Analysis of a Multi-level Inverse Iteration Procedure for Eigenvalue Problems

Randolph E. Bank

December, 1980

Yale University Technical Report # 199

Analysis of a Multi-level Inverse Iteration Procedure for Eigenvalue Problems

Randolph E. Bank

December, 1980

Yale University Technical Report # 199

YALE UNIVERSITY

DEPARTMENT OF COMPUTER SCIENCE

Abstract

We define and analyze a procedure for computing approximate eigenvalues and eigenfunctions for self-adjoint elliptic operators using a combination of inverse iteration and a multi-level iterative technique. This algorithm achieves the optimal order work estimates typical of multi-level techniques.

Analysis of a Multi-level Inverse Iteration Procedure for Eigenvalue Problems

Randolph E. Bank

December, 1980

Yale University Technical Report # 199

This work was supported in part by the Office of Naval Reasearch under grants N00014-80-C-0645, and N00014-76-C-0277.

Table of Contents

List of Figures

List of Tables

1. Introduction

In this work we consider the solution of the self-adjoint elliptic eigenvalue problem

$$\begin{aligned} -\nabla a \nabla u + b\,u &= \lambda\,u && \text{in } \Omega \subset R^2, \\ u&=0 && \text{on } \partial\Omega_1, \\ \partial u/\partial n &= 0 && \text{on } \partial\Omega - \partial\Omega_1, \end{aligned} \qquad (1.1)$$

where Ω is a polygonal region in R^2, $a \in C^1(\bar{\Omega})$, $b \in C^0(\bar{\Omega})$, with

$$\begin{aligned} &0 < \underline{a} \leq a(x) \leq \bar{a}, \\ &0 \leq b(x) \leq \bar{b}\ , \qquad \text{for } x \in \bar{\Omega}. \end{aligned}$$

When one discretizes the weak form of (1.1) using a finite element discritization one obtains a generalized algebraic eigenvalue problem of the form

$$A\,U = \lambda\,M\,U \quad , \qquad (1.2)$$

where the stiffness matrix A and the mass matrix M are large, sparse, symmetric, and positive definite (A may be only semi-definite).

Some effective ways to solve (1.2) are through the use of inverse iteration and its extensions and generalizations (e.g., the Rayleigh quotient method, block inverse power method and subspace iteration) [10] [11] [12], or more recent procedures based on the Lanczos method [6]. These methods all require the solution of one or more linear systems of the form

$$(A - \mu M) U = B \tag{1.3}$$

in each iteration. In many instances, these systems are solved by sparse direct methods based on Gaussian elimination.

In this work, we analyze a procedure for computing the eigenvector correspnding to the smallest eigenvalue by means of inverse iteration, using a multi-level iterative method to approximately solve the required sets of linear equations. This was done to simplify the analysis as much as possible rather than to advocate any specific algorithm. Our real hope is that this work will suggest that multi-level techniques can be advantageously incorporated into many of the usual algorithms which are used to solve finite element eigenvalue problems. Our approach to the problem was motivated by the work of Blue, Wilson and their coworkers [4], [5] in combining an multi-level code for linear boundary value problems [3] and a Rayleigh quotient method.

When the shift μ is equal to an eigenvalue, it is hard to make quantitative assessment of the success of the inverse iteration procedure since the condition number of $A - \mu M$ becomes infinite. In our analysis, we must bound the shift away from the eigenvalues by a small ammount in order to establish norm estimates for the rate of convergence. However, our convergence proof shows that when this assumption is violated, one should expect the customary rapid convergence of the procedure, for essentially the same reasons as in the case of Gaussian elimination.

As in the case of linear boundary value problems, the application of the multi-level iteration to (1.3) results in an algorithm in which the smallest eigenvalue and corresponding eigenfunction can asymptotically be computed in O(N) operations (N is the order of A and M). Another effective multi-level technique for eigenvalue problems has been proposed by Hackbusch [7]. His approach is somewhat different from ours in that he develops a more self-contained multi-level procedure specifically for eigenvalue problems, as opposed to incorparating multi-level techniques, as linear equations solvers, into existing algorithms. A Rayleigh quotient scheme similar in some respects to ours is described in McCormick [8].

In Section 2, we define terms, establish notation, state our assumptions, and prove some preliminary Lemmas. In Section 3, we define the j-level iteration and analyze its convergence when μ is close to the smallest eigenvalue. In Section 4, we consider an idealized inverse iteration procedure using the j-level iteration to approximately solve the linear systems. In Section 5, we present a numerical example and make some concluding remarks.

2. Preliminaries

We seek a numerical approximation of the weak form of (1.1): find $u \in H^1_E$ and $\lambda \in R$ such that

$$a(u,v) = \lambda\ (u,v) \qquad \text{for all } v \in H^1_E, \tag{2.1}$$

where

$$H^1_E = \{\ v \mid v \in H^1(\Omega)\ ,\ v = 0 \text{ on } \partial\Omega_1\ \}\ ,$$

$$a(u,v) = \int_\Omega a\ \nabla u\ \nabla v + b\ u\ v\ dx\ ,$$

$$(u,v) = \int_\Omega u\ v\ dx\ . \tag{2.2}$$

The problem (2.1) will in general have an infinite sequence of non-decreasing eigenvalues

$$0 \leq \lambda_1 \leq \lambda_2 \ldots$$

and corresponding eigenfunctions ξ_j. Without loss of generality, we assume that λ_1 is positive, and that

$$a(\xi_i,\xi_j) = \lambda_i \, \delta_{ij} \quad ,$$

$$(\xi_i,\xi_j) = \delta_{ij} \quad , \tag{2.3}$$

where δ_{ij} is the Kronecker delta.

In this work we will focus attention on the approximation of λ_1 and ξ_1. In particular, we note that the minimax characterization of λ_1 is given by

$$\lambda_1 = \inf_{\substack{u \neq 0 \\ u \,\varepsilon\, H^1_E}} a(u,u) \,/\, (u,u) \ . \tag{2.4}$$

For non-negative integral k, let $H^k(\Omega)$ be the usual Sobelev space equipped with the norm

$$\|u\|_k^2 = \sum_{|\beta| \leq k} \|D^\beta u\|_0^2$$

$$= \sum_{|\beta| \leq k} (D^\beta u, D^\beta u) \quad .$$

For positive non-integral k, define $H^k(\Omega)$ by interpolation, and for k negative, define $H^k(\Omega)$ as the dual of $H^{-k}(\Omega)$. We define the energy norm, $|||u|||^2 = a(u,u)$, and note that for some positive constant $C = C(a,b,\Omega)$,

$$C^{-1} \|u\|_1^2 \leq a(u,u) \leq C \|u\|_1^2 \quad , \text{ for all } u \in H_E^1 \tag{2.5}$$

We assume a modest ammount of elliptic regularity for the solution $u \in H_E^1$ of the boundary value problem

$$a(u,v) = (f,v) \qquad \text{for all } v \in H_E^1. \tag{2.6}$$

Specifically, we assume there exists $0 < \alpha \leq 1$ such that $u \in H^{1+\alpha}(\Omega)$ provided $f \in H^{\alpha-1}(\Omega)$ and

$$\|u\|_{1+\alpha} \leq C(a,b,\Omega) \|f\|_{\alpha-1} . \tag{2.7}$$

Let $\mathcal{T}_1$ be a triangulation of Ω. We assume $\mathcal{T}_1$ to be both quasi-uniform and shape-regular. Letting h_t denote the diameter of $t \in \mathcal{T}_1$, we set

$$h_1 = \max_{t \in \mathcal{T}_1} h_t .$$

Let $h_t \, d_t$ denote the diameter of the inscribed circle for t. We define the positive constants δ_0 and δ_1 by

$$\delta_0 = \min_{t \in \mathcal{T}_1} d_t \quad ,$$
$$\delta_1 = \min_{t \in \mathcal{T}_1} h_t \, / \, h_1 . \tag{2.8}$$

We inductively construct a nested sequence of triangulations $\mathcal{T}_j$, $j=1,2,\ldots$, starting from $\mathcal{T}_1$ in the usual fashion. For each $t \in \mathcal{T}_{j-1}$, construct four triangles in $\mathcal{T}_j$ by pairwise connecting the midpoints of the edges of t. Each triangulation $\mathcal{T}_j$ will then have the same shape regularity and quasi-uniformity constants, δ_0 and δ_1, as $\mathcal{T}_1$ and will have

$$h_j = \max_{t \in \mathcal{T}_j} h_t = h_1 \, 2^{1-j} \ .$$

With each triangulation $\mathcal{T}_j$, we associate the N_j-dimensional space $\mathcal{M}_j \subset H^1_E$ of C^0-piecewise linear polynomials. Since the triangulations are nested we have the inclusion property

$$\mathcal{M}_j \subset \mathcal{M}_k \, , \qquad 1 \leq j \leq k.$$

Also,

$$N_j \cong 4 \, N_{j-1} \, , \qquad j > 1.$$

We assume that the spaces $\mathcal{M}_j$ satisfy the following standard approximation property [9], [12]: if $u \varepsilon H^s(\Omega)$, $1 \leq s \leq 1 + \alpha$, then there exists $u_j \, \varepsilon \, \mathcal{M}_j$ such that

$$\|u-u_j\|_0 + h_j \, \|u-u_j\|_1 \leq C \, h_j^s \, \|u\|_s \tag{2.9}$$

where $C = C(\delta_0, \delta_1, \Omega)$.

The finite element analogue of (2.1) is: find $u \, \varepsilon \, \mathcal{M}_j$ and $\lambda \, \varepsilon \, R$ such that

$$a(u,v) = \lambda \, (u,v) \qquad \text{for all } v \, \varepsilon \, \mathcal{M}_j \ . \tag{2.10}$$

The problem (2.10) will have N_j real eigenvalues

$$0 < \lambda_{1j} \leq \lambda_{2j} \leq \dots \leq \lambda_{N_j j}$$

and corresponding eigenfunctions ξ_{ij}. We assume without loss of generality

$$a(\xi_{ij}, \xi_{kj}) = \lambda_{ij} \delta_{ik} \quad ,$$

$$(\xi_{ij}, \xi_{kj}) = \delta_{ik} \; . \tag{2.11}$$

The minimax characterization of the smallest eigenvalue λ_{1j} is given by

$$\lambda_{1j} = \min_{\substack{u \, \varepsilon \, \mathcal{M}_j \\ u \neq 0}} a(u,u) \,/\, (u,u) \quad . \tag{2.12}$$

A simple homogeneity argument, coupled with (2.5) and (2.8) shows

$$\lambda_{N_j j} \leq C(a,b,\delta_0,\delta_1,\Omega) \; h_j^{-2} \quad . \tag{2.13}$$

We shall assume that λ_1 is a simple eigenvalue well separated from λ_2. Then, using (2.7), (2.9), and standard finite element error analysis [12], one can show that

$$|||\xi_1 - \xi_{1j}||| \leq \mathcal{X}_0 \, h_j^{\alpha} \, \lambda_1^{\alpha/2},$$

$$\lambda_1 \leq \lambda_{1j} \leq \lambda_1 + \mathcal{X}_1 \, h_j^{2\alpha} \;\; \lambda_1^{1+\alpha} \; . \tag{2.14}$$

If h_1 is sufficiently small λ_{1j} is simple and well separated from λ_{2j} (independent of j). Finally, note that the minimax characterizations (2.4) and (2.11) can be used to show

$$\lambda_1 \leq \lambda_{1k} \leq \lambda_{1j} \; , \qquad 1 \leq j \leq k. \tag{2.15}$$

Let $u \, \varepsilon \, \mathcal{M}_j$. Then

$$u = \sum_{i=1}^{N_j} c_i \, \xi_{ij}.$$

We define the norms $|||u|||_s$, $-2 \leq s \leq 2$, by

$$|||u|||_s^2 = \sum_{i=1}^{N_j} c_i^2 \, \lambda_{ij}^s \ . \tag{2.16}$$

Note $|||u||| = |||u|||_1$, and $\|u\|_0 = |||u|||_0$ for $u \in \mathcal{M}_j$.

The following lemma is proved in [2], [1].

<u>Lemma 2.1</u>: Let $u \in \mathcal{M}_j$. Then there exists $C = C(a,b,\delta_0,\delta_1,\Omega)$ such that, for $0 \leq s \leq 1$,

$$C^{-1} \|u\|_s \leq |||u|||_s \leq C \|u\|_s. \tag{2.17}$$

The following Lemma follows from results in [1].

<u>Lemma 2.2</u>: Assume that (2.7) and (2.9) hold and let $u \in \mathcal{M}_j$. Then

$$\inf_{v \in \mathcal{M}_{j-1}} \|u-v\|_0 \leq C \, h_j^{1+\alpha} \, |||u|||_{1+\alpha} \ , \tag{2.18}$$

where $C = C(a,b,\delta_0,\delta_1,\Omega)$.

Let μ be an approximation of λ_1. Ultimately, μ will be the shift in the inverse iteration procedure.

__Lemma 2.3__: Let $u \in \mathcal{M}_j$ and $\bar{u} \in \mathcal{M}_{j-1}$ satisfy

$$a(u-\bar{u},v) - \mu\,(u-\bar{u},v) = 0 \qquad , \text{ for all } v \in \mathcal{M}_{j-1}. \tag{2.19}$$

If $\mu < \lambda_{1j}$ satisfies

$$\{\, |||\xi_{1j-1}-\xi_{1j}|||^2 - \mu\, |||\xi_{1j-1}-\xi_{1j}|||_0^2 \,\}^{1/2} \le \lambda_{1j} - \mu , \tag{2.20}$$

then

$$|||\bar{u}||| \le C\, |||u||| . \tag{2.21}$$

Proof: Lemma 2.3 states that the projection with respect to the inner product given in (2.19) is stable with respect to the energy norm, provided that μ satisfies (2.20). Let

$$\begin{aligned} \bar{u} &= \beta_1\, \xi_{1j-1} + \sum_{i=2}^{N_{j-1}} \beta_i\, \xi_{ij-1} \\ &= u_1 + u_2 . \end{aligned}$$

Then

$$\begin{aligned} |||u_2|||^2 &= a(u_2,u_2) \\ &\le (1-\mu/\lambda_{2j-1})^{-1}\, \{\, a(u_2,u_2) - \mu\,(u_2,u_2) \,\} \\ &\le (1-\mu/\lambda_{2j-1})^{-1}\, \{\, a(\bar{u},\bar{u}) - \mu\,(\bar{u},\bar{u}) \,\} \\ &= (1-\mu/\lambda_{2j-1})^{-1}\, \{\, a(\bar{u},u) - \mu\,(\bar{u},u) \,\} \\ &\le C\, |||\bar{u}|||\ |||u||| \end{aligned} \tag{2.22}$$

where we have used (2.19) with $v = \bar{u}$. Also,

$$\begin{aligned}
|||u_1||| &= \lambda_{1j-1}^{1/2}(\lambda_{1j-1}-\mu)^{-1} \; | \; a(\bar{u},\xi_{1j-1}) - \mu \; (\bar{u},\xi_{1j-1}) \; | \\
&\leq \lambda_{1j-1}^{1/2}(\lambda_{1j-1}-\mu)^{-1} \; \{ \; | \; a(u,\xi_{1j}) - \mu \; (u,\xi_{1j}) \; | \\
&\quad + | \; a(u,\xi_{1j-1}-\xi_{1j}) - \mu \; (u,\xi_{1j-1}-\xi_{1j}) \; | \; \} \\
&\leq C \; (\lambda_{1j-1}-\mu)^{-1} \; \{ \; (\lambda_{1j}-\mu) + \\
&\quad [\; |||\xi_{1j}-\xi_{1j-1}|||^2 - \mu \; |||\xi_{1j}-\xi_{1j-1}|||_0^2 \;]^{1/2} \; \} \; |||u||| \\
&\leq C \; |||u||| \qquad (2.23)
\end{aligned}$$

where we have used (2.15) and (2.20). The lemma now follows since

$$|||\bar{u}|||^2 = |||u_1|||^2 + |||u_2|||^2.$$

3. The j-level Iteration

Let μ be an approximation of λ_1, to be taken as a shift for the inverse iteration procedure. The algorithm outlined in Section 4 consists of solving a sequence of problems of the form (1.3), or, in finite element notation, find $z \in \mathcal{M}_j$ such that

$$a(z,v) - \mu \; (z,v) = G(v) \qquad \text{for all } v \in \mathcal{M}_j, \qquad (3.1)$$

where $G(v)$ is a linear functional. The exact interpretation of $G(v)$ will vary in different situations.

The j-level scheme we will analyze here is addressed to problems of the form (3.1) and can be defined inductively as follows:

1. If j=1 (3.1) is solved exactly, typically by a direct method.

2. If $j > 1$, one iteration of the j-level scheme takes an initial guess $z_0 \varepsilon \mathcal{M}_j$ to a final guess $z_{m+1} \varepsilon \mathcal{M}_j$ as follows: for $1 \le k \le m$,

$$(z_k - z_{k-1}, v) = (\lambda_{N_j j} - \mu)^{-1} \{G(v) - a(z_{k-1}, v) + \mu \; (z_{k-1}, v)\} \tag{3.2}$$

for all $v \; \varepsilon \mathcal{M}_j$. Let $q \; \varepsilon \mathcal{M}_{j-1}$ be the approximation of $\bar{q} \; \varepsilon \mathcal{M}_{j-1}$ generated by applying p iterations of the j-1 level scheme to the residual equations

$$\begin{aligned} a(\bar{q}, v) - \mu \; (\bar{q}, v) &= G(v) - a(z_m, v) + \mu \; (z_m, v) \\ &= \bar{G}(v) \end{aligned} \tag{3.3}$$

for all $v \; \varepsilon \mathcal{M}_{j-1}$, starting from initial guess zero. Then set

$$z_{m+1} = z_m + q \; . \tag{3.4}$$

This scheme is analagous to the j-level scheme described in [2] and reduces to that scheme in the case $\mu = 0$. As usual, the particular form of the smoothing iteration (3.2) is chosen for convenience, and can be replaced by computationally more attractive iterations as outlined in [2]. In practice, we take $p = 2$, since this choice leads to optimal order work estimates.

The following Theorem is the analogue of Theorem 1 in [2].

<u>Theorem 3.1</u>: Let the assumptions detailed in Section 2 hold and let $p > 1$ be any integer. Let $\mu < \lambda_{1j}$ satisfy

$$\{ \; |||\xi_{1j-1} - \xi_{1j}|||^2 - \mu \; |||\xi_{1j-1} - \xi_{1j}|||_0^2 \; \}^{1/2} \le \delta \; \{ \; \lambda_{1j} - \mu \; \} \tag{3.5}$$

for some $\delta > 0$. Then there exists a constant $0 \le \gamma < 1$, and an integer $m \ge 1$, both independent of j, such that, if δ is sufficiently small and

$$|||\bar{q}-q||| \le \gamma^p \; |||\bar{q}||| \;, \tag{3.6}$$

then

$$|||z_{m+1}-z||| \le \gamma \; |||z_0-z||| \;. \tag{3.7}$$

Proof: Let $e_k = z_k - z$, $0 \le k \le m+1$, and let

$$e_0 = \sum_{i=1}^{N_j} c_i \, \xi_{ij} .$$

Then a straightfoward computation shows that for $1 \le k \le m$

$$e_k = \sum_{i=1}^{N_j} c_i \, \xi_{ij} \; \{ \; (1-\lambda_{ij}/\lambda_{N_j j}) \; / \; (1 - \mu/\lambda_{N_j j}) \quad \}^k \;. \tag{3.8}$$

From (3.8), it is evident that

$$|||e_m||| \le |||e_0||| . \tag{3.9}$$

Next, note that (3.3) can be written as

$$a(\bar{q}-e_m, v) - \mu \; (\bar{q}-e_m, v) = 0 \;, \tag{3.10}$$

for all $v \, \varepsilon \, \mathcal{M}_{j-1}$. From Lemma 2.3, we have

$$|||\bar{q}||| \le C \; |||e_m||| \le C \; |||e_0||| . \tag{3.11}$$

We must now obtain a bound for $|||\bar{q}-e_m|||$. Let

$$\begin{aligned}\bar{q} - e_m &= \beta_1 \xi_{1j} + \sum_{i=2}^{N_j} \beta_i \xi_{ij} \\ &= \rho_1 + \rho_2 ,\end{aligned}$$

and choose $\bar{\rho} \varepsilon \mathcal{M}_{j-1}$ to satisfy

$$a(\rho_2-\bar{\rho},v) - \mu (\rho_2-\bar{\rho},v) = 0$$

for all $v \varepsilon \mathcal{M}_{j-1}$. Then, for $\eta \varepsilon \mathcal{M}_{j-1}$,

$$\begin{aligned}|||\rho_2|||^2 &= a(\rho_2,\rho_2) \\ &\leq (1 - \mu/\lambda_{2j})^{-1} \{ a(\rho_2,\rho_2) - \mu (\rho_2,\rho_2) \} \\ &= (1 - \mu/\lambda_{2j})^{-1} \{ a(\bar{q}-e_m,\rho_2) - \mu (\bar{q}-e_m,\rho_2) \} \\ &= (1 - \mu/\lambda_{2j})^{-1} \{ a(\bar{q}-e_m,\rho_2-\bar{\rho}) - \mu (\bar{q}-e_m,\rho_2-\bar{\rho}) \} \\ &= (1 - \mu/\lambda_{2j})^{-1} \{ a(\eta-e_m,\rho_2-\bar{\rho}) - \mu (\eta-e_m,\rho_2-\bar{\rho}) \} \\ &\leq C (\lambda_{N_j j}-\mu)^{1/2} \|\eta-e_m\|_0 \; |||\rho_2-\bar{\rho}||| \\ &\leq C (\lambda_{n_j j}-\mu)^{1/2} h_j^{1+\alpha} \; |||e_m|||_{1+\alpha} \; |||\bar{q}-e_m||| \end{aligned} \tag{3.12}$$

where we have used Lemmas 2.2 and 2.3. A standard argument given in [2] shows

$$|||e_m|||_{1+\alpha} < C m^{-\alpha/2} \lambda_{N_j j}^{\alpha/2} |||e_0||| . \tag{3.13}$$

Thus, (3.12), (3.13), together with (2.13) show

$$|||\rho_2|||^2 \leq C m^{-\alpha/2} |||\bar{q}-e_m||| \; |||e_0||| . \tag{3.14}$$

To estimate $|||\rho_1|||$, note

$$|||\rho_1||| = \lambda_{1j}^{1/2} (\lambda_{1j}-\mu)^{-1} \mid a(\bar{q}-e_m,\xi_{1j}) - \mu (\bar{q}-e_m,\xi_{1j}) \mid$$

$$= \lambda_{1j}^{1/2} (\lambda_{1j}-\mu)^{-1} \mid a(\bar{q}-e_m,\xi_{1j}-\xi_{1j-1}) - \mu (\bar{q}-e_m,\xi_{1j}-\xi_{1j-1}) \mid$$

$$\leq C \delta |||\bar{q}-e_m||| \tag{3.15}$$

where we have used (3.5). Thus

$$|||\bar{q}-e_m|||^2 = |||\rho_1|||^2 + |||\rho_2|||^2$$

$$\leq C m^{-\alpha/2} |||e_0||| \; |||\bar{q}-e_m||| + C' \delta |||\bar{q}-e_m|||^2$$

or

$$|||\bar{q}-e_m||| \leq C m^{-\alpha/2} (1 - C'\delta)^{-1} |||e_0||| . \tag{3.16}$$

Thus, from (3.6), (3.11), and (3.16)

$$|||e_{m+1}||| \leq |||e_m-\bar{q}||| + |||\bar{q}-q|||$$

$$\leq C \{ \gamma^p + m^{-\alpha/2} (1-C'\delta)^{-1} \} |||e_0||| . \tag{3.17}$$

Choosing γ sufficiently small that $C\gamma^p \leq \gamma/2$ and m sufficiently large that $Cm^{-\alpha/2} (1-C'\delta)^{-1} \leq \gamma/2$ completes the proof.

We now consider in detail the role of assumption (3.5) (and (2.20) of Lemma 2.3). Basically (3.5) provides the ability to control the error in the direction of ξ_{1j} in (3.15). As μ approaches λ_{1j}, the argument given in (3.15) fails. Thus, in the term $|||\bar{q}-e_m|||$ in (3.17), we are only able to control the portion of the error orthogonal to ξ_{1j}. However, since we expect to incorporate this scheme in an inverse iteration procedure, we are really only interested in computing a vector in the direction of ξ_{1j}, and as a practical matter we can tolerate (or even applaud) errors in this particular direction.

The term $|||\bar{q}-q|||$ in (3.17) measures the error due to the recursive application of the j-level iteration. As μ approaches λ_{1j}, we lose the ability to control the error in the direction of ξ_{1j-1} in Lemma 2.3. However, this is also not serious in the present context for several reasons. First, most of this error lies in the direction of ξ_{1j}; by (2.14), one can see that the error in ξ_{1j-1} orthogonal to ξ_{1j} is smaller by $O(h_j^{\alpha})$ than the part lying in the direction of ξ_{1j}. Second, with $\mu \leq \lambda_{1j} \leq \lambda_{1j-1}$ by (2.15) (with $\lambda_{1j} < \lambda_{1j-1}$ unless $\xi_{1j} = \xi_{1j-1}$) the deterioration of the bound (2.23) is generally much less severe than in (3.15).

Finally, as a practical matter, the proofs of Lemma 2.3 and Theorem 3.1 suggest that it might be advantageous to choose shifts μ which satisfy $\mu < \lambda_{1j}$. For example, if one were to choose a shift satisfying $\lambda_{1j} < \mu \leq \lambda_{1j-1}$ with μ closer to λ_{1j-1} than to λ_{1j}, convergence to the direction of ξ_{1j} might be retarded somewhat, since the term $|||\bar{q}-q|||$ would then be the term over which we have least control.

4. A Multi-level Inverse Iteration Algorithm

In this section we consider an idealized inverse iteration procedure in which the j-level scheme is used to solve (approximately) the resulting linear systems. Suppose λ_{1j} is known and we have some initial guess $\sigma_{0j} \in \mathcal{M}_j$. Let $|||\sigma_{0j}||| = 1$, and assume

$$| a(\sigma_{0j},\xi_{1j}) | \geq \lambda_{1j}^{1/2}/2 \quad , \tag{4.1}$$

and that the shift μ_j satisfies (3.5). Then the inverse iteration procedure for computing σ_{kj}, k=1,2,... is defined as follows: Compute $\tilde{\sigma}_{kj}$ $\varepsilon\ \mathcal{M}_j$, an approximation of $\bar{\sigma}_{kj}$ $\varepsilon\ \mathcal{M}_j$, where

$$a(\bar{\sigma}_{kj},v) - \mu_j\ (\bar{\sigma}_{kj},v) = \{\lambda_{1j}-\mu_j\}\ (\sigma_{k-1j},v) \tag{4.2}$$

for all v $\varepsilon\ \mathcal{M}_j$, using r $\geq$ 1 iterations of the j-level scheme and initial guess σ_{k-1j}. Then set

$$\sigma_{kj} = \tilde{\sigma}_{kj}\ /\ |||\tilde{\sigma}_{kj}|||\ . \tag{4.3}$$

We shall choose r and μ_j to satisfy

$$0 < (\lambda_{1j}-\mu_j)\ /\ (\lambda_{2j}-\mu_j) < 1/2 - 2\gamma^r \tag{4.4}$$

where γ is given in Theorem 3.1. This can be done consistantly with (3.5) since λ_{1j} and λ_{2j} are assumed to be well separted independent of j. Let ω_{kj}, $\bar{\omega}_{kj}$, and $\tilde{\omega}_{kj}$ denote the restrictions of σ_{kj}, $\bar{\sigma}_{kj}$, and $\tilde{\sigma}_{kj}$, respectively, to the orthogonal complement of ξ_{1j}.

Theorem 3.1 tells us that

$$\begin{aligned} |||\tilde{\sigma}_{kj}-\bar{\sigma}_{kj}||| &\leq \gamma^r\ |||\sigma_{k-1j}-\bar{\sigma}_{kj}||| \\ &\leq \gamma^r\ |||\omega_{k-1j}|||\ . \end{aligned} \tag{4.5}$$

We now seek to estimate $|||\omega_{kj}||| = |||\tilde{\omega}_{kj}|||\ /\ |||\tilde{\sigma}_{kj}|||$ in terms of $|||\omega_{k-1j}|||$. First note that

$$\begin{aligned} |||\tilde{\omega}_{kj}||| &\leq |||\bar{\omega}_{kj}||| + |||\tilde{\omega}_{kj}-\bar{\omega}_{kj}||| \\ &\leq |||\bar{\omega}_{kj}||| + |||\tilde{\sigma}_{kj}-\bar{\sigma}_{kj}||| \\ &\leq \{\ (\lambda_{1j}-\mu_j)\ /\ (\lambda_{2j}-\mu) + \gamma^r\ \}\ |||\omega_{k-1j}||| \end{aligned} \tag{4.6}$$

where we have used (4.5). Also

$$| a(\tilde{\sigma}_{kj},\xi_{1j}) | \geq | a(\bar{\sigma}_{kj},\xi_{1j}) | - | a(\tilde{\sigma}_{kj}-\bar{\sigma}_{kj},\xi_{1j}) |$$

$$\geq (1/2 - \gamma^{r}) \lambda_{1j}^{1/2} \quad , \qquad (4.7)$$

where the second inequality follows if we assume (as an induction hypothesis) that

$$| a(\sigma_{k-1j},\xi_{1j}) | \geq \lambda_{1j}^{1/2}/2 \ . \qquad (4.8)$$

Noting that (4.4) shows $\gamma^{r} < 1/2$, we have

$$|||\tilde{\sigma}_{kj}||| \geq 1/2 - \gamma^{r} > 0 \ , \qquad (4.9)$$

and

$$|||\omega_{kj}||| \leq \{ (\lambda_{1j}-\mu_{j}) \ / \ (\lambda_{2j} - \mu_{j}) + \gamma^{r} \} (1/2 - \gamma^{r})^{-1} |||\omega_{k-1j}|||$$

$$\leq \varepsilon \ |||\omega_{k-1j}||| \quad , \qquad (4.10)$$

where $\varepsilon < 1$ by (4.4). Note that for an appropriate choice of shift μ_{j}, both ε and r can be taken independent of j. Also note that (4.10) validates the induction hypothesis (4.8). Thus we have shown

Theorem 4.1: Let the hypotheses of Theorem 3.1 hold,let r and μ_{j} satisfy (4.4), and let σ_{0j} satisfy (4.1). Then the iterates σ_{kj} defined by (4.2)-(4.3) converge to $\pm \lambda_{1j}^{-1/2} \ \xi_{ij}$.

As the final step in our analysis, we consider an overall procedure in which we sequentially compute approximations to ξ_{1j}, $j=1,2,\ldots$, using the approximation of ξ_{1j-1} as the initial guess for ξ_{1j}.

1. For j=1, compute σ_1, an approximation of $\pm \lambda_{11}^{-1/2} \xi_{11}$, using s_1 iterations of (4.2)-(4.3) using the 1-level scheme (direct solution) and some initial guess not deficient in ξ_{11}.

2. Then for j > 1 we compute $\sigma_j \; \varepsilon \mathcal{M}_j$, an approximation of $\pm \lambda_{1j}^{-1/2}$ ξ_{1j}, using s iterations of (4.2)-(4.3) and initial guess $\sigma_{j-1} \; \varepsilon$ $\mathcal{M}_{j-1} \subset \mathcal{M}_j$.

To analyze this procedure, we first consider the extent to which σ_{j-1} lies in the direction of ξ_{1j}. Letting ω_j be the component of σ_j orthogonal to ξ_{1j}, and assuming $|||\sigma_j||| = 1$, we have

$$\begin{aligned} \sigma_{j-1} &= \{ (1-|||\omega_{j-1}|||^2) \lambda_{1j-1}^{-1} \}^{1/2} \xi_{1j-1} + \omega_{j-1} \\ &= \{ (1-|||\omega_{j-1}|||^2) \lambda_{1j-1}^{-1} \}^{1/2} \xi_{1j} \qquad (4.11) \\ &\quad + \{ (1-|||\omega_{j-1}|||^2) \lambda_{1j-1}^{-1} \}^{1/2} \{ \xi_{1j-1} - \xi_{1j} \} + \omega_{j-1} \quad . \end{aligned}$$

The last two terms in (4.11) may have some non-trivial component in the direction of ξ_{1j}. Nonetheless, on the basis of (4.10) and (2.14), we have, for j > 1

$$\begin{aligned} |||\omega_j||| &\leq \varepsilon^s \{ (1-|||\omega_{j-1}|||^2) \lambda_{1j-1}^{-1} \}^{1/2} |||\xi_{1j-1}-\xi_{1j}||| \\ &\quad + \varepsilon^s |||\omega_{j-1}||| \\ &\leq \varepsilon^s \{ \mathcal{K}_0 h_j^{\alpha} \lambda_1^{(\alpha-1)/2} (1+2^{\alpha}) + |||\omega_{j-1}||| \}. \qquad (4.12) \end{aligned}$$

Solving the majorizing difference equation, under the assumption that $2^{\alpha}\varepsilon^s < 1$, we obtain

$$|||\omega_j||| \leq \varepsilon^{s(j-1)} \; |||\omega_1||| + \frac{\theta}{2} \mathcal{K}_0 \; h_j^{\alpha} \; \lambda_1^{(\alpha-1)/2}$$

$$\theta = 2 \; \varepsilon^{s} \; (1+2^{\alpha}) \; (1-\varepsilon^{s} 2^{\alpha})^{-1} \tag{4.13}$$

Choosing s_1 sufficiently large that

$$|||\omega_1||| \leq \frac{\theta}{2} \mathcal{K}_0 \; h_1^{\alpha} \; \lambda_1^{(\alpha-1)/2} \tag{4.14}$$

we obtain for $j \geq 1$,

$$|||\omega_j||| \leq \theta \, \mathcal{K}_0 \; h_j^{\alpha} \; \lambda_1^{(\alpha-1)/2}. \tag{4.15}$$

Finally, we note from (4.12) that for $j > 1$,

$$| \; a(\sigma_{j-1},\xi_{1j}) \; | = \lambda_{1j} \; \lambda_{1j-1}^{-1/2} \; (1-|||\omega_{j-1}|||^2)^{1/2} - C \, \mathcal{K}_0 \; h_j^{\alpha} \; \lambda_1^{\alpha/2}$$
$$> \lambda_{1j}^{1/2}/2$$

for h_1 sufficiently small.

We next consider the extent to which σ_j lies in the direction of ξ_1. In particular,

$$\sigma_j - \xi_1\lambda_1^{-1/2} = \{ \; \xi_{1j} - \xi_1 \; \} \; \lambda_1^{-1/2}$$
$$+\{ \; (1-|||\omega_j|||^2)^{1/2} \; \lambda_{1j}^{-1/2} - \lambda_1^{-1/2} \; \} \; \xi_{1j} + \omega_j \; . \tag{4.16}$$

Since all three terms are of the same size, we have from (2.14), (4.15), and (4.16)

$$|||\sigma_j-\xi_1\lambda_1^{-1/2}||| \leq \mathcal{R}_0 \; h_j^{\alpha} \; \lambda_1^{(\alpha-1)/2} \; \rho \tag{4.17}$$

for $\rho = O(1)$ independent of j.

Theorem 4.2: Let s be chosen such that $\varepsilon^s 2^\alpha < 1$, and let the hypotheses of Theorem 4.1 hold. Then the cost of computing σ_j satisfying (4.17) is $O(N_j)$ for j sufficiently large.

Proof: By theorems proved in [2], [1], the cost of a single j-level iteration is asymptotically bounded by $C_0 N_j$. The total number of j-level iterations on each level for $j > 1$ is r·s, where r and s can be chosen to be independent of j. Thus the cost of computing σ_j is bounded by

$$C_1 + \sum_{k=2}^{j} C_0 \; r \; s \; N_k \leq C_2 \; N_j$$

where C_1 is the (fixed) cost of comuting σ_1, and we have used the fact that $N_k \cong 4 \; N_{k-1}$.

5. Numerical Example and Concluding Remarks

In practical computations the values of the integer parameters m, r, and s can usually be chosen to be relatively small, on the order of 1 - 4. Note that the procedure for sequentially computing the σ_j as in Theorem 4.2 consists of r·s j-level iterations ($j > 1$), with the right hand side updated every r-th iteration.

As an example, consider the problem

$$\begin{aligned} -\Delta u &= \lambda \; u \qquad \text{in } \Omega = (0,1) \times (0,1), \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{5.1}$$

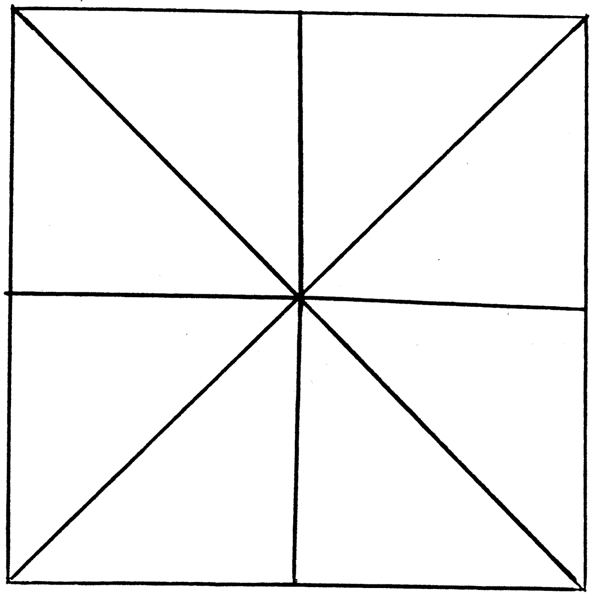

Figure 5-1:$\mathcal{T}_1$

with $_1$ the 3 X 3 grid pictured in figure 5-1. We solved this problem using four levels of traingulation, as described in Section 2, with the finest mesh being 17 X 17. We chose m = 2, r = 1, and s = 2. Empirically, this results in $\gamma^r \cong 10^{-2}$. The smoothing iteration for the j-level scheme was symmetric Gauss-Seidel. The value of λ_{1j} on the right hand side of (4.3) is approximated by a Rayleigh quotient. If u $\varepsilon \mathcal{M}_j$, u $\neq$ 0, then the Rayleigh quotient R(u) is given by

$$R(u) = a(u,u) \;/\; (u,u)$$

The initial guess for level j is given by a local piecewise quadratic approximation constructed from the the level j-1 solution (in most cases, this appears to be slightly better than just using the level j-1 solution). In other respects, the multi-level method used is very close to the scheme described here. Some results of the computation are shown in Table 5-1. The true eigenvalue $\lambda_1 = 2\ \pi^2 \cong 19.7392$. The approximate eigenvalues λ_{1js} are the Rayleigh quotients of the approximate eigenvector iterates. In

Table 5-1: Convergence of the Inverse Iteration Procedure

j	N_j	μ_j	s	λ_{1js}	$\lvert(\lambda_{1js-1}-\lambda_{1js})/\lambda_{1js}\rvert$
1	1	-216.0	0	24.0	
2	9	18.0	0	24.0	
			1	21.659	1.1 E 0
			2	21.658	5.2 E-5
3	49	16.2	0	21.272	
			1	20.272	4.9 E-2
			2	20.270	8.3 E-5
4	225	15.2	0	19.944	
			1	19.876	3.4 E-3
			2	19.876	6.2 E-7

terms of the continuous problem, s = 1 is more than adaquate to compute an approximation of λ_1 to the order of the discretization error.

One can note the rapid convergence of the inverse iteration procedure despite the relatively poor shift policy (the computations were done on a DEC-20 using single precision arithmetic - a 27 bit mantissa). The conservative shift policy was used to bias the Rayleigh quotient iteration in favor of convergence to a multiple of ξ_{1j}. The initial shift for the level j problem is given by

$$\mu_j = \tilde{\lambda}_{1j-1} - \tilde{\lambda}_{1j-1}/4 , \qquad j > 1 \tag{5.2}$$

where $\tilde{\lambda}_{1j}$ is the accepted value of λ_{1j}. This imparts some flavor of the power method and thus favors convergence to ξ_{1j}. In this particular program, a new shift is computed every 5-th iteration (in s), and is given by the average of the old shift and current eigenvalue estimate (only rarely does one need s > 5). Thus the sequence of shifts is chosen to (hopefully) approach the true eigenvalue from below and thus bias the overall iteration in favor of convergence to ξ_{1j}. A shift policy this conservative was not necessary for for this particular problem but has proved necessary when λ_1 is close to λ_2 and ξ_1 is not very smooth.

There are situations in which we may be interested in computing several eigenvalues and corresponding eigenvectors. One approach to this problem is to used the method of subspace iteration. Suppose λ_j, $1 \leq j \leq n$, are well separated from the remaining eigenvalues. We begin the procedure with n orthonormal vectors. We perform an inverse power step on each vector, and use the resulting vectors to form an n X n matrix. We then solve the resulting (generally dense) eigenvalue problem to obtain approximate eigenvalues and to re-orthonormalize the vectors. This sequence is then repeated until convergence [12], [10].

The analysis in Sections 2 and 3 generalizes in straightfoward fashion to cover this situation. The j-level iteration is used in all of the inverse power steps. The strategy of solving the problem sequentially on the grids $\mathcal{T}_1, \mathcal{T}_2, \ldots$, and using the solution on the j-th grid as the initial guess at the solution on the j+1-st grid is also applicable. However, the condition that h_1 be sufficiently small becomes more restrictive an n increases. This is because bounds analagous to (2.14) for other eigenfunctions and eigenvalues generally deterioriate as λ_j increases; for example, the approximation of λ_{N_j} and ξ_{N_j} in the space $\mathcal{M}_j$ is typically O(1) [12]. Thus as n increases, the fineness of $\mathcal{T}_1$ must increase in order to insure adaquate approximation of all the desired eigenvalues and eigenfunctions.

Another procedure for computing several eigenvalues and eigenvectors has recently been proposed by Ericsson and Ruhe [6]. Their scheme is based on the Lanczos algorithm. In their work, they wish to compute all eigenvalues lying in a specified interval and the corresponding eigenvectors. Multi-level techniques can be applied to this situation in a fashion analagous to subspace iteration. A common technique for determining the number of eigenvalues less than a given number μ, is to compute an LDL^T composition of the matrix $A - \mu M$, and count the number of negative pivots [6], [12]. This can still be done in the present case since the matrix corresponding to the coarsest grid $\mathcal{T}_1$ is factored as part of the multi-level procedure (success here again requires that h_1 be small enough that all the eigenvalues and eigenvectors of interest are well approximated on the coarsest grid).

The stratgedy of sequentially solving the problem on $\mathcal{T}_1, \mathcal{T}_2, \ldots$ is again advantageous, since the issue of how many eigenvalues lie in the given interval and their distribution can be resolved once and for all on the coarsest grid, where computation is relatively inexpensive. The solution of the remaining problems can be viewed primarily as a means of increasing the accuracy of the eigenvalues and eigenvectors as approximations of the continuous problem. One may even wish to switch or modify the algorithm after the first grid, since much of its power would no longer be necessary.

In both this scheme and subspace iterration, it may be desirable to choose shifts for which $A - \mu M$ is indefinite, although μ may not be especially close to any eigenvalue. The multi-level schemes for indefinte systems described in [1] could be used. If the shift were close to an eigenvalue, the analysis given there would have to be modified in a fashion analogue to Theorem 3.1. As in the case of λ_{1j}, choosing a shift to be slightly less than the the given eigenvalue is probably safer than choosing it to be slightly larger, since corresponding eigenvalues on different grids will satisfy inequalities similar to (2.15).

If the integer parameters m, r, and s are fixed, then $O(r \cdot s \cdot n)$ j-level iterations will be used in computing an approximate solution of the level j problem, where n is the number of approximate eigenvectors, and s is the number of subspace iterations per level, or the number of Lanczos steps per eigenvector per level, depending on the algorithm. In any event, m, r, and s can be selected independent of j and the work to solve linear systems will be $O(n \cdot N_j)$. Both procedures require the solution of small eigenvalue problems of fixed size. The number of such problems per level is fixed, so this will contribute a term like $O(F(n) \cdot \log(N_j))$ to the work estimate, where F(n) is a bound on the cost of solving one of these problems (note j is proportional to $\log(N_j)$ for spaces that increase geometrically in size). In the subspace iteration procedure, one must re-orthonormalize the trial vectors at each step. This cost is essentially the same as the cost of multiplying an n X n matrix and an n X N_j matrix. Since a fixed number of subspace iterations will be used per level, this will contribute a term like $O(n^2 \cdot N_j)$ to the work estimate. A similar sort of cost will be required in the Lanczos scheme. There may well be good hueristics for reducing the orthonormalization cost, since as j incresaes, the initial guesses for the eigenvectors improve.

REFERENCES

[1] Randolph E. Bank. A Comparison of Two Multi-level Iterative Methods for Non-symmetric and Indefinite Finite Element Equations. *SIAM Journal on Numerical Analysis* to appear:, .

[2] Randolph E. Bank and Todd F. Dupont. An Optimal Order Process for Solving Finite Element Equations. *Mathematics of Computation* 36:, 1981.

[3] Randolph E. Bank and Andrew H. Sherman. *PLTMG Users' Guide*. Technical Report CNA152, Center for Numerical Analysis, University of Texas, 1979.

[4] J. L. Blue, A. Kahn, J. E. Lowney, and C. L. Wilson. Disappearance of Impurity Levels in Silicon and Germanium due to Screening. *Journal of Applied Physics* :, submitted.

[5] J. L. Blue and C. L. Wilson. Calculating Eigenvalues and Eigenfunctions Using an Interior Constraint. *Journal of Computational Physics* :, submitted.

[6] Thomas Ericsson and A. Ruhe. The Spectral Transformation Lanczos Method for the Numerical Solution of Large Sparse Generalized Symmetric Eigenvalue Problems. *Mathematics of Computation* 35:1251,1268, 1980.

[7] Wolfgang Hackbusch. On the computation of Approximate Eigenvalues and Eigenfunctions of Elliptic Operators by Means of a Multi-grid Method. *SIAM Journal on Numerical Analysis* 16:201,215, 1979.

[8] Stephen F. McCormick. *A Mesh Refinement Method for* $Ax = \lambda Bx$. Technical Report , Colorado State University, 1979.

[9] J. T. Oden and J. R. Reddy. *An Introduction to the Mathematical Theory of Finite Elements*. Interscience, New York, 1976.

[10] B. N. Parlett. *The Symmetric Eigenvalue Problem*. Prentice-Hall, Englewood Cliffs, New Jersey, 1980.

[11] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press, New York, 1973.

[12] Gilbert Strang and George J. Fix. *An Analysis of the Finite Element Method*. Prentice-Hall, Englewood Cliffs, New Jersey, 1973.