Some Comments on Accelerating Convergence of Iterative Sequences Using Direct Inversion of the Iterative Subspace (DIIS)

C. David Sherrill
School of Chemistry and Biochemistry
Georgia Institute of Technology

May 1998

Introduction

Most standard textbook approaches to solving systems of linear equations or diagonalizing matrices are described as "direct" methods, and they typically require a fixed number of mathematical operations which depends on the dimensions of the problem. These methods generally require access to matrix elements in random order, which poses serious difficulties for the very large matrices typically encountered in computational quantum chemistry: random access of large disk files becomes prohibitively expensive, and often the matrices are too large even to store on disk! In such cases, one may avoid the need for random access to individual matrix elements by turning to iterative techniques, which require only the repeated evaluation of matrix-vector products. Unfortunately, iterative methods are not guaranteed to converge, and they can have difficulties when the matrix is not diagonally dominant or when there are nearly degenerate solutions.

The well-known Davidson method [1] for the iterative solution of the lowest few eigenvalues and eigenvectors of large, symmetric matrices combines some of the features of direct and iterative techniques. Although only matrix-vector operations are required, and there is no need to explicitly store the Hamiltonian matrix, Davidson's method also uses direct methods to diagonalize a small Hamiltonian matrix formed in the subspace of all trial CI vectors that have been considered up to the present iteration. The current estimates of the eigenvalues of the full Hamiltonian matrix are obtained as the eigenvalues the small Hamiltonian matrix, and the current CI vectors are obtained as the linear combinations of the trial vectors whose coefficients are given by the eigenvectors of the small Hamiltonian matrix. Pople and co-workers later used related ideas to iteratively solve the large systems of linear equations occuring in the coupled-perturbed Hartree-Fock method [2].

In 1980, Pulay published a somewhat similar method [3] known as the direct inversion of the iterative subspace (DIIS). Like the Davidson method, DIIS applies direct methods to a small linear algebra problem (now a system of linear equations instead of an eigenvalue problem) in a subspace

formed by taking a set of trial vectors from the full-dimensional space. Pulay found that DIIS could be useful for accelerating the convergence of self-consistent-field (SCF) procedures and, to a lesser extent, geometry optimizations.

The Mathematics of DIIS

Suppose that we have a set of trial vectors $\{\mathbf{p}^i\}$ which have been generated during the iterative solution of a problem. Now let us form a set of "residual" vectors defined as

$$\Delta\mathbf{p}^i = \mathbf{p}^{i+1} - \mathbf{p}^i. \tag{1}$$

The DIIS method assumes that a good approximation to the final solution $\mathbf{p}^f$ can be obtained as a linear combination of the previous guess vectors

$$\mathbf{p} = \sum_i^m c_i \mathbf{p}^i, \tag{2}$$

where m is the number of previous vectors (in practice, only the most recent few vectors are used). The coefficients c_i are obtained by requiring that the associated residual vector

$$\Delta\mathbf{p} = \sum_i^m c_i \left(\Delta\mathbf{p}^i\right) \tag{3}$$

approximates the zero vector in a least-squares sense. Furthermore, the coefficients are required to add to one,

$$\sum_i^m c_i = 1. \tag{4}$$

The motivation for the latter requirement can be seen as follows. Each of our trial solutions $\mathbf{p}^i$ can be written as the exact solution plus an error term, $\mathbf{p}^f + \mathbf{e}^i$. Then, the DIIS approximate solution is given by

$$\begin{aligned} \mathbf{p} &= \sum_i^m c_i \left(\mathbf{p}^f + \mathbf{e}^i\right) \\ &= \mathbf{p}^f \sum_i^m c_i + \sum_i^m c_i \mathbf{e}^i. \end{aligned} \tag{5}$$

Hence, we wish to minimize the *actual error*, which is the second term in the equation above (of course, in practice, we don't know $\mathbf{e}^i$, only $\Delta\mathbf{p}^i$); doing so would make the second term vanish, leaving only the first term. For $\mathbf{p} = \mathbf{p}^f$, we must have $\sum_i^m c_i = 1$.

Thus, we wish to minimize the norm of the residuum vector

$$\langle \Delta\mathbf{p}|\Delta\mathbf{p}\rangle = \sum_{ij}^{m} c_i^* c_j \langle \Delta\mathbf{p}^i|\Delta\mathbf{p}^j\rangle, \tag{6}$$

subject to the constraint (4). These requirements can be satisfied by minimizing the following function with Lagrangian multiplier λ

$$\mathcal{L} = \mathbf{c}^\dagger \mathbf{B}\mathbf{c} - \lambda\left(1 - \sum_{i}^{m} c_i\right), \tag{7}$$

where $\mathbf{B}$ is the matrix of overlaps

$$B_{ij} = \langle \Delta\mathbf{p}^i|\Delta\mathbf{p}^j\rangle. \tag{8}$$

We can minimize $\mathcal{L}$ with respect to a coefficient c_k to obtain (assuming real quantities)

$$\begin{aligned} \frac{\partial\mathcal{L}}{\partial c_k} = 0 &= \sum_{j}^{m} c_j B_{kj} + \sum_{i}^{m} c_i B_{ik} - \lambda \\ &= 2\sum_{i}^{m} c_i B_{ki} - \lambda. \end{aligned} \tag{9}$$

We can absorb the factor of 2 into λ to obtain the following matrix equation, which is eq. (6) of Pulay [3]:

$$\begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1m} & -1 \\ B_{21} & B_{22} & \cdots & B_{2m} & -1 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ B_{m1} & B_{m2} & \cdots & B_{mm} & -1 \\ -1 & -1 & \cdots & -1 & 0 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \cdots \\ c_m \\ \lambda \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdots \\ 0 \\ -1 \end{pmatrix} \tag{10}$$

Programming DIIS

The DIIS procedure seems so simple that further comment on specific computational implementations might appear superfluous. However, I have found that the precise computational details are absolutely crucial for effective interpolation. Hence, I will describe here my implementation of DIIS for the optimization of orbitals in a two-step CASSCF program. There are probably many variations on this implementation which would also work, but often seemingly inconsequential chnages can make dramatic differences in efficiency.

In the two-step CASSCF procedure, one begins with a set of guess orbitals, solves the full CI problem in the active space, determines the gradient for orbital rotations, takes a step in orbital

rotation (theta) space down the gradient direction (i.e., obtains new guess orbitals), and repeats the process iteratively until convergence. To allow DIIS interpolation, one can express the current set of guess orbitals as the result of the multiplication of a set of Givens rotation matrices by a matrix of "reference" orbitals ($C_{\mu p} = \sum_q C^o_{\mu q} U_{qp}$, see [4]). The rotation angles which define the unitary transformation U (a product of Givens rotation matrices) comprise a vector of parameters, $\mathbf{p}$.

In this case, one can define the error vectors as the differences between subsequent sets of orbital rotation angles, or one could also reasonably choose the orbital gradient vector. In my DETCAS program, the regular theta step is determined using a Newton-Raphson approach with an approximate, diagonal orbital Hessian. This is equivalent to scaling the orbital gradient to a new coordinate system. Since the step in theta space is just the scaled gradient, the scaled gradient is the same as the difference between successive theta vectors (apart from a sign) before the DIIS procedure starts. However, I have found it much better to associate the gradient vector with the *next* iteration's theta vector, *not* with the theta vector from which it was computed. In other words, it is best to change eq. (1) to the following:

$$\Delta \mathbf{p}^{i+1} = \mathbf{p}^{i+1} - \mathbf{p}^{i}. \tag{11}$$

Another general consideration is that one does not want to add an interpolated vector to the list of vectors $\{\mathbf{p}\}$ unless it contains some new character to add to the subspace. Otherwise, linear dependencies can result.

An outline of my DIIS procedure for the DETCAS program is given below:

1. Using current orbitals $\mathbf{p}^i$, obtain scaled orbital gradient $\mathbf{g}^i$.
2. Take Newton-Raphson step $\mathbf{p}^{i+1} = \mathbf{p}^i - \mathbf{g}^i$.
3. Add $\mathbf{p}^{i+1}$ to list of vectors. Add $\Delta\mathbf{p}^{i+1} = -\mathbf{g}^i$ to list of error vectors.
4. Perform DIIS interpolation to obtain new guess vector. Overwrite $\mathbf{p}^{i+1}$ with DIIS interpolant. (This vector will never be added to list of vectors).
5. Increment i, begin new cycle.

References

[1] E. R. Davidson, J. Comp. Phys. **17**, 87 (1975).

[2] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. Symp. **13**, 225 (1979).

[3] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

[4] M. Head-Gordon and J. A. Pople, J. Phys. Chem. **92**, 3063 (1988).