Theor Chim Acta (1996) 94: 233–245

Theoretica Chimica Acta

Exact transformations of CI spaces, VB representations of CASSCF wavefunctions and the optimization of VB wavefunctions

Thorstein Thorsteinsson*, **David L. Cooper**

Department of Chemistry, University of Liverpool, P.O. Box 147, Liverpool L69 3BX, UK

Received April 3, 1996/Final revision received June 5, 1996/Accepted June 5, 1996

Summary. It is demonstrated how exact transformations of full CI spaces may be carried out for general, non-unitary orbital transformations, and a detailed description of an efficient implementation of this scheme is presented for Slater determinants. It is then shown how this technology may be employed in the optimization of general VB wave functions with respect to both orbital and structure coefficients. This may be done in a straightforward manner, incorporating first and second derivatives of the variational parameters.

Key words: CASSCF – VB – Orbital optimization

1 Introduction

The present popularity of the complete active space self-consistent field (CASSCF) method [1] is beyond dispute, and most standard *ab initio* packages now provide the facility to carry out such calculations. The CASSCF method is conceptually very simple, unbiased (in the sense that only the number of active orbitals, m, and active electrons, N, need to be specified for a given molecular system), and typically leads to reliable energy surfaces and molecular properties. The underlying reason for the *practicality* of these calculations, and hence their ubiquity, clearly lies in the vast theoretical and computational simplifications that are possible when the CI space is complete.

The present work shows how some of these extraordinary properties of full CI spaces may be utilized to determine very efficiently the exact structure transformation corresponding to a general, non-unitary orbital transformation. We can then use this approach to 'circumvent' the non-orthogonality problem normally associated with the optimization of VB wavefunctions.

Valence bond methods have now firmly established themselves as serious alternatives to traditional MO methods. Exemplified by the spin-coupled valence bond approach [2], high quality, very compact wavefunctions may be obtained which have obvious advantages with respect to their interpretation, because of the

---

* *Current address:* School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK

close correspondence to traditional chemical concepts of bonding. At the present time, however, only a small minority of quantum chemists employ VB methods on a regular basis, at least in part because of the limited availability of efficient general program packages. Indeed, developing code comparable in sophistication to the current MO packages would require an immense effort and so it might seem that there is unlikely to be any significant shift towards modern VB methods in the foreseeable future.

The motivation behind this research is twofold, as follows. Firstly, the general availability of inexpensive procedures for generating compact VB representations of CASSCF wavefunctions could help bridge the existing conceptual gap between these two competing approaches. Secondly, the strategies described here also represent a very significant advance in the general development of VB methodology. In principle, *all* the available technology associated with CASSCF wavefunctions now becomes available for the variational optimization of general types of VB wavefunction. Among the many examples one might mention are: utilization of point group symmetry; optimization of excited states (or a weighted average thereof); gradients and geometry optimization; direct approaches, with on on-the-fly evaluation of integrals.

Representations of CASSCF wavefunctions based on the spin-coupled wavefunction have already been obtained for a range of systems [3, 4]. The CASSCF/VB overlap has always been close to unity, and the converged solutions are similar to the corresponding variational results. We plan to publish further applications of this approach in the near future.

The structure of this paper is as follows. In Sect. 2 we develop the underlying theory for the transformations of full CI spaces induced by general transformations of the defining orbitals. Sect. 3 describes our specific implementation of this, using Slater determinants. In Sect. 4 we consider optimization of valence bond wavefunctions and two examples of criteria that might be defined for this purpose. Sect. 5 describes, on the basis of these, how our general strategy may be incorporated into a general second-order optimization procedure. A final discussion is presented in Sect. 6, and a short consideration of some of the consequences of molecular point group symmetry is included as an appendix.

2 Transformations of full CI spaces

We consider in this section the effect of changing the orbital representation of a wavefunction of full CI type, such as a CASSCF wavefunction. Similar considerations have been presented by Malmqvist [5], in the context of the so-called "CASSCF state interaction method" [6]. The transformation of the full CI space is of considerable general utility and it is of potential interest to anyone considering the optimization of MCSCF or VB wavefunctions.

For any general (non-unitary) linear transformation of the m active orbitals[1]

$$\{\phi'\} = \{\phi\}\mathbf{O}, \tag{1}$$

---

[1] The wavefunction may, without loss of generality, contain a "core" part, provided that this part is identical in all structures. It is then necessary to consider only the orbitals with variable occupations, i.e., the active set.

there exists a corresponding transformation of the structure space

$$\{\Phi'\} = \{\Phi\}\mathbf{T}(\mathbf{O}), \tag{2}$$

in which $\{\Phi\}$ is the row-vector of the complete set of structures (configurations) with appropriate symmetry and spin, defined in terms of the orbitals $\{\phi\}$, and $\{\Phi'\}$ is the corresponding row-vector defined in terms of the $\{\phi'\}$. The square matrix $\mathbf{O}$ is thus of dimension $m \times m$ and the square matrix $\mathbf{T}(\mathbf{O})$ is of dimension $N_{\mathrm{CI}} \times N_{\mathrm{CI}}$, with N_{CI} being the number of functions, Φ, that define the full CI space. Equation (2) is a well-known property of full CI spaces, but one may find analogous identities for certain other types of CI space. An expansion including all configurations up to a given excitation level from the closed-shell reference function, for example, will be invariant under linear transformations of *either* the reference orbitals *or* the virtuals. We next take advantage of a simple property of the structure transformations that, assuming non-singular transformations, they must form a faithful representation of the group of orbital transformations:

$$\mathbf{O} = \mathbf{O}_1\mathbf{O}_2 \Leftrightarrow \mathbf{T}(\mathbf{O}) = \mathbf{T}(\mathbf{O}_1)\mathbf{T}(\mathbf{O}_2). \tag{3}$$

A useful result deriving from this, since $\mathbf{T}(\mathbf{1}) = \mathbf{1}$ (where $\mathbf{1}$ and $\mathbf{1}$ are the $m \times m$ and $N_{\mathrm{CI}} \times N_{\mathrm{CI}}$ identity matrices, respectively), is

$$\mathbf{T}^{-1}(\mathbf{O}) = \mathbf{T}(\mathbf{O}^{-1}). \tag{4}$$

Our aim is then to write the orbital transformation as a product of $m \times m$ simple "updates" of the form

$$\mathbf{O}_{\mu\nu}(\lambda)\colon \phi_\nu \rightarrow \phi_\nu + \lambda\phi_\mu \tag{5}$$

for which the corresponding structure transformation will be straightforward to evaluate. The diagonal case, $\mu = \nu$, represents a scaling of ϕ_μ by $1 + \lambda$. If the following identity holds for a given set of λ parameters and ordering of updates

$$\mathbf{O} = \mathbf{O}_{11}(\lambda_1)\mathbf{O}_{12}(\lambda_2)\mathbf{O}_{13}(\lambda_3)\ \ldots\ \mathbf{O}_{mm}(\lambda_{m^2}), \tag{6}$$

then the total transformation of the structure space is just the corresponding product of $m \times m$ simple structure transformations. These will be defined in detail later. Recasting Eq. (6) as

$$\mathbf{O}_{mm}^{-1}(\lambda_{m^2})\ \ldots\ \mathbf{O}_{11}^{-1}(\lambda_1)\mathbf{O} = \mathbf{1}, \tag{7}$$

and noting that $\mathbf{O}_{\mu\nu}^{-1}(\lambda_i)$ represents a "row operation" on $\mathbf{O}$ (adding a multiple of row ν to row μ), we realize that a solution may be found by any of the numerous standard numerical methods for solving linear problems. Examples of commonly used procedures include Gaussian elimination with back-substitution, Jordan's method, and Crout's factorization algorithm (for LU decomposition) – each of these essentially reduces a square matrix to the identity by a sequence of row operations [7]. The overriding consideration in our choice of algorithm is the numerical stability, since any inaccuracies in the factorization (6) are likely to accumulate when constructing the corresponding structure transformation. In particular, algorithms will full pivoting are to be preferred.

We note that a λ parameter used for the $\nu \rightarrow \mu$ update in Eq. (6) will in general be different from the corresponding element of the orbital transformation matrix, $(\mathbf{O})_{\mu\nu}$, although they may coincide for certain simple transformations. This is a consequence of using consecutive single-orbital updates to achieve the effect of a simultaneous transformation of all the orbitals.

With the usual assumption of a "restricted" set of spin orbitals, so that $\mathbf{O}^{\alpha} = \mathbf{O}^{\beta}$, the $\nu \rightarrow \mu$ update of the *spatial* orbital (Eq. (5)) can be realized by identical, consecutive updates of the α and β spin orbitals. The corresponding transformation of the structure space is thus realized by application of the operator

$$(\hat{I} + \lambda \hat{E}^{\alpha}_{\mu\nu})(\hat{I} + \lambda \hat{E}^{\beta}_{\mu\nu}) = \hat{I} + \lambda \hat{E}^{(1)}_{\mu\nu} + \lambda^2 \hat{E}^{(2)}_{\mu\nu}. \tag{8}$$

The orbitals defining the structures may in general be non-orthogonal and so the excitation operator (for the α orbital) must be defined according to

$$\hat{E}^{\alpha}_{\mu\nu} = a^{\alpha\dagger}_{\mu} \tilde{a}^{\alpha}_{\nu}, \tag{9}$$

with a dual annihilation operator (cf. Ref. [8]). This is necessary to preserve the usual anticommutation relation between creation and annihilation operators, such that the action of $\hat{E}^{\alpha}_{\mu\nu}$ is a simple replacement of ϕ_{ν} by ϕ_{μ}. We must remark, however, that the role of the dual set in this context should be viewed merely as one of notational convenience. $\hat{E}^{(1)}_{\mu\nu}$ in Eq. (8) describes single excitations $\nu \rightarrow \mu$ and $\hat{E}^{(2)}_{\mu\nu}$ describes double excitations. Similarly, $\hat{E}^{(1)}_{\mu\mu}$ is the usual "number operator" for orbital μ, and $\hat{E}^{(2)}_{\mu\mu}$ has eigenvalues 1 or 0 according to whether orbital μ is *doubly* occupied or not.

Clearly, the matrix representations of the excitation operators inherent in Eq. (2) are independent of orbital overlap. This is most readily seen by inserting a resolution of the identity, $\sum_K |\Phi_K\rangle \langle \tilde{\Phi}_K|$, into the definition of, say $\mathbf{E}^{(1)}_{\mu\nu}$:

$$\hat{E}^{(1)}_{\mu\nu} \Phi_I = \sum_J (\mathbf{E}^{(1)}_{\mu\nu})_{IJ} \Phi_J, \tag{10}$$

so as to obtain

$$(\mathbf{E}^{(1)}_{\mu\nu})_{IJ} = \langle \tilde{\Phi}_I | \hat{E}^{(1)}_{\mu\nu} | \Phi_J \rangle. \tag{11}$$

The elements of this matrix coincide with the standard one-particle coupling coefficients.

It should be stressed that the formulae given above do not involve any approximations whatsoever. The *exact* transformation of the structure space is thus determined simply by $m \times m$ consecutive applications of Eq. (8), using the ordering of updates and λ parameters found from Eq. (6).

3 Expansion in Slater determinants

In view of our earlier remarks, a full CI transformation is likely to be of greatest utility in the context of one of the existing CASSCF modules. This heavily influences the choice of expansion functions. One of our aims was to write standard code which could easily be incorporated into standard molecular packages such as MOLPRO [9], and it was therefore natural to consider Slater determinants as the functions spanning the CASSCF space. Another consideration is that an integral-driven (or "direct") CI code might be available, in which case the structure transformation can be implemented fairly straightforwardly from the code determining the one-electron contribution to the CI vector. The direct CI scheme originally due to Siegbahn [10] has been adapted to full CI Slater determinant spaces by Knowles and Handy [11].

A significant computational convenience associated with using Slater determinants to span the full CI space is that the α and β parts may be considered separately, i.e., $\Phi = \Phi^\alpha \Phi^\beta$ (see also Ref. [11]). If the α orbitals occur first in all determinants, for example, any phase factor relating to the excitation of an α orbital will be independent of the nature of the β string (and vice versa). For the same reason, it is natural to store the CI vector as $\mathbf{c}(I^\alpha, I^\beta)$ (essentially a rectangular matrix), where I^α is the index of the α string only, and I^β is the index of the β string. For the special case $S = 0$, the matrix $\mathbf{c}$ is square symmetric, and only its triangular part has to be stored. One must also mention the most serious disadvantage of this approach, which is the larger number of determinants required, relative to the number of configuration state functions (CSFs).

The basic strategy that we use for the update of an α orbital involves creating an intermediate string of $N_\alpha - 1$ α orbitals, so as to minimize the need for large indexing arrays. The scheme may be summarized as follows:

1. Loop over all α strings containing the orbital ϕ_ν.
2. For each α string rearrange so as to get ϕ_ν first – the associated phase factor is P_1.
3. "Annihilate" ϕ_ν. This will give an intermediate $N_\alpha - 1$ string.
4. "Create" ϕ_μ. Obtain the index for the final α string.
5. Rearrange the α string so as to get ascending orbitals – the associated phase factor is P_2.
6. Loop over all β strings. Update according to $\mathbf{c}(I^\alpha_{\text{to}}, I^\beta) := \mathbf{c}(I^\alpha_{\text{to}}, I^\beta) + \lambda P_1 P_2 \times \mathbf{c}(I^\alpha_{\text{from}}, I^\beta)$.

Note that in-place updating of the CI vector is possible because no recurrence ever occurs in the loop – the I^α_{from} and I^α_{to} indices form disjoint sets. Steps 2 to 5 may be omitted for the case $\mu = \nu$. The indexing and phase information may be conveniently kept in precalculated arrays because of the modest associated storage requirement.

The loop structure must be modified slightly for the special case $S = 0$, since only the upper triangle of $\mathbf{c}$ will be updated. The most important consideration is how to retain the in-place updating of the CI vector while still avoiding the problem of recurrence. If $\mu = \nu$, this is simply achieved by restricting the length of the inner loop. Otherwise, it is important that the α and β updates are done in pairs, in order, as far as is possible, to retain the $\alpha \leftrightarrow \beta$ symmetry. One possible scheme is shown in Fig. 1, in which the inner loop structure replaces item 6 above.

The algorithm we have just outlined has an extremely simple loop structure which should enable efficient implementation on modern computer architectures. The length of the inner loop will be 10^2–10^3 for typical systems, each iteration consisting of a simple multiply-add or, for the case, $\mu = \nu$ a simple multiply. The effort associated with a complete transformation of a CI vector, as defined by the total number of inner loop iterations N_{L}, is given in Table 1 for a range of representative combinations of N, m and S. It has been assumed here that all the λ parameters occurring in Eq. (6) are non-zero. This would not be the case if, for example, some of the orbitals were orthogonal, perhaps because of symmetry considerations.

We have also shown in Table 1 the corresponding numbers for bonded structures, based on Rumer functions, although in this case each iteration is slightly more involved than for determinants. The number of determinants is simply the product of the number of α and β strings:

$$N_{\text{det}}(N, m, S) = \binom{m}{N_\alpha} \times \binom{m}{N_\beta}, \tag{12}$$

```
Alpha update

If ($I^\alpha_{\mathrm{to}} > I^\alpha_{\mathrm{from}}$) then
        Do $I^\beta = I^\alpha_{\mathrm{to}}, N^\beta_{\mathrm{det}}$
            $c(I^\alpha_{\mathrm{to}}, I^\beta) := c(I^\alpha_{\mathrm{to}}, I^\beta) + \lambda P_1 P_2 \times c(I^\alpha_{\mathrm{from}}, I^\beta)$
        End Do ($I^\beta$)

Else
        Do $I^\beta = I^\alpha_{\mathrm{to}}, I^\alpha_{\mathrm{from}}$
            $c(I^\alpha_{\mathrm{to}}, I^\beta) := c(I^\alpha_{\mathrm{to}}, I^\beta) + \lambda P_1 P_2 \times c(I^\beta, I^\alpha_{\mathrm{from}})$
        End Do ($I^\beta$)
        Do $I^\beta = I^\alpha_{\mathrm{from}} + 1, N^\beta_{\mathrm{det}}$
            $c(I^\alpha_{\mathrm{to}}, I^\beta) := c(I^\alpha_{\mathrm{to}}, I^\beta) + \lambda P_1 P_2 \times c(I^\alpha_{\mathrm{from}}, I^\beta)$
        End Do ($I^\beta$)

End If

Beta update

If ($I^\beta_{\mathrm{to}} > I^\beta_{\mathrm{from}}$) then
        Do $I^\alpha = 1, I^\beta_{\mathrm{from}} - 1$
            $c(I^\alpha, I^\beta_{\mathrm{to}}) := c(I^\alpha, I^\beta_{\mathrm{to}}) + \lambda P_1 P_2 \times c(I^\alpha, I^\beta_{\mathrm{from}})$
        End Do ($I^\alpha$)
        Do $I^\alpha = I^\beta_{\mathrm{from}}, I^\beta_{\mathrm{to}}$
            $c(I^\alpha, I^\beta_{\mathrm{to}}) := c(I^\alpha, I^\beta_{\mathrm{to}}) + \lambda P_1 P_2 \times c(I^\beta_{\mathrm{from}}, I^\alpha)$
        End Do ($I^\alpha$)
Else
        Do $I^\alpha = 1, I^\beta_{\mathrm{to}}$
            $c(I^\alpha, I^\beta_{\mathrm{to}}) := c(I^\alpha, I^\beta_{\mathrm{to}}) + \lambda P_1 P_2 \times c(I^\alpha, I^\beta_{\mathrm{from}})$
        End Do ($I^\alpha$)

End If
```

Fig. 1. Schematic inner loop structure for the $\phi_\nu \to \phi_\mu$ update in the case of $S = 0$

with $N_\alpha = \frac{1}{2}N + S$ and $N_\beta = \frac{1}{2}N - S$, whereas the dimension of the space formed from bonded structures, N_{CAS}, can be shown to be [12]

$$N_{\mathrm{CAS}}(N, m, S) = \frac{2S+1}{m+1}\binom{m+1}{\frac{1}{2}N - S}\binom{m+1}{\frac{1}{2}N + S + 1}. \tag{13}$$

It is noticeable (see Table 1) that the effort associated with evaluating **T(O)** is not very large relative, say, to that involved in a simple scalar product between two CI vectors – for the *largest* systems given here, the value of N_{L} for a complete structure transformation is approximately 110 times the number of loop iterations needed for a scalar product.

4 Optimizing VB wavefunctions

In this and the next section we consider one particular application of the technology of structure transformations, namely the optimization of valence bond functions – either independently or as a means of providing VB representations of full CI wavefunctions. The general idea is simply that the non-orthogonality problem

Table 1. Number of loop iterations (N_L) associated with one structure transformation of a single 'N in N' CI vector. N_{det} is the total number of determinants and N_{CAS} is the numbers of CSFs in the CI space: these values could be reduced by making use of any molecular point group symmetry

N	S	N_{CAS}	N_{det}	N_L (determinants)	N_L (Rumer functions)
2	0	3	4	8	10
2	1	1	1	2	2
3	1/2	8	9	63	56
4	0	20	36	216	228
4	1	15	16	160	162
5	1/2	75	100	1.70×10^3	1.44×10^3
6	0	175	400	4.80×10^3	4.50×10^3
6	1	189	225	4.95×10^3	5.00×10^3
7	1/2	784	1225	3.80×10^4	3.07×10^4
8	0	1764	4900	9.80×10^4	8.48×10^4
8	1	2352	3136	1.19×10^5	1.21×10^5
9	1/2	8820	15876	7.78×10^5	6.08×10^5
10	0	19404	63504	1.91×10^6	1.56×10^6
10	1	29700	44100	2.56×10^6	2.61×10^6
11	1/2	104544	213444	1.52×10^7	1.16×10^7
12	0	226512	853776	3.59×10^7	2.83×10^7
12	1	382239	627264	5.14×10^7	5.28×10^7
13	1/2	1288287	2944656	2.86×10^8	2.15×10^8
14	0	2760615	11778624	6.60×10^8	5.07×10^8
14	1	5010005	9018009	9.92×10^8	1.03×10^9

may be circumvented once the structure transformation $\mathbf{T}(\mathbf{O})$ is known. We have, for example,

$$\mathbf{H}' = \mathbf{T}^{\dagger}(\mathbf{O})\mathbf{H}\mathbf{T}(\mathbf{O}), \tag{14}$$

$$\mathbf{S}' = \mathbf{T}^{\dagger}(\mathbf{O})\mathbf{S}\mathbf{T}(\mathbf{O}), \tag{15}$$

and

$$\mathbf{c}'_{\text{CAS}} = \mathbf{T}^{-1}(\mathbf{O})\mathbf{c}_{\text{CAS}}. \tag{16}$$

So, if $\mathbf{H}$, $\mathbf{S}$ and $\mathbf{c}_{\text{CAS}}$ are first evaluated using orthogonal orbitals they may then subsequently be transformed straightforwardly to the corresponding quantities in terms of non-orthogonal orbitals.

We shall assume that the valence bond function can be expressed in terms of structures:

$$\Psi_{\text{VB}} = \sum_{I}^{N_{\text{VB}}} b_I \Psi_I^{\text{VB}}, \tag{17}$$

in which the Ψ_I^{VB} are formed from a set of m linearly independent, in general non-orthogonal, (active) orbitals $\{\Phi^{\text{VB}}\}$. Clearly, the valence bond wavefunction will always lie in the "N in m" full CI space formed from any set of orbitals spanning the same space as the $\{\phi^{\text{VB}}\}$. The valence bond structures will in general be related by a linear transformation to a subset (often quite small) of transformed structures

$$\{\Psi^{\text{VB}}\} = \{\Phi^{\text{VB}}\}\mathbf{D}, \tag{18}$$

where $\mathbf{D}$ is a rectangular matrix of dimension N_{CAS} by N_{VB} or N_{det} by N_{VB}, depending on the type of structures used. In the case of the spin-coupled wavefunction, for example, $\mathbf{D}$ would consist of the transformation coefficients from covalent determinants to spin eigenfunctions.

We have previously considered two main types of wavefunction optimization [3, 4]: straightforward minimization of the energy expression

$$E_{\mathrm{VB}}=\frac{\langle\Psi_{\mathrm{VB}}|\hat{H}|\Psi_{\mathrm{VB}}\rangle}{\langle\Psi_{\mathrm{VB}}|\Psi_{\mathrm{VB}}\rangle}=\frac{\mathbf{b}^{\dagger}\mathbf{D}^{\dagger}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{H}\mathbf{T}(\mathbf{O})\mathbf{D}\mathbf{b}}{\mathbf{b}^{\dagger}\mathbf{D}^{\dagger}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{S}\mathbf{T}(\mathbf{O})\mathbf{D}\mathbf{b}}, \tag{19}$$

or, alternatively, maximization of the overlap with a previously optimized CASSCF wavefunction

$$S_{\mathrm{VB}}=\frac{\langle\Psi_{\mathrm{CAS}}|\Psi_{\mathrm{VB}}\rangle}{\langle\Psi_{\mathrm{VB}}|\Psi_{\mathrm{VB}}\rangle^{1/2}}=\frac{\mathbf{c}_{\mathrm{CAS}}^{\dagger}\mathbf{S}\mathbf{T}(\mathbf{O})\mathbf{D}\mathbf{b}}{(\mathbf{b}^{\dagger}\mathbf{D}^{\dagger}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{S}\mathbf{T}(\mathbf{O})\mathbf{D}\mathbf{b})^{1/2}}. \tag{20}$$

Many other criteria might be envisaged, but the two given here are sufficient to illustrate the underlying ideas. Neither $\mathbf{H}$ nor $\mathbf{S}$ will in general be stored explicitly, the crucial point being that, as they are formed from orthogonal orbitals, they may be constructed very efficiently.

Both criteria given above can be used to obtain very compact approximations to CASSCF wavefunctions, in that if either quantity is optimized with respect to the orbital transformation matrix $\mathbf{O}$ and the vector $\mathbf{b}$ in Eq. (17) (giving $m\times(m-1)$ and $N_{\mathrm{VB}}-1$ free parameters respectively), then we have $\mathbf{Db}\approx\mathbf{T}(\mathbf{O}^{-1})\mathbf{c}_{\mathrm{CAS}}$ (cf. Eq. (16)) at convergence [3, 4]. With this in mind, it is natural to introduce the projection operator $\hat{P}_{\mathrm{VB}}$ (with matrix representation $\mathbf{P}_{\mathrm{VB}}$), defined so that $\mathbf{P}_{\mathrm{VB}}\mathbf{T}(\mathbf{O}^{-1})\mathbf{c}_{\mathrm{CAS}}$ is a valid valence bond wavefunction. This may then be substituted for $\mathbf{Db}$ in Eqs. (19) and (20), so that the complexity of the optimization problem is somewhat reduced (by the number of linear variational parameters, $N_{\mathrm{VB}}-1$) [3, 4].

If a representation of a CASSCF wavefunction is sought, then the orbital set in Eq. (1) will be identical to the optimized active CASSCF orbital set. However, independent optimization of the valence bond wavefunction is also possible. Of course, the active-virtual orbital rotations must be included in the optimization procedure, and, if the wavefunction contains also a core part, the rotations involving the core orbitals. This is most simply achieved by a two-step procedure with alternating updates of the "internal" variational parameters defining $\mathbf{O}$ and $\mathbf{b}$, and the "external" orbital rotation parameters, as has previously been done in Ref. [13]. A more efficient procedure will entail incorporating also the coupling between these two parameter sets, and we consider this briefly in Sect. 5.

The quantities E_{VB} and S_{VB} are most efficiently evaluated by constructing each term sequentially left-to-right (or right-to-left), so that at most two vectors need be stored in memory at any particular time. A simplification is possible when $\mathbf{S}$ is the identity (for determinants or structures based on Kotani functions) in that $\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{T}(\mathbf{O})=\mathbf{T}(\mathbf{O}^{\dagger}\mathbf{O})=\mathbf{T}(\mathbf{s})$, $\mathbf{s}$ being the orbital overlap matrix. Assuming orthogonal structures, evaluating S_{VB} in Eq. (19) can then clearly be achieved with two transformations of the structure space. Similar E_{VB} in Eq. (20) may be evaluated using just two structure transformations, but the evaluation will normally be dominated by the effort involved in multiplying the Hamiltonian, $\mathbf{H}$, on the vector $\mathbf{T}(\mathbf{O})\mathbf{Db}$. To our knowledge, the most efficient approach for achieving this is represented by further developments of Siegbahn's algorithm [10], as may be found in Refs. [14–16]. The operation count is approximately

$$N_{\mathrm{H}} = N_{\mathrm{det}} \times \left(\tfrac{1}{4}N_\alpha^2(m - N_\alpha)^2 + \tfrac{1}{4}N_\beta^2(m - N_\beta)^2 + N_\alpha N_\beta (m - N_\alpha)(m - N_\beta)\right) \quad (21)$$

multiply-adds, assuming that all of the two-electron integrals are non-negligible. The three terms in Eq. (21) are associated with the $\alpha\alpha$, $\beta\beta$ and $\alpha\beta$ parts of the two-electron integrals, respectively. Thus, for the larger systems considered in Table 1, the effort associated with applying the Hamiltonian is roughly thirty times greater than that involved in applying $\mathbf{T}(\mathbf{s})$.

5 Derivative expressions

Orbital optimization in both MCSCF and VB is generally sufficiently complex to warrant the use of full second-order methods, and thus evaluation of first and second derivatives. This may be done straightforwardly in the present scheme by considering the effect of variations in $\mathbf{O}$ or $\mathbf{b}$ on the quantities being optimized (Eq. (19) or (20)). For the linear coefficients, this is entirely trivial, in that we have

$$\frac{\partial \mathbf{b}}{\partial b_I} = \boldsymbol{\delta}^{(I)} A \quad (22)$$

in which $\boldsymbol{\delta}^{(I)}$ is unity at position I, and is otherwise zero. From this it is then straightforward to find the corresponding component of the gradient for E_{VB} or S_{VB}.

The first-order transformation matrix corresponding to $\mathbf{O}_{\mu\nu}$ can be determined easily from Eq. (8) as

$$\left[\frac{\partial \mathbf{T}(\mathbf{O}_{\mu\nu}(\lambda))}{\partial \lambda}\right]_{\lambda=0} = \mathbf{E}_{\mu\nu}^{(1)}, \quad (23)$$

i.e., simply the one-particle coupling coefficients. For the purpose of evaluating the associated gradient we write the total orbital transformation as $\mathbf{O}\mathbf{O}_{\mu\nu}(\lambda)$, giving a corresponding first-order structure transformation as $\mathbf{T}(\mathbf{O})\mathbf{E}_{\mu\nu}^{(1)}$.

Similarly, it is straightforward to show that the second-order change corresponding to the simultaneous updates

$$\mathbf{O}_{\mu\nu,\sigma\tau}(\lambda_1, \lambda_2)\colon \phi_\nu \rightarrow \phi_\nu + \lambda_1 \phi_\mu \wedge \phi_\tau \rightarrow \phi_\tau + \lambda_2 \phi_\sigma \quad (24)$$

reduces to the two-particle coupling coefficients (cf. Ref. [17])

$$\left[\frac{\partial^2 \mathbf{T}(\mathbf{O}_{\mu\nu,\sigma\tau}(\lambda_1, \lambda_2))}{\partial \lambda_1 \partial \lambda_2}\right]_{\lambda_1, \lambda_2 = 0} = \mathbf{E}_{\mu\nu}^{(1)}\, \mathbf{E}_{\sigma\tau}^{(1)} - \delta_{\nu\sigma} \mathbf{E}_{\mu\tau}^{(1)}. \quad (25)$$

This may be shown trivially by writing $\mathbf{O}_{\mu\nu,\sigma\tau}$ as a product of single-orbital updates of the form in Eq. (5). It is not in fact necessary to evaluate the last term in Eq. (25) provided that the correct ordering of single-excitation matrices is ensured in the product; for the case $\mathbf{E}_{\mu\nu}^{(1)}\mathbf{E}_{\nu\mu}^{(1)} - \mathbf{E}_{\mu\mu}^{(1)}$, the last term will give an *overall* contribution of zero, since the optimization expression will normally be independent of orbital normalization. An expression for the total Hessian may then be constructed by considering second-order changes arising from simultaneous updates (Eq. (24)), as well as from combinations of single orbital updates (Eq. (5)) and linear variations (Eq. (22)).

We now consider the computational effort associated with the evaluation of first and second derivatives. For both $\min(E_{\text{VB}})$ and $\max(S_{\text{VB}})$, most of the required computational effort will arise from the combinations of first-order updates contributing to the total Hessian matrix. For the numerator defining E_{VB}, for example, these will be of the form

$$\mathbf{b}^{\dagger}\mathbf{D}^{\dagger}\mathbf{E}_{\mu\nu}^{\dagger(1)}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{H}\mathbf{T}(\mathbf{O})\mathbf{E}_{\sigma\tau}^{(1)}\mathbf{D}\mathbf{b}, \tag{26}$$

$$\mathbf{b}^{\dagger}\mathbf{D}^{\dagger}\mathbf{E}_{\mu\nu}^{\dagger(1)}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{H}\mathbf{T}(\mathbf{O})\mathbf{D}\delta^{(I)}, \tag{27}$$

and

$$\delta^{(I)\dagger}\mathbf{D}^{\dagger}\mathbf{T}^{\dagger}(\mathbf{O})\mathbf{H}\mathbf{T}(\mathbf{O})\mathbf{D}\delta^{(J)}, \tag{28}$$

and so a total of $m\times(m-1)+N_{\text{VB}}-1$ multiplications by **H** will be required (variations associated with normalization conditions may be omitted). The same number of structure transformations, **T**(s), will be required for variations of the denominator (as well as for variations of S_{VB}). In general, N_{var} multiplications by **H** and/or structure transformations will be required for the evaluation of this part of the Hessian, where N_{var} is the number of free, non-redundant variables.

It is clear that only a small part of the Hessian evaluation is responsible for the vast majority of the computational effort, and it is appropriate to take this into account in the optimization procedure. This could, for example, be achieved by updating this part of the Hessian only every so many iterations, or testing more than one update at a time based on the exact Hessian matrix.

In the course of a standard CASSCF calculation, an important step is constructing and solving the full-CI secular problem in order to determine the optimal CI coefficients. This is then followed by the optimization of the external orbital rotations. This second phase may or may not incorporate relaxation of the CI vector, i.e., coupling between the CI coefficients and orbital rotations. For further details we refer the reader to the literature, e.g., Ref. [18]. In our case, the CI optimization step will clearly be replaced by the optimization of internal VB variables as described above. For the coupling between internal and external parameters, the first-order change in the CI vector corresponding to changes in the internal parameters must be gauged. The first-order change of the CI vector will be of the form

$$\mathbf{T}(\mathbf{O})\mathbf{E}_{\mu\nu}^{(1)}\mathbf{D}\mathbf{b} \tag{29}$$

for an orbital update, or

$$\mathbf{T}(\mathbf{O})\mathbf{D}\delta^{(I)} \tag{30}$$

for the first-order change of the linear coefficient, b_I. Thus, if the coupling between CI coefficients and external parameters is known, it is also straightforward to obtain the coupling between the internal VB parameters and the external orbital rotations. A compromise strategy would be to vary the CI coefficients freely in response to external orbital rotations, then eliminate variations not of the form $\mathbf{T}(\mathbf{O})\mathbf{E}_{\mu\nu}^{(1)}\mathbf{D}\mathbf{b}$ or $\mathbf{T}(\mathbf{O})\mathbf{D}\delta^{(I)}$, before reoptimizing the external orbital parameters using the CI coefficients thus obtained.

6 Discussion

The simple idea, outlined in Sect. 2 and Refs. [3, 5], of arranging *sequential* updates so as to obtain the effect of simultaneous ones, may be a useful alternative to the approaches generally adopted in MCSCF theory, most of which do not go beyond the (admittedly very useful) 'exponential-i-lambda' method [19, 20] of describing the structure transformation. For a full CI space, the most efficient treatment of the structure transformation is achieved by using Slater determinants. In our specific implementation, the following characteristics are particularly worthy of mention:

1. The total effort scales approximately *linearly* with the size of the CI space (cf. Table 1).
2. In-place updating of the CI vector.
3. Proper treatment of the case $S = 0$ in order to minimize memory and CPU requirements.
4. Intermediate $N^{\alpha} - 1$ and $N^{\beta} - 1$ strings to avoid large indexing arrays.

The second point is particularly important in order to avoid superfluous copying or zeroing of the CI vector. If CSFs are used instead of determinants, in-place updating can similarly be achieved provided that care is taken to eliminate the possibility of recurrence. In most cases where an integral-driven, direct CI algorithm exists, implementation of the structure transformation scheme should be very straightforward.

The strategy of structure transformations is very well suited to the optimization of valence bond wavefunctions, provided that the numbers of active electrons and orbitals are of comparable magnitudes to what may be treated by CASSCF. The formalism presented here is very flexible, both regarding variations of the type of VB wavefunction that may be optimized, as well as the optimization criteria that can be used. Independent optimization of the wavefunction is possible, or the strategy may be employed for obtaining a compact representation of a previously optimized CASSCF wavefunction. The computational effort associated with an iteration, employing a complete second-order optimization scheme, will then be approximately

$$N_{\mathrm{var}} \times N_{\mathrm{H}} \tag{31}$$

for optimizing quantities containing the term $\langle \Psi_{\mathrm{VB}} | \hat{H} | \Psi_{\mathrm{VB}} \rangle$ (see Eq. (21)), and

$$N_{\mathrm{var}} \times N_{\mathrm{L}} \tag{32}$$

for quantities containing only $\langle \Psi_{\mathrm{VB}} | \Psi_{\mathrm{VB}} \rangle$, N_{var} being the number of free variational parameters defining the wavefunction.

Overall, the strategy we have presented here has proved to be very competitive compared with the standard codes for non-orthogonal orbital optimization (as described, for example, in Ref. [21]), and for systems with molecular symmetry it should be clearly superior. The effects of point group symmetry will be a reduction in the effort associated with multiplication by **H**, a reduction in N_{var} (and hence in the total effort), and possibly a reduction in the effort associated with applying **T**(**O**), depending on the structure of **O**.

The previously reported applications of this method [3, 4] have been concerned only with VB representations of CASSCF wavefunctions, but should amply demonstrate the viability of the scheme we have described here. We except to publish further applications in the very near future.

Appendix: consequences of point group symmetry

Provided that the full CI vector is symmetry-pure, it may be assumed that the orthogonal active orbitals belong to irreducible representations of the molecular point group. Particularly if the symmetry group is Abelian, it is then a trivial matter to determine the symmetry property of each individual structure, and thus to restrict the CI space according to the required overall symmetry. Since each irreducible representation is one-dimensional for an Abelian group, the irreducible representation of the total structure (or determinant) may be identified trivially from the simple product of characters of each MO. If a graphical indexing scheme is employed [22], it is particularly straightforward to ensure that only structures of correct symmetry are included.

The case is far more complicated for VB wavefunctions, as it is usually not possible (in advance of seeing the results) to place any restrictions on the individual orbitals. A very thorough analysis of the conditions on the single-configuration spin-coupled wavefunction has been carried out by Gerratt for the case where the total wavefunction belongs to a non-degenerate irreducible representation [23]. The corresponding case of a degenerate irreducible representation has been considered in Ref. [24]. It would be straightforward to extend these considerations to a wavefunction consisting of more than one configuration. Summarizing the results, we note that, with the exception of special cases, the point group operations must induce simple permutations of the orbital set. This leads to a corresponding permutational symmetry of the spin function, so that the symmetry condition may be used to restrict the number of spin functions used. Short of explicitly creating symmety-adapted structures, there is however no straightforward way of reducing the total CI space.

For the purpose of optimizing CASSCF wavefunctions, it is very convenient to assume a particular distribution of the active MOs among the irreducible representations. This is equally convenient for the optimization of VB wavefunctions, and we have therefore adopted this approach. This will be particularly important for the optimization of the external variables, but it has the added advantage that it is very simple to ensure the correct symmetry of the total VB wavefunction. A symmetry operation will induce a change of phase in certain orbital coefficients, as determined by the linear characters, and it is then trivial to establish if this corresponds to an orbital permutation.

Inherently, symmetry will not in general reduce the effort associated with applying **T**(**O**). This statement is analogous to the observation that intermediate states of all symmetries are required in the scheme of Siegbahn [10]. However, symmetry considerations may affect the computational effort indirectly. This would be the case, if, for example, **O** possessed partial orthogonality such that a subset of λ parameters in Eq. (6) became zero.

References

1. For a review of the CASSCF method see, for example: Roos BO (1987) Adv Chem Phys 69:399
2. For a review of the spin-coupled method see, for example: Cooper DL, Gerratt J, Raimondi M (1987) Adv Chem Phys 69:319
3. Thorsteinsson T (1995) Ph.D. Thesis, University of Liverpool, UK
4. Thorsteinsson T, Cooper DL, Gerratt J, Karadakov PB, Raimondi M (1996) Theor Chim Acta 93:343
5. Malmqvist PÅ (1986) Int J Quant Chem 30:479
6. Malmqvist PÅ, Roos BO (1989) Chem Phys Lett 155:189
7. See for example: Press WH, Teukolsky SA, Vetterling WT, Flannery BT (1986) Numerical recipes, 1st ed., Cambridge University Press, New York

8. Moshinsky M, Seligman TH (1971) Ann Phys 66:311
9. MOLPRO is a package of *ab initio* programs written by Werner H-J and Knowles PJ, with contributions from Almlöf J, Amos R, Elbert S, Hampel K, Meyer W, Peterson K, Pitzer R and Stone AJ
10. Siegbahn PEM (1984) Chem Phys Lett 109:417
11. Knowles PJ, Handy NC (1984) Chem Phys Lett 111:315
12. Paldus J (1976) in: Eyring H and Henderson DG (eds) Theoretical chemistry: advances and perspectives, Vol. 2. Academic Press, New York
13. Murphy RB, Messmer RP (1993) J Chem Phys 98:7958
14. Olsen J, Roos BO, Jørgensen P, Jensen HJAa (1988) J Chem Phys 89:2185
15. Zarrabian S, Sarma CR, Paldus J (1989) Chem Phys Lett 155:183
16. Harrison RJ, Zarrabian S (1989) Chem Phys Lett 158:393
17. Paldus J (1974) J Chem Phys 61:5321
18. Werner H-J, Knowles PJ (1985) J Chem Phys 82:5053
19. Levy B (1969) Chem Phys Lett 4:17
20. Thouless DJ (1961) The quantum mechanics of many-body systems. Academic Press, New York
21. Cooper DL, Gerratt J, Raimondi M, Sironi M, Thorsteinsson T (1993) Theor Chim Acta 85:261
22. Duch W (1986) GRMS or graphical representation of model spaces. Springer, Berlin
23. Gerratt J (1971) Adv Atom Mol Phys 7:141
24. Wright SC, Cooper DL, Gerratt J, Raimondi M (1992) J Phys Chem 96:7943