Novel orthogonalization and biorthogonalization algorithms

Towards multistate multiconfiguration perturbation theory

Zsuzsanna Tóth · Péter R. Nagy · Péter Jeszenszki · Ágnes Szabados

Abstract Orthogonalization with the prerequisite of keeping several vectors fixed is examined. Explicit formulae are derived both for orthogonal and biorthogonal vector sets. Calculation of the inverse or square root of the entire overlap matrix is eliminated, allowing computational time reduction. In this

Tóth Zsuzsanna - Péter Jeszenszki - Ágnes Szabados

Laboratory of Theoretical Chemistry

Eötvös University,

H-1518 Budapest, POB 32, Hungary

Tel.: +36-1-3722500

Fax: +36-1-3722909

E-mail: szabados@chem.elte.hu

Péter R. Nagy

MTA-BME Lendület Quantum Chemistry Research Group

Department of Physical Chemistry and Materials Science,

Budapest University of Technology and Economics,

H-1521 Budapest, POB 91, Hungary

special situation it is found sufficient to evaluate functions of matrices of the dimension matching the number of fixed vectors.

The (bi)orthogonal sets find direct application in extending multiconfigurational perturbation theory to deal with multiple reference vectors.

Keywords overlap · orthogonalisation · biorthogonal sets · multiconfiguration perturbation theory · multistate theory

1 Introduction

There are two different, equivalent approaches for treating nonorthogonality of a nonredundant vector set in a linear algebraic problem. One way is creating biorthogonal vectors to the overlapping set, the other more common way is orthogonalising the basis set. Let us assume that $\left\{\mathbf{c}^i\right\}_{i=1}^N$ is a nonredundant, nonorthogonal set of N-dimensional vectors. Arranging vectors $\mathbf{c}^i$ as columns in matrix $\mathbf{C}$, the overlap matrix

$$\mathbf{S} = \mathbf{C}^\dagger \mathbf{C}$$

is a positive definite, nonunit matrix. Orthonormalization implies a linear transformation $\mathbf{B} = \mathbf{C}\mathbf{L}_o$, obeying

$$\mathbf{B}^\dagger \mathbf{B} = \mathbf{L}_o^\dagger \mathbf{S} \mathbf{L}_o = \mathbf{I} \ . \tag{1}$$

Vector set $\left\{\widetilde{\mathbf{c}}^i\right\}_{i=1}^N$, biorthonormal to $\left\{\mathbf{c}^i\right\}_{i=1}^N$ is obtained by the transformation $\widetilde{\mathbf{C}} = \mathbf{C}\mathbf{L}_{bo}$, fulfilling

$$\widetilde{\mathbf{C}}^\dagger \mathbf{C} = \mathbf{L}_{bo}^\dagger \mathbf{S} = \mathbf{I} \ , \tag{2}$$

where $\widetilde{\mathbf{C}}$ contains vectors $\widetilde{\mathbf{c}}^i$ in columns. In contrast to orthonormalization, the task of biorthonormal or reciprocal set construction has the uniquely defined solution

$$\mathbf{L}_{bo} = \mathbf{S}^{-1}$$

as a direct consequence of Eq.(2).

Of the possible orthonormalization procedures, Löwdin's symmetric scheme[1, 2], operating with $\mathbf{L}_o = \mathbf{S}^{-1/2}$ is widely exploited in quantum chemistry. Gram-Schmidt orthogonalization is a popular, less costly alternative that lacks the symmetry conservation and resemblance[2–4] properties of Löwdin's symmetric treatment. Gram-Schmidt orthogonalization is known to depend on the ordering of vectors $\mathbf{c}^i$, which is not necessarily a shortcoming. It is a deliberate advantage e.g. if one, selected vector is meant to be fixed.

This situation may be met when aiming to describe dynamic electron correlation starting from a single, multideterminantal reference vector. Correction schemes based on perturbation theory (PT) have been applying successive Gram-Schmidt orthogonalization in such circumstances[5–7], occasionally combined with Löwdin's symmetrical[8] or canonical procedure[9,10]. Biorthogonal treatment of the overlap when correcting a single, multideterminantal reference has been advocated and investigated extensively in the laboratory of Péter Surján[11–13]. The framework termed multiconfigurational PT (MCPT) collects several approaches differing in the treatment of overlap and choice for the zero-order Hamiltonian (see Ref.[14] for an elaboration on the perturba-

tive partitioning). Focusing on the handling of overlap, options exploited so far, given a single reference vector are

1. orthogonal projection to the reference (first step of a Gram-Schmidt procedure)

    (a) biorthonormal set construction in the $(N-1)$-dimensional space

    (b) Löwdin-orthogonalization in the $(N-1)$-dimensional space

2. omit orthogonal projection to the reference

    (a) biorthonormal set construction in the N-dimensional space

A notable feature of all three strategies is that the (bi)orthonormal set can be given explicitly, eliminating the need of numerical overlap treatment. Option 2b is lucidly missing from the above list just for the reason that no close form of the underlying inverse square root of the overlap could be constructed. Notation 2a is hence somewhat superfluous, which we however keep to stress analogy with 1a. The procedure of option 1b is equivalent to the Jacobi rotations' inspired orthogonalisation, designed by István Mayer[15,16]. For applications of the above (bi)orthogonal schemes in PT strategies alternative to MCPT see Refs.[17,18].

In the present work we extend the above overlap treatments for the case of multiple reference vectors. This involves derivation of explicit formulae for the (bi)orthonormal sets in cases analogous to 1a, 1b and 2a above, with reference vectors, $m > 1$ in number. Construction of $\mathbf{S}^{-1}$ and $\mathbf{S}^{-1/2}$ in case 1 for the $(N-m) \times (N-m)$ overlap matrix and $\mathbf{S}^{-1}$ in case 2 for the $N \times N$ overlap matrix is presented in Section 2. These results facilitate a multistate

extension of the MCPT framework. The pertinent formulae are given in Section 3, followed by an illustrative numerical study in Section 4.

2 (Bi)orthogonal vector sets

Let $\{\mathbf{e}^i\}_{i=1}^N$ be the set of unit vectors and let us replace the first m vectors with the orthonormal set[1] of vectors $\{\mathbf{c}^i\}_{i=1}^m$. The new N-dimensional set, $\{\mathbf{c}^i\}_{i=1}^m \cup \{\mathbf{e}^i\}_{i=m+1}^N$ is not orthogonal as its subsets $\{\mathbf{c}^i\}_{i=1}^m$ and $\{\mathbf{e}^i\}_{i=m+1}^N$ overlap.

Let us assume that coefficients C_{ji} of the expansion

$$\mathbf{c}^i = \sum_{j=1}^N \mathbf{e}^j\, C_{ji}$$

are arranged in matrix $\mathbf{C}$, of dimension $N \times m$. We now introduce notation $\mathbf{C}_1$ for the upper $m \times m$ block of $\mathbf{C}$, and $\mathbf{C}_2$ for the lower $(N-m) \times m$ block. This allows to write

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_1 \\ \mathbf{C}_2 \end{pmatrix} . \tag{3}$$

Orthonormality of vectors $\mathbf{c}^i$ is reflected by

$$\mathbf{I}_m = \mathbf{C}_1^\dagger \mathbf{C}_1 \;+\; \mathbf{C}_2^\dagger \mathbf{C}_2 \;=\; \mathbf{A} + \; \mathbf{C}_2^\dagger \mathbf{C}_2 \;, \tag{4}$$

where $\mathbf{I}_m$ denotes the m-dimensional unit matrix and shorthand $\mathbf{A}$ is introduced for subsequent use. We note here, that the union of sets $\{\mathbf{c}^i\}_{i=1}^m$ and $\{\mathbf{e}^i\}_{i=m+1}^N$ being N-dimensional relies on the tacit assumption that matrix $\mathbf{A}$ is positive definite.

---

[1] Orthonormality of vectors $\mathbf{c}^i$ is assumed since orthonormalising m vectors is relatively cheap for $m << N$.

To extend option 1, projector $\mathbf{P}$ corresponding to the vector set $\{\mathbf{c^i}\}_{i=1}^{m}$ is formulated as

$$\mathbf{P} \;=\; \mathbf{C}\mathbf{C}^{\dagger} \;=\; \left(\begin{array}{c|c} \mathbf{C}_1\mathbf{C}_1^{\dagger} & \mathbf{C}_1\mathbf{C}_2^{\dagger} \\ \hline \mathbf{C}_2\mathbf{C}_1^{\dagger} & \mathbf{C}_2\mathbf{C}_2^{\dagger} \end{array} \right) \;.$$

As the next step, vectors $\{\mathbf{e^i}\}_{i=m+1}^{N}$, arranged as the last $N-m$ columns of unit matrix $\mathbf{I}_N$, are projected orthogonal to $\{\mathbf{c^i}\}_{i=1}^{m}$ to generate the set $\{\mathbf{e}^{i\prime}\}_{i=m+1}^{N}$. Denoting the corresponding matrix $\mathbf{D}'$, we get

$$\mathbf{D}' \;=\; (\mathbf{I}_N - \mathbf{P}) \left(\begin{array}{c} 0 \\ \mathbf{I}_{N-m} \end{array} \right) \;=\; \left(\begin{array}{c} -\,\mathbf{C}_1\mathbf{C}_2^{\dagger} \\ \mathbf{I}_{N-m} - \mathbf{C}_2\mathbf{C}_2^{\dagger} \end{array} \right) \;, \tag{5}$$

with obvious notation for $\mathbf{I}_{N-m}$. Overlap of vectors $\mathbf{e}^{i\prime}$ can be expressed as

$$\mathbf{D}'^{\dagger}\mathbf{D}' = I_{N-m} \;-\; \mathbf{C}_2\mathbf{C}_2^{\dagger} \;, \tag{6}$$

using Eqs.(4) and (5). Matrix $\mathbf{D}'^{\dagger}\mathbf{D}'$ represents the nontrivial part of the $N \times N$ overlap matrix

$$\mathbf{S} = \left(\begin{array}{c|c} \mathbf{I}_m & \\ \hline & \mathbf{D}'^{\dagger}\mathbf{D}' \end{array} \right) \tag{7}$$

of the set $\{\mathbf{c^i}\}_{i=1}^{m} \cup \{\mathbf{e}^{i\prime}\}_{i=m+1}^{N}$.

In the following paragraphs the inverse and inverse square root of the overlap matrix of Eq.(7) is constructed. Clearly, it suffices to focus on $\mathbf{D}'^{\dagger}\mathbf{D}'$ of Eq.(6). We approach the problem in a general manner, by expressing any analytic function, $f :\; (0,1] \to \mathbb{R}$ of matrix $\mathbf{S}$. Matrix function $f(\mathbf{S})$ is defined via Taylor-expansion, written as

$$f\left(\mathbf{I}_{N-m} - \mathbf{C}_2\mathbf{C}_2^{\dagger} \right) = \sum_{n=0}^{\infty} d_n \left(\mathbf{C}_2\mathbf{C}_2^{\dagger} \right)^{n} \;, \tag{8}$$

with $d_n = (-1)^n f^{(n)}(1)/n!$. Since $||\mathbf{C}_2\mathbf{C}_2^\dagger|| < 1$ (taking e.g. the 2-norm of the matrix), the Taylor-series is convergent if the coefficients d_n are bounded. Explicit form of the Taylor-coefficients d_n is not necessary for further derivation, all we need to know is that they are indeed bounded for the inverse and inverse square root function.

We now recognize the following recursion for the powers of $\mathbf{C}_2\mathbf{C}_2^\dagger$

$$\left(\mathbf{C}_2\mathbf{C}_2^\dagger\right)^n = \mathbf{C}_2\left(\mathbf{I}_m - \mathbf{A}\right)^{n-1}\mathbf{C}_2^\dagger \tag{9}$$

that is easy to prove by induction. While the $n = 1$ case is trivial, case $n = 2$ is obtained as

$$\left(\mathbf{C}_2\mathbf{C}_2^\dagger\right)^2 = \mathbf{C}_2\mathbf{C}_2^\dagger\mathbf{C}_2\mathbf{C}_2^\dagger = \mathbf{C}_2(\mathbf{I}_m - \mathbf{A})\mathbf{C}_2^\dagger ,$$

an obvious consequence of Eq.(4) . To take the induction step, let us suppose that statement (9) holds for $n - 1$ and examine the case $n > 2$. We find

$$\begin{aligned}
\left(\mathbf{C}_2\mathbf{C}_2^\dagger\right)^n &= \mathbf{C}_2(\mathbf{I}_m - \mathbf{A})^{n-2}\mathbf{C}_2^\dagger\mathbf{C}_2\mathbf{C}_2^\dagger \\
&= \mathbf{C}_2(\mathbf{I}_m - \mathbf{A})^{n-2}(\mathbf{I}_m - \mathbf{A})\mathbf{C}_2^\dagger , \\
&= \mathbf{C}_2(\mathbf{I}_m - \mathbf{A})^{n-1}\mathbf{C}_2^\dagger
\end{aligned}$$

which completes the proof.

Let us now substitute Eq.(9) into the Taylor-expansion of Eq.(8) and utilize $d_0 = f(1)$. We obtain

$$
\begin{aligned}
f &\left(\mathbf{I}_{N-m} - \mathbf{C}_2 \mathbf{C}_2^\dagger\right) \\
&= f(1)\mathbf{I}_{N-m} + \mathbf{C}_2 \left[\sum_{n=1}^{\infty} d_n (\mathbf{I}_m - \mathbf{A})^n\right] (\mathbf{I}_m - \mathbf{A})^{-1} \mathbf{C}_2^\dagger \\
&= f(1)\mathbf{I}_{N-m} \\
&\quad + \mathbf{C}_2 \left\{\left[\sum_{n=0}^{\infty} d_n (\mathbf{I}_m - \mathbf{A})^n\right] - f(1)\mathbf{I}_m\right\} (\mathbf{I}_m - \mathbf{A})^{-1} \mathbf{C}_2^\dagger \\
&= f(1)\mathbf{I}_{N-m} + \mathbf{C}_2 \left(f(\mathbf{A}) - f(1)\mathbf{I}_m\right)(\mathbf{I}_m - \mathbf{A})^{-1} \mathbf{C}_2^\dagger \; .
\end{aligned}
\tag{10}
$$

In the case $\mathbf{I}_m = \mathbf{A}$ the second term of the expression above is zero. The role of the analytical treatment can be clearly pointed out at this step. Instead of evaluating function f of an $(N-m) \times (N-m)$ matrix, it is sufficient to calculate the inverse and the same function f of $m \times m$ matrices. As long as $m << N$, much can be gained in computational time, the eventual speedup depending on the structure of $\mathbf{A}$ and the nature of f. As $m \to N$, the computational advantage evidently disappears.

2.1 Extended option 1a

Let us now work out the general formula of Eq.(10) for our two functions of interest. Starting with the inverse function we obtain:

$$
\left(\mathbf{I}_{N-m} - \mathbf{C}_2 \mathbf{C}_2^\dagger\right)^{-1} =
$$

$$
\mathbf{I}_{N-m} + \mathbf{C}_2 \underbrace{\left(\mathbf{A}^{-1} - \mathbf{I}_m\right)(\mathbf{I}_m - \mathbf{A})^{-1}}_{\mathbf{A}^{-1}} \mathbf{C}_2^\dagger
$$

The inverse of the overlap matrix of Eq.(7) therefore reads

$$
\mathbf{S}^{-1} = \begin{pmatrix} \mathbf{I}_m & \\ \hline & \mathbf{I}_{N-m} + \mathbf{C}_2 \left(\mathbf{C}_1^\dagger \mathbf{C}_1 \right)^{-1} \mathbf{C}_2^\dagger \end{pmatrix} .
$$

Evaluating the biorthogonal counterpart of vectors $\mathbf{e}^{i\prime}$ one obtains

$$
\widetilde{\mathbf{D}}' \;=\; \mathbf{D}' \left(\mathbf{I}_{N-m} + \mathbf{C}_2 \mathbf{A}^{-1} \mathbf{C}_2^\dagger \right) \;=\; \begin{pmatrix} -\mathbf{C}_1 \mathbf{A}^{-1} \mathbf{C}_2^\dagger \\ \\ \mathbf{I}_{N-m} \end{pmatrix} ,
$$

having utilized Eqs.(4) and (5). Noting, that $\mathbf{C}_1 \mathbf{A}^{-1} = \mathbf{C}_1^{\dagger -1}$, the special case of $m = 1$ is recovered as

$$
\widetilde{\mathbf{e}}^{i\prime} = \mathbf{e}^i - \frac{c_i^*}{c_1^*} \mathbf{e}^1 \; , \quad i = 2, \ldots, N \; , \tag{11}
$$

c_i denoting the components of the single column vector $\mathbf{C}$ of Eq.(3). While the biorthonogal vectors of Eq.(11) were introduced in Ref.[11], Eq.(3) of Ref.[12] allows for a more transparent comparison.

2.2 Extended option 1b

Let us step now to the inverse square root of Eq.(7). Based on the general result of Eq.(10) we obtain

$$
\left(\mathbf{I}_{N-m} - \mathbf{C}_2 \mathbf{C}_2^\dagger \right)^{-1/2} = \mathbf{I}_{N-m} + \mathbf{C}_2 \underbrace{\left(\mathbf{A}^{-1/2} - \mathbf{I}_m \right) \left(\mathbf{I}_m - \mathbf{A} \right)^{-1}}_{\left[\left(\mathbf{I}_m + \mathbf{A}^{1/2} \right) \mathbf{A}^{1/2} \right]^{-1}} \mathbf{C}_2^\dagger \; . \tag{12}
$$

The $-1/2$ power of the overlap matrix of Eq.(7) hence becomes

$$
\mathbf{S}^{-1/2} = \left(\begin{array}{c|c} \mathbf{I}_m & \\ \hline & \mathbf{I}_{N-m} + \mathbf{C}_2 \left[\left(\mathbf{I}_m + \left(\mathbf{C}_1^\dagger \mathbf{C}_1 \right)^{1/2} \right) \left(\mathbf{C}_1^\dagger \mathbf{C}_1 \right)^{1/2} \right]^{-1} \mathbf{C}_2^\dagger \end{array} \right) .
$$

The special case of $m = 1$, derived in Ref.[16] is obtained as

$$
(S^{-\frac{1}{2}})_{ij} = \delta_{ij} + \frac{c_i c_j^*}{|c_1|(1 + |c_1|)}, \quad i,j > 1 ,
$$

utilizing that $\mathbf{C}$ of Eq.(3) is composed of a single column.

Matrix $\mathbf{S}^{-1/2}$ above facilitates to construct the Löwdin-orthogonalized counterpart of vectors $\mathbf{e}^{i\prime}$ as

$$
\begin{aligned}
\mathbf{D}'^L &= \mathbf{D}' \left(\mathbf{I}_{N-m} + \mathbf{C}_2 \left(\mathbf{A}^{-1/2} - \mathbf{I}_m \right) \left(\mathbf{I}_m - \mathbf{A} \right)^{-1} \mathbf{C}_2^\dagger \right) \\
&= \left(\begin{array}{c} -\mathbf{C}_1 \mathbf{A}^{-1/2} \mathbf{C}_2^\dagger \\ \mathbf{I}_{N-m} - \mathbf{C}_2 \left(\mathbf{I}_m + \mathbf{A}^{1/2} \right)^{-1} \mathbf{C}_2^\dagger \end{array} \right) ,
\end{aligned}
$$

having made use of Eqs.(4), (5) and (12).

2.3 Extended option 2a

Let us finally consider the almost trivial case of reciprocal set construction to the set $\{\mathbf{c^i}\}_{i=1}^m \cup \{\mathbf{e^i}\}_{i=m+1}^N$. Note, that the orthogonal projection of Eq.(5) is

now omitted. The $N \times N$ overlap matrix of this set reads

$$\mathbf{S} = \left(\begin{array}{c|c} \mathbf{I}_m & \mathbf{C}_2^\dagger \\ \hline \\ \mathbf{C}_2 & \mathbf{I}_{N-m} \\ \\ \end{array} \right) .$$

Partitioning $\mathbf{S}^{-1}$ into the same structure as $\mathbf{S}$ and solving for the individual

blocks, the inverse overlap is readily found to be

$$\mathbf{S}^{-1} = \left(\begin{array}{c|c} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}_2^\dagger \\ \hline \\ -\mathbf{C}_2\mathbf{A}^{-1} & \mathbf{I}_{N-m} + \mathbf{C}_2\mathbf{A}^{-1}\mathbf{C}_2^\dagger \\ \\ \end{array} \right) . \tag{13}$$

The $N \times m$ matrix $\widetilde{\mathbf{C}}$ and $N \times (N-m)$ matrix $\widetilde{\mathbf{D}}$ collecting reciprocal column

vectors are obtained as

$$\left(\widetilde{\mathbf{C}} \middle| \widetilde{\mathbf{D}} \right) = \left(\begin{array}{c|c} \mathbf{C}_1 & 0 \\ \hline \mathbf{C}_2 & \mathbf{I}_{N-m} \end{array} \right) \mathbf{S}^{-1} ,$$

their explicit form reading as

$$\widetilde{\mathbf{C}} = \left(\begin{array}{c} \mathbf{C}_1\mathbf{A}^{-1} \\ \\ 0 \end{array} \right) ,$$

and

$$\widetilde{\mathbf{D}} = \left(\begin{array}{c} -\mathbf{C}_1\mathbf{A}^{-1}\mathbf{C}_2^\dagger \\ \\ \mathbf{I}_{N-m} \end{array} \right) .$$

Similarly to Eq.(10), the most demanding computational task (inversion of $\mathbf{A}$)

is connected to an $m \times m$ matrix, instead of the entire, $N \times N$ overlap matrix.

It is interesting to observe, that the reciprocal vectors $\widetilde{\mathbf{D}}'$ of extended option 1a and $\widetilde{\mathbf{D}}$ of extended option 2a match explicitly. In the special case of $m = 1$, the first reciprocal vector reads $\widetilde{\mathbf{e}}^1 = c_1^{*-1}\,\mathbf{e}^1$, while $\widetilde{\mathbf{e}}^i$ are the same as $\widetilde{\mathbf{e}}^{i\prime}$ of Eq.(11) for $i > 1$, in accordance with [12].

3 Multistate extension of MCPT

Electronic structure description often reaches its target in two successive steps. A first, qualitative approximation is corrected in a second shot to incorporate the so-called dynamic correlation. Situations where a single vector – even if multiconfigurational in character – does not represent an adequate first approximation, call for two or more reference vectors. Targeting more states of the system at a time is another example where multiple reference vectors are necessary. Many correction schemes have a version designed for such situations, assuming multiple reference states that form a so-called model space (MS). Equations of Rayleigh-Schrödinger PT[19,20] for instance can be regarded as a special case of the Bloch-equation[21], corresponding to a one-dimensional model space.

Bloch-equation based multistate PT formulations, termed quasidegenerate PT (QDPT)[22] largely assume an orthonormal set of vectors in the configuration interaction (CI) space, that is partitioned for a model space and its complement. This restricts applicability to model spaces easily separable from the rest, e.g. formed by simple determinants. While determinants facilitate a transparent derivation of many-body QDPT formulae[23,24], identifying the

determinants that need to be included in the model space is not always trivial. Though complete active space (CAS) appears a simple way out, CAS based QDPT is unfortunately prone to the so-called intruder problem, especially for large active spaces.

The idea of picking multiconfigurational vectors to span the model space is appealing for two reasons. On one hand the dimension of the space is reduced as compared to the case of using determinants. On the other hand it may have a beneficial effect on intruder sensitivity, due to the internally coupled nature of reference vectors. Such an approach is however hard to find for the simple reason, that the orthogonal complement of multiple multiconfiguration vectors is not easy to construct. It is at this point, where the overlap treatments elaborated in Section 2 can be relied upon. Detailed derivation of the multistate extension of the MCPT framework is out of the scope of the present report. In what follows we confine ourselves to the key formulae necessary for the illustrative application of Section 4.

Multiconfigurational reference functions, $m > 1$ in number, constitute the starting point of our approximation. Vectors $\{\mathbf{c}^i\}_{i=1}^m$ of Section 2 are associated with these reference functions, C_{ji} denoting the jth component in the determinantal expansion of reference $\mathbf{c}^i$. In accordance with the generally applicable philosophy of MCPT, we do not assume any special structure of the reference functions, apart from being orthonormal. Unit vectors $\{\mathbf{e}^i\}_{i=1}^N$ of Section 2 now represent determinants spanning the CI space. The first m among

determinants, $\mathbf{e}^i$ are selected based on their projection to the model space. In particular, rows j for which $\sum_i C_{ji}^2$ are the largest, constitute $\mathbf{C}_1$.

The central quantity of PT approaches, the zero-order Hamiltonian is formulated in MCPT via its spectral resolution. Extended overlap treatment options 1a and 2a both imply a nonsymmetrical operator, 1b works with a symmetrical expression. The zero-order Hamiltonian of option 1a, introduced in projected MCPT (pMCPT)[11,13], can be extended as

$$\mathbf{H}^{(0)}_{\text{MS-pMCPT}} = \mathbf{C}\,\mathbf{E}^{(0)}_P\,\mathbf{C}^\dagger \;+\; \mathbf{D}'\,\mathbf{E}^{(0)}_{P\perp}\,\widetilde{\mathbf{D}}'^{\,\dagger} \;,$$

MS standing for 'multistate'. Overlap treatment option 1b, applied e.g. in Ref.[16], can be extended for multiple reference vectors as

$$\mathbf{H}^{(0)}_{\text{MS-pLMCPT}} \;=\; \mathbf{C}\,\mathbf{E}^{(0)}_P\,\mathbf{C}^\dagger \;+\; \mathbf{D}'^L\,\mathbf{E}^{(0)}_{P\perp}\,\mathbf{D}'^{L\,\dagger} \;.$$

Finally, the complete biorthogonal treatment of option 2a, termed originally SC2-MCPT or unprojected MCPT (uMCPT)[12,13] can be extended to the multistate case as

$$\mathbf{H}^{(0)}_{\text{MS-uMCPT}} = \mathbf{C}\,\mathbf{E}^{(0)}_P\,\widetilde{\mathbf{C}}^\dagger \;+\; \mathbf{D}\,\mathbf{E}^{(0)}_{P\perp}\,\widetilde{\mathbf{D}}^\dagger \;.$$

Diagonal matrices $\mathbf{E}^{(0)}_P$ and $\mathbf{E}^{(0)}_{P\perp}$ above contain zero-order energies corresponding to the model space and its complement, respectively. Zero-order energies in the Epstein-Nesbet (EN) partitioning[25,26] are collected in Table 1 for completeness.

Energy corrections are calculated in MS theories as eigenvalues of an effective Hamiltonian, obtained relying on the Bloch-equation and the zero-order

operator. Matrix elements for the MS-pLMCPT variant at order 2 read as

$$\left(\mathbf{H}_{\text{MS-pLMCPT}}^{\text{eff}(2)}\right)_{ij} = \delta_{ij}\mathbf{E}_{P,i}^{(0)} \;-\; \sum_{k=m+1}^{N} \frac{\langle\mathbf{c}^i|H|\mathbf{e}^{k\prime L}\rangle\langle\mathbf{e}^{k\prime L}|H|\mathbf{c}^j\rangle}{\mathbf{E}_{P\perp,k}^{(0)} - \mathbf{E}_{P,j}^{(0)}} \;. \quad (14)$$

Corrected energies, $E^{(2)}$ and reference functions, $\mathbf{c}^{i(2)}$ arise as

$$\sum_j \mathbf{H}_{ij}^{\text{eff}(2)} d_{jk}^{(2)} = E_k^{(2)} d_{ik}^{(2)} \;,$$

and

$$\mathbf{c}^{i(2)} = \sum_j \mathbf{c}^j d_{ji}^{(2)} \;,$$

respectively.

Closing this Section let us note, that the MS-MCPT variants would merit a detailed formal and numerical examination respecting e.g. size-consistency, intruder sensitivity or dependence on the zero-order eigenvalues. We intentionally withdraw from such a study presently, as MS-MCPT merely serves an illustration purpose here.

4 Numerical illustration

Choosing suitable reference function(s) is a persisting challenge for multireference theories. Geminal based approaches, continuously cultivated by Péter Surján are promising in this respect[27–29] as they are more economic than CAS, still they often reflect the multiconfigurational nature of the target function correctly. Obviously, geminal wavefunctions have their own shortcomings. We focus here on one of these, the process of switching between two Lewis-structures of the same molecule. A simple case study is provided by the rectangular to square distortion of the H_4 system. Antysimmetrized Product of

Strongly Orthogonal Geminals (APSG) produces a cusp on the energy curve of this system at square geometry, as reflected by Fig. 1. The problem is connected to the fact, that geminals (assigned to bonds) are reordered at the switching point, thereby capturing the correct, dominant Lewis-structure at both rectangular arrangements.

There exist solutions to this problem[30–32], APSG based PT is hovever notably not among them, since a qualitative defect of the reference can not be cured by PT. This is reflected by the MCPT curve in Fig. 1. If wishing to proceed by PT, it appears straightforward to construct two reference functions, corresponding to either of the Lewis-structures and follow a MS-MCPT strategy with two-dimensional model space. The energy profile obtained by such a procedure is shown for the MS-pLMCPT variant in Fig. 1. Multistate theory apparently results a smooth energy curve with a correct, zero derivative at $90\,^o$. One can also observe in Fig. 1 a characteristic overshooting of EN partitioning at order 2.

Due to the small system size, the example of Fig. 1 can not illustrate the gain in computational time, brought about by the results of Section 2. To give an impression in this line, let us consider the ground state of p-benzyne. Description by a geminal based MS-MCPT approach necessitates three reference vectors, corresponding to the three dominant Lewis-structures. Value for N of Section 2 is given by the length of the determinantal expansion of the reference vectors. Depending on the geminal scheme chosen, N may range from a couple of hundreds (in case of a minimalistic APSG) to astronomical dimensions. The

all pair coupled cluster doubles wavefunction, a geminal type function[33] e.g. includes deteminants on the order of 10^{16}, in a double zeta, polarized basis set, cores assumed frozen. Accordingly, analytic handling of overlap allows for dealing with 3×3 matrices instead of the numerical treatment of the $N \times N$ overlap matrix, with the above values for N.

References

1. P.O. Löwdin, J. Chem. Phys. **18**, 365 (1950)

2. P.O. Löwdin, Adv. Quantum Chem. **5**, 185 (1970)

3. I. Mayer, Int. J. Quantum Chem. **90**(1), 63 (2002). DOI 10.1002/qua.981

4. I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry* (Kluwer, New York, 2003)

5. K. Wolinski, H. Sellers, P. Pulay, Chem. Phys. Letters **140**, 225 (1987)

6. K. Wolinski, P. Pulay, J. Chem. Phys. **90**, 3647 (1989)

7. H.J.J. van Dam, J.H. van Lenthe, Mol. Phys. **93**, 431 (1998)

8. H.J. Werner, Mol. Phys **89**, 645 (1996)

9. K. Andersson, P.Å. Malmqvist, B.O. Roos, A.J. Sadlej, K. Wolinski, J. Phys. Chem. **94**, 5483 (1990)

10. K. Andersson, P.Å. Malmqvist, B.O. Roos, J. Chem. Phys. **96**, 1218 (1992)

11. Z. Rolik, Á. Szabados, P.R. Surján, J. Chem. Phys. **119**, 1922 (2003)

12. Á. Szabados, Z. Rolik, G. Tóth, P.R. Surján, J. Chem. Phys. **122**, 114104 (2005)

13. M. Kobayashi and Á. Szabados and H. Nakai and P. R. Surján, J. Chem. Theory. Comput. **6**, 2024 (2010)

14. P. Surján, Z. Rolik, Á. Szabados, D. Kőhalmi, Ann. Phys. (Leipzig) **13**, 223 (2004)

15. I. Mayer, Theor. Chim. Acta **104**, 163 (2000)

16. P. Nagy, P. Surján, Á. Szabados, Theoretical Chemistry Accounts (Theoretica Chimica Acta) **131**, 1109 (2012). URL http://dx.doi.org/10.1007/s00214-012-1109-y

17. P.A. Limacher, P.W. Ayers, P.A. Johnson, S. De Baerdemacker, D.V. Neck, P. Bultinck, Phys. Chem. Chem. Phys. **16**, 5061 (2014)

18. P.R. Nagy, Á. Szabados, Int. J. Quantum Chem. **113**, 230 (2012)

19. J.W.S. Lord Rayleigh, *The Theory of Sound*, vol. 1 (Dover, New York, 1976)

20. E. Schrödinger, Ann. Physik **80**, 437 (1926)

21. I. Lindgren, J. Morrison, *Atomic Many-Body Theory* (Springer, Berlin, 1986)

22. I. Shavitt, R.J. Bartlett, *Many-Body Methods in Chemistry and Physics* (Cambridge University Press, Cambridge, 2009)

23. G. Hose, U. Kaldor, J. Phys. B **12**, 3827 (1979)

24. L. Meissner, R.J. Bartlett, J. Chem. Phys. **91**, 4800 (1989)

25. P. Epstein, Phys. Rev. **28**, 695 (1926)

26. R. Nesbet, Proc. Roy. Soc. (London) **A230**, 312 (1955)

27. P.R. Surján, Topics in current chemistry **203**, 63 (1999)

28. Péter R. Surján, Ágnes Szabados, Péter Jeszenszki, and Tamás Zoboki, J. Math. Chem. **50**, 534 (2012)

29. Péter Jeszenszki, Péter R. Nagy, Tamás Zoboki, Ágnes Szabados and Péter R. Surján, Int. J. Quantum Chem. **114**, 1048 (2014)

30. V.A. Rassolov, F. Xu, J. Chem. Phys. **127**, 044104 (2007)

31. D.W. Small, M. Head-Gordon, The Journal of Chemical Physics **130**, 084103 (2009)

32. P. Jeszenszki, V. Rassolov, P.R. Surján, Á. Szabados, Molecular Physics **113**(3-4), 249 (2015)

33. P.A. Johnson, P.W. Ayers, P.A. Limacher, S.D. Baerdemacker, D.V. Neck, P. Bultinck, Computational and Theoretical Chemistry **1003**, 101 (2013)

Table 1 Zero-order eigenvalues within the Epstein-Nesbet partitioning in multistate MCPT variants. Subscript P refers to the model space, $P \perp$ stands for the complementary space. See text for further notations.

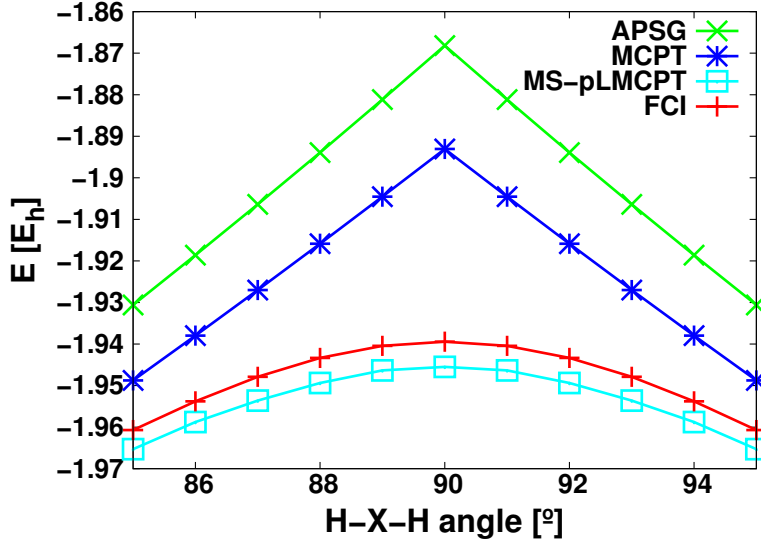|	MS-pMCPT	MS-pLMCPT	MS-uMCPT
$\mathbf{E}_{P,i}^{(0)}$	$\langle \mathbf{c}^i|H|\mathbf{c}^i \rangle$	$\langle \mathbf{c}^i|H|\mathbf{c}^i \rangle$	$\langle \widetilde{\mathbf{c}}^i|H|\mathbf{c}^i \rangle$
$\mathbf{E}_{P\perp,i}^{(0)}$	$\langle \widetilde{\mathbf{e}}^{i\,\prime}|H|\mathbf{e}^{i\,\prime} \rangle$	$\langle \mathbf{e}^{i\,\prime\,L}|H|\mathbf{e}^{i\,\prime\,L} \rangle$	$\langle \widetilde{\mathbf{e}}^i|H|\mathbf{e}^i \rangle$

Fig. 1 Total energy for H_4, in STO-3G basis set. The four hydrogen atoms are confined to a circle with a radius of $\sqrt{2}$ bohr. Angle of two neighboring hydrogen atoms (H) and the center of mass (X) is labeled angle(H-X-H). The APSG wavefunction, involving 2 geminals, with two orbitals assigned to each represents one of the reference states. The other reference function is generated by (i) localising orbitals to atoms within the Arai subspaces; (ii) assigning these orbitals to geminals representing the the longer HH bonds, instead of the shorter; (iii) optimizing geminal coefficients but not the orbitals. The MS-pLMCPT energy is obtained as the the lower lying root of the effective Hamiltonian of Eq.(14). The curve MCPT is obtained in EN partitioning following the APSG based PT strategy of Ref.[11]. Full CI is shown for comparison.