Chemical Physics 356 (2009) 64–75

Contents lists available at ScienceDirect

Chemical Physics

journal homepage: www.elsevier.com/locate/chemphys

A priori identification of configurational deadwood

Laimutis Bytautas, Klaus Ruedenberg *

Department of Chemistry and Ames Laboratory USDOE, Iowa State University, Ames, IA 50011, USA

ARTICLE INFO

Article history:
Received 30 October 2008
Accepted 17 November 2008
Available online 30 November 2008

Keywords:
Ab initio quantum chemistry
Electronic structure
Electron correlation
Configuration interaction
Multi-configurational valence spaces

ABSTRACT

A method is given for the *a priori* assessment of those terms in a configuration interaction expansion that can be deleted if a given error in the energy is tolerated. The truncations are performed independently for the quadruple, quintuple and sextuple excitations on the basis of information derived from the double and triple excitations. The effectiveness of the method is illustrated by application to SDTQ56-CI wavefunctions of the molecules HNO, N_2 and NCCN.

1. Introduction

Accurate descriptions of chemical processes require an accurate accounting of electron correlation. This became already apparent early on when James and Coolidge [1] (and later Kołos and Roothaan [2] and Kołos and Wolniewicz [3]) formulated wavefunctions for the accurate calculation of the potential energy curve and the vibrational spectrum of the hydrogen molecule. Remarkable and successful progress in the further development of methods that include inter-electronic distances in the wavefunctions of heavier molecules has been made in recent years by Kutzelnigg, Klopper and others [4–15].

Nonetheless, the configuration interaction (CI) method [16,17] has so far remained the most widely used approach, notably also because of its variational properties. Unfortunately, it is extremely inefficient in its brute force version because the primitive expansions contain huge numbers of ineffective configurations. A large amount of this deadwood is eliminated by constructing correlated wavefunctions via successive excitations with respect to a dominant reference function and many approaches have been developed to accomplish this in an efficient manner [18–39]. In this context, the ability to select the important configurations *a priori* is essential. Some approaches [38] treat all excitation levels on the same footing and select important determinants by using a threshold for the estimated contribution to the FCI energy in the next iteration step of the Davidson procedure [40]. Other methods use arguments based on the linked cluster expansion [41–44] and estimate [34,36,45–47] the importance of the determinants for the higher excitation levels by using the information on the determinants representing the lower excitation levels. The most successful approach so far has been the coupled cluster method [48–51], which has greatly extended the range of molecules that can now be treated with the inclusion of correlation.

As yet, the effective application of coupled cluster methods has been limited to systems whose wavefunctions are dominated by a single Hartree–Fock reference determinant. Its extension to cases where the dominant reference function is multi-configurational has so far proved elusive although progress is continuing [52–55]. On the other hand, the CEEIS method [56–63] has made it possible to generate accurate approximations to full CI energies by successive excitations from multi-configurational reference functions.

In any event, it has become apparent that multi-configurational reference functions are essential along reaction paths and one problem is how to define or choose them. A unique prescription is to obtain them as the MCSCF wavefunction in a full valence space or in the configurational space of N electrons in N orbitals [64–66]. We have denoted these as FORS1 and FORS2 models [36,64,67]. However, these configuration spaces are too large to serve as reference spaces and they still contain much deadwood. The significance of well chosen compact reference spaces has been emphasized in the recent developments of multi-reference model chemistries (see, e.g. Refs. [37,68]).

Thus, the problem of eliminating configurational deadwood is also relevant for the construction of multi-configurational reference spaces. One approach to this truncation consists of formulating a reference space on the basis of certain model concepts [37]. The aim of the present study, on the other hand, is to develop a *quantitative* procedure for the separation of the "livewood" from

* Corresponding author. Tel.: +1 515 294 5253; fax: +1 515 294 4709.
E-mail address: ruedenberg@iastate.edu (K. Ruedenberg).

0301-0104/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemphys.2008.11.021

the "deadwood" in a given configuration space. Moreover, this assessment should be accomplished *before* making the CI calculation rather than *after* it has been made, i.e. *a priori* rather than *a posteriori*. The truncation method is illustrated by applications to three molecules. While it is in principle applicable to any CI expansion constructed by successive excitations, our immediate target is the use for obtaining compact reference functions.

2. Orbital determination

2.1. Strongly and weakly occupied approximate FORS molecular orbitals

The number N_o of occupied valence SCF orbitals in a molecule is typically less than the total number N_{mb} of orbitals in the minimal valence basis sets of all atoms. The full valence MCSCF wavefunction is the optimal expansion in terms of all configurations that can be generated from N_{mb} molecular orbitals. Closely related is the full MCSCF wavefunction of all configurations that can be generated from N_e orbitals, where N_e is the number of valence electrons, i.e. each occupied valence orbital has a correlating orbital, as first postulated by Foster and Boys [69] and also presumed in perfect pairing models [70,71]. We have called these two types of full spaces FORS1 and FORS2 [36,67]. In both, the inner shell remains closed.

For the present purpose, viz. the establishment of truncation criteria, approximate FORS orbitals are adequate. The task is thus the determination of a good approximation to the FORS orbital space without a full valence space MCSCF calculation. The following procedure [36] has proven successful.

First, the SCF wavefunction is determined. Then, the SD-CI wavefunction is calculated in the determinantal space of *all* single and double excitations from the occupied to the virtual SCF orbitals. Then, the natural orbitals [72,73] of this SD-CI wavefunction are found. If the FORS wavefunction is to be based on M molecular orbitals, then the M natural orbitals with the highest occupation numbers are being chosen as first approximations to FORS configuration space generators. The M identified SD-CI natural orbitals are then further separated into two groups, the *strongly* and the *weakly* occupied NOs.

Alternatively, one can also chose the occupied SCF orbitals as the strongly occupied FORS MOs and determine the "virtual natural orbitals" by diagonalizing only that sub-block of the SD-CI first-order density matrix that is spanned by the virtual SCF orbitals [74].

If desired an MCSCF improvement of all M orbitals can be performed after completion of the configurational truncation. The effect of the MCSCF optimization on the computed energies has been discussed in Ref. [36]. A recent study by Abrams and Sherrill [75] has demonstrated the effectiveness of the natural orbitals in CAS-CI active space when used in place of the CASSCF optimized orbitals in the bond breaking process of ethylene.

2.2. Split-localized FORS molecular orbitals

As we have recently demonstrated, the conventional presumption that the configurations generated from natural orbitals always yield the most rapid configurational convergence is in error. We showed [67] that, in fact, the orbitals we called *split-localized* FORS orbitals are often considerably more effective in this respect. They are obtained by *localizing the strongly and the weakly occupied FORS orbitals separately without mixing the two orbital types* [36,67]. Thereby, one obtains two groups: the strongly occupied localized molecular orbitals (SOLMOs) and the weakly occupied localized molecular orbitals (WOLMOs). Fig. 1 exhibits the split-localized molecular orbitals for the NCCN molecule for a MCSCF-optimized FORS2 wavefunction of 18 electrons in 18 orbitals. The SOLMOs and WOLMOs for the HNO molecule were displayed in Fig. 2 of Ref. [67].

There exist other ways of determining acceptable weakly occupied localized MOs. One way would be to determine for each SOLMO a WOLMO* orbital in the weakly occupied space by maximizing the exchange integral with the SOLMO and, then, to orthogonalize all these WOLMO*s symmetrically. Another would be to localize the weakly *and* strongly occupied (in the FORS1 or FORS2 orbital space) natural orbitals *all together*, which will yield orbitals that are localized on the *atoms*. From the bond order matrix of these quasiatomic MOs one can then readily deduce SOLMOs and WOLMOs. Such alternative determinations of split-localized orbitals may prove useful [76–78] under bonding situations that differ from those examined here.

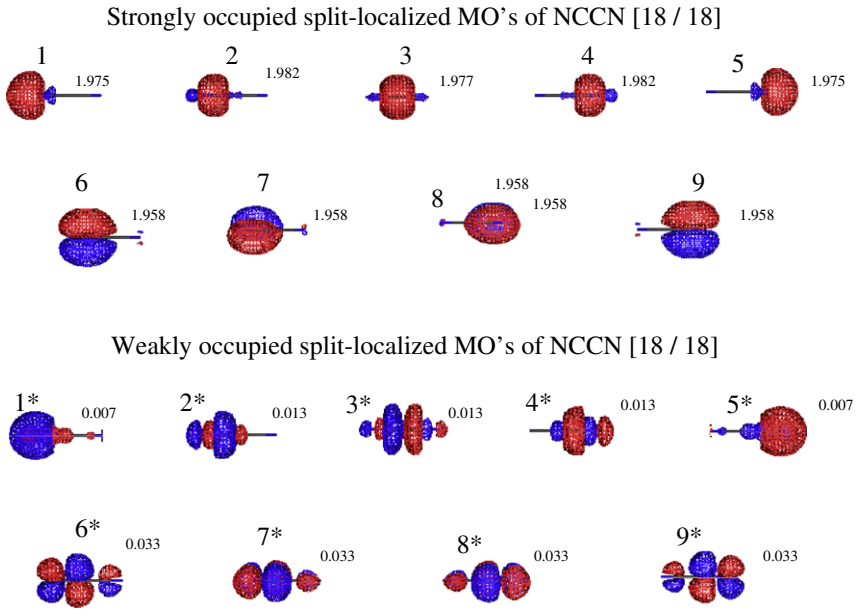

Fig. 1. Split-localized molecular orbitals of the NCCN molecule, MCSCF optimized in the space of 18 valence electrons in 18 orbitals. The orbital labeled n^* is the correlating orbital for the orbital labeled n. Occupation numbers are shown for each orbital.