J. KSIAM **Vol.5**, No.2, 17-23, 2001

PERTURBATION ANALYSIS OF DEFLATION TECHNIQUE FOR SYMMETRIC EIGENVALUE PROBLEM

HO-JONG JANG

ABSTRACT. The evaluation of a few of the smallest eigenpairs of large symmetric eigenvalue problem is of great interest in many physical and engineering applications. A deflation-preconditioned conjugate gradient(PCG) scheme for a such problem has been shown to be very efficient. In the present paper we provide the numerical stability of a deflation-PCG with partial shifts.

1. Introduction

In this paper, we are concerned with the perturbation analysis of the deflation-PCG scheme with partial shifts for computing a few of the smallest eigenvalues and their corresponding eigenvectors of the generalized eigenvalue problem. The partial eigenanalysis of large sparse symmetric matrices is a common task in many scientific applications, e.g. structural mechanics [1], hydrodynamics [5], and plasma physics [12].

Several techniques have been developed for the solution of the partial eigenproblem, including subspace iteration [1], Lanczos scheme [3], and multigrid [8]. A preconditioned conjugate gradient(PCG) method based on the optimization of successive deflated Rayleigh quotients also works well for such a problem [5,7,12], and proves to be competitive with respect to other more commonly used schemes, in particular with respect to the Lanczos algorithm when the dimension of the eigenproblem is large [6].

Two different types of deflation techniques, which employ a PCG method to minimized the Rayleigh quotient, are typically used for computing a few of the smallest eigenpairs. Those are deflation-PCG with partial shifts [5,13,14] and an orthogonal deflation-PCG [7].

In [13], Schwartz proposed the numerical stability of the deflation-PCG with partial shifts. Here we continue this study for general updating procedures.

2. Minimization of Rayleigh Quotients via PCG Scheme

Consider the generalized eigenvalue problem

$$Ax = \lambda Bx, \tag{1}$$

Key words: symmetric eigenproblem, preconditioned conjugate gradients, deflation

where A and B are large sparse symmetric positive definite matrices of dimension n. Let

$$0 < \lambda_1 < \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_n$$

be the eigenvalues (1), and let $z_1, z_2, \cdots, z_n$ be the corresponding eigenvectors, which satisfy

$$Az_i = \lambda_i B z_i, \ \ {z_i}^T B z_i = 1, \ \ i = 1, 2, \cdots, n.$$

The eigenvectors of (1) are the stationary points of the Rayleigh quotient

$$R(x) = \frac{x^T Ax}{x^T Bx}, \tag{2}$$

and the gradient of $R(x)$ is given by

$$g(x) = \frac{2}{x^T Bx}[Ax - R(x)Bx].$$

For an iterate $x^{(k)}$, the gradient of $R(x^{(k)})$,

$$\nabla R(x^{(k)}) = g^{(k)} = g(x^{(k)}) = \frac{2}{{x^{(k)}}^T Bx^{(k)}}\left[Ax^{(k)} - R(x^{(k)})Bx^{(k)}\right],$$

is used to fix the direction of descent $p^{(k+1)}$ in which $R(x)$ is minimized. These directions of descent are defined by

$$p^{(1)} = -g^{(0)}, \ \ p^{(k+1)} = -g^{(k)} + \beta^{(k)} p^{(k)}, \ \ k = 1, 2, \cdots,$$

where $\beta^{(k)} = \dfrac{{g^{(k)}}^T g^{(k)}}{{g^{(k-1)}}^T g^{(k-1)}}$ [11]. The subsequent iterate $x^{(k+1)}$ along $p^{(k+1)}$ through $x^{(k)}$ is written as

$$x^{(k+1)} = x^{(k)} + \alpha^{(k+1)} p^{(k+1)}, \ \ k = 0, 1, \cdots,$$

where $\alpha^{(k+1)}$ is obtained by minimizing $R(x^{(k+1)})$ [9],

$$R(x^{(k+1)}) = \frac{{x^{(k)}}^T Ax^{(k)} + 2\alpha^{(k+1)}{p^{(k+1)}}^T Ax^{(k)} + {\alpha^{(k+1)}}^2 {p^{(k+1)}}^T Ap^{(k+1)}}{{x^{(k)}}^T Bx^{(k)} + 2\alpha^{(k+1)}{p^{(k+1)}}^T Bx^{(k)} + {\alpha^{(k+1)}}^2 {p^{(k+1)}}^T Bp^{(k+1)}}.$$

The performance of the CG scheme can be improved by using a preconditioner [2,4]. The idea behind the PCG is to apply the "regular" CG scheme to the transformed system

$$\tilde{A}\tilde{x} = \lambda \tilde{B}\tilde{x},$$

where $\tilde{A} = C^{-1}AC^{-1}$, $\tilde{B} = C^{-1}BC^{-1}$, $\tilde{x} = Cx$, and C is nonsingular symmetric matrix. By substituting $x = C^{-1}\tilde{x}$ into (2), we obtain

$$R(\tilde{x}) = \frac{\tilde{x}^T C^{-1}AC^{-1}\tilde{x}}{\tilde{x}^T C^{-1}BC^{-1}\tilde{x}} = \frac{\tilde{x}^T \tilde{A}\tilde{x}}{\tilde{x}^T \tilde{B}\tilde{x}}, \tag{3}$$

where the matrices $\tilde{A}$ and $\tilde{B}$ are symmetric positive definite. The transformation (3) leaves the stationary values of (2) unchanged, which are eigenvalues of (1), while the corresponding stationary points are obtained from $\tilde{x}_j = Cz_j, \ j = 1, 2, \cdots, n$.

3. Perturbation Analysis of Higher Eigenvalue Computation

3.1. **Deflation-PCG with partial shifts.** In most applications not only the smallest but some of the smallest stationary values of the Rayleigh quotient are wanted. The PCG scheme in §2 can be modified using a deflation based on a partial shift of the spectrum, so that the next higher eigenvalues can be computed by essentially the same process.

When the first $r-1$ eigenpairs are approximately known, the next eigenpair (λ_r, z_r) could be obtained by minimizing the Rayleigh quotient $R(x)$ of the modified eigenproblem $A_r x = \lambda Bx$, where A_r is defined by

$$A_r = A + \sum_{k=1}^{r-1} \sigma_k (Bz_k)(Bz_k)^T, \tag{4}$$

with σ_k is the shift that satisfies $\sigma_k > 0$ and $\lambda_k + \sigma_k > \lambda_r,\ k = 1, 2, \cdots, r-1$.

It is clear that the eigenvalues and eigenvectors of $A_r x = \lambda Bx$ satisfy, because of the B-orthonormality of the z_j,

$$\begin{aligned} A_r z_j &= Az_j + \sum_{k=1}^{r-1} \sigma_k (Bz_k)(Bz_k)^T z_j \\ &= \begin{cases} (\lambda_j + \sigma_j) Bz_j, & j = 1, 2, \cdots, r-1; \\ \lambda_j Bz_j, & j = r, r+1, \cdots, n. \end{cases} \end{aligned}$$

The eigenpair (λ_r, z_r) could then be determined from the PCG in §2 by replacing A by A_r.

In the proposed method we assume that the shifts σ_i are chosen properly. Some ways of deterimining the shifts σ_i are reported in [14].

If the preconditioner M is kept fixed for minimizing the Rayleigh quotient of the modified eigenproblem $A_r x = \lambda Bx$, the preconditioning effect is lost for increasing r in general. Thus it is necessary to use an equivalent preconditioner for the matrix A_r that takes into account the deflation steps [13].

3.2. **Numerical stability.** In this section we present a numerical stability of the deflation process (4). We first cite the theorem in [10]. It provides a error bound on Ritz value which approximates a eigenvalue of the symmetric eigenvalue problem.

Lemma 3.1. *Let A be a symmetric matrix with eigenpairs $(\lambda_i,\ z_i)$. Let y be a 2-normalized vector with $\theta = y^T Ay$ and residual $r(y) = Ay - \theta y$. Let λ be the eigenvalue of A closest to θ, let z be its 2-normalized eigenvector, and let $\psi = \angle(y, z)$. Then*

$$|\sin\psi| = \frac{\|r(y)\|_2}{d} \quad \text{and} \quad |\theta - \lambda| \le \frac{\|r(y)\|_2^2}{d},$$

where $d = \min|\lambda_i - \lambda|$ over all $\lambda_i \ne \lambda$.

The straightforward extention of Lemma 3.1, with the appropriate pair of vector norms $\|x\|_B = \sqrt{x^T Bx}$ and $\|x\|_B^{-1} = \sqrt{x^T B^{-1} x}$, to the generalized eigenvalue problem yields the following theorem [13].

THEOREM 3.2. *Let A and B be symmetric matrices and B positive definite and $(\lambda_i,\ z_i)$ be the eigenpairs of $Ax = \lambda Bx$. Let x be a B-normalized vector with $\theta = x^T Ax$ and the residual $r(x) = Ax - \theta Bx$. Let λ be the eigenvalue of the matrix pair (A, B) closest to θ, let z be its B-normalized eigenvector, and let $\psi = \angle(x, z)$. Then*

$$|\sin\psi| = \frac{\|r(x)\|_{B^{-1}}}{d} \quad \text{and} \quad |\theta - \lambda| \le \frac{\|r(x)\|_{B^{-1}}^2}{d}, \tag{5}$$

where $d = \min|\lambda_i - \lambda|$ over all $\lambda_i \neq \lambda$.

For the assumption that an approximation $\hat{z}_k$ of the eigenvector z_k has been determined with a relative accuracy ε_k and being B-normalized, the approximations $\hat{z}_k$ can be expressed as

$$\hat{z}_k = c_k{}^{(k)} z_k + \varepsilon_k \sum_{\substack{i=1\\ i\neq k}}^{n} c_i{}^{(k)} z_i \text{ with } \| \sum_{\substack{i=1\\ i\neq k}}^{n} c_i{}^{(k)} z_i \|_B = 1,\ k = 1, \cdots, r-1. \tag{6}$$

Here the coefficients $c_k{}^{(k)}$ satisfy

$$c_k{}^{(k)2} + \varepsilon_k^2 (\sum_{\substack{i=1\\ i\neq k}}^{n} c_i{}^{(k)2}) = c_k{}^{(k)2} + \varepsilon_k^2 = 1 \text{ and } c_k{}^{(k)} \cong 1 - \tfrac{1}{2}\varepsilon_k^2.$$

To make the statements below neatly, we define ε and c_r as

$$|\varepsilon| = \max_{1\le k\le r-1} |\varepsilon_k|,\ |c_r| = \max_{1\le k\le r-1} |c_r^{(k)}|. \tag{7}$$

We now show the influence of the approximations $\hat{z}_k,\ k = 1, \cdots, r-1,$ to the next higher eigenvalue λ_r.

THEOREM 3.3. *Let (λ_r, z_r) be the eigenpair of the matrix A_r in (4), and let $\hat{z}_k$ be the approximations of the eigenvectors z_k, for $k = 1, \cdots, r-1$, as in (6). And let $\hat{\lambda}_r$ be the computed eigenvalue of $\hat{A}_r = A + \sum_{k=1}^{r-1} \sigma_k (B\hat{z}_k)(B\hat{z}_k)^T$ with the same shifts σ_k in (4). Let ε and c_r be defined as in (7). Then*

$$|\hat{\lambda}_r - \lambda_r| \le \frac{1}{d_r} \varepsilon^2 c_r^2 \sum_{k=1}^{r-1} \sigma_k^2, \tag{8}$$

where $d_r = \min_{i \neq r} |\hat{\lambda}_r - \hat{\lambda}_i|$ *and* $\hat{\lambda}_i s$ *are all eigenvalues computed from* $\hat{A}_r$*.*

Proof. We have

$$
\begin{aligned}
\hat{A}_r &= A + \sum_{k=1}^{r-1} \sigma_k (B\hat{z}_k)(B\hat{z}_k)^T \\
&= A + \sum_{k=1}^{r-1} \sigma_k ({c_k}^{(k)} B z_k + \varepsilon_k \sum_{\substack{i=1 \\ i \neq k}}^{n} {c_i}^{(k)} B z_i)({c_k}^{(k)} B z_k + \varepsilon_k \sum_{\substack{j=1 \\ j \neq k}}^{n} {c_j}^{(k)} B z_j)^T \\
&= A + \sum_{k=1}^{r-1} \sigma_k {c_k^{(k)}}^2 (B z_k)(B z_k)^T \\
&\quad + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k c_k^{(k)} \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} {c_i}^{(k)} \left\{ (B z_k)(B z_i)^T + (B z_i)(B z_k)^T \right\} \right] \\
&\quad + \sum_{k=1}^{r-1} \sigma_k {\varepsilon_k}^2 \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} {c_i}^{(k)} {c_j}^{(k)} (B z_i)(B z_j)^T \right] \\
&= A + \sum_{k=1}^{r-1} \sigma_k (1 - {\varepsilon_k}^2)(B z_k)(B z_k)^T \\
&\quad + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k c_k^{(k)} \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} {c_i}^{(k)} \left\{ (B z_k)(B z_i)^T + (B z_i)(B z_k)^T \right\} \right] \\
&\quad + \sum_{k=1}^{r-1} \sigma_k {\varepsilon_k}^2 \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} \sum_{\substack{j=1 \\ j \neq k}}^{n} {c_i}^{(k)} {c_j}^{(k)} (B z_i)(B z_j)^T \right] \\
&= A_r + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} {c_i}^{(k)} \left\{ (B z_k)(B z_i)^T + (B z_i)(B z_k)^T \right\} \right] + O(\varepsilon^2)
\end{aligned}
$$

Now, we get the Ritz value $\theta_r = z_r^T \hat{A}_r z_r$ and the residual $r(z_r) = \hat{A}_r z_r - \theta B z_r$ by applying Theorem 3.2 with $A = \hat{A}_r$ and $x = z_r$. We first consider

$$\begin{aligned}
\hat{A}_r z_r &= A_r z_r + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k \left[\sum_{\substack{i=1 \\ i \neq k}}^{n} {c_i}^{(k)} \left\{ (Bz_k)(Bz_i)^T + (Bz_i)(Bz_k)^T \right\} \right] z_r + O(\varepsilon^2) \\
&= A_r z_r + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) + O(\varepsilon^2),
\end{aligned}$$

and get

$$\begin{aligned}
\theta_r = {z_r}^T \hat{A}_r z_r &= {z_r}^T A_r z_r + {z_r}^T \{ \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) \} + O(\varepsilon^2) \\
&= \lambda_r + O(\varepsilon^2).
\end{aligned}$$

We have

$$\begin{aligned}
r(z_r) &= \hat{A}_r z_r - \theta_r B z_r \\
&= A_r z_r + \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) - \lambda_r B z_r + O(\varepsilon^2) \\
&= \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) + O(\varepsilon^2)
\end{aligned}$$

and

$$\begin{aligned}
\|r(z_r)\|_{B^{-1}}^2 &= \{ \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) \}^T B^{-1} \{ \sum_{k=1}^{r-1} \sigma_k \varepsilon_k {c_r}^{(k)} (Bz_k) \} \\
&= \sum_{k=1}^{r-1} {\sigma_k}^2 {\varepsilon_k}^2 {c_r}^{(k)^2}.
\end{aligned}$$

Now from (5), it follows that

$$|\hat{\lambda}_r - \lambda_r| \leq \frac{\|r(z_r)\|_{B^{-1}}^2}{d_r} = \frac{1}{d_r} \sum_{k=1}^{r-1} {\sigma_k}^2 {\varepsilon_k}^2 {c_r}^{(k)^2} \leq \frac{1}{d_r} \varepsilon^2 c_r^2 \sum_{k=1}^{r-1} \sigma_k^2,$$

where $d_r = \min_{i \neq r} |\hat{\lambda}_r - \hat{\lambda}_i|$. $\square$

In [13] they considered the bounds of $\hat{\lambda}_k,\ k \geq 2$, based only on the $\hat{A}_2$ while the bound we obtained in (8) concerns for general updating procedure. Furthermore, we only need to focus on the bound of $\hat{\lambda}_r$, which is the smallest eigenvalue of $\hat{A}_r$, rather than the bounds of eigenvalues $\hat{\lambda}_k,\ k > 2$, of $\hat{A}_2$ as in [13].

References

[1] K. J. Bathe and E. Wilson, *Solution methods for eigenvalue problems in structural dynamics*, Internat. J. Numer. Methods Engrg., 6(1973), pp. 213–226.

[2] Y. Cho and Y. K. Yong, *A multi-mesh, preconditioned conjugate gradient solver for eigenvalue problems in finite element models*, Comput. Struct., 58(1996), pp. 575–583.

[3] J. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations Vol. 1, Theory*, Birkhauser, Boston, 1985.

[4] Y. T. Feng and D. R. J. Owen, *Conjugate gradient methods for solving the smallest eigenpair of large symmetric eigenvalue problems*, Internat. J. Numer. Methods Engrg., 39(1996), pp. 2209–2229.

[5] G. Gambolati, G. Pini and F. Sartoretto, *An improved iterative optimization technique for the leftmost eigenpairs of large symmetric matrices*, J. Comput. Phys., 74(1988), pp. 41–60.

[6] G. Gambolati and M. Putti, *A comparison of Lanczos and optimization methods in the partial solution of sparse symmetric eigenproblems*, Internat. J. Numer. Methods Engrg., 37(1994), pp. 605–621.

[7] G. Gambolati, F. Sartoretto, and P. Florian, *An orthogonal accelerated deflation technique for large symmetric eigenproblems*, Comput. Methods Appl. Mech. Engrg., 94(1992), pp. 13–23.

[8] W. Hackbusch, *Multi-Grid Methods and Applications*, Springer-Verlag, New York, 1985.

[9] D.E. Longsine and S.F. McCormick, *Simultaneous Rayleigh quotient minimization methods for $Ax = \lambda Bx$*, Linear Algebra Appl., 34(1980), pp. 195–234.

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

[11] A. Ruhe, *Computation of eigenvalues and eigenvectors*, in Sparse Matrix Techniques, V. A. Baker, ed., Springer-Verlag, Berlin(1977), pp. 130–184.

[12] F. Sartoretto, G. Pini and G. Gambolati, *Accelerated simultaneous iterations for large finite element eigenproblems*, J. Comput. Phys., 81(1989), pp. 53–69.

[13] H. R. Schwarz, *Eigenvalue problems and preconditioning*, ISNM, 96(1991), pp. 191–208.

[14] H. R. Schwarz, *The eigenvalue problem $(A - \lambda B)x = 0$ for symmetric matrices of high order*, Comput. Methods Appl. Mech. Engrg., 3(1974), pp. 11–28.

Department of Mathematics
Hanyang University, Seoul 133-791, Korea