Journal of Computational Physics **161,** 653–667 (2000)
doi:10.1006/jcph.2000.6520, available online at http://www.idealibrary.com on IDEAL®

New Methods for Calculations of the Lowest Eigenvalues of the Real Symmetric Generalized Eigenvalue Problem

Alexander V. Mitin[1,2]

Bergische Universität—Gesamthochschule Wuppertal, Fachbereich 9—Theoretische Chemie, Gaußstraße 20, D-42097 Wuppertal, Germany

Received May 25, 1999; revised March 20, 2000

A new iterative method based on a Newton correction vector for extension of the Krylov subspace, its diagonal, and band versions are proposed for calculation of selected lowest eigenvalues and corresponding eigenvectors of the generalized symmetric eigenvalue problem. Additionally, diagonal and band Jacobi–Davidson methods are introduced. Test calculations show that the new iterative method usually converges faster than quadratic near a solution. The new iterative method along with its band version uses a smaller number of iterative steps to obtain a solution compared to the Jacobi–Davidson, band Jacobi–Davidson method, and generalized Davidson method correspondingly. The diagonal version of the new method preserves an advantage over the diagonal Jacobi–Davidson and the Davidson method.

I. INTRODUCTION

The determination of the extreme eigenvalues and corresponding eigenvectors of the generalized eigenvalue problem

$$AX = \lambda BX, \tag{1}$$

with real symmetric matrices A, B and a positive definite matrix B, is one of the main problems of linear algebra, which is very important in many fields of natural sciences. Among them we can mention, for example, such fields of physical and chemical investigations as

[1] Present address: Department of Chemistry, Kansas State University, 111 Willard Hall, Manhattan, KS 66506-3701. E-mail: mitin@hecuba.chem.ksu.edu.

[2] Permanent address: Institute for High Temperatures, Russian Academy of Sciences, Izhorskaya ul. 13/19, Moscow 127412, Russia.

0021-9991/00 $35.00

calculations of the low-lying electronic states of atoms and molecules by the configuration interaction method [1, 2], numerical solution of the Hartree–Fock equation [3], and determination of the rotational-vibrational energy levels of molecules [4]. As a rule, the precision of these methods increases with the growing matrix dimension in Eq. (1). Normally the dimension of matrices in modern calculations could be significantly large than 10^3. In this case, iterative methods are usually used to find extreme and/or all eigenvalues of Eq. (1). The present investigation deals with the calculation of the extreme eigenvalues of Eq. (1) by iterative methods.

In the case of the unit matrix B

$$AX = \lambda X, \tag{2}$$

the Lanczos [5] and Davidson [6] methods are used mostly for the determination of the lowest eigenvalues and corresponding eigenvectors of Eq. (2). Different generalizations of these methods have been given in Refs. [7–45], while some efficient computer programs employing them have been published in Refs. [46–50]. Modifications of the Lanczos and Davidson methods for calculation of eigenpairs of Eq. (1) have been proposed in Refs. [12–14, 21, 32, 39, 51–55].

The main goal of many of these modifications was improvement of iterative convergence. However, only recently has it been recognized that the Boys–Nesbet correction formula [56, 57] used in the Davidson method often generates basis vectors that are in the same direction as the desired eigenvector, rather than orthogonal to it [42]. This results in many convergence problems in the Davidson method. On this basis the new Jacobi–Davidson method has been introduced and investigated in Refs. [39, 42, 45]. To keep the orthogonal condition, a new correction vector in this method is calculated from the solution of the system of linear equations

$$(I - XX^+)(A - \varepsilon B)(I - XX^+)U = -R, \tag{3}$$

where $R = AX - \varepsilon BX$ is the residual vector; ε, X are the current approximate eigenvalue and eigenvector for Eq. (1); and U is a new correction vector. Numerical complexity of a solution of this linear equation system is proportional N^3, where N is the matrix dimension. For this reason application of the Jacobi–Davidson method for calculation of eigenvalues of large matrices is time consuming. However, note that the correction formula of the Davidson [6] and the generalized Davidson [22] method has been derived from the equation

$$(A - \lambda B)\delta X = (A - \lambda B)X \tag{4}$$

by introducing the diagonal and band approximation of the left hand side matrices. Numerical complexity of these methods is proportional N and Nn^2 correspondingly, where n is the bandwidth. This is much less than compared to N^3. Similar simplifications could be applied for Eq. (3). They lead to the diagonal and band Jacobi–Davidson methods that are introduced in the present investigation.

The other problem of the Davidson method connects with its classification. It is well known that Davidson has classified his method as the Newton–Raphson type method [58]. However, it is possible to give other interpretations. It can be considered as a combination of the Boys–Nesbet correction formula with the Lanczos [5] and Karush [7] iteration methods,

as well as a generalization of the optimal relaxation method [59] to include the corrections from a large number of correction vectors. Wood and Zunger [21] have noted also that the Boys–Nesbet correction formula can be derived straightforwardly if it is assumed that a variation of the residual vector is equal to zero

$$R(X + \delta X) = (A - \lambda B)X + (A - \lambda B)\delta X = 0.$$

The expression for the correction vector δX is obtained then by introducing the diagonal approximation to the A and B matrices of the denominator

$$\delta X_i = -\frac{[(A - \lambda B)X]_i}{A_{ii} - \lambda B_{ii}}. \tag{5}$$

Thus, we can see that the Davidson method could be interpreted in a different way. The main difficulty with the Davidson method is, however, that it does not have quadratic convergence near a solution. On the other hand, we know that the quadratic convergence is an inherent property of the Newton type methods. This permits us to conclude that the Davidson method is not a Newton–Raphson type method.

This conclusion stimulates the development of the Newton type iterative methods for determination of selected lowest eigenvalues of Eq. (1) presented below in the following order. The new Newton type iterative method is proposed in Section II of this work. The diagonal and band approximations of the Hessian matrix, introduced in Section III, have permitted us to construct appropriate versions of the new iterative method that are given in the same section. The diagonal and band Jacobi–Davidson methods are presented in Section IV. Some numerical tests and comparison of the new methods with the Jacobi–Davidson, generalized Davidson, and the Davidson methods are presented in Section V, while conclusions are given in Section VI.

II. A NEW ITERATIVE METHOD

Let us write Eq. (1) in the form

$$\rho(X) = \frac{(X, AX)}{(X, BX)} \tag{6}$$

and apply the Newton–Raphson method for minimization of this functional. Then, a correction vector δX for the initial vector X can be found from the Newton–Raphson equation [60]

$$H\delta X = -G, \tag{7}$$

where H is a Hessian matrix and G is a gradient vector. The gradient of the functional (6) is equal to

$$G = AX - \rho(X)BX, \tag{8}$$

while the second derivative matrix H is

$$H = A - \rho(X)B - BXG^{+} - G(BX)^{+}. \tag{9}$$

Substituting (8) and (9) into (7), we obtain

$$(A - \rho(X)B - BXG^{+} - G(BX)^{+})\delta X = -G. \tag{10}$$

This is the Newton–Raphson equation for the correction vector δX. It was used by Roothaan and Bagus in the single vector diagonalization method [61]. This method was developed for finding eigenvalues and eigenvectors of small matrices.

Equation (10) is an inhomogeneous system of linear equations. By direct verification, we can find that X is a solution of the homogeneous linear equation system

$$(A - \rho(X)B - BXG^{+} - G(BX)^{+})Y = 0. \tag{11}$$

This means that the Hessian matrix (9) is ill conditioned. According to the general theory of the linear equation system [62], the system of linear equations (10) has a solution in this case only if X is orthogonal to the right hand side of Eq. (10), i.e., to the gradient vector G. But this is true. Thus, we can see that the Newton–Raphson equation (10) gives a correction vector that is orthogonal to the current approximate eigenvector X. The Jacobi–Davidson correction vector has a similar orthogonality property. However, they differ: the first one is derived by assuming that a variation of the quadratic approximation of the energy functional (6) is equal to zero, while the second one can be obtained by assuming that a variation of the residual vector of Eq. (1) is equal to zero. In contrast to this the Boys–Nesbet correction vector (5) does not have a similar orthogonality property.

Equation (10) permits us to construct a new iterative method for calculation of selected lowest eigenvalues and corresponding eigenvectors of the generalized eigenvalue problem (1). It is based on the solution of Eq. (1) in the Krylov subspace by the Galerkin method [55, 63]. The critical moment in this approach is the choice of a Krylov space basis. It is known that the Krylov vectors form a basis in this space. However, other bases can be used as well. At least two of them are well known. These are the gradient vector system used in the Lanczos method [5] that is obtained by employing the orthogonalization procedure to the Krylov vector system and the Boys–Nesbet vector system used in the Davidson method [6]. The Newton correction vectors system, generated by Eq. (10), also forms a basis of the Krylov space. Therefore, it can be used in a solution of Eq. (1) by the Galerkin method. The algorithm of the new iterative method based on the Newton correction vectors system is given as follows:

Step 1. Select an orthogonal set of approximate vectors to the first eigenvectors $X_1, X_2, \ldots, X_k$ $(k \geq m)$. Form and save $AX_1, AX_2, \ldots, AX_k$, and $F_{ij} = (AX_i, X_j)$, $S_{ij} = (BX_i, X_j)$, $i, j = 1, \ldots, k$.

Step 2. Solve the equation $Fd = \varepsilon Sd$ and select the eigenvector $d = (d, \ldots, d_k)^T$ corresponding to the mth eigenvalue ε_m of this equation.

Step 3. Form $Z = \sum_{i=1}^{k} d_i X_i$ as an approximation to the mth eigenvector. Form $G_m = AZ - \varepsilon_m BZ$ and check convergence on $\|G_m\|$.

Step 4. Solve the equation

$$(A - \varepsilon_m B - BZG^{+} - G(BZ)^{+})X_{k+1} = -G_m \tag{12}$$

to define new vector X_{k+1}.

Step 5. Orthogonalize X_{k+1} to the vector $X_1, \ldots, X_k$, i.e., for $n = 1, \ldots, k$ do

$$\sigma_n = (X_n, X_{k+1}),$$

$$X_{k+1} = X_{k+1} - \sigma_n X_n.$$

Step 6. Compare the maximum of the absolute value of σ_n $(n = 1, \ldots, k)$ with an orthogonality criterion and possibly repeat *Step* 5. Normalize the final X_{k+1}.

Step 7. Put $k = k + 1$. Form $F_{ik+1} = (AX_i, X_{k+1})$, $S_{ik+1} = (BX_i, X_{k+1})$, $i = 1, \ldots, k + 1$, and return to *Step* 2.

Usually all intermediate vectors are stored on external files in this type of algorithm. In the case of large matrices and/or slow convergence of the iterations, this results in some computational difficulties connected with keeping and using large arrays. To overcome these difficulties the number of saved vectors could be limited and a restart procedure could be intorduced. It can use a vector calculated in Step 4 as a new vector and the calculation can be continued with Step 2.

The main problem of this algorithm, however, is connected with the singularity of the Hessian matrix in Eq. (9) and, consequently, in Eq. (10). A possible approach to the solution of this equation has been proposed in Ref. [43]. A modern approach to a solution of the singular linear equation system consists of using the incomplete Cholesky factorization method [62, 64]. In the present study, it was found that Eq. (10) could be solved successfully also by the Cholesky factorization method [65], because the condition number of the Hessian matrix was less than 10^{30}, at least in all numerical tests presented below. The DECOMP and SOLVE subroutines given in Ref. [66] have been used for this purpose in the present investigation.

III. THE BAND AND DIAGONAL ITERATIVE METHODS

Solution of the linear equations system (12) is the most important step in the new method. New vectors generated at this step are used for extending the generalized Krylov subspace of an initial generalized eigenvalue problem (1). When this subspace is formed by the Newton correction vectors, then an iterative method based on it will have convergence faster than quadratic near a solution, because it will use correction terms higher than quadratic for the construction of a new vector. With other vectors, the obtained method could converge more slowly; nevertheless, the convergence speed could be high enough. Particularly, appropriate vectors can be generated by employing approximation methods for the solution of the Newton–Raphson equation (12).

Many different approximation methods can be used for this purpose. This opens extensive possibilities in constructing new efficient iterative methods. In the present study, the two simple approximations of the Hessian matrix, band and diagonal, are considered.

The band approximation leads to the band iterative method. In this method, the band submatrix with n diagonals of the full Hessian matrix is used in a solution of Eq. (12). This reduces significantly the numerical complexity of a solution of the Newton–Raphson equation (12). In all others, the band iteration method is similar to the original one.

The diagonal approximation of the A, B, and Hessian matrix by the main diagonal in Eq. (10) results in a new simple formula of a correction vector

$$\delta X_i = -\frac{G_i}{A_{ii} - \rho B_{ii} - B_{ii} X_i G_i - G_i B_{ii} X_i}.$$

It permits us to propose the diagonal iterative method for calculation of extreme selected eigenvalues of Eq. (1) without solution of a linear equation system. The algorithm of this

method is as follows:

Step 1. Select an orthogonal set of approximate vectors to the first eigenvectors $X_1, X_2, \ldots, X_k$ $(k \geq m)$. Form and save $AX_1, AX_2, \ldots, AX_k$, and $F_{ij} = (AX_i, X_j)$, $S_{ij} = (BX_i, X_j)$, $i, j = 1, \ldots, k$.

Step 2. Solve the equation $Fd = \varepsilon Sd$ and select the eigenvector $d = (d_1, \ldots, d_k)^T$ corresponding to the mth eigenvalue ε_m of this equation.

Step 3. Form $Z = \sum_{i=1}^{k} d_i X_i$ as an approximation to the mth eigenvector. Form $G_m = AZ - \varepsilon_m BZ$ and check convergence on $\|G_m\|$.

Step 4. Form new vector X_{k+1}

$$X_{i,k+1} = -\frac{G_{i,m}}{A_{ii} - \varepsilon_m B_{ii} - B_{ii} Z_i G_i - G_i B_{ii} Z_i}.$$

Step 5. Orthogonalize X_{k+1} to the vectors $X_1, \ldots, X_k$, i.e., for $n = 1, \ldots, k$ do

$$\sigma_n = (X_n, X_{k+1}),$$
$$X_{k+1} = X_{k+1} - \sigma_n X_n.$$

Step 6. Compare the maximum of the absolute value of σ_n $(n = 1, \ldots, k)$ with an orthogonality criterion and possibly repeat *Step* 5. Normalize the final X_{k+1}.

Step 7. Put $k = k + 1$. Form $F_{ik+1} = (AX_i, X_{k+1})$, $S_{ik+1} = (BX_i, X_{k+1})$, $i = 1, \ldots, k + 1$ and return to *Step* 2.

The diagonal iterative method is similar to the new iterative method and the band iterative method. However, the main difference consists in the absence of a linear equation solution. This simplification leads to more slow convergence of the diagonal iterative method compared to the new iterative method and the band iterative method. Nevertheless, when a calculation of the A and B matrices is faster than a solution of a linear equation system, then the diagonal iterative method could have an advantage over the band and new iterative methods, because the diagonal iterative method could be less time consuming in this case. However, on the other hand, when calculation of the A and B matrices takes longer than a solution of a linear equation system, then the band and the new iterative methods outperform the diagonal iterative method due to more faster convergence of the iterations. Thus, we can see that the problem of computational efficiency of the new methods depends on some external factors so that it must be considered separately.

IV. DIAGONAL AND BAND JACOBI–DAVIDSON METHODS

The diagonal approximations of all left hand side matrices of Eq. (3) lead to another new simple formula of a basis vector

$$U_i = -\frac{[(A - \varepsilon B)X]_i}{(A_{ii} - \varepsilon B_{ii})(B_{ii} - (B_{ii} X_i)^2)^2}.$$

This equation permits us to construct the diagonal Jacobi–Davidson iterative method for calculation of selected lowest eigenvalues and corresponding eigenvectors of the generalized eigenvalue problem (1). A numerical algorithm of this method is similar to the diagonal

iterative method introduced in the previous section and given below:

Step 1. Select an orthogonal set of approximate vectors to the first eigenvectors $X_1, X_2, \ldots, X_k$ $(k \geq m)$. Form and save $AX_1, AX_2, \ldots, AX_k$, and $F_{ij} = (AX_i, X_j)$, $S_{ij} = (BX_i, X_j)$, $i, j = 1, \ldots, k$.

Step 2. Solve the equation $Fd = \varepsilon Sd$ and select the eigenvector $d = (d_1, \ldots, d_k)^T$ corresponding to the mth eigenvalue ε_m of this equation.

Step 3. Form $Z = \sum_{i=1}^{k} d_i X_i$ as an approximation to the mth eigenvector. Form $G_m = AZ - \varepsilon_m BZ$ and check convergence on $\|G_m\|$.

Step 4. Form new vector X_{k+1}

$$X_{i,k+1} = -\frac{G_{i,m}}{(A_{ii} - \varepsilon_m B_{ii})(B_{ii} - (B_{ii} Z_i)^2)^2}.$$

Step 5. Orthogonalize X_{k+1} to the vectors $X_1, \ldots, X_k$, i.e., for $n = 1, \ldots, k$ do

$$\sigma_n = (X_n, X_{k+1}),$$
$$X_{k+1} = X_{k+1} - \sigma_n X_n.$$

Step 6. Compare the maximum of the absolute value of σ_n $(n = 1, \ldots, k)$ with an orthogonality criterion and possibly repeat *Step* 5. Normalize the final X_{k+1}.

Step 7. Put $k = k + 1$. Form $F_{ik+1} = (AX_i, X_{k+1})$, $S_{ik+1} = (BX_i, X_{k+1})$, $i = 1, \ldots, k + 1$, and return to *Step* 2.

The band Jacobi–Davidson algorithm is obtained when a band submatrix with n diagonals of the full matrix

$$(I - XX^+)(A - \varepsilon B)(I - XX^+) \tag{13}$$

is used in solution of Eq. (3). Therefore, the difference between the diagonal and band Jacobi–Davidson algorithms is only in Step 4. In the band Jacobi–Davidson algorithm a new vector X_{k+1} is calculated from a solution of the linear equation system with the band matrix

$$(I - ZZ^+)(A - \varepsilon B)(I - ZZ^+)X_{k+1} = -G_m. \tag{14}$$

Note that the Jacobi–Davidson algorithm used in the present investigation is obtained when a calculation of a new vector at Step 4 in the diagonal Jacobi–Davidson method is replaced by the solution of Eq. (14) with the full matrix.

Note also that Eq. (3) for a correction vector of the Jacobi–Davidson method has been derived from Eq. (4). However, this equation is degenerate when no approximations are used for left-hand side matrices. Equation (3) preserves in some sense this property. In particular, by direct verification we can find that matrix (13) is singular if X is a solution of Eq. (1). Nevertheless, the system of linear equation (3) is consistent because the residual vector is orthogonal to a solution of Eq. (1). Thus, we can see that numerical problems in a solution of a linear equation system in the Jacobi–Davidson and in the new iterative method are similar and connected with singularity of the linear system matrix.

V. NUMERICAL TESTS

In all numerical tests presented below iterations were terminated when the Euclidean norm of the residual vector was less than 10^{-10}. The Euclidean norm of a residual vector or

number of iterations needed for calculation of an eigenvalue are presented in tables mainly to demonstrate convergence of different methods in these calculations. The maximal dimension of the Krylov space was equal to 50. The following notations are used below: NIM is for the new iterative method proposed in this work; BIM(n) is for the band iterative method with n diagonals; DIM is for the diagonal iterative method; JD is used for the Jacobi–Davidson method; BJD(n) is for the band Jacobi–Davidson method with n diagonals; DJD is for the diagonal Jacobi–Davidson method; D is for the Davidson method; and GD(n) is used for the generalized Davidson method [22] with n diagonals. The algorithm of the Davidson method used in the present calculation differs from that of the diagonal iterative method and the diagonal Jacobi–Davidson method described above in Step 4 only. At this step, a new basis vector in the Davidson method was calculated by the Boys–Nesbet formula for a correction vector (4).

EXAMPLE 1. A test matrix A of order 20 was taken from the first example of Ref. [22]. Matrix A was tridiagonal except that A_{1n} and A_{n1} were nonzero, with $A_{ij} = i$ while all other nonzero elements were equal to 1. The starting vector was $(1.0, 0.1, 0.1, \ldots, 0.1)^{\mathrm{T}}$. The structure of this matrix is similar to the Hückel matrix [1] of a ring polymer. The convergence of the D, GDM(3), DJD, BJD(3), JD, DIM, BIM(3), and NIM methods is presented in Table I. Comparison of these results shows that convergence of the GD, BJD, and BIM methods as well JD and NIM methods is similar while DJD and DIM used one iteration step less when compared to the Davidson method.

EXAMPLE 2. The lowest eigenvalue of the diagonal matrix of order 100

$$A_{ij} = \begin{cases} i/55, & i = 1, \ldots, 8 \\ 19/55 + i/55, & i = 9, \ldots, 16 \\ i - 16, & i = 16, \ldots, 100 \end{cases}$$

proposed in Ref. [43] was determined by the Davidson, DJD, and DIM methods in this example. The last two methods converged at the 2nd iteration while the Davidson method reached the solution at the 56th iteration. Iterations started with the $(1.0, 0.05, \ldots, 0.05)^{\mathrm{T}}$

TABLE I

Convergence on the Norm of the Residual Vector for Different Methods in Example 1

Iteration	D	GD(3)	DJD	BJD(3)	JD	DIM	BIM(3)	NIM
1	0.52735+01	0.52735+01	0.52735+01	0.52735+01	0.52735+01	0.52735+01	0.52735+01	0.52735+01
2	0.55473+01	0.37773+01	0.13455+01	0.18826+01	0.23692+01	0.17915+01	0.18826+01	0.23692+01
3	0.18025+01	0.12863+01	0.13239+01	0.11818+01	0.14567+01	0.12866+01	0.11818+01	0.14567+01
4	0.17788+01	0.11214+01	0.38056+00	0.14981+01	0.15298+01	0.42053+00	0.14981+01	0.15298+01
5	0.95384+00	0.10343+01	0.83907-01	0.59038-01	0.22739-01	0.78438-00	0.59038-01	0.22739-01
6	0.76416-01	0.15143-01	0.22301-01	0.17246-02	0.65026-07	0.15242-01	0.17246-02	0.65026-07
7	0.11769-01	0.10149-07	0.42753-02	0.51518-07	0.19118-14	0.29721-02	0.51518-07	0.26635-14
8	0.24070-02	0.56239-14	0.67498-03	0.33977-14		0.42528-03	0.23743-14	
9	0.22914-03		0.85420-04			0.52728-04		
10	0.24869-04		0.92272-05			0.50383-05		
11	0.28416-05		0.95266-06			0.44750-06		
12	0.23595-06		0.87736-07			0.37375-07		
13	0.20861-07		0.71991-08			0.41318-08		
14	0.14084-08		0.33083-09			0.28808-09		
15	0.14697-09		0.31498-10			0.14483-10		
16	0.60098-11							

TABLE II
Convergence on the Norm of the Residual Vector of Different Methods for the Test Matrix of Example 2

Iteration	D	DJD	DIM
1	0.18748+02	0.18748+02	0.18748+02
2	0.17025+02	0.29466-14	0.38262-14
3	0.93305+01		
4	0.90111+01		
5	0.84227+01		

initial vector. The convergence of these iterations is presented in Table II. We can see that the convergence of the diagonal Jacobi–Davidson and diagonal iterative methods is faster than quadratic at the end of the iterations. This is because for the diagonal matrix these diagonal methods are equivalent to the Jacobi–Davidson and the new iterative methods.

EXAMPLE 3. In this example the number of iterations that were needed to determine the lowest eigenvalue and the corresponding eigenvector of a symmetric matrix of order 500 by different methods were compared. This matrix is defined by the relation

$$A_{ij} = \begin{cases} i, & i = j \\ W/abs(i-j), & i \neq j. \end{cases}$$

TABLE III
A Comparison of the Different Methods for Example 3

	W				
Method	1.0	0.5	0.1	0.01	0.001
D	26	26	22	21	21
GD(3)	24	21	22	20	24
GD(5)	26	30	22	21	25
GD(11)	39	46	26	36	30
GD(21)	82	61	65	68	69
GD(51)	40	43	58	52	50
DJD	17	13	10	7	6
BJD(3)	14	12	9	7	6
BJD(5)	13	13	9	7	7
BJD(11)	13	11	8	7	7
BJD(21)	14	11	9	8	7
BJD(51)	13	11	9	9	8
JD	9	8	7	6	6
DIM	15	13	9	7	6
BIM(3)	14	12	9	7	6
BIM(5)	13	13	9	7	7
BIM(11)	13	11	8	7	7
BIM(21)	14	11	9	8	7
BIM(51)	13	11	9	9	8
NIM	9	8	7	6	6

Note. The number of iterations is reported that are needed to find the lowest eigenvalue.

TABLE IV
Convergence of the Different Methods of Example 4

	Dimension				Dimension				Dimension		
Method	400	900	1600	Method	400	900	1600	Method	400	900	1600
D	50	52	52	DJD	50	52	52	DIM	52	52	52
GD(3)	38	38	38	BJD(3)	38	38	38	BIM(3)	38	38	38
GD(5)	36	36	36	BJD(5)	37	37	37	BIM(5)	36	36	36
GD(21)	36	36	36	BJD(21)	36	37	37	BIM(21)	36	36	36
GD(51)	28	36	36	BJD(51)	20	36	37	BIM(51)	18	36	36
GD(101)	27	22	22	BJD(101)	12	20	20	BIM(101)	11	18	18
GD(151)	53	27	22	BJD(151)	16	12	20	BIM(151)	11	11	18
GD(201)	15	60	27	BJD(201)	20	19	12	BIM(201)	9	11	11
				JD	7	6	6	NIM	6	6	6

Note. The number of iterations is reported that are needed to find the lowest eigenvalue.

All iteration methods started with initial vector equal to $(1.0, 0.001, 0.001, \ldots, 0.001)^{\mathrm{T}}$. Calculations were performed for five different values of W. This parameter changes the diagonal dominance of this matrix. The results of these calculations are presented in Table III. They show that the DIM has a small advantage over the DJD method and both of them are better than the Davidson method. It should be noted also that the advantage of the DIM and DJD methods over the Davidson method increases with increasing diagonal dominance of the test matrix. Convergence of the BIM and BJD methods is very similar and significantly better compared to the GD method. The NIM and JD methods display the best convergence in this example.

Additionally, the lowest eigenvalue of the same matrix of order 2000 with $W = 0.001$ was calculated by the DIM, DJD, and Davidson methods. In the latter case, 175 iterations were needed to obtain the correct solution. The DJD used 7 iterations, while only 6 iterations were used by the NIM method.

EXAMPLE 4. The lowest eigenvalue of the generalized eigenvalue problem (1) of order 400, 900, and 1600 with the matrices proposed in Ref. [53] was determined in this example by different methods. Results of these calculations are presented in Tables IV and V. Their

TABLE V
Convergence of the New Iterative Method with the Matrices of Order 400 of Example 4

	JD		NIM	
Iteration	Eigenvalue	Res. norm	Eigenvalue	Res. norm
1	−12.26625649283094	0.35255+01	−12.26625649283094	0.35255+01
2	−12.29216863017134	0.17875+01	−12.28740689800787	0.16976+01
3	−15.51688085881869	0.16415+01	−15.49796410187801	0.19020+01
4	−15.93619336403492	0.68650-01	−15.93693829254018	0.30147-01
5	−15.93704612682402	0.11697-03	−15.93704612831661	0.92399-07
6	−15.93704612831663	0.10616-09	−15.93704612831660	0.29008-13
7	−15.93704612831662	0.27146-13		

comparison shows that the DIM, DJD, and Davidson methods as well the BIM, BJD, and GD methods with a small number of band rows display similar convergence correspondingly. However, for a large number of band rows the BIM method outperforms the BJD method, which in one's turn outperforms the Davidson method. The convergence of the NIM is a little better than that of the JD method. For the matrices of order 400, the convergence of the NIM and JD methods is given in TableV. We can see that it is faster than quadratic.

EXAMPLE 5. The largest eigenvalue of a test matrix proposed in Ref. [35] whose non-zero elements are defined by the relation

$$A_{ij} = \begin{cases} i, & i = j, \\ 0.5, & i = j+1, i = j-1 \\ 0.5, & (i,\ j) \in [(1,n),(n,1)] \end{cases}$$

was calculated by different methods. Dimension of the test matrix was equal to 1000, while an initial vector was equal to $(1.0, 0.01, \ldots, 0.01)^{\mathrm{T}}$. Obtained results are given in Table VI. They show that the convergence property of the NIM and JD as well BIM and BJD methods is similar, while the behavior of the diagonal methods is quite different. The best convergence was displayed by the DIM and the slowest by the Davidson method. The condition number of the matrix in the JD method in this example was extremely high at the end of iterations. This results in a numerical problem. Therefore, only the 0.27065-09 value of the residual vector was reached by this method.

EXAMPLE 6. In this example, the zero eigenvalue of the Hilbert matrix with dimension of order 1000 was determined by different methods. The Hilbert matrix is defined in the following way

$$A_{ij} = \begin{cases} \frac{-1}{2i-1}, & i = j \\ \frac{-1}{i+j-1}, & i \neq j. \end{cases}$$

Results of these test calculations obtained with the starting vector equal to $(0.001, 0.001, \ldots, 1.0)^{\mathrm{T}}$ are given in Table VII. They show only a small advantage of the BIM and BJD

TABLE VI

Convergence on the Norm of the Residual Vector for Different Methods in Example 5

Iteration	D	GD(3)	DJD	BJD(3)	JD	DIM	BIM(3)	NIM
1	0.16766+03	0.16766+03	0.16766+03	0.16766+03	0.16766+03	0.16766+03	0.16766+03	0.16766+03
2	0.79458+02	0.16710+03	0.70974+00	0.56094+00	0.77235+02	0.97330+00	0.56094+00	0.77235+02
3	0.79174+02	0.51283+00	0.70974+00	0.94622+00	0.32774+02	0.56030+00	0.94622+00	0.32774+02
4	0.57425+02	0.29706-01	0.77239+00	0.14526-01	0.65083+01	0.66127-01	0.14526-01	0.65083+01
5	0.12659+02	0.52532-05	0.20686-00	0.47811-04	0.21372+01	0.15521-01	0.47811-04	0.21372+01
6	0.12829+02	0.25793-11	0.59411-01	0.13194-10	0.70261+00	0.19505-02	0.10282-10	0.70261+00
7	0.21416+02		0.79522-02		0.49001-02	0.11102-03		0.49001-02
8	0.65544+01		0.11531-02		0.27065-09	0.96683-05		0.67689-10
9	0.19410+02		0.92830-04			0.57259-06		
10	0.44205+01		0.90118-05			0.48091-07		
11	0.24229+01		0.52165-06			0.34047-08		
12	0.22197+01		0.38238-07			0.18247-09		
13	0.34601+01		0.17295-08			0.11505-10		
14	0.57337+01		0.10136-09					
15	0.94285+01		0.37496-10					

TABLE VII
A Comparison of the Different Methods for Example 6

Method	Num. of iter.	Method	Num. of iter.	Method	Num. of iter.
D	18	DJD	18	DIM	16
GD(3)	17	BJD(3)	18	BIM(3)	18
GD(5)	18	BJD(5)	19	BIM(5)	19
GD(21)	15	BJD(21)	15	BIM(21)	15
GD(51)	16	BJD(51)	16	BIM(51)	15
GD(101)	13	BJD(101)	11	BIM(101)	11
GD(151)	16	BJD(151)	14	BIM(151)	14
GD(201)	11	BJD(201)	10	BIM(201)	10
		JD	8	NIM	8

methods over the GD method. The DIM used less number of iterations when compared to the DJD and Davidson methods. Convergence of the NIM and JD methods was similar.

EXAMPLE 7. The total energy of the ground state of the BC molecule was calculated by the configuration interaction (CI) method as implemented in the MRD-CI program [67–71]. For correct comparison, the DJD and DIM methods were used in this program directly without modification to calculate the lowest eigenvalue of the CI matrix along with the Davidson method, which is usually employed in this program. The present investigation points out, however, that a more efficient method could be proposed for calculation of lowest roots of the CI matrix [72]. The order of the CI matrix was equal to 267,985. The iterations were terminated when the Euclidean norm of the residual vector was less than 10^{-6}. All three methods showed similar convergence in these calculations. A solution was obtained at 18th iteration step by each method.

VI. CONCLUSIONS

A new iterative method for calculation of selected lowest eigenvalues and corresponding eigenvectors of the generalized eigenvalue problem is proposed in this work together with its band and diagonal versions. The band and diagonal Jacobi–Davidson methods have been proposed also by employing the corresponding approximation to the Jacobi–Davidson method. These five new methods together with the Davidson, generalized Davidson, and Jacobi–Davidson methods form three groups of methods. The first group includes the Davidson and generalized Davidson methods, the second one is formed by the Jacobi–Davidson method and its simplified versions, while the new iterative method together with its diagonal and band versions forms the third group of methods.

Numerical tests show that the convergence property in the Jacobi–Davidson and the Newton type methods is improved from diagonal to the full methods. However, for the Davidson-type methods this is not correct. The generalized Davidson method outperforms the Davidson method only for narrow band matrices. With wide band matrices, convergence of the generalized Davidson method is bad. This is a consequence of the fact that with increasing bandwidth Eq. (4) becomes degenerate. Summarizing the results of all test calculations we can conclude that the convergence of the Jacobi–Davidson and Newton type methods is better when compared to the Davidson type method. This is due to the

orthogonality property of the correction vector of the Jacobi–Davidson and the new iterative methods. An advantage of the Newton type methods over the Jacobi–Davidson type methods that can be observed in the presented results is connected with the general approach to the derivation of a correction vector. The equation of a correction vector in the Jacobi–Davidson type method can be derived by assuming that variation of the residual vector is equal to zero, while the equation of a correction vector in the Newton type methods is obtained by assuming that a variation of the quadratic approximation of the functional (6) is equal to zero.

ACKNOWLEDGMENTS

The author is grateful to Professors P. Jensen and R. J. Buenker for valuable discussion.

REFERENCES

1. R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics* (Academic Press, London/New York, 1969).
2. I. Shavitt, The method of configuration interaction, in *Mehtods of Electronic Structure Theory*, edited by H. F. Shaefer III (Plenum, New York, 1977), p. 189.
3. C. F. Fischer, *The Hartee–Fock Method for Atoms: A Numerical Approach* (Wiley, New York, 1977).
4. E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations* (McGraw–Hill, New York, 1955).
5. C. Lanczos, An iterative method for the solution of the eigenvalue problem of linear differential and integral operators, *J. Res. Natl. Bur. Stand.* **45**, 255 (1950).
6. 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**, 87 (1975).
7. W. Karush, An iterative method for finding characteristic vectors of a symmetric matrix, *Pacific J. Math.* **1**, 233 (1951).
8. C. C. Page, Computational variants of the Lanczos method for the eigenproblem, *J. Inst. Math. Appl.* **10**, 373 (1972).
9. W. Butscher and W. E. Kammer, Modification of Davidson's method for the calculation of eigenvalues and eigenvectors of large real-symmetric matrices: "Root homing procedure," *J. Comput. Phys.* **20**, 313 (1976).
10. B. Liu, The simultaneous expansion method, in *Numerical Algorithms in Chemistry: Algebraic Methods*, edited by C. Moler and I. Shavitt (Lawrence Berkley Lab., Univ. of California, 1978), p. 49.
11. B. N. Parlett, The Lanczos algorithm with selective orthogonalization, *Math. Comput.* **33**, 217 (1979).
12. B. N. Parlett, *The Symmetirc Eigenvalue Problem* (Prentice Hall, Englewood Cliffs, NJ, 1980).
13. T. Ericsson and A. Ruhe, The spectral transformation Lanczos method for the numerical solution of large sparse generalized symmetric eigenvalue problems, *Math. Comput.* **35**, 1251 (1980).
14. T. Z. Kalamboukis, Davidson's algorithm with and without perturbation correction, *J. Phy. A: Math. Gen.* **13**, 57 (1980).
15. E. R. Davidson, Comments on the Kalamboukis tests of the Davidson agorithm, *J. Phys. A: Math. Gen.* **13**, L179 (1980).
16. J. Cullum and R. A. Wiloughby, Computing eigenvalues of very large symmetric matrices—An implementation of a Lanczos algorithm with no reorthogonalization, *J. Comput. Phys.* **43**, 329 (1981).
17. H. D. Simon, Analysis of the symmetric Lanczos algorithm with reorthogonalization methods, *Linear Algebra Appl.* **61**, 103 (1984).
18. H. D. Simon, The Lanczos algorithm with partial reorthogonalization, *Math. Comput.* **42**, 115 (1984).
19. N. Kosugi, Modification of the Liu–Davidson method for obtaining one or simultaneously several eigensolutions of large real-symmetric matrix, *J. Comput. Phys.* **55**, 426 (1984).
20. J. K. Cullum and R. A. Wiloughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (Birkhäuser, Boston, 1985), Vol. 1.

21. D. M. Wood and A. Zunger, A new method for diagonalising large matrices, *J. Phys. A: Math. Gen.* **18**, 1343 (1985).
22. R. B. Morgan and D. S. Scott, Generalizations of Davidson's method for computing eigenvalues of sparse symmetric matrices, *SIAM J. Sci. Stat. Comput.* **7**, 817 (1986).
23. V. M. Umar and C. F. Fischer, Multitasking the Davidson algorithm for the large, sparse eigenvalue problem, *Int. J. Supercomput. Appl.* **3**, 28 (1989).
24. J. D. Kress, S. W. Woodruff, G. A. Parker, and R. T. Pack, Some strategies for enhancing the performance of the block Lanczos method, *Comput. Phys. Comm.* **53**, 109 (1989).
25. B. Noud-Omid, Applications of the Lanczos method, *Comput. Phys. Comm.* **53**, 157 (1989).
26. B. N. Parlett and B. Nour-Omid, Towards a block Lanczos program, *Comput. Phys. Comm.* **53**, 169 (1989).
27. J. H. van Lenthe and P. Pulay, A space-saving modification of Davidson's eigenvector algorithm, *J. Comput. Chem.* **11**, 1164 (1990).
28. J. Olsen, P. Jørgensen, and J. Simons, Passing the one-billion limit in full configuration-interaction (FCI) calculations, *Chem. Phys. Lett.* **169**, 463 (1990).
29. C. W. Murray, S. C. Racine, and E. R. Davidson, Improved algorithms for the lowest few eigenvalues and associated eigenvectors of large matrices, *J. Comput. Phys.* **103**, 382 (1992).
30. R. B. Morgan and D. S. Scott, Preconditioning the Lanczos algorithm for sparse symmetric eigenvalue problems, *SIAM J. Sci. Comput.* **14**, 585 (1993).
31. T. Koslowski and W. von Nissen, Linear combination of Lanczos vectors: A storage-efficient algorithm for sparse matrix eigenvector computations, *J. Comput. Chem.* **14**, 769 (1993).
32. R. G. Grimes, J. G. Levis, and M. D. Simon, A shifted block Lanczos algorithm for solving sparse symmetric generalized eigenproblems, *SIAM J. Matrix Anal. Appl.* **15**, 228 (1994).
33. D. Calvetti, L. Reichel, and D. Sorensen, An implicitly restarted Lanczos method for large symmetric eigenvalue problems, *Electron. Trans. Numer. Anal.* **2**, 1 (1994).
34. J. M. Bofill and J. M. Anglada, Some remarks on the use of the three-term recurrence method in the configuration interaction eigenvalue problem, *Chem. Phy.* **183**, 19 (1994).
35. M. Crouzeix, B. Philippe, and M. Sadkane, The Davidson method, *SIAM J. Sci. Comput.* **15**, 62 (1994).
36. F. X. Gadea, Large matrix diagonalization, comparison of various algorithms and a new proposal, *Chem. Phys. Lett.* **227**, 201 (1994).
37. A. Stathopoulos, Y. Saad, and C. F. Fischer, Robust preconditioning of large, sparse, symmetric eigenvalue problems, *SIAM J. Comput. Appl. Math.* **64**, 197 (1995).
38. H. Dachsel and H. Lischka, An efficient data compression method for the Davidson subspace diagonalization scheme, *Theor. Chim. Acta* **92**, 339 (1995).
39. G. L. G. Sleijpen, J. G. L. Booten, D. R. Fokkema, and H. A. Van der Vorst, A Jakobi–Davidson type methods for generalized eigenproblems and polynomial eigenproblems, *BIT* **36**, 595 (1996).
40. F. Webster and G.-C. Lo, Projective block Lanczos algorithm for dense, hermitian eigensystems, *J. Comput. Phys.* **124**, 146 (1996).
41. H. J. J. van Dam, J. H. van Lenthe, G. L. G. Sleijpen, and H. A. van Der Vorst, An improvement of Davidson's iteration method: Applications to MRCI and MRCEPA calculations, *J. Comput. Chem.* **17**, 267 (1996).
42. G. L. G. Sleijpen and H. A. van Der Vorst, A Jakobi–Davidson iteration method for linear eigenvalue problems, *SIAM J. Matrix Anal. Appl.* **17**, 401 (1996).
43. A Strathopoulos, Y. Saad, and K. Wu, Dynamic thick restarting of the Davidson and the implicitly restarted Arnoldi method, *SIAM J. Sci. Comput.* **19**, 227 (1998).
44. L. Borges and S. Oliveira, A parallel Davidson-type algorithm for several eigenvalues, *J. Comput. Phys.* **144**, 727 (1998).
45. M. Genseberger and G. L. G. Sleijpen, Alternative correction equation in the Jakobi–Davidson method, *Numer. Linear Algebra Appl.* **6**, 235 (1999).
46. J. K. Cullum and R. A. Wiloughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* (Birkhäuser, Boston, 1985), Vol. 2.
47. J. Weber, R. Lacroix, and G. Wanner, The eigenvalue problem in configuration interaction calculations: S computer program based on a new derivation of the algorithm of Davidson, *Comput. Chem.* **4**, 55 (1980).

48. G. Cisneros and C. F. Bunge, An improved computer program for eigenvector and eigenvalues of large configuration interaction matrices using the algorithm of Davidson, *Comput. Chem.* **8**, 157 (1984).
49. G. Cisneros, M. Berrondo, and C. F. Bunge, DVDSON: A subroutine to evaluate selected sets of eigenvalues and eigenvectors of large symmetric matrices, *Comput. Chem.* **10**, 281 (1986).
50. A Stathopoulos and C. F. Fischer, A Davidson program for finding a few selected extreme eigenpairs of a large, sparse, real, symmetric matrix, *Comput. Phys. Comm.* **79**, 268 (1994).
51. D. S. Scott, Solving sparse symmetric generalized eigenvalue problems without factorization, *SIAM J. Numer. Anal.* **18**, 102 (1981).
52. H. A. van der Vorst, A generalized Lanczos scheme, *Math. Comput.* **39**, 559 (1982).
53. G. A. Gallup, Iterative calculation of eigenvalues and eigenvectors of large, real matrix system with overlap, *J. Comput. Chem.* **3**, 127 (1982).
54. R. B. Morgan, Davidson's method and preconditioning for generalized eigenvalue problems, *J. Comput. Phys.* **89**, 241 (1990).
55. A. V. Mitin, Iterative methods for the calculation of a few of the lowest eigenvalues and corresponding eigenvectors of the $AX = \lambda BX$ equation with real symmetric matrices of large dimension, *J. Comput. Chem.* **15**, 747 (1994).
56. S. F. Boys, Electronic wave functions. II. A calculation for the ground state of the beryllium atom, *Proc. R. Soc. London A* **201**, 125 (1950).
57. R. K. Nesbet, Algorithm for diagonalization of large matrices, *J. Chem. Phys.* **43**, 311 (1965).
58. E. R. Davidson, Monster matrices: Their eigenvalues and eigenvectors, *Comput. Phys.* **7**, 519 (1993).
59. D. K. Faddeev and V. N. Faddeeva, *Computational Methods of Linear Algebra* (Freeman, San Francisco, 1963). [English translation]
60. R. Fletcher, *Practical Methods of Optimization* (Wiley, Chichester/New York, 1987).
61. C. C. J. Roothaan and P. S. Bagus, Atomic self-consistent field calculations by the expansion method, in *Methods in Computational Physics*, edited by B. Alder, S. Fernbach, and M. Rotenberg (Academic Press, New York/London, 1963), Vol. 2, p. 47.
62. O. Axelsson, *Iterative Solution Methods* (Cambridge Univ. Press, Cambridge, UK, 1994).
63. B. G. Galerkin, Sterzni I plastinki. Rjadi v nekotorih voprosah yprygogo ravnovesija sterznei i plastinok, *Vestn. Ingenerov* **1**, 897 (1915).
64. T. F. Chan and H. A. Van der Vorst, Approximate and incomplete factorization, in *Parallel Numerical Algorithms*, edited by D. E. Keyes, A. Sameh, and V. Venkatakrishnan (Kluwer Academic, Dordrecht, 1997), p. 167.
65. J. H. Wilkinson, *The Algebraic Eigenvalue Problem* (Clarendon, Oxford, 1965).
66. G. E. Forsythe, M. A. Malcolm, and C. B. Mouler, *Computer Methods for Mathematical Calculations* (Prentice Hall, Englewood Cliffs, NJ, 1977).
67. R. J. Buenker and S. D. Peyerimhoff, Energy extrapolation in CI calculations, *Theor. Chim. Acta* **39**, 217 (1975).
68. R. J. Buenker and S. D. Peyerimhoff, *Ab initio* calculations close to the full CI level of accuracy and their use for the interpretation of molecular spectra, in *New Horizons of Quantum Chemistry*, edited by P.-O. Löwdin and B. Pullman (Reidel, Dordrecht, 1976), p. 183.
69. R. J. Buenker, S. D. Peyerimhoff, and W. Butcher, Applicability of the multi-reference double-exitation CI (MRD-CI) method to the calculation electronic wavefunctions and comparison and related techniques, *Mol. Phys.* **35**, 771 (1978).
70. R. J. Buenker, Combining perturbation theory techniques with variational CI calculations to study molecular exited states, *Int. J. Quantum Chem.* **29**, 435 (1986).
71. D. B. Knowles, J. R. Alvarez-Collado, G. Hirsch, and R. J. Buenker, Comparison of perturbatively corrected energy results from multiple reference double-excitation configuration-interaction method calculations with exact full configuration-interaction benchmark values, *J. Chem. Phys.* **92**, 585 (1990).
72. A. V. Mitin, to be published.