Russ. J. Numer. Anal. Math. Modelling, Vol.7, No.6, pp. 473–486 (1992)

On an iterative method for finding lower eigenvalues

E. G. D'YAKONOV and A. V. KNYAZEV

Abstract – We present the analysis of the errors involved in approximate orthogonalization with respect to previously found eigenvectors in preconditioned iterations of a subspace for simultaneous determination of a cluster of eigenvalues and the corresponding eigenvectors of a large sparse symmetrical eigenvalue problem.

1. INTRODUCTION

Efficient preconditioned iterative methods have been lately gaining acceptance for solving partial symmetric sparse eigenvalue problems arising in grid discretization of spectral problems in mathematical physics (see, e.g. [1], the review [4], and the references therein). In addition to conventional preconditioned vector iterations for calculation (in turns) of the eigenvectors, there have been developed [3] preconditioned iterations of a subspace for simultaneous calculation of a group of the eigenvectors corresponding to a cluster of eigenvalues: using the subspace H_p^n one constructs a new p-dimensional subspace H_p^{n+1} and thereby determines more exact approximations to λ_p and U_{λ_p}, where λ_p is the p-th, in increasing order, eigenvalue of the problem:

$$Lu = \lambda Mu, \qquad L \in C(H), \qquad M \in C(H) \tag{1.1}$$

while U_{λ_p} is the corresponding eigen subspace

$$C(H) \equiv \{A: A \in \mathscr{L}(H,H), A = A^* > 0\}$$

H is the N-dimensional Euclidean space with the inner product (u,v). It has been proved that

$$\beta^n \equiv \beta(H_p^n) \equiv \max_{u \in H_p^n, u \neq 0} \left\{ \frac{(Lu,u)}{(Mu,u)} \right\} \tag{1.2}$$

converges to λ_p at a rate of the geometric progression with the ratio of

$$q_p \equiv q \equiv q\left(\delta, \frac{\beta^0}{\lambda_p}, \frac{\beta^0}{\nu_p}, \frac{\lambda_p}{\nu_p}\right) < 1$$

provided $\beta^0 < \nu_p$, where

$$\nu_p \equiv \lambda_{p+1} > \lambda_p, \qquad \delta \equiv \delta_0/\delta_1, \qquad \delta_0 B \leqslant L \leqslant \delta_1 B, \qquad \delta_0 > 0 \tag{1.3}$$

$$H_p^{n+1} = R_{\beta^n} H_p^n, \qquad R_{\beta^n} \equiv E - \gamma B^{-1}(L - \beta^n M). \tag{1.4}$$

When implementing the method, it is necessary to solve systems of the form of $Bv_i^{n+1} = g_i^n$ p-times at each iteration. In the case of grid problems an appropriate choice of the operator B is usually associated with the spectral equivalence of the operators L and B, and the operator B can be selected out of known samples.

It is also easy to calculate the corresponding approximation to U_{λ_p} in the subspace H_p^n. Using orthogonalization with respect to some approximately determined eigen subspaces, one can construct similar iterative methods of finding all λ_i, $i \leqslant p$, and the corresponding subspaces U_{λ_i}. The convergence rates of these methods are equal to those of geometric progressions with a ratio of $q_i \leqslant q_p = q$. Thus, the availability of some closely-spaced λ_i with $i \leqslant p$ does not decrease the convergence rate of the iterative method being discussed. The urgency of the problem of finding iterative methods possessing the above property has been pointed out, for instance, by G. I. Marchuk and V. I. Lebedev.

This paper is concerned with studying the convergence of the method mentioned above, with allowance made for the fact that orthogonalization with respect to the previously determined eigen subspaces is approximate. This study is, to a considerable extent, analogous to that made for vector iterations in [1] where more detailed proofs are available. The reader frightened by cumbersome calculations typical of the direct error analysis is referred to [1,4] where a backward error analysis is available, though for a somewhat more tedious orthogonalization procedure. The case of the singular operator M of unfixed sign is also studied in [1,4].

Here we present the results of numerical experiments on implementing the method under consideration for model difference problems.

This work has been published in Russian (see [2]), the main results have been discussed in [1] without detailed proofs.

2. DETERMINATION OF λ_p

For $A \in C(H)$ let H_A be an Euclidean space that differs from H only in the inner product

$$(u,v)_{H_A} \equiv (u,v)_A \equiv (Au,v)\,, \qquad \|u\|_A^2 \equiv (Au,u)\,.$$

By $u_1, u_2, \ldots, u_N$ we shall denote an orthonormal basis H_M comprising the eigenvectors of (1.1):

$$Lu_i = \lambda_i M u_i\,, \qquad i = 1,\ldots,N$$
$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N\,. \tag{2.1}$$

Let ω_0 be a subset of the indices i for which there have been found the approximations $\bar{u}_i$ to u_i, where

$$\|u_i - \bar{u}_i\|_L \leqslant \varepsilon_{L,i}\,, \qquad \|u_i - \bar{u}_i\|_M \leqslant \varepsilon_{M,i}$$
$$\max_{i \in \omega_0} \{\varepsilon_{L,i}, \varepsilon_{M,i}\} = \varepsilon \tag{2.2}$$

$Q \equiv Q_{\omega_0}$ and $\bar{Q} \equiv \bar{Q}_{\omega_0}$ are, respectively, the linear spans of u_i and $\bar{u}_i$, $i \in \omega_0$, $\dim Q = m$, $Q^\perp$ and $\bar{Q}^\perp$ are orthogonal in H_M complements to Q and $\bar{Q}$, P and $\bar{P}$ are orthogonal in H_M projectors onto Q and $\bar{Q}$, $P^\perp \equiv E - P$, $\bar{P}^\perp \equiv E - \bar{P}$, and E is the identity operator. The proximity of the subspaces Q and $\bar{Q}$ will be defined from (2.2) and with the use of the openings $\Theta_M \equiv \Theta_M(Q;\bar{Q})$ and $\Theta_L \equiv \Theta_L(Q;\bar{Q})$ of these subspaces into H_M and H_L, respectively. In this case (see [1]) we have

$$\Theta_M = \max_{n \in \bar{Q},\, \|u\|_M = 1} \|u - Pu\|_M = \|P - \bar{P}\|_M . \tag{2.3}$$

For $u \neq 0$ the following notation is used:

$$\mu(u) \equiv \mu \equiv \frac{(Lu,u)}{(Mu,u)} . \tag{2.4}$$

Let

$$\lambda_p^{\perp} \equiv \min_{H_p \subseteq Q^{\perp}} \max_{u \in H_p} \mu(u) , \quad \bar{\lambda}_p^{\perp} \equiv \min_{H_p \subseteq \bar{Q}^{\perp}} \max_{u \in H_p} \mu(u) , \qquad p = 1,\dots,N-m \tag{2.5}$$

where H_p is a p-dimensional subspace of H. It is easily seen that $\lambda_p^{\perp}$ coincides with the p-th, in increasing order, eigenvalue λ_i that has remained after eliminating all λ_i with $i \in \omega_0$ from the set $\lambda_1,\dots,\lambda_N$; if $\lambda_i \geqslant \lambda_p^{\perp}$ for all $i \in \omega_0$, then $\lambda_p^{\perp} = \lambda_p$, and it is this case that will be most important in what follows.

Let $H_{p_n}^n$ be a prescribed p-dimensional subspace with the basis $a_1^n,\dots,a_p^n$. Consider the following algebraic eigenvalue problem:

$$\bar{L}_p^n \bar{X} = \lambda \bar{M}_p^n \bar{X} \tag{2.6}$$

where

$$\bar{X} \equiv (x_1,\dots,x_p)^{\mathrm{T}}$$

$$\bar{L}_p^n \equiv [(La_i^n, a_j^n)] , \qquad 1 \leqslant i,j \leqslant p$$

$$\bar{M}_p^n \equiv [(Ma_i^n, a_j^n)] , \qquad 1 \leqslant i,j \leqslant p .$$

If $a_1^n,\dots,a_p^n$ is the orthonormal in H_M basis of H_p^n, then $\bar{M}_p^n$ becomes an identity matrix and problem (2.6) is simplified. Let $\lambda_1^n \leqslant \lambda_2^n \leqslant \dots \leqslant \lambda_p^n$ be its eigenvalues, with their eigenvectors $\bar{X}^{(i)} \equiv (x_1^{(i)},\dots,x_p^{(i)})^{\mathrm{T}}$ forming an orthonormal system in accordance with the inner product $(\bar{M}_p^n \bar{X}, \bar{Y}) = \bar{Y}' \bar{M}_p^n \bar{X}$. Then the vectors

$$v_i \equiv \sum_{j=1}^{p} x_j^{(i)} a_j^n , \qquad i = 1,\dots,p \tag{2.7}$$

form an orthonormal in H_M basis of H_p^n and

$$\mu(v_i) = \lambda_i^n = \min_{H_i \subseteq H_p^n} \max_{u \in H_i} \mu(u) \geqslant \lambda_i , \qquad \lambda_p^n = \beta(H_p^n) . \tag{2.8}$$

Besides, the orthogonality of $\bar{X}^{(1)},\dots,\bar{X}^{(p)}$ in accordance with the inner product $(\bar{L}_p^n, \bar{X}, \bar{Y})$ entails the orthogonality of the vectors $v_1,\dots,v_p$ in H_L [see (2.7)].

Lemma 2.1. Let conditions (1.3) and (2.2) be satisfied, $H_p^n \subseteq \bar{Q}^{\perp}$, $0 < \gamma < 4\delta_1^{-1}$, and

$$H_p^{n+1} = R_{\beta^n} H_p^n \tag{2.9}$$

be an image of H_p^n in mapping $R_{\beta^n} \equiv E - \gamma \bar{P}^{\perp} B^{-1}(L - \beta^n M)$. Then H_p^{n+1} is a p-dimensional subspace of $\bar{Q}^{\perp}$.

Proof. Let $u \in H_p^n$ and $\|u\|_M = 1$, $\beta^n \equiv \beta$ [see (1.2)]. Then $u = \bar{P}^{\perp} u$ and

$$(R_{\beta^n} u, u)_M = 1 - \gamma(B^{-1}Lu, Mu) + \gamma\beta\|Mu\|^2_{B^{-1}}$$
$$\geqslant 1 - \gamma\|B^{-1/2}L^{1/2}\| \, \|L^{1/2}u\| \, \|Mu\|_{B^{-1}} + \gamma\beta\|Mu\|^2_{B^{-1}}.$$

In view of (1.3) it is easy to verify that $\|B^{-1/2}L^{1/2}\| \leqslant \delta_1^{1/2}$. In addition, as $u \in H_p^n$, then $\|L^{1/2}u\| \leqslant \beta^{1/2}$. Therefore,

$$(R_\beta u, u)_M \geqslant 1 - \gamma\delta_1^{1/2}\beta^{1/2}t + \gamma\beta t^2$$

where $t \equiv \|Mu\|_{B^{-1}}$. Hence,

$$(R_\beta u, u)_M \geqslant 1 - \tfrac{1}{4}\gamma\delta_1 \geqslant K_0 > 0, \qquad \|R_\beta u\|_M \geqslant K_0$$

which proves the lemma. □

Note that if ω_0 is empty, then the operator $R_\beta = E - \gamma B^{-1}(L - \beta M)$ is self-adjoint in H_B and $(R_\beta u, u)_B \geqslant \gamma\beta\|u\|^2$ for $0 \leqslant \gamma \leqslant \delta_1^{-1}$ and for any $u \in H$. If $u \in H_p^n$, then $(R_\beta u, u)_B \geqslant \|u\|_B^2$ for any $\gamma \geqslant 0$. To find the basis $a_1, \ldots, a_p$ of the subspace H_p^{n+1} it is sufficient to solve the system of equations:

$$Bz_i = -\gamma(Lv_i - \beta Mv_i), \qquad i = 1, \ldots, p \tag{2.10}$$

and to set $a_i = v_i - \bar{P}^{\perp} z_i$. Now we can again construct the problem of the type of (2.6) for H_p^{n+1} and find an orthonormal in H_M basis of H_p^{n+1} consisting of the vectors $v_1^{n+1}, \ldots, v_p^{n+1}$ [see (2.7)].

Now we proceed to a study of the convergence of the quantity $\beta^n \equiv \beta(H_p^n)$ from (1.2) to $\lambda_p^{\perp}$ with H_p^n determined by iterative method (2.9) at $n = 0, 1, \ldots$.

Let $\dim Q_{\omega_0} = m$ and $\lambda_1^{\perp} \leqslant \lambda_2^{\perp} \leqslant \ldots \leqslant \lambda_{N-m}^{\perp}$ [see (2.5)]. Let $\nu_k^{\perp}$ be an eigenvalue $\lambda_i^{\perp}$ closest to $\lambda_k^{\perp}$ and strictly greater than $\lambda_k^{\perp}$.

Lemma 2.2 (see [1]). Let the conditions of Lemma 2.1 be satisfied, $u \in \bar{Q}^{\perp}$, $\|u\|_M = 1$, $\lambda \equiv \lambda_k^{\perp} \leqslant \mu(u) \equiv \mu < \nu_k^{\perp} \equiv \nu$, $\beta > 0$, $r_\beta \equiv Lu - \beta Mu$. Then

$$\|r_\beta\|^2_{L^{-1}} \geqslant \frac{\beta^2}{\mu}\left[\frac{(\mu - \lambda)(\nu - \mu)}{\lambda\nu} + 1 - \frac{\mu}{\beta}\right] - \frac{\beta^2}{\lambda\nu} \sum_{i \in \omega_{0,0}} \varepsilon^2_{M,i} \frac{(\nu - \lambda_i)(\lambda_i - \lambda)}{\lambda_i} \tag{2.11}$$

where $\omega_{0,0}$ is a subset of the indices i from ω_0 such that $\lambda < \lambda_i < \nu$.

Lemma 2.3. Let the conditions of Lemma 2.1 be satisfied, $0 < \gamma < 2\delta_1^{-1}$, $u \equiv u^n \in H_p^n$, $\|u\|_M = 1$, $\mu(u) \equiv \mu^n \equiv \mu$, $\beta^n \equiv \beta$, $\delta \equiv \delta_0/\delta_1$, $u^{n+1} \equiv R_\beta u^n$ [see (2.9)], $\mu^{n+1} \equiv \mu(u^{n+1})$. Then, for sufficiently small ε there exist such numbers $\bar{\gamma} \equiv \bar{\gamma}(\gamma, \delta_1) > 0$, $K_2 > 0$ that the following inequality holds:

$$\beta^n - \mu^{n+1} \geqslant t(u)[1 + \beta^{-1}t(u)]^{-1} - K_2\varepsilon^2 \tag{2.12}$$

where $t(u) \equiv \bar{\gamma}\|r_\beta\|^2_{B^{-1}} + \beta - \mu$.

Proof. Let $w \equiv \bar{P}^{\perp} B^{-1} r_\beta$. Then

$$\mu^n - \mu^{n+1} = \{2\gamma(w, r_\beta) - \gamma^2\|w\|_L^2 + \gamma^2\mu\|w\|_M^2 + 2\gamma(\beta - \mu)(u, w)_M\}\|u^{n+1}\|_M^{-2}.$$

Let us represent the term $(w,r_\beta) = (\bar{P}^\perp B^{-1} r_\beta, r_\beta)$ in the form:

$$(w,r_\beta) = \|r_\beta\|^2_{B^{-1}} - (PB^{-1}r_\beta,r_\beta) + ((\bar{P}^\perp - P^\perp)B^{-1}r_\beta,r_\beta)$$

and find a lower bound, by analogy with the estimate for $(\bar{w},\bar{r})$ given in [1]. Then

$$|(PB^{-1}r_\beta,r_\beta)| \leqslant \|PM^{-1}r_\beta\|_M \|B^{-1}r_\beta\| \leqslant 2K_\omega \varepsilon \|r_\beta\|_{B^{-1}} \leqslant a_1 \|r_\beta\|^2_{B^{-1}} + a_1^{-2}(K_\omega \varepsilon)^2$$

where

$$2K_\omega \geqslant \lambda_1^{-1/2}\delta^{1/2}\Big[\sum_{i\in\omega_0}(\lambda_i-\beta)^2\Big]^{1/2}, \qquad a_1 \in (0,1).$$

Besides,

$$|((\bar{P}^\perp - P^\perp)B^{-1}r_\beta,r_\beta)| \leqslant \|P-\bar{P}\|_B \|r_\beta\|^2_{B^{-1}} \leqslant \delta^{-1/2}æ_L \varepsilon \|r_\beta\|^2_{B^{-1}}$$

where

$$æ_L \geqslant \lambda_1^{-1/2}\Big(m^{1/2} + \sum_{i\in\omega_0}\lambda_i^{1/2}\Big).$$

Therefore,

$$(w,r_\beta) \geqslant \|r_\beta\|^2_{B^{-1}}(1 - \varepsilon\delta^{-1/2}æ_L - a_1) - a_1^{-1}K_\omega^2\varepsilon^2. \tag{2.13}$$

For $-\|w\|^2_L$, similar to [1], we find

$$-\|w\|^2_L \geqslant -(1+æ_L\varepsilon)^2\|B^{-1}r_\beta\|^2_L \geqslant -\delta_1(1+æ_L\varepsilon)^2\|r_\beta\|^2_{B^{-1}}$$

which, combined with (2.13), yields the inequality:

$$2\gamma(w,r_\beta) - \gamma^2\|w\|^2_L \geqslant y(\gamma,\varepsilon,a_1)\|r_\beta\|^2_{B^{-1}} - K_1\varepsilon^2$$

where

$$K_1 \equiv 2\gamma K_\omega^2 a_1^{-1}, \qquad y(\gamma,\varepsilon,a_1) \equiv 2\gamma(1-\varepsilon æ_L\delta^{-1/2} - a_1) - \gamma^2\delta_1(1+æ_L\varepsilon)^2 \geqslant \bar{y} > 0$$

for $\gamma < 2\delta_1^{-1}$ and small ε and a_1. Hence,

$$\mu^n - \mu^{n+1} \geqslant \frac{\bar{y}\|r_\beta\|^2_{B^{-1}} + \gamma^2\mu\|w\|^2_M + 2\gamma(\beta-\mu)(u,w)_M}{1-2\gamma(u,w)_M + \gamma^2\|w\|^2_M} - K_1\varepsilon^2$$

$$\geqslant \frac{\bar{y}\|r_\beta\|^2_{B^{-1}} + \gamma^2\mu\|w\|^2_M + \beta - \mu}{1-2\gamma(u,w)_M + \gamma^2\|w\|^2_M} + \mu - \beta - K_1\varepsilon^2.$$

Let us also take into consideration that, in view of Lemma 2.1, $\|u^{n+1}\|_M = \|R_\beta u\|_M \geqslant K_0 > 0$. Therefore, setting $K_2 \equiv K_1/K_0$ we obtain

$$\beta - \mu^{n+1} + K_2\varepsilon^2 \geqslant \frac{\bar{y}\|r_\beta\|^2_{B^{-1}} + \gamma^2\mu\|w\|^2_M + \beta - \mu}{1-2\gamma(u,w)_M + \gamma^2\|w\|^2_M}. \tag{2.14}$$

Let us increase the denominator in the right-hand side of (2.14) by replacing the term $-2\gamma(u,w)_M$ with its upper bound $a\|u\|^2_M + a^{-1}\|w\|^2_M$, $a > 0$. Then, allowing for $\beta \geqslant \mu$, from (2.14) we have

$$\beta - \mu^{n+1} - K_2\varepsilon^2 \geqslant \min\left\{\frac{t(u)}{1+a}; \frac{\beta}{1+a^{-1}}\right\}.$$

We may assume that $r_\beta \neq 0$ and take $a = t(u)/\beta$. In view of this, we obtain lower bound maximal in a for $\beta - \mu^{n+1}$, which leads to (2.12). □

Theorem 2.1. Let ω_0, $\bar{Q}_{\omega_0}$ and $p > 1$ be such that we have $\lambda_i \geqslant \lambda_p$ for any $i \in \omega_0$ and $\Theta_L(Q;\bar{Q}) \leqslant \varepsilon$, $\Theta_M(Q;\bar{Q}) \leqslant \varepsilon$. Assume that we have an initial p-dimensional subspace $H_p^0 \subseteq Q^\perp$ with $\beta(H_p^0) \equiv \beta^0 \leqslant \bar{\beta} < \nu_p^\perp \equiv \nu$, where $\nu_p^\perp$ coincides with the eigenvalue λ_i, $i \notin \omega_0$, closest to $\lambda_p \equiv \lambda$ and strictly greater than λ. Let inequalities (1.3) and $0 < \gamma < 2\delta_1^{-1}$ be true. Let ε be sufficiently small and $\bar{\gamma}\delta_0 \leqslant 1$ (see Lemma 2.3). Assume that there have been defined the constants K_ω, K_1, and K_2 from Lemma 2.3, the same for all β belonging to $[\lambda_1, \bar{\beta}]$, and

$$K_3 \geqslant \frac{\bar{\beta}}{\lambda\nu} \sum_{\substack{i\in\omega_0 \\ \lambda\leqslant\lambda_i\leqslant\nu}} \frac{(\lambda_i - \lambda)(\nu - \lambda_i)}{\lambda_i}, \qquad K_4 \geqslant K_3\bar{\beta} + K_2(1 + K_3\varepsilon^2)$$

$$\rho(\beta) \equiv \frac{1 - \bar{\gamma}\delta_0(1 - \beta/\nu)}{1 + \bar{\gamma}\delta_0(\beta/\lambda - 1)(1 - \beta/\nu)}, \qquad q_p \equiv \max_{\lambda\leqslant\beta\leqslant\beta^0} \rho(\beta) < 1, \qquad K_5 \geqslant K_4(1 - q_p)^{-1}.$$

Then for iterative method (2.9) the following estimate is valid:

$$\beta^{n+1} - \lambda \leqslant \rho(\beta^n)(\beta^n - \lambda) + K_4\varepsilon^2 \tag{2.15}$$

and the method enables one to obtain $\beta^r \in [\lambda_p, \lambda_p + \varepsilon^2(K_4 + K_5)]$ in $s \equiv [\log_{q_p}(\varepsilon^2 K_4(\beta^0 - \lambda)^{-1})] + 1$ iterations for a certain value of $r \leqslant s$.

Proof. Substitute $u \equiv u^n \in H_p^n$ in (2.12) so that $\mu(u^{n+1}) = \beta(H_p^{n+1}) \equiv \beta^{n+1}$. Then, taking into account the monotonicity in t of the function $t(1 + at)^{-1}$ with $a > 0$, $t > 0$, from (2.12) we obtain

$$\beta^n - \beta^{n+1} \geqslant \varphi(1 + \beta^{-1}\varphi)^{-1} - K_2\varepsilon^2 \tag{2.16}$$

where

$$0 < \varphi \leqslant \min_{u\in H_p^n} t(u).$$

Let us determine the explicit form of φ.

First of all, note that, in view of (1.3), $\|r_\beta\|^2_{B^{-1}} \geqslant \delta_0\|r_\beta\|^2_{L^{-1}}$. The lower bound of the right-hand side of this inequality can be found using (2.11). Then, using the fact that $\lambda_k^\perp = \lambda_k$ for $k \leqslant p$, we obtain

$$t(u) \geqslant \bar{\gamma}\delta_0[\beta^2(\lambda_k + \nu_k - \mu)(\lambda_k\nu_k)^{-1} + \mu - 2\beta] + \beta - \mu \equiv g_k(\mu)$$

for

$$\lambda_k \leqslant \mu < \nu_k \leqslant \lambda_p \equiv \lambda$$

and

$$t(u) \geqslant \bar{\gamma}\delta_0[\beta^2(\lambda+\nu-\mu)(\lambda\nu)^{-1}+\mu-2\beta]+\beta-\mu-K_3\varepsilon^2 = g_p(\mu)-K_\beta\varepsilon^2$$

for $\mu \geqslant \lambda$ due to the assumption made about ω_0, the form of $\omega_{0,0}$ [see (2.11)], and the validity of inequalities (2.2). As

$$\bar{\gamma}\delta_0[1-\beta^2(\lambda_k\nu_k)^{-1}] < 1\,, \qquad g_k(\lambda_{k+1}) = g_{k+1}(\lambda_{k+1})$$

then

$$\min_{k\leqslant p}\ \min_{\mu\in[\lambda_k,\min(\nu_k,\beta)]} g_k(\mu) = g_p(\beta)$$

and

$$t(u) \geqslant \varphi \equiv \bar{\gamma}\delta_0\beta(1-\beta/\nu)(\beta/\lambda-1)-K_3\varepsilon^2$$

which, together with (2.16), leads to inequality (2.15).

It is easy to verify that $\beta^n \geqslant \bar{\lambda}_p^{\perp} \geqslant \lambda_p$. Besides, from (2.15) it follows that

$$\beta^n - \lambda \leqslant q_p^n(\beta^0-\lambda)+K_5\varepsilon^2$$

and, hence, the last s-th iteration will have the accuracy required. In fact, such an accuracy could have been reached previously, since there could be a violation in the monotonicity of decrease of β^n, which could happen only for $\beta^r \in [\lambda_p, \lambda_p + K_5\varepsilon^2)$ [see (2.15)]. It is easily seen that for small ε the best value of γ is $\gamma \approx \delta_1^{-1}$, since with this value we obtain the least values of $\rho(\beta)$ and q_p.

In the particular case, when $\varepsilon = 0$ ($P = \bar{P}$) or when ω_0 is empty, the investigation of the method is significantly simplified. Then it can be proved [3] that for $\gamma = \delta_1^{-1}$ the inequalities $\beta^n - \lambda_p \leqslant \sigma(n)(\beta^0-\lambda_p)$ hold, where

$$\sigma(n) \equiv \prod_{i=0}^{n-1}\rho(\beta^i) \leqslant q_p^n$$

$$\rho(\beta) = \frac{1-\delta(1-\beta/\nu_p)}{1+\delta(\beta/\lambda_p-1)(1-\beta/\nu_p)}\,, \qquad \delta \equiv \frac{\delta_0}{\delta_1}\,, \qquad q_p \equiv \max\{\rho(\beta^0),\rho(\lambda_p)\}\,.$$

3. DETERMINATION OF U_{λ_p}

Lemma 3.1. Let $H_p^r \subseteq \bar{Q}^{\perp}$ be a subspace such that, in view of Theorem 2.1, the following estimate is valid:

$$0 \leqslant \beta^r - \lambda_p \leqslant K_6\varepsilon^2 \tag{3.1}$$

and $\lambda_i \geqslant \beta^r$ for any $i \in \omega_0$. Let p' be a dimension of a direct sum of the eigen subspaces U_{λ_i} of problem (2.1), provided $\lambda_i < \lambda_p$. Then, in H_p^r there exists a $(p-p')$-dimensional subspace $\tilde{U}^r$ such that

$$\Theta_M(\tilde{U}^r; U_{\lambda_p}) \leqslant [(\beta^r-\lambda_p)(\nu_p^{\perp}-\lambda_p)^{-1}+m\varepsilon^2]^{1/2} \leqslant K_7\varepsilon \tag{3.2}$$

where $K_7 \equiv [K_6(\nu_p^{\perp}-\lambda_p)^{-1}+m]^{1/2} \geqslant 1$.

Proof. Take $\widetilde{U}^r \equiv H_p^r \cap L(u_{p'+1},\ldots,u_N)$. Then, for any $u \in \widetilde{U}^r$ [see (2.2)] with $\|u\|_M = 1$, we have

$$u = \sum_{i=1}^{N} c_i u_i, \qquad \sum_{i=p'+1}^{N} c_i^2(\lambda_i - \mu) = 0$$

where $\mu \equiv \mu(u) \in [\lambda_p, \beta^r]$. Therefore,

$$\sum_{i\in\omega_2} (\lambda_i - \mu)c_i^2 = \sum_{i=p'+1}^{p} c_i^2(\mu - \lambda_p) + \sum_{i\in\omega_0} (\mu - \lambda_i)c_i^2$$

where $\omega_2 \equiv \{p+1,\ldots,N\}\backslash\omega_0$. Take into account that $\lambda_i - \mu \geqslant \nu_p^{\perp} - \beta^r$ for $i \in \omega_2$ and that $\mu - \lambda_i \leqslant 0$ for $i \in \omega_0$. Then

$$\sum_{i\in\omega_2} c_i^2 \leqslant (\nu_p^{\perp} - \beta^r)^{-1}(\beta^r - \lambda_p)$$

and

$$\rho_M^2(u;U_{\lambda_p}) = \sum_{i\in\omega_0\cup\omega_2} c_i^2 \leqslant (\beta^r - \lambda_p)(\nu_p^{\perp} - \beta^r)^{-1} + m\varepsilon^2 \tag{3.3}$$

since

$$\sum_{i\in\omega_0} c_i^2 \leqslant m\varepsilon^2 .$$

Relation (3.3), together with (2.3) and (3.1), gives (3.2). □

For H_p^r consider the numbers $\lambda_1^r \leqslant \ldots \leqslant \lambda_p^r$ [see (2.6)], $\lambda_i^r \geqslant \lambda_i$, $\lambda_p \leqslant \lambda_{p'+1}^r + 1 \leqslant \ldots \leqslant \lambda_p^r = \beta^r$.

Theorem 3.1. Provided (3.1) is true, for $\lambda_{p'}^{r+1}$ the following estimate holds for sufficiently small ε:

$$\lambda_{p'}^{r+1} \leqslant \beta^r[1 + \varphi_1/\beta^r]^{-1} + K_2\varepsilon^2 \tag{3.4}$$

where

$$\varphi_1 \equiv \overline{\gamma}\delta_0\beta^r(1 - \beta^r/\nu_p^{\perp})(\beta^r/\lambda_{p'} - 1) - K_3\varepsilon^2 > 0 . \tag{3.5}$$

Proof. Let

$$T^r \equiv H_p^r \cap L(u_1,\ldots,u_{p'},u_{p+1},\ldots,u_N), \qquad T^{r+1} \equiv R_{\beta^r}T^r$$

$$\dim T^r = \dim T^{r+1} = p'$$

and $u \in T^r$, with $\|u\|_M = 1$, be such that

$$\mu(R_{\beta^r}u) \equiv \mu^{r+1} = \beta(T^{r+1}) \geqslant \lambda_{p'}^r \geqslant \lambda_{p'} .$$

Then, in view of (2.12),

$$\begin{aligned} \lambda_{p'}^{r+1} &\leqslant \mu^{r+1} \leqslant \beta^r + K_2\varepsilon^2 - t(u)[1 + t(u)/\beta^r]^{-1} \\ &\leqslant \beta^r + K_2\varepsilon^2 - \widetilde{\varphi}_1[1 + \widetilde{\varphi}_1/\beta^r]^{-1} = \beta^r[1 + \widetilde{\varphi}_1/\beta^r]^{-1} + K_2\varepsilon^2 \end{aligned} \tag{3.6}$$

where $\tilde{\varphi}_1 \geqslant 0$ is the lower bound for $t(u)$, $u \in T^r$, $\|u\|_M = 1$. Let us find the expression for $\tilde{\varphi}_1$. If we drop the values $\lambda_i^{\perp}$ with $i = p' + 1,\ldots,p$, from the set $\lambda_1^{\perp},\ldots,\lambda_{N-m}^{\perp}$ [see (2.5)] then, as $T^r \subseteq \bar{Q}^{\perp} \cap L^{\perp}(u_{p'+1},\ldots,u_p)$, inequality (2.11) is valid and the same reasoning, as used in the proof of Theorem 2.1, applies to the lower bound of $t(u)$. As a result, we obtain: $t(u) \geqslant \tilde{\varphi}_1 = \varphi_1$. Combining the lower bound with (3.6) we obtain the required inequality (3.4). From inequality (3.4) it follows that for sufficiently small ε we have

$$\lambda_{p'}^r < \lambda_p \leqslant \lambda_{p'+1}^r \leqslant \ldots \leqslant \lambda_p^r = \beta^r \leqslant \lambda_p + K_6 \varepsilon^2 .$$

Therefore, $\lambda_{p'+1}^r,\ldots,\lambda_p^r$, with an accuracy of $K_6\varepsilon^2$ can be combined into a separate group that is a finite distance apart from $\lambda_{p'}^r$. Thus, the number p' can be determined from the results of the calculations.

Theorem 3.2. Let the conditions of Lemma 3.1 be satisfied and $\lambda_{p'}^r < \lambda_p$ [see (3.6)]. Let problem (2.6) correspond to H_p^r and $\bar{X}^{(i)}$, $i = 1,\ldots,p$, be the eigenvectors of (2.6) orthonormalized in accordance with the inner product $(\bar{Y},\bar{X})_{\bar{M}_p^r} \equiv \bar{Y}'\bar{M}_p^r\bar{X}$. Let these eigenvectors be put in correspondence with the vectors $v_i^r \equiv v_{i_r}$ [see (2.7)] forming an orthonormal in H_M basis of H_p^r. Let U^r be a linear span of the vectors $v_{p'+1},\ldots,v_p$. Then

$$\Theta_M(U^r;U_{\lambda_p}) \leqslant K_7\varepsilon + (\beta^r - \lambda_p)^{1/2}(\beta^r - \lambda_{p'}^r)^{-1/2} \leqslant K_8\varepsilon . \tag{3.7}$$

Proof. Let us estimate $\Theta_M(\tilde{U}^r;U^r)$, where $\tilde{U}^r$ has been defined in Lemma 3.1. Both these $(p - p')$-dimensional subspaces belong to H_p^r. To estimate $\rho_M(z;U^r)$, where $x \in \tilde{U}^r$ and $\|z\|_M = 1$, we perform an isometric in H_M transition from H_p^n to the p-dimensional Euclidean space $\bar{H}_{\bar{M}}$ consisting of $\bar{X} \equiv (x_1,\ldots,x_p)^{\mathrm{T}}$ with the inner product $(\bar{Y},\bar{X})_{\bar{M}}$, where $\bar{M} \equiv \bar{M}_p^r$ [see (2.6)]. In this case, if

$$z = \sum_{i=1}^{p} z_i v_i$$

then

$$\bar{Z} = \sum_{i=1}^{p} z_i \bar{X}^{(i)}, \qquad \rho_M^2(z;U^r) = \sum_{i=1}^{p'} z_i^2 .$$

Since

$$\sum_{i=1}^{p} z_i^2(\lambda_i^r - \mu) = 0, \qquad \mu \equiv \mu(z) \geqslant \lambda_p, \qquad \lambda_i^r \geqslant \lambda_i$$

[see (2.4) and (2.8)], we have

$$\sum_{i=1}^{p'} z_i^2(\mu - \lambda_i^r) = \sum_{i=p'+1}^{p} z_i^2(\lambda_i^r - \mu)$$

and

$$(\mu - \lambda_{p'}^r)\sum_{i=1}^{p'} z_i^2 \leqslant (\beta^r - \mu)\left(1 - \sum_{i=1}^{p'} z_i^2\right).$$

Hence,

$$\rho_M^2(z;U^r) \leqslant (\beta^r - \mu)(\beta^r - \lambda_{p'}^r)^{-1} \leqslant (\beta^r - \lambda_p)(\beta^r - \lambda_{p'}^r)^{-1}$$

and

$$\Theta_M(\tilde{U}^r;U^r) \leqslant (\beta^r - \lambda_p)^{1/2}(\beta^r - \lambda_{p'}^r)^{-1/2}. \tag{3.8}$$

As the opening of the subspaces possesses the properties of a metric, then inequality (3.7) follows from (3.2) and (3.8). Now it is easy to estimate $\Theta_L(U^r;U_{\lambda_p})$, since for any $u \in U_{\lambda_p}$ with $\|u\|_M = 1$ and for any v with $\|v\|_M = 1$, we have the equality:

$$\|v - u\|_L^2 = \mu(v) - \lambda_p + \lambda_p\|u - v\|_M^2.$$

Therefore

$$\Theta_L(U^r;U_{\lambda_p}) = \{\beta^r - \lambda_p + \lambda_p\Theta_M^2(U^r;U_{\lambda_p})\}^{1/2} \leqslant K_9\varepsilon. \tag{3.9}$$

Having found U_{λ_p} with an accuracy of $\mathcal{O}(\varepsilon)$ [see (3.7) and (3.8)], we are in a position to change the structure of ω_0 and $Q_{\omega_0}^{\perp}$ and to turn to the determination of $\lambda_{p'}$ and $U_{\lambda_{p'}}$. By virtue of (3.4), we may take the linear span of the vectors $v_1,\ldots,v_{p'}$ belonging to H_p^r as an initial iteration $H_{p'}^0$. If $\lambda_{p'}^r$ is sufficiently far away from $\lambda_{p'+1}^r \approx \lambda_p$, orthogonalization can turn out to be unnecessary. In case $\lambda_{p'}^r$ differs from $\lambda_{p'+1}^r$ only slightly, the use of orthogonalization with respect to U^r can be an essential factor.

Inserting the indices $p'+1,\ldots,p$ into ω_0 with possible elimination of some higher indices enables one to obtain the convergence rate estimates of such iterative methods no worse than the given estimates for the case of finding λ_p and U_{λ_p} considered above. Thus, it provides a possibility of determining λ_i and U_{λ_p} with the accuracy required for all $i \leqslant p$, given, for example, an initial iteration H_p^0 with $\beta(H_p^0) < \lambda_{p+1}$.

4. NUMERICAL EXPERIMENTS

As previously mentioned, the most important applications of modified gradient methods are connected with the solution of grid eigenvalue problems that approximate, for example, the corresponding ones with elliptic operators. In order to carry out a larger number of numerical experiments on the study of the effect of various conditions and parameters on the convergence, while preserving reasonable computational costs, the calculations were performed only for model problems. These were the simplest difference schemes on a uniform grid with the step $h \equiv \pi/(N+1)$ for the problems of the type:

$$-(a(x)u'(x))' = \lambda q(x)u(x), \quad x \in (0,\pi), \quad u(0) = u(\pi) = 0.$$

The problems were of the form: $u_0 = u_{N+1} = 0$,

$$-h^{-2}[a_{i+1/2}(u_{i+1} - u_i) - a_{i-1/2}(u_i - u_{i-1})] = \lambda q_i u_i, \quad i = 1,\ldots,N \tag{4.1}$$

and, on eliminating U_0 and U_{N+1}, they were reduced to standard algebraic problems (1.1). The operator L with $a \equiv 1$ defined the operator B. For the examples considered, we took L with $a(x) \in (0,1]$ and the constants δ_0 and δ_1 were determined from the formulae:

$$\delta_0 = \min_{x \in [0,1]} a(x), \qquad \delta_1 = 1.$$

For this reason, the iterative parameter γ was always taken equal to $\gamma = \delta_1^{-1} = 1$. The main attention was paid to the problem of determining λ_1 and λ_2 with an accuracy of ε by two methods: the conventional modified gradient method (method 1) and the group method with $p = 2$ (method 2). With both methods, the initial approximations were determined by the same subspace H_2^0. In method 1 the initial approximation was usually u_1^0 with $\|u_1^0\|_M = 1$ and $\mu(u_1^0) = \min_{u \in H_2^0} \mu(u)$. Then we performed the iterations:

$$\bar{u}^{n+1} = u^n - B^{-1}(Lu^n - \mu(u^n)Mu^n), \qquad u^{n+1} = \bar{u}^{n+1}\|\bar{u}^{n+1}\|_M^{-1} \tag{4.2}$$

with $u^0 \equiv u_1^0$ until we obtained the inequality:

$$\mu(u^{k_1}) - \mu(u^{k_1+1}) \leqslant \varepsilon. \tag{4.3}$$

Then at the second stage of the method, for u_2^0 with $\|u_2^0\|_M = 1$ and $\mu(u_2^0) = \beta(H_2^0)$, we constructed a new approximation:

$$u^0 = \bar{u}^0\|\bar{u}^0\|_M^{-1}, \qquad \bar{u}^0 \equiv u_2^0 - (u_2^0, \bar{u}_1)_M \bar{u}_1$$

with $\bar{u}_1$ coinciding with previously determined u^{k_1+1}. After this, we performed next k_2 iterations:

$$\bar{u}^{n+1} = u^n - \bar{P}^{\perp}(Lu^n - \mu(u^n)Mu^n), \qquad u^{n+1} = \bar{u}^{n+1}\|\bar{u}^{n+1}\|_M^{-1}$$

with orthogonalization with respect to $\bar{u}_1$ until we obtained an inequality of the type of (4.3) for a number k_2. The value u^{k_2+1} obtained was taken equal to $\bar{u}_2$.

In the group method 2, (without orthogonalization) the set ω_0 was taken empty and we performed iterations until we obtained the inequality:

$$\beta(H_2^{r_1}) - \beta(H_2^{r_1+1}) < \varepsilon.$$

Then the vector $\bar{u}_2 \in H_2^{r_1+1}$ with $\|\bar{u}_2\|_M = 1$ and $\mu(\bar{u}_2) = \beta(H_2^{r_1+1})$ yielded an orthogonal in H_M complement $\bar{Q}^{\perp}$. Then method (4.2) with orthogonalization with respect to $\bar{u}_2$, and with the initial iteration $u^0 \in H_2^{r_1+1}$ and

$$\mu(u^0) = \min_{u \in H_2^{r_1+1}} \mu(u)$$

was applied. With this method, r_2 iterations were required for an inequality of the form of (4.3) to be obtained. In this case the computational costs can be estimated by the number $k_1 + k_2$ for method 1 and $2r_1 + r_2$ for method 2.

For $N = 10$ and $N = 20$ the results were compared with the available information on eigenvalues and eigenvectors obtained by using the conventional program of the method of orthogonal rotations. The subspace H_2^0 usually coincided with the linear span of the grid functions $\sin x$ and $\sin 2x$.

Let us give some conclusions and illustrations.

(1) The choice of an initial approximation that does not obey the conditions of the type of $\beta^0 < \lambda_3$ as a rule did not prevent us from obtaining the convergence to the required values λ_1 and λ_2, when using both method 1 and method 2. For some highest eigenvalues λ_i we did sometimes observe the convergence slowing-down. The iterative process in the vicinity of λ_3 came to a halt only when the initial iteration virtually coincided with the linear span of certain eigenfunctions. For example, Fig. 1 demonstrates the behaviour of $\mu(u^n)$ at the first stage of method 1 (determination of λ_1), while Fig. 2 shows the behaviour of β^n at the first stage of method 2 (determination of λ_2) for problem (4.1) with $N = 10$, $a(x) = 10^{-3}$ for $0 \leqslant x \leqslant 1$ and $a(x) = 1$ for $1 \leqslant x \leqslant \pi$, $q \equiv 1$, $\delta_0 = 10^{-3}$, $\varepsilon = 10^{-4}$, and $\lambda_1 = 0.0072$, $\lambda_2 = 0.0245$, $\lambda_3 = 0.0418$, $\lambda_4 = 0.5391$, $\lambda_5 = 4.686$, $\beta(H_2^0) = 2.91$, $r_1 = 81$, $r_2 = 3$, $k_1 = 27$, $k_2 = 69$.

From Fig. 2 it is seen that the convergence becomes slower in the vicinity of the point λ_4. In Figs. 1, 2 for simplicity, instead of functions defined on the integers, continuous functions are depicted.

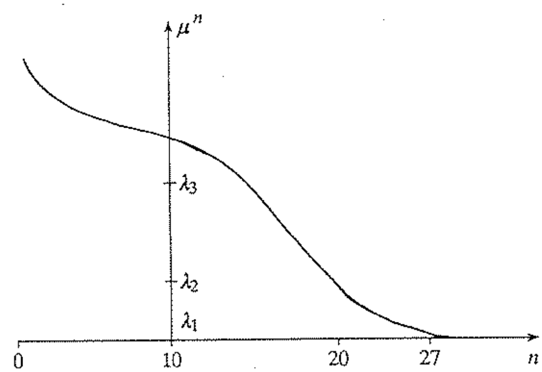

Figure 1.

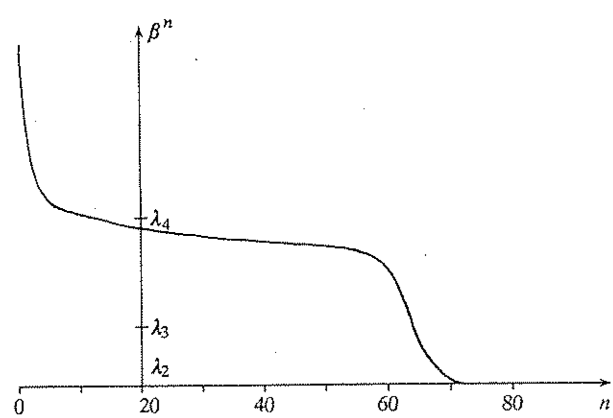

Figure 2.

The most aggravated situation (we did this intentionally) occurred for methods 1 and 2 with a special choice of H_2^0 coinciding with the linear span of $u_1^0 = u_3 + u_6$ and $u_2^0 = u_4 + u_5$, where u_i denote the corresponding eigenfunctions. In this case we had $a(x) = 10^{-2}$ on the segment $[0,1]$ and $a(x) = 1$ on the segment $(1,\pi]$, $q = 1$, $\delta_0 = 10^{-2}$, $\varepsilon = 10^{-4}$, $\lambda_1 = 0.06952$, $\lambda_2 = 0.2387$, $\lambda_3 = 0.4126$, $\lambda_4 = 0.5814$, and $\lambda_5 = 4.7133$. An example of the dependence of μ^n on n method 1 gave under these conditions is provided in Table 1.

Method 1, stopped after 100 iterations, gave a rather crude approximation to λ_1 ($\mu^{100} = 0.0890$). Method 2 after 100 iterations at the first stage gave quite an accurate approximation to λ_3 (instead of λ_2) while at the second stage as little as 16 iterations were required to obtain $\mu^{16} = 0.0699$ close to λ_1.

(2) Almost in all the experiments it was noted that the strongest decrease in β^n and μ^n happened at several first iterations. Frequently, this change at the first 3–5 iterations accounted for as much as 90% of the total initial error.

(3) While applying method 2, we noticed that, in general, λ_1^n changed non-monotonically. However, the last iterations at the second stage gave as a rule very good approximations to u_1. Therefore, the second stage of the method usually required a significantly smaller number of iterations (sometimes, 10 times smaller) than at the first stage.

(4) The effect of the value of ε on the required number of iterations in methods 1 and 2 was not profound. A 100-fold decrease in ε often called for mere doubling of the number of iterations. For example, in the problem with $a(x) = 1 - x/\pi$, $q \equiv 1$, and $\lambda_1 \approx 0.29$, $\lambda_2 \approx 1.19$, $\lambda_3 \approx 2.66$, $\lambda_4 \approx 7.25$, $\lambda_5 \approx 10.5$ method 1 required $k_1 = 17$ and $k_2 = 26$ iterations ($\mu^0 = 0.5$) for $\varepsilon = 10^{-3}$ and it required $k_1 = 27$ and $k_2 = 32$ iterations for $\varepsilon = 10^{-5}$ and at the worst initial value $\mu^0 \approx 3$. In the same situation, method 2 required respectively $r_1 = 26$, $r_2 = 3$ and $r_1 = 48$, $r_2 = 8$.

(5) The calculations were performed with the values of δ_1/δ_0 varying between 2 and 1000. An increase in the value of δ_1/δ_0 was not profound either.

(6) In all the above examples method 2 was used in the situations when the determination of λ_2 was a more difficult problem than the determination of λ_1. For instance, in the example from item 4 we have $\lambda_2/\lambda_1 \approx 4$, but $\lambda_3/\lambda_2 \approx 2$. Probably, it is this fact that was a reason why method 2 usually had no advantages over method 1. Recall that method 2 is, first of all, aimed at the cases when $\lambda_1 \approx \lambda_1 \ll \lambda_3$. Presumably, the method may be of particular assistance when one has to separate λ_1 and λ_2 (u_1 and u_2) given a relatively good approximation to $U_{\lambda_1} + U_{\lambda_2}$. Evidently, it would be reasonable to obtain such an approximation using method 1.

Table 1.

Stage	n	0	4	10	30	50	70	90	100
1	μ^n	6.35	0.36	0.297	0.173	0.13	0.107	0.0925	0.0890
2	μ^n	2.65	0.573	0.572	0.562	0.55	0.53	0.482	0.4148

REFERENCES

1. E. G. D'yakonov, *Minimization of Computational Work. Asymptotically Optimal Algorithms for Elliptic Problems*. Nauka, Moscow, 1989 (in Russian).
2. E. G. D'yakonov and A. V. Knyazev, On an iterative method for finding lower eigenvalues. *Preprint No. 26*, Dept. Numer. Math., USSR Acad. Sci., 1981 (in Russian).
3. E. G. D'yakonov and A. V. Knyazev, A group iterative method for finding lower-order eigenvalues. *Moscow University, Ser. 15, Computational Mathematics and Cybernetics* (1982) 2, 32–40.
4. A. V. Knyazev, Convergence rate estimates for iterative methods for a grid symmetric eigenvalue problem. *Sov. J. Numer. Anal. Math. Modelling* (1987) 2, No. 5, 371–396.