数理解析研究所講究録
483巻 1983年 181-195

分子科学における大次元行列固有値問題に対するアレイプロセッサ向け解法

東大理　小杉信博（Nobuhiro Kosugi）

分子科学において，大次元行列は多電子問題を扱う際に出現する。多電子関数は一電子問題の解である分子軌道関数を含む規格直交系（有限）を成す一電子関数の積（電子配置関数）で展開される。その展開係数を変分的に求める際に，ハミルトニアン行列（配置間相互作用行列）と呼ばれる大次元行列の固有値問題が出現する。この分野でよく用いられるのは緩和法に基づく方法であり，最近では$10^6 \times 10^6$次元の行列の最低固有値問題が解かれている。ここでは，多根を同時に求めることにより，従来の解法では本質的に避けることのできないI/Oバウンドの問題を解消させ，しかも，非零要素だけ扱っていながらベクトル処理も可能になる新しいアルゴリズムを紹介する。この分野における従来の解法の紹介と本アルゴリズムの特徴，そのプログラムEMOR1（東大大型計算機センターに登録済）の性能についても述べる。

I. INTRODUCTION

The relaxation method, which had been originally developed by Southwell[1] for solving the coupled linear equation within the framework of the iteration (successive approximation) method, was applied to the eigenvalue problem by Cooper[2]. Large-scale matrix eigenproblems are familiar to quantum chemists in configuration interaction calculations of electronic wavefunctions of atoms and molecules. Several quantum chemists [3-5] have adopted the relaxation method for diagonalization of large Hamiltonian matrices. The original relaxation method, i.e., coordinate relaxation method, improves one element in the trial vector at a time based on the linear (first-order) minimization of the Rayleigh quotient (expectation value of Hamiltonian, in quantum mechanics) or the linear vanishing of the residual. Within the coordinate relaxation algorithm, Shavitt et al.[6] corrected the trial vector based on the quadratic (exact) minimization of the Rayleigh quotient. Their method, which is called the method of optimal (coordinate) relaxations (MOR), is the most widely used algorithm for the non-degenerate lowest eigenproblem in quantum chemistry.

Some modification of the coordinate relaxation method has been presented: the simultaneous improvement of several elements in the trial vector[7,8]. This group-coordinate relaxation method is a natural extention of the two-by-two Ritz iteration algorithm[2], and is effective to resolve convergence difficulties for nearly degenerate eigensolutions.

Algorithms to generate only one correction vector to improve all elements in the trial vector at a time have also been presented. The gradient method by Hestenes and Karush[9] is based on the two-by-two Ritz iteration algorithm, but gives much poorer convergence than the above-mentioned relaxation methods. In order to remedy the gradient method

generation of a series of correction vectors (expansion vectors) is necessary. The Lanczos method[10], which originally is a modification of the power method based on the Krylov sequence, can be regarded as an extention of the gradient method. The Lanczos method has some weak points but is widely used, especially, by the nuclear physicists.

Another algorithm to generate only one correction vector to improve all elements in the trial vector is hit on, in which the correction vector has individual coordinate relaxations. This algorithm may be better than the gradient method, but undoubtedly gives slower convergence than the relaxation methods which update continuously a trial vector as soon as one or several elements are corrected. In order to remedy this algorithm and the Lanczos method, Davidson[11] proposed the use of this type of correction vectors as expansion vectors in a Lanczos-like algorithm. Davidson's method, which is also called a modified MOR (MMOR), overcomes the convergence difficulties for nearly degenerate eigensolutions by repeating the Ritz iteration every time to obtain a new correction vector, and is widely used by quantum chemists as well as the MOR[6].

Some schemes for obtaining higher eigensolutions have been proposed within the above-mentioned methods. In the methods adopting the Ritz iteration, the orthogonality-constraint procedure[6,9] of the trial vector to lower eigenvectors is effective; in the other methods, the root-shifting procedure[2,9] to mimic the deflation is effective. Against these procedures, the origin-shift version of the MOR[6] by Feler[12] and the root-homing version of the MMOR[11] by Butscher and Kammer[13] make possible the direct determination of higher eigensolutions without knowledge of the exact lower ones.

Recently new methods to obtain several or many eigensolutions, which

are not based on the sequential iteration algorithms to need knowledge of the exact lower solutions in advance but on a simultaneous iteration algorithm for all the trial solutions to be corrected, have been developed by taking into account the computational efficiency[8,14-19]. Clint and Jennings[14] combined the power method with a Ritz-like diagonalization algorithm. Cheung and Bishop[8] corrected simultaneously several trial vectors within the group-coordinate relaxation method. Raffenetti[15] and Liu[16] (See ref.[17], too.) have proposed simultaneous versions of the MOR by Shavitt et al.[6] and of the MMOR by Davidson[11], respectively. Golebiewski[18] combined the power method with a new orthogonalization procedure, but his method is basically equivalent to the single-premultiplication version of the Jennings method[14]. Iwata[19] extended the gradient method based on the two-by-two Ritz iteration algorithm[9] with implicit incorporation of the orthogonalization among trial and correction vectors and even previously solved eigenvectors.

In the present work, a new algorithm for solving one or simultaneously several eigensolutions is presented by taking into account both advantages of the group-coordinate relaxation method[7,8] and of the simultaneous expansion method by Liu[16]. The present algorithm is especially suitable for array processors equipped with recent high-speed (super) computers.

II. AN EXTENDED METHOD OF OPTIMAL RELAXATION

1. Scheme

The Ritz iteration algorithm is getting powerful as the correction-vector space is increasing[8]. The group-coordinate relaxation method keeps the number of correction vectors fixed in each group[7,8]. On the other hand, Liu's method increases the number of correction vectors by the

number of sought solutions each iteration cycle[16]. His method has the advantage of the complete use of the Ritz iteration but a disadvantage to be liable to suffer an overflow of the main-memory space required.

The method proposed in the present work keeps the number of correction vectors fixed through the iteration cycles. Correction vectors for the renewed trial vectors are individually obtained in each iteration cycle after Davidson's method[11] in the same way as in Liu's method[16], but unimportant constituents of the correction vectors obtained in the previous iteration cycle are neglected in the next cycle. This point is different from the Davidson-Liu method in which the correction-vector space used in an iteration cycle is basically always (within the limits of the main-memory space admitted) included in the successive iteration cycles. The meaning to keep constituents of old correction vectors in the Ritz iteration does not lie in the concentration of sought solutions as in the Lanczos method[10] but in the optimal extrapolation within the framework of the Ritz iteration algorithm. A criterion for important constituents of correction vectors is the contribution to the correction for the trial vectors; that is, the trial vectors renewed in the previous iteration cycles include important constituents of correction vectors. The present algorithm does not keep old correction vectors but keeps old trial vectors in the successive iteration cycles.

2. Procedures

A. Initialization

1. Form initial guess vectors $\mathbf{C}^{(0)}$ (dimension $N \times n_{solv}$; n_{solv} is the number of eigensolutions to be sought.) from $\mathbf{c}^{(0)}$ by solving a small eigenvalue problem: $\mathbb{X}^{(0)}\mathbf{c}^{(0)} = \mathbf{c}^{(0)}\mathbb{E}^{(0)}$, where the matrix $\mathbf{X}^{(0)}$ of dimension N_{guess} has important elements selected from the full matrix $\mathbf{X}$ of dimension

N ($N_{guess} \ll N$), and $\mathbb{E}^{(0)}$ is an initial guess for sought eigenvalues of $\mathbb{X}$.

2. Form $\mathbb{D}^{(0)} = \mathbb{X}\,\mathbb{C}^{(0)}$, where the dimension of $\mathbb{D}^{(0)}$ is $N \times n_{solv}$.

3. Form the initial residual vectors $\mathbf{q}_k^{(0)} = \mathbb{D}_k^{(0)} - E_k^{(0)}\mathbb{C}_k^{(0)}$, ($k = 1 \sim n_{solv}$).

4. Form the initial correction vectors $\mathbf{b}_k^{(0)} = (\mathbf{X}^{diag} - E_k^{(0)}\mathbf{1})^{-1}\cdot\mathbf{q}_k^{(0)}$, where the diagonal matrix $\mathbf{X}^{diag}$ has the diagonal elements of $\mathbf{X}$. If the number of correction vectors, n_{corr}, is greater than n_{solv}, scatter $\mathbf{b}_k^{(0)}$ ($k = 1 \sim n_{solv}$) over $\mathbb{b}^{(0)}$ (dimension $N \times n_{corr}$) by dividing $\mathbf{b}_k^{(0)}$ of dimension N into several parts.

B. Iteration ($i = 0, 1, \ldots$)

5. Form the orthonormalized correction vectors $\mathbf{C}_{corr}^{(i)}$ from $\mathbb{b}^{(i)}$ by Schmidt orthogonalization; then, ${}^t(\mathbb{C}^{(i)}, \mathbb{C}_{corr}^{(i)})\cdot(\mathbb{C}^{(i)}, \mathbb{C}_{corr}^{(i)}) = \mathbf{1}$.

6. Form $\mathbb{D}_{corr}^{(i)} = \mathbf{X}\,\mathbb{C}_{corr}^{(i)}$.

7. Form $\mathbf{H}^{(i)} = {}^t(\mathbb{C}^{(i)}, \mathbb{C}_{corr}^{(i)})\cdot(\mathbb{D}^{(i)}, \mathbb{D}_{corr}^{(i)})$, and solve the reduced eigenvalue problem (Ritz iteration) of dimension $n_{solv} + n_{corr}$: $\mathbf{H}^{(i)}\mathbf{T}^{(i,i+1)} = \mathbf{T}^{(i,i+1)}\mathbf{E}^{(i+1)}$.

8. Form $\mathbb{C}^{(i+1)} = (\mathbb{C}^{(i)}, \mathbf{C}_{corr}^{(i)})\cdot\mathbf{T}^{(i,i+1)}$ and $\mathbb{D}^{(i+1)} = (\mathbb{D}^{(i)}, \mathbb{D}_{corr}^{(i)})\cdot\mathbf{T}^{(i,i+1)}$

9. Form $\mathbf{q}_k^{(i+1)} = \mathbb{D}_k^{(i+1)} - E_k^{(i+1)}\mathbb{C}_k^{(i+1)}$, and check convergence by $|\mathbf{q}_k^{(i+1)}|^2 / |E_k^{(i+1)}|^2$.

10. Form $\mathbf{b}_k^{(i+1)} = (\mathbf{X}^{diag} - E_k^{(i+1)}\mathbf{1})^{-1}\cdot\mathbf{q}_k^{(i+1)}$ for unconverged $\mathbf{C}_k^{(i+1)}$. If $n_{corr} > n_{solv} - n_{conv}$ (n_{conv} is the number of converged solutions), fill $\mathbb{b}^{(i+1)}$ with $\mathbf{C}^{(i)}$ and, if necessary, $\mathbf{C}_{corr}^{(i)}$.

11. Return to step 5 with $i = i + 1$.

3. Some Comments

The large, real-symmetric matrix $\mathbf{X}$ to be diagonalized should be

arranged so that the diagonal elements $\mathbb{X}^{diag}$ may be in ascending order, especially when the matrix $\mathbf{X}$ is diagonally dominant, because it is easy to construct initial trial vectors for lowest eigensolutions by solving the small eigenvalue problem for the submatrix $\mathbf{X}^{(0)}$ of dimension N_{guess} (step 1). The off-diagonal elements of $\mathbb{X}$ need not be used in any specific sequence; then, only nonzero elements can be stored in auxiliary memory in arbitrary order.

The $\mathbb{X}$ matrix multiplication steps (1 and 6) are extremely time-consuming, even if limited to nonzero multiplication, and are bound by I/O processing. In Davidson's method and its modifications multiplication and accumulation per one off-diagonal element are performed only twice per iteration cycle. The present algorithm, in which the multiplication and accumulation are performed simultaneously in proportion to n_{solv} or n_{corr}, is capable of avoiding the I/O-binding and is suitable for the recent and near-future array processors which have very short vector/scalar-crossing loop-length ($\lesssim 5$).

In order to avoid round-off errors, explicit orthogonalization is required even among trial vectors $\mathbb{C}^{(i)}$ in step 5 and the procedure $\mathbb{D}^{(i)} = \mathbf{X}\,\mathbb{C}^{(i)}$ (step 6') is required before step 7. In the program EMOR1[20], the explicit orthogonalization and step 6' are performed every iteration cycle and every five cycle, respectively. Because the Schmidt orthogonalization procedure is completely adapted for vector processing and its computational time is negligible compared with the $\mathbf{X}$ matrix multiplication, the implicit orthogonalization algorithms[10,19] will not be useful.

The computer program (named EMOR1) coded in FORTRAN 77 by the present author is opened in the Computer Centre, the University of Tokyo[20]. The small eigenvalue problems in steps 1 and 7 are solved with a Householder-

bisection-QR-inverse-iteration routine modified from the original version[21] by the present author so as to adapt for vector processing. Because the root-homing pattern search[13] is incorporated in EMOR1, correction vectors only for a specified trial vector(s) can be formed (but they are not ineffective to the other solutions). In the present version of EMOR1 on the HITAC M-280H computer with an integrated array processor of the Computer Centre, the University of Tokyo, the maximum values for parameters are as follows: $N_{guess} \leqq 750$; $(N, n_{solv} + n_{corr}) \leqq (13000, 26)$, $(20000, 17)$, or $(30000, 10)$. The extension to non-symmetric matrix eigenproblems and generalized eigenproblems ($\mathbf{X}\, \mathbf{C}_k = E_k\, \mathbb{Y}\, \mathbb{C}_k$) is under consideration.

III. TEST PROBLEMS

1. Results

The program EMOR1 was applied to five test problems. Structure of five test matrices, which are the original Nesbet[4] (matrix $\mathbb{A}$) and modified Nesbet matrices, is summarized in Table I. Eigenvalues for the test matrices are summarized in Table II.

In Tables III, IV and V, test results for many sets of parameters in the matrices $\mathbb{A}$, $\mathbb{B}$ and $\mathbb{C}$ are shown, respectively. Through the problems N_{guess} is equal to n_{solv}; therefore, the maximum value of $\mathbf{q}^2$ for initial guess vectors, $\mathbf{q}^2_{guess}$, is rather large. We dared to solve the problems from poor initial guesses because eigenvalue problems for matrices of dimension ~300 can be easily solved by standard direct-diagonalization algorithms. For the matrix $\mathbb{A}$ the condition $n_{corr} = n_{solv}$ is quite enough except that $n_{solv} = 1$. The condition $n_{corr} = n_{solv}/2$ is sufficient when $n_{solv} \gtrsim 8$. On the other hand, the condition $n_{corr} = n_{solv}$ is not sufficient when $n_{solv} \lesssim 6$ for $\mathbb{B}$ and even when $n_{solv} = 10$ for $\mathbb{C}$. For $\mathbb{C}$ the parameter n_{corr} should be

chosen to be at least $4 \cdot n_{solv}$ ($n_{solv} \leqq 2$) and about 12 ($n_{solv} \geqq 4$); that is, the number of effective correction vectors does not depend on n_{solv} when n_{solv} becomes large.

Results of eigenvalue problems for the matrices $\mathbb{D}$ and $\mathbb{E}$ are shown in Tables VI and VII, respectively. Through the problems, the parameter n_{solv} is fixed to 10; the parameters N_{guess} and n_{corr} are varied. For the matrix $\mathbb{D}$ when $N_{guess} = 10$, the conditions $n_{corr} = 20$ and 30 give the same result. The reason is that nearly linear dependence among correction vectors is encountered when $n_{corr} \gtrsim 20$. The condition that $(N_{guess}, n_{corr}) = (50, 20)$ or (100, 10) is sufficient. When $N_{guess} = 200$, the problem is nearly solved within the initial guess. For the matrix $\mathbb{E}$, lowest two eigenvalues are unusual(Table II). The parameter N_{guess} should be chosen to be at least ~100 and in practice ~300.

2. Discussion

Choice of the parameters N_{guess} and n_{corr} is quite important for efficiently solving eigenproblems. The larger is N_{guess}, the better is convergence; on the other hand, when n_{corr} is too large convergence is ill-behaved, because nearly linear dependence is encountered. An optimum value for the parameter n_{corr} is strongly dependent on structure of a matrix in problem. The value $\mathbf{q}^2_{guess}$ is a criterion in determining an optimum n_{corr}. When n_{solv} is greater than ~10, we may choose n_{corr} to be nearly constant (less than n_{solv}). This feature might originate in a similar principle as in the group-coordinate relaxation method[7,8].

Even when only the lowest eigensolution is desired, it is preferred to choose n_{solv} to be greater than 1; here, correction vectors of number n_{corr} are all for the lowest solution. Trial vectors for the other solutions will work in order to exclude constituents of the other solutions from a trial

vector for the lowest one. This discussion is also valid for cases that only an eigensolution except the lowest one or some interior eigensolutions are desired. The present algorithm for directly determining higher eigensolutions are superior to the method proposed by Butscher and Kammer[13].

If the number n_{solv} of eigensolutions to be sought is too large to store all the vectors of $n_{solv} + n_{corr}$ in the main-memory space admitted, the partition such as $n_{solv} = n_{solv1} + n_{solv2} + \ldots$ is required and the problem is solved for n_{solv1}, n_{solv2}, and so on, sequentially from the lower solutions. Solved eigenvectors are saved in auxiliary memory and are read out and used only at Schmidt orthogonalization among them and initial guess vectors and among them and current trial and correction vectors (step 5). Considering the above test results, the number n_{solv} should be partitioned so as to include as many vectors $(n_{solvi} + n_{corri})$ as possible.

As shown in the above test results, the Ritz iteration algorithm to use the "trial" and "correction" vectors is powerful for many types of eigenvalue problems.

REFERENCES

1. A.Pellew and R.V.Southwell, Proc.Roy.Soc.(London) A175 (1940), 262.
2. J.L.B.Cooper, Quart.Appl.Math. 6 (1948), 179.
3. S.F.Boys, Proc.Roy.Soc.(London) A201 (1950), 125.
4. R.K.Nesbet, J.Chem.Phys. 43 (1965), 311.
5. I.Shavitt, J.Computational Phys. 6 (1970), 124.
6. I.Shavitt, C.F.Bender, A.Pipano and R.P.Hosteny, J.Computational Phys. 11 (1973), 90.
7. Z.Falk, Z.Angew.Math.Mech. 53 (1973), 73.
8. L.P.Cheung and D.M.Bishop, Computer Phys.Commun. 13 (1977), 247.
9. M.R.Hestenes and W.Karush, J.Res.Nat.Bur.Stand. 47 (1951), 45.
10. C.Lanczos, J.Res.Nat.Bur.Stand. 45 (1950), 255.
11. E.R.Davidson, J.Computational Phys. 17 (1975), 87.
12. M.G.Feler, J.Computational Phys. 14 (1974), 341.
13. W.Butscher and W.E.Kammer, J.Computational Phys. 20 (1976), 313.
14. M.Clint and A.Jennings, Computer J. 13 (1970), 76.
15. R.C.Raffenetti, J.Computational Phys. 32 (1979), 403.
16. B.Liu, The Simultaneous Expansion Method, in "Report on the Workshop: Numerical Algorithms in Chemistry: Algebraic Methods" (C.Moler and I. Shavitt, ed.). P.49, Lawrence Berkeley Lab., Univ. of California, 1978.
17. E.R.Davidson, J.Phys. A13 (1980), L179.
18. A.Golebiewski, Int.J.Quantum Chem. XV (1979), 693.
19. S.Iwata, J.Computational Phys., to be published.
20. N.Kosugi, Program EMOR1, Program Library, The Computer Centre, The Univ. of Tokyo, 1982.
21. Y.Beppu and I.Ninomiya, Computer Phys.Commun. 23 (1981), 123; Computers & Chemistry 6 (1982), 87.

TABLE I

Structure of Test Matrices[a]

matrix dimension		matrix elements		No. of non-zero elements	diagonal dominance [%]	density [%]
X	N	X_{ii}	$X_{ij}(=X_{ji})$			
A	300	2i-1	1	45150	100	100
B	300	1.0+0.1×(2i-1)	1	45150	96.0	100
C	300	1.00+0.01×(2i-1)	1	45150	68.8	100
D	1000	2i-1	$\begin{cases} 0, \lvert i-j\rvert \geq 50 \\ 1, \lvert i-j\rvert < 50 \end{cases}$	48775	100	9.75
E	1000	1.0+0.1×(2i-1)		48775	98.8	9.75

[a] diagonal dominance = ratio of the case which satisfies that $|X_{ij}|<|X_{ij}-X_{jj}|$; density = ratio of non-zero elements

TABLE II

Lowest Eigenvalues of Test Matrices

n	A	B	C	D	E
1	0.2355346	0.1296170	0.01303906	0.2791881	-4.456670
2	2.262109	0.3336875	0.03346562	2.316219	-2.594780
3	4.278451	0.5362786	0.05373813	4.339914	0.07319100
4	6.290699	0.7382596	0.07394690	6.358201	0.2732267
5	8.300687	0.9398978	0.09411976	8.373496	0.4739468
6	10.30922	1.141313	0.1142692	10.38687	0.6756589
7	12.31674	1.342569	0.1344020	12.39891	0.8781389
8	14.32349	1.543706	0.1545223	14.40997	1.081195
9	16.32966	1.744750	0.1746327	16.42027	1.284691
10	18.33535	1.945719	0.1947352	18.42997	1.488534

TABLE III

Test Result for Matrix **A**

n_{solv}	n_{corr}	N_{guess}	$\mathbf{q}^2_{guess}$	$n_{it}(-6)$	$n_{it}(-10)$
1	1	1	$0.299\ 10^{+3}$	>20	
	2			6	9
	3			5	7
2	2	2	$0.508\ 10^{+3}$	4	6
	3			3	4
4	2	4	$0.917\ 10^{+3}$	8	9
	4			3	5
6	3	6	$0.132\ 10^{+4}$	5	8
	6			2	3
8	4	8	$0.173\ 10^{+4}$	5	6
	8			2	3
10	5	10	$0.213\ 10^{+4}$	4	5
	10			2	2
15	5	15	$0.311\ 10^{+4}$	5	6
	10			3	5
20	5	20	$0.407\ 10^{+4}$	5	6
	10			4	5

Notation : n_{solv}, number of eigensolutions to be sought simultaneously; n_{corr}, number of correction vectors used in each iteration; N_{guess}, dimension of the small matrix for an initial guess; $\mathbf{q}^2_{guess}$, maximum value of $\mathbf{q}^2$ for initial guess vectors; $n_{it}(-x)$, number of iterations required for converging $\mathbf{q}^2$ to less than 10^{-x}.

TABLE IV

Test Result for Matrix **B**

n_{solv}	n_{corr}	N_{guess}	q^2_{guess}		$n_{it}(-6)$	$n_{it}(-10)$
1	1	1	0.299	10^{+3}	>20	
	2				10	16
	3				7	12
	4				6	9
2	2	2	0.594	10^{+3}	>20	
	4				5	11
	6				4	6
4	4	4	0.118	10^{+4}	7	11
	6				3	5
	8				3	3
6	6	6	0.175	10^{+4}	4	8
	9				2	3
8	8	8	0.232	10^{+4}	3	5
	12				2	2
10	10	10	0.289	10^{+4}	2	4
	15				2	3

TABLE V

Test Result for Matrix **C**

n_{solv}	n_{corr}	N_{guess}	q^2_{guess}		$n_{it}(-6)$	$n_{it}(-10)$
1	2	1	0.299	10^{+3}	13	>20
	3				11	>20
	4				9	17
2	4	2	0.595	10^{+3}	9	>20
	6				6	>20
	8				5	9
4	6	4	0.119	10^{+4}	>20	
	8				3	9
	12				3	4
6	6	6	0.176	10^{+4}	>20	
	9				3	6
	12				3	5
8	8	8	0.233	10^{+4}	>20	
	12				3	4
	16				2	3
10	10	10	0.289	10^{+4}	9	15
	15				2	4

TABLE VI

Test Result for Matrix D

n_{solv}	n_{corr}	N_{guess}	q^2_{guess}		$n_{it}(-6)$	$n_{it}(-10)$
10	10	10	0.327	10^{+3}	13	15
	20				6	8
	30				6	8
10	10	50	0.457	10^{+1}	8	10
	20				6	8
10	10	100	0.481	10^{-1}	5	8
10	10	200	0.188	10^{-7}	1	2

TABLE VII

Test Result for Matrix E

n_{solv}	n_{corr}	N_{guess}	q^2_{guess}		$n_{it}(-6)$	$n_{it}(-10)$
10	10	100	0.369	10^{+2}	>20	
	20				17	>20
	30				13	17
10	10	200	0.801	10^{0}	17	>20
	20				10	16
10	10	300	0.154	10^{-2}	6	12
	20				4	8
10	10	400	0.153	10^{-5}	2	4