A STUDY OF THE COMPUTATION AND CONVERGENCE BEHAVIOR OF EIGENVALUE BOUNDS FOR SELF–ADJOINT OPERATORS

by

Gyou-Bong Lee

Dissertation submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematics

APPROVED:

Christopher A. Beattie, Chairman

J. A. Burns

K. B. Hannsgen

J. U. Kim

W. E. Kohler

May, 1991

Blacksburg, Virginia

A STUDY OF THE COMPUTATION AND CONVERGENCE BEHAVIOR OF EIGENVALUE BOUNDS FOR SELF–ADJOINT OPERATORS

by

Gyou–Bong Lee

Committee Chairman: Christopher A. Beattie
Mathematics Department

(ABSTRACT)

The convergence rates for the method of Weinstein and a variant method of Aronszajn known as "truncation including the remainder" are derived in terms of the containment gaps between exact and approximating subspaces, using analytical techniques that arise in part in the convergence analysis of finite element methods for differential eigenvalue problems. An example of a one dimensional Schrödinger operator with a potential is presented which arises in quantum mechanics.

Examples using the recent eigenvector–free (EVF) method of Beattie and Goerisch are considered. Since the EVF method uses finite element trial functions as approximating vectors, it produces sparse and well–structured coefficient matrices. For these large–order sparse matrix eigenvalue problems, we adapt a spectral transformation Lanczos algorithm for finding a few wanted eigenvalues. For a few particular examples of vibration in beams and plates, convergence behavior is experimentally evaluated.

ACKNOWLEDGEMENTS

It is very pleasure for me to thank my advisor, Professor Christopher A. Beattie for his guidance and support. In the last three years I have been really indebted to him for having provided me a lot of knowledge on mathematics.

It is also my pleasure to thank my wife, Hyun-Sook, for her constant aid and patience.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

PRELIMINARIES

1.1 Introduction.

It is important to compute accurately the eigenvalues and eigenvectors of differential operators in order to analyze successfully various natural phenomena. We easily find many examples including the frequencies of bars, beams and plates, critical values of the Reynolds number in hydrodynamics, and bound state energy levels of atoms and molecules. The importance of such problems has encouraged mathematicians to study methods for finding the eigenvalues of differential operators. However the eigenvalues are not explicitly known in most cases, and thus several methods for their approximation have been presented and developed over many years. Since there is no method that provides precise error estimation in approximation, the only reliable way may be to use two ancillary methods that give upper and lower bounds, respectively, to the eigenvalues considered. In their analysis, we meet equations of the style $Au = \lambda u$ in Ω, where A is considered as a semi-bounded self-adjoint operator on a Hilbert space, having eigenvalues of finite multiplicity below the lowest limit point(if any) of the spectrum.

Historically, in the last quarter of the 19th century, Lord Rayleigh had initiated a development in the approximation of eigenvalues, based on the so-called *Rayleigh Principle* which states that if one limits the freedom of vibration of a mechanical system, the frequencies of the obtained system can not be lower than those of the original system [55]. In 1909 W. Ritz illustrated that by choosing a constrained system with a finite but sufficiently large number of degrees of freedom, arbitrarily close approximations to the lower eigenvalues of the original continuous system could

be obtained. This observation leads to the oldest method for obtaining numerical upper bounds called the *Rayleigh-Ritz method*[44]. A much more difficult problem is that of finding accurate lower bounds, for which we will consider the *method of intermediate eigenvalue problems*, which gives a sequence of improvable lower bounds.

In 1937 A. Weinstein developed a method for finding lower bounds for the eigenvalues of certain differential operators [72]. This method was extended and simplified by N. Aronszajn in 1948 by use of the properties of compact self-adjoint operators in Hilbert space [2]. However, it initially proved to be very difficult to implement Aronszajn's method numerically. In 1959 N. Bazley achieved the first major innovation in the implementation of Aronszajn's method with the development of the method of special choice [7]. In the same year H. Weinberger published a method for improvable lower bounds and a method for simplifying the calculations involved in Weinstein's method [67]. Subsequently Bazley together with D. Fox developed a number of means for implementing Aronszajn's method for differential problems [8–14]. Very recently, Beattie and Goerisch have developed a method for finding lower bounds without having knowledge of eigenvectors of a base problem which otherwise are necessary in most intermediate eigenvalue problems [17]. Using both the Rayleigh-Ritz method and the intermediate problem method, one is able to find an interval, whose length can be made as small as desired, guaranteed to contain a selected eigenvalue

This dissertation concentrates on both the usual method of intermediate problems as well as Beattie and Goerisch's *eigenvector free*(EVF) *method*, which may be found to be of use in classical and quantum mechanical eigenvalue problems that involve complex domain geometry or realistic potentials. Since we limit our attention to problems which can be formulated in terms of self-adjoint operators in Hilbert space, our approach will be operator theoretic in nature.

Section 2 presents some background for the variational approaches and Section 3

gives a brief explanation of the intermediate problem method. Section 4 contains some remarks. Chapter 2 introduces new results about convergence rates for a sequence of semi–bounded operators, which are applied to intermediate problems together with the method of *truncation including remainder* and also presents an example which comes from quantum mechanics. In Chapter 3 we deal with the Beattie and Goerisch EVF method with numerical examples that arise from the vibration of beams, and also analyze how to take advantage of the sparsity of a large–order matrix which comes from the EVF method as using finite element trial functions with an example of the vibration of a rectangular clamped plate.

1.2 Variational Principles of Eigenvalue Approximation.

In this section we outline the development of the variational principles for eigenvalues. Let $\mathcal{H}$ be a separable complex Hilbert space with norm $\|u\|$ and inner product $\langle u, v\rangle$. Let A be a self adjoint operator with domain $Dom(A)$ dense in $\mathcal{H}$. We suppose that A is bounded below and that the lower part of its spectrum consists of a finite or infinite number of isolated eigenvalues

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_\infty$$

each having finite multiplicity. Here λ_∞ denotes the lowest limit point (if any) of the spectrum of A. For convenience we denote such a class of operators by $\mathcal{S}$. If A has compact resolvent, then we set $\lambda_\infty = \infty$ (We say that A has compact resolvent if $(A - z)^{-1}$ is compact for any $z \in \rho(A)$.).

We note that many operators that arise in the eigenvalue problems of mathematical physics and engineering are in $\mathcal{S}$ for some choice of Hilbert space $\mathcal{H}$. The lower eigenvalues of such operators have classical characterizations for which the oldest one is originally due to Lord Rayleigh [55] and Weber [63].

THEOREM 1.2.1. *(Rayleigh-Weber) The eigenvalues of $A \in \mathcal{S}$ are given by the equations*

$$\lambda_1 = \min_{u \in Dom(A)} \frac{\langle Au, u\rangle}{\langle u, u\rangle} \quad \text{and} \quad \lambda_n = \min_{\substack{u \in Dom(A) \\ \langle u, u_i\rangle = 0 \\ i=1,\ldots,n-1}} \frac{\langle Au, u\rangle}{\langle u, u\rangle},$$

where $u_1, u_2, \ldots, u_{n-1}$ denote eigenvectors corresponding to $\lambda_1, \lambda_2, \ldots, \lambda_{n-1}$.

We may find a modern proof of Theorem 1.2.1 in [73] and an extension of this result to semi-bounded, closed quadratic forms in [68] that is frequently more useful. It directly follows from Theorem 1.2.1 that for a unit vector $v_0 \in Dom(A)$, the value $\langle Av_0, v_0\rangle$ provides an upper bound to the lowest eigenvalue λ_1. For more improved bounds to λ_1, we can take a sequence of orthonormal vectors $\{v_i\}_{i=1}^{\infty} \subset Dom(A)$

and compute the lowest eigenvalue of the matrices $[\langle Av_i, v_j\rangle]_{i,j=1}^n$ for $n = 1, 2, \ldots$ successively.

While the classical characterization is very important as an analytical device, it has the disadvantage that it may not be used to determine higher eigenvalues without employing explicitly all preceding eigenvectors. Nearly a quarter of a century after the classical principle was given, the situation was considerably improved by Poincaré [50], who developed the Rayleigh-Weber result into a set of inequalities relating the eigenvalues of A to the eigenvalues of a finite-dimensional restriction of A. For this purpose, we let $\mathcal{P}_n$ be a n-dimensional subspace of $Dom(A)$ with P_n representing the related orthogonal projection onto $\mathcal{P}_n$. Then P_nAP_n is self-adjoint as a transformation from the finite-dimensional space $\mathcal{P}_n$ into itself. If we consider P_nAP_n as an operator on $\mathcal{H}$, its spectrum consists of the eigenvalues $\Lambda_1, \Lambda_2, \ldots, \Lambda_n$ as well as the eigenvalue $\Lambda = 0$ with infinite multiplicity.

THEOREM 1.2.2. *(Poincaré) For any n-dimensional space $\mathcal{P}_n$, the eigenvalues $\{\Lambda_i\}_{i=1}^n$ of P_nAP_n satisfy the inequalities*

$$\lambda_1 \leq \Lambda_1, \lambda_2 \leq \Lambda_2, \ldots, \lambda_n \leq \Lambda_n.$$

Fischer [34] applied the Poincaré's inequalities for finite-dimensional spaces while Pólya [51] applied them to operators in infinite-dimensional spaces. The inequalities were formulated as a so-called *minimum-maximum principle.*

THEOREM 1.2.3. *(Fischer-Pólya : Minimum-Maximum Principle) The eigenvalues of $A \in \mathcal{S}$ may be characterized as*

$$\lambda_n = \min_{\substack{\mathcal{P}_n \subset Dom(A) \\ dim\mathcal{P}_n = n}} \max_{u \in \mathcal{P}_n} \frac{\langle Au, u\rangle}{\langle u, u\rangle}.$$

As a related application of Theorem 1.2.2 or Theorem 1.2.3, we have an outstanding method for obtaining upper bounds for eigenvalues known as the *Rayleigh-Ritz*

method. This method provides an efficient means of computing nonincreasing upper bounds for an arbitrary but finite number of eigenvalues of any operator in class $\mathcal{S}$. The main idea of this method is to restrict a given operator to a finite-dimensional subspace of its domain, yielding a matrix problem for which the eigenvalues are numerically computable. It follows then from Theorem 1.2.2 that the computed eigenvalues are all upper bounds to the corresponding eigenvalues of the given operator [73].

THEOREM 1.2.4. *(Rayleigh-Ritz Method) Let* $\mathcal{P}_n = span\{p_1, p_2, \ldots, p_n\}$. *Then the eigenvalues* $\{\Lambda_i\}_{i=1}^n$ *of* $P_n A P_n|_{\mathcal{P}_n}$ *are the solutions to the general matrix eigenvalue problem in* $\mathbb{C}^n$

$$[\langle Ap_i, p_j\rangle]\mathbf{x} = \Lambda[\langle p_i, p_j\rangle]\mathbf{x}$$

for all $i, j = 1, \ldots, n$.

We have from this result an easy approach for finding upper bounds to the lower eigenvalues of A. With only these upper bounds is it difficult to realize how close they are to the eigenvalues of A. Thus we need rigorous lower bounds subsidiary to the upper bounds. For this we present an alternate characterization of the lower eigenvalues of A that originally comes from an inequality of Weyl [74], which later was applied by Courant [31].

THEOREM 1.2.5. *(Weyl's Inequality) For any choice of vectors* $p_1, p_2, \ldots, p_{n-1} \in \mathcal{H}$, *we have the inequality*

$$\min_{\substack{u \in Dom(A) \\ \langle u, p_i\rangle = 0 \\ i=1,2,\ldots,n-1}} \frac{\langle Au, u\rangle}{\langle u, u\rangle} \leq \lambda_n.$$

Let us note in passing that the value in the left of Theorem 1.2.5 is the lowest eigenvalue of a problem $Au - PAu = \lambda u$ restricted to $Pu = 0$, where P is the orthogonal projection onto the space spanned by $\{p_1, p_2, \ldots, p_{n-1}\}$, say $\mathcal{P}$ (cf. [73]).

THEOREM 1.2.6. *(Courant-Weyl : Maximum-Minimum Principle) The eigenvalues of* $A \in \mathcal{S}$ *are given by the equation*

$$\lambda_n = \max_{\substack{p_1,p_2,\ldots,p_{n-1} \\ \in \mathcal{H}}} \min_{\substack{u \in Dom(A) \\ \langle u,p_i \rangle = 0 \\ i=1,2,\ldots,n-1}} \frac{\langle Au, u \rangle}{\langle u, u \rangle}.$$

for $n = 2, 3, \ldots$.

If $\mathcal{H}$ is a finite–dimensional space, the maximum–minimum principle and the minimum–maximum principle are in a sense equivalent, which is essentially due to the fact that in this case the orthogonal complement of a finite-dimensional space is itself finite-dimensional. But Theorem 1.2.6 (Max-min principle) is very different from Theorem 1.2.3 (Min-max principle) in spite of similarities in statement. The reason is mainly the usual infinite-dimensionality of the orthogonal complement to the space $\mathcal{P}$. The existence of the minimum in Theorem 1.2.5 has been proved recently for $A \in \mathcal{S}$ [73]. The computational difficulties in obtaining rigorous lower bounds come from the infinite-dimensionality of $\mathcal{P}^{\perp}$. For example, let us take a finite-dimensional subspace $\mathcal{R}_m \subset Dom(A)$ with $dim\mathcal{R}_m = m \leq n-1$ (cf. [15]). Then

$$\lambda_n \geq \max_{\substack{dim\mathcal{P}=n-1 \\ \mathcal{P} \supset \mathcal{R}_m}} \min_{u \in \mathcal{P}^{\perp}} \frac{\langle Au, u \rangle}{\langle u, u \rangle} = \max_{dim\mathcal{P}=n-m-1} \min_{u \in \mathcal{P}^{\perp} \cap \mathcal{R}_m^{\perp}} \frac{\langle Au, u \rangle}{\langle u, u \rangle}.$$

The right-hand side may be identified with the $(n-m)$-th eigenvalue of

$$Au - R_m Au = \lambda u \quad \text{with} \quad R_m u = 0$$

where $R_m : \mathcal{H} \longrightarrow \mathcal{R}_m$ is the orthogonal projection. However $(A - R_m A)|_{\mathcal{R}_m^{\perp}}$ is generally not a finite rank operator and does not usually have clear finite dimensional reducing spaces. Hence this approach does not seem to result in a computationally practical strategy. We now present an important application of the minimum-maximum principle.

DEFINITION. *Let $\mathcal{P}$ be a closed subspace of $\mathcal{H}$ and let $P : \mathcal{H} \longrightarrow \mathcal{P}$ be the orthogonal projection onto $\mathcal{P}$. Let $Q = I - P$. We say that $A - PA$ on $\mathcal{P}^{\perp}$ is the part of A in $\mathcal{P}^{\perp}$, and that QAQ is the projection of A onto $\mathcal{P}^{\perp}$.*

THEOREM 1.2.7. *(Rayleigh's Theorem for r Constraints [44]) Let $\mathcal{P}$ be an r dimensional subspace of $\mathcal{H}$ and let P be the orthogonal projection onto $\mathcal{P}$. Let $\{\lambda_i'\}$ be the eigenvalues of the part of A in $\mathcal{P}^{\perp}$ arranged in increasing order according to multiplicity. Then for all $i = 1, 2, \ldots,$*

$$\lambda_i \leq \lambda_i' \leq \lambda_{i+r}.$$

If A is self-adjoint, then the part of A is also self-adjoint [73]. The following theorems come from Theorems 1.2.3 or 1.2.7 and have important roles in the analysis of intermediate problems which follows in the next sections.

THEOREM 1.2.8. *(First Monotonicity Principle) Let A be an operator in $\mathcal{S}$ and A' be a part of A in the subspace Q of $\mathcal{H}$. Then the eigenvalues λ_i' and λ_i of A' and A, respectively, satisfy the inequalities*

$$\lambda_i \leq \lambda_i'$$

for all $i = 1, 2, \ldots,$

DEFINITION. *For symmetric operators S and T we define $S \leq T$ if $Dom(T) \subset Dom(S)$ and $\langle Su, u\rangle \leq \langle Tu, u\rangle$, for all $u \in Dom(T)$.*

THEOREM 1.2.9. *(Second Monotonicity Principle) Let A' and A be operators of class $\mathcal{S}$ satisfying $A \leq A'$. Then the eigenvalues λ_i' and λ_i of A' and A, respectively, satisfy the inequalities*

$$\lambda_i \leq \lambda_i'$$

for all $i = 1, 2, \ldots.$

1.3 Construction of Intermediate Eigenvalue Problems.

In this section we review the methods presented by Weinstein in 1935 with his work on the estimation of buckling loads and vibration frequencies for plates [68–71] and by Aronszajn in 1951 [2] who proposed a similar estimation procedure for obtaining lower bounds that was applicable to a much wider class of eigenvalue problems than Weinstein's procedure. As a variant of Aronszajn's method, we present the method of truncation including the remainder which was first analyzed by Greenlee [43] and developed further by Greenlee and Beattie [19,20]. Finally we also present a variant of the Aronszajn method which was initiated by Bazley and Fox [10].

The scheme of intermediate problems is the following: Given an eigenvalue problem for an operator A of type $\mathcal{S}$, the first step is to find a base operator A_0 in $\mathcal{S}$ whose eigenvalues are not greater than the corresponding eigenvalues of the given operator. The next step is to construct a sequence of eigenvalue problems, called *intermediate eigenvalue problems*, in such a way that they yield computable eigenvalues which are not smaller than those of the preceding problem in the sequence, not greater than those of the succeeding problem, and never greater than the eigenvalues of the original problem. The base problem

$$A_0 u = \lambda u$$

is picked so that A_0 is in $\mathcal{S}$ and $A_0 \leq A$. We assume that the isolated eigenvalues of the base problem

$$\lambda_1^0 \leq \lambda_2^0 \leq \cdots \leq \lambda_\infty^0$$

are computable to arbitrary precision. The closure of the quadratic form $\langle A_0 u, u \rangle$ is denoted by $a_0(u)$. Then $a_0(u) \leq a(u)$ for all $u \in Dom(a) \subset Dom(a_0)$. The second monotonicity principle implies that $\lambda_\infty^0 \leq \lambda_\infty$, and that for each i such that $\lambda_i < \lambda_\infty^0$, λ_i^0 exists and $\lambda_i^0 \leq \lambda_i$. Without loss of generality we may assume that the difference

between a_0 and a is strictly positive, that is,

$$b(u) = a(u) - a_0(u) \geq \alpha \|u\|^2,$$

for some $\alpha > 0$ and all $u \in Dom(b) = Dom(a) \subset Dom(a_0)$.

We should note that most suitable base problems having computable eigenvalues and eigenvectors produce very poor and fixed bounds. The intermediate problem methods provide an approach for adding back incrementally what was lost in passing from A to A_0 in a way that permits explicit resolution of the intermediate eigenvalue problems to improve lower bounds to the eigenvalues of A.

1.3.1 On the method of Weinstein. We suppose that the quadratic forms a_0 and a are closed, densely defined and coercive in $\mathcal{H}$ such that $a_0(u) \leq a(u)$ for all $u \in Dom(a) \subset Dom(a_0)$. Then the corresponding self-adjoint operator A_0 is positive definite and the Hilbert space $\mathcal{H}_{a_0}$ which is the completion of $Dom(A_0)$ with respect to norm generated by $a_0(u, v)$ is continuously embedded in $\mathcal{H}$. The similarly defined Hilbert space $\mathcal{H}_a$ may be considered as a closed subspace of $\mathcal{H}_{a_0}$(cf. [27]).

We assume that $P : \mathcal{H}_{a_0} \longrightarrow \mathcal{H}_{a_0} \ominus \mathcal{H}_a$ is the a_0–orthogonal projection onto $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$ and that $A = A_0 - PA_0$ on $Dom(A) \subset Dom(A_0)$. We note [45,73] that the spectral resolution of the projection of A_0 to $\mathcal{H}_a$, QA_0Q, is obtained from the spectral theorem for the part of A_0 in $\mathcal{H}_a$ and adjoining the eigenvalue zero on $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$. Here $Q = I - P : \mathcal{H}_{a_0} \longrightarrow \mathcal{H}_a$ is a projection. Thus the positive eigenvalues of QA_0Q are just those of $QA_0|_{\mathcal{H}_a}$. Hence A may be considered as QA_0Q, and the first monotonicity theorem implies that the base problem, $A_0u = \lambda u$, yields lower bounds to the eigenvalues of A. Let us take a sequence of finite dimensional subspaces, $\{\mathcal{P}_i\}$, in the orthogonal complement of $\mathcal{H}_a$ in $\mathcal{H}_{a_0}$,

$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_k \subset \mathcal{P}_{k+1} \subset \cdots \subset \mathcal{H}_{a_0} \ominus \mathcal{H}_a$$

and let $P_k : \mathcal{H}_{a_0} \longrightarrow \mathcal{P}_k$ be the orthogonal projection. We now construct the intermediate operators as

$$A_k = Q_k A_0 Q_k$$

where $Q_k = I - P_k$. It follows [73] that if A_0 is compact, $Q_k A_0 Q_k$ is compact, and if $A \in \mathcal{S}$, so is $Q_k A_0 Q_k$. The minimum-maximum principle provides that the eigenvalues of A_k satisfy the inequality,

$$\lambda_i^0 \leq \cdots \leq \lambda_i^{(k)} \leq \lambda_i^{(k+1)} \leq \cdots \leq \lambda_i$$

for each i. That is, the intermediate operators provide improvable lower bounds to the eigenvalues of A with increasing k .

The intermediate eigenvalue problem $A_k u = \lambda u$ on $\mathcal{P}_k^{\perp}$ yields the so–called *Weinstein matrix*,

$$W_n(\lambda) = (\langle R_\lambda^0 p_i, p_j \rangle),$$

where $R_\lambda^0 = (A_0 - \lambda)^{-1}$ and $\mathcal{P}_k = span\{p_1, p_2, \ldots, p_k\}$, because $P_k A_0 u \in \mathcal{P}_k$ and Q_k is the identity on $\mathcal{P}_k^{\perp}$. The zeros and poles of the determinant of the matrix provide the eigenvalues of A_k. But a direct computation is obstructed by the difficulty in obtaining a functional expression for $R_\lambda^0 p_i$ in terms of λ. Using a truncation of A_0 as a base operator, we may overcome the difficulty [73].

1.3.2 On the method of Aronszajn. The method is designed for a different problem setting from the previous case. We recall that we have quadratic forms a, a_0 and b such that

$$b(u) = a(u) - a_0(u) \geq \alpha \|u\|^2$$

for some $\alpha > 0$ and all $u \in Dom(b) = Dom(a) \subset Dom(a_0)$. Suppose that $b(u)$ is closable in $\mathcal{H}$ and denote its closure as b again. There are densely defined self-adjoint operators, A, A_0 and B, associated with $a(u), a_0(u)$ and $b(u)$, respectively, such that

$b(u,v) = \langle u, Bv \rangle$ for all $u \in Dom(b)$ and $v \in Dom(B)$. We assume that the operator A can be decomposed as the sum of two operators

$$A = A_0 + B,$$

where A_0 is a resolvable base operator. The theoretical basis for this scheme lies in the second monotonicity principle while that of the Weinstein lies in the first monotonicity principle. The main notion behind Aronszajn's method is to approximate B with finite rank perturbations of the resolvable operator A_0.

For this purpose we introduce a new Hilbert space $\mathcal{H}_b$ which is the completion of $Dom(B)$ in the norm generated by the new inner product $\langle u, Bv \rangle$. Let a sequence of finite dimensional subspaces

$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_k \subset \mathcal{P}_{k+1} \subset \cdots \subset Dom(B)$$

be given, and let $P_k : \mathcal{H}_b \longrightarrow \mathcal{P}_k$ be the projection that is orthogonal with respect to the inner product $\langle u, Bv \rangle$. That is, for any u,

$$P_k u = \sum_{i,j=1}^{k} \langle u, Bp_i \rangle \, b_{ij} \, p_j$$

where the matrix (b_{ij}) is the inverse to the Gram matrix $(\langle p_i, Bp_j \rangle)$ of order k.

We now form the intermediate quadratic forms as

$$a_k(u) = a_0(u) + b(P_k u)$$

for all $u \in Dom(a_k) = Dom(a_0)$ with the corresponding self-adjoint operators

$$A_k = A_0 + BP_k,$$

where k is called order of the intermediate operator. Since the operator BP_k is symmetric and bounded, it follows from [45,64] that the operators A_k are self-adjoint

and have the same domain as A_0. Moreover, since BP_k is a compact operator, each of the operators A_k has exactly the same limit points in its spectrum as does A_0 [58]. It follows from the boundedness of BP_k that the operator BP_k may be considered as an operator on the space $\mathcal{H}$ and thus we have

$$a_0(u) \leq \cdots \leq a_k(u) \leq a_{k+1}(u) \leq \cdots \leq a(u)$$

for all $u \in Dom(a) \subset Dom(a_k) = Dom(a_0)$. The second monotonicity theorem implies that the eigenvalues of A_k satisfy the inequality,

$$\lambda_i^0 \leq \cdots \leq \lambda_i^{(k)} \leq \lambda_i^{(k+1)} \leq \cdots \leq \lambda_i$$

for all i such that $\lambda_i \leq \lambda_\infty^0$. The eigenvalues of A_k thus give lower bounds to the corresponding eigenvalues of A that improve with increasing k.

We now turn to the problem of determining the eigenvalues and eigenvectors of the intermediate operators A_k. We will sketch the procedure; one may refer to [2,73] for details. First, let $\mathcal{P}_k = span_{1 \leq i \leq k}\{p_i\}$ and consider the eigenvalue problem,

$$A_k u = \lambda u.$$

Then for λ that is not in the spectrum of A_0, we have

$$u = -\sum_{j=1}^{k} \alpha_j R_\lambda^0 B p_j$$

where the coefficients α_j's satisfy the matrix equation

$$\sum_{j=1}^{k} \alpha_j \left\langle p_j + R_\lambda^0 B p_j, B p_l \right\rangle = 0$$

for all $l = 1, 2, \ldots, k$. All of the α_j's cannot vanish if u is to be nontrivial. Thus λ must satisfy the determinantal equation

$$det(\left\langle p_i + R_\lambda^0 B p_i, B p_j \right\rangle) = 0.$$

For the case that λ is in the spectrum of A_0, one may refer to [73]. We call the matrix $(\langle p_i + R_\lambda^0 B p_i, B p_j \rangle)_{i,j=1}^k$ the *Weinstein-Aronszajn* (W-A) matrix of order k and denote it again by $W_k(\lambda)$. The matrix $W_k(\lambda)$ has a meromorphic character with singularities at the isolated eigenvalues of A_0, but direct computation is obstructed by the problem of not having a functional expression for $R_\lambda^0 B p_i$ in terms of λ when the choice of vectors p_i is left general.

By a special choice of the vectors p_i, Bazley first recognized [7] that the meromorphic function may be reduced to a rational function which can be written in an explicit form. In other words if we choose p_i so that Bp_i is a unit eigenvector of A_0, say u_i^0, the W-A matrix $W_k(\lambda)$ can be represented by

$$((\lambda_i^0 - \lambda)\delta_{ij} + \langle B^{-1} u_i^0, u_j^0 \rangle).$$

Hence if the inverse of the operator B is explicitly known, the eigenvalue problem for A_k is easily resolvable. Bazley and Fox also extended this to the case where each p_i could be chosen so that Bp_i was a known linear combination of eigenvectors of A_0 [8].

The method of special choice is not always possible. Thus it is important to consider a general choice of the p_i. But in this case we may meet the difficulty described previously. That is, the resolvent operator R_λ^0 for the base operator is rarely known in closed form. In many cases it can be expressed by infinite sums of integrals in general,

$$R_\lambda^0 u = \sum \frac{\langle u, u_i^0 \rangle \, u_i^0}{\lambda_i^0 - \lambda} + \int_{\lambda_\infty^0}^{\infty} \frac{dE_\mu^0 u}{\mu - \lambda}$$

where E_μ^0 is the spectral projection of A_0 for μ. To overcome this difficulty, Bazley and Fox proposed the use of the truncation of the base operator in [8] which Weinberger had initiated in [67] to his way, defined in terms of a spectral projection as

$$A_0^{(n)} = A_0 E_{\lambda_n^0}^0 + \lambda_{n+1}^0 (I - E_{\lambda_n^0}^0)$$

which is called a *truncation of A_0 of order n*. Clearly, it satisfies

$$A_0^{(n)} \leq A_0^{(n+1)} \leq A_0$$

for $n = 1, 2, \ldots$. The new intermediate operators $A_{n,k}$ having $A_0^{(n)}$ as a base operator are defined by

$$A_{n,k} = A_0^{(n)} + BP_k$$

for $n, k = 1, 2, \ldots$. These are bounded, symmetric and monotonically increasing in n and k. That is,

$$A_{n,k} \leq \begin{bmatrix} A_{n+1,k} \\ A_{n,k+1} \end{bmatrix} \leq A$$

for $n, k = 1, 2, \ldots$. Thus they provide lower bounds to eigenvalues of A which improve with increasing n and k.

The W-A matrix for this method may be represented by

$$(\left\langle p_i + R_\lambda^{(n)} B p_i, B p_j \right\rangle)_{i,j=1}^k$$

in which the resolvent operator $R_\lambda^{(n)}$ of $A_0^{(n)}$ is given by the closed expression,

$$R_\lambda^{(n)} v = \sum_{i=1}^{n} \frac{\langle v, u_i^0 \rangle \, u_i^0}{\lambda_i^0 - \lambda} + \frac{1}{\lambda_{n+1}^0 - \lambda}(v - \sum_{i=1}^{n} \langle v, u_i^0 \rangle \, u_i^0).$$

Thus the W-A determinant, *det* $W_k(\lambda)$, is a rational form instead of a (generally) transcendental function, which reduces the difficulty of determining roots. But we pay a price in that we are using a cruder base operator $A_0^{(n)}$ than A_0.

As another method to overcome the difficulty of a special choice, Bazley and Fox used another projection, called *the method of second projection*. We will sketch this method (see [9] for further detail).

For any constant δ, the operator A_k may be rewritten by

$$A_k = (A_0 - \delta^2) + (BP_k + \delta^2).$$

Let $B_k = BP_k + \delta^2$, for each δ and k. The operator B_k produces a new inner product $\langle u, B_k v\rangle$ on the Hilbert space $\mathcal{H}$. Let a sequence of finite dimensional subspaces,

$$\hat{\mathcal{P}}_1 \subset \hat{\mathcal{P}}_2 \subset \cdots \subset \hat{\mathcal{P}}_n \subset \hat{\mathcal{P}}_{n+1} \subset \cdots \subset \mathcal{H}.$$

be given, and let $\hat{P}_n : \mathcal{H} \longrightarrow \hat{\mathcal{P}}_n$ be the projection that is orthogonal with respect to this inner product $\langle u, B_k v\rangle$. We form the intermediate operators as

$$A_{k,n} = (A_0 - \delta^2) + B_k \hat{P}_n.$$

The operators $B_k \hat{P}_n$ are then bounded, symmetric and positive semidefinite such that

$$B_k \hat{P}_n \leq B_k \hat{P}_{n+1} \leq B_k \hat{P}_n.$$

It follows that the inequality holds

$$A_0 - \delta^2 \leq A_{k,n} \leq \begin{bmatrix} A_{k,n+1} \\ A_{k+1,n} \end{bmatrix} \leq \begin{bmatrix} A_k \\ A_{k+1} \end{bmatrix} \leq A.$$

Hence they provide lower bounds to eigenvalues of A which improve increasingly in n and k.

The W-A matrix for this method may be expressed by

$$(\langle \hat{p}_i + R^0_{\lambda+\delta^2} B_k \hat{p}_i, B_k \hat{p}_j \rangle)^n_{i,j=1}$$

in which the operator B_k and the inverse have the explicit forms. In fact the inverse of B_k is expressed as

$$\begin{aligned} B_k^{-1} u &= \frac{1}{\delta^2}[I - B(\delta^2 + P_k B)^{-1} P_k] u \\ &= \frac{1}{\delta^2}[u - \sum_{i,j=1}^{k} \langle u, Bp_i \rangle \, c_{ij} Bp_j] \end{aligned}$$

where (c_{ij}) is the matrix inverse to $(\delta^2 \langle p_i, Bp_j \rangle + \langle Bp_i, Bp_j \rangle)$. Therefore the operators $A_{k,n}$ have been constructed so that a special choice of the $\hat{p}_i$ is always possible. That is, $\hat{p}_i = B_k^{-1} u_i^0$. Thus we have the W–A matrix,

$$W_{k,n}(\lambda) = ((\lambda_i^0 - \delta^2 - \lambda)\delta_{ij}) + c_{ij})$$

where (c_{ij}) is the inverse to $(\langle u_i^0, B_k u_j^0 \rangle)$.

It follows from [9] that the operator $A_{k,n}$ is monotonically increasing in δ^2 on the space spanned by $\{u_1^0, \ldots, u_n^0\}$, but is decreasing on the orthogonal complement. The eigenvalues $\lambda_i^{(k,n)}$ considered as functions of δ converges to the eigenvalue λ_i^0 as δ goes to zero. For each k and n, the best value of δ^2 for the estimation of λ_i, $i \leq n$, is that

$$\delta^2 = \lambda_{n+1}^0 - \lambda_i^{(k,n)}.$$

Börsch-Supan first compared the methods of truncation and second projection. According to [25], if we take $\delta^2 = \lambda_{n+1}^0 - \lambda$ and $\hat{p}_i = B_k^{-1} u_i^0$, for $i = 1, 2, \ldots, n$, then they have the same eigenvalues with slightly different eigenvectors. In the case of the second projection, the best value of δ^2 for λ_i is $\delta^2 = \lambda_{n+1}^0 - \lambda_i^{(k,n)}$ and generally the method of truncation will produce better lower bounds.

1.3.3 On the method of truncation including the remainder. Bazley and Fox first introduced this method in [13], Greenlee analyzed it in [43] and later Beattie and Greenlee have developed further this method in [19,20]. For the following we adopt notations directly from [43,19,20]. Let us take a real number γ satisfying $\lambda_1(A_0) < \gamma \leq \lambda_\infty(A_0)$, with the restriction that $\gamma < \lambda_\infty(A_0)$ if A_0 has an infinity of eigenvalues below $\lambda_\infty(A_0)$. Define the truncation of A_0 at γ by

$$A_0^{(\gamma)} = A_0 E_{\gamma-}[A_0] + \gamma(I - E_{\gamma-}[A_0])$$

where $E_\lambda[A_0]$ is the right continuous resolution of the identity for A_0. We note that if $\gamma = \lambda_{n+1}^0$, then the $A_0^{(\gamma)}$ is the same as the previously defined $A_0^{(n)}$. But we use the notation $A_0^{(\gamma)}$ thereafter in order to follow their notations. We note that $A_0^{(\gamma)}$ has the same action as A_0 on the finite dimensional subspace, $\mathcal{U}_0^{(\gamma)} = Ran(E_{\gamma-}[A_0])$, and acts as a scalar multiplication by γ on $(\mathcal{U}_0^\gamma)^\perp$. The corresponding quadratic form $a_0^{(\gamma)}$ may be used to define a quadratic form

$$\tilde{a}(u) = a(u) - a_0^{(\gamma)}(u) \geq b(u) \geq \alpha \|u\|^2.$$

One may observe that $Dom(\tilde{a}) = Dom(a)$ where $\tilde{a}$ is a closed quadratic form and the corresponding self adjoint operator is given by

$$\tilde{A} = A - A_0^{(\gamma)}$$

with $Dom(\tilde{A}) = Dom(A)$. The main notion behind this method is to approximate $\tilde{A}$ with a finite rank operator which consequently produces intermediate operators that are finite rank perturbations of the resolvable operator $A_0^{(\gamma)}$.

For this purpose, we introduce a new Hilbert space $\mathcal{H}_{\tilde{a}}$ which is the completion of $Dom(\tilde{A})$ in the norm generated by the new inner product $\langle u, \tilde{A}u \rangle$. Let a sequence of finite dimensional subspaces

$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_k \subset \mathcal{P}_{k+1} \subset \cdots \subset Dom(\tilde{A})$$

be given, and let $P_k : \mathcal{H}_{\tilde{a}} \longrightarrow \mathcal{P}_k$ be the projection that is orthogonal with respect to the inner product $\langle u, \tilde{A}v \rangle$. For each k, we now define the intermediate form,

$$a_k(u) = a_0^{(\gamma)}(u) + \tilde{a}(P_k u)$$

for $u \in Dom(a_k) = \mathcal{H}$, with the corresponding self adjoint operator

$$A_k = A_0^{(\gamma)} + \tilde{A}P_k.$$

By construction, we have

$$a_0(u) = a_0^{(\gamma)}(u) \leq a_k(u) \leq a_{k+1}(u) \leq a(u)$$

for all k and $u \in Dom(a)$ where the second monotonicity principle implies that

$$\lambda_i(A_0) = \lambda_i(A_0^{(\gamma)}) \leq \lambda_i(A_k) \leq \lambda_i(A_{k+1}) \leq \lambda_i(A),$$

for all k and i such that $\lambda_i(A) < \gamma$.

1.3.4 On the method of Bazley-Fox. Let $a(u)$ and $a_0(u)$ be the quadratic forms which are the closures of $\langle u, Au\rangle$ and $\langle u, A_0u\rangle$, respectively, such that

$$a_0(u) \leq a(u)$$

for all $u \in Dom(a) \subset Dom(a_0)$. We assume that the quadratic form $a(u)$ is decomposed as

$$a(u) = a_0(u) + \|Tu\|_*^2$$

where T is a closed operator on $\mathcal{H}$ to another Hilbert space $\mathcal{H}_*$.

Let a sequence of finite dimensional spaces

$$\mathcal{P}_1 \subset \mathcal{P}_2 \subset \cdots \subset \mathcal{P}_k \subset \mathcal{P}_{k+1} \subset \cdots \subset Dom(T^*) \subset \mathcal{H}_*$$

be given, and let $P_k : \mathcal{H}_* \longrightarrow \mathcal{P}_k$ be the projection that is orthogonal with respect to the inner product $\langle u, v\rangle_*$. We construct the intermediate quadratic forms $a_k(u)$ as

$$a_k(u) = a_0(u) + \|P_kTu\|_*^2$$

for all $u \in Dom(a_k) = Dom(a_0) \cap Dom(T)$. Since $Ran(P_k) \subset Dom(T^*)$, we may extend $\|P_kTu\|_*$ to all of $\mathcal{H}$ by continuity where $a_k(u)$ may be associated with a self-adjoint operator given by

$$A_k = A_0 + T^*P_kT$$

with $Dom(A_k) = Dom(A_0)$. By an argument similar to Section 1.3.2, the second monotonicity theorem with Bessel's inequality yields that the eigenvalues of A_k provide lower bounds to the corresponding eigenvalues of A that improve increasingly in k.

The W-A matrix, $W_k(\lambda)$, whose zeros of determinant provide the eigenvalues of A_k, may be represented as

$$(\langle p_i, p_j\rangle_* + \langle R_\lambda^0 T^*p_i, T^*p_j\rangle).$$

By a special choice of the vectors p_i, i.e., $T^*p_i = u_i^0$, the matrix $W_k(\lambda)$ is compactly expressed as

$$(\frac{1}{\lambda_i^0 - \lambda}\delta_{ij} + \langle p_i, p_j \rangle_*).$$

For more information, one may refer to [19].

If we take a truncation of A_0 at λ_n^0, the intermediate operators are written by

$$A_{n,k} = A_0^n + T^* P_k T$$

and the W-A matrix is obtained by

$$W_{n,k}(\lambda) = \left(\langle p_i, p_j \rangle_* + \left\langle R_\lambda^{(n)} T^* p_i, T^* p_j \right\rangle \right)$$

from the equation $A_{n,k}u = \lambda u$. Notice that $A_{n,k}$ has λ_{n+1}^0 as an eigenvalue of infinite multiplicity. Following [17], if we define for some fixed $\delta \neq 0$

$$B_k = T^* P_k T + \delta^2 I$$

then the operator B_k is bounded, self-adjoint and positive definite. The intermediate operators for the second projection are defined to be the same form as in the previous section. The W-A matrix is also expressed as

$$(\langle \hat{p}_i + R_{\lambda+\delta^2}^0 B_k \hat{p}_i, B_k \hat{p}_j \rangle)$$

in which B_k has an explicit inverse given by

$$B_k^{-1} v = \frac{1}{\delta^2}(v - \sum_{i,j=1}^{k} \langle v, T^* p_i \rangle \, c_{ij} T^* p_j)$$

where (c_{ij}) is the matrix inverse to $(\delta^2 \langle p_i, p_j \rangle_* + \langle T^* p_i, T^* p_j \rangle)$. Therefore if we take a special choice of the $\hat{p}_i$, i.e., $\hat{p}_i = B_k^{-1} u_i^0$, the W-A matrix is explicitly computable as

$$\left[(\frac{1}{\delta^2} + \frac{1}{\lambda_i^0 - \lambda - \delta^2})\delta_{ij} - \frac{1}{\delta^2} \sum_{l,m=1}^{k} \left\langle u_i^0, T^* p_l \right\rangle c_{lm} \left\langle T^* p_m, u_j^0 \right\rangle \right].$$

This expression(cf. [17]) is applied to build the EVF method.

1.4 Remarks.

We note that the Rayleigh–Ritz method does not always strictly improve previously obtained bounds at each successive stage [73]. For instance, if we take eigenvectors of the given operator as test functions, there is no improvement. But if trial functions are chosen from a set complete in a sufficiently strong topology, the bounds will converge to the eigenvalues. In the method of Weinstein, we can obtain the base problem by the removal of constraints. In the method of Aronszajn, the base problem is found by neglecting a positive term in the expression of the given operator. In applications we often have the advantage of the method of Bazley and Fox that the approximating functions, $\{p_i\}$, may be chosen from $Dom(T^*)$ rather than $Dom(B)$ [44]. This usually means that the vectors $\{p_i\}$ satisfy fewer boundary conditions.

In the method of Aronszajn, the essential spectrum of A_k is the same as that of A_0 since the operator BP_k is compact. Assume that $\lambda_\infty^0 = \lambda_\infty$, and that A_0 and A have spectrum which begins with isolated eigenvalues of finite multiplicity, then so does A_k. In order to succeed with Aronszajn's method, the base problem must be selected in such a way that it has no essential spectrum below an eigenvalue to be approximated. There are important classes of problems for which rigorous lower bounds are of interest but for which Aronszajn's method generally gives no more knowledge than was initially available from the base operator for the fixed essential spectrum. For such a problem, Fox presented [36] a method which can move the essential spectrum of the base operator. He used techniques on the tensor product structure of the underlying Hilbert space and on separation of variables so that the operator BP_k is noncompact. That is, the projecting space $\mathcal{P}_k$ is of infinite dimension. Significantly, he showed how this could be done in such a way that it still allows A_k to be computationally resolvable.

CHAPTER 2

A STUDY OF CONVERGENCE RATES FOR SEMI–BOUNDED OPERATORS AND INTERMEDIATE PROBLEMS

2.1 Introduction.

In this chapter we present conditions sufficient to guarantee the convergence of eigenvalues of an increasing sequence of operators in S and also derive convergence rates for the sequence of operators. These results will be applied to the methods of intermediate problems including a variant of a method of Aronszajn known as truncation including the remainder which was analyzed by Beattie and Greenlee [19,20,43]. We note that though the conditions for convergence we give may be the same as those of Brown [28] and Beattie and Greenlee [18], the convergence rates obtained are slightly improved over those obtained very recently by Beattie and Greenlee [20]. Moreover the convergence theorem for the method of truncation including the remainder follows as a special case. For the method of Weinstein, our result of convergence rate appears to be the first one for a general choice of approximating vectors p_i. For this purpose, we discuss the convergence rate of the eigenvalues and eigenvectors of the increasing sequence, $\{A_k\}_{k=0}^{\infty}$, of semi-bounded operators using techniques to analyze finite–element methods for the differential eigenvalue problem [6].

Throughout this chapter we denote by $\mathcal{U}$ the eigenspace of A corresponding to the eigenvalue $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+m-1}$ with multiplicity m which is less than λ_∞^0, the lowest point of the essential spectrum of A_0. Similarly, $\mathcal{U}^{(k)}$ denotes the eigenspace of A_k corresponding to the eigenvalues $\lambda_i^{(k)}, \lambda_{i+1}^{(k)}, \ldots, \lambda_{i+m-1}^{(k)}$. We also represent the spectral projections of A and A_k onto $\mathcal{U}$ and $\mathcal{U}^{(k)}$ as E and E_k respectively.

In Section 2 we review a result of Weidmann and the relevant theory for the finite element method usually used for differential eigenvalue problems. With the aid of these results, we will provide sufficient conditions for the convergence of eigenvalues and also derive the corresponding rate for a sequence of semi-bounded operators in $\mathcal{S}$. Section 3 deals with application of the derived results to the problem types of Aronszajn and Bazley–Fox as well as those of Weinstein. We derive a convergence rate for the method of truncation including the remainder in Section 4. Finally in Section 5 we present a numerical example of a one dimensional Schrödinger operator with a potential for the method of truncation including the remainder.

2.2 Convergence Rates for Semi-bounded Operators.

In this section we present some convergence results and estimates of convergence rates for the sequence,$\{A_k\}_{k=0}^{\infty}$, of operators and A which are in $\mathcal{S}$ as well as sufficient conditions for the convergence of their eigenvalues. We first assume that the A_k and A are bounded. It is then well known [58] that if A_k converges to A uniformly, then $\lambda_i^{(k)}$ converges to λ_i. For a sequence of compact operators we need only strong convergence to get convergence of their eigenvalues by an analog of Dini's theorem [8].

THEOREM 2.2.1. *Let A be a compact and self-adjoint operator and let $\{A_k\}$ be a sequence of compact and self-adjoint operators such that $A \leq A_{k+1} \leq A_k$. If A_k converges to A strongly, then A_k converges uniformly to A.*

PROOF: We assume that A_k does not converge to A uniformly. Since $A_k - A$ is symmetric,

$$\|A_k - A\| = \sup_{\|u\|=1} \langle (A_k - A)u, u \rangle .$$

Thus there exists a positive number δ and a sequence $\{u_k\}$ with $\|u\| = 1$ such that $\langle (A_k - A)u_k, u_k \rangle \geq \delta$, for any k. Since the sequence $\{A_k\}$ is decreasing, we have for any fixed N,

$$\langle A_N u_k, u_k \rangle \geq \langle A_k u_k, u_k \rangle \geq \langle A u_k, u_k \rangle + \delta$$

for any $k \geq N$. Since a Hilbert space is weakly compact, there is a subsequence of $\{u_k\}$, denoted again by $\{u_k\}$, and u such that u_k converges to u weakly. Since A and A_N are compact, it follows that $A_N u_k$ and $A u_k$ converge strongly to $A_N u$ and Au, respectively. Consequently we have

$$\langle A_N u, u \rangle \geq \langle Au, u \rangle + \delta$$

which is a contradiction to the assumption.■

We introduce some convergence rates for the sequence of bounded operators whose proof may be found in [6].

THEOREM 2.2.2(BABUŠKA AND OSBORN). *Let (A_k) be a sequence of bounded operators which converges to A uniformly. Then for any i and $j = i, i+1, \ldots, i+m-1$ and $u \in \mathcal{U}$, we have a sufficiently large k such that*

(1) $|\lambda_i - \lambda_j^{(k)}| \leq \max_{u\in\mathcal{U},\|u\|=1} |\langle (A_k - A)u, u\rangle| + C_1 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k - A)u\|^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k - A)u\|$

for some constants C_1 and C_2 independent of k.

We now assume that A_k and A are bounded below such that $A_k \leq A_{k+1} \leq A$, for all $k \geq 0$. We recall the following definition. (e.g., [33, 45])

DEFINITION. *Let A_k be a sequence of self-adjoint operators acting in a Hilbert space $\mathcal{H}$. We say that the A_k converges to A in the strong resolvent sense if*

$$(A_k - z)^{-1} \longrightarrow (A - z)^{-1} \quad \textit{strongly}$$

for some z which is bounded away from the spectra of the A_k and A.

If the A_k and A are all coercive, convergence in the strong resolvent sense is equivalent to the strong convergence of A_k^{-1} to A^{-1}. It has been well known [44,58] that if the self–adjoint operators A and B are compact (even bounded), then the differences of the corresponding eigenvalues of A and B are dominated by the norm of the difference of the operators. Thus the uniform convergence of a sequence of bounded operators implies the convergence of the corresponding eigenvalues. For a sequence of semi–bounded operators, we have

THEOREM 2.2.3. *Let A be in $\mathcal{S}$ and let $\{A_k\}$ be a sequence of operators in $\mathcal{S}$ such that $A \leq A_{k+1} \leq A_k$ and $Dom(A) = Dom(A_k)$ for all k. If $\mathcal{U}$ is the eigenspace of A corresponding to the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_l$, then*

$$0 \leq \lambda_l^{(k)} - \lambda_l \leq \max_{u\in\mathcal{U},\|u\|=1} \|(A_k - A)u\|.$$

PROOF: It easily follows from the minimum-maximum principle that

$$\begin{aligned}\lambda_l^{(k)} &= \min_{\substack{\mathcal{P}_l \subset Dom(A_k) \\ dim\, \mathcal{P}_l = l}} \max_{\substack{u\in\mathcal{P}_l \\ \|u\|=1}} \langle A_k u, u\rangle \le \max_{\substack{u\in\mathcal{U} \\ \|u\|=1}} \langle A_k u, u\rangle \\ &\le \lambda_l + \max_{\substack{u\in\mathcal{U} \\ \|u\|=1}} \langle (A_k - A)u, u\rangle \\ &\le \lambda_l + \max_{\substack{u\in\mathcal{U} \\ \|u\|=1}} \|(A_k - A)u\|.\end{aligned}$$

Hence we have the result.■

The goal of this section is to get the same conclusion as in Theorem 2.2.2 for a sequence of semi–bounded operators in $\mathcal{S}$. We modify a result of Weidmann for our problem setting. Notice that the result already had been applied to get sufficient conditions for the convergence of eigenvalues in [19–21,27,28].

LEMMA 2.2.4(WEIDMANN). *Let (A_k) be an increasing sequence of operators in $\mathcal{S}$ which converges to A in the strong resolvent sense. Let $\lambda_1^{(k)} \le \lambda_2^{(k)} \le \cdots \le \lambda_\infty^{(k)}$ and $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_\infty$ be the isolated eigenvalues of A_k and A, respectively. Then for all i such that $\lambda_i < \lambda_\infty^{(0)}$, $\lambda_i^{(k)}$ converges to λ_i, where λ_∞^0 denotes the lowest point of the essential spectrum of A_k.*

With the aid of Lemma 2.2.4 and the proof of Theorem 2.2.2 in [6], we have the main estimate result which plays a crucial role in our estimates.

THEOREM 2.2.5. *Let (A_k) be an increasing sequence of $\mathcal{S}$ which converges to A in the strong resolvent sense. Then for all i such that $\lambda_i < \lambda_\infty^{(0)}$, $\lambda_i^{(k)}$ converges to λ_i as k becomes large. Furthermore, if λ_i has multiplicity m with $\lambda_i = \lambda_{i+1} = \cdots = \lambda_{i+m-1}$, we have the following estimates,*

(1) $|\lambda_i - \lambda_j^{(k)}| \le \max_{u\in\mathcal{U},\|u\|=1} |\langle (A_k - A)u, u\rangle| + C_1 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k - A)u\|^2$

(2) $|\frac{1}{\lambda_i} - \frac{1}{\lambda_j^{(k)}}| \le \max_{u\in\mathcal{U},\|u\|=1} |\langle (A_k^{-1} - A^{-1})u, u\rangle| + C_2 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k^{-1} - A^{-1})u\|^2$

(3) $\|u - E_k u\| \le C_3 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k^{-1} - A^{-1})u\|^2 \le C_4 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(A_k - A)u\|^2$

for k sufficiently large and for some constants C_i's independent of k.

PROOF: We note that the spectral projection associated with λ_i is denoted by

$$E = \frac{1}{2\pi i} \int_{\Gamma} R_z(A)\, dz,$$

where Γ is a circle in the complex plane centered at λ_i which lies in the resolvent set, $\rho(A)$, of A and which encloses no other points of the spectrum, $\sigma(A)$, of A and $R_z(A)$ is the resolvent operator of A at z, i.e. $R_z(A) = (z - A)^{-1}$. Since A_k converges monotonically to A in the strong resolvent sense, Lemma 2.2.4 implies that $\lambda_j^{(k)}$ converges to λ_i as k goes to ∞ for $j = i, i+1, \ldots, i+m-1$. There is thus a sufficiently large k such that Γ lies also in $\rho(A_k)$ enclosing only λ_i and $\{\lambda_j^{(k)}\}_{j=i}^{i+m-1}$. Thus the spectral projection E_k associated with A_k and $\{\lambda_j^{(k)}\}_{j=i}^{i+m-1}$ may be expressed as

$$E_k = \frac{1}{2\pi i} \int_{\Gamma} R_z(A_k)\, dz.$$

Hence for any $u \in \mathcal{U}$, we have

$$\begin{aligned}
\|(E - E_k)u\| &= \|\frac{1}{2\pi i} \int_{\Gamma} (R_z(A_k) - R_z(A))u\, dz\| \\
&\leq \frac{1}{2\pi} \| \int_{\Gamma} R_z(A_k)(A - A_k)R_z(A)u\, dz\| \\
&\leq \frac{1}{2\pi} \cdot \ell(\Gamma) \cdot \max_{z\in\Gamma} \|R_z(A_k)\| \cdot \max_{u\in\mathcal{U}, \|u\|=1} \|(A_k - A)u\| \\
&\cdot \max_{z\in\Gamma} \|R_z(A)\| \|u\| \\
&\leq C \cdot \max_{u\in\mathcal{U}, \|u\|=1} \|(A_k - A)u\|, \quad \text{for some } C \text{ independent of } k
\end{aligned}$$

because $R_z(A_k)$ and $R_z(A)$ are uniformly bounded on Γ. Here $\ell(\Gamma)$ is the *arc length* of Γ. This gives (3). Since A_k converges to A in the strong resolvent sense with $\mathcal{U}$ as a finite dimensional space, E_k converges to E on the space $\mathcal{U}$.

Let $\hat{E}_k : \mathcal{U} \longrightarrow \mathcal{U}^{(k)}$ be the restriction of E_k to the space $\mathcal{U}$. Suppose that $\hat{E}_k u = 0$ for some $u \in \mathcal{U}$. Then

$$\|u\| = \|(E - E_k)u\| \leq \max_{u\in\mathcal{U}, \|u\|=1} \|(E_k - E)u\| \|u\|.$$

Since E_k converges to E on the space $\mathcal{U}$, we have that $u = 0$ for k sufficiently large. Since $dim\,\mathcal{U} = dim\,\mathcal{U}^{(k)}$, it follows that $\hat{E}_k : \mathcal{U} \to \mathcal{U}^{(k)}$ is bijective for sufficiently large k. Furthermore

$$\|\hat{E}_k^{-1}\| \leq 2$$

for k sufficiently large, since for any $u \in \mathcal{U}$,

$$\|u\| - \|\hat{E}_k u\| \leq \max_{u \in \mathcal{U}, \|u\|=1} \|(E_k - E)u\| \cdot \|u\| \leq \frac{1}{2}\|u\|$$

for k sufficiently large. For convenience, let $T_k = \hat{E}_k^{-1} A_k \hat{E}_k$. Then T_k is an operator from $\mathcal{U}$ onto $\mathcal{U}$ having eigenvalues which are

$$\sigma(T_k) = \{\lambda_j^{(k)}\}_{j=i}^{i+m-1}.$$

Let $w_k \in \mathcal{U}$ be defined so that $T_k w_k = \lambda_j^{(k)} w_k$ for some fixed $i \geq j \geq i + m - 1$ and $\|w_k\| = 1$. Then

$$\lambda_i - \lambda_j^{(k)} = \langle (A - T_k) w_k, w_k \rangle.$$

Since $\hat{E}_k^{-1} E_k$ is the identity on $\mathcal{U}$, we have for any $v \in \mathcal{U}$ with $\|v\| = 1$,

$$\begin{aligned}
\langle (A - T_k)v, v \rangle &= \langle \hat{E}_k^{-1} E_k A v, v \rangle - \langle \hat{E}_k^{-1} A_k \hat{E}_k v, v \rangle \\
&= \langle \hat{E}_k^{-1} E_k (A - A_k) v, v \rangle \\
&= \langle (I - \hat{E}_k^{-1} E_k)(A_k - A)v, v \rangle - \langle (A_k - A)v, v \rangle.
\end{aligned}$$

Since $E_k \hat{E}_k^{-1} = I$ on $\mathcal{U}^{(k)}$, we have that $I - \hat{E}_k^{-1} E_k = (I - E_k)(I - \hat{E}_k^{-1} E_k)$ and $(I - E_k)v = (E - E_k)v$ for any $v \in \mathcal{U}$. Thus it follows from (3) that

$$\begin{aligned}
|\langle (I - \hat{E}_k^{-1} E_k)(A_k - A)v, v \rangle| &= |\langle (I - \hat{E}_k^{-1} E_k)(A_k - A)v, (E - E_k)v \rangle| \\
&\leq \|I - \hat{E}_k^{-1} E_k\| \|(A_k - A)v\| \|(E - E_k)v\| \\
&\leq 3C \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(A_k - A)u\|^2
\end{aligned}$$

for sufficiently large k. It leads to (1). By the same way (2) also follows. ∎

COROLLARY 2.2.6. *Let (A_k) be an increasing sequence of operators in S and let A be in S such that $A_k \leq A$ for all k. If $A_k v$ converges strongly to Av for any $v \in Dom(A)$, then A_k converges to A in the strong resolvent sense. Thus we have the same results as in Theorem 2.2.5.*

PROOF: It easily follows from the fact that

$$A_k^{-1} - A^{-1} = A_k^{-1}(A - A_k)A^{-1}. \quad \blacksquare$$

For any self adjoint operator A_k, the corresponding closed quadratic form is denoted by $a_k(u)$. It follows from [33] and Lemma 2.2.4 that we have the following theorem. One may also refer to Kato [45] and Simon [61].

THEOREM 2.2.7. *Let (A_k) be an increasing sequence of operators in S which is dominated by $A \in S$ from above. We assume that for u in $\cap_{k\geq 0} Dom(a_k)$ such that $a_k(u)$ is uniformly bounded, the vector u is in $Dom(a)$ and $a_k(u)$ converges to $a(u)$. Then A_k converges to A in the strong resolvent sense and thus for all i such that $\lambda_i < \lambda_\infty^{(0)}$, $\lambda_i^{(k)}$ converges to λ_i as k goes to ∞ .*

The set in the second hypothesis of Theorem 2.2.7 can be expressed as the domain of a_∞. That is,

$$Dom(a_\infty) = \{u \in \cap_{k\geq 1} Dom(a_k) : \sup a_k(u) < \infty\}$$

and $a_\infty(u) = \lim_{k\to\infty} a_k(u)$, for all $u \in Dom(a_\infty)$. Both Theorem 2.2.5 and 2.2.7 will be applied to get the sufficient conditions for the convergence of eigenvalues and also its rate for the intermediate problems with the method of truncation including remainder. If $Dom(a) = \cap_{k\geq 1} Dom(a_k)$ is to be assumed, $a_k(u)$ is uniformly bounded for $u \in Dom(a)$ for $a_k \leq a$. Thus we get the following useful corollaries.

COROLLARY 2.2.8. *Let (A_k) be an increasing sequence of operators in S. Let A be in S such that $A_k \leq A$ and assume that $Dom(a) = \cap_{k\geq 1} Dom(a_k)$. If $a_k(u) \longrightarrow a(u)$ for all $u \in Dom(a)$, then A_k converges to A in the strong resolvent sense and thus we have the same results as in Theorem 2.2.5.*

Let (A_k) be an increasing sequence of bounded and self adjoint operators such that A_k converges weakly to a bounded and self-adjoint operator A. It follows from Corollary 2.2.8 that A_k converges to A in the strong resolvent sense. Since $A_k - A = A_k(A^{-1} - A_k^{-1})A$, we have the strong convergence of A_k to A but not the uniform convergence. However we obtain the convergence of eigenvalues with its rate. One may compare this with Theorem 2.2.2.

COROLLARY 2.2.9. *Let (A_k) be an increasing sequence of bounded and self adjoint operators. If A_k converges weakly to A, then $\lambda_i^{(k)}$ converges to λ_i. Thus we have the same estimates as in Theorem 2.2.5.*

We review the following basic results because they may be used in intermediate problems.

THEOREM 2.2.10([30]). *Let A and B be self-adjoint operators. Then*

$$\alpha < A \leq B \quad \text{if and only if} \quad (A-\alpha)^{-1} \geq (B-\alpha)^{-1} > 0.$$

This allows us to transform a monotone increasing sequence of unbounded operators into an equivalent monotone decreasing sequence of bounded operators. That is,

$$\begin{aligned} & 0 < \alpha \leq A_0 \leq \cdots \leq A_k \leq A_{n+1} \leq \cdots \leq A \\ \Longleftrightarrow\ & \frac{1}{\alpha} \geq A_0^{-1} \geq \cdots \geq A_k^{-1} \geq A_{n+1}^{-1} \geq \cdots \geq A^{-1}. \end{aligned}$$

THEOREM 2.2.11. *Let A and B be self-adjoint such that $0 < B < A$. If A is compact, then B is compact.*

PROOF: We note that $A^{\frac{1}{2}}$ is compact because A is positive and compact. Let $\{x_k\}$ be a sequence of vectors such that x_k converges to 0 weakly. Since $A^{\frac{1}{2}}$ is compact, $A^{\frac{1}{2}}x_k$ converges to 0 strongly. Since

$$\begin{aligned}\|B^{\frac{1}{2}}x_k\|^2 = \left\langle B^{\frac{1}{2}}x_k, B^{\frac{1}{2}}x_k \right\rangle &= \langle B_k x, x_k\rangle \\ &\leq \langle Ax_k, x_k\rangle = \|A^{\frac{1}{2}}x_k\|,\end{aligned}$$

$B^{\frac{1}{2}}$ is compact and thus B is compact.∎

This implies that if the base operator A_0 has compact resolvent, then all intermediate operators with the given operator have compact resolvent.

2.3 Convergence Rate for Intermediate Problem Methods.

We introduce the following notation in order to lay out convergence rate results. For any densely defined closed positive coercive quadratic form $c(u)$ on $\mathcal{H}$, let $\mathcal{M}$ and $\mathcal{N}$ be subspaces of $Dom(c)$ with $dim\,\mathcal{N} > 0$. Beattie and Greenlee [19] define the containment gap relative to $c(u)$ for the approximation of $\mathcal{M}$ by $\mathcal{N}$ as

$$\delta_c(\mathcal{M},\mathcal{N}) = \sup_{0\neq u\in\mathcal{N}} \inf_{v\in\mathcal{M}} \frac{\|u-v\|_c}{\|u\|}.$$

We note that $\delta_c(\mathcal{M},\mathcal{N})$ is not symmetric in $\mathcal{N}$ and $\mathcal{M}$, and $\delta_c(\mathcal{M},\mathcal{N}) = 0$ if and only if $\mathcal{M} \supset \mathcal{N}$. Likewise we denote

$$\delta_{\mathcal{M}}(\mathcal{N}) = \sup_{0\neq u\in\mathcal{N}} \inf_{v\in\mathcal{M}} \frac{\|u-v\|}{\|u\|}.$$

We note that this is unlike the gap of Kato [45].

For the speed of the convergence, Weinberger gave an error estimation for the convergence in 1952 [66] which is historically the first example of convergence rate for intermediate problems. For the basic convergence rate for the Rayleigh-Ritz method, one may refer to [6,20,35].

2.3.1 On the Weinstein type. We recall that the operator A and the intermediate operators A_k are written in terms of A_0 as

$$A = QA_0Q \text{ and } A_k = Q_kA_0Q_k,$$

where Q is the orthogonal projection of $\mathcal{H}_{a_0}$ onto $\mathcal{H}_a$ and $Q_k : \mathcal{H}_{a_0} \longrightarrow \mathcal{P}_k^{\perp}$ is the orthogonal projection onto $\mathcal{P}_k^{\perp}$. We note that the sequence, $\{\mathcal{P}_k^{\perp}\}$, of subspaces with codimension k satisfies the inequality,

$$\mathcal{H}_a \subset \cdots \subset \mathcal{P}_{k+1}^{\perp} \subset \mathcal{P}_k^{\perp} \subset \cdots \subset \mathcal{P}_0^{\perp} = \mathcal{H}_{a_0}.$$

Thus the corresponding projections have the property

$$I = Q_0 \geq Q_1 \geq \cdots \geq Q_k \geq Q_{k+1} \geq \cdots \geq Q.$$

We suppose that the sequence of vectors, $\{p_i\}$, in $\mathcal{H}_{a_0}$ is selected such that it is complete in $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$. Each vector u in $\mathcal{H}_{a_0}$ may be uniquely decomposed as $u = v + w$, where $v \in \mathcal{H}_a$ and $w \in \mathcal{H}_{a_0} \ominus \mathcal{H}_a$. Thus

$$\begin{aligned}\|(Q_k - Q)u\|_{a_0} &= \|(P_k - P)u\|_{a_0} \\ &= \|(P_k - I)w\|_{a_0} \longrightarrow 0 \quad as \quad n \to \infty,\end{aligned}$$

where $P : \mathcal{H}_{a_0} \longrightarrow \mathcal{H}_{a_0} \ominus \mathcal{H}_a$ is the orthogonal projection and $Q_k = I - P_k$.

Aronszajn and Weinstein [4] showed in 1949 the convergence of the Weinstein method under the assumption that the base operator A_0 has a compact inverse and that $\{p_i\}$ is complete in $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$. One may refer to [44,73] for the proof. In 1984 Brown [27] showed the convergence without any compactness assumption on A_0.

LEMMA 2.3.1. *Let the sequence, $\{p_i\}$, of vectors be complete in $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$ and let A_0 be compact. Then $Q_k A_0 Q_k$ converges uniformly to $Q A_0 Q$.*

PROOF: Since $Q_k A_0 Q_k$ and $Q A_0 Q$ are compact, it suffices to show only the strong convergence. For any $u \in \mathcal{H}_{a_0}$, we have

$$\begin{aligned}\|(Q_k A_0 Q_k - Q A_0 Q)u\|_{a_0} &\le \|(Q_k - Q)A_0 Q u\|_{a_0} + \|Q_k A_0 (Q_k - Q)u\|_{a_0} \\ &\le \|(Q_k - Q)A_0 Q u\|_{a_0} + \|A_0\|_{a_0} \|(Q_k - Q)u\|_{a_0} \\ &\longrightarrow 0 \quad \text{as } k \to \infty. \quad \blacksquare\end{aligned}$$

Let $\mathcal{U}$ be the eigenspace of A corresponding to λ_i with multiplicity m. Then

$$\begin{aligned}\max_{u \in \mathcal{U}, \|u\|_{a_0}=1} \|(Q_k A_0 Q_k - Q A_0 Q)u\|_{a_0} &= \max_{u \in \mathcal{U}, \|u\|_{a_0}=1} \|(Q_k - Q)A_0 u\|_{a_0} \\ &\le \max_{u \in \mathcal{U}} \frac{\|(Q_k - Q)A_0 u\|_{a_0}}{\|A_0 u\|_{a_0}} \cdot \frac{\|A_0 u\|_{a_0}}{\|u\|_{a_0}} \\ &\le \max_{u \in \mathcal{U}, \|u\|_{a_0}=1} \|A_0 u\|_{a_0} \max_{v \in A_0 \mathcal{U}} \frac{\|(Q_k - Q)v\|_{a_0}}{\|v\|_{a_0}}\end{aligned}$$

and since Q_k and Q are identities on the space $\mathcal{U}$, it follows that for any $u \in \mathcal{U}$,

$$\begin{aligned}\langle (Q_k A_0 Q_k - Q A_0 Q)u, u\rangle_{a_0} &= \langle (Q_k - Q)A_0 u, u\rangle_{a_0} \\ &= \langle A_0 u, (Q_k - Q)u\rangle_{a_0} \\ &= 0.\end{aligned}$$

Therefore it follows directly from Theorem 2.2.2 that we have

THEOREM 2.3.2. *Let the sequence, $\{p_i\}$, of vectors be complete in $\mathcal{H}_{a_0} \ominus \mathcal{H}_a$ and let A_0 be compact. Then for $j = i, i+1, \ldots, i+m-1$, we have*

$$|\lambda_i - \lambda_j^{(k)}| \leq C_i \cdot \delta^2_{A_0\mathcal{U}}(\mathcal{P}_k)$$

for some constant C_i independent of k.

PROOF: It is enough to show that

$$\begin{aligned}\max_{v \in A_0\mathcal{U}} \frac{\|(Q_k - Q)v\|_{a_0}}{\|v\|_{a_0}} &= \max_{v \in A_0\mathcal{U}} \frac{\|(P_k - I)v\|_{a_0}}{\|v\|_{a_0}} \\ &= \max_{v \in A_0\mathcal{U}} \min_{p \in \mathcal{P}_k} \frac{\|v - p\|_{a_0}}{\|v\|_{a_0}} \\ &= C_i \cdot \delta_{A_0\mathcal{U}}(\mathcal{P}_k). \quad \blacksquare\end{aligned}$$

2.3.2 On the Aronszajn type. We recall that the intermediate operators of the Aronszajn method may be expressed as $A = A_0 + B$ and $A_k = A_0 + BP_k$ such that

$$0 < A_0 \leq A_k \leq A_{k+1} \leq A$$

with $Dom(A) \subset Dom(A_k) = Dom(A_0)$. Here B is assumed to be coercive and $P_k : \mathcal{H}_b \longrightarrow \mathcal{P}_k$ is a b–orthogonal projection. For any $v \in Dom(A)$, we have

$$\|(A_k - A)v\| = \|B(P_k - I)v\| \text{ and } \langle (A_k - A)v, v\rangle = \|(P_k - I)v\|_b^2.$$

THEOREM 2.3.3. *If the set of vectors $\{p_i\}$ is chosen such that span$\{p_i\}$ is dense in $Dom(B)$ with respect to the norm $\|Bu\|$, then $\lambda_i^{(k)}$ converges to λ_i for any i satisfying $\lambda_i < \lambda_\infty^0$. Also for $j = i, i+1, \ldots, i+m-1$ and for any $u \in \mathcal{U}$,*

(1) $|\lambda_i - \lambda_j^{(k)}| \leq \max_{u \in \mathcal{U}, \|u\|=1} \|(I - P_k)u\|_b^2 + C_1 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|B(I - P_k)u\|^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|B(I - P_k)u\|$

for some C_i's independent of k.

For this we need the following lemmas.

LEMMA 2.3.4. *If the set of vectors $\{p_i\}$ span a dense subspace in $Dom(B)$ with respect to the graph norm $\|Bu\|$, then span$\{Bp_i\}$ is dense in $\mathcal{H}$ and thus $\{p_i\}$ is a core of B.*

PROOF: Suppose that there is a vector u in $\mathcal{H}$ such that $\langle u, Bp_i \rangle = 0$ for all i. Since B is coercive and $Ran(B) = \mathcal{H}$, there is a bounded inverse B^{-1} such that $u = BB^{-1}u$. Since $\{p_i\}$ is dense in $Dom(B)$ with respect to $\|Bu\|$, it follows that $B^{-1}u = 0$. Thus $u = 0$.∎

LEMMA 2.3.5. *Let $\{p_i\} \subset Dom(B)$ be chosen such that span$\{Bp_i\}$ is dense in $\mathcal{H}$. Then the set of all vectors for which $b(P_k u)$ is uniformly bounded with respect to k is the domain of b. That is,*

$$\{u \in \mathcal{H} \mid b(P_k u) < \infty, \text{ for all } k\} = Dom(b).$$

PROOF: Note that $P_k u = \sum_{i,j=1}^{k} \langle u, Bp_i \rangle b_{ij}\, p_j$ where $[b_{ij}]$ is the inverse to the matrix $[\langle p_i, Bp_j \rangle]$. Thus we have

$$\begin{aligned}
\langle P_k u, v \rangle &= \sum_{i,j=1}^{k} \langle u, Bp_i \rangle b_{ij} \langle p_j, v \rangle \\
&= \langle u, \sum_{i,j=1}^{k} Bp_i b_{ij} \langle p_j, v \rangle \rangle \\
&= \langle u, \sum_{i,j=1}^{k} \langle v, p_j \rangle b_{ij} Bp_i \rangle \\
&= \langle u, P_k^* v \rangle.
\end{aligned}$$

That is,

$$P_k^* v = \sum_{i,j=1}^{k} \langle v, p_i \rangle b_{ij} Bp_j.$$

Hence

$$\begin{aligned} P_k^* B v &= \sum_{i,j=1}^{k} \langle Bv, p_i \rangle b_{ij} B p_j \\ &= \sum_{i,j=1}^{k} \langle v, Bp_i \rangle b_{ij} B p_j \\ &= B P_k v. \end{aligned}$$

We note that for any n and m, we have

$$\lim_{n,m\to\infty} b(P_n u - P_m u) = \lim_{n,m\to\infty} |b(P_n u) - b(P_m u)| = 0$$

since $P_n P_m = P_m P_n = P_{min(m,n)}$ and $b(P_k u) < \infty$ for all k. Thus there is a vector w in H_b such that $P_k u$ converges to w. For every i, it follows that

$$\begin{aligned} \langle w - u, Bp_i \rangle &= \lim_{n\to\infty} \langle P_n u, Bp_i \rangle - \langle u, Bp_i \rangle \\ &= \langle u, Bp_i \rangle - \langle u, Bp_i \rangle \\ &= 0 \quad \text{for sufficiently large } n. \end{aligned}$$

Since $\{Bp_i\}$ is dense in $\mathcal{H}$, it follows that $w = u$. The converse follows from the fact that P_k is a projection with respect to the norm induced by $\langle u, Bv \rangle$. ∎

Proof of Theorem 2.3.3: By combining Lemmas 2.3.4 and 2.3.5 with Theorems 2.2.7 and 2.2.5, the result is obtained.∎

We note that it may not be easy to interpret the expression $\|B(I - P_k)|_{\mathcal{U}}\|$ since the projection P_k is orthogonal with respect to the inner product $\langle u, Bv \rangle$. However if $\|(P_k - I)|_{\mathcal{U}}\|_b = O(k^{-s})$ and $\|B(P_k - I)|_{\mathcal{U}}\| = O(k^{-t})$ as k becomes large for some constants s and t , then we have

$$|\lambda_i - \lambda_j^{(k)}| = O(k^{-2p}), \text{ where } p = min(s,t)$$

for all $j = i, i+1, \ldots, i+m-1$. This is similar to the result Fix obtained in [24], although he assumed that B was bounded and used a special choice of the vectors p_i's.

We assume that the operator B is relatively bounded with respect to the base operator A_0 with bound m, i.e., $\|Bu\| \leq m\|A_0 u\|$ and $Dom(A_0) \subset Dom(B)$. Thus $Dom(A) = Dom(A_k) = Dom(A_0)$. By the Heinz theorem [64], it follows that $\|B^{\frac{1}{2}} A_0^{-\frac{1}{2}}\| \leq m$. Now we consider the following:

$$\begin{aligned}
\|(A_k^{-1} - A^{-1})u\| &= \|A_k^{-1} B(I - P_k)A^{-1}u\| \\
&\leq \|A_k^{-1} B^{\frac{1}{2}}\| \|(I - P_k)A^{-1}u\|_b \\
&\leq \|B^{\frac{1}{2}} A_o^{-\frac{1}{2}}\| \|A_0^{\frac{1}{2}} A_k^{-\frac{1}{2}}\| \|A_k^{-\frac{1}{2}}\| \|(I - P_k)A^{-1}u\|_b \\
&\leq \frac{m\|A_0^{-\frac{1}{2}}\|}{\lambda_i} \|(I - P_k)u\|_b
\end{aligned}$$

and

$$\begin{aligned}
|\langle (A_k^{-1} - A^{-1})u, u\rangle| &= |\langle (A_k^{-1} B(I - P_k)A^{-1}u, u\rangle| \\
&= |\langle B(I - P_k)A^{-1}u, (I - P_k)A_k^{-1}u\rangle| \\
&\leq \|(I - P_k)A^{-1}u\|_b \|(I - P_k)A_k^{-1}u\|_b.
\end{aligned}$$

Since

$$\begin{aligned}
\|(I - P_k)A_k^{-1}u\|_b &\leq \|(I - P_k)(A_k^{-1} - A^{-1})u\|_b + \|(I - P_k)A^{-1}u\|_b \\
&\leq \|A_k^{-1} B(I - P_k)A^{-1}u\|_b + \|(I - P_k)A^{-1}u\|_b \\
&\leq \|B^{\frac{1}{2}} A_k^{-1} B^{\frac{1}{2}}\| \|B^{\frac{1}{2}}(I - P_k)A^{-1}u\| + \|(I - P_k)A^{-1}u\|_b \\
&\leq \|B^{\frac{1}{2}} A_0^{-1} B^{\frac{1}{2}}\| \|(I - P_k)A^{-1}u\|_b + \|(I - P_k)A^{-1}u\|_b \\
&= (\|B^{\frac{1}{2}} A_0^{-\frac{1}{2}}\|^2 + 1)\|(I - P_k)A^{-1}u\|_b,
\end{aligned}$$

it follows that

$$|\langle (A_k^{-1} - A^{-1})u, u\rangle| \leq \frac{m^2 + 1}{\lambda_i^2} \|(I - P_k)u\|_b^2.$$

In 1961, Bazley and Fox showed that if the perturbation operator B is relatively bounded with respect to A_0, A_0 has a compact inverse and the set of vectors p_i is

complete in $\mathcal{H}_b$, then the inverses of the intermediate operators converge uniformly to the inverse of the given operator, thus guaranteeing the convergence of the eigenvalues. In 1980, Weidmann introduced a weaker condition sufficient to guarantee the convergence of eigenvalues of a sequence of operators in $\mathcal{S}$. He showed that the strong resolvent convergence of a monotone operator sequence is enough for the convergence of the eigenvalues. In 1982, Beattie proved that the completeness of $\{Bp_i\}$ in $\mathcal{H}$ implied strong resolvent convergence of A_k to A and so gave conditions for convergence of intermediate problems.

THEOREM 2.3.6. *Let B be relatively bounded with respect to A_0 and let $\{p_i\}$ be chosen to be complete in $Dom(B)$ with respect to the quadratic form $b(u)$. Then for any i satisfying $\lambda_i < \lambda_\infty^0$, $\lambda_i^{(k)}$ converges to λ_i, and for $j = i, i+1, \ldots, i+m-1$ and for any $u \in \mathcal{U}$,*

(1) $|\lambda_i - \lambda_j^{(k)}| \leq C_1 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(I - P_k)u\|_b$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(I - P_k)u\|_b$

for some constant C_i's independent to k.

PROOF: Since P_k is the orthogonal projection with respect to $b(u)$, P_k converges strongly to I with respect to $b(u)$. Since $\|A_k^{-1} - A^{-1}\| = O(\|(P_k - I)A^{-1}u\|_b)$ for any $u \in \mathcal{H}$, it follows that A_k^{-1} converges strongly to A^{-1}. The conclusion follows from Theorem 2.2.5 and the fact that $\frac{1}{\lambda_j^{(k)}} - \frac{1}{\lambda_i} \geq \frac{\lambda_i - \lambda_j^{(k)}}{\lambda_i^2}$. ∎

Theorem 2.3.6 may be considered as an extension of the results of Bazley and Fox [8] and Poznyak [52] because we do not assume that A and A_0 have compact inverses. We note that the density condition in Theorem 2.3.6 is weaker than that in Theorem 2.3.3 because the latter implies the former. If B is not relatively bounded with respect to A_0, we need the latter condition for the convergence of the eigenvalues. For counter examples, one may refer to [28]. We note in passing that we need only the

relative form boundedness of B to A_0 in the proof instead of the relative boundedness itself. Moreover, the latter is stronger than the former (Heinz Theorem). Beattie and Greenlee [21] showed that the relative boundedness may be replaced by the relative form boundedness under the assumption that $A_0 + B$ is essentially self-adjoint with unique self-adjoint extension A.

COROLLARY 2.3.7. *Let $Q_k : \mathcal{H}_a \longrightarrow \mathcal{P}_k$ be an a–orthogonal projection onto $\mathcal{P}_k$. Then the bound in Theorem 2.3.6 may be replaced by $\|(I - Q_k)|_{\mathcal{U}}\|_a^2$.*

PROOF: The result follows from the fact that

$$\begin{aligned}\|(I - P_k)u\|_b &= \|(I - P_k)(I - Q_k)u\|_b \\ &\leq \|(I - Q_k)u\|_b \\ &\leq \|B^{\frac{1}{2}}A_0^{-\frac{1}{2}}\|\|(I - Q_k)u\|_a \\ &\leq m\|(I - Q_k)u\|_a.\end{aligned}$$

REMARK.

$$\begin{aligned}\|(I - P_k)|_{\mathcal{U}}\|_b &= \max_{u\in\mathcal{U},\|u\|=1} \|u - P_k u\|_b \\ &= \max_{u\in\mathcal{U},\|u\|=1} \min_{p\in\mathcal{P}_k} \|u - p\|_b \\ &= \max_{u\in\mathcal{U}} \min_{p\in\mathcal{P}_k} \frac{\|B^{\frac{1}{2}}u - B^{\frac{1}{2}}p\|}{\|B^{\frac{1}{2}}u\|} \frac{\|B^{\frac{1}{2}}u\|}{\|u\|} \\ &\leq \|B^{\frac{1}{2}}|_{\mathcal{U}}\| \cdot \max_{v\in B^{\frac{1}{2}}\mathcal{U}} \min_{q\in B^{\frac{1}{2}}\mathcal{P}_k} \frac{\|v - q\|}{\|v\|} \\ &= O(\delta_{B^{\frac{1}{2}}\mathcal{U}}(B^{\frac{1}{2}}\mathcal{P}_k)),\end{aligned}$$

since $\mathcal{U}$ is of finite dimension. Likewise it follows from $A^{\frac{1}{2}}\mathcal{U} \subset \mathcal{U}$ that

$$\begin{aligned}\|(I - Q_k)u\|_a &= \max_{u\in\mathcal{U},\|u\|=1} \min_{p\in\mathcal{P}_k} \|u - p\|_a \\ &= O(\delta_{\mathcal{U}}(A^{\frac{1}{2}}\mathcal{P}_k)).\end{aligned}$$

We note that if p_i is selected to be a linear combination of u_i which are the eigenvalues of A corresponding to λ_i, then we have a bound $O(\delta_{\mathcal{U}}(\mathcal{P}_k))$. Also, if B is bounded, then we have $O(\delta_{\mathcal{U}}(\mathcal{P}_k))$ as a bound.

COROLLARY 2.3.8. *Let $\{p_i\}$ be selected to be complete in $\mathcal{H}$. If B is bounded, then for $j = i, i+1, \ldots, i+m-1$ and for any $u \in \mathcal{U}$,*

(1) $|\lambda_i - \lambda_j^{(k)}| = O(\delta_{\mathcal{U}}^2(\mathcal{P}_k))$

(2) $\|u - E_k u\| = O(\delta_{\mathcal{U}}(\mathcal{P}_k))$.

2.3.3 On the Bazley-Fox Type. We recall that the intermediate operators of the Bazley-Fox method are represented by $A_k = A_0 + T^* P_k T$ such that

$$0 < A_0 \leq A_k \leq A_{k+1} \leq A$$

with $Dom(A) \subset Dom(A_k) = Dom(A_0)$. Here $P_k : \mathcal{H}_* \longrightarrow \mathcal{P}_k$ is the $\mathcal{H}_*$–orthogonal projection onto $\mathcal{P}_k$. For any $v \in Dom(A)$, we have

$$\|(A_k - A)v\| = \|T^*(P_k - I)Tv\| \text{ and } \langle (A_k - A)v, v\rangle = \|(P_k - I)Tv\|_*^2.$$

THEOREM 2.3.9. *Let $\{P_k\}$ be an increasing sequence of orthogonal projections in $\mathcal{H}_*$ such that for each k, $Ran(P_k T) \subset Dom(T^*) \cap Ran(T)$. If $\cup Ran(P_k T)$ is a core of T^*, then for any i satisfying $\lambda_i < \lambda_\infty^0$, $\lambda_i^{(k)}$ converges to λ_i, and for $j = i, i+1, \ldots, i+m-1$ and for any $u \in \mathcal{U}$,*

(1) $|\lambda_i - \lambda_j^{(k)}| \leq \max\limits_{u \in \mathcal{U}, \|u\|=1} \|(I - P_k)Tu\|_*^2 + C_1 \cdot \max\limits_{u \in \mathcal{U}, \|u\|=1} \|T^*(I - P_k)Tu\|^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max\limits_{u \in \mathcal{U}, \|u\|=1} \|T^*(I - P_k)Tu\|$

for some constants $\{C_i\}$ independent of k.

PROOF: We note that $\cup Ran(P_k T)$ is a core of T^* if and only if $\cup Ran(P_k T)$ is dense in $Dom(T^*)$ with respect to the graph norm of T^*(cf. [45]). It follows then from [28,18] that A_k converges to A in the strong resolvent sense. Thus Theorem 2.2.5 implies the result. ∎

We assume that the operator T^*T is relatively bounded with respect to the base operator A_0. Then $Dom(a_0) \subset Dom(T)$ and thus $Dom(a) = Dom(a_k) = Dom(a_0)$.

Now we consider that for any $u \in \mathcal{U}$,

$$\begin{aligned}
|\langle (A_k^{-1} - A^{-1})u, u\rangle| &= |\langle (A_k^{-1}T^*(I - P_k)TA^{-1}u, u\rangle| \\
&= |\langle (I - P_k)TA^{-1}u, (I - P_k)TA_k^{-1}u\rangle_*| \\
&\leq \|(I - P_k)TA^{-1}u\|_*(\|T(A_k^{-1} - A^{-1})u\|_* + \|(I - P_k)TA^{-1}u\|_*) \\
&\leq \|(I - P_k)TA^{-1}u\|_*(m^2\|(I - P_k)TA^{-1}u\|_* \\
&\quad + \|(I - P_k)TA^{-1}u\|_*) \\
&\leq (m^2 + 1)\|(I - P_k)TA^{-1}u\|_*^2 \\
&\leq \frac{m^2 + 1}{\lambda_i^2}\|(I - P_k)Tu\|_*^2
\end{aligned}$$

with $m = \|TA_0^{-\frac{1}{2}}\|_*$ and

$$\begin{aligned}
\|(A_k^{-1} - A^{-1})u\| &= \|A_k^{-1}T^*(I - P_k)TA^{-1}u\| \\
&\leq m\|A_0^{\frac{1}{2}}A_k^{-\frac{1}{2}}\|\|A_k^{-\frac{1}{2}}\|\|(I - P_k)TA^{-1}u\|_* \\
&\leq \frac{m\|A_0^{-\frac{1}{2}}\|}{\lambda_i}\|(I - P_k)Tu\|_*.
\end{aligned}$$

Hence we have the following estimation.

THEOREM 2.3.10. *Let T^*T be relatively bounded to A_0 and let P_k converge strongly to I in $\mathcal{H}_*$. Then for any i satisfying $\lambda_i < \lambda_\infty^0$, $\lambda_i^{(k)}$ converges to λ_i , and for $j = i, i+1, \ldots, i+m-1$ and for any $u \in \mathcal{U}$,*

(1) $|\lambda_i - \lambda_j^{(k)}| \leq C_1 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(I - P_k)Tu\|_*^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|(I - P_k)Tu\|_*$

for some constants C_i's independent of k.

PROOF: Since the assumption implies that A_k converges to A in the strong resolvent sense, the conclusion follows from Theorem 2.2.5 and the fact that $\frac{1}{\lambda_j^{(k)}} - \frac{1}{\lambda_i} \geq \frac{\lambda_i - \lambda_j^{(k)}}{\lambda_i^2}$.

■

This may be considered as an extension of the result Poznyak [53] obtained because we do not assume that A and A_0 have compact inverses.

REMARK.

$$\begin{aligned}
\max_{u\in\mathcal{U},\|u\|=1} \|(I-P_k)Tu\|_* &= \max_{u\in\mathcal{U}} \frac{\|Tu-P_kTu\|_*}{\|Tu\|_*}\frac{\|Tu\|_*}{\|u\|} \\
&\le \max_{u\in\mathcal{U},\|u\|=1} \|Tu\|_* \cdot \max_{v\in T\mathcal{U}} \frac{\|v-P_kv\|_*}{\|v\|_*} \\
&= \max_{u\in\mathcal{U},\|u\|=1} \|Tu\|_* \cdot \max_{v\in T\mathcal{U}} \min_{p\in\mathcal{P}_k} \frac{\|v-p\|_*}{\|v\|_*} \\
&= O(\delta_{T\mathcal{U}}(\mathcal{P}_k))
\end{aligned}$$

since $\mathcal{U}$ is of finite dimension.

2.4 Convergence Rates for the Method of Truncation including the Remainder.

We recall that the intermediate forms are

$$a_k(u) = a_0^{(\gamma)}(u) + \tilde{a}(P_k u),$$

for $u \in Dom(a_k) = \mathcal{H}$, with the corresponding self adjoint operator

$$A_k = A_0^{(\gamma)} + \tilde{A}P_k.$$

Here the quadratic form, $\tilde{a}$, is

$$\tilde{a}(u) = a(u) - a_0^{(\gamma)}(u) \geq b(u) \geq \alpha\|u\|^2$$

and the corresponding self adjoint operator

$$\tilde{A} = A - A_0^{(\gamma)},$$

where $A_0^{(\gamma)}$ is the truncation of A_0 at γ. Also, the operator P_k is the projection from $\mathcal{H}_{\tilde{a}}$ onto $\mathcal{P}_k$ that is orthogonal with respect to the inner product $\langle u, \tilde{A}v\rangle$. For clarity, we denote that $\lambda(A)$ is an eigenvalue corresponding to the operator A.

Suppose that the set of vectors $\{p_i\}$ is taken to be dense in $Dom(\tilde{A})$ with respect to the graph norm $\|\tilde{A}u\|$. Then it follows from Lemmas 2.3.4 and 2.3.5 that the set of all vectors with which $\tilde{a}(P_k u)$ is uniformly bounded with respect to k is the domain of $\tilde{a}$. Application of Theorem 2.2.7 implies that A_k converges to A in the strong resolvent sense. Since $\langle u, (A - A_k)u\rangle = \|(P_k - I)u\|_{\tilde{a}}^2$ and $\|(A - A_k)u\| = \|\tilde{A}(P_k - I)u\|$, we have the following estimate for this method.

LEMMA 2.4.1. *If the set of vectors $\{p_i\}$ is dense in $Dom(\tilde{A})$ with respect to the norm $\|\tilde{A}u\|$, then $\lambda_i^{(k)}$ converges to λ_i for any i satisfying $\lambda_i < \gamma$, and for all $j = i, i+1, \ldots, i+m-1$ and $u \in \mathcal{U}$,*

(1) $|\lambda_i(A) - \lambda_j(A_k)| \leq \max_{u\in\mathcal{U},\|u\|=1} \|(I - P_k)u\|_{\tilde{a}}^2 + C_1 \cdot \max_{u\in\mathcal{U},\|u\|=1} \|\tilde{A}(I - P_k)u\|^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|\tilde{A}(I - P_k)u\|$,

for some C_i's independent of k.

It may not be easy to interpret the expression $\max_{u \in \mathcal{U}, \|u\|=1} \|\tilde{A}(I - P_k)u\|$. In order to get an interpretation for the expression, we adopt the following from [19]. We first assume that A is bounded, then

$$\|\tilde{A}(I - P_k)u\| = \|\tilde{A}^{\frac{1}{2}}\| \|(I - P_k)u\|_{\tilde{a}}.$$

Thus A_k converges strongly to A.

Define $Q_k : \mathcal{H} \longrightarrow span_{1 \leq i \leq k}\{\tilde{A}p_i\}$ to be the orthogonal projection. Then

$$\begin{aligned}
\|(I - P_k)u\|_{\tilde{a}} &\leq \|\tilde{A}^{-\frac{1}{2}}\| \|(I - P_k^*)\tilde{A}u\| \\
&\leq \|\tilde{A}^{-\frac{1}{2}}\| \|(I - P_k^*)\| \|(I - Q_k)\tilde{A}u\| \\
&\leq \|\tilde{A}^{-\frac{1}{2}}\| (1 + \|\tilde{A}P_k\tilde{A}^{-1}\|) \|(I - Q_k)\tilde{A}u\| \\
&\leq \|\tilde{A}^{-\frac{1}{2}}\| (1 + \kappa) \|(I - Q_k)\tilde{A}u\|,
\end{aligned}$$

where $\kappa = \|\tilde{A}^{\frac{1}{2}}\| \|\tilde{A}^{-\frac{1}{2}}\|$. It follows that we have

THEOREM 2.4.2. *Assume the hypotheses of Lemma 2.4.1. If A is bounded, then for $j = i, i+1, \ldots, i+m-1$ and $u \in \mathcal{U}$,*

(1) $|\lambda_i(A) - \lambda_j(A_k)| \leq C_1 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(I - Q_k)\tilde{A}u\|^2$

(2) $\|u - E_k u\| \leq C_2 \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(I - Q_k)\tilde{A}u\|$

for some constants C_i's independent of k.

REMARK.

$$\begin{aligned}
\max_{u \in \mathcal{U}} \frac{\|(I - Q_k)\tilde{A}u\|}{\|u\|} &= \max_{u \in \mathcal{U}} \frac{\|\tilde{A}u - Q_k\tilde{A}u\|}{\|\tilde{A}u\|} \frac{\|\tilde{A}u\|}{\|u\|} \\
&\leq \max_{u \in \mathcal{U}, \|u\|=1} \|\tilde{A}u\| \cdot \max_{v \in \tilde{A}\mathcal{U}} \min_{p \in \tilde{A}\mathcal{P}_k} \frac{\|v - p\|}{\|v\|} \\
&= O(\delta_{\tilde{A}\mathcal{U}}(\tilde{A}\mathcal{P}_k)).
\end{aligned}$$

For the following argument we are indebted to Greenlee [43] and Beattie and Greenlee [19,20]. We assume that A is unbounded. It could happen that P_k fails to converge strongly to I so that $\{P_k\}$ may not be uniformly bounded. In order to bypass this difficulty, Greenlee introduced the auxiliary operator

$$\tilde{\tilde{A}} = A^{(\mu)} - A_0^{(\gamma)}$$

where μ is chosen sufficiently large so that the corresponding quadratic form satisfies $\tilde{\tilde{a}}(u) \geq \frac{\alpha}{2}\|u\|^2$. See [43] for a proof that such a μ exists. We then have

$$a^{(\mu)}(u) = a_0^{(\gamma)}(u) + \tilde{\tilde{a}}(u),$$

applying the Aronszajn method to this decomposition of $a^{(\mu)}$.

Given the approximating vectors $\{p_i\}$, we define $\{\hat{p}_i\}$ by $\hat{p}_i = \tilde{\tilde{A}}^{-1}\tilde{A}p_i$, for each $i = 1, 2, \ldots$. Then the following lemma easily follows.

LEMMA 2.4.3. *If $\{p_i\}$ is dense in $Dom(\tilde{A})$ with respective to the norm $\|\tilde{A}u\|$, then the set of $\{\hat{p}_i\}$ is dense in $Dom(\tilde{\tilde{A}})$ with respective to $\|\tilde{\tilde{A}}u\|$.*

PROOF: We assume that $\left\langle \tilde{\tilde{A}}u, \tilde{\tilde{A}}\hat{p}_i \right\rangle = 0$ for some $u \in Dom(\tilde{\tilde{A}})$, then

$$0 = \left\langle \tilde{\tilde{A}}u, \tilde{A}p_i \right\rangle = \left\langle \tilde{A}\tilde{A}^{-1}\tilde{\tilde{A}}u, \tilde{A}p_i \right\rangle.$$

Since $\{p_i\}$ is complete in $Dom(\tilde{A})$ with respect to the graph norm $\|\tilde{A}u\|$, it follows that $\tilde{A}^{-1}\tilde{\tilde{A}}u = 0$. Hence $u = 0$.∎

Now we define $\hat{P}_k : \mathcal{H}_{\tilde{\tilde{a}}} \longrightarrow \hat{\mathcal{P}}_k$ to be the orthogonal projection, where $\hat{\mathcal{P}}_k = span_{1 \leq i \leq k}\{\hat{p}_i\}$. Since $\tilde{\tilde{A}}$ is bounded, the projections $\hat{P}_k$ and $\hat{P}_k^*$ converges to I strongly. Furthermore

$$Ran(I - \hat{P}_k) = Ker\hat{P}_k = (\tilde{\tilde{A}}\hat{\mathcal{P}}_k)^{\perp} = (\tilde{A}\mathcal{P}_k)^{\perp} = KerP_k = Ran(I - P_k).$$

We define the intermediate operators as

$$A_k'' = A_0^{(\gamma)} + \tilde{\tilde{A}}\hat{P}_k.$$

Then we have for $u \in Dom(\tilde{a})$,

$$\tilde{\tilde{a}}(\hat{P}_k u) = \tilde{\tilde{a}}(u - (I - \hat{P}_k)u) \leq \tilde{\tilde{a}}(u - (I - P_k)u) = \tilde{\tilde{a}}(P_k u) \leq \tilde{a}(P_k u) \leq \tilde{a}(u),$$

since $Ran(I - \hat{P}_k) = Ran(I - P_k)$ and $I - \hat{P}_k$ is orthogonal with respective to $\tilde{\tilde{a}}$, but $I - P_k$ is not.

REMARK. *We note that if $\{p_i\}$ is chosen to be dense with respect to $\tilde{a}(u)$ and $\hat{p}_i$ is defined by $\tilde{\tilde{A}}^{-\frac{1}{2}} \tilde{A}^{-\frac{1}{2}} p_i$, then the set $\{\hat{p}_i\}$ is complete in $Dom(\tilde{\tilde{A}})$ with respect to $\tilde{\tilde{a}}(u)$. But the set of $\{\hat{p}_i\}$ does not produce $Ran(I - \hat{P}_k) = Ran(I - P_k)$. The reason is that since $(\tilde{\tilde{A}}\hat{\mathcal{P}}_k)^{\perp} \neq (\tilde{A}\mathcal{P}_k)^{\perp}$, we may not have*

$$\tilde{\tilde{a}}(u - (I - \hat{P}_k u)) \leq \tilde{\tilde{a}}(u - (I - P_k u)).$$

The above inequality yields that for any i with $\lambda_i(A) < \gamma$,

$$\lambda_i(A_k'') \leq \lambda_i(A_k) \leq \lambda_i(A)$$

so that

$$|\lambda_i(A) - \lambda_j(A_k)| \leq |\lambda_i(A) - \lambda_j(A_k'')|.$$

THEOREM 2.4.4. *If the set of vectors $\{p_i\}$ is dense in $Dom(\tilde{A})$ with respect to the norm $\|\tilde{A}u\|$, then for $j = i, i+1, \ldots, i+m-1$,*

$$|\lambda_i(A) - \lambda_j(A_k)| \leq C \cdot \delta^2_{\tilde{\tilde{A}}\mathcal{U}}(\tilde{A}\mathcal{P}_k)$$

for some constants C independent of k.

PROOF: Let $\hat{Q}_k : \mathcal{H} \longrightarrow span_{1 \leq i \leq k}\{\tilde{\tilde{A}}\hat{p}_i\}$ be the orthogonal projection. Since $\tilde{\tilde{A}}$ is bounded, it follows from Theorem 2.4.2 and Lemma 2.4.3 that we have

$$|\lambda_i(A) - \lambda_j(A_k)| \leq C \cdot \max_{u \in \mathcal{U}, \|u\|=1} \|(I - \hat{Q}_k)\tilde{\tilde{A}}u\|^2.$$

By the same argument as the remark of Theorem 2.4.2, we get

$$\max_{u \in \mathcal{U}} \frac{\|(I - \hat{Q}_k)\tilde{\tilde{A}}u\|}{\|u\|} = O(\delta_{\tilde{\tilde{A}}\mathcal{U}}(\tilde{\tilde{A}}\hat{\mathcal{P}}_k)).$$

Since $\tilde{\tilde{A}}\mathcal{U} = \tilde{A}\mathcal{U}$ and $\tilde{\tilde{A}}\hat{\mathcal{P}}_k = \tilde{A}\mathcal{P}_k$, we have the results.■

REMARK. *With the same conditions as Theorem 2.4.4 has, Beattie and Greenlee obtained a similar result to Theorem 2.4.4 [20]:*

$$|\lambda_i(A) - \lambda_j(A_k)| \leq C \cdot \{\delta^2_{\mathcal{U}^\gamma}(\tilde{A}\mathcal{P}_k) + \delta^2_{\mathcal{U}+\mathcal{U}_0^\gamma}(\tilde{A}\mathcal{P}_k)\}.$$

where $\mathcal{U}^\gamma$ and $\mathcal{U}_0^\gamma$ are the eigenspaces of A and A_0, respectively, corresponding to the eigenvalues less than γ.

2.5 Application to a Schrödinger Operator.

In order to apply the preceding estimates to differential eigenvalue problems, it is convenient to dominate the containment gap of Theorems 2.4.4 in terms of spectral projections of an auxiliary operator B. For this we cite Beattie and Greenlee's papers [19,20]. Let B be a positive definite and self adjoint operator in $\mathcal{H}$ such that $Dom(B) \subset Dom(\tilde{A})$ and $\|\tilde{A}u\| \leq \beta\|Bu\|, \beta \geq 0$, for all $u \in Dom(B)$ with B^{-1} compact. Let

$$0 \leq \mu_1 \leq \mu_2 \leq \cdots \nearrow \infty$$

be the eigenvalues of B with corresponding eigenvectors $\{p_i\}$ orthonormal in $\mathcal{H}$. If these vectors $\{p_i\}$ are employed as the trial vectors to construct the projection operators $\{P_k\}$, then the following estimation is obtained.

THEOREM 2.5.1(BEATTIE AND GREENLEE [19]). *If the eigenspace $\mathcal{U}$ is contained in $Dom(B^\tau)$ with $\tau > 1$, then*

$$\delta_{\tilde{A}\mathcal{U}}(\tilde{A}\mathcal{P}_k) = o(\mu_{k+1}^{1-\tau}), \text{ as } k \longrightarrow \infty,$$

where o is the usual Landau symbol and B^τ denotes the unique positive definite τ^{th} power of B.

Theorem 2.5.1 implies that if $\mathcal{U} \subset Dom(B^\tau)$, then

$$|\lambda_i(A) - \lambda_i(A_k)| = o(\mu_{k+1}^{2-2\tau})$$

$$\|u - E_k u\| = o(\mu_{k+1}^{1-\tau}),$$

as $k \longrightarrow \infty$.

As an example, we experimentally verify the rate of convergence of a differential problem with non-trivial continuous spectrum that was considered in [19,20]. The eigenvalue problem is for a one-dimensional Schrödinger operator with potential defined by

$$q(x) = b(x^2 - a^2)exp(-cx^2),$$

where b and c are positive constants. That is, the operator A is given by

$$Au = -u'' + qu$$

for $u \in H^2(\mathbb{R})$ with the corresponding form,

$$a(u) = \int_{-\infty}^{\infty} (|u'|^2 + q|u|^2)dx,$$

for $u \in H^1(\mathbb{R})$. Let the square well potential q_0 be

$$q_0(x) = \begin{cases} q(0) + \gamma, & -a < x < a \\ \gamma, & \text{otherwise,} \end{cases}$$

where $\gamma < 0$ is so big that all negative eigenvalues of A are less than γ. The negative number γ will be our truncation point. We define the base operator A_0 by

$$A_0 u = -u'' + q_0 u,$$

for $u \in H^2(\mathbb{R})$ with the corresponding form

$$a_0(u) = \int_{-\infty}^{\infty} (|u'|^2 + q_0|u|^2)dx,$$

for $u \in H^1(\mathbb{R})$. The base problem $A_0 u = \lambda u$ is explicitly solvable. In fact, if we consider only the even symmetry class of functions for convenience, the lower spectrum of A_0 consists of simple eigenvalues which are the solutions in λ of

$$\tan(a\sqrt{ba^2 - \gamma + \lambda}) = \sqrt{\frac{\gamma - \lambda}{ba^2 - \gamma + \lambda}}$$

lying in the interval $(\gamma - ba^2, \gamma)$. The lowest point of the essential spectrum of A_0 is given by γ and the number of eigenvalues of A_0 smaller than γ is equal to the biggest integer, say N, smaller than $\frac{a^2\sqrt{b}}{\pi} + 1$. These eigenvalues below γ are labeled as $\lambda_1^0 \leq \lambda_2^0 \leq \cdots \leq \lambda_N^0$. The corresponding (unnormalized) eigenvectors of A_0 are given by

$$\begin{cases} \exp(-a\sqrt{\gamma - \lambda_i^0})\cos(\sqrt{ba^2 + \lambda_i^0 - \gamma}x), & -a < x < a \\ \cos(a\sqrt{ba^2 + \lambda_i^0 - \gamma})\exp(-\sqrt{\gamma - \lambda_i^0}|x|), & \text{otherwise.} \end{cases}$$

For the auxiliary operator B, we take the harmonic oscillator, that is,

$$B = -\frac{d^2}{dx^2} + \alpha^2 x^2,$$

with $Dom(B) = H^2(\mathbb{R}) \cap Dom(x^2)$ Then B is self adjoint, and $\mu_k = \alpha(2k-1)$ for $k = 1, 2, \ldots$ [32]. Moreover $\mathcal{U} \subset Dom(B^\tau)$, for all $\tau > 0$ [20]. It follows that we have an estimation

$$|\lambda_i(A) - \lambda_i(A_k)| = o(k^{-\delta}) \text{ as } k \to \infty, \text{ for all } \delta > 0,$$

which is called infinite order convergence.

For the intermediate problem, we must choose $\{p_i\}$ so that the set becomes dense in $H^2(\mathbb{R})$. A useful choice appears to be the solutions to $Bp = \mu p$. The eigenvalues and eigenvectors are known as

$$p_k(x) = \left(\frac{1}{2^k k!}\sqrt{\frac{\alpha}{\pi}}\right)^{\frac{1}{2}} H_k(\sqrt{\alpha}x)\exp(-\frac{1}{2}\alpha x^2)$$

and $\mu_k = \alpha(2k+1)$ for $k = 0, 1, 2, \ldots$, where $H_k(\xi)$ is the Hermite polynomials satisfying the following recursion:

$$H_k(\xi) = 2\xi H'_{k-1}(\xi)$$

$$H_{k+1}(\xi) = 2\xi H_k(\xi) - 2kH_{k-1}(\xi).$$

Since we restrict ourselves to the even symmetry subspace of $L^2(\mathbb{R})$, we need to consider only the even functions $\{p_{2k}\}$.

Now we consider the intermediate operators

$$A_k = A_0^{(\gamma)} + \tilde{A}P_k$$

where $A_0^{(\gamma)}u = \sum_{l=1}^N \lambda_l^0 \langle u, u_l^0\rangle u_l^0 + \gamma(u - \sum_{l=1}^N \langle u, u_l^0\rangle u_l^0)$ and $\tilde{A} = A - A_0^{(\gamma)}$. Since the space $span\{u_1^0, \ldots, u_N^0\} \oplus span\{\tilde{A}p_1, \ldots, \tilde{A}p_k\}$ reduces the operator A_k, the intermediate problem $A_k u = \lambda u$ produces the matrix equation:

$$\begin{bmatrix} (\langle A_k u_i^0, u_j^0 \rangle) & (\langle A_k u_i^0, \tilde{A} p_j \rangle) \\ (\langle \tilde{A} p_i, A_k u_j^0 \rangle) & (\langle A_k \tilde{A} p_i, \tilde{A} p_j \rangle) \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

$$= \lambda \begin{bmatrix} (\langle u_i^0, u_j^0 \rangle) & (\langle u_i^0, \tilde{A} p_j \rangle) \\ (\langle \tilde{A} p_i, u_j^0 \rangle) & (\langle \tilde{A} p_i, \tilde{A} p_j \rangle) \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

whose rank is $k + N$. We recall that $P_k u = \sum_{i,j=1,k}^{k} \left\langle u, \tilde{A} p_i \right\rangle b_{ij} p_j$ where the matrix (b_{ij}) is the inverse to the Gram matrix $(\left\langle p_i, \tilde{A} p_j \right\rangle)$. Notice then that if we assume the eigenvectors, u_i^0, of A_0 are normalized, the inner products in the above equation are expressed as the following:

$$(\langle A_k u_i^0, u_j^0 \rangle) = \Lambda + B.C.^{-1} B^*$$

$$(\langle A_k u_i^0, \tilde{A} p_j \rangle) = \Lambda B + B.C.^{-1} A$$

$$(\langle A_k \tilde{A} p_i, \tilde{A} p_j \rangle) = B^*(\Lambda - \gamma) B + \gamma A + A C^{-1} A$$

where

$$A = (\langle \tilde{A} p_i, \tilde{A} p_j \rangle), \qquad B = (\langle u_i^0, \tilde{A} p_j \rangle)$$

$$C = (\langle p_i, \tilde{A} p_j \rangle) \quad \text{and } \Lambda = diag(\lambda_i^0).$$

It follows from [16] that the above equation can be represented as a compact equation:

$$\begin{bmatrix} I & B \\ B^* & A \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (\lambda - \gamma) \begin{bmatrix} (\Lambda - \gamma)^{-1} & 0 \\ 0 & C \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

where I is the identity matrix.

The matrices A, B and C may be expressed as

$$A = \langle A p_i, A p_j \rangle - \sum_{l=1}^{N} (\lambda_l^0 - \gamma)[\langle p_i, u_l^0 \rangle \langle u_i^0, A p_j \rangle + \langle p_j, u_l^0 \rangle \langle u_l^0, A p_i \rangle]$$
$$+ \sum_{l=1}^{N} (\lambda_l^{0^2} - \gamma^2) \langle p_i, u_l^0 \rangle \langle p_j, u_l^0 \rangle - 2\gamma \langle A p_i, p_j \rangle + \gamma^2 \langle p_i, p_j \rangle$$

$$B = \langle u_i^0, A p_j \rangle - \lambda_i^0 \langle u_i^0, p_j \rangle$$

$$C = \langle A p_i, p_j \rangle - \sum_{l=1}^{N} (\lambda_l^0 - \gamma) \langle p_i, u_l^0 \rangle \langle u_l^0, p_j \rangle - \gamma \langle p_i, p_j \rangle .$$

Thus the inner products involved may be expressed in terms of the four basic ones

$$\langle u_i^0, p_j \rangle, \quad \langle u_i^0, Ap_j \rangle, \quad \langle Ap_i, Ap_j \rangle \quad \text{and} \quad \langle Ap_i, p_j \rangle.$$

Analytical expressions may be obtained for $\langle Ap_i, p_j \rangle$ and $\langle Ap_i, Ap_j \rangle$. But the inner products $\langle u_i^0, p_j \rangle$ and $\langle u_i^0, Ap_j \rangle$ must be approximated with numerical quadratures. For reasons of economy and precision they are determined from recurrence relations that are derived from the basic three-term recurrence for Hermite polynomials. In the appendix we will express how to compute the basic four inner products in detail using the recurrence relations of Hermite polynomials. The transcendental integral evaluations are reduced to the evaluation of the complementary error function and the quadrature of

$$\int_0^a \cos(\sqrt{ba^2 - \gamma + \lambda_i^0}\, x) \exp(-\frac{x^2}{2})\, dx$$

for $i = 1, 2, \ldots, N$.

Calculations were performed on a Vax 8800 in double precision carrying a unit roundoff$\approx 1.4 \times 10^{-17}$. Numerical quadratures were carried out using Gauss-Kronrod scheme to an estimated relative accuracy of 10^{-14}. The matrix eigenvalue problem was solved using the $\mathcal{QZ}$ method of Moler and Stewart with eigenvalues participating in bounds to a relative accuracy of better than 10^{-11}. An order 30 Rayleigh-Ritz calculation using even-ordered Hermite trial functions were performed to provide complementary upper bounds. The computational results are given in Table 1 and Table 2 and the difference between upper and lower bounds are plotted against intermediate problem order k, on a log-log scale in Figures 1 and 2. We observe that no linear asymptote is apparent for any of the error curves, which is consistent with the predicted infinite order convergence.

Table 1. Radial Schrödinger Equation: $\gamma = 0$

N	λ_1	λ_2	λ_3
Base	-17.764406220	-15.885358216	-12.162742720
5	-16.108567768	-9.0407459021	-3.1574282306
10	-16.108530500	-9.0354830733	-3.0515306038
15	-16.108530475	-9.0354751904	-3.0510018349
Ritz	-16.108530475	-9.0354751845	-3.0510013156

Table 2. Radial Schrödinger Equation: $\gamma = -0.1$

N	λ_1	λ_2	λ_3
Base	-17.864406220	-15.985358216	-12.262742720
5	-16.108571114	-9.0413075976	-3.1734022489
10	-16.108530500	-9.0354832125	-3.0515521939
15	-16.108530475	-9.0354751905	-3.0510018553
Ritz	-16.108530475	-9.0354751845	-3.0510013156

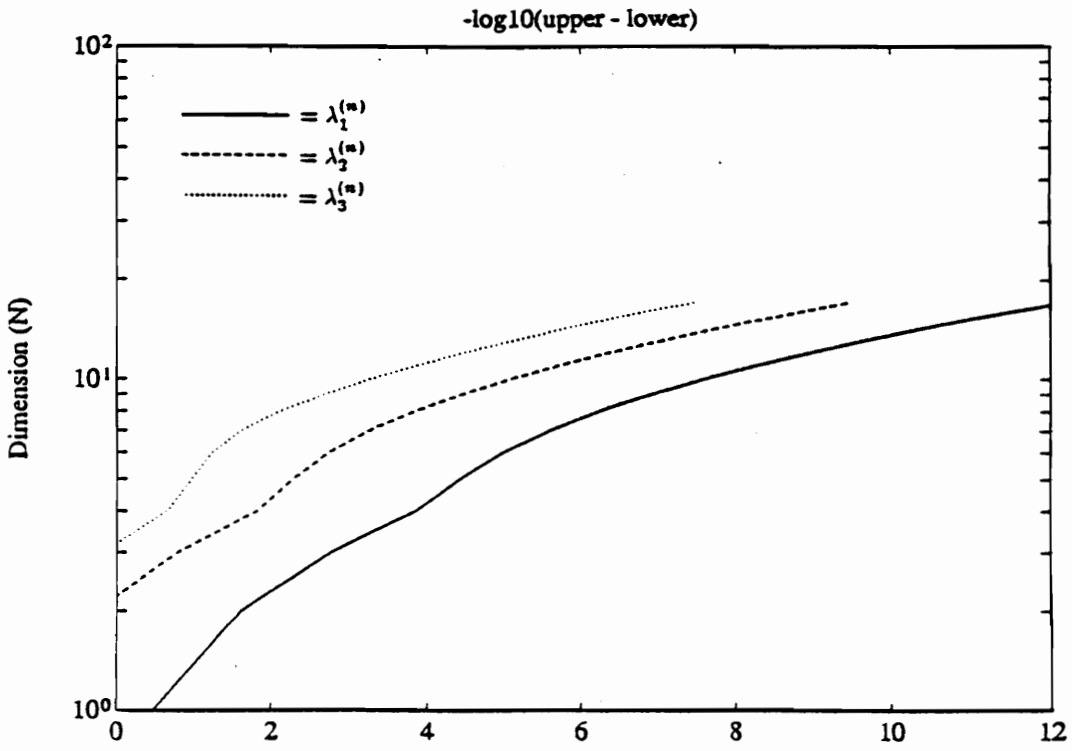

Figure 1: Schrodinger equation; gamma=0

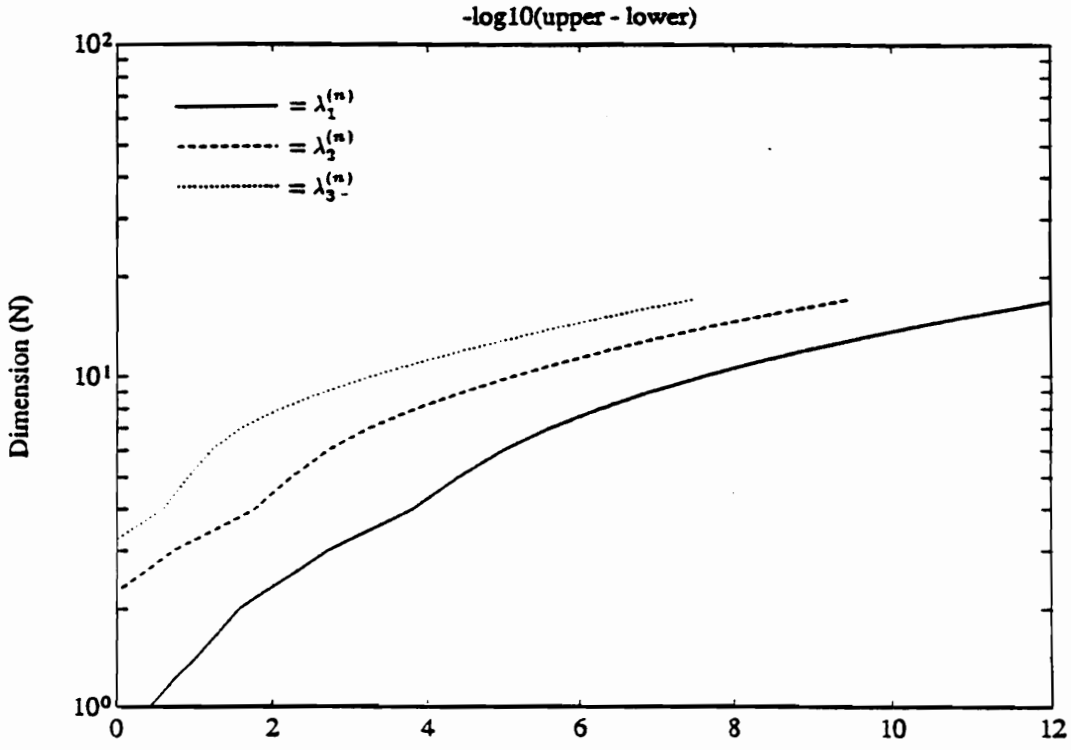

Figure 2: Schrodinger equation; gamma=-0.1

2.6 Appendix.

In this section we express how to compute the four basic inner products

$$\langle Ap_i, p_j\rangle, \langle Ap_i, Ap_j\rangle, \langle u_i^0, p_j\rangle \text{ and } \langle u_i^0, Ap_j\rangle$$

and how to relate with the matrix pencil.

For the computations of $\langle Ap_i, p_j\rangle$ and $\langle Ap_i, Ap_j\rangle$, we denote that

$$P_{ij} = \langle p_i', p_j'\rangle \qquad A1_{ij} = \langle p_i'', p_j''\rangle \qquad A4_{ij} = \left\langle x^2 e^{-cx^2} p_i, e^{-cx^2} p_j\right\rangle$$
$$S_{ij} = \left\langle e^{-cx^2} p_i, p_j\right\rangle \quad A2_{ij} = \left\langle p_i'', x^2 e^{-cx^2} p_j\right\rangle \quad A5_{ij} = \left\langle e^{-cx^2} p_i, e^{-cx^2} p_j\right\rangle$$
$$T_{ij} = \left\langle x^2 e^{-cx^2} p_i, p_j\right\rangle \quad A3_{ij} = \left\langle p_i'', e^{-cx^2} p_j\right\rangle \quad A6_{ij} = \left\langle x^2 e^{-cx^2} p_i, x^2 e^{-cx^2} p_j\right\rangle$$

for $i, j = 0, 1, \ldots, 2(k-1)$. Then the inner products $\langle Ap_i, p_j\rangle$ and $\langle Ap_i, Ap_j\rangle$ can be expressed as

$$\langle Ap_i, p_j\rangle = P_{ij} + bT_{ij} - ba^2 S_{ij}$$
$$\langle Ap_i, Ap_j\rangle = A1_{ij} - b(A2_{ij} + A2_{ji}) + ba^2(A3_{ij} + A3_{ji}) - b^2a^2(A4_{ij} + A4_{ji}) + b^2a^4 A5_{ij} + b^2 A6_{ij}.$$

Thus we need to compute the matrices $P, S, T, A1, A2, A3, A4$ and $A5$. For i and $j \geq 0$ we have

$$P_{ij} = \frac{\alpha}{2}[(\sqrt{ij} + \sqrt{(i+1)(j+1)})\delta_{ij} - \sqrt{i(j+1)}\delta_{i-1,j+1} - \sqrt{j(i+1)}\delta_{i+1,j-1}]$$
$$S_{ij} = \frac{\alpha}{\alpha+c}\sqrt{\frac{j}{i}} S_{i-1,j-1} - \frac{c}{\alpha+c}\sqrt{\frac{i-1}{i}} S_{i-2,j}$$
$$T_{ij} = \frac{1}{2\alpha}[\sqrt{(i+1)(j+1)} S_{i+1,j+1} + \sqrt{i(j+1)} S_{i-1,j+1} + \sqrt{j(i+1)} S_{i+1,j-1} + \sqrt{ij} S_{i-1,j-1}]$$
$$A1_{ij} = \frac{\alpha^2}{4}[\sqrt{(i+1)(i+2)(j+1)(j+2)} + (2i+1)(2j+1) + \sqrt{i(i-1)j(j-1)}\delta_{ij} - (2j+1)\sqrt{(i+1)(i+2)} + (2i+1)\sqrt{j(j-1)}\delta_{i+2,j} + \sqrt{(i+1)(i+2)j(j-1)}\delta_{i+2,j-2}]$$

$$A2_{ij} = \frac{\alpha}{2}[\sqrt{(i+1)(i+2)}T_{i+2,j} - (2i+1)T_{ij} + \sqrt{i(i-1)}T_{i-2,j}]$$
$$A3_{ij} = \frac{\alpha}{2}[\sqrt{(i+1)(i+2)}S_{i+2,j} - (2i+1)S_{ij} + \sqrt{i(i-1)}S_{i-2,j}]$$
$$A4_{ij} = \hat{T}_{ij} \quad \text{and} \quad A5_{ij} = \hat{S}_{ij}$$
$$A6_{ij} = \frac{1}{2\alpha}[\sqrt{(i+1)(i+2)}A4_{i+2,j} + (2i+1)A4_{ij} + \sqrt{i(i-1)}A4_{i-2,j}]$$

with $S_{00} = \sqrt{\frac{\alpha}{\alpha+c}}$. Here $\hat{S}_{ij}$ and $\hat{T}_{ij}$ are the same as S_{ij} and T_{ij} except that c is replaced by $2c$ and we use the symmetry of S_{ij}, i.e. $S_{ij} = S_{ji}$, to get the value S_{0j}.

For the computation of $\langle u_i^0, p_j \rangle$ and $\langle u_i^0, Ap_j \rangle$, we define by

$$\lambda_i = \lambda_i^0 - \gamma, \quad \beta_i = e^{-\sqrt{-\lambda_i}a}, \quad \alpha_i = \cos(\sqrt{ba^2 + \lambda_i}a) \quad \text{and} \quad d_i = \sqrt{ba^2 + \lambda_i},$$

for $i = 1, 2, \ldots, N$. We denote again by

$$A_{ij} = \int_0^a \cos(d_i x) \cdot H_j(\sqrt{\alpha}x) \cdot e^{-\frac{1}{2}\alpha x^2}\, dx$$
$$B_{ij} = \int_a^\infty e^{-\sqrt{-\lambda_i}x} \cdot H_j(\sqrt{\alpha}x) \cdot e^{-\frac{1}{2}\alpha x^2}\, dx,$$
$$A'_{ij} = \int_0^a \cos(d_i x) \cdot H_j(\sqrt{\alpha}x) \cdot e^{-(\frac{1}{2}\alpha+c)x^2}\, dx$$
$$B'_{ij} = \int_a^\infty e^{-\sqrt{-\lambda_i}x} \cdot H_j(\sqrt{\alpha}x) \cdot e^{-(\frac{1}{2}\alpha+c)x^2}\, dx$$
$$A''_{ij} = \int_0^a \cos(d_i x) \cdot H_j(\sqrt{\alpha}x) \cdot x^2 \cdot e^{-(\frac{1}{2}\alpha+c)x^2}\, dx$$
$$B''_{ij} = \int_a^\infty e^{-\sqrt{-\lambda_i}x} \cdot H_j(\sqrt{\alpha}x) \cdot x^2 \cdot e^{-(\frac{1}{2}\alpha+c)x^2}\, dx,$$

for $i = 1, 2, \ldots, N$ and $j = 0, 1, \ldots, 2(k-1)$. Then we have

$$\langle u_i^0, p_j \rangle = 2C_j(\beta_i A_{ij} + \alpha_i B_{ij})$$
$$\langle u_i^0, Ap_j \rangle = 2C_j(d_i^2 \beta_i A_{ij} + \lambda_i \alpha_i B_{ij}) - 2ba^2 C_j(\beta_i A'_{ij} + \alpha_i B'_{ij})$$
$$+ 2b.C._j(\beta_i A''_{ij} + \alpha_i B''_{ij}),$$

where $C_j = (\frac{1}{2^j j!}\sqrt{\frac{\alpha}{\pi}})^{\frac{1}{2}}$. The recurrence formula for $A_{ij}, A'_{ij}, A''_{ij}, B_{ij}, B'_{ij}$ and B''_{ij} are

for $i \geq 1$ and $j \geq 0$,

$$
\begin{aligned}
A_{i,j+2} &= (4j+2-\frac{4}{\alpha}d_i^2)A_{ij} - 4j(j-1)A_{ij-2} + \frac{4}{\alpha}d_i H_j(\sqrt{\alpha}a)e^{-\frac{1}{2}\alpha a^2}\sin(d_i a) \\
&\quad - \frac{1}{\sqrt{\alpha}}[-8jH_{j-1}(0) + 2H_{j+1}(\sqrt{\alpha}a) - 2jH_{j-1}(\sqrt{\alpha}a)e^{-\frac{1}{2}\alpha a^2}\cos(d_i a)] \\
A'_{ij+2} &= \frac{1}{\alpha}(\frac{2\alpha}{\alpha+2c})^2[d_i H_j(\sqrt{\alpha}a)\sin(d_i a)e^{-(\frac{1}{2}\alpha+c)a^2} + \sqrt{\alpha}H_{j+1}(0) \\
&\quad + 2j\sqrt{\alpha}(\frac{1}{2}-\frac{c}{\alpha})H_{j-1}(\sqrt{\alpha}a) - \sqrt{\alpha}(\frac{1}{2}+\frac{c}{\alpha})H_{j+1}(\sqrt{\alpha}a)e^{-(\frac{1}{2}\alpha+c)a^2}\cos(d_i a) \\
&\quad - 4j(j-1)\alpha(\frac{1}{2}-\frac{c}{\alpha})^2 A'_{ij-2} + 2(2j+1)\alpha(\frac{1}{2}-\frac{c}{\alpha})(\frac{1}{2}+\frac{c}{\alpha}) - d_i^2 A'_{ij}] \\
A''_{ij} &= \frac{1}{\alpha}[\frac{1}{4}A'_{i,j+2} + (j+\frac{1}{2})A'_{ij} + j(j-1)A'_{i,j-2}],
\end{aligned}
$$

and for $i \geq 1$ and $j \geq 1$,

$$
\begin{aligned}
B_{ij} &= \frac{2}{\sqrt{\alpha}}[H_{j-1}(\sqrt{\alpha}a)e^{-\frac{1}{2}\alpha a^2 - \sqrt{-\lambda_i}a} + (j-1)\sqrt{\alpha}B_{ij-2} - \sqrt{-\lambda_i}B_{ij-1}] \\
B'_{i,j} &= \frac{2\sqrt{\alpha}}{\alpha+2c}[H_{j-1}(\sqrt{\alpha}a)e^{-(\frac{1}{2}\alpha+c)a^2 - \sqrt{-\lambda_i}a} + (j-1)\frac{\alpha-2c}{\sqrt{\alpha}}B'_{ij-2} - \sqrt{-\lambda_i}B'_{ij-1}] \\
B''_{ij} &= \frac{1}{(\alpha+2c)^2}[(a(\alpha+2c) - \sqrt{-\lambda_i})H_j(\sqrt{\alpha}a) + 2j\sqrt{\alpha}H_{j-1}(\sqrt{\alpha}a) \\
&\quad e^{-(\frac{1}{2}\alpha+c)a^2 - \sqrt{-\lambda_i}a} + (\alpha+2c-\lambda_i)B'_{ij} - 4j\sqrt{-\lambda_i\alpha}B'_{ij-1} \\
&\quad + 4j(j-1)\alpha B'_{ij-2}]
\end{aligned}
$$

with

$$
\begin{aligned}
A_{i0} &= \int_0^a \cos(d_i x)e^{-\frac{1}{2}\alpha x^2}\,dx \\
A'_{i0} &= \int_0^a \cos(d_i x)e^{-(\frac{1}{2}\alpha+c)x^2}\,dx \\
B_{i0} &= \sqrt{\frac{2}{\alpha}}e^{-\frac{\lambda_i}{2\alpha}}(\frac{\sqrt{\pi}}{2} - \int_0^{\sqrt{\frac{\alpha}{2}}a + \sqrt{-\frac{\lambda_i}{2\alpha}}} e^{-x^2}\,dx) \\
B'_{i0} &= \sqrt{\frac{2}{\alpha+2c}}e^{-\frac{\lambda_i}{2\alpha+4c}}(\frac{\sqrt{\pi}}{2} - \int_0^{\sqrt{\frac{\alpha+2c}{2}}a + \sqrt{-\frac{\lambda_i}{2\alpha+4c}}} e^{-x^2}\,dx).
\end{aligned}
$$

We note here that it may not be necessary to compute if j is odd because of even symmetry.

CHAPTER 3

A STUDY OF THE EIGENVECTOR FREE METHOD WITH CONVERGENCE BEHAVIOR

3.1 Introduction.

With the method of intermediate eigenvalue problems we may consider the original operator eigenvalue problem as a perturbation of a simpler, resolvable, self-adjoint eigenvalue problem, called a base problem, that gives rough lower bounds. The full perturbation is approximated systematically by related finite-rank perturbations. The associated intermediate eigenvalue estimates are obtained by computing the spectrum of the base operator summed with a positive semi-definite finite rank operator approximating the full perturbation. Intermediate problem methods have some limitations. In practice, they require not only explicit knowledge of reducing spaces and spectrum of the base operator but also special choices for the range space of the approximating finite rank operators. This makes the resulting problem involve dense matrices so that heavy burdens may be imposed on available computational resources. These practical obstructions primarily come from the explicit involvement of the base problem eigenfunctions which are typically supported throughout the problem domain and consequently may be difficult to handle practically. In the case of the Lehmann-Maehly method, even if one uses finite-element trial functions which make the computational matrix banded and well-structured, the method still depends on the knowledge of numbers that are known to separate adjacent eigenvalues of the given operator.

The so–called eigenvecter free method (EVF) which has been developed by Beattie and Goerisch [17] may overcome such problems since it does not need information of eigenvectors of the base problem and permits the effective use of finite-element trial

functions so that it yields final computational matrices which are sparse and well-structured. Moreover it needs only information about separation of the spectrum of the base operator instead of the given operator itself, so it encompasses both the Weinstein–Aronszajn theory and the Lehmann–Maehley's theory in a certain sense.

Bounds are derived ultimately from eigenvalues of generalized symmetric matrix eigenvalue problems. Highly accurate bounds require large matrix order which may make impractical the use of the QZ method [46] which, in spite of great stability, is unable to exploit any existing sparsity in the original coefficient matrices. For large-order problems, the necessity of retaining the sense of these derived bounds in the face of finite-precision arithmetic and finite computing resources leads to the consideration of iterative algorithms having a variational component. Such a component provides, at every step, intermediate results that may be used to deduce rigorous bounds, even if the method terminates prematurely.

In Section 2 we review the eigenvector free method (EVF) of Beattie and Goerisch. With EVF we explore experimentally in Section 3 convergence behavior of two one–dimensional examples which arise from the vibrations of beams. Section 4 deals with how to take advantage of the sparsity of large–order matrix eigenvalue problems as well as how to choose shifts to make the number of iterations small. With these shifts we explore experimentally in Section 5 convergence behavior of a two–dimensional example of vibrational frequencies of a clamped plate on rectangular domains. Section 6 presents convergence behavior of bounds for each problem and contains some remarks on this topic.

3.2 The Eigenvector Free Method of Beattie and Goerisch.

In this section we present a brief description of the EVF method. For more detail, one should refer to [17]. We recall from Chapter 1 that the associated $n \times n$ W-A matrix of the operator $A_{k,n} = A_0 - \delta^2 + B_k \hat{P}_n$ is given by

$$W_{k,n}(\lambda) = [\langle \hat{p}_i + R^0_{\lambda+\delta^2} B_k \hat{p}_i, B_k \hat{p}_j \rangle] \tag{2.1}$$

for $i, j = 1, \ldots, n$. If we let $\mu = \lambda + \delta^2$ and introduce the change of variable $q_i = R^0_\mu B_k \hat{p}_i$ into the W-A matrix (2.1), we get

$$W_{k,n}(\lambda) = [\langle B_k^{-1}(A_0 - \mu)q_i, (A_0 - \mu)q_j \rangle + \langle q_i, (A_0 - \mu)q_j \rangle]$$

which may be further simplified with the aid of the formula for B_k^{-1} to get

$$\begin{aligned} W_{k,n}(\lambda) =& [\langle q_i, (A_0 - \mu)q_j \rangle + \frac{1}{\mu - \lambda}\{\langle (A_0 - \mu)q_i, (A_0 - \mu)q_j \rangle \\ & - \sum_{l,m=1}^{k} \langle (A_0 - \mu)q_i, T^* p_l \rangle c_{lm} \langle T^* p_m, (A_0 - \mu)q_j \rangle\}]. \end{aligned}$$

If we define the matrices as

$$F_1 = [\langle q_i, (A_0 - \mu)q_j \rangle] \in \mathbb{C}^{n \times n}, \quad F_2 = [\langle p_i, p_j \rangle_*] \in \mathbb{C}^{k \times k},$$

$$G_1 = [\langle (A_o - \mu)q_i, (A_0 - \mu)q_j \rangle] \in \mathbb{C}^{n \times n}, \quad G_2 = [\langle T^* p_i, T^* p_j \rangle] \in \mathbb{C}^{k \times k},$$

and

$$H = [\langle (A_0 - \mu)q_i, T^* p_j \rangle] \in \mathbb{C}^{n \times k},$$

then the W-A matrix , $W_{k,n}(\lambda)$, may be compactly expressed as

$$W_{k,n}(\lambda) = F_1 + \frac{1}{\mu - \lambda}\{G_1 - H[(\mu - \lambda)F_2 + G_2]^{-1} H^*\}. \tag{2.2}$$

Based on this W–A matrix, Beattie and Goerisch introduced the following method which is called the eigenvector free method.

THEOREM 3.2.1(BEATTIE AND GOERISCH). *Let μ and r be chosen so that $\lambda_{r-1}^0 < \mu \leq \lambda_r^0$. Suppose that $\{p_i\}_{i=1}^k \subset Dom(T^*)$ and $\{q_i\}_{i=1}^n \subset Dom(A_0)$. If the generalized matrix eigenvalue problem*

$$\begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \zeta \begin{bmatrix} G_1 & H \\ H^* & G_2 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{2.3}$$

has discrete finite eigenvalues ordered as

$$\zeta_1 \leq \zeta_2 \leq \cdots \leq \zeta_l < 0 \leq \zeta_{l+1} \leq \ldots$$

($l = 0$ if all discrete eigenvalues are either nonnegative or infinite), then for each eigenvalue ζ_p with $p \leq l$ we have a corresponding lower bound for an eigenvalue of A,

$$\mu + \frac{1}{\zeta_p} \leq \lambda_{r-d-m(p)} \leq \lambda_{r-d-p} \leq \lambda_{r-p} \tag{2.4}$$

where $m(p) = max\{m|\zeta_m = \zeta_p\}$ and d is the number of negative eigenvalues of $[\mathcal{V}_1^* F_1 \mathcal{V}_1 + \mathcal{V}_2^* F_2 \mathcal{V}_2]$ for $\mathcal{V} = \begin{pmatrix} \mathcal{V}_1 \\ \mathcal{V}_2 \end{pmatrix}$ having columns that form a basis for $ker \begin{bmatrix} G_1 & H \\ H^* & G_2 \end{bmatrix}$.

Notice that the number of negative eigenvalues of (2.3) is less than r because the Gram matrix and F_2 are positive semi–definite and because the number of eigenvalues of A_0 less than μ is at most $r-1$. We then arrive at the following useful result.

COROLLARY 3.2.2. *In addition to the assumptions of Theorem 3.2.1, if $\{p_i\}$ and $\{q_j\}$ are chosen such that $\{(A_0 - \mu)q_i\}_{i=1}^n$ and $\{T^* p_i\}_{i=1}^k$ are jointly linearly independent, then*

$$\mu + \frac{1}{\zeta_p} \leq \lambda_{r-m(p)} \leq \lambda_{r-p}.$$

Eigenvector-Free Method ([17]).

(i) Select trial vectors $\{q_i\}_{i=1}^n \subset Dom(A_0)$ and $\{p_j\}_{j=1}^k \subset Dom(T^*)$.

(ii) Pick a value $\mu \in (\lambda_{r-1}^0, \lambda_r^0]$ for a selected $r > 1$.

(iii) Form and solve the matrix eigenvalue problem defined by (2.3).

(iv) The finite negative discrete eigenvalues computed from (2.3) may each be associated with eigenvalue bounds as given in (2.4).

We note that if $\{q_i\}_{i=1}^n$ and $\{p_j\}_{j=1}^k$ are chosen to have local support as with finite-element trial functions, then the resulting matrices will be sparse and the matrix eigenvalue problem may be efficiently handled using sparse techniques, even for quite large values of n and k. Furthermore, the only need for *a priori* spectral information comes through the selection of μ as a sufficiently good lower bound to λ_r^0 to separate it from λ_{r-1}^0. No eigenvector data for A_0 are necessary nor are exact values for the eigenvalues of A_0 needed for appropriate selection of μ.

We next consider a relation between eigenpairs of the matrix pencil (2.3) and those of intermediate operators $A_{k,n}$. Let u be an eigenvector of $A_{k,n}$ corresponding to an eigenvalue λ. Then λ satisfies the determinant equation of W–A matrix (2.1) and $u = -\sum_{j=1}^n \alpha_j R_{\lambda+\delta^2}^0 B_k \hat{p}_j$ with $\alpha \in ker W_{k,n}(\lambda)$. Since $q_i = R_\mu^0 B_k \hat{p}_i$, we have

$$F_1\alpha = \zeta\{G_1 - H[-\frac{1}{\zeta}F_2 + G_2]^{-1}H^*\}\alpha$$

with $\zeta = \frac{1}{\lambda-\mu}$ and $u = -\sum_{j=1}^n \alpha_j q_j$. If we define

$$\beta = \zeta(F_2 - \zeta G_2)^{-1}H^*\alpha, \tag{2.5}$$

the vector $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ is an eigenvector of (2.3) corresponding to an eigenvalue ζ. Conversely, let $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ be an eigenvector of (2.3) corresponding to an eigenvalue ζ. Then

$$F_1\alpha = \zeta\{G_1 - H[-\frac{1}{\zeta}F_2 + G_2]^{-1}H^*\}\alpha.$$

Therefore, $u = -\sum_{j=1}^n \alpha_j q_j$ is an eigenvector of $A_{k,n}$ corresponding to an eigenvalue $\lambda = \mu + \frac{1}{\zeta}$, which leads to the following.

THEOREM 3.2.3. *If* $\begin{pmatrix}\alpha\\ \beta\end{pmatrix}$ *is an eigenvector of the matrix pencil (2.3) with a corresponding eigenvalue* ζ, *then* $-\sum_{j=1}^n \alpha_j q_j$ *is an eigenvector of intermediate operator* $A_{k,n}$ *with a corresponding eigenvalue* $\mu + \frac{1}{\zeta}$. *The converse holds for* β *in (2.5).*

3.3 Application to Beam Problems.

We consider the free vibration of a uniform rotating beam clamped at one end and free at the other. This may be modeled by the differential equation

$$EI\frac{d^4u}{dz^4} - \frac{mC\Omega^2}{2}\frac{d}{dz}(l^2 - z^2)\frac{du}{dz} - 4\pi^2 f^2 mCu = 0, \quad 0 < z < l$$

with boundary conditions

$$u(0) = \frac{du}{dz}(0) = \frac{d^2u}{dz^2}(l) = \frac{d^3u}{dz^3}(l) = 0,$$

where u is the transverse displacement of the beam, I is the moment of inertia of the cross section about the principal axis in the plane of rotation, E is the modulus of elasticity, m is the mass per unit volume, l is the length of the beam, C is the cross-section area, Ω is the angular velocity of rotation, and f is the natural frequency. This problem has been treated previously with other methods in [12,59].

For convenience, we introduce the nondimensional variable $x = \frac{z}{l}$ and write the above differential equation as an eigenvalue problem,

$$\frac{d^4u}{dx^4} - \frac{a^2}{2}\frac{d}{dx}(1 - x^2)\frac{du}{dx} = \lambda u, \quad 0 < x < 1 \tag{3.1}$$

with boundary conditions

$$u(0) = \frac{du}{dx}(0) = \frac{d^2u}{dx^2}(1) = \frac{d^3u}{dx^3}(1) = 0$$

Here the parameter a^2 is proportional to the angular velocity of rotation, $a^2 = \frac{mCl^4\Omega^2}{EI}$, and the eigenvalue λ is related to the natural frequency f by $\frac{4\pi^2 mCl^4 f^2}{EI}$. We denote by A the differential operator of the equation (3.1) and by $Dom(A)$ its domain, i.e.,

$$Au = \frac{d^4u}{dx^4} - \frac{a^2}{2}\frac{d}{dx}(1 - x^2)\frac{du}{dx}$$

and

$$Dom(A) = \{u \in H^4(0,1) | u(0) = \frac{du}{dx}(0) = \frac{d^2u}{dx^2}(1) = \frac{d^3u}{dx^3}(1) = 0\}.$$

The quadratic form associated with the operator A is given by

$$a(u) = \int_0^1 (|\frac{d^2u}{dx^2}|^2 + \frac{a^2}{2}(1-x^2)|\frac{du}{dx}|^2)\, dx$$

with the domain

$$Dom(a) = \{u \in H^2(0,1) |\ u(0) = \frac{du}{dx}(0) = 0\}.$$

If we take the base operator A_0 as

$$A_0 u = -\frac{a^2}{2}\frac{d}{dx}(1-x^2)\frac{du}{dx}$$

with

$$Dom(A_0) = \{u \in H^2(0,1) | u(0) = \lim_{x \to 1^-}(1-x)\frac{du}{dx} = 0\},$$

and the perturbation operator T as

$$Tu = -\frac{d^2u}{dx^2}$$

with

$$Dom(T) = \{u \in H^2(0,1) | u(0) = \frac{du}{dx}(0) = 0\},$$

then the quadratic forms associated with the operators A and A_0 are

$$a(u) = a_0(u) + \|Tu\|^2, \quad a_0(u) = \frac{a^2}{2}\int_0^1 (1-x^2)|\frac{du}{dx}|^2\, dx$$

with $Dom(a_0) = \{u \in H^1(0,1) | u(0) = 0\}$, and the adjoint operator of T is obtained as

$$T^* u = -\frac{d^2u}{dx^2}$$

with

$$Dom(T^*) = \{u \in H^2(0,1) | u(1) = \frac{du}{dx}(1) = 0\}.$$

The eigenvalues of A_0 are easily found by

$$\lambda_l^0 = a^2 l(2l-1), \quad for\ l = 1, 2, 3, \ldots.$$

It follows that for a given function $u \in Dom(a)$, the quadratic forms a and a_0 satisfy the inequality

$$a_0(u) \leq a(u).$$

The eigenvalues associated with these quadratic forms thus satisfy the inequality

$$\lambda_\nu^0 \leq \lambda_\nu, \quad \text{for } \nu = 1, 2, 3, \ldots.$$

Define a uniform mesh on $[0, 1]$ with a mesh size $h = \frac{1}{N}$. Furthermore, define cubic spline functions on this mesh, $B_i(x)$, centered at $x_i = ih$ for $i = -1, 0, 1, \ldots, N+1$ so that

$$B_i(x_i) = 1, \quad B_i(x_{i\pm1}) = \frac{1}{4} \quad \text{and} \quad B_i(x_{i\pm2}) = 0.$$

In order to take the projecting vectors $\{q_i\}$ and $\{p_j\}$ within $Dom(A_0)$ and $Dom(T^*)$ respectively, we define them by

$$q_1 = B_0 - 4B_{-1}, \quad q_2 = B_0 - 4B_1, \quad q_i = B_{i-1}, \quad \text{for } i = 3, \ldots, N+2$$

and

$$p_j = B_{j-2}, \quad \text{for } j = 1, \ldots, N, \quad p_{N+1} = B_{N-1} - \frac{1}{2}B_N + B_{N+1}.$$

This provides an $(N+2, N+1)$-order problem. Since the order is dependent only on the mesh size, the eigenvalue estimates will be denoted by $\lambda_\nu^{(N)}$. The elements of

matrices F_1, F_2, G_1, G_2 and H of (2.3) are given by the inner products:

$$
\begin{aligned}
F_1^{ij} &= \frac{a^2}{2}\Big(\int_0^1 q_i' \cdot q_j'\,dx - \int_0^1 q_i' \cdot x^2 q_j'\,dx\Big) - \mu \int_0^1 q_i \cdot q_j\,dx,\\
F_2^{ij} &= \int_0^1 p_i \cdot p_j\,dx,\\
G_1^{ij} &= \frac{a^4}{4}\Big(\int_0^1 q_i'' \cdot q_j''\,dx - 2\int_0^1 q_i'' \cdot x^2 q_j''\,dx + \int_0^1 q_i'' \cdot x^4 q_j''\,dx - 2\int_0^1 q_i'' \cdot x q_j'\,dx,\\
&\quad + 2\int_0^1 q_i'' \cdot x^3 q_j'\,dx - 2\int_0^1 q_i' \cdot x q_j''\,dx + 2\int_0^1 q_i' \cdot x^3 q_j''\,dx + 4\int_0^1 q_i' \cdot x^2 q_j'\,dx,\\
&\quad - a^2\mu\Big(\int_0^1 q_i' \cdot q_j'\,dx - \int_0^1 q_i' \cdot x^2 q_j'\,dx\Big) + \mu^2 \int_0^1 q_i \cdot q_j\,dx,\\
G_2^{ij} &= \int_0^1 p_i'' \cdot p_j''\,dx,\\
H^{ij} &= \frac{a^2}{2}\Big(\int_0^1 q_i'' \cdot p_j''\,dx - 2\int_0^1 a q_i' \cdot p_j''\,dx - \int_0^1 x^2 q_i'' \cdot p_j''\,dx\Big) - \mu \int_0^1 q_i' \cdot p_j'\,dx.
\end{aligned}
$$

For the upper bounds the basis functions are chosen as

$$\phi_1 = B_{-1} - \frac{1}{2}B_0 + B_1 \quad \text{and} \quad \phi_i = B_i, \quad \text{for } i = 2, \ldots, N+1$$

to satisfy the boundary conditions. Upper bounds to the eigenvalues of the rotating beam are obtained as the eigenvalues λ of $(N+1)$-st order symmetric generalized algebraic eigenvalue problem,

$$(\langle A_0\phi_i, \phi_j\rangle + \langle T\phi_i, T\phi_j\rangle)x = \lambda(\langle \phi_i, \phi_j\rangle)x$$

for $i, j = 1, \ldots, N+1$. Here

$$
\begin{aligned}
\langle A_0\phi_i, \phi_j\rangle &= \frac{a^2}{2}\Big(\int_0^1 \phi_i' \cdot \phi_j'\,dx - \int_0^1 \phi_i' \cdot x^2 \phi_j'\,dx\Big),\\
\langle T\phi_i, T\phi_j\rangle &= \int_0^1 \phi_i'' \cdot \phi_j''\,dx \quad \text{and} \quad \langle \phi_i, \phi_j\rangle = \int_0^1 \phi_i \cdot \phi_j\,dx.
\end{aligned}
$$

The result is contained in Table 3. Here the upper bounds come from Rayleigh–Ritz problem of $N = 200$.

Table 3. Clamped Beam Problem (CBP)

$a^2 = 200 \quad \mu = 65000 \quad r = 13$

N	λ_1	λ_2	λ_3	λ_4	λ_5
Base	200.000000	1200.00000	3000.00000	5600.00000	9000.00000
40	233.793442	1771.61203	7305.16111	21716.2501	51978.7670
70	233.793442	1771.61204	7305.16292	21716.3288	51982.8794
100	233.793442	1771.61205	7305.16308	21716.3359	51983.2470
130	233.793442	1771.61205	7305.16311	21716.3373	51983.3221
Ritz	233.793442	1771.61205	7305.16315	21716.3383	51983.3639

$a^2 = 10000 \quad \mu = 910000 \quad r = 7$

N	λ_1	λ_2	λ_3
Base	10000.0000	60000.0000	150000.000
40	10215.0332	61582.3109	159783.051
70	10215.0885	61582.6118	159783.980
100	10215.0905	61582.6228	159784.017
130	10215.0907	61582.6242	159784.022
Ritz	10215.0932	61582.6373	159784.059

N	λ_4	λ_5	λ_6
Base	280000.000	450000.000	660000.000
40	319654.833	557153.539	884797.692
70	319658.096	557170.293	885609.127
100	319658.267	557171.431	885677.091
130	319658.298	557171.643	885690.672
Ritz	319658.387	557171.897	885697.824

Next we consider the free vibration of a uniform rotating beam simply supported at one end and free at the other. The differential equation governing this problem is the same as the clamped case but with a different boundary condition

$$u(0) = \frac{d^2u}{dx^2}(0) = \frac{d^2u}{dx^2}(1) = \frac{d^3u}{dx^3}(1) = 0.$$

Then the base operator A_0 and its domain are the same as in the clamped case, but the domain of a is $Dom(a) = \{u \in H^1(0,1) \,|\, u(0) = 0\}$ and the domain of T^* is

$$Dom(T^*) = \{u \in H^2(0,1) \,|\, u(0) = u(1) = \frac{du}{dx}(1) = 0\}.$$

The projecting vectors $\{q_i\}$ are the same as those of the previous case, but the vectors p_j are defined as

$$p_1 = B_0 - 4B_{-1}, \quad p_2 = B_1 - 4B_{-1}$$

and

$$p_i = B_{i-1}, \quad \text{for } i = 3, \ldots, N-1, \quad p_N = B_{N-1} - \frac{1}{2}B_N + B_{N+1}.$$

This yields $(2N+2)$-order eigenvalue problem.

For the upper bounds the trial functions are chosen to be

$$\phi_1 = B_0 - 4B_{-1}, \quad \phi_2 = B_0 - 4B_1, \phi_i = B_{i-1}, \quad \text{for } i = 3, \ldots, N+2$$

in order to satisfy boundary conditions. Table 4 contains the result. Here the upper bounds come from Rayleigh–Ritz problem of $N = 200$.

Table 4. Simply Supported Beam Problem (SBP)

$a^2 = 5 \quad \mu = 2805 \quad r = 17$

N	λ_1	λ_2	λ_3
Base	5.00000000	30.00000000	75.0000000
40	5.00000000	269.67028572	1853.3382756
70	5.00000000	269.67028886	2585.9300079
100	5.00000000	269.67028915	2585.9333458
130	5.00000000	269.67028921	2585.9340304
Ritz	5.00000013	269.67028923	2585.9344046

$a^2 = 500 \quad \mu = 138000 \quad r = 12$

N	λ_1	λ_2	λ_3
Base	500.0000	3000.000000	7500.0000
40	500.0000	3316.362162	10958.2998
70	500.0000	3316.362184	10958.3014
100	500.0000	3316.362185	10958.3015
130	500.0000	3316.362186	10958.3016
Ritz	500.0000	3316.362186	10958.3016

N	λ_4	λ_5	λ_6
Base	14000.0000	22500.0000	33000.000
40	28159.2102	61332.9650	119134.256
70	28159.2528	61333.7617	119162.440
100	28159.2567	61333.8332	119164.951
130	28159.2575	61333.8478	119165.464
Ritz	28159.2580	61333.8567	119165.745

3.4 Numerical Realization of Eigenvalue Bounds.

In this section we deal with large order matrix eigenvalue problem which comes from the EVF method. For this purpose, we consider the generalized matrix eigenvalue problem

$$\mathcal{A}\mathbf{x} = \Lambda\mathcal{B}\mathbf{x} \tag{4.1}$$

where $\mathcal{A}$ is a symmetric positive-definite matrix and $\mathcal{B}$ is a symmetric positive semi-definite matrix. A variety of approaches exist for computing selected eigenpairs of (4.1) when $\mathcal{A}$ and $\mathcal{B}$ are very large and very sparse. The simplest of these is subspace iteration (cf. [48]). Starting with full rank $S^{(0)} \in \mathbb{R}^{n\times m}(m \ll n)$ and $\theta_i^{(0)} = 1$, iterate

$$\begin{aligned}
\text{Form } \bar{S}^{(k)} &= (\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}S^{(k-1)}diag(\theta_i^{(k-1)}) \\
\text{For } \mathcal{A}^{(k)} &= \bar{S}^{(k)*}\mathcal{A}\bar{S}^{(k)} \text{ and } \mathcal{B}^{(k)} = \bar{S}^{(k)*}\mathcal{B}\bar{S}^{(k)} \\
&\text{Solve } \mathcal{A}^{(k)}G^{(k)} = \mathcal{B}^{(k)}G^{(k)}diag(\theta_i^{(k)}) \\
\text{Form } S^{(k)} &= \bar{S}^{(k)}G^{(k)}
\end{aligned}$$

for $k = 1, 2, \ldots$. Since for each $k \geq 1$, $\{\theta_i^{(k)}\}_{i=1}^m$ are the result of Ritz approximations to (4.1) out of $span(S^{(k)})$, it is clear that $\lambda_i \leq \Lambda_i \leq \theta_i^{(k)}$, independent of σ. While the convergence rate is linear, it will be more rapid to those eigenvalues of (4.1) closest to the shift σ since it depends on the ratio $\max_j \frac{|\theta_j^{(k-1)}|}{|\lambda_i - \sigma|}$.

The slow linear rate of convergence and the necessity for solving m linear systems, where m should be selected larger than the number of eigenvalues actually wanted, at each step ultimately make subspace iteration less appealing than the Lanczos method. Since we are interested in a few eigenvalues, the spectral transformation Lanczos method (STLM) may be useful. Thus if one is willing to live with the expense of an occasional factorization of $\mathcal{A} - \sigma\mathcal{B}$, STLM is often substantially more effective than subspace iteration. Moreover, if $\mathcal{B}$ is singular, STLM does not suffer the same

degradation of the accuracy [47]. Eqn. (4.1) is transformed to

$$(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}\mathbf{x} = \frac{1}{\Lambda - \sigma}\mathbf{x}. \tag{4.2}$$

For convenience, let $\mathcal{M} = (\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$. All eigenvectors of (4.1) corresponding to finite eigenvalues are also eigenvectors of $\mathcal{M}$, and they lie in the range of $\mathcal{M}$. The semi–inner product induced by $\mathcal{B}$ is a true inner product on the range of $\mathcal{M}$, and also the eigenvalue problem (4.2) is self–adjoint with respect to this inner product even though the problem is not symmetric [47]. STLM requires calculating the action of $\mathcal{M}$ on a vector of the range of $\mathcal{M}$ at each iteration step. It constructs a symmetric tridiagonal matrix, $T_j \in \mathbb{R}^{j\times j}$, in the course of j iteration steps, whose eigenvalues approximate those of (4.2). A set of Lanczos vectors $\{\mathbf{q}_i\}_{i=1}^{j}$ that form a $\mathcal{B}$-orthogonal basis for the order j Krylov subspace is generated by $\mathcal{M}$ and $\mathbf{q}_1$. In floating point arithmetic, $\mathcal{B}$-orthogonality is volatile and expensive to maintain, but so long as the $\{\mathbf{q}_i\}_{i=1}^{j}$ are kept robustly independent ($\mathcal{B}$-"semiorthogonal"), one can guarantee up to terms on the order of the machine precision that T_j is the Rayleigh-Ritz restriction of (4.2) to $span(\{\mathbf{q}_i\}_{i=1}^{j})$ with respect to the $\mathcal{B}$-inner product [49]. The eigenvalues of T_j will be associated with **upper** bounds to corresponding eigenvalues of (4.2) and thus it will be associated with **lower** bounds to corresponding eigenvalues of (4.1).

We recall from the EVF method that eigenvalues of the given operator may be associated with lower bounds according to

$$\mu + \frac{1}{\zeta_p} \leq \lambda_{r-p}$$

for each $p = 1, 2, \ldots, r-1$. If $\hat{\zeta}_p \geq \zeta_p$ is an estimate of ζ_p, then $\mu + (1/\hat{\zeta}_p) \leq \mu + (1/\zeta_p) \leq \lambda_{r-p}$. Hence we seek **upper** bounds to the negative eigenvalues of (2.3) or (4.1) so as to maintain consistent lower bounds to $\{\lambda_i\}_1^{r-1}$. Practically, if m is such that $\Lambda_m < \mu < \Lambda_{m+1}$, then it makes sense to find lower bounds only to $\{\lambda_i\}_1^m$. In

other words we need only the m biggest negative eigenvalues of $\{\zeta_i\}$ instead of the entire set of negative eigenvalues.

Let $\mathcal{A} = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}$ and $\mathcal{B} = \begin{bmatrix} G_1 & H \\ H^* & G_2 \end{bmatrix}$. Now we consider how to select the shift σ for the equation (4.1). It is desirable to choose a zero shift in order to preserve the sparsity of $\mathcal{A}$. However, small magnitude eigenvalues may need many iterations to get a reasonable accuracy. In our model of the clamped plate problem the negative eigenvalues of $(\mathcal{A}, \mathcal{B})$ have very small magnitudes compared to the extreme positive eigenvalues. Moreover, the number of Lanczos steps required exceeds half of the size of its computational matrix to get a reasonable accuracy. In order to overcome such trouble, it may be possible to take the shift so that the wanted eigenvalue of $\mathcal{M}$ has the biggest magnitude. For this purpose, let m be such that $\Lambda_m < \mu$ and let $p = r - m$. If we take $\sigma = (\Lambda_m - \mu)^{-1}$, then $\zeta_{r-m-1} < \sigma < \zeta_{r-m}$. Without loss of generality we may assume σ is closer to ζ_{r-m} than to ζ_{r-m-1} since Λ_m can be taken to be closer to λ_m than to λ_{m+1}. For clarity we may refer to Figure 3.

If ν_i's are the ordered eigenvalues of $(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$, then we have

$$\zeta_{r-m} = \frac{1}{\nu_S} + \sigma, \ldots, \zeta_{r-1} = \frac{1}{\nu_{S-m+1}} + \sigma,$$

where S is the rank of $\mathcal{B}$. We note that the eigenvalue ν_S has the biggest magnitude. Since we only need few extreme eigenvalues, $\nu_S, \ldots, \nu_{S-m+1}$, of $(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$, the Lanczos method is expected to be quite efficient. Moreover, if $\mathcal{A}$ and $\mathcal{B}$ are large and sparse, we can efficiently reduce the storage for $\mathcal{A}$ and $\mathcal{B}$ as storing only their nonzero entries because the STLM requires calculating the action of $(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$ on a vector at each iteration step, even if it needs additional storage for factorization of $(\mathcal{A} - \mathcal{B})$.

Since we seek upper bounds to the negative eigenvalues of (2.3) or (4.1), we have to find lower bounds to the corresponding eigenvalues of $(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$. It is appropriate to comment here that the modifications to Rutishauser's subspace iteration *ritzit*

(cf. [48]) that extend its applicability to (4.1) are straight-forward, but the resulting eigenvalue estimates are **lower** bounds to Λ_i of (4.1). Hence it is impossible to directly deduce upper bounds to λ_i. Remarkably, it can be recovered with a rank-one modification of T_j and so regain the sense of derived bounds for λ_i. We give a brief description. For more detail, one may refer to [47]. Let $T_j = Q_j^* \mathcal{B}(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}Q_j$ be the tridiagonal matrix in STLM and define $W_j = Q_j^*(\mathcal{A} - \sigma\mathcal{B})Q_j$, where Q_j is a matrix whose columns are Lanczos vectors. Then the eigenvalues of W_j are the Ritz value approximations to $\zeta_i - \sigma$ and thus the eigenvalues of W_j^{-1} are lower bounds to ν_i which we want. Moreover, the matrix W_j^{-1} differs from T_j only in the last diagonal entry. Hence we easily modify the Lanczos algorithm for our goal. The following is a modified algorithm with full reorthogonalization.

Set $q_0 = 0$ and take $r_1 \in ran(\mathcal{M})$ and let $\beta_1 = \|r_1\|$.

For $j = 1, \ldots,$ maxit, do

$q_j \leftarrow \dfrac{r_j}{\beta_j}$ (normalization)

$\alpha_j \leftarrow q_j^t \mathcal{B}\mathcal{M} q_j$

if $j = 1$; $\omega_1 \leftarrow q_1^t(\mathcal{A} - \sigma\mathcal{B})q_1$ and $\mu_1 \leftarrow \dfrac{1}{\omega_1} - \alpha_1$

else; $\mu_j \leftarrow -\alpha_j - \dfrac{\beta_j^2}{\mu_{j-1}}$

$r \leftarrow (\mathcal{M} - \alpha_j)q_j - \beta_j q_{j-1}$

$r_{j+1} \leftarrow r - \sum_{i=1}^{j} q_i(q_i^t \mathcal{B} r)$ (orthogonalization)

$\beta_{j+1} \leftarrow (r_{j+1}^t \mathcal{B} r_{j+1})^{\frac{1}{2}}$ (norm of r_{j+1} with respect to $\mathcal{B}$)

$\alpha_{\text{maxit}} \leftarrow \alpha_{\text{maxit}} + \mu_{\text{maxit}}$.

We next consider that the mapping from the eigenvalues of (2.3) to the final lower bounds to λ_i introduces potential error magnification. The relative error in a bound

to λ_{r-p} caused by approximating ζ_p with $\hat{\zeta}_p > \zeta_p$ may be expressed as

$$\begin{pmatrix} \text{relative error in the} \\ \text{estimate to } \lambda_{r-p} \end{pmatrix} = \left(\frac{1}{\mu|\zeta_p| - 1}\right) \frac{|\hat{\zeta}_p - \zeta_p|}{|\hat{\zeta}_p|}.$$

The coefficient $1/(\mu|\zeta_p| - 1)$ may be assumed to be positive without loss of generality since all eigenvalues $-1/\mu \leq \zeta_p < 0$ produce nonpositive lower bounds to λ_{r-p} (which are already known to be positive). The error in ζ_p will be either magnified or diminished depending on whether $\zeta_p > -2/\mu$ or $\zeta_p < -2/\mu$.

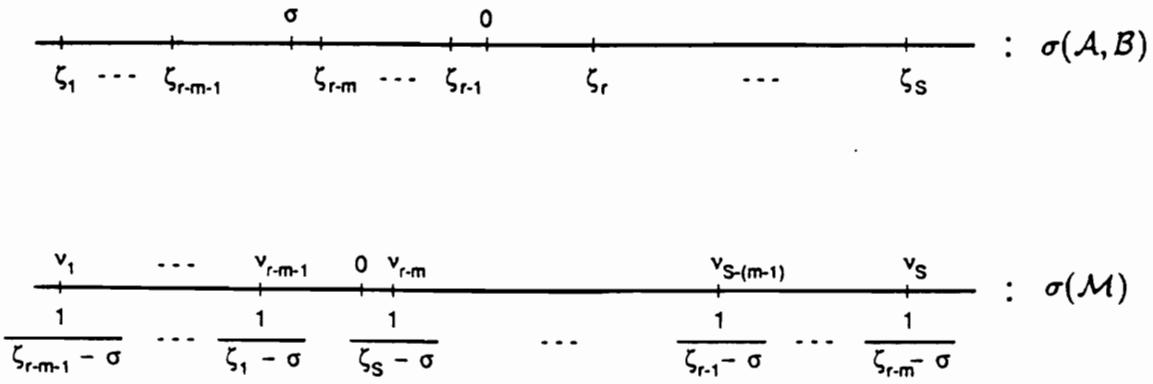

Figure 3 : Relation Between Spectra

3.5 Application to a Clamped Plate Problem.

In this section we give rigorous upper and lower bounds to vibrations of uniform clamped plates on a rectangular domain. The estimation of these vibrations has been treated previously in [14,73]. The lower bounds are obtained by EVF using bicubic spline functions as trial functions, while the upper bounds are obtained by the finite element method using the same trial functions.

Let Ω denote the open rectangle $(-\frac{a}{2}, \frac{a}{2}) \times (-\frac{b}{2}, \frac{b}{2})$ in $\mathbb{R}^2$. Consider the following simple model of vibration of a clamped plate:

$$\Delta^2 u = \lambda u \text{ on } \Omega \quad \text{with} \quad u = \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega.$$

That is, the operator A is defined on a core of $C_0^\infty(\Omega) \subset L^2(\Omega) = \mathcal{H}$ by

$$Au = \frac{\partial^4 u}{\partial^4 x} + \frac{\partial^4 u}{\partial^4 y} + 2\frac{\partial^4 u}{\partial x^2 \partial y^2} \quad \text{with} \quad u = \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega.$$

We now define a base operator A_0 on a core of $C_0^\infty(\Omega) \subset L^2(\Omega)$ by

$$A_0 u = 2\frac{\partial^4 u}{\partial x^2 \partial y^2} \quad \text{with} \quad u = 0 \text{ on } \partial\Omega$$

and T on a core of $C_0^\infty(\Omega) \subset L^2(\Omega)$ into $L^2(\Omega) \times L^2(\Omega) = \mathcal{H}_*$ by

$$Tu = \left\{ \frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial y^2} \right\} \quad \text{with} \quad u = \frac{\partial u}{\partial n} = 0 \text{ on } \partial\Omega.$$

The adjoint operator T^* of T is then defined on sufficiently smooth functions of $L^2(\Omega) \times L^2(\Omega)$ by

$$T^*(v, w) = -(\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 w}{\partial y^2})$$

with free boundary conditions.

Notice that the region, the differential equation and the boundary conditions, share common properties of symmetry. Thus we can take advantage of this so that

we restrict the problem on the space of functions even with respect to both x-axes and y-axes. Then we need extra boundary conditions of

$$\frac{\partial u}{\partial n} = 0 \text{ on } \{(0,y),(x,0) \,|\, 0 < x < \frac{a}{2} \text{ and } 0 < y < \frac{b}{2}\}.$$

We define $\Omega = (0, \frac{a}{2}) \times (0, \frac{b}{2}), \Gamma_1 = \{(\frac{a}{2}, y), (x, \frac{b}{2}) \mid 0 < x < \frac{a}{2} \text{ and } 0 < y < \frac{b}{2}\}$ and

$$\Gamma_2 = \{(0,y),(x,0) \mid 0 < x < \frac{a}{2} \text{ and } 0 < y < \frac{b}{2}\}.$$

The boundary conditions for A, A_0 and T^* restricted to even–even symmetry class are as follows:

(1) For $A, u = \dfrac{\partial u}{\partial n} = 0$ on Γ_1 and $\dfrac{\partial u}{\partial n} = 0$ on Γ_2

(2) For $A_0, u = 0$ on Γ_1 and $\dfrac{\partial u}{\partial n} = 0$ on Γ_2

(3) For $T^*, \dfrac{\partial u}{\partial n} = 0$ on Γ_2.

The eigenvalues of A_0 with these boundary conditions are easily found to be

$$\frac{2\pi^4}{a^2b^2}(2i-1)^2(2j-1)^2 \text{ for } i,j \geq 1.$$

Now we are in a position to construct approximating vectors for both the EVF and the Rayleigh–Ritz methods. Let $N \times N$ finite-element mesh be overlaid on Ω. Trial functions will be constructed from the associated set of bicubic splines so as to satisfy necessary boundary conditions. Let B_i be cubic spline functions on $[0,1]$ for $i = -1, \ldots, N+1$. For approximating vectors within $Dom(A_0)$, we define

$$\tilde{B}_0 = B_0, \quad \tilde{B}_1 = B_1 + B_{-1}$$

$$\tilde{B}_j = B_j, \quad \text{for } j = 2, \ldots, N-2$$

$$\tilde{B}_{N-1} = 4B_{N-1} - B_N, \quad \tilde{B}_N = 4B_{N+1} - B_N.$$

Then the approximating vectors are defined as $q_{ij}(x,y) = \tilde{B}_i(x)\tilde{B}_j(y)$ for $0 \le i,j \le N$ so that the dimension of the finite element space for $Dom(A_0)$ is $(N+1)^2$. For approximating vectors within $Dom(T^*)$, define

$$\hat{B}_0 = B_0, \quad \hat{B}_1 = B_1 + B_{-1}$$

$$\hat{B}_j = B_j, \quad \text{for } j = 2, \ldots, N+1.$$

Then the approximating vectors are defined as $\{\hat{B}_i(x)B_j(y), 0\}$ and $\{0, B_k(x)\hat{B}_l(y)\}$ for $0 \le i,l \le N+1$ and $-1 \le j,k \le n+1$ so that the dimension is $2(N+2)(N+3)$. Thus we have $n = (N+1)^2$ and $k = 2(N+2)(N+3)$ for the EVF method.

For upper bounds, we define

$$\bar{B}_0 = B_0, \quad \bar{B}_1 = B_1 + B_{-1}$$

$$\bar{B}_j = B_j, \quad \text{for } j = 2, \ldots, N-2$$

$$\bar{B}_{N-1} = B_{N-1} - \frac{1}{2}B_N + B_{N+1}.$$

The approximating vectors for $Dom(a)$ are defined by $\phi_{ij}(x,y) = \bar{B}_i(x)\bar{B}_j(y)$ for $0 \le i,j \le N-1$ so that we have $n = N^2$ for the Rayleigh–Ritz problem.

For the computation of each entry of the matrices of EVF and Rayleigh–Ritz problem, we need not compute all the integrations that come from the inner products of approximating vectors directly. Instead, we need only find 4 local overlap matrices of dimension 4×4 and later compute matrix entries by assembly. For this purpose we denote by S_{-1}, S_0, S_1 and S_2 the cubic spline functions on $[0,1]$ with mesh size of 1. Let S_i' and S_i'' be the first and second derivatives of S_i. Then we have the following local 4×4 matrices:

$\langle S_i, S_j \rangle$			
1/7	129/140	3/7	1/140
129/140	297/35	933/140	3/7
3/7	933/140	297/35	129/140
1/140	3/7	129/140	1/7

$\langle S_i, S_j'' \rangle$			
1.2	−2.1	0.6	0.3
9.9	−13.2	−3.3	6.6
6.6	−3.3	−13.2	9.9
0.3	0.6	−2.1	1.2

$\langle S_i', S_j' \rangle$			
1.8	2.1	−3.6	−0.3
2.1	10.2	−8.7	−3.6
−3.6	−8.7	10.2	2.1
−0.3	−3.6	2.1	1.8

$\langle S_i'', S_j'' \rangle$			
12	−18	0	6
−18	36	−18	0
0	−18	36	−18
6	0	−18	12

Let B_i's be the cubic spline functions on $[0, \ell]$ with N uniform meshes and let $h = \frac{\ell}{N}$. Then the global $N+3$ by $N+3$ matrices $[\langle B_i, B_j \rangle]$, $[\langle B_i'', B_j'' \rangle]$, $[\langle B_i, B_j'' \rangle]$ and $[\langle B_i', B_j' \rangle]$ are obtained by assembling the corresponding local matrices and multiplying by $h, \frac{1}{h^3}, \frac{1}{h}$ and $\frac{1}{h}$, respectively. Moreover each entry of the matrices $[\langle \tilde{B}_i, \tilde{B}_j \rangle]$, $[\langle \bar{B}_i, \bar{B}_j \rangle]$ and $[\langle \hat{B}_i, \hat{B}_j \rangle]$ with matrices of their derivatives are formed to be a linear combination of each entries of $[\langle B_i, B_j \rangle]$ with matrices of its derivatives. From these, the final matrices F_1, F_2, G_1, G_2, and H are built. The matrix $\mathcal{A} - \sigma\mathcal{B}$ and $\mathcal{B}$ have at most $441N^2 + 196N + 103$ nonzero entries since $\langle B_i, B_j \rangle = 0$ if $|i - j| \geq 4$. If we store only the nonzero entries of the matrix, then the size of storage may be reduced from $O(N^4)$ to $O(N^2)$ even though additional storage for factorization of $(\mathcal{A} - \sigma\mathcal{B})$ is needed. In Tables 5 and 6 we show upper and lower bounds for rectangular clamped plate problem and square clamped plate problem, respectively. Here the upper bounds come from Rayleigh–Ritz problem of $N = 30$.

Table 5: Vibration of a clamped rectangular plate; $a = 2$ and $b = 3$
(even-even symmetry class)

shift(σ)	45.6	276.6	981.2	1299.7
N	λ_1	λ_2	λ_3	λ_4
Base	5.41161632	48.7045469	48.7045469	135.290408
8	45.57898297	276.5527261	980.6751978	1287.381572
12	45.57909607	276.5678153	981.0796336	1297.112053
16	45.57912595	276.5708077	981.1509026	1298.862718
20	45.57913693	276.5717255	981.1703925	1299.314289
Ritz	45.57915548	276.5728822	981.1859568	1299.639980

Table 6: Vibration of a clamped square plate; $a = b = \pi$
(even-even symmetry class)

shift(σ)	13.3	177.8	179.5	497.1
N	λ_1	λ_2	λ_3	λ_4
Base	2.00000000	18.0000000	18.0000000	50.0000000
8	13.29371618	177.7164393	179.4028827	496.1218673
12	13.29375278	177.7352417	179.4238800	496.8199162
16	13.29376216	177.7385483	179.4278173	496.9514904
20	13.29376564	177.7394530	179.4289743	496.9901598
Ritz	13.29377089	177.7403470	179.4302061	497.0222355

STLM was used with a random starting vector and shifts derived from the corresponding Ritz values estimating $\lambda_1, \lambda_2, \lambda_3$, and λ_4. For simplicity, full reorthogonalization was used. The sparse LU factorization needed by STLM was performed with the Harwell subroutine MA28. Calculations were performed on a Vax 3800 in double precision. The single biggest eigenvalue of $(\mathcal{A} - \sigma\mathcal{B})^{-1}\mathcal{B}$ stabilized to full machine accuracy within 2 Lanczos steps, independent of N. It should be noted again that if we use zero shift, the number of Lanczos steps required exceeds half of the size of its computational matrix, i.e. $(N+1)^2 + 2(N+2)(N+3)$, to get the same accuracy as nonzero shift has.

3.6 Concluding Remarks.

We note that the inner products of cubic spline functions B_i and B_j vanish if the difference between i and j is greater than or equal to 4(i.e.,$|i-j| \geq 4$). The inner product matrices F_1, F_2, G_1, G_2 and H have a full-band width of 7. The (i,j) entry of each matrix is expressible as the sum of $4-|i-j|$ integrals of polynomials of degree 6 or less over some,up to 4, consecutive subintervals $[x_k, x_{k+1}]$ with $x_k = \frac{k}{n}$ and $0 \leq k \leq n-1$. These integrals may be computed analytically in principle but this may be highly tedious. Since each integrand is a polynomial of degree no greater than 6, a Gauss quadrature rule with 4-points is adequate to compute exactly each subinterval integration.

In the case of the simply supported beam problem, each Gram matrix we have encountered is always positive definite while the clamped beam problem is singular. Both problems have the parameter $d = 0$ and $m(p) = p$ which implies the negative eigenvalue of the computational eigenvalue problem is simple.

In the case of simply supported beam problem, the smallest eigenvalue of A is the same as that of A_0 so that we don't actually need to compute it because the lower bound is between those of A_0 and of A. Thus convergence is moot in this case.

In the case of clamped plate problem, we need a couple of Lanczos steps (exactly two steps) to get better than 10 digit accuracy so that it does not make sense to use reorthogonalization. In our model problem we have used Rayleigh–Ritz values when taking shifts. But if we have *a priori* knowledge of separation of eigenvalues of the given operator, we don't need to find upper bounds before computing lower bounds. On the other hand, if we have some information on eigenvectors of the matrix pencil (2.3) corresponding to the wanted eigenvalues, it might be desirable to try to use zero shift in STLM. As seen in Section 4, we know the relationship between eigenvectors of the matrix pencil and those of the intermediate operators $A_{k,n}$, and we may think

that the eigenvectors of intermediate eigenvalue problems and those of Rayleigh–Ritz problem are close in a sense.

Since the EVF has been developed very recently, it still has open questions. First, we don't know analytically how fast (or slowly) the bounds converge as the N increases. From Figures 4, 5, 6 and 7, we may predict its behavior is like $O(N^{-\alpha})$ for some positive α. The rate suggested from CBP, SBP and CPP is always $O(N^{-4})$, which is the same as the rate obtained from the Rayleigh–Ritz method using the same trial functions. Second, we don't know how sensitive the bounds are depending on the choice of μ. As seen in Table 7, the bounds seem to converge to some value less than the corresponding upper bounds for fixed N as increasing μ,which make us have a conjecture that there is a close relationship between μ and the convergence of bounds. In Table 7 we have used $\sigma = 45.6$ as a shift fixed (i.e. $m = 1$ in Section 4 is used).

Table 7: Bounds depending on μ

$N \setminus \mu$	48.7	438.3	1217.6	1563.9
6	45.56417025	45.57833434	45.57873757	45.57881867
10	45.57510808	45.57894476	45.57904766	45.57906020
14	45.57737952	45.57906744	45.57910952	45.57911501
18	45.57845109	45.57910860	45.57912984	45.57913276
20	45.57845649	45.57911893	45.57913488	45.57913686

Table 7 suggest us to take μ as big as possible to get high accuracy for the fixed N.

Rayleigh-Ritz calculations of order 200 and 30 were performed to provide complementary upper-bounds for the beam problem and plate problems, respectively. We note that as the parameter a in beam problems gets larger and larger, we need smaller N for the mesh to get the same accuracy. The reason is that if a gets bigger and bigger, the eigenvalues of the base operator A_0 get closer to those of the given operator A so that small value of r is enough to get some lower bounds, directly making the dimension of coefficient matrices smaller. The relative differences between upper

bounds and lower bounds (i.e., $\frac{\text{upper}-\text{lower}}{\text{upper}}$ for a fixed upper bound and $\frac{\text{upper}-\text{lower}}{\text{lower}}$ for a fixed lower bound) are plotted against the number, N, of mesh element on a log–log scale in Figures 4, 5, 6, and 7. Linear asymptotes are evident in each case. For a few large values of N, a slight deterioration in convergence rate occurs which apparently is an artifact of insufficiently accurate Ritz calculation. The graphs are all drawn using MATLAB on a VAXstation 3800.

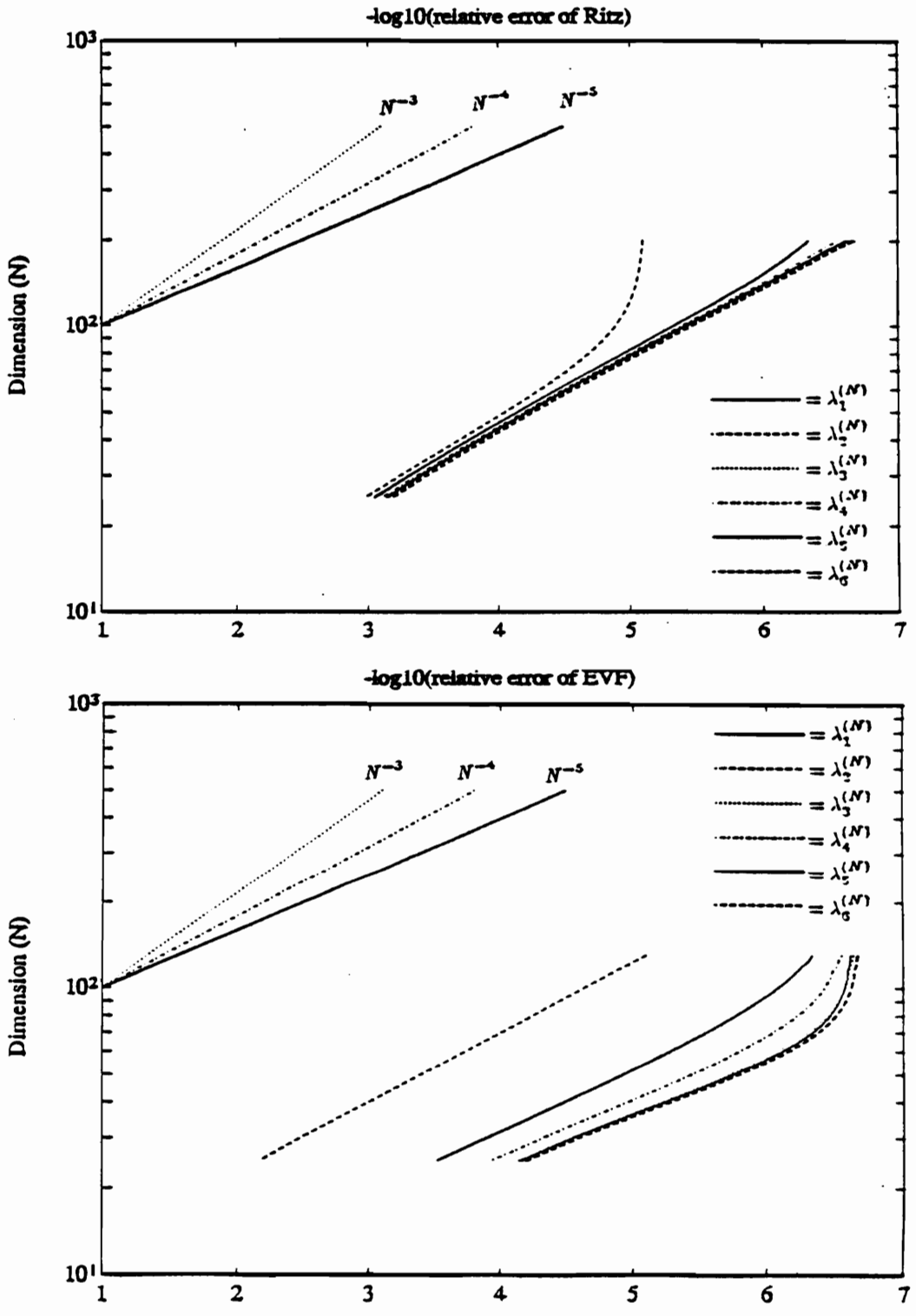

Figure 4: Error Behavior for CBP

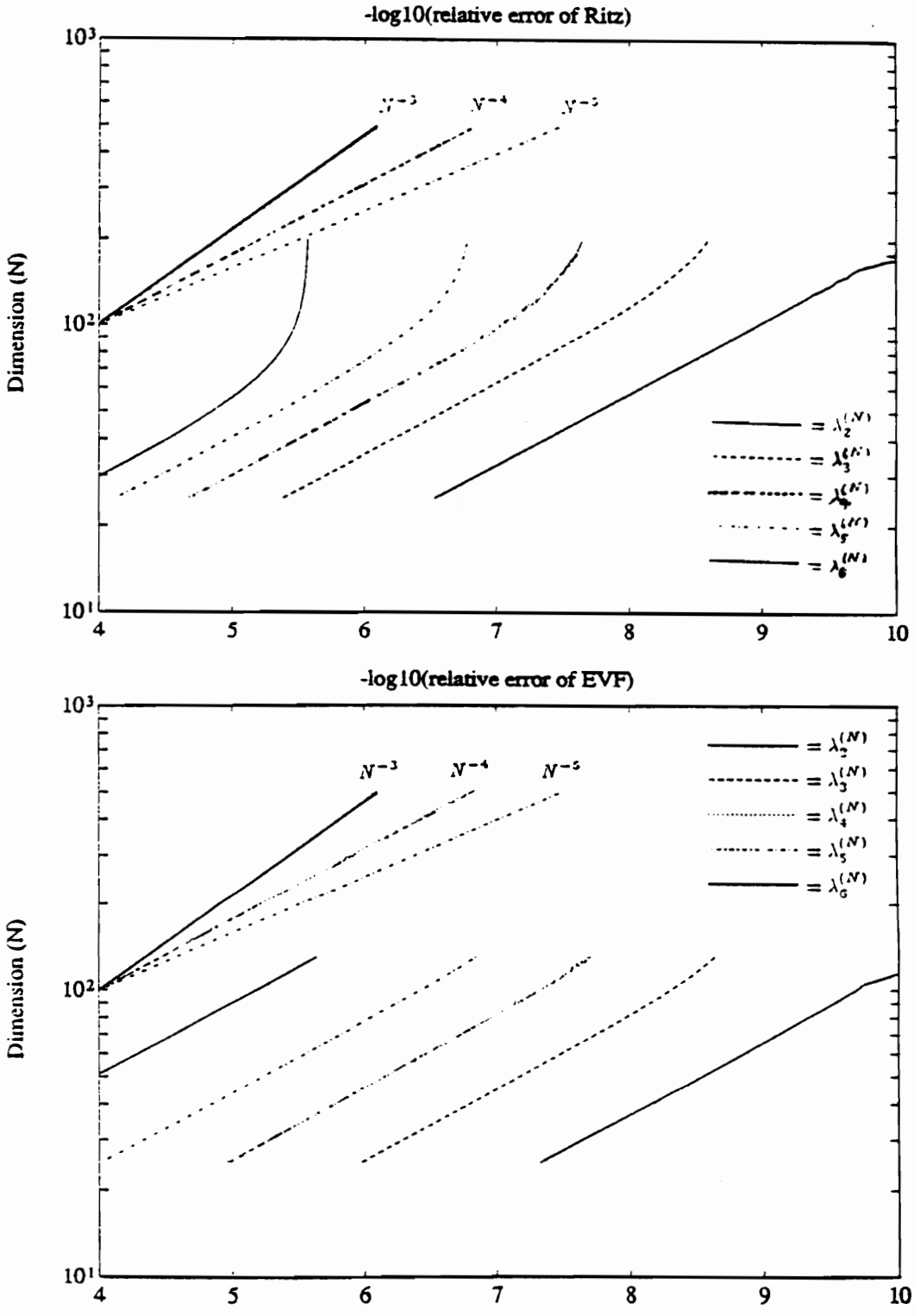

Figure 5: Error Behavior for SBP

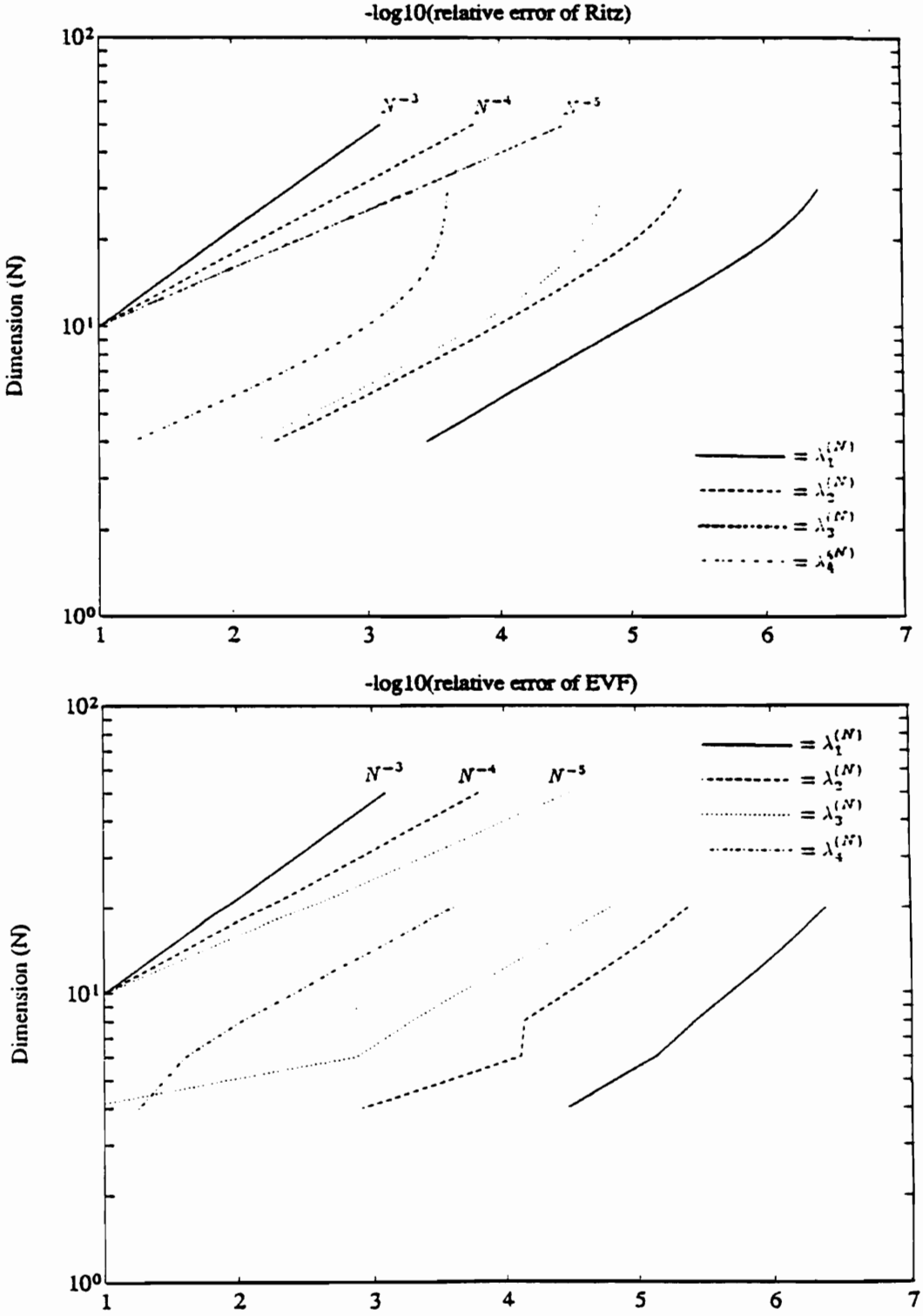

Figure 6: Error Behavior for Rectangular CPP

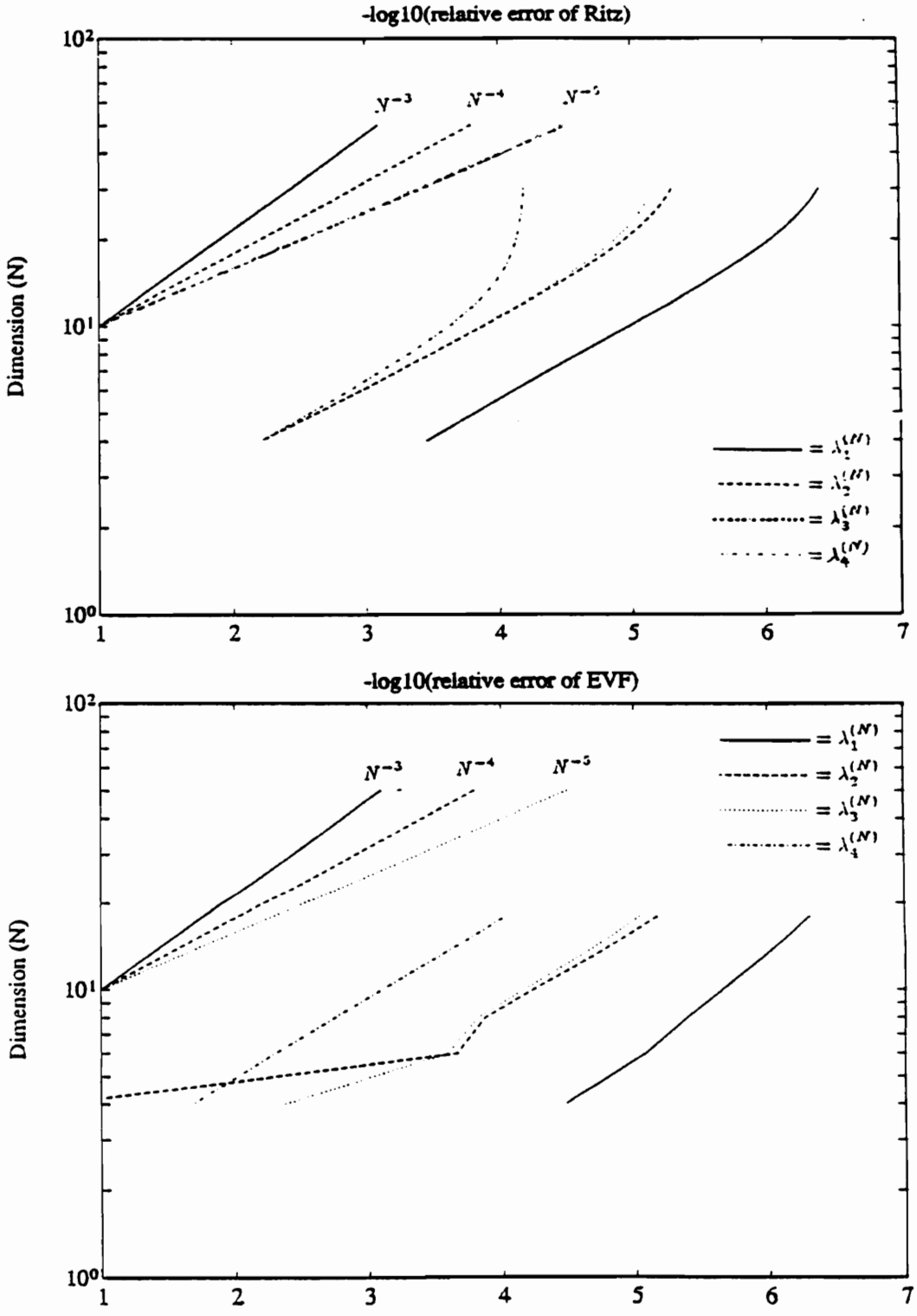

Figure 7: Error Behavior for Square CPP

REFERENCES

1. N. Aronszajn, The Rayleigh-Ritz and A. Weinstein methods for approximation of eigenvalues I, II, Proc. Nat. Acad. Sci., U.S.A., 34, 474–480, 594–601, 1948
2. N. Aronszajn, Approximation methods for eigenvalues of completely continuous symmetric operators. Proc. Symposium on Sectral Theory and Differential Problems, Oklahoma A&M College, Stillwater, 179–202, 1951
3. N. Aronszajn and A. Weinstein, Sur la convergence d'un procédé variationnel d'approximation dans la theorie des plaques encastrées, C. R. Accd. Sci. Paris 204, 96–98, 1937
4. N. Aronszajn and A. Weinstein, Existence, convergence and equivalence in the unified theory of eigenvalues of plates and membranes, Proc. Nat. Acad. Sci. U.S.A. 27, 188–191, 1941
5. N. Aronszajn and A. Weinstein, On the unified theory of eigenvalues of plates and membranes, Amer. J. Math. 64, 623–645, 1942
6. I. Babuška and J. Osborn, *Eigenvalue problems*, in Handbook of Numerical Analysis, Amsterdam, North-Holland press, 1991
7. N. Bazley, Lower bounds for eigenvalues with application to the helium atom, Proc. Nat. Acad. Sci., 45, 1959
8. N. Bazley and D. W. Fox, Truncation in the method of intermediate problems for lower bounds to eigenvalues. J. Res. Nat. Bru. Stds. 65B, 105–111, 1961
9. N. Bazley and D. W. Fox, A procedure for estimating eigenvalues, J. Math. and Phys. 3, 469–471, 1962
10. N. Bazley and D. W. Fox, Lower bounds to eigenvalues using operator decompositions of the form B^*B, Arch. Rat. Mech. Anal. 10, 352–360, 1962 .
11. N. Bazley and D. W. Fox, Improvement of bounds to eigenvalues of operators of the form T^*T, J. Res. Nat. Bur. Sta. Sect. B. 68,173–183 , 1964
12. N. Bazley and D. W. Fox, Methods for lower bounds to frequencies of continuous elastic systems. J. Appl. Math. and Phys. , ZAMP 17(1), 1–37, 1966
13. N. Bazley and D. W. Fox, Comparison operators for lower bounds to eigenvalues. J. Reine Angew. Math, 223, 142–149, 1966
14. N. Bazley, D. W. Fox and J. T. Stadter,Upper and lower bounds for the frequencies of rectangular clamped plates, ZAMM, 47, 191–198, 1967

15. C. A. Beattie, Some convergence results for intermediate problems that displace essential spectra, Applied Physics Laboratory, The Johns Hopkins Univ., 1982
16. C .A. Beattie and A. Banach, Rapid resolution of truncated intermediate problems, Eigenvalue Problems in Engineering Science and their Numerical Treatment, ISNM series, Birkhauser, 1986
17. C. Beattie and F. Goerisch, Extensions of Weinstein-Aronszajn and Lehmann-Maehly Methods for Computing Lower Bounds to Eigenvalues of Self-adjoint operators. ICAM Technical Report 90-09-01, (September, 1990). Virginia Polytechnic Institute and State University, Blacksburg.
18. C. Beattie and W. M. Greenlee, Convergence theorems for intermediate problem, Proc. Roy. Soc. Edinburgh Sect. A 100, 107–122, 1985 . Also correction, Proc. Roy. Soc. Edinburgh Sect. A104 349–350, 1986
19. C. Beattie and W. M. Greenlee, Convergence rates for intermediate problems, Manuscripta Math. 59, 209–227, 1987
20. C. Beattie and W. M. Greenlee, Improved convergence rates for intermediate problems, appear in Math. Comp., April, 1992
21. C. Beattie and W. M. Greenlee, Some remarks concerning closure rates for Aronszajn's method, to be appeared in Eigenvalue Problems in Engineering Science and their Numerical Treatment, ISNM series, Birkhauser.
22. H. Behnke, The Determination of Guaranteed Bounds to Eigenvalues with the use of Variational Methods II. In Computer Arithmetic and Self-validating Numerical Methods (ed: Ch. Ullrich) Academic Press (1990)
23. S. K. Berberian, *Notes on Spectral Theory*, D. Van Nostrand Co., 1966
24. G. Birkhoff and G. Fix, Accurate eigenvalue computations for elliptic problems. Proc. Numerical Solution of Field Problems in Continuum Physics. Symp. on Appl. Math., Vol. 2, American Mathematical Society,111–151, 1970
25. W. Börsch-Supan, Comparison of two methods for lower bounds to eigenvalues, J. Math. Phys., 5, 1787-1788, 1964
26. R. D. Brown, Convergence of approximation methods for eigenvalues of completely continuous quadratic forms, Rocky Mountain J. Math. 10, 199–215, 1980
27. R. D. Brown, Variational approximation methods for eigenvalues; convergence theorems, Banach Center Publications, Vol.13, Warsaw,543–558, 1984
28. R. D. Brown, Convergence criteria for Aronszajn's method and for the Bazley-Fox method, Proc. Roy. Soc. Edinburgh, 108A, 91–108, , 1988

29. F. Chatelin, *Spectral Approximation of Linear Operators*, Academic Press, 1983
30. J. B. Conway, *A course in Functional Analysis*, Springer-verlag,, 1985
31. R. Courant, Über die Eigenwerte bei den Differential gleichungen der Mathematischen, Physik Mathematika,7, 1–57 , 1920
32. R. Courant and D. Hilbert, *Methods of mathematical physics*, Vol.1, New York, Interscience, 1953
33. W. G. Faris, *Self-Adjoint Operators*, Springer-Verlag, 1975
34. E. Fischer, Über quadratische formen mit reellen Koeffizienten, Monatsch. Math. phys., 16, 234–249, 1905
35. G. Fix, Orders of convergence of the Rayleigh-Ritz and Weinstein-Bazley methods. Proc. Nat. Acad. Sci. U.S.A. 61, 1219–1223, 1968
36. D. W. Fox, Lower bounds for eigenvalues with displacement of essential spectra, SIAM J. Math. Anal. 3, 4, 617–624, 1972
37. D. W. Fox and W. C. Rheinboldt, Computational methods for determining lower bounds for eigenvalues of operators in Hilbert space, SIAM Rev. 8, 427–462, 1966
38. F. Goerisch and H. Haunhorst, Eigenwertschranken für Eigenwertaufgaben mit partiellen Differential gleichungen. Z. Angew. Math. Mech. 65 (1985) pp. 129-135.
39. S. Goldberg, *Unbounded Linear Operators, Theory and Applications*, New York, Dover, 1966
40. G. H. Golub and C. F. Van Loan,*Matrix Computations*, 2ed, The Johns Hopkins University Press, 1989
41. W. M. Greenlee, Rate of convergence in singular perturbations, Ann. Inst. Fourier, Grenoble 18, 135–191, 1968
42. W. M. Greenlee, On fractional powers of operators in Hilbert space, Acta Sci. Math. , Szeged 36, 55–61, 1974
43. W. M. Greenlee, A convergent variational method of eigenvalue approximation. Arch. Rational Mech. 81, 3 .279–287, 1983
44. W. M. Greenlee, *Approximation of Eigenvalues by Variational Methods*, Communication of the Mathematical Institute, Rijksuniversiteit Utrecht, V10, 1979
45. T. Kato, *Perturbation Theory for Linear Operators.* Springer, 1976
46. C. B. Moler and G. W. Stewart, An Algorithm for Generalized Matrix Eigenvalue Problems. SIAM J. Num. Anal. 10 (1973) pp. 241-256.

47. B. Nour-Omid, B. N. Parlett, T. Ericsson, and P. S. Jensen, How to Implement the Spectral Transformation, Math. Comp. 48 (1987) pp. 663-673.
48. B. N. Parlett, The Symmetric Eigenvalue Problem. (Englewood Clifts: Prentice Hall, 1980).
49. B. N. Parlett, B. Nour-Omid, and Z. S. Liu, How to Maintain Semiorthogonality among Lanczos Vectors. Report PAM-74, Center for Pure and Applied Mathematics, University of California at Berkeley (1988)
50. H. Poincaré, Sur les équations sux dérivées partielles de la physique mathématique, Amer. J. Math. 12,211-294, 1890
51. G. Pólya, *Estimates for Eigenvalues, Studies in Mathematics and Mechanics*, 200-207, Academic Press, New York, 1954
52. L. T. Poznyak, Estimation of the rate of convergence of a variant of the method of intermediate problems. Z. Vycisl. Mat. i. Mat. Fiz. 8,1117–1126, 1968 ; USSR Computational Math. and Math. Phys.8,167–184, 1969
53. L. T. Poznyak, The convergence of the Bazley-Fox process and an estimate of the rate of this convergence. Z. Vycisl. Mat. i. Mat. Fiz. 9,860–872, 1968 ;USSR Computational Math. and Math. Phys. 9,167–184, 1969
54. L. T. Poznyak, The rate of convergence of the Bazley-Fox method of intermediate problems in the generalized eigenvalue problem of the form $Au = \lambda Cu$, Zh. Vycisl. Mat. i. Mat. Fiz. 10,326–339, 1970 ;USSR Computational Math. and Math. Phys.10,56–74, 1970
55. L. Rayleigh, *The Theory of Sound*, 2nd ed.,Dover, New York, 1945
56. M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol.1, Functional Analysis, New York, Academic Press, 1975
57. M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol.2, Fourier Analysis and Self-Adjoint Operators, New York, Academic Press, 1975
58. F. Riesz and B. Sz-Nazy, *Functional Analysis*, New York, Frederick Ungar pub., 1965
59. N. Rubinstein and J. T. Stader, Bounds to bending frequencies of a rotating beam, J. Franklin Institute, v.294, No. 4, 1972
60. B. Simon, Lower semicontinuity of positive quadratic forms, Proc. Roy. Soc. Edinburgh Sect. A, 79, 267–273, 1977
61. B. Simon, A canonical decomposition for quadratic forms with applications to monotone convergence theorem, J. Funct. Anal., 28, 377–385, 1978

62. G. F. Simmons, *Introduction to Topology and Modern Analysis*, Kogakusha, McGraw-Hill, 1963
63. H. Weber, Über die Integration der Partiellen Differentialgleichung, Math. Ann. 1, 1–36, 1869
64. J. Weidmann, *Linear Operators in Hilbert Spaces*, Springer–Verlag, New York, 1980
65. J. Weidmann, Monotone continuity of the spectral resolution and the eigenvalues, Proc. Royal Soc. Edinburgh, 85A, 131-136, 1980
66. H.Weinberger, Error estimation in the Weinstein method for eigenvalues. Proc. Amer. Math. Soc. 3,643-646, 1952
67. H.Weinberger, *A Theory of Lower Bounds for Eigenvalues.* Tech. Note BN-103, Inst. for Fluid Dyn. and Appl. Math. Univ. of Maryland, College Park 1959
68. H. Weinberger, *Variational Methods for Eigenvalues Approximation*, Philadelphia, SIAM, 1974
69. A. Weinstein, On a Minimal Problem in the Theory of Elasticity, J. London Math. Soc. 10, 184-192, 1935
70. A. Weinstein, Sur la Stabilité de Plaques encastrées, Comp. Rend. 200, 107-109, 1935
71. A. Weinstein, On the Symmetries of the Solutions of a Certain Variational Problem, Proc. Cambridge, Philos. Soc. 32, 96-101, 1936
72. A. Weinstein, Études des Spectres des Équations aux Derivées Partilles de la Théorie des Plaques Élastiques, Mémor. Sci. Math. 88, 1937
73. A. Weinstein and W. Stenger, *Methods of Intermediate Problems for Eigenvalues, Theory and Ramifications*, New York, academic Press, 1972
74. H. Weyl, Das Asymtotische Verteilungsgesetz der Eigenwerte Linearer Partieller Differential Gleichungen, mit einer Anwendung auf die Theorie der Hohlraumstrahlung, Ann. Math. 71, 441–479, 1912

VITA

Gyou-Bong Lee was born in Chunchon, Korea on 14 February 1957. He graduated from Sogang University at Seoul in 1979 and received M. S. in Mathematics from Sogang University in 1982 as well as from Virginia Polytechnic Institute and State University in 1988. He received the Ph. D. in Mathematics from Virginia Polytechnic Institute and State University in 1991.