ELSEVIER

Optimization of MCSCF excited states using directions of negative curvature

Mark R. Hoffmann [a,*], C. David Sherrill [b], Matthew L. Leininger [c,1],
Henry F. Schaefer III [c]

[a] *Chemistry Department, University of North Dakota, P.O. Box 9024, Grand Forks, ND 58202-9024, USA*
[b] *School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA*
[c] *Center for Computational Quantum Chemistry, University of Georgia, Athens, GA 30602, USA*

Abstract

A line search method that uses directions of negative curvature for the optimization of ground and excited state multiconfigurational self-consistent field (MCSCF) wavefunctions is suggested. The method is applicable to general MCSCF wavefunctions and not restricted to specific classes of model spaces, such as CASSCF functions. It is shown that the approach can be implemented with Newton or quasi-Newton methods for determination of descent directions. Thus, the method is particularly promising for cases in which it is inconvenient or costly to calculate exact curvature matrices. We demonstrate the viability of the approach by numerical example on the difficult BeO problem. © 2002 Published by Elsevier Science B.V.

1. Introduction

The Multiconfigurational Self-Consistent Field (MCSCF) approximation to nondynamical correlation is well established in molecular electronic structure theory (see, for example, Reviews [1–5]). Nonetheless, the issue of convergence cannot be considered wholly solved. There are at least two classes of convergence problems that continue to be investigated actively: first, those characterized as topologically simple but of very large dimension parameter space, and second, those characterized as of more moderate parameter space but topologically challenging. The ground electronic states of many moderate to large molecules fit into the first category; we are not interested in these in this Letter. We consider problems of the type generated from moderate-sized one-electron active spaces and substantial multiconfigurational character. Convergence is particularly challenging for excited electronic states, especially in the proximity of other states.

On discussion of optimization of MCSCF excited states, the issue of 'best' representation of an excited state is necessarily relevant. The question of the 'best' representation of the Nth exact state by an MCSCF wavefunction is a subtle question. One possibility is the Nth state of a CI calculation using the converged MCSCF orbitals. As was

---

[*] Corresponding author. Fax: +1-701-777-2331.
E-mail address: mhoffmann@chem.und.edu (M.R. Hoffmann).
[1] Present address: Sandia National Laboratories, Livermore, CA 94551, USA.

demonstrated in [6,7], a set of parameters based on other criteria may be more advantageous in some cases. The method we discuss can be applied in principle to a variety of definitions of 'best'. It is most straightforwardly applied to the situation where the Nth state of the MCSCF CI, guarded against uncontrolled deterioration of low-lying surfaces, is the 'best' representation and it is in this context that the method is described.

The topological problem is that of a nonlinear equality-constrained problem (NEP) (see e.g., [8]): we seek the lowest value of an energy-like function in the parameter space of orbital rotations *subject to* the energy-like function adequately representing the Nth excited state. In this Letter, we explore the ability of a step-length algorithm, which actively uses directions of negative curvature, in the context of the MCSCF minimization problem on a manifold. To the best of our knowledge, such investigations have not been previously described in the literature. However, it is interesting to note that the issue of directionality along modes of curvature contrary to the desired motion has been addressed in molecular geometry optimization (see, e.g., [9–11]). Although, the specific approach we discuss herein has not been considered in such context. The remainder of this Letter is divided into six additional sections. The issue of search direction is described in the following section. Section 3 discusses step length. Section 4 contains a description of the computational details of our algorithm. Section 5 introduces a model problem that illustrates a possible nondifferentiability and a modification of the optimality function. Section 6 provides illustrative calculations on the three lowest $^1\Sigma^+$ states of BeO, which have been considered by other methods. These states are known to be difficult problems for MCSCF description and provide a severe test. A final section summarizes this work.

2. Search direction

Let us choose as a basis for describing variations of the molecular orbitals used to construct the wavefunction the set of nonredundant orbital rotations [12–15]. More specifically, if we have a set of molecular orbitals, $\{|\phi^{(0)}\rangle\}$, which are re-

lated to a set of atomic orbitals by a coefficient matrix, $\mathbf{C}^{(0)}$, then all sets of molecular orbitals can be described as

$$\mathbf{C}^{(k)} = \mathbf{C}^{(0)}\mathrm{e}^{\Xi^{(k)}}, \tag{1a}$$

where $\Xi^{(k)}$ is an anti-hermitian real matrix, $[\Xi^{(k)}]^{\mathrm{T}} = -\Xi^{(k)}$. Alternatively, we can represent the orbitals in terms of the orbitals of the preceding iterations,

$$\mathbf{C}^{(k)} = \mathbf{C}^{(k-1)}\mathrm{e}^{\mathbf{X}^{(k)}}, \tag{1b}$$

where the $\Xi^{(k)}$ of Eq. (1a) and the $\mathbf{X}^{(k)}$ of Eq. (1b) are related. We choose the latter representation. We are cognizant of the issue of changing reference frames and take it into account in our updating procedure (vide infra). A direction expressed in this basis is a vector, $\mathbf{x}$, whose coefficients are the components of the lower triangle of $\mathbf{X}$ arranged as a vector. Throughout this Letter, the iteration number is suppressed when there is no ambiguity.

Imposing our constraint to the manifold defined by the Nth MCSCF CI root, the energy function can be written as,

$$E(\mathbf{x}) = E_n(\mathbf{x}) = \frac{\langle \psi_n(\mathbf{x})|H(\mathbf{x})|\psi_n(\mathbf{x})\rangle}{\langle \psi_n(\mathbf{x})|\psi_n(\mathbf{x})\rangle}. \tag{2}$$

A more general energy-like function can be constructed by use of a weighting vector,

$$E(\mathbf{x}) = \sum_i w_i E_i(\mathbf{x}), \tag{3}$$

where each of the energies are defined as in Eq. (2). Eq. (3) is to be recognized as the energy function of a state-averaged MCSCF approach [15–18]. It is at this stage that our suggested methodology differs from existing approaches. A Newton approach expands Eq. (2) or (3) to second order, truncates, differentiates and obtains a set of linear equations. The pitfalls of simplistic application of this procedure are well known and well documented. Jørgensen et al. [19] introduced successful trust-region implementations of the Newton approach. In such approaches, the problem of unconstrained minimization is replaced by a constrained subproblem [8]:

$$\text{minimize } \mathbf{g}_k^{\mathrm{T}}\mathbf{p}_k + \frac{1}{2}\mathbf{p}_k^{\mathrm{T}}\mathbf{G}_k\mathbf{p}_k, \tag{4a}$$

subject to $\|\mathbf{p}_k\|_2 \leqslant \Delta$, (4b)

where $\mathbf{g}_k \equiv \mathbf{g}(\mathbf{x}_k)$ is the gradient after the kth step, $\mathbf{G}_k \equiv \mathbf{G}(\mathbf{x}_k)$ is the hessian, $\mathbf{p}_k$ is the step, i.e., $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k$, and $\| \bullet \|_2$ is the Euclidean norm. If necessary, trust-region methods (see, e.g., [8]) solve first for a level shift, μ_k,

$$\phi(\mu_k) \equiv \|(\mathbf{G}_k + \mu_k I)^{-1}\mathbf{g}_k\|_2 = \Delta, \qquad (5)$$

whence the set of linear equations

$$(\mathbf{G}_k + \mu_k I)\mathbf{p}_k = -\mathbf{g}_k \qquad (6)$$

is solved.

Alternatives to trust-region methods, which are widely used in the optimization literature, are step-length-based methods (see, e.g., [8,20]). To define the problem, let $\mathbf{s}_k$ designate a nonascent direction, i.e.,

$$\mathbf{s}_k^{\mathrm{T}}\mathbf{g}(\mathbf{x}_k) \leqslant 0 \qquad (1f)$$

and let $\mathbf{d}_k$ designate a direction of nonpositive curvature,

$$\mathbf{d}_k^{\mathrm{T}}\mathbf{G}(\mathbf{x}_k)\mathbf{d}_k \leqslant 0 \qquad (1g)$$

so that $(\mathbf{s}_k, \mathbf{d}_k)$ is a descent pair. The search is along a curve of the form [21]

$$\begin{aligned} C = \{\mathbf{x}_k(\alpha) : \mathbf{x}_k(\alpha) \\ = \mathbf{x}_k + \phi_1(\alpha)\mathbf{s}_k + \phi_2(\alpha)\mathbf{d}_k, \ \alpha \geqslant 0\}. \end{aligned} \qquad (7)$$

Though there is active debate on the specific functions $\phi_1(\alpha)$ and $\phi_2(\alpha)$ (cf. [21–23] and [8,20,24] vis-à-vis modified Cholesky decomposition), it appears that the general idea of utilizing descent pairs in regions of indefiniteness has been firmly established. In this work, we have adopted the suggestion of Goldfarb [23] and take $\phi_1(\alpha) = \alpha$ and $\phi_2(\alpha) = \alpha^2$. In particular, such algorithms address two issues of central importance to MCSCF optimization. Firstly, under very mild conditions, they *guarantee* convergence to a minimum, as opposed to other critical points (e.g., saddle points). The frequent neglect of proper characterization of MCSCF wavefunctions has been commented on repeatedly in the literature. Secondly, whereas many optimization techniques are useful in regions that are described well by (local) quadratic approximation to the surface, the issue of convergence far from a stationary point is

more problematic. It is then the primary task of the optimization method to expeditiously search towards the correct region.

Although in principle *any* descent pair can be used in a line search algorithm with negative curvature, some pairs will be better than others. On the other hand, determination of the best descent pair must be considered in light of computational cost. In Section 4, we describe details of the hessian or approximate hessian that we use; here we describe aspects independent of the specific hessian. The algorithm is straightforwardly described in terms of spectral decomposition, and so it is implemented in our pilot computer code, but it must be emphasized that an efficient alternative starting from a modified Cholesky factorization is possible [20].

Suppose that

$$\mathbf{GU} = \mathbf{U}\lambda, \qquad (8)$$

where λ is the diagonal matrix of eigenvalues of the hessian matrix $\mathbf{G}$. Then a step that minimizes the local quadratic approximation (N.B. (Eqs. (4a)–(6)) is

$$\tilde{s}_i = \frac{-1}{\lambda_i + \mu_i}\tilde{g}_i, \qquad (9)$$

where $\tilde{g}_i = \sum_j U_{ij}^{\mathrm{T}}g_j$ and $s_i = \sum_j U_{ij}\tilde{s}_j$. In principle, the level shifts, μ_i, could vary, e.g., Gill-Murray modified Cholesky decomposition [8], but there is good theoretical reason [20] not to do so and we adopt a constant shift. However, the step from Eq. (9) is not completely satisfactory in all cases. Let us suppose that we are cognizant that steps based on negative denominators in Eq. (9) correspond to directions of increasing function value, e.g., towards a saddle point, and that we would choose a sufficiently positive level shift to prevent such occurrence. Consequently, even though the directions relative to an undesirable stationary point are acceptable, the curvatures may be severely inaccurate. Consider that if the level shift is chosen such that $\lambda_1 + \mu \geqslant \varepsilon$, ε a small positive number, then μ may in fact be relatively large, so that 'soft modes', i.e., $\lambda_2, \lambda_3, \ldots \lambda_i$ small (either positive or negative), which should have large responses to gradients, will not. In a step-length method that uses directions of negative curvature, if there is at

least one negative eigenvalue then the eigenvector of lowest eigenvalue is taken to be $\mathbf{d}$, the negative curvature direction. The phase of $\mathbf{d}$ is chosen such that $(\mathbf{g}, \mathbf{d}) \leqslant 0$. Of course, if all the eigenvalues are positive, then $\mathbf{d} \equiv \mathbf{0}$. Having eliminated the largest negative curvature direction, the requirement on the level shift is that $\lambda_2 + \mu \geqslant \varepsilon$, which will introduce less perturbation than the other scheme (unless $\lambda_1 = \lambda_2$, in which case the perturbation is the same). With this specification, Eq. (9) is used for obtaining the descent direction; except that, in the event of a negative curvature direction, $s_1 = 0$, and in the event that there is a null space of nonzero dimension (i.e., there exist 'redundant variables'), $s_j = 0$, for $|\lambda_j| < \varepsilon_0$, with $\varepsilon_0 \sim 10^{-6}$.

3. Linesearch

Once a search direction is obtained, as discussed in the preceding section, a univariate minimization is to be performed,

$$\min f(\mathbf{x}_k(\alpha)), \tag{10}$$

where $\mathbf{x}_k(\alpha)$ is the curve specified in Eq. (7), with the descent pair $(\mathbf{s}_k, \mathbf{d}_k)$ discussed above. However, there is a tradeoff between the amount of effort expended to find α_k and the amount the function is lowered. It is now generally accepted that an inexact minimization is preferable [8,20]. The subject of criteria for line searches is discussed extensively in the optimization literature and, herein, we limit consideration to necessities. At the most basic level, one wishes to guarantee monotonic decreases, i.e.,

$$f(\mathbf{x}_{k+1}) < f(\mathbf{x}_k); \tag{11}$$

although such a straightforward criterion is not sufficient. Specifically, one must safeguard that the decreases are sufficiently large [21]:

$$\Phi(\alpha) \leqslant \Phi(0) + \varepsilon_1 \left[\alpha^i \Phi'(0) + \frac{1}{2} \alpha^{2i} \Phi''(0) \right]; \tag{12}$$

and, secondly, that the step sizes are not too small. In Eq. (12), $\Phi(\alpha)$ represents the univariate function $f(\mathbf{x}_k(\alpha))$; explicit expressions for required derivatives can be found in [21]. These conditions are of Armijo and Goldstein type. It can be shown [20]

that *any* method that generates a sequence $\{\mathbf{x}_k\}$ satisfying Eqs. (12) and (13), and $\mathbf{g}(\mathbf{x}_k)^{\mathrm{T}}(\mathbf{x}_{k+1} - \mathbf{x}_k) < 0$ at each iteration is essentially globally convergent *to a minimum*.

Efficiency considerations mandate that the number of steps in a line search be minimized, while still obeying the global optimization requirements. To that end, we take $\alpha = (\sqrt{5} - 1)/2$ and begin the search at the (modified) Newton (or quasi-Newton) length, i.e, $i = 0$. If Eq. (12) is not obeyed, we backtrack, i.e., take $i = 1, 2, \ldots$ until it is satisfied. Systematic backtracking, with an initial step that attempts to take the full (quasi-)Newton step, is currently accepted as the best general strategy for line searching [20].

4. Algorithm

The step-length algorithm for MCSCF that we implemented uses directly the level-shifted Newton (or quasi-Newton) direction vector and Goldstein–Armijo principle line search described in the preceding sections. In this section, we discuss specific details of our implementation.

The most critical issue that has not yet been addressed in this Letter is the specific hessian matrix used. In principle, any approximation to the hessian matrix at iteration k, $\mathbf{G}_k$, could be used in a minimization algorithm. We adopted the following scheme. Far from the minimum, i.e., in early iterations, we compute the orbital hessian for a specific set of orbitals, $\mathbf{G}_k \equiv \mathbf{G}^{\mathrm{oo}}(\mathbf{x}_k)$. Once the minimization is proceeding in a well-behaved manner, we begin Broyden–Fletcher–Goldfarb–Shanno (BFGS) update,

$$\mathbf{G}_{k+1} \equiv \mathbf{G}_k + \frac{\gamma_k^{\mathrm{T}} \gamma_k}{\alpha_k \mathbf{p}_k^{\mathrm{T}} \gamma_k} - \frac{(\mathbf{G}_k \mathbf{p}_k)^{\mathrm{T}}(\mathbf{G}_k \mathbf{p}_k)}{\mathbf{p}_k^{\mathrm{T}} \mathbf{G}_k \mathbf{p}_k}, \tag{13}$$

where $\gamma_k = \mathbf{g}_{k+1} - \mathbf{g}_k$, $\mathbf{p}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$, and other symbols have been defined earlier. Note that the kth iteration hessian, $\mathbf{G}_k$, is used in Eq. (13) (and not the level-shifted hessian) so that the product $\mathbf{G}_k \mathbf{p}_k$ is not necessarily expressible in terms of the gradient.

The choice of the hessian can be understood as follows. Far from convergence, the orbitals

are changing rapidly and the orbital hessian from the preceding iteration is not necessarily a good approximation to that of the current iteration. Moreover, attempting to correct such differences using updates may take a large number of iterations. However, near convergence, when the orbital hessian is no longer changing (much), it is *not* the orbital hessian that is required for second-order convergence, but rather the *effective*, or partitioned, orbital hessian,

$$\mathbf{G}^{\mathrm{eff}} \equiv \mathbf{G}^{\mathrm{oo}} + \mathbf{G}^{\mathrm{oc}}(\mathbf{G}^{\mathrm{cc}})^{-1}\mathbf{G}^{\mathrm{co}}, \tag{14}$$

where $\mathbf{G}^{\mathrm{oc}}$ and $\mathbf{G}^{\mathrm{cc}}$ are the orbital–state coupling and state blocks of the hessian. Our suggestion then is to approximate this effective orbital hessian starting from a nearly converged orbital hessian. The algorithm is given in Fig. 1. Indeed, it is this idea that potentially gives our method second-order convergence in the final steps. Formally, far from convergence our method is a Newton method, actually a modified Newton method, but close to convergence it is a (modified) quasi-Newton method.

5. Objective function for excited state surfaces: a model problem

The difficulties of finding the 'best' MCSCF function for an excited state can be illustrated by a simple one-dimensional model problem, which embodies many of the essential features of an MCSCF calculation. Let us suggest a two-state, one-parameter, problem, as shown in Fig. 2a. The wavefunctions are the diabatic states of this model problem and correspond to the MCSCF adiabatic states. The recent work of Hoffmann and Schatz [25] discusses a similar identification of diabatic states in the context of reactive scattering. Let us assume that a small constant coupling connects the functions; the corresponding adiabatic energies are graphed in Fig. 2b. Here, the coupling corresponds to the effect of dynamic electron correlation, which is neglected in the MCSCF optimization. The argument of Golab et al. [6,7] concerning 'best' representation can be stated in the context of this model problem: It is not at all clear that the lowest point on the upper curve, i.e., *A*, is a better representation of the true wavefunction (which in-

0. Initialization.
1. Calculate gradient: $\mathbf{g}(\mathbf{x}_k) = \nabla f(\mathbf{x}_k)$
2. If update enabled, check probity of update:
    2a. Transform gradient to old coordinate system
    2b. Obtain change of gradient change for last step: $\gamma_k = \tilde{g}_k - g_{k+1}$
    2c. Calculate overlap between gradient change and step vector: $(\mathbf{p}_k, \boldsymbol{\gamma}_k)$
    2d. If $(\mathbf{p}_k, \mathbf{g}_k) < 0$, cancel update
3. If update enabled, calculate update:
    3a. Calculate BFGS update
    3b. Rotate effective orbital hessian to new orbitals
4. If not update, then calculate hessian:
    4a. Calculate orbital hessian
    4b. If second-order, then calculate state contribution to partitioned orbital hessian
5. Calculate eigenpairs of hessian
6. Test for convergence
7. If a negative eigenvalue exists, then determine length and phase of negative curvature vector
8. Determine descent vector
    8a. Choose level shift, μ, so that lowest relevant eigenvalue obeys tolerance
    8b. Calculate descent vector for given level shift
    8c. If length of descent vector exceeds tolerance, then increase shift until tolerance is met
9. Line search
    9a. Set i=0
    9b. Generate $\mathbf{x}(\alpha^i)$ in accord with univariate curve [Eq. (7)]
    9c. Calculate energy-like function
    9d. If decrease is sufficiently large [Eq. (12)], then accept step and terminate line search
    9e. $\alpha^{i+1} = \alpha^i * \alpha$; loop to Step 9b
10. Set new reference point for molecular orbitals
11. Update control
    11a. If update is allowed, then activate update if start criteria are met
    11b. If update is active and hessian becomes indefinite, cancel update
12. Increment counter and loop to Step 1
13. Diagnostics and output

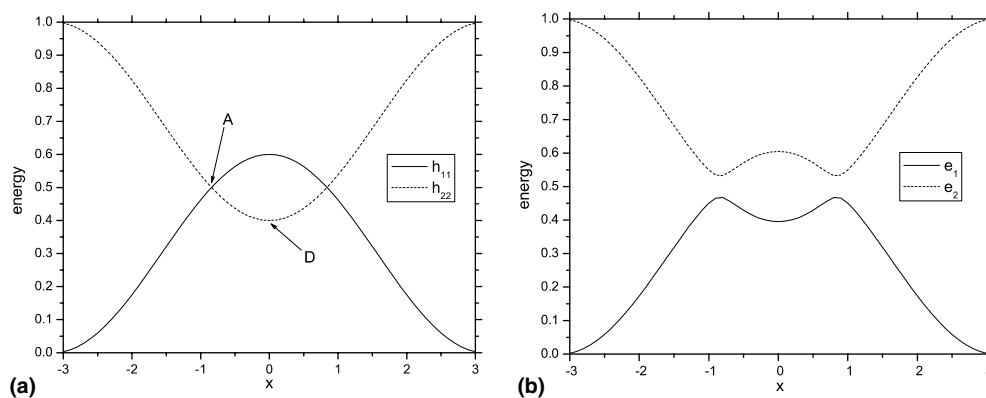Fig. 1. Suggested algorithm for MCSCF energy optimization.

Fig. 2. (a) Diabatic and (b) adiabatic energies of model problem as functions of parameter.

cludes more than just two states) than a point such as D. If the point A is deemed to be a 'better' representation, then our approach, as outlined in the preceding sections, can be used more-or-less straightforwards. On the other hand, if a D-like point is best, our algorithm can be applied with modification. Specifically, a manifold can be constructed e.g., by applying a level shift, static or dynamic, to the Nth root surface. This situation can arise in practice when one 'knows' the nature of the excited state from a larger MCSCF calculation, or from symmetry considerations. One may wish to obtain the smaller MCSCF antecedent to a post-MCSCF calculation.

On the other hand, if, as we choose, the criteria of lowest point on a manifold is of paramount importance, the issue of the nondifferentiability at the point of crossing of the diabatic curves must be addressed. Recall that Newton, or quasi-Newton, methods are only appropriate for surfaces with continuous derivatives. As is common practice in optimization, optimization of a nondifferentiable function is accomplished by optimization of a differentiable function that is known, or believed, to have common minima with the nondifferentiable function [8]. One such possibility is to add a small amount of the function value from the lower curve to the function of interest (and, thereby, approximate, albeit crudely, the coupling). In fact, this produces the well-established state-average MCSCF procedure, with unequal state weighting, to produce a manifold with the desired topological stationary point. Specific use

of this technique is discussed below in a concrete context.

6. Numerical example

The three lowest $^1\Sigma^+$ states of BeO were first identified as challenging problems for an MCSCF description by Bauschlicher and Yarkony (BY) [26], and have served as a severe test for optimization since, as in the extensive studies by Golab, Yeager and Jørgensen (GYJ) [6]. At issue is the fact that the different states are optimized with substantially different orbitals. Consequently, optimization of an excited state degrades the lower lying states to the point of variational collapse or 'root flipping'. Although this was recognized and a solution found (by expanding the active set of orbitals) [27], obtaining a reasonable MCSCF solution within the original active space remains a classic problem. We performed three sets of calculations on this system. The first set duplicates the earlier four configuration state function (CSF) studies of BY and GYJ, except that our restriction to Abelian point groups required us to span the $1\pi^3 2\pi^1$ configuration with an additional function. The four $^1\Sigma^+$ states and the one component of the $^1\Delta$ were noninteracting and easily identifiable. The second set consists of all spin- and space-adapted CSFs that can be generated from single- and double-excitations from the SCF reference; i.e., a so-called MCSCF(SD); this results in 19 functions. The third set is a CASSCF in the same orbital

space; this gives 55 CSF. The one-electron basis functions used in all calculations are identical to that of the earlier BY and GYJ studies, to all published digits, and can be found in the literature.

The first study investigates some general convergence properties, but is otherwise unchallenging: we investigate the ground state using CASSCF and using MCSCF(SD) starting from the Natural Orbitals of an MP2 calculation [28]. A step-length restriction of $1/\sqrt{2}$ was used in this study. Examination of Table 1 shows that while the first-order convergence proved adequate for the initial descent, the final convergence is essentially linear. Finally, consider the results from our update method, with a maximum rotation generator of 0.1 as the criterion for switching from first order to update. Starting from the third iteration, in this case, the BFGS update was applied. We see convergence in a total of 8 iterations. While convergence is not truly quadratic, it is unmistakably superlinear.

A slightly more difficult task, and more useful for assessment purposes, is provided by the optimization of the MCSCF(SD) wavefunction for the lowest $^1\Sigma^+$ state, again starting from the Natural Orbitals of an MP2 calculation. The problem remains topologically simple as there are no negative eigenvalues in the orbital (and, eventually, effective orbital) hessian at any iteration. The convergence

rate of the second-order procedure is essentially unchanged from the CASSCF. Although we are able to converge this wavefunction using a purely first-order procedure, it requires an astounding 115 iterations, which is a substantial increase from the 9 required for converging the CASSCF wavefunction. The behavior of the updated partitioned orbital hessian is intermediate: 14 iterations are now required rather than 8. Evidently, superlinear convergence is maintained for incomplete model spaces, and it appears that convergence is substantially more like second order than like first order.

The second study uses the updated effective hessian for the five CSF MCSCF wavefunction of BY and GYJ for the ground and first two excited states of BeO. We calculate first the ground state starting from canonical SCF orbitals(!). The initial orbital hessian has a concave subspace of dimension 8 and a null space $(|\lambda| \leqslant 1. \times 10^{-6})$ of dimension 5. A region of positive semi-definiteness is reached after the 4th iteration, and BFGS updating is started after the 5th iteration. The updating is unstable [by $(\gamma, \mathbf{x}) < 0.$] and is cancelled immediately. Curvature and step-length criteria are met again at the end of the 12th iteration and updating is begun again. Convergence is to -89.507164393764 a.u. in 29 iterations, and agrees exactly to the μh level reported by BY and to the value reported by GYJ to within 137 nh.

In the basis of the ground state MCSCF orbitals, the first excited state of $^1A_1$ symmetry is one component of the lowest $^1\Delta$ state. So, the next calculation we performed is a state-averaged (SA) MCSCF, with equal weighting, on the lowest three states of $^1A_1$ symmetry (cf. Table 2). The initial orbital hessian has a concave subspace of dimension 7, but no null space. A region of positive curvature is reached immediately after the first step, but the step lengths are too large to meet our criterion for updating until after the 3rd iteration. Convergence to 10^{-12} required 9 iterations. Initial roots 2 and 3 flipped, which is not of consequence since both roots were weighted equally in the optimization. We optimize for $2\,^1\Sigma^+$ by starting from the 3-state SA-MCSCF and performing a calculation with unnormalized weights of 0.01 and 1 for the lowest and second roots, respectively. Varia-

Table 1
Energy errors (a.u.) of $1\,^1\Sigma^+$ BeO as a function of iteration number for different MCSCF calculations

Iteration	Second order	First order	Update
0	5.2363 (-03)	5.2363 (-03)	. . .
1	1.0731 (-03)	3.9877 (-04)	. . .
2	2.1553 (-05)	4.5687 (-06)	. . .
3	3.2031 (-08)	7.2492 (-07)	7.0424 (-07)
4	1.2790 (-13)	2.7935 (-07)	1.6986 (-07)
5		1.1404 (-07)	2.2733 (-08)
6		4.6922 (-08)	8.5896 (-10)
7		1.9353 (-08)	2.3761 (-11)
8		7.9881 (-09)	1.2079 (-12)
9		3.2963 (-09)	
10		1.3582 (-09)	
11		5.5736 (-10)	
12		2.2635 (-10)	
13		8.9571 (-11)	
14		3.3026 (-11)	
15		9.6634 (-12)	

Table 2
Energies (a.u.) of various MCSCF calculations for BeO

Calculation[a]	5-csf MCSCF				19-csf MCSCF			
	$1\,^1\Sigma^+$	$2\,^1\Sigma^+$	$1\,^1\Delta$	$3\,^1\Sigma^+$	$1\,^1\Sigma^+$	$2\,^1\Sigma^+$	$1\,^1\Delta$	$3\,^1\Sigma^+$
SA (3-state)	−89.467301252	−89.390185914	−89.266437476	...	−89.477200137	−89.396644977	−89.264130883	...
$2\,^1\Sigma^+$	−89.406945281	−89.399962935	...	...	−89.435637398	−89.424117509	...	...
SA (4-state)	−89.435137114	−89.373590261	−89.259133765	−89.128080572	−89.441408381	−89.376779575	−89.263313760	−89.122950384
$3\,^1\Sigma^+$	−89.367799988	−89.305801445	−89.203368458	−89.189725834	−89.371054119	−89.305768328	−89.203817855	−89.189924640

[a] See text for details.

tional collapse or root-flipping does not occur. This calculation begins with a positive definite orbital hessian and updating commences with the second iteration. Updating is cancelled in the 10th iteration because of a $(\gamma, \mathbf{x}) < 0$. It is reinitialized in the 11th iteration. The calculation converges to less than 1×10^{-12} in a total of 28 iterations. Our obtained value (cf. Table 2) is very close (1×10^{-4}) to the Fletcher algorithm result of GYJ (−89.40006310) and within 2 mh of the mode controlled result of these authors (−89.39863110) and of BY (−89.398552). The substantial deterioration of the ground state is worth noting. Explicit calculation of the partitioned orbital hessian for state 2 at the final orbitals showed no negative eigenvalues.

In the basis of our SA-MCSCF for the lowest three states, the third excited state was of $^1\Sigma^+$ symmetry. We next started our optimization of $3\,^1\Sigma^+$ by performing a 4-state SA-MCSCF starting from the orbitals of our 3-state SA-MCSCF. The initial orbital hessian is positive definite and updating commenced with the 4th iteration. Convergence to less than 10^{-12} was rapid: 6 iterations. Root flipping relative to the 3-state SA-MCSCF did not occur. The fourth state is qualitatively described by $1\pi \rightarrow 2\pi$, as expected. Following the protocol described above, we seek the $3\,^1\Sigma^+$ as the fourth root of a state-averaged calculation with an unnormalized weighting vector of $\{0.01, 0.01, 0.01, 1\}$. The initial orbital hessian has two negative eigenvalues. A positive definite orbital hessian is obtained after the third iteration and updating commences with the 7th iteration. Convergence to 10^{-12} is achieved in 17 iterations. This optimization was less straightforward than the preceding ones. Specifically, the line search algorithm required several backtracks.

The value for the fourth root (cf. Table 2) is to be compared with the result of −89.188404 by BY and the results of −89.19013440 and −89.19298010 by GYJ for Fletcher and mode control algorithms, respectively. GYJ state that neither of their results satisfies all of their criteria and do not favor one result over the other. Likewise, when we calculated the partitioned orbital hessian for the (pure) fourth root at our converged orbitals, we obtain one negative eigenvalue. A calculation for the $3\,^1\Sigma^+$

state starting from the 4-state SA-MCSCF result and with a weighting vector of (0.1, 0.1, 0.1, 1.) produced noticeably better representations of the ground and first excited states, but is qualitatively similar to the 0.01 weighting results.

Since we favor MCSCF(SD) wavefunctions when practicable, the third study performed the analogous series of calculations with such wavefunctions. These calculations paralleled the 5-csf study to great extent and are not described herein. Energy results are given in Table 2.

7. Conclusion

A line search method that uses directions of negative curvature actively has been suggested for optimization of ground and excited states of MCSCF wavefunctions. The method was implemented using previously well-established concepts in MCSCF optimization, such as level shifting the hessian to produce reasonable descent directions, were appropriate, together with the novel idea of the negative curvature direction. This method has addressed two of the most difficult problems in MCSCF optimization: effective search directions far from a stationary point and prevention of convergence to a stationary point of incorrect curvature. We showed, by numerical experiment, that the proposed method performs well for the difficult problem of optimization of the lowest three states of $^1\Sigma^+$ symmetry of BeO. It was also shown that the method performs well for various MCSCF wavefunctions, including, not surprisingly, CASSCF, but also for an MCSCF using a very small number of selected CSF; additional calculations of the MCSCF(SD)-type corroborate the findings.

The suggested method can be implemented with a variety of schemes for obtaining descent pairs. In particular, first order, second order (Newton) and updated (quasi-Newton) hessians were investigated at some level in the scope of this work. We found that a protocol that recomputes the orbital hessian far from convergence, but uses an update scheme to generate an approximate partitioned orbital hessian, is quite effective and to be recommended.

Our numerical experiments on the lowest three states of $^1\Sigma^+$ symmetry of BeO agree well with the earlier works of Golab, Yeager and Jørgensen [6] (and of Bauschlicher and Yarkony [26]) concerning both the quantitative and qualitative natures of the wavefunctions. However, our suggested scheme is based on a rigorous minimization without consideration of auxillary conditions arising, e.g., from post-MCSCF calculations. Most importantly, the suggested approach contradicts, as least for the system studied, the assertion that a so-called MCSCF CI criterion for the wavefunction leads to an unnecessarily constrained wavefunction and poor convergence. The emergence of a robust minimization scheme for excited state MCSCF wavefunctions appears to be the most important theoretical result from the current effort.

As negative convergence results are rarely described in the literature, we welcome workers to bring such problems to our attention.

References

[1] B. Roos, in: G.H.F. Diercksen, S. Wilson (Eds.), Methods of Computational Molecular Physics, Reidel, Dordrecht, 1983.
[2] J. Olsen, D.L. Yeager, P. Jørgensen, Adv. Chem. Phys. 54 (1983) 1.
[3] H.-J. Werner, Adv. Chem. Phys. 69 (1987) 1.
[4] R. Shepard, Adv. Chem. Phys. 69 (1987) 63.
[5] M.W. Schmidt, M.S. Gordon, Annu. Rev. Phys. Chem. 49 (1998) 233.
[6] J.T. Golab, D.L. Yeager, P. Jørgensen, Chem. Phys. 78 (1983) 175.

[7] J.T. Golab, D.L. Yeager, P. Jørgensen, Chem. Phys. 93 (1985) 83.

[8] P.E. Gill, W. Murray, M.H. Wright, Practical Optimization, Academic Press, New York, 1981.

[9] D. O'Neal, H. Taylor, J. Simons, J. Phys. Chem. 88 (1984) 1510.

[10] A. Banerjee, N. Adams, J. Simons, R. Shepard, J. Phys. Chem. 89 (1985) 52.

[11] J. Simons, J. Nichols, Int. J. Quantum Chem. S24 (1990) 263.

[12] L. Yaffe, W.A. Goddard, Phys. Rev. A 13 (1976) 1682.

[13] E. Dalgaard, P. Jørgensen, J. Chem. Phys. 69 (1978) 3833.

[14] C.C.J. Roothaan, J. Detrich, D.G. Hopper, Int. J. Quantum Chem. Symp. 13 (1979) 93.

[15] B.H. Lengsfield, J. Chem. Phys. 77 (1982) 4073.

[16] K.K. Docken, J. Hinze, J. Chem. Phys. 57 (1972) 4928.

[17] H.J. Werner, W. Meyer, J. Chem. Phys. 74 (1981) 5802.

[18] R.N. Diffenderfer, D.R. Yarkony, J. Phys. Chem. 86 (1982) 5098.

[19] P. Jørgensen, P. Swanstrøm, D.L. Yeager, J. Chem. Phys. 78 (1983) 347.

[20] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall, Englewood Cliffs, NJ, 1983.

[21] J.J. Moré, D.C. Sorensen, Math. Prog. 16 (1979) 1.

[22] G.P. McCormick, Math. Prog. 13 (1977) 111.

[23] D. Goldfarb, Math. Prog. 18 (1980) 31.

[24] J.J. Moré, S.J. Wright, Optimization Software Guide Frontiers in Applied Mathematics, vol. 14, SIAM, Philadelphia, PA, 1994.

[25] (a) M.R. Hoffmann, G.C. Schatz, J. Chem. Phys. 113 (2000) 9456;
(b) M.R. Hoffmann, G.C. Schatz, in: M.R. Hoffmann, K.G. Dyall (Eds.), Low-Lying Potential Energy Surfaces, ACS Symposium Series, American Chemical Society, Washington, DC, in press.

[26] C.W. Bauschlicher Jr., D.R. Yarkony, J. Chem. Phys. 72 (1980) 1138.

[27] C.W. Bauschlicher Jr., B.H. Lengsfield III, D.R. Yarkony, J. Chem. Phys. 73 (1980) 5702.

[28] H.J.Aa. Jensen, P. Jørgensen, H. Ågren, J. Olsen, J. Chem. Phys. 88 (1988) 3834.