Advanced Software for the Calculation of Thermochemistry, Kinetics, and Dynamics

R. Shepard, A. F. Wagner, and S. K. Gray
Chemistry Division, Argonne National Laboratory, Argonne, IL 60439
shepard@tcg.anl.gov

Abstract. The Born-Oppenheimer separation of the Schrodinger equation allows the electronic and nuclear motions to be solved in three steps. 1) The solution of the electronic wave function at a discrete set of molecular conformations; 2) the fitting of this discrete set of energy values in order to construct an analytical approximation to the potential energy surface (PES) at all molecular conformations; 3) the use of this analytical PES to solve for the nuclear motion using either time-dependent or time-independent formulations to compute molecular energy values, chemical reaction rates, and cumulative reaction probabilities. This project involves the development of technology to address all three of these steps. This report focuses on our recent work on the optimization of nonlinear wave function parameters for the electronic wave functions.

1. Introduction

Our work this past year has focused on the optimization of nonlinear wave function parameters for electronic wave functions and on the fitting of computed potential energy surface (PES) values using the interpolating moving least-squares (IMLS) approach. This report summarizes our work on the electronic wave function optimization problem. Our IMLS work is reported in Ref. [1] in this issue.

We have recently proposed a new expansion form for molecular electronic wave functions [2-4]. The wave function is written as a linear combination of product basis functions, and each product basis function in turn is formally equivalent to a linear combination of configuration state functions (CSFs) that comprise the underlying linear expansion space of dimension $N_{csf}$. The CSF coefficients that define the basis functions are nonlinear functions of a smaller number of variables $N_{csf}/N_{ef}$. This new approach addresses directly the main bottlenecks of previous electronic structure methods. Much of the computational effort and data storage requirements with this new approach depends only on $\log(N_{csf})$ rather than $N_{csf}$. The method is formulated in terms of spin-eigenfunctions using the Graphical Unitary Group Approach (GUGA) [5], and consequently it does not suffer from the spin contamination or spin instability that is often associated with single-reference methods. The expansion form is appropriate for both ground and excited states and to closed- and open-shell molecules. The method is not based on the idea of expansion about a “reference” wave function, so its accuracy is not inherently limited due to failures of the Hartree-Fock method or to artificially imposed excitation-level restrictions. Our initial applications to some small molecules are encouraging. As seen in Figs. 1 and 2, even a single product function $N_{ef}=1$ based on a $6^6$ Shavitt graph is sufficient to dissociate the N≡N triple bond correctly to ground state 4S atom fragments, two product functions are sufficient to approach the full-valence CI electronic energy to within 3 mh, and three product functions...
reproduce exactly the full-valence CI energy at all bond lengths. For a larger, more challenging molecule, we find that ten expansion terms based on a 12\textsuperscript{18} Shavitt graph are sufficient to reproduce the full-valence CI energy of the ground state \(^1\text{A}_1\) O\(_3\) molecule near its equilibrium geometry to within 1 \(\text{mhartree}\) (see Fig. 3).

![Energy curves for the dissociation of ground state \(^1\Sigma^+_g\) N\(_2\) using various wave functions. The \(N_{\alpha}=1, 2,\) and 3 curves correspond to a 6\textsuperscript{6} Shavitt graph with the orbitals ordered the same way as for the PP-GVB graph.](image)

**Fig. 1.** Energy curves for the dissociation of ground state \(^1\Sigma^+_g\) N\(_2\) using various wave functions. The \(N_{\alpha}=1, 2,\) and 3 curves correspond to a 6\textsuperscript{6} Shavitt graph with the orbitals ordered the same way as for the PP-GVB graph.

![Energy curves for the dissociation of ground state \(^1\Sigma^+_g\) N\(_2\) using various wave functions near the R\(_e\) region.](image)

**Fig. 2.** Energy curves for the dissociation of ground state \(^1\Sigma^+_g\) N\(_2\) using various wave functions near the R\(_e\) region.

![Convergence of the energy with respect to \(N_{\alpha}\). The N\(_2\) calculations are with the 6\textsuperscript{6} Shavitt graph, with two different orderings of the orbitals. The O\(_3\) calculations are with the 12\textsuperscript{18} Shavitt graph.](image)

**Fig. 3.** Convergence of the energy with respect to \(N_{\alpha}\). The N\(_2\) calculations are with the 6\textsuperscript{6} Shavitt graph, with two different orderings of the orbitals. The O\(_3\) calculations are with the 12\textsuperscript{18} Shavitt graph.

### 2. Wave Function Optimization

Although the results in the previous section are very encouraging, the most remarkable feature of our new method is the relatively small effort required to construct hamiltonian matrix elements and transition density matrices in the product function basis\([3,4]\). Given a wave function expansion of the form

\[
|\psi\rangle = \sum_{\mathcal{M}} c_{\mathcal{M}} |\mathcal{M}\rangle
\]

in which the \(N_{\alpha}\) product functions are denoted \(|\mathcal{M}\rangle\), each basis function \(|\mathcal{M}\rangle\) depends on a corresponding set of nonlinear parameters \(\mathbf{\varphi}_{\mathcal{M}}\). The wave function thereby depends on the linear expansion coefficients \(c\) and on the nonlinear parameters \(\varphi\). Optimization of the expansion coefficients \(c\) to minimize the energy results in the symmetric generalized eigenvalue equation

\[
H \mathbf{c} = S \mathbf{c} E
\]

with \(H_{\mathcal{MN}} = \langle \mathcal{M} | \hat{H} | \mathcal{N}\rangle\) and \(S_{\mathcal{MN}} = \langle \mathcal{M} | \mathcal{N}\rangle\). An efficient procedure to compute hamiltonian matrix elements and reduced one- and two-particle density matrices for this nonlinear expansion form has been developed\([3]\). The effort required to construct an individual hamiltonian matrix element between two product basis functions \(H_{\mathcal{MN}}\) scales as \(O(\beta n^4)\) for a wave function expanded in \(n\) molecular orbitals. The prefactor \(\beta\) itself scales between \(N^6\) and \(N^7\), for \(N\) electrons, depending on the complexity of the underlying Shavitt graph. The corresponding metric matrix element \(S_{\mathcal{MN}} = \langle \mathcal{M} | \mathcal{N}\rangle\) requires effort that scales as \(O(\beta n)\). There is no component of the effort or storage for matrix element computation that scales as \(N_{\text{cf}}\). Hamiltonian matrix element timings with our initial implementation of this method are very promising. Wave function expansions that are orders of magnitude larger than can be treated with traditional CI methods require only modest effort with our new method. Table 1 shows some timing statistics for a series of Shavitt graphs for singlet wave functions with \(n=N\). A matrix element involving product basis functions with \(n=N=46\) corresponding to an underlying linear expansion space...
dimension $N_{\text{csf}}=5.5 \times 10^{24}$, or over 9.2 mol of CSFs, requires only a few seconds with our new method. The computation of that same Hamiltonian matrix element using traditional full-CI technology is estimated to require about $10^{24}$ seconds, which is about a million times longer than the age of the universe. Of course, a traditional full-CI code has the capability of computing $H_{MN}=\mathbf{x}^{M}H\mathbf{x}^{N}$ for arbitrary vectors $\mathbf{x}^{M}$ and $\mathbf{x}^{N}$, whereas our new method is restricted to vectors that can be represented in our graphical-based nonlinear expansion form. This restriction affects eventually $N_{\varphi}$, the number of basis functions that are required to achieve convergence to chemical accuracy. We note at this point that our new method does not depend on approximating $H_{MN}$ in any way (e.g., local density or fast multipole approximations); our recursive procedure computes exactly the same $H_{MN}$ value as the traditional full-CI method. This timing comparison shows the tremendous potential of our new method. $H_{MN}$ computation scales approximately the same as the simple Hartree-Fock method, yet the method is capable of approaching the accuracy of full-CI wave functions.

Table I also gives various Shavitt graph statistics for this sequence of product wave functions. $N_{\text{row}}$ is the number of nodes of the graph, and this value determines the storage requirements and computational effort required for the recursive construction of $S_{MN}$ and $H_{MN}$. $N_{\varphi}$ is the number of essential nonlinear parameters in each product basis function. It is clear from this table that $N_{\varphi} \ll N_{\text{csf}}$ for increasing $n$. The optimization of these nonlinear parameters is the current outstanding challenge in the implementation and application of this new method.

| $n=N$ | $N_{\text{csf}}$ | $N_{\text{row}}$ | $N_{\varphi}$ | $t(H_{MN})$ | $t(E)\phi$ | $t(E;\text{FD})$ |
|-------|-----------------|-----------------|-------------|-------------|-------------|-------------|
| 2     | 3               | 5               | 2           | 0.00        | 0.00        | 0.00        |
| 4     | 20              | 14              | 13          | 0.00        | 0.00        | 0.00        |
| 6     | 175             | 30              | 39          | 0.00        | 0.00        | 0.02        |
| 8     | 1,764           | 55              | 86          | 0.00        | 0.01        | 0.10        |
| 10    | 19,404          | 91              | 160         | 0.00        | 0.05        | 0.64        |
| 12    | 226,512         | 140             | 267         | 0.00        | 0.16        | 3.20        |
| 14    | 2,760,615       | 204             | 413         | 0.01        | 0.44        | 12.39       |
| 16    | 34,763,300      | 285             | 604         | 0.04        | 1.24        | 36.24       |
| 18    | 449,141,836     | 385             | 846         | 0.07        | 3.48        | 118.44      |
| 20    | 5,924,217,936   | 506             | 1145        | 0.13        | 9.29        | 297.70      |
| 22    | 79,483,257,308  | 650             | 1507        | 0.21        | 25.67       | 632.94      |
| 24    | 1,081,724,803   | 819             | 1938        | 0.34        | 65.49       | 1,328.3     |
| 26    | 14,901,311,070  | 1015            | 2444        | 0.54        | 140.61      | 2,648.5     |
| 28    | 207,426,250,094| 1240            | 3031        | 0.82        | 250.45      | 4,978.3     |
| 30    | 2,913,690,606   | 1496            | 3705        | 1.21        | 423.87      | 8,978.3     |
| 32    | 41,255,439,318  | 1785            | 4472        | 1.75        | 676.76      | 1,578.4     |
| 34    | 588,272,005,095| 2109            | 5338        | 2.49        | 1,073.4     | 2,668.4     |
| 36    | 8,441,132,926   | 2470            | 6309        | 3.46        | 1,628.3     | 4,378.4     |
| 38    | 121,805,548,126 | 2870            | 7391        | 4.66        | 2,388.3     | 6,898.4     |
| 40    | 1,766,594,752   | 3311            | 8590        | 6.27        | 3,488.3     | 10,888.5    |
| 42    | 25,739,723,541  | 3795            | 9912        | 8.25        | 4,938.3     | 1,648.5     |
| 44    | 376,607,675,256 | 4324            | 11363       | 11.19       | 6,888.3     | 2,548.5     |
| 46    | 46,531,425,230  | 4900            | 12949       | 14.43       | 9,478.3     | 3,748.5     |

$^{a}$ Times are in seconds on a 2.5GHz PowerMac G5 (PPC 970) to construct a single $H_{MN}$ matrix element. $^{b}$ Times in seconds to construct the analytic gradient vector $E'(\varphi) = \partial E(\varphi)/\partial \varphi_{\text{ref}}|_{\varphi}$ for $N_{\varphi}=1$ using the $G^{(i)}$ and $S^{(i)}$ arrays. $^{c}$ Times in seconds to construct the gradient with a finite-difference approximation, $t(E;\text{FD})=2N_{\varphi}t(H_{MN})$. 
A major challenge presented by the new method is the optimization of the nonlinear parameters. A straightforward application of standard optimization methods, such as conjugate-gradient and quasi-newton approaches, results in relatively slow convergence requiring hundreds or thousands of gradient evaluations. Each gradient evaluation requires effort $O(n^3)$, which is significantly more expensive than an energy evaluation, so it is critical to develop other optimization methods that require less effort. Table 1 shows some timings for our analytic gradient computation along with estimates of the finite difference gradient approximation. In general, we find that our analytic gradient computation is 10 to 100 times faster than a finite difference approximation, and it is 10 to 100 times slower than a single energy computation. Fig. 4 shows the convergence of some of these methods.

The LBFGS curves in Fig. 4 use the limited-memory quasi-newton method of Liu and Nocedal [6]. Each step of this procedure requires a gradient and an energy evaluation. Various numbers of history vectors are stored, and it is clear from the curves that increasing the number of vectors improves the convergence in this case. However, none of the convergence curves are entirely satisfactory.

The conjugate gradient curve shown uses the CG_DESCENT code of Hager and Zhang [7]. This method requires on average two to three function evaluations for each gradient evaluation. Like the quasi-newton curves, the convergence is relatively slow, with stalls of several hundred iterations.

The substitution curve is based on a special purpose optimization method for this specific problem. If all of the nonlinear arc factors in the Shavitt graph are frozen except for those at a single orbital level $u$, then the optimal values of the arc factors $\alpha^{[u]}$ at that level are given by the elements of the lowest eigenpair of the generalized symmetric eigenproblem

$$G^{[u]} \alpha^{[u]} = S^{[u]} \alpha^{[u]} E$$

Furthermore, if these optimal arc factors are substituted for the original arc factors at level $u$, then the computed energy is exactly the eigenvalue of this eigenproblem. That is, the Rayleigh quotient

$$E = \frac{\alpha^{[u]^T} G^{[u]} \alpha^{[u]}}{\alpha^{[u]^T} S^{[u]} \alpha^{[u]}}$$

is an exact representation of the energy as a function of this subset of nonlinear parameters. The computation of the orbital level matrices $G^{[u]}$ and $S^{[u]}$ is comparable in effort to a gradient evaluation. The substitution approach leads to improved convergence compared to the quasi-newton and CG
methods, but several hundred iterations are required in order to attain convergence to within a $\mu h$, so improved convergence is still necessary before wave functions such as those at the bottom half of Table 1 can be considered.

Our final convergence method in Fig. 4 uses the optimal $\alpha^{(u)}$ values from the orbital eigenproblem to define sets of arc factors, one set for each orbital level $u$. These $n$ sets of arc factors are then used to define a derivative-free subspace search. We use the PRAXIS code of Brent [8] to optimize within this subspace. The convergence of this method with respect to the number of gradient evaluations is much improved over the other methods, but our current implementation of this approach can require up to several hundred function evaluations for each step. Consequently, we feel that this approach shows promise, but we have not yet solved the efficiency and convergence problems of the arc factor optimization in order to allow us to handle large wave function expansions such as those in the bottom half of Table 1. We hope to improve on this approach by including selected analytic line derivatives (which are relatively cheap to compute [4]) in order to accelerate the line-search optimizations.

In the future, we plan to explore other optimization approaches, including those based on stepwise optimization of nested subgraphs and on various interpolation approaches [4,9,10]. We will also explore the possibility of improved analytic gradient computation methods [11].

3. Conclusions

We have presented initial results of a general formulation for energy-based optimization of the arc factors for our recently developed nonlinear wave function expansion form for electronic wave functions. The energy-based optimization is formulated in terms of analytic energy gradients and orbital-level Hamiltonian matrices which correspond to a specific kind of uncontraction of each of the product basis functions. These orbital-level Hamiltonian matrices give an intuitive representation of the energy as a function of disjoint subsets of the arc factors, they provide for an efficient computation of gradients of the energy with respect to the arc factors, and they allow optimal arc factors to be determined in closed form for subspaces of the full variation problem. Our analytic gradient computations are between a factor of 10 and 100 times faster than a finite-difference approximation.

We observe slow convergence for the arc factor optimization, and this suggests that the optimization problem should be a main focus of future effort with this method. Some initial applications using our current optimization methods show very promising results for the difficult $N_2$ dissociation problem and for the larger $O_3$ molecule. Energy and gradient computations with our new method require relatively little effort compared to other electronic structure methods. Timings for energy and arc factor gradient computations involving expansion spaces of over $10^{24}$ CSFs have been reported.

References

[1] D. L. Thompson, “Chemical Applications of Multi-Dimensional Moving Least Squares Fitting Methods for Hypersurfaces”, this volume.
[2] R. Shepard, J. Phys. Chem. A 109, 11629 (2005)
[3] R. Shepard, J. Phys. Chem. A 110, xxx (2006)
[4] R. Shepard and M. Minkoff, Int. J. Quantum Chem. (in press, 2006).
[5] I. Shavitt, The Graphical Unitary Group Approach and its Application to Direct Configuration Interaction Calculations, in The Unitary Group for the Evaluation of Electronic Energy Matrix Elements (Lecture Notes in Chemistry 22), pp. 51-99, J. Hinze, ed., (Springer-Verlag, Berlin, 1981).
[6] D. Liu and J. Nocedal, Mathematical Programming B 45, 503 (1989).
[7] W. W. Hager and H. Zhang, http://www.math.ufl.edu/~hager/papers/CG.
[8] R. P. Brent, in Algorithms For Minimization without Derivatives, (Prentice-Hall, Englewood Cliffs, New Jersey, 1973).
[9] P. Pulay, Chem. Phys. Letts. 73, 393 (1980); P. Pulay, J. Comput. Chem. 3, 556 (1982).
[10] J. More´ and M. Anitescu, MCS Division, ANL, private communication (2006).
[11] P. Hovland, J. Utke, and B. Norris, MCS Division, ANL, private communication (2006).