The Jacobi-Davidson Method and Correction Equations in Electronic-Structure Calculations

Akitaka SAWAMURA

The Jacobi-Davidson method consists of two major parts: the Davidson part, where an eigenproblem is projected on a small subspace, and the correction equation part, where orthogonalization is operated. However, the orthogonalization can be a bottleneck when many eigensolutions are sought at once. In this study, electronic structure calculations are tested with the aim of reducing orthogonalization costs without sacrificing computational efficiency. As a result, the correction equation without the right projector appears to be promising. Two linear solvers, conjugate-gradient method and symmetric quasi-minimum residual method, are also examined.

Keywords: numerical linear algebra, eigenvalue problem, simultaneous linear equations, iterative method

1. Introduction

First-principles calculations based on density functional theory[1] have gained enormous interest among solid-state physicists, materials scientists, and quantum chemists. The main reason for the enormous success of first-principles methods lies in the fact that an explosion of computer performance and developments in mathematical schemes to solve the Kohn-Sham (KS) equation[2], which plays a vital role within the density functional theory, have allowed us to apply them to solids and molecules which contain increasingly many atoms.

Of discretization approaches, a plane-wave formalism with pseudopotentials[3],[4] is frequently adopted, because it has many advantages including a simple but accurate evaluation of forces acting on atoms and stresses on a unit cell. When this is the case a matrix eigenvalue problem, into which the KS equation is discretized, is inevitably of large dimension in comparison with the number of eigensolutions necessary to determine the materials properties. In order to solve such a matrix eigenvalue problem, iterative methods are usually used[5]-[11]. The Jacobi-Davidson method[11] is probably the most sophisticated one, because the whole eigenvalue problem is broken down into subproblems including a linear equation, frequently referred to as a correction equation, which is easier to solve in principle.

Since the correction equation involves orthogonal projectors, however, solving it for all the desired eigensolutions still requires $O(N^3_{\mathrm{atom}})$ floating-point operations in practice, where N_{atom} is the number of atoms. In the present study, four versions of the correction equation and two iterative solution methods for linear equations are examined to reduce this computational load.

2. Method

In this section, The author introduces four correction equations and two iterative solution methods. The matrix eigenvalue problem into which the Kohn-Sham equation is transformed is given by

$$A\mathbf{x} = \varepsilon\mathbf{x},$$

where A, ε, and $\mathbf{x}$ are the Hamiltonian matrix, an eigenvalue, and a corresponding eigenvector, respectively. When multiple eigensolutions are to be sought at once, the correction equation in the Jacobi-Davidson method is a block variant expressed as[12]

$$\left(I - \sum_{1\le j\le N_{\mathrm{eig}}} \mathbf{x}_j \mathbf{x}_j^H\right)(A-\mu_i)\left(I - \sum_{1\le j\le N_{\mathrm{eig}}} \mathbf{x}_j \mathbf{x}_j^H\right)\mathbf{t}_i = \mathbf{r}_i, \quad \cdots\cdots(1)$$

where the superscript ^{H} stands for the Hermitian conjugate, $\mathbf{x}_i$ is an orthonormalized current approximation to the ith eigenvector, N_{eig} the number of eigensolutions, μ_i a current guess for the ith eigenvalue, $\mathbf{t}_i$ an augmentation vector to $\mathbf{x}_i$, and $\mathbf{r}_i$ the ith residual vector defined as

$$\mathbf{r}_i = (A-\lambda_i)\,\mathbf{x}_i,$$

with λ_j the Rayleigh quotient $\lambda_j = \mathbf{x}_i^H A\mathbf{x}_i$. After **Eq.(1)** is solved, roughly speaking, coefficients α and β contained in

$$\mathbf{x}'_i = \alpha\mathbf{x}_i + \beta\mathbf{t}_i$$

are so adjusted that $\mathbf{x}'_i$ becomes a closer approximation. In the present study **Eq.(1)** is picked up as the first correction equation.

The arithmetic operations associated with the projector

$$\left(I - \sum_{1\le j\le N_{\mathrm{eig}}} \mathbf{x}_j \mathbf{x}_j^H\right)$$

in solving **Eq.(1)** are as follows: **Eq.(1)** should be solved for the N_{eig} eigensolutions. The projector contains $O(N_{\mathrm{eig}})$ vectors. The dimension of the vectors is proportional to N_{atoms} · N_{eig} is also proportional to N_{atoms}. This leads to $O(N^3_{\mathrm{atom}})$ computational load.

Zhou[13] has proposed an alternative correction equation with only the left projector retained for computing eigensolutions in a sequential fashion. The second is a block variant of his equation written as

$$\left(I-\sum_{1\le j\le N_{\mathrm{eig}}}\mathbf{x}_j\mathbf{x}_j^H\right)(A-\mu_i)\,\mathbf{t}_i=\mathbf{r}_i. \quad \cdots\cdots(2)$$

Formally, the computational load associated with the projector is halved down by replacing **Eq.(1)** with **Eq.(2)**. This may cause slower convergence of the eigensolution, however.

Genseberger and Sleijpen[14] have shown that an original form[11] of the correction equation which contains two projectors with minimum orthogonalization,

$$(I-\mathbf{x}_i\mathbf{x}_i^H)\;(A-\mu_i)\;(I-\mathbf{x}_i\mathbf{x}_i^H)\,\mathbf{t}_i=\mathbf{r}_i, \quad \cdots\cdots(3)$$

is more efficient with respect to the number of iterations than the block variant. This suggests that **Eq.(3)** would be far more efficient from the viewpoint of CPU time, because consumption of arithmetic operations by applying the projectors in **Eq.(3)** is suppressed down to $O(N^2_{\mathrm{atom}})$. **Eq.(3)** is chosen as the third form.

If **Eqs.(2) and (3)** are just as effective as **Eq.(1)**, Zhou's original form,

$$(I-\mathbf{x}_i\mathbf{x}_i^H)\;(A-\mu_i)\,\mathbf{t}_i=\mathbf{r}_i, \quad \cdots\cdots(4)$$

should work as well. **Eq.(4)** is the fourth choice.

The conjugate-gradient (CG) method[15] is usually employed to solve the correction equation and appears to be effective even when the correction equation is not a Hermitian one[16]. Since the correction equation can be an indefinite one, however, Stathopoulos[17] has proposed using the symmetric quasi-minimum residual (sQMR) method[18] to improve robustness and efficiency. In the present study eight implementations of the Jacobi-Davidson method, to which the four forms of the correction equation and the two iterative methods to solve them lead, are compared.

3. Test calculations

In the present study two test calculations are performed within the plane-wave formalism. The first is a small-scale one. The lowest 144 eigenstates are computed until self-consistent interelectron potential and equilibrium structure are found by means of a multiple-secant algorithm[19] for a doped compound semiconductor represented by a supercell containing 64 atoms at a cut-off energy of 340eV, which is a measure of spatial resolution of the plane-wave basis set. Subsequently density of states corresponding to the lowest 256 is obtained.

The second test calculation is a large-scale one. The lowest 486 eigenstates are computed until self-consistent interelectron potential and equilibrium structure are found for a compound semiconductor alloy represented by a supercell of 216 atoms at a cut-off energy of 109eV.

For each test calculation, the eight variations of the Jacobi-Davidson method are examined in conjunction with a preconditioner derived from the Neumann expansion of an operator[20] for efficiency. The actual small-and large-scale calculations are carried out on a cluster of Intel Pentium CPUs at 3.4GHz and of Intel Xeon 5140 CPUs at 2.33GHz, respectively.

The timing results are summarized in **Table 1**. For the small-scale calculation, the total amount of CPU time does not vary much with the selection of the forms of the correction equation and of the solution methods. A closer look reveals, however, that **Eqs.(1) and (2)** have a slight edge over **Eqs.(3) and (4)**. This difference is more significant for the large-scale calculation. When the CG method is chosen, the total amount of CPU time with **Eqs.(3) and (4)** grows nine times as large as with **Eqs.(1) and (2)**, probably because of denser spectrum of the eigenvalues. This burden is mitigated to some extent by employing the sQMR method, though much larger amount of CPU time is still required. Therefore at least the projector with minimum orthogonalization should be avoided. Furthermore, when either **Eq.(1)** or **Eq.(2)** is chosen, it is solved by the CG method so stably (and slightly more efficiently) that advanced methods including the sQMR are unnecessary.

Within the scope of the present test calculations, employing **Eq.(2)** instead of **Eq.(1)** leads to reduction only in the amount of CPU time corresponding to the projection. Since the percentage of it increases with

Table 1. Relationships of CPU time with the selection of forms of the correction equation and of the solution methods for the (a) small-scale and (b) large-scale calculations. The digits in the parenthesis are the percentages of CPU time.

(a) small-scale

Correction	CPU time ($\times 10^3$s)	
	Total	Projection
CG		
(1)	47.59	8.49 (17.8)
(2)	51.17	5.91 (11.5)
(3)	62.75	9.05 (14.4)
(4)	63.39	5.54 (8.7)
sQMR		
(1)	52.71	9.57 (18.2)
(2)	50.77	6.53 (12.9)
(3)	60.93	8.53 (14.0)
(4)	58.84	5.60 (9.5)

(b) large-scale

Correction	CPU time ($\times 10^3$s)	
	Total	Projection
CG		
(1)	63.86	42.94 (67.2)
(2)	60.15	25.63 (42.6)
(3)	565.25	13.47 (2.3)
(4)	587.76	12.80 (2.2)
sQMR		
(1)	73.96	44.58 (60.3)
(2)	70.04	28.30 (40.4)
(3)	170.66	9.76 (5.7)
(4)	170.62	9.51 (5.6)

the number of atoms, however, the projection is likely to be the bottleneck in the calculation of far larger models. In such a situation, **Eq.(2)** is preferred.

4. Conclusion

The goal of the present study was to examine several variants of the correction equation in the Jacobi-Davidson method for solving the matrix eigenvalue problem arising in the first-principles plane-wave pseudopotential framework from the viewpoint of faster calculations. After a review the four forms of the correction equation, test calculations were performed. The CG and sQMR methods to solve the correction equation were also compared. The results of the test calculations indicate that block forms of the correction equation should be employed. While the correction equation without the right projector did not lead to significant reduction in the CPU time, solving it by the CG method is likely to be preferred for dealing with larger systems than in the present study. To confirm this, the author will continue my examination for the larger systems, where reducing the required CPU time is of greater importance.

* Intel, Pentium and Xeon are trademarks or registered trademarks of Intel Corporation.

References

(1) P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev. 136 (1964), B859-863.
(2) W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev. 140 (1965), A1133-1138.
(3) J. Ihm, A. Zunger, and M. L. Cohen, "Momentum-space formalism for the total energy of solids," J. Phys. C, 12 (1979), 4409-4422.
(4) W. E. Pickett, "Pseudopotential methods in condensed matter applications," Comput. Phys. Rep., 9 (1989), 115-197.
(5) C. Lanczos, "An iteration method for the solution of the eigenvalue problem of linear differential and integral operators," J. Res. Nat. Bur. Standards, 45 (1950), 255-282.
(6) 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), 45-61.
(7) W. E. Arnoldi, "The principle of minimized iterations in the solution of the matrix eigenvalue problem," Quart. Appl. Math., 9 (1951), 17-29.
(8) W. W. Bradbury and R. Fletcher, "New iterative methods for solution of the eigenproblem," Numer. Math., 9 (1966), 259-267.
(9) E. R. Davidson, "The iterative calculation of a few of the lowest eigenvalues and corresponding eigenvectors of large real symmetric matrices," J. Comput. Phys., 17 (1975), 87-94.
(10) R. Car and M. Parrinello, Unified approach for molecular dynamics and density-functional theory, Phys. Rev. Lett., 55(1985), 2471-2474.
(11) G. L. G. Sleijpen and H. A. van der Vorst, "A Jacobi-Davidson iteration method for linear eigenvalue problems," SIAM J. Matrix Anal. Appl., 17 (1996), 401-425.
(12) P. Arbenz, U. L. Hetmaniuk, R. B. Lehoucq, and R. S. Tuminaro, "A comparison of eigensolvers for large-scale 3D modal analysis using AMG-preconditioned iterative methods," Int. J. Numer. Meth. Eng. 64 (2005), 204-236.
(13) Y. Zhou, "Studies on Jacobi-Davidson, Rayleigh quotient iteration, inverse iteration generalized Davidson and Newton updates," Numer. Lin. Alg. Appl., 13 (2006), 621-642.
(14) M. Genseberger and G. L. G Sleijpen, "Alternative correction equations in the Jacobi-Davidson method," Numer. Linear Algebra Appl., 6 (1999), 235-252.
(15) M. R. Hestenes and E. Stiefel, "Methods of conjugate gradients for solving linear systems," J. Res. Nat. Bur. Standards, 49 (1952), 409-436.
(16) K. Wu, Y. Saad, and A. "Stathopoulos, Inexact Newton preconditioning techniques for eigenvalue problems," Elec. Trans. Numer. Anal., 7 (1998), 202-214.
(17) A. Stathopoulos, "Nearly optimal preconditioned methods for Hermitian eigenproblems under limited memory. Part I: Seeking one eigenvalue," Tech. Report: WM-CS-2005-03, July, 2005.
(18) R. W. Freundand and N. M. Nachtigal, "A new Krolov-subspace method for symmetric indefinite linear systems," Proceedings of the 14th IMACS World Congress on Computational and Applied Mathematics, ed. by Ames W. F., IMACS, 1994, 1253.1256; R. W. Freundand and N. M. Nachtigal, Software for simplified Lanczos and QMR algorithms, Appl. Numer. Math., 19 (1995), 319-341.
(19) A. Sawamura, M. Kohyama, T. Keishi, and M. Kaji, "Acceleration of self-consistent electronic-structure calculations: storage-saving and multiple-secant implementation of the Broyden method," Mater. Trans., JIM, 40 (1999), 1186-1192.
(20) A. Sawamura, M. Kohyama, and T. Keishi, "An efficient preconditioning scheme for plane-wave-based electronic structure calculations," Comput. Mater. Sci., 14 (1999), 4-7.

Contributor

A. SAWAMURA
• Dr. Eng.
Assistant General Manager,
Analysis Technology Research Center
Involved in computational materials science and related fields