Trends in Mathematics
Information Center for Mathematical Sciences
Volume 4, Number 2, December 2001, Pages 103–109

PRECONDITIONED CONJUGATE GRADIENT METHOD FOR LARGE GENERALIZED EIGENPROBLEMS

HO-JONG JANG

Abstract. A short survey of some results on preconditioned conjugate gradient(PCG) scheme, with the aid of deflation, for computing a few of the smallest eigenpairs of the large generalized eigenproblems is presented. The survey is by no means complete and reflects the author's personal interests and biases.

1. Introduction

The partial eigenanalysis of large sparse symmetric matrices is a common task in many scientific applications that make use of finite difference or finite element models. Typical examples are offered by the dynamical analysis of elastic structures[1], the design of optimal waveguides[37], and the spectral superposition approach for the solution of large sets of differential equations[14]. Several techniques have been developed for the solution of the partial eigenproblem, including subspace iteration[4,36], Lanczos scheme[6], multigrid[19], and Davidson's method[7,8].

An alternative to those techniques employs iterative schemes for minimizing the Rayleigh quotient

$$R(x) = \frac{x^T A x}{x^T B x}, \tag{1}$$

based on the fact that the minimum value of (1) is equal to the smallest eigenvalue of the generalized eigenvalue problem

$$Ax = \lambda Bx, \tag{2}$$

where A and B are large sparse symmetric positive definite matrices.

The idea of transforming the eigenproblem (2) into a minimum optimization problem, first proposed by Hesteness and Karush[20], opens a wide prospect for the evaluation of eigenvalues with the aid of the optimization procedures which have become well developed in recent decades. Certainly, some special conditions are needed in selecting the particular methods.

Several methods, such as the steepest descent method[20], coordination relaxation scheme[12] and CG method[3,17,21,22,25,30,33,34] were adopted to assess the smallest eigenpairs based on the minimization of the Rayleigh quotient. These methods can evaluate not only the smallest eigenpairs but also several smallest eignepairs with the aid of deflation[16,18,32,35] or subspace techniques[36].

2000 *Mathematics Subject Classification.* 65F50, 65F15.
Key words and phrases. Generalized eigenvalue problem, PCG method, Deflation.
Received April 27, 2001.

The most attractive characteristic of these methods is that the iterative process mainly consists of vector-vector operations or matrix-vector multiplications which are hightly suitable for being parallelized or vectorized. Moreover, these methods can easily handle large scale problmes due to their much reduced storage requirements.

Among the methods mentioned above for minimizing the Rayleigh quotient, the CG scheme appears to be the most efficient and robust providing relatively faster convergence for large sparse eigenproblems. As in the case of a system of linear equations, successful application of the CG scheme to eigenproblems depends also upon the preconditioning techniques[9,10,11]. A proper choice of the preconditioner significantly improves the convergence of the CG scheme. Among the various preconditioning techniques, incomplete Cholesky(IC) factorization preconditioner is a reliable and efficient tool for the solution to both linear systems and eigenproblems in a finite element context [5,13,31].

As we mentioned above, the PCG can also be used to evaluate several smallest eigenpairs with the aid of deflation. Two different types of deflation techniques are typically used to compute a few of the smallest eigenpairs, deflation-PCG with partial shifts [16,32,35] and an orthogonal deflation-PCG(ODPCG) [18].

2. Generalized eigenproblem via preconditioned CG

2.1. Conjugate gradient scheme

We here look for the m smallest eigenvalues

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

and for the corresponding eigenvectors $z_1, z_2, \cdots, z_m$ of (2) such that

$$Az_j = \lambda_j Bz_j,\ {z_j}^T Bz_j = 1,\ j = 1, 2, \cdots, m.$$

The number m of the desired eigenpairs (λ_j, z_j) is small compared with the order n of the matrices.

The minimum of the Rayleigh quotient $R(x)$ corresponding to (2) is equal to λ_1 and is attained at z_1:

$$\min_{x \neq 0} R(x) \ = \ \min_{x \neq 0} \frac{x^T Ax}{x^T Bx} \ = \ \lambda_1 \ = \ \frac{{z_1}^T Az_1}{{z_1}^T Bz_1}$$

The minimum of $R(x)$ is determined iteratively by means of the CG scheme. For an iterate x_k the corresponding gradient of $R(x)$

$$g_k \ = \ g(x_k) \ = \ \nabla\, R(x_k) \ = \ \frac{2}{{x_k}^T Bx_k} \{Ax_k - R(x_k)Bx_k\}$$

is used to fix the direction of descent p_{k+1} in which $R(x)$ is minimized. If x_0 denotes the initial vector, the directions of descent are defined as follows[24,27]:

$$p_1 = -g_0,\ p_{k+1} = -g_k + \frac{{g_k}^T g_k}{{g_{k-1}}^T g_{k-1}} p_k,\ k = 1, 2, 3, \cdots .$$

The subsequent iterate

$$x_{k+1} = x_k + \delta_{k+1} p_{k+1},\ k = 0, 1, 2, \cdots \tag{3}$$

is obtained from, setting $\delta_{k+1} = \delta$,

$$R(x_{i+1}) = \frac{{x_k}^T Ax_k + 2\delta {p_{k+1}}^T Ax_k + \delta^2 {p_{k+1}}^T Ap_{k+1}}{{x_k}^T Bx_k + 2\delta {p_{k+1}}^T Bx_k + \delta^2 {p_{k+1}}^T Bp_{k+1}} = \min! \tag{4}$$

A detailed explanation to get the values for $\delta^{(k+1)}$ can be found in [22,23].

2.2. Preconditioned CG scheme

The convergence of the sequence of iterates x_k in (3) towards the direction of z_1 depends on the condition number of the Hessian matrix $H(x)$ of $R(x)$

$$H(x) = \frac{2}{x^T Bx}\left[A - R(x)B - g(x)(Bx)^T - (Bx)g(x)^T\right]$$

evaluated at z_1 [28,29]. Due to the B-orthonormality of z_1 and $g(z_1) = 0$, $H(z_1) = 2(A-\lambda_1 B)$ holds and the norm of $H(z_1)$, subordinate to the norms $\|x\|_B = \sqrt{x^T Bx}$ and $\|x\|_{B^{-1}} = \sqrt{x^T B^{-1}x}$, is given by

$$\|H(z_1)\|_{B,B^{-1}} = \sup_{x\neq 0} \frac{\|H(z_1)x\|_{B^{-1}}}{\|x\|_B} = 2(\lambda_n - \lambda_1).$$

Since $H(z_1)$ is positive semidefinite, the corresponding condition number is defined by

$$\kappa_{B,B^{-1}}(H(z_1)) = \frac{\lambda_n - \lambda_1}{\lambda_2 - \lambda_1},$$

However, the condition number of the Hessian matrix can be essentially decreased by the use of suitable preconditioning techniques. The idea behind the PCG is to apply 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 (1), we obtain

$$R(\tilde{x}) = \frac{\tilde{x}^T C^{-1}AC^{-1}\tilde{x}}{\tilde{x}^T C^{-1}BC^{-1}\tilde{x}} = \frac{\tilde{x}^T \tilde{A}\tilde{x}}{\tilde{x}^T \tilde{B}\tilde{x}}, \tag{5}$$

where the matrices $\tilde{A}$ and $\tilde{B}$ are symmetric positive definite. The transformation (5) leaves the stationary values of (1) unchanged, which are eigenvalues of (2), while the corresponding stationary points are obtained from $\tilde{x}_j = Cz_j,\ j = 1, 2, \cdots, n$.

There are a number of choices of M, ranging from simple to complicated forms, among which the IC decomposition that preserves exactly the nonzero pattern of A exhibits both efficiency and robustness. The preconditioning matrix $M = HH^T$ is used, H being the pointwise IC factor of A. The PCG algorithm for solving the smallest eigenpair with implicit preconditioning is summarized as follows.

The PCG algorithm for computing the smallest eigenpair

Choose the preconditioner M
Start : Choose $x_0 \neq 0;\ v_0 = Ax_0,\ \hat{v}_0 = Bx_0;\ s_0 = 0;$
$\alpha_0 = {x_0}^T v_0,\ \rho_0 = {x_0}^T \hat{v}_0;\ q_0 = \alpha_0/\rho_0;\ \zeta_a = 1;$
Iteration ($k = 1, 2, 3, \cdots$) :

$$
\begin{aligned}
& g_{k-1} = (v_{k-1} - q_{k-1}\hat{v}_{k-1})(2/\rho_{k-1}); \\
& Mh_{k-1} = g_{k-1};\ \text{(preconditioning step)} \\
& \zeta = {g_{k-1}}^T h_{k-1};\ \varepsilon_{k-1} = \zeta/\zeta_a; \\
& s_k = -h_{k-1} + \varepsilon_{k-1} s_{k-1}; \\
& w_k = As_k,\ \hat{w}_k = Bs_k; \\
& \beta = {x_{k-1}}^T w_k,\ \gamma = {s_k}^T w_k,\ \sigma = {x_{k-1}}^T \hat{w}_k,\ \tau = {s_k}^T \hat{w}_k; \\
& \text{compute } \delta_k \text{ from (4)}; \\
& x_k = x_{k-1} + \delta_k s_k, \\
& v_k = v_{k-1} + \delta_k w_k,\ \hat{v}_k = \hat{v}_{k-1} + \delta_k \hat{w}_k, \\
& \alpha_k = {x_k}^T v_k,\ \rho_k = {x_k}^T \hat{v}_k;\ q_k = \alpha_k/\rho_k;\ \zeta_a = \zeta; \\
& \text{test on convergence}
\end{aligned}
\tag{6}
$$

3. Higher eigenvalues computation

Although the PCG scheme in §2 only produces the smallest eigenpair of (2), this algorithm can also be used to evaluate a few of the smallest eigenvalues and their corresponding eigenvectors together with the aid of deflation.

3.1. The deflation with a partial shift

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

$$A_r = A + \sum_{i=1}^{r-1} \sigma_i (Bz_i)(Bz_i)^T, \tag{7}$$

with σ_i is the shift that satisfies $\sigma_i > 0$ and $\lambda_i + \sigma_i > \lambda_r,\ i = 1, 2, \cdots, r-1$. More details and the numerical stability of the deflation process (7) are reported in [32].

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. It is natural to define the corresponding preconditioner M_r, that is exact one in case of $M = A$, as follows

$$M_r = M + \sum_{i=1}^{r-1} \sigma_i (Bz_i)(Bz_i)^T.$$

The replacement of M with M_r of the PCG scheme in §2 requires the solution of the linear equation

$$\left\{ M + \sum_{i=1}^{r-1} \sigma_i (Bz_i)(Bz_i)^T \right\} h = g. \tag{8}$$

To solve for h in (8), a Sherman-Morrison formula is applied.

In the proposed method we assume that the shifts σ_i are chosen properly. A too small choice of σ_i, such that the shifted eigenvalue $\lambda_1 + \sigma_1$ is not much greater than λ_2 and even much smaller than λ_3 causes a considerable reduction of the convergence rate. On the other hand, too large shifts have also an unfavorable effect on the convergence rate. In general most satisfactory results are achieved, if the already computed eigenvalues $\lambda_1, \lambda_2, \cdots \lambda_{r-1}$ are shifted to values which are at least greater than λ_r, but not too much greater [35]. In fact, the proper way of determining the shifts σ_i which gives better convergence will be affected by several factors and may not be theoretically obtained.

Some ways of deterimining the shifts σ_i are reported in [35], but their influence seems not to be very significant upon the convergence. A possible strategy in [16], which is based on the assumption that the preconditioner M is kept fixed for A_r, is computationally expensive along with the modification of M_r.

Another possible strategy defines the σ_i to be a multiple of the lastly computed eigenvalue λ_{r-1}[32], where the factor decreases from a starting value to a limiting value larger than one with increasing index r.

3.2. Orthogonal deflation-PCG

Although the DPCGB is numerically superior to some other deflation techniques with a partial shift, in case of pathological eigenvalue distributions every rule for defining the shifts may happen to fail. In a series of papers [2,15,18,31], Gambolati et al. have developed schemes which can avoid this difficulty.

Following [18], the basic idea underlying ODPCG is as follows. Assume that the eigenpairs $(\lambda_i, z_i),\ i = 1, \cdots, r-1$, have been computed. To avoid convergence toward one of the computed eigenvectors $z_i,\ i = 1, \cdots, r-1$, the next initial vector $\tilde{x}_0^r$ is chosen to be B-orthogonal to $Z_{r-1} = \text{span}\{z_i \mid i = 1, \cdots, r-1\}$. And the direction vector $\tilde{p}_k^r$ is evaluated by B-orthogonalizing p_k^r with respect to Z_{r-1}. Also the new approximation vector $\tilde{x}_k^r$ is evaluated by B-normalizing x_k^r (in (3)). Now from the characterization of the eigenvectors [26]

$$R(z_r) = \min_{x \perp_B Z_{r-1}} R(x),$$

$\tilde{x}_k^r$ converges toward z_r as k increases. That is, after $z_i,\ i = 1, \cdots, r-1$ have been evaluated, z_r can be determined by minimizing $R(x)$ over the vector space which is the B-orthogonal complement to Z_{r-1}. The minimization is performed by the PCG scheme in (6).

References

[1] [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] L. Bergamaschi, G. Gambolati, and G. Pini, *Asymptotic convergence of conjugate gradient methods for the partial symmetric eigenproblem*, Numer. Linear Algebra Appl., 4(2)(1997), pp. 69–84.

[3] W. W. Bradbury and R. Fletcher, *New iterative methods for the solution of the eigenproblem*, Numer. Math., 9(1966), pp. 259–267.

[4] J. H. Bramble, J. E. Pasciak, and A. V. Knyazev, *A subspace preconditioning algorithm for eigenvector/eigenvalue computation*, Adv. Comput. Math., 6(2)(1997), pp. 159–189.

[5] 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.

[6] J. K. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations*, Birkhäuser, Boston(1985).

[7] E. R. Davidson, *The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices*, J. Computational Physics, 17(1975), pp. 817–825.
[8] E. R. Davidson, *Matrix eigenvector methods*, in G.H.F.Direcksen and S. Wilson, editors, Methods in Computational Molecular Physics, Reidel, Boston(1983), pp. 95–113.
[9] D. J. Evans, *Preconditioned iterative methods for the generalized eigenvalue problem*, In Matrix Pencils, Springer, Berlin(1983), pp. 189–194.
[10] D. J. Evans, *On preconditioned iterative methods for solving* $(A - \lambda B)x = 0$, Computing, 32(2)(1984), pp.139–152.
[11] D. J. Evans and J. Shanehchi, *Preconditioned iterative methods for the large sparse symmetric eigenvalue problem*, Comp. Meth. Appl. Mech. and Eng., 31(3)(1982), pp. 252–264.
[12] D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra*, Freeman, San Francisco, 1963.
[13] 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.
[14] G. Gambolati, *On time integration of groundwater flow equations by spectral methods*, Water Resource Res., 29(1993), pp. 1257–1267.
[15] G. Gambolati, G. Pini, and M. Putti, *Nested iterations for symmetric eigenproblems*, SIAM J. Sci. Comput., 16(1995), pp. 173–191.
[16] 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.
[17] 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.
[18] 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.
[19] W. Hackbush, *Multi-Grid Methods and Applications*, Springer-Verlag, New York, 1985.
[20] M. R. Hestenes and W. Karush, *A method of gradients for the calculation of the characteristic roots and vectors of a real symmetric matrix*, J. Res. Nat. Bur. Standards, 47(1951), pp. 45–61.
[21] M. R. Hestenes and W. Karush, *Solutions of* $Ax = \lambda Bx$, J. Res. Nat. Bur. Standards, 49(1951), pp. 471–478.
[22] D.E. Longsine and S.F. McCormick, *Simultaneous Rayleigh quotient minimization methods for* $Ax = \lambda Bx$, Linear Algebra Appl., 34(1980), pp. 195–234.
[23] S. F. McCormick, *Some convergence results on the method of gradients for* $Ax = \lambda Bx$, J. Compute. and Syst. Sci., 13(2)(1976), pp. 213–222.
[24] S. F. McCormick and T. Noe, *Simultaneous iteration for the matrix eigenvalue problem*, Linear Algebra and Applications., 16(1)(1977), pp. 43–56.
[25] M. Papadrakakis and M. Yakoumidakis, *A partial preconditioned conjugate gradient method for large eigenproblems*, Comp. Meth. appl. Mech. Engrg., 62(1987), pp. 195–207.
[26] B. N. Parlett, *The Symmetric Eigenvalue Problem*, Prentice-Hall, Englewood Cliffs, NJ, 1980.
[27] E. Polak, *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York(1971).
[28] A. Ruhe, *SOR-methods for the eigenvalue problem with large sparse matrices*, Math. Comput., 28(127)(1974), pp. 695–710.
[29] A. Ruhe, *Iterative eigenvalue algorithm based on convergent splittings*, J. Comp. Phys., 19(1975), pp. 110–120.
[30] A. Ruhe, *Computation of eigenvalues and eigenvectors*, in Sparse Matrix Techniques, V. A. Baker, ed., Springer-Verlag, Berlin(1977), pp. 130–184.
[31] F. Sartoretto, G. Pini and G. Gambolati, *Accelerated simultaneous iterations for large finite element eigenproblems*, J. Comput. Phys., 81(1989), pp. 53–69.
[32] H. R. Schwarz, *Eigenvalue problems and preconditioning*, ISNM, 96(1991), pp. 191–208.
[33] H. R. Schwarz, *Rayleigh-Quotient-Minimierung mit Vorkonditionierung*, in: Collatz L. et al, ed. (Birkhäuser, Basel), ISNM, 81(1987), pp. 229–245.

[34] H. R. Schwarz, *in Lecture Notes in Mathematics*, Vol. 968, J. Hinze, ed., Springer-Verlag, Berlin(1982).
[35] 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.
[36] Henk van der Vorst, *Subspace iteration for eigenproblems*, CWI Quarterly, 9(1-2)(1996), pp. 151–160, SMC 50 jubilee (Amsterdam, 1996).
[37] [37] M. Zoboli and P. Bassi, *The finite element method for anisotropic optical waveguides*, in Anisotropic and Nonlinear Optical Waveguides, C. Someda and G. Stegeman, eds., Elsevier, Amsterdam, 1992, pp. 77–116.

Department of Mathematics, Hanyang University, Seoul 133-791, Korea
E-mail address: hjang@hanyang.ac.kr