j \leq s \leq l_j \). Lemma 1 implies that

$$\forall \delta > 0 \exists j_0 \forall j > j_0 \quad |\lambda_{m_j} - r| < \delta.$$ 

Hence, either the sequence \( \{\lambda_j\} \) converges or it contains subsequences converging to every point in some interval \([r_1, r_2] \subset \Lambda \).

Let us suppose that \( \{\lambda_j\} \) does not converge. Let \( r \) be arbitrary element of \([r_1, r_2]\) and let \( \lambda_{r_j} \to r \) as \( j \to \infty \). We have \( \mu_N(\lambda_{r_j}) \to \mu_N \) for any \( r \in [r_1, r_2] \), where \( \mu_N \) is the limit of \( \{\mu_N(\lambda_j)\} \). For any fixed basis \( v_1, \ldots, v_N \) in \( V_N \)

$$\det[\{B_{2r_j, v_k, v_j} - \delta_{kl}\mu_N(\lambda_{r_j})\}] = 0$$

holds for \( j = 0, 1, 2, \ldots \); hence, by continuity,

$$\det[\{B_{r_k, v_k, v_j} - \delta_{kl}\mu_N\}] = 0$$
for any \( r \in [r_1, r_2] \subseteq \Lambda \). But \( r \in \Lambda \) implies \( r = u(t) + iv(t) \) for all \( t, \theta_1 \leq t \leq \theta_2 \), with some \( \theta_1, \theta_2, \theta_1 < \theta_2 \). Observe that the left-hand side of the last inequality is a nonzero polynomial with respect to \( t \in \mathbb{R} \). This means that this polynomial vanishes for any \( t \in [\theta_1, \theta_2] \), which is impossible. Hence the sequence \( \{ \lambda_j \} \) has to converge.

**Remark 1.** Let \( B_\lambda \equiv (\lambda - A)^2 \), where \( A \) is a selfadjoint linear operator with a dense domain \( D(A) \) in \( H \) and \( \lambda \in \mathbb{R} \). As it was mentioned in Introduction,

\[
\text{dist}(\hat{\lambda}, \sigma(A))^2 \leq \frac{(B_\lambda x, x)}{(x, x)}
\]

for arbitrary \( x \in D(A) \), \( x \neq 0 \). The iteration process (4.2), (4.3) for \( X_N \subset D(A) \) and \( \Lambda = \mathbb{R} \) was used by Kuttler and Sigillito in [5] in order to obtain better computable bounds for eigenvalues of \( A \). Namely, by (4.4),

\[
\text{dist}(\lambda_j, \sigma(A))^2 \leq \mu_N(\lambda_j) \leq \mu_N(\lambda_{j-1}) \leq \ldots \leq \mu_N(\lambda_0), \quad j = 1, 2, \ldots
\]

for \( \lambda_0 = \hat{\lambda} \). According to Theorem 4, we have \( \lambda_j \to \hat{\lambda} \). This convergence result was first obtained in [6].

**Remark 2.** Let \( A \) be a selfadjoint differential operator with a domain \( D(A) \subset L^2(\Omega), \Omega \subset \mathbb{R}^n \), and let \( \hat{A} \) be its extension generated by the differential expression only (without boundary conditions). Let \( \partial \Omega \) be Lipschitz continuous. Assume that we know an explicit a priori inequality, i.e. that there are known constants \( c_1 \) and \( c_2 \) such that

\[
\forall w \in D(\hat{A}) \quad \| w \|_0^2 \leq c_1 \| \hat{A} w \|_0 + c_2 \| l w \|_2^2 (\partial \Omega)
\]

where \( l \) is an operator corresponding to boundary conditions generating \( A \). For this case and for \( \lambda \in \mathbb{R} \), Kuttler and Sigillito proved the following \textit{a priori—a posteriori} inequality (cf. [5]):

\[
(4.7) \quad \inf_{s \in \sigma(A)} \left| \frac{\lambda - \sigma^2}{s} \right| \leq \frac{\Phi(\lambda, x)}{(x, x)} \quad \text{for} \quad x \in D(\hat{A})
\]

where \( \Phi(\lambda, x) = c_1 \| \hat{A} x - \lambda x \|_0 + c_2 \| l x \|_2^2 (\partial \Omega) \).

To improve this bound they propose to apply the iteration process (4.2), (4.3) for \( A = \mathbb{R} \) and \( X_N \subset D(\hat{A}) \). It should be mentioned that functions belonging to \( X_N \) need not satisfy the boundary conditions. This fact simplifies the algorithm and is one of the most important advantages of their method. Convergence of this method was proved and a certain characterization of a limit point of \( \{ \lambda_j \} \) is given in [6]. Let \( \hat{\lambda} \) denote the limit of \( \lambda_j \) and let \( \hat{x} \) be an arbitrary accumulation point of \( \{ x_j \} \). Then \( \{ \hat{\lambda}, \hat{x} \} \) is a critical point for the functional \( \Phi(\lambda, x) \) on the set \( \mathbb{R} \times S_N \), \( S_N = \{ \psi \in X_N, \| \psi \| = 1 \} \) (cf. [6], th. 3).
5. A posteriori inequality for eigenvalues

In [5], for the case of selfadjoint operator $A$ with compact inverse, a priori-a posteriori inequalities for eigenvalues of $A$ were considered (cf. (4.7)). Using the completeness of the system of eigenfunctions of selfadjoint operator with compact inverse, the authors proved a posteriori inequality of the form:

(5.1) \[ \forall \lambda \in \mathbb{R} \quad \inf_{s \in \sigma(A)} \frac{|\lambda - s|}{s} \leq \frac{\|w\|}{\|x\|}, \]

where $x \neq 0$ is an arbitrary element of the domain of a certain extension $\tilde{A}$ of $A$, and $w$ is a solution of an auxiliary equation

(5.2) \[ \tilde{A}w = \tilde{A}x - \lambda x, \quad w - x \in D(A). \]

For some differential operators and their extensions generated by the differential expression only there exists an explicit a priori estimation for the solution of (5.2) (cf. [7]). In these cases we have the so-called a priori-a posteriori inequality of the form (4.7), for which the iteration method (4.2), (4.3) can be used.

Now, we are going to show that inequality (5.1) holds not only for selfadjoint operators with compact inverse but also for normal operators. It is a simple consequence of the spectral mapping theorem.

**Theorem 5.** Let $A$ be a closed normal operator with a dense domain in a Hilbert space $H$. If $0 \notin \sigma(A)$, then $\forall \lambda \in \mathbb{R} \ \forall x \in H, \ x \neq 0$,

(5.3) \[ \inf_{s \in \sigma(A)} \frac{|\lambda - s|}{s} \leq \frac{\|w\|}{\|x\|}, \]

where $w \in H$ is a solution of the following auxiliary problem:

(5.4) \[ A(w - x) = -\lambda x, \ w - x \in D(A) \]

**Proof.** The function $f(z) \overset{df}{=} (\lambda - z)/\lambda$ is holomorphic in $C \setminus \{0\}$. Since, by the assumption, $\sigma(A) \subset C \setminus \{0\}$, thus $f(A)$ is well defined and

\[ f(A) = (\lambda - A)A^{-1}. \]

Since $A$ is normal, $f(A)$ is normal too. So, we have

\[ \min_{\tau \in \sigma(f(A))} |\tau| = \frac{1}{\|f(A)^{-1}\|}. \]

Using now the spectral mapping theorem, which says that $\sigma(f(A)) = f(\sigma(A))$, we conclude that

\[ \min_{s \in \sigma(A)} \frac{|\lambda - s|}{\lambda} = \frac{1}{\|A(\lambda - A)^{-1}\|}. \]
Thus for arbitrary \( x \in H, \ x \neq 0 \) we have
\[
\min_{s \in \sigma(A)} \left| \frac{\lambda - s}{\lambda} \right| \leq \frac{\| (\lambda - A) A^{-1} x \|}{\| x \|}.
\]
Now, it is enough to observe that the element
\[
w = (\lambda - A) A^{-1} x
\]
is the solution of (5.4).

**Remark 3.** If \( w \) is the solution of (5.4) then \( w \) is also the solution of (5.2) for an arbitrary extension \( \tilde{A} \) of \( A \) such that \( w \in D(\tilde{A}) \subset H \).

Theorem 5 allows us to apply the iteration method (4.2), (4.3) for computing bounds for eigenvalues of a wider class of operators. Namely, if there exists explicit a priori estimation for the solution \( w \) of the equation (5.2)
\[
\| w \|^2 \leq \alpha \Phi(\lambda, x)
\]
with a known constant \( \alpha \) and \( \Phi \) of the form (4.1) then we can apply the method (4.2), (4.3) for improving the bound (5.3). If \( A \) is a straight line then \( \{ \lambda_j \} \) is convergent (Theorem 3).

## 6. The selfadjoint case

Now, we are going to discuss a behaviour of the process (4.2), (4.3) for
\[
\Phi(\lambda, x) = \| (\lambda - A) x \|^2, \quad \lambda \in \mathbb{R}, \ x \in X_N
\]
when the starting point \( \lambda_0 \) is in a neighbourhood of a single eigenvalue of \( A \).

For simplicity, in the following we assume that \( A \) is a linear bounded selfadjoint operator on a real Hilbert space \( H \) and \( A : H \rightarrow V \subset H \) where \( V \) is a dense subspace with a norm \( \| \cdot \|_1 \) stronger than the norm \( \| \cdot \| \) in \( H \). It should be mentioned that many results discussed below hold in more general situations, and the above restrictions are chosen so as to avoid some tedious calculations in deriving and presenting results.

Using the notation of Section 2, let us introduce the following auxiliary matrices:
\[
\tilde{B}_N = r_N A^2 p_N, \quad \tilde{R}_N = \tilde{B}_N - \tilde{A}_N^2.
\]
They are symmetric since \( A \) is selfadjoint.

**Lemma 2.** Let \( \{ X_N \} \) be such that \( \forall v \in V \inf_{x \in X_N} \| v - x \| \leq \varepsilon_N \| v \|_1 \) and \( \varepsilon_N \rightarrow 0 \) when \( N \rightarrow \infty \). If \( r_N, p_N \) are defined by (2.1) and \( \Pi_N = p_N r_N \) then
\[
\forall x \in H \quad (r_N x)^T \tilde{R}_N r_N x = \| (I - \Pi_N) A \Pi_N x \|
\]
and
\[
| \tilde{R}_N x |_N \leq \| A \|^2 \varepsilon_N^2 \rightarrow 0 \quad \text{when} \ N \rightarrow \infty.
\]
Proof. For $\alpha = r_N x$ we have
\[
\|(1 - \Pi_N)\lambda \Pi_N x\|^2 = \|(1 - \Pi_N)\lambda \Pi_N x, A \Pi_N x\rangle = (A p_N x - p_N \lambda x, A p_N x) \\
= \sum_{i,j=1}^N (A^2 \xi_j, \xi_i) x_i x_j - \sum_{i,j,k=1}^N (A \xi_i, \xi_k) (A \xi_j, \xi_k) x_j x_i = \alpha^T (\tilde{B}_N - \tilde{A}_N) \alpha,
\]
since $A = A^*$. Moreover, since $\Pi_N$ is an orthogonal projection and $A \Pi_N x \in V$,
\[
\|(1 - \Pi_N)\lambda \Pi_N x\| = \inf_{y \in \Pi_N x} \|\lambda \Pi_N x - y\| \leq \varepsilon_N \|\lambda \Pi_N x\| \leq \varepsilon_N \|\lambda \Pi_N x\| r_N x = \varepsilon_N \|\lambda \Pi_N x\| r_N x.
\]
Thus, for symmetric matrix $\tilde{R}_N$ we have
\[
\forall \alpha \in \Pi_N x, \quad \alpha^T \tilde{R}_N \alpha \leq \varepsilon_N^2 \|\lambda \Pi_N x\|^2 r_N^2,
\]
what ends the proof.

Let us consider the function $\mu_N$ of real variable $\lambda \in \mathbb{R}$
\[
(6.3) \quad \mu_N(\lambda) \overset{df}{=} \inf \{\|\lambda - A\| x\|^2 | x \in \Pi_N x, \|x\| = 1\}.
\]
This function appears in the process (4.2), (4.3) and, as it was mentioned in
Section 4, it is equal to the minimal eigenvalue of the operator $\Pi_N (\lambda - A)^2 |_{\Pi_N x}$.

Since
\[
\Pi_N (\lambda - A)^2 |_{\Pi_N x} = p_N [(\lambda - \tilde{A}_N)^2 + \tilde{R}_N] r_N |_{\Pi_N x},
\]
$\mu_N(\lambda)$ is also the minimal eigenvalue of the matrix
\[
(6.4) \quad (\lambda - \tilde{A})^2 + \tilde{R}_N.
\]
Let $u_N(\lambda)$ denote a normed eigenvector of $(\lambda - \tilde{A}_N)^2 + \tilde{R}_N$ corresponding to the
eigenvalue $\mu_N(\lambda)$. Let $s$ be a single eigenvalue of $A$, and let $\{s_N\}, N = N_0$, be a sequence of single eigenvalues of $\tilde{A}_N$ converging to $s$ (the existence of $\{s_N\}$ follows from Theorem 1). Let us denote by $s_N, s_N^1, \ldots, s_N^N$ the eigenvalues of
$\tilde{A}_N$ and by $v_N, v_N^1, \ldots, v_N^N$ the corresponding orthonormal eigenvectors. Similarly, denote by $\mu_N(\lambda), \mu_N^1(\lambda), \ldots, \mu_N^{N-1}(\lambda)$ eigenvalues of the symmetric positively semidefinite matrix $(\lambda - \tilde{A}_N)^2 + \tilde{R}_N$, repeated according to their multiplicity, and by $u_N(\lambda), u_N^1(\lambda), \ldots, u_N^{N-1}(\lambda)$ corresponding orthonormal eigenvectors.

**Lemma 3.** Let $C$ be $N \times N$ symmetric matrix with eigenvalues $\mu, \mu_1, \ldots, \mu_{N-1}$
and corresponding orthonormal eigenvectors $u, u_1, \ldots, u_{N-1}$. Let
\[
M = \begin{bmatrix}
\mu - C & u \\
2u^T & 0
\end{bmatrix}.
\]
If $\mu$ is single, then
\[
M^{-1} = \begin{bmatrix}
Q & u/2 \\
u^T & 0
\end{bmatrix},
\]
where

$$Q = \sum_{j=1}^{N-1} \frac{1}{\mu - \mu_j} u_j u_j^T,$$

Proof. We have

$$\begin{bmatrix} Q & u/2 \\ u^T & 0 \end{bmatrix} \begin{bmatrix} \mu - C & u^T \\ 2u^T & 0 \end{bmatrix} = \begin{bmatrix} Q(\mu - C) + uu^T & Qu \\ u^T(\mu - C) & u^T u \end{bmatrix},$$

$Qu = 0$ and $u^T u = 1$ by orthonormality of the system $u, u_1, \ldots, u_{N-1}$. Moreover, $u^T(\mu - C) = 0$ because $u$ is an eigenvector corresponding to $\mu$. Finally,

$$Q(\mu - C) = \left( \sum_{j=1}^{N-1} \frac{1}{\mu - \mu_j} u_j u_j^T \right)(\mu - C) = \sum_{j=1}^{N-1} u_j u_j^T$$

since $u_j^T(\mu - C) = (\mu - \mu_j)u_j^T$. Hence

$$Q(\mu - C) + uu^T = \sum_{j=1}^{N-1} u_j u_j^T + uu^T = I$$

and the result follows.

As in Section 3, let $B_w(w, \delta)$ denote the ball in $X$ with a center $w$ and a radius $\delta$. Let us define the function $F_N$:

$$(6.5) \quad F_N(\lambda, \hat{R}, u, \mu) = \begin{bmatrix} (\mu - (\lambda - \hat{A}_N)^2 - \hat{R})^2 \\ u^T u - 1 \end{bmatrix}$$

of the variables $\lambda \in R, \hat{R} \in S^N, \mu \in R, u \in R^N$ with values in $R^{N+1}$, where $S^N$ is the space of $N \times N$ real symmetric matrices.

**Theorem 6.** Let $s$ be a single eigenvalue of $A$ and let $\{s_n\}$ be a sequence of eigenvalues of $\hat{A}_N$ converging to $s$. If $N$ is so large that $s_n$ is single, then there exist constants $a_1, b$ independent of $N$ such that, for any $\lambda \in B_{R}(s_n, a_1), \mu(\lambda)$ is a single eigenvalue of (6.4),

$$\mu_n: B_R(s_n, a_1) \to B_R(0, b), \quad u_n: B_R(s_n, a_1) \to B_{X^n}(v_n, b))$$

and $u_n$ and $\mu_n$ are arbitrarily differentiable in $B_{R}(s_n, a_1)$.

Proof. Apply Theorem 2 for $G = F_N, X = R \times S^N, Y = R^N \times R, Z = R^{n+1}$ with the following norms:

for $x = \begin{bmatrix} \lambda \\ \hat{R} \end{bmatrix} \in X, \quad \|x\| = |\lambda| + |\hat{R}|_N$;

for $y = \begin{bmatrix} u \\ \mu \end{bmatrix} \in Y, \quad \|y\| = (|u|^2_N + |\mu|^2)^{1/2}$;

for $z \in Z, \quad \|z\| = (z^T z)^{1/2}$. 

Let \( x_0 = \begin{bmatrix} s_N \\ 0 \end{bmatrix} \in X, \ y_0 = \begin{bmatrix} v_N \\ 0 \end{bmatrix} \in Y \). Observe that we can choose \( \delta > 0 \) arbitrarily, because \( F_N \) is everywhere defined and regular. Clearly, we can find positive constants \( c_1, \ldots, c_p \) (for any \( p \)) depending uniquely on \( K \) and \( \delta \), where \( \forall N \lvert \hat{A}_N \rvert_N < K \) (see (2.6)). These constants majorize the norms of derivatives of \( F_N \), as in Theorem 2. Let us find now the constant \( c_0 \) majorizing \( \left\| \frac{\partial}{\partial y} F_N(x_0, y_0)^{-1} \right\| \). We have
\[
\frac{\partial}{\partial y} F_N = \begin{bmatrix} \frac{\partial}{\partial \mu} F_N & \frac{\partial}{\partial \mu} F_N \\ \frac{\partial}{\partial u} F_N & \frac{\partial}{\partial u} F_N \end{bmatrix} = \begin{bmatrix} \mu - (\lambda - \hat{A}_N)^2 - \hat{R} & u \\ 2u^T & 0 \end{bmatrix}
\]
and
\[
\frac{\partial}{\partial y} F_N(x_0, y_0) = \begin{bmatrix} -(s_N - \hat{A}_N)^2 & v_N \\ 2u^T & 0 \end{bmatrix} = M_N.
\]
Since, for \( \lambda = s_N \) and \( \hat{R}_N = 0 \), we have \( \mu_N(\lambda) = 0, \mu_N = (s_N - s_N^2)^2, u_N = v_N, u_N^j = v_N^j, j = 1, 2, \ldots, n - 1 \), from Lemma 3 we get
\[
M_N^{-1} = \begin{bmatrix} Q_N & v_N/2 \\ v_N^T & 0 \end{bmatrix}
\]
with
\[
Q_N = - \left[ \sum_{j=1}^{N-1} (v_N^j v_N^j T)/(s_N - s_N^j)^2 \right].
\]
We need to majorize the norm \( \lvert M_N^{-1} \rvert_{N+1} \). For \( \zeta = \begin{bmatrix} x \\ q \end{bmatrix}, x \in \mathbb{R}^N, q \in \mathbb{R} \)
\[
M_N^{-1} \zeta = \begin{bmatrix} Q_N x + v_N q/2 \\ v_N^T x \end{bmatrix}.
\]
Since \( Q_N v_N = 0 \), we have
\[
\lvert M_N^{-1} \zeta \rvert^2_{N+1} = \lvert Q_N x \rvert^2 + \lvert v_N \rvert^2 q^2/4 + (v_N^T x)^2 \leq \lvert Q_N \rvert^2 \lvert x \rvert^2 + q^2 + \lvert x \rvert^2,
\]
\[
\leq \lvert Q_N \rvert^2 + 1 \lvert x \rvert^2.
\]
If \( y \in \mathbb{R}^N \) and \( \lvert y \rvert_N = 1 \) then \( y = \sum_{j=1}^{N-1} \alpha_j v_N^j + z v_N \) with \( \sum_{j=1}^{N-1} \alpha_j^2 + x^2 = 1 \). Thus
\[
\lvert Q_N y \rvert^2 = \sum_{j=1}^{N-1} \alpha_j^2 (s_N - s_N^j)^2 \leq \max_{1 \leq j \leq N-1} 1/(s_N - s_N^j)^2.
\]
Finally,
\[
\lvert M_N^{-1} \rvert_{N+1} \leq \lvert Q_N \rvert^2 + 1 \lvert x \rvert^2 \leq 1 + \lvert Q_N \rvert_N = 1 + \max_{1 \leq j \leq N-1} 1/(s_N - s_N^j)^2.
\]
Taking into account Theorem 1, we see that \( |M_N^{-1}|_{N+1} \) can be majorized by some \( c_0 > 0 \) independently of \( N \). Hence Theorem 2 implies existence of functions \( u_N(\lambda, \tilde{R}) \) and \( \mu_N(\lambda, \tilde{R}) \) for \( |\lambda-s|+|\tilde{R}|_N < a, a > 0 \), where \( a \) does not depend on \( N \). We shall treat the variable \( \tilde{R} \) as a small parameter (see Lemma 2). We can assume that \( |\tilde{R}|_N < a_1 = a/2 \). Then the theorem follows.

In order to analyze behaviour of the function \( \mu_N(\lambda) \) in \( (s_N-a_1, s_N+a_1) \) we need some additional information about \( u_N(\lambda) \).

**Lemma 4.** Let the assumptions of Theorem 6 be satisfied. If \( v_N \) is a normed eigenvector of \( \tilde{A}_N \) corresponding to \( s_N \) then for any \( \lambda \in (s_N-a_1, s_N+a_1) \)

\[
  u_N(\lambda) = v_N + \eta(\lambda) \quad \text{and} \quad |\eta(\lambda)|_N \leq k_0|\tilde{R}_N|_N
\]

with constant \( k_0 \) independent of \( N \).

**Proof.** To get this result we apply Theorem 3 for \( G = F_N \) and

\[
  q(\lambda) = \begin{bmatrix} v_N \\ (\lambda-s_N)^2 \end{bmatrix}
\]

We have

\[
  |\eta(\lambda)|_N = |u_N(\lambda)-v_N|_N \leq \left[ |u_N(\lambda)-v_N|_N^2 + |u_N(\lambda)-(\lambda-s_N)^2|^2 \right]^{1/2}
\]

\[
  = \left[ \begin{bmatrix} u_N(\lambda) \\ \mu_N(\lambda) \end{bmatrix} - q(\lambda) \right]_{N+1} \leq k |F_N(\lambda, \tilde{R}_N, \hat{\gamma}_N, (\lambda-s_N)^2)|_{N+1}
\]

\[
  = k \left[ \left[ (\lambda-s_N)^2-(\hat{\lambda}-\tilde{A}_N)^2 + \hat{R}_N \right] v_N \right]_{N+1}
\]

\[
  = k \left[ \left[ (\lambda-s_N)^2-(\hat{\lambda}-\tilde{A}_N)^2 + \hat{R}_N \right] v_N \right]_{N+1} + (v_N^T v_N - 1)^2 \]

where \( k \) is some constant. Finally, since \( [(\lambda-s_N)^2-(\hat{\lambda}-\tilde{A}_N)^2] v_N = 0 \) and \( v_N^T v_N = 1 \), we obtain \( |\eta(\lambda)|_N \leq k_0|\tilde{R}_N v_N|_N \).

Now, let us calculate the first and second derivative of \( \mu_N \). For first derivatives of \( \mu_N \) and \( u_N \) we can apply the equation \( F_N(\lambda, \hat{R}_N, u_N(\lambda), \mu_N(\lambda)) = 0 \) for fixed \( \hat{R}_N \). We get

\[
  \frac{\partial F_N}{\partial \lambda} + \left[ \begin{array}{c} \frac{\partial F_N}{\partial u} \\ \frac{\partial F_N}{\partial \mu_N} \end{array} \right] \frac{d}{d\lambda} \left[ \begin{array}{c} u_N \\ \mu_N \end{array} \right] = 0
\]

or

\[
  \left[ \begin{array}{ccc} -2(\lambda-\hat{A}_N) \mu_N & -2(\lambda-\hat{A}_N) \mu_N & \frac{d}{d\lambda} \left[ \begin{array}{c} u_N \\ \mu_N \end{array} \right] \end{array} \right] = 0.
\]

Theorem 2 implies that for \( \lambda \in (s_N-a_1, s_N+a_1) \) the matrix

\[
  M_N = \frac{\partial F_N}{\partial \lambda}(x_0, y_0) = \left[ \begin{array}{ccc} \mu_N & \frac{d}{d\lambda} u_N \end{array} \right]
\]

35 Banach Center 2 24
is invertible and (see Lemma 3)

\[ M_N^{-1} = \begin{bmatrix} Q_N & u_N/2 \\ u_N^T & 0 \end{bmatrix} \]

where

\[ Q_N = \sum_{j=1}^{N-1} (u_N u_N^T)/(\mu_N - \mu_j^\dagger). \]

Finally:

\[ \frac{d}{d\lambda} u_N = 2Q(\lambda - \hat{A}_N)u_N, \]

\[ \frac{d}{d\lambda} \mu_N = 2u_N^T(\lambda - \hat{A}_N)u_N, \]

\[ \frac{d^2}{d\lambda^2} \mu_N = 2 + 8u_N^T(\lambda - \hat{A}_N)Q_N(\lambda - \hat{A}_N)u_N. \]

**Theorem 7.** Let \( s \) be a single eigenvalue of \( A \) and \( \{s_N\} \) be a sequence of single eigenvalues of \( \tilde{A}_N \) converging to \( s \). Then there exists \( N_0 \) such that for \( N > N_0 \) the function \( \mu_N \) is convex in \( (s_N - a_1, s_N + a_1) \), and attains in this interval the unique minimum at \( s_N^* \), and

\[ s_N^* = s_N + O(|\tilde{R}_N|^2) \rightarrow s_N \quad \text{as} \quad N \rightarrow \infty. \]

**Proof.** From (6.9) and Lemma 4 it follows that

\[ \frac{d^2}{d\lambda^2} \mu_N = 2 + \epsilon_N(\lambda) \]

where

\[ \epsilon_N(\lambda) \overset{df}{=} 8\eta^T(s_N - \hat{A}_N)Q(s_N - \hat{A}_N)\eta. \]

Using now (6.6) and the result obtained above: \( |M_N^{-1}|_{n+1} \leq 2c_0 \) for \( |\lambda - s_N| \leq a_1 \) (see the proof of Theorem 6) we easily get that for any \( y \in \mathbb{R}^N, \ |y|_N = 1 \)

\[ |Q_N y|_N^2 \leq |Q_N y|_N^2 + |u_N^T y|_N^2 \leq 4c_0^2. \]

Thus

\[ |Q_N(\lambda)| \leq 8|\eta|_N^2|s_N - \hat{A}_N|_N^2|Q_N|_N^2 \leq k|\tilde{R}_N|^2 \rightarrow 0 \quad \text{as} \quad N \rightarrow 0 \]

which is implied by Lemma 4. So, we have

\[ \frac{d^2}{d\lambda^2} \mu_N > 0 \]

for \( |\lambda - s_N| < a_1 \) and \( N \) large enough, i.e. \( \frac{d}{d\lambda} \mu_N \) is strictly increasing in \( (s_N - a_1, s_N + a_1) \). Similarly, since \( u_N^T(s_N - \hat{A}_N) = \eta^T(s_N - \hat{A}_N) \), from (6.8) and Lemma 4 we get

\[ \frac{d}{d\lambda} \mu_N = 2(\lambda - s_N) + 2\eta^T(s_N - \hat{A}_N)\eta \]
and

\[ \sup_{|\lambda - s_N| < a_1} |\eta^T(s_N - \hat{A}_N)\eta| \leq k_1 |\hat{R}_N|_N^2. \]

For \( N \) large enough \( \frac{d}{d\lambda} u_N \) is strictly increasing function, negative for some \( \lambda' < s_N \) and positive for some \( \lambda'' > s_N \). Hence, there exists \( s_N^* \) unique in \((s_N - a_1, s_N + a_1)\) satisfying

\[ \frac{d}{d\lambda} \mu_N(s_N^*) = 0 \quad \text{and} \quad \frac{d^2}{d\lambda^2} \mu_N(s_N^*) > 0. \]

Finally, let us observe that

\[ |s_N^* - s_N| = |\eta^T(s_N - \hat{A}_N)\eta| \leq k_1 |\hat{R}_N|_N^2 \to 0 \quad \text{as} \quad N \to \infty, \]

which ends the proof of the theorem.

Now, let us observe that since

\[ \frac{d}{d\lambda} \mu_N(\lambda) = 2u_N^T(\lambda) (\lambda - \hat{A}_N) u_N(\lambda), \]

we have

\[ s_N^* = u_N^T(s_N^*) \hat{A}_N u_N(s_N^*). \]

So \( s_N^* \) is the fixpoint of the function

\[ \Psi(\lambda) \overset{\text{def}}{=} u_N^T(\lambda) \hat{A}_N u_N(\lambda) \]

on the interval \((s_N - a_1, s_N + a_1)\).

Let us consider the method of successive approximations for \( \Psi \)

(6.10) \[ \lambda_{k+1} = \Psi(\lambda_k), \quad k = 0, 1, 2, \ldots \]

**Theorem 8.** The process (6.10) of successive approximations for \( \Psi \) is equivalent to the process (4.2), (4.3) for \( \Phi(\lambda, x) = \| (\lambda - A) x \|^2 \) and \( \Lambda = \mathbb{R} \).

**Proof.** If \( \lambda_{k+1} \) is defined by (6.10) then

\[ \lambda_{k+1} = u_N^T(\lambda_k) \hat{A}_N u_N(\lambda_k). \]

Here \( u_N(\lambda_k) \) is a normalized eigenvector corresponding to the smallest eigenvalue of the matrix \((\lambda_k - \hat{A}_N)^2 + \hat{R}_N\), i.e. it minimizes the quadratic form

\[ \alpha^T [(\lambda_k - \hat{A}_N)^2 + \hat{R}_N] \alpha = \| (\lambda_k - A) x \|^2 = \Phi(\lambda_k, x) \]

for \( x = \sum_{j=1}^N \xi_j x_j \in X_N = p_N \hat{X}_N \). Let

\[ f(\lambda) \overset{\text{def}}{=} \Phi(\lambda, x) = u_N^T(\lambda_k) [(\lambda - \hat{A}_N)^2 + \hat{R}_N] u_N(\lambda_k) \]
for \( x_k = p_N g_N (\lambda_k) \). We have \( f'(x) = 2(\lambda - \Psi (\lambda_k)) \), \( f''(\lambda) = 2 \). It is clear that \( f(\lambda) \) attains its local minimum for \( \lambda_{k+1} = \Psi (\lambda_k) \); hence the two processes are equivalent.

**Theorem 9.** Let \( s_N \) be a Galerkin approximation of a single eigenvalue \( s \) of \( A \). Then there exist constants \( d < a_1/2 \) and \( N_0 \) such that for \( \lambda_0 \) satisfying \( |\lambda_0 - s_N| < d \), and for \( N > N_0 \) (fixed), the sequence \( \{ \lambda_k \} \) defined by (6.10) converges as \( k \to \infty \) to the point \( s_N^* \) in which \( \mu_N \) attains its local minimum.

**Proof.** If \( \lambda', \lambda'' \in (s_N - a_1, s_N + a_1) \) then by (6.7) we have

\[
    u_N (\lambda') - u_N (\lambda'') = 2 \int_{\lambda'}^{\lambda''} Q_N (\lambda - A_N) u_N (\lambda) d \lambda.
\]

Since \( Q_N (\lambda - A_N) u_N (\lambda) = Q_N (s_N - A_N) \eta \) (see Lemma 4) and \( |Q_N|_{\infty} \leq 2 c_0 \), thus applying once more Lemma 4 we get

\[
    |u_N (\lambda') - u_N (\lambda'')| \leq |\lambda' - \lambda''| 4 c_0 |s_N - A_N|_{\infty} \sup_{\lambda \in [\lambda', \lambda'']} |\eta(\lambda)|
\]

\[
    \leq k_2 |\lambda' - \lambda''| |R_N|_{\infty}.
\]

This implies that for any \( \lambda', \lambda'' \in (\lambda_0 - d, \lambda_0 + d) \subset (s_N - a_1, s_N + a_1) \) we have, by the definition of \( \Psi (\lambda) \),

\[
    |\Psi (\lambda') - \Psi (\lambda'')| \leq L |\lambda' - \lambda''|
\]

where \( L = k_3 |R_N|_{\infty} < 1 \) for \( N \) sufficiently large (cf. Lemma 2). Moreover, Lemma 4 implies that

\[
    \lambda_0 - \Psi (\lambda_0) = \lambda_0 - s_N + u_N^T (\lambda_0)(s_N - A_N) \eta = \lambda_0 - s_N + \eta^T (s_N - A_N) \eta.
\]

Thus

\[
    |\lambda_0 - \Psi (\lambda_0)| \leq \gamma \overset{\text{df}}{=} |\lambda_0 - s_N| + k_4 |R_N|_{\infty}
\]

for some constant \( k_4 \) independent of \( N \). It is easy to verify that if \( 0 < L < 1 \) and \( \gamma < (1 - L) a_1/2 \), then the process (6.10) converges to the unique fixpoint of the function \( \Psi \) in \( (\lambda_0 - a_1/2, \lambda_0 + a_1/2) \). Let us observe that the last inequality can be written as follows:

\[
    d + k_4 |R_N|_{\infty} < (1 - k_3 |R_N|_{\infty}) a_4/2.
\]

So, it holds for \( N \) large enough.

The residuum \( \Phi (\lambda, x) = \| (\lambda - A) x \| \) attains its local (or global) minimum for all \( x \in X_N, \| x \| = 1 \), and \( \lambda \) in a neighbourhood of the Galerkin point \( s_N \), at \( \lambda = s_N^* \). One could presume that \( \lambda = s_N^* \) is a better approximation of the eigenvalue \( s \) of \( A \) than its Galerkin approximation \( s_N \). However, as was observed by J. Descloux [private communication], this presumption is in general false.
Assume that the single eigenvalue $s$ of $A$ is the smallest element of the spectrum $\sigma(A)$ of $A$. We have

$$ s = \inf_{x \in H, \|x\| = 1} (Ax, x) \leq \inf_{x \in X_N, \|x\| = 1} (Ax, x) \leq \inf_{\alpha \in \mathbb{R}^N, \|\alpha\| = 1} \alpha^T \tilde{A}_N \alpha = s_N $$

$$ \leq u_N^T(s_N^*) \tilde{A}_N u_N(s_N^*) = \Psi(s_N^*) = s_N^*. $$

Hence $s \leq s_N \leq s_N^*$. Similar inequality can easily be derived when $s$ is the largest element of $\sigma(A)$: $s_N^* \leq s \leq s_N$. In both the cases, Galerkin point $s_N$ approximates better the point $s$ than $s_N^*$ does. Hence, starting in the process (6.10) with $\lambda_0 = s_N$ (which is allowed and reasonable) we may arrive, after iterations to the worse approximation of $s$ then the starting point was, even in extremally regular situation. Examples show that when $s$ is an intermediate eigenvalue of $A$ this phenomenon can appear or not.

At any rate $\mu_N(s_N^*)$ is always some computable error bound. For example in the case of the smallest eigenvalue $s$ of $A$ we have

$$ |s - s_N| \leq |s - s_N^*| \leq \mu_N(s_N^*). $$

On the other hand, the method may be applied to computation of approximated eigenvalues $\lambda_j$ starting with arbitrary point $\lambda_0$. It should be stressed that on each step the bound for distance of $\lambda_j$ from $\sigma(A)$ is computed.

Similar algorithm to (6.10), but always looking for the global minimum of $\mu_N$, is discussed in details in [4] for large matrices.

References


Presented to the Semester
Numerical Analysis and Mathematical Modelling
February 25 - May 29, 1987