Efficient evaluation of the error vector in the direct inversion in the iterative subspace scheme

Rustam Z. Khaliullin [a,b,*], Martin Head-Gordon [a,b], Alexis T. Bell [b,c]

[a] *Department of Chemistry, University of California, Head-Gordon group, Berkeley, CA 94720, United States*
[b] *Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, United States*
[c] *Department of Chemical Engineering, University of California, Berkeley, CA 94720, United States*

Abstract

Faster computation of the Pulay error vector for the direct inversion in the iterative subspace acceleration scheme is achieved by expanding the density matrix in terms of the occupied molecular orbitals and by performing matrix multiplications in an optimal order. The predicted speed-up over the usual implementation is N/O, where N is the size of the atomic orbital basis set, and O is number of occupied molecular orbitals, which is confirmed by test calculations.
© 2005 Elsevier B.V. All rights reserved.

Direct inversion in the iterative subspace (DIIS) is one of the most efficient methods for accelerating the convergence of the self-consistent field (SCF) iterative procedure [1,2]. In this method, the approximate solution on the current iteration is found by an interpolation in the subspace spanned by the results of several previous iterations. The interpolation coefficients are obtained from the so-called 'error vectors', or residuals, which are evaluated on every iteration.

The calculation of the error vector is the most computationally expensive part of the DIIS scheme. The evaluation of the most widely used error vector proposed by Pulay in 1982 [2] scales as N^3, where N is the basis set size. This step takes significant time in each SCF iteration, when performing calculations on large systems. For large systems, linear scaling methods make evaluation of the Coulomb [3–6], exact exchange [7], and exchange-correlation [8,9] parts of the Fock matrix inexpensive compared to matrix manipulations. For example, the DIIS error vector formation takes from 13% to 34% of iteration time in the Hartree–Fock/DFT calculations on

288 water molecules with 6-311g(d,p) or cc-pVDZ basis (around 7000 basis functions) depending on the platform architecture. In this Letter, we propose a faster algorithm for calculating the Pulay error vector.

Usually, the DIIS error vector elements for the closed-shell systems are expressed as:

$$\text{err}_{\mu\nu} = 2\langle \phi'_\mu \mid (\hat{1} - \hat{\rho})\hat{F}\hat{\rho} \mid \phi'_\nu \rangle - \text{h.c.}, \tag{1}$$

where $\mid \phi'_\nu \rangle$ are orthogonalized atomic orbitals, $\hat{\rho}$ is the one-electron density matrix operator, and $\hat{F}$ is the Fock operator. Orthogonalization of the atomic basis set is used to produce a more balanced error vector [2].

The error vector can be re-written:

$$\text{err}_{\mu\nu} = 2\sum_{\tau\lambda\sigma\eta\rho}^{N} X_{\mu\tau}(\delta_{\tau\lambda} - \sum_{\pi}^{N} S_{\tau\pi}R_{\pi\lambda})F_{\lambda\sigma}R_{\sigma\eta}S_{\eta\rho}X_{\rho\nu} - \text{h.c.}$$

$$= 2\sum_{\lambda\sigma\eta\rho}^{N} X_{\mu\lambda}F_{\lambda\sigma}R_{\sigma\eta}S_{\eta\rho}X_{\rho\nu} - \text{h.c.}, \tag{2}$$

where $\mathbf{F}$, $\mathbf{R}$, $\mathbf{S}$ are the Fock, density, and overlap matrices, and $\mathbf{X}$ is the orthogonalization matrix, which usually is taken to be $\mathbf{S}^{-1/2}$.

Traditionally the DIIS error vector is computed according to Eq. (2) and requires four matrix multiplications,

---

*   Corresponding author.
    E-mail address: rustam@khaliullin.com (R.Z. Khaliullin).

Table 1
Speed-ups of the DIIS error vector computation

# of molecules	Basis			
	6-311G	cc-pVDZ	6-311G(d,p)	6-311(++)G(d,p)
9	1.67	2.00	2.25	2.73
18	2.33	2.76	3.12	3.39
36	2.73	3.22	3.84	4.49
72	3.19	3.91	4.66	5.41
144	3.64	4.46	5.34	6.21
288	3.88	4.75	5.82	6.92
$8N^3/(4N^2O + 4NVO)$	**4.38**	**5.36**	**6.55**	**7.74**

each of which has approximately $2N^3$ floating point operations (f.p.o.). When counting f.p.o.'s, we neglect quadratic steps, like matrix subtraction or transpositions.

Eq. (1) can be represented in a slightly different form to achieve better scaling:

$$\text{err}_{\mu\nu} = 2 \sum_a^V \sum_i^O \langle \phi'_\mu \mid \psi_a \rangle \langle \psi_a \mid \hat{F} \mid \psi_i \rangle \langle \psi_i \mid \phi'_\nu \rangle - \text{h.c.}, \quad (3)$$

$$= 2 \sum_{\sigma\lambda}^N \sum_a^V \sum_i^O C'_{\mu a} C_{\lambda a} F_{\lambda\sigma} C_{\sigma j} C'_{\nu j} - \text{h.c.}, \quad (4)$$

where $|\psi_k\rangle$ are spatial molecular orbitals, O, V are the number of occupied and virtual spatial molecular orbitals, C, C' are molecular orbital coefficients in the atomic orbital and orthogonalized atomic orbital basis sets, respectively.

Eq. (4) can be written in the following matrix form:

$$\mathbf{err} = 2((\mathbf{C}'_V(\mathbf{C}^\dagger_V(\mathbf{FC}_O)))\mathbf{C}'^\dagger_O) - \text{h.c.} \quad (5)$$

Again, according to Eq. (5), it is necessary to perform four matrix multiplications to calculate the error vector. However, if the multiplications are done in the order specified by the parentheses, the first and the last matrix multiplies require only $2NNO$ f.p.o. each, and the other two multiplies have $2NVO$ f.p.o. each. Therefore, the total number of f.p.o. is $4NNO + 4NVO \approx 8NNO$. Compared to the conventional $8N^3$ f.p.o., the expected speed-up is N/O.

It should be noted that the evaluation of the molecular orbital coefficients in the orthogonalized atomic orbital basis set C', which enters Eq. (5), does not require an additional matrix multiplication. C' matrix is an intermediate result in the conventional diagonalization scheme. It can be saved and re-used in the error vector construction.

In the case of the spin-unrestricted wavefunction, the error vector can be constructed from the separate error vectors for α and β spins using the same trick.

The proposed algorithm has been implemented in a development version of the Q-Chem software package [10]. Hartree–Fock calculations on two-dimensional water clusters are used to test the performance of the scheme. A series of basis sets, 6-311G, 6-311G(d,p), 6-311(++)G(d,p), cc-pVDZ, were used to illustrate how the speed-ups change with basis set size. Linear scaling algorithms are employed for the formation of the Fock matrix [3,5,7].

As one can see from Table 1 the new method for the DIIS error vector calculation is effective even for small systems. As the size of the system increases the speed-ups are approaching the value predicted from the f.p.o. analysis (last row in bold). As expected, larger speed-ups are obtained for larger basis sets.

Thus, we achieve faster computation of the DIIS error vector by expanding the one-electron density matrix in terms of the occupied molecular orbitals and by using the associativity property to perform matrix multiplications in an optimal order for efficiency. While we have employed dense linear algebra routines to evaluate the error vector, the formulation we have introduced here (as well as the original formulation) can in principle also be used with sparse matrix routines to lower the computational scaling with system size from cubic towards linear, similar to linear scaling diagonalization replacements [11–13].

References

[1] P. Pulay, Chem. Phys. Lett. 73 (1980) 393.
[2] P. Pulay, J. Comput. Chem. 3 (1982) 556.
[3] C.A. White, B.G. Johnson, P.M.W. Gill, M. Head-Gordon, Chem. Phys. Lett. 253 (1996) 268.
[4] M.C. Strain, G.E. Scuseria, M.J. Frisch, Science 271 (1996) 51.
[5] Y. Shao, M. Head-Gordon, Chem. Phys. Lett. 323 (2000) 425.
[6] L. Fusti-Molnar, J. Kong, J. Chem. Phys. 122 (2005) 74108.
[7] C. Ochsenfeld, C.A. White, M. Head-Gordon, J. Chem. Phys. 109 (1998) 1663.
[8] J.M. Perez-Jorda, W. Yang, Chem. Phys. Lett. 241 (1995) 469.
[9] R.E. Stratmann, G.E. Scuseria, M.J. Frisch, Chem. Phys. Lett. 257 (1996) 213.
[10] J. Kong, C.A. White, A.I. Krylov, et al., J. Comput. Chem. 21 (2000) 1532.
[11] S. Goedecker, Rev. Mod. Phys. 71 (1999) 1085.
[12] G.E. Scuseria, J. Phys. Chem. A 103 (1999) 4782.
[13] D.R. Bowler, T. Miyazaki, M.J. Gillan, J. Phys.: Condens. Matter 14 (2002) 2781.