Advances in the density matrix renormalization group method

for use in quantum chemistry

(spin adaptation, two–body density matrix evaluation, orbital optimization)

by

Dominika Zgid

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemistry

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis.

This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. I understand that my thesis may be made electronically available to the public.

Abstract

Despite the success of modern quantum chemistry in predicting properties of organic molecules, the treatment of inorganic systems, which have many close lying states, remains out of quantitative reach for current methods. To treat non-dynamic correlation, we take advantage of the density matrix renormalization group (DMRG) method that has become very successful in the field of solid state physics. We present a detailed study of the DMRG method, and we pay special attention to the evolution of the understanding behind the mathematical structure of the DMRG wave function. Our primary goal is to develop a density matrix renormalization group self–consistent–field (DMRG-SCF) approach, analogous to the complete active space self–consistent field (CASSCF) method, but dealing with large active spaces that are too demanding for the full configuration interaction (FCI) method. As a first step towards such a DMRG-SCF procedure, we present a spin-adapted DMRG algorithm designed to target spin- and spatial-symmetry states that are hard to obtain while using an unrestricted algorithm. Our next step is a modification of the DMRG algorithm to obtain decreasing energy at every step during the sweep. This monotonically convergent DMRG scheme lets us obtain the two-body density matrix as a by–product of the existing procedure without any additional cost in storage. Additionally, the two-body density matrix produced at convergence using this scheme is free from the N-representability problem that is present when the two–body density matrix is produced with the two-site DMRG scheme without additional storage cost. Finally, taking advantage of the modifications developed herein, we present results obtained from our DMRG-SCF method. Lastly, we discuss possible ways of incorporating dynamical correlation into the DMRG scheme, in order to obtain a modern multireference approach.

Acknowledgements

This work would not be possible without the everyday kindness of those by whom I was surrounded. I have been given much more than I will ever possibly be able to return. Those numerous people who I have met, during the years when I was a graduate student, have shaped me into what I am now. Although it is impossible to list here all these persons and numerous occasions, I will always remember them. Here, I will list all those that were the most directly involved in the process of creating this thesis.

I would like to thank my advisor Professor Marcel Nooijen, who gave me the chance to work on the density matrix renormalization group (DMRG) method, which back then when I was starting my Ph.D., was not a very well known method in electronic structure. I will always appreciate that he had the courage and eagerness for adventure to let me work on this topic. I am greatly indebted for the numerous discussions we have had that deepened my understanding of the theory and allowed me to understand which questions are important to be asked by scientists when trying to solve a problem. During my Ph.D., Marcel gave me a lot of independence. In this way he helped to equip me with the most valuable set of skills a scientist should possess: the determination to reach a goal and the judgment to choose the theoretical means that are required to reach it. However, he was always there to help me, when I was having difficulties with some problems and complaining that something cannot be done or implemented. To be his student meant not only to investigate the problems connected with this thesis, but also to develop both the "scientific backbone" as well as a broader view of science. To me, Marcel will always remain one of the most essential persons in development of my knowledge and scientific personality.

I thank Professor Bob LeRoy and Professor Fred McCourt for creating a wonderful lab atmosphere in our theory group in Waterloo. Both of them were a marvelous example of devotion to students. They have always listened to problems and tried to seek solutions. I benefited from observing in them that research excellency can be connected with fantastic teaching skills. Additionally, I am very grateful to Professor Bob LeRoy for correcting my thesis; I know that it required a lot of time and patience. I learned quite a bit in this process.

I thank Professors Paldus and Goddard for agreeing to be my committee members and for

their helpful comments during my Ph.D. I am indebted to Professor Paldus for his constant encouragement to pursue the investigation of the DMRG method.

During these years, I have interacted with fantastic colleagues that were in our theory group. I am especially thankful to Anirban and Alex, who were the only two peers that I knew in Waterloo when I arrived. I deeply appreciate all the interactions I have had with Liguo, who was also Marcel's graduate student, and thus a "brother in arms" for most of the duration of my Ph.D. I am greatly indebted for all good scientific (and not only) discussions we had with Yije, Sham, Ondrej, Bijoy and Hui. I have also learned a lot from visitors such as Professor Michael Hanrath and Professor Debashis Mukherjee, who came to Waterloo. I also appreciate a lot Jason, Aneta and Chris, who were undergraduate students working in our group. They also added a lot of color and joy to many grey days.

I am also very thankful to Cathy van Esch, who is an excellent graduate student secretary. She has always helped me to deal smoothly with all the formalities.

Before coming to Waterloo, I had met wonderful scientists and teachers who brought my attention to the importance of the multireference problem. Professors Jeziorski and Chałasiński have partially sparked my interests in electronic structure. I am thankful for all discussions I had with Professor Bartlett. I am also greatly indebted to Rafał for his constant faith and support during some of the difficult moments I have experienced.

I am thankful to Professor Paul Ayers for the discussions about density functional theory (DFT) and encouragement to pursue the work on the DFT-DMRG hybrid.

I am also greatly indebted to Professor Garnet Chan, whom I met during conferences, for his encouragement, enthusiastic and adventurous scientific attitude, and faith that one is able to solve even the most difficult problems.

During all this time, my strongest emotional supporters have been my grandma and my parents. It would be impossible to finish this work without them. I know that it was difficult for them to see me home very rarely, hence I appreciate the more their efforts to encourage me to seek in this experience how to become a better person and scientist.

Dla mojej babci i moich rodziców

(For my grandma and my parents)

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Statement of the problem

Every new student to the field of quantum chemistry, or more precisely electronic structure, will probably try to identify the challenges existing in this field. Soon, she or he can learn that even though many problems have been solved, there are some that still require attacking. One of the most visible challenges is the development of methods that have an advantageous computational scaling and scale linearly with the growth of the system. Another challenge is the treatment of phenomena that require wave functions with many configurations in order to qualitatively describe the system of interest. Let us focus on the second problem, and let us understand how important it may be to provide some solutions to it. This would allow modern computational chemistry to extend into areas of inorganic and metal chemistry with a predictive level of accuracy and provide a systematically improvable and reliable treatment.

It is well known that for many molecules the restricted Hartree-Fock (RHF) model gives a qualitatively correct description near the equilibrium geometry. The great success of a variety of single reference methods, such as configuration interaction (CI) [3], second order Møller-Plesset perturbation theory (MP2) [4] and coupled cluster (CC) [5, 6, 7, 4], in describing small closed shell systems near their equilibrium geometry with a high degree of accuracy is mostly based on having

a good starting point that is a qualitatively correct RHF wave function. Even though it may seem that existing single reference methods can deal with quite a broad range of chemical systems, their break down can be illustrated by a very simple example. Let us imagine the dissociation of a single bond in the hydrogen molecule, H_2. The RHF model cannot describe correctly the dissociation process resulting in the open-shell products, H + H. A possible solution of the problem would be to use the multireference approach to introduce coefficients in front of different covalent and ionic contributions to the wave function, illustrating appropriately the dissociation to $H^+ + H^-$ and H + H.

$$\Psi = c_{\text{ion}}\Phi_{\text{ion}} + c_{\text{cov}}\Phi_{\text{cov}}. \tag{1.1}$$

Thus, a multiconfigurational formulation can correctly describe the whole potential curve.

Another example that shows the inability of the RHF method for giving an even qualitatively correct description is the simultaneous dissociation of a multiple bond. The best known example, which is also a benchmark case for almost every method, is the dissociation of the triple bond in the nitrogen molecule, N_2. Almost none of the currently existing single reference methods give the correct description of the entire potential energy curve for N_2. These two cases illustrate the degeneracies that occur at the dissociation limit for a covalent bond.

Similar near–degeneracy effects may appear for some molecules even in their ground state close to the equilibrium geometry [7]. RHF has problems with the description of molecules with competing valence structures. When several structures are possible, as for the ozone molecule O_3 or for the NO_2 radical, the RHF model has problems taking into account the redistribution of the electrons within possible valence structures for radicals and biradicals.

The calculation of energy surfaces for chemical reactions is another case in which a simple single configurational Hartree-Fock description breaks down. These surfaces may have a barrier that is caused by the change of the dominant electronic configuration in the wave function. Consequently, it is not uncommon that a quantum chemical calculation of a system passing through the saddle point region needs to incorporate multiconfigurational character in the wave function.

A multiconfigurational description is also necessary for compounds containing transition metals, where the electronic configurations $d^n s^2, d^{n+1}s^1$ and d^{n+2} often characterize a variety of

electronic states with very similar energies.

Strong configurational mixing also occurs in excited states of molecular systems. Different singly excited configurations with respect to the RHF ground state are often very close in energy. When such configurations are of the same symmetry, even a small interaction term leads to strong mixing. Typical examples are the so called alternant hydrocarbons. Hence, the Fermi level degeneracy makes the study of all (1D) systems, such as conjugated polymers (e.g. polyacetylene) and annulenes ($C_M H_M$, where $M = 4m_0 + 2, m_0 = 1, 2, \cdots$), as models of (1D) metals, a real challenge for quantum chemistry.

The above examples show that many phenomena frequently encountered in chemistry require the inclusion of some multireference description, even if one only aims for a qualitative accuracy. The demand for multireference methods from users who are primarily interested in experimental and computational chemistry will grow due to increasing interest in molecular magnets, surface chemistry, catalysis, or metalloproteins. Many of these examples involve, of course, molecules that have too many orbitals and electrons for any *ab initio* calculation. However, the most chemically significant and reactive parts of these molecules are often restricted to 20–30 atoms. The major difficulty lies in describing the electronic states of such active parts or centers, while the rest of the molecule can be treated with some other suitable method with a lower computational scaling. Hence, in present day quantum chemistry, on the one hand one tends to work on methods that can have low computational scaling with the cost of the computation growing linearly with the size of the system. However, there is simultaneously a need to work on methods that can treat many of the multireference phenomena, since there are many experimental properties (e.g. from inorganic chemistry) that cannot be calculated with existing methods.

The quest for a general method that treats cases in which the mean-field theory breaks down is still of high importance. Of course, providing a good description of multireference phenomena is not the only quality that is demanded from a method. In particular, one can list a number of criteria that are required to balance the applicability and generality of a theoretical method [8].

1. *Size-consistency, size-extensivity and separability.* The energy should scale properly with the number of electrons and converge to proper separated fragments by bond breaking.

2. *Smoothness of description of different regions of the potential energy surface*[1] with only a single type of calculation that utilizes the same parameters for the calculation of the whole surface.

3. *Symmetry.* The model should not violate the spatial or spin symmetry required by the system.

4. *Effective description of "static" (near degeneracy effects) and "dynamic" (cusp region) correlation*[2].

5. *Generality I.* The model should be able to work not only for a special configuration and symmetry, and at the very least, future generalizations should be possible.

6. *Generality II.* The model should be flexible enough to be easily extended to ground, excited, ionized or open shell states, as well as being able to provide gradients and properties.

7. *Systematic improvement.* The model should converge to a full CI energy in a systematic way after inclusion of all needed parameters.

8. *Black-box usage.* The method should be accessible for non-expert users.

Providing a suitable description of multireference phenomena and satisfying the criteria listed above will be important guidelines that the density matrix renormalization group (DMRG) method studied in this thesis will be judged against. One can ask now, why DMRG should be used in chemistry and what was the motivation of this work?

The density matrix renormalization group method was presented for the first time in 1992 by White [9], and it soon become very successful in the field of solid state physics. Some years passed before DMRG was introduced to chemistry, but a promise of the potential of the method could be seen in 1998 when Fano *et al.* [10] performed some studies for cyclic polyenes (C_MH_M, $M = 4m + 2$, $m = 1, 2, \ldots$) called [M]annulenes with the DMRG method within the Pariser-Parr-Pople (PPP) model [11, 12, 13] and obtained excellent results. Similarly good results were obtained

[1] Different portions of the potential energy surface often require different dominant configurations.

[2] All correlation is included in the full CI, so "static" ("non-dynamic") and "dynamic" are only used to distinguish between approximations.

by Bendazzoli *et al.* [14] for polyacetylene. The significance of the message that has been sent by these papers can be brought to light, if one realizes that [M]annulenes were difficult multireference systems that have been studied since 1984 by Paldus and coworkers [15, 16]. They discovered and documented the breakdown of the CCSD method for these cyclic polyenes systems attributing the failure to the omission from the CCD method of terms providing the coupling between the D and Q excitations. Later studies (in 2002), of such highly correlated systems with an added triples correction (CCSD(T)) and subsequently quadruples correction (CCSDTQ) conducted by Podeszwa and Stolarczyk [17] revealed that even the addition of high order excitations was not able to remedy the difficulties encountered in their treatment of these systems in the strongly correlated limit, due to an inability to handle highly multireference effects by a single reference CC, even after inclusion of high order excitations. Although the early DMRG studies in quantum chemistry by Fano *et al.* [10] and Bendazzoli *et al.* [14] were still in the spirit of the Hubbard model [18], it was clear that the DMRG method, even though seemingly single reference in nature, could include multireference effects that were impossible to deal with at a single reference CC level with high excitations.

In 1999, the first application of the DMRG method to a molecule without invoking any approximate Hamiltonians was made by White and Martin [19]. In that paper, a water molecule was studied and a foundation was provided for an efficient procedure of multiplication of the Hamiltonian matrix by a vector. The next study of a molecule was published by Daul *et al.* [20] in 2000. They investigated the singlet–triplet gap in HHeH, and the role of the localized orbitals was examined for the first time without yielding remarkable improvements (perhaps due to examination on a water molecule). A 2001 paper by Mitrushenkov *et al.* [21] brought attention to the importance of orbital ordering, a question that did not arise in the applications in physics. Then in 2002, Chan and Head–Gordon [22] published the first in–depth study of the DMRG method as applied to quantum chemistry. Their paper contained extensive discussion of the algorithm, implementation details, and computational scaling of the method, thus shaping the terminology used for DMRG in quantum chemistry. Chan and Head–Gordon [22] also presented numerous benchmark calculations and proposed use of the symmetric reverse Cuthill-McKee ordering [23, 24] of the

orbitals.

In 2003, there appeared papers by Legeza *et al.* [25, 26, 27] in which DMRG was examined for the ionic - neutral curve crossing between the two lowest $^{(1)}\Sigma^{(+)}$ states of LiF [26], showing excellent performance for the ground and several low-lying excited state energies. Legeza *et al.* (in Ref. [27]) also studied the issue of ordering by the criteria of maximization of the entropy of entanglement. Chan has pushed the limit of DMRG applicability by developing a parallel version of the code [28] and performing a series of calculations [28, 29] that showed that the DMRG method could outperform the CCSDTQ method by 0.019 mH at $M = 3000$ for a water molecule at equilibrium in the TZ2P basis, and could yield an extrapolated exact result for this basis with 41 orbitals and 10 electrons, while the full configuration interaction for such a basis was far beyond reach. In recent years, papers from the Reiher group were published, of which the most well known address the analysis of the orbital ordering by means of the genetic algorithm [30] and analysis of the DMRG wave function in terms of Slater determinants [31]. Chan's group also continued to publish on the DMRG topic, including a remarkable paper [32] in which a "duel" between the DMRG method and CC was presented for nitrogen bond breaking, which showed that DMRG method yielded better results than CCSDTQPH. Another interesting study by Chan's group included DMRG with quadratic scaling for analyzing linear polyenes [33]. In that study, due to the inherent locality of the DMRG ansatz, a remarkable result was achieved without imposing the domains that are usually used with localized orbitals by other methods. In the same paper, it was also shown that with the DMRG method it is possible to include 100 orbitals and 100 electrons in the active space.

Although there has been a significant progress in application of the DMRG method to quantum chemistry, some disbelief and misunderstanding has been still hindering its full acceptance by the quantum chemistry community. Perhaps the most important reason for this was the algorithmic description presented in the early papers, which lead to a false belief that there is no underlying wave function in the DMRG method. However, recently due to work of Östlund and Rommer [34, 35] and efforts of other authors [33, 36, 37, 38] to popularize the wave function perspective, the DMRG ansatz has been explained better with the help of the ansatz known in statistical

physics as a matrix product state. One hopes that these papers will provide a better foundation for understanding the DMRG method in the quantum chemistry community. It is also worth mentioning that in physics, DMRG emerged on a firmer footing recently due to many connections between DMRG and quantum information theory which gave rise to the time–dependent [39] formulation of the DMRG method as well as generalization of DMRG to 2D lattices [40, 41, 42, 43] making many new applications possible.

In the last year, formulations allowing the orbitals to be optimized while using the DMRG method in the active space (DMRG-SCF) were developed by Chan [37] and by Zgid and Nooijen [44] . Such an approach, which is analogous to the complete active space self consistent field (CASSCF) method, should allow one to apply DMRG to many interesting molecules from inorganic chemistry. Additionally, Chan's group has published a method that targets excited states with DMRG [45], allowing access to exited states of molecules at moderate cost, while Zgid and Nooijen presented a spin–adapted version of the DMRG method [46] making DMRG even more suitable for the difficult metal systems.

I believe that one can safely say that the development of the (1D) formulation of DMRG has reached an advanced state, and that the dream of treating complicated multireference systems is within reach. The remaining problem is the inclusion of dynamical correlation that is not well covered in the DMRG calculations with a small number of states. Several solutions can be proposed to such a problem, among which the most suitable for DMRG includes the work connected with canonical diagonalization (CD) [47, 48, 49]. Another problem that needs to be tackled is that of orbital ordering; however, it is unclear if such problem can be handled efficiently using 1D DMRG, and consequently there is a need to develop the higher dimensional formulation of the DMRG method. Such formulations may also help to deal with the inclusion of the effect of dynamical correlation. Additionally, the analysis of these types of ansatz can help to classify phenomena according to their wave function ansatz, thus forming a classification of possible treatments. Finally, we note that time–dependent DMRG can be used for spectra simulation, hence bringing a remedy to very highly–dimensional bosonic problems.

1.2 Scope of this thesis

Since the structure of this thesis partially follows the evolution of our understanding of the DMRG method, let me briefly discuss the content of the chapters in order to provide an overview of the material to be presented.

Chapter 2 introduces the historical language commonly adopted by DMRG practitioners, and the presentation therein highlights the algorithmic aspects of the problem rather than the DMRG wave function structure. This chapter starts by outlining the reasons for the breakdown of the traditional renormalization group (RG) methods, as applied to the particle in a box problem, hence setting the foundation for consideration of the DMRG as a remedy to the problem. The introduction of the DMRG algorithm is not intended to be presented in an *ad hoc* manner, and three arguments justifying the DMRG truncation procedure are presented to familiarize the reader with the concept. Next, the DMRG algorithm is introduced and all the components of a single DMRG step are discussed in detail. The implementation details are presented, the most important part being the multiplication of the Hamiltonian matrix by a vector, a step that is essential for understanding DMRG computational scaling. Finally, a discussion of the memory and disk requirements is presented, together with an operation count for all the stages of a single step.

In Chapter 3, the inclusion of spin and spatial symmetry in the DMRG method is presented. Hence, the language used is rather practical, and no connection is made with modern terminology. The algorithm that deals with spin and spatial symmetry within the DMRG method, is presented, and all necessary modifications of the traditional algorithm are explained in detail. The structure of the DMRG wave function, as compared to the usual FCI wave function, is only mentioned in order to indicate that DMRG has an underlying wave function. Finally, numerical examples are presented illustrating the spin-adapted DMRG performance when applied to a system with intersecting spin states.

Chapter 4 is a turning point of this thesis due to the introduction there of the modern DMRG language. The DMRG ansatz is discussed at length, and the wave function form of DMRG is highlighted. Different parametrizations for one– and two–site DMRG are presented, leading to

a discussion of the convergence of the DMRG method. The space in which the minimization of the energy is performed is considered, and the problem of possibly getting stuck in local minima is studied at length. A perturbative correction that prevents the DMRG minimization from converging to a local minimum is introduced. The features of the DMRG method are listed and discussed, and an examination of the degree to which it satisfied the criteria presented in the Introduction is attempted. Numerical examples presented in this chapter involve investigations of the convergence of the two possible DMRG schemes, as well as an investigation of the effectiveness of the perturbative correction in preventing the algorithm from getting stuck in local minima.

In Chapter 5, the modern DMRG language is extended to the multidimensional formulation of the DMRG method. This chapter should be treated as a glimpse of possible new developments involving the multidimensional formulation of the DMRG method known as projected entangled pair states (PEPS) that was developed during the last two years in the solid state physics community [41, 42, 43]. Although the author of this thesis never worked on the PEPS, the multidimensional formulation illustrates how one can benefit from the knowledge of the mathematical structure of the DMRG ansatz since the PEPS formulation would not be possible while using only practical algorithmic language. Additionally, it is very likely that there will be a lot of new development originating from PEPS, since it can be used to investigate model crystal lattices of some super-conductors, and together with the time–dependent formulation it can be used to simulate spectra. Hence, this is a "future outlook" chapter, and its location after Chapter 4 is justified by the introduction of language that is an extension of that for simple 1D DMRG.

Chapter 6 returns to the more usual quantum chemistry ground and takes advantage of the formalism developed in Chapter 4. An approach is presented that allows one to produce the two–body density matrix during the DMRG run without any additional increase in the current disk and memory requirements. The method is based on an assumption that different elements of the two–body density matrix can be calculated during different steps of a sweep. Hence, it is desirable that the wave function at the convergence does not change during a sweep. From the knowledge of the theoretical structure of the DMRG ansatz developed in Chapter 4, we know that during the one–site DMRG procedure [50] the energy and the wave function is converging

monotonically at every step of the sweep. Thus, the one–site algorithm provides an opportunity to obtain the two–body density matrix free from the N-representability problem.

In Chapter 7, the knowledge gained in Chapters 3, 4, and 6 is used to develop the density matrix renormalization group self–consistent field (DMRG-SCF) approach that is analogous to the complete active space self–consistent field (CASSCF) method, but instead of using the FCI method for the description of the active space, DMRG-SCF uses the DMRG method. The DMRG-SCF approach, like CASSCF, properly describes the multiconfigurational character of the wave function, but avoids the exponential scaling of the FCI method and replaces it with a polynomial scaling. Hence, calculations for a larger number of orbitals and electrons in the active space are possible since the DMRG method provides an efficient tool to automatically select from the full Hilbert space the many–body contracted basis states that are the most important for the description of the wave function.

Since the qualitative means of analysis of the electronic structure are presented in Chapter 7, in Chapter 8, we start to consider the inclusion of dynamical correlation effects after the DMRG-SCF calculations. This chapter should be treated as a possible outline of future developments that may be attempted. The two simplest ways of adding the dynamical correlation are discussed, namely, **(i)** n–electron valence perturbation theory (NEVPT) [51, 52, 53] obtained while using the DMRG-SCF as a zero order wave function, and **(ii)** a possible DMRG-DFT hybrid in which the non–dynamic correlation effects are covered by the DMRG calculations while the dynamic correlation can be dealt with by DFT due to the splitting of the Coulomb two–electron integrals into the short– and long–range parts [54, 55, 56].

Chapter 2

Algorithmic beginnings

The density matrix renormalization group (DMRG) [9] was conceived as a numerical approach for finding accurate approximations to the ground state and the low-lying excited states of strongly interacting (1D) systems such as the Heisenberg [57], $t-J$ [18], and Hubbard [58] models. DMRG was a remedy that eliminated the ill treatment of boundaries from the original renormalization group (RG) technique [59, 60] developed by Wilson. After the first paper published in 1992 by White [9], DMRG started to be widely popular in physics and was applied to calculate dynamic properties [61, 62], conjugated polymers [63, 64, 65], thermodynamic properties [66], and electron-phonon systems [67, 68]. DMRG was proven to be an amazingly powerful method for the description of these phenomena since it was able to treat efficiently correlation beyond the mean-field theory level. That is also essentially the challenge faced by all the post–Hartree–Fock methods in quantum chemistry.

In this chapter, we attempt to describe the DMRG method from a historical perspective. Hence, we will be using the language of the standard (real–space) renormalization group (RG), and consequently the description presented will focus on the algorithmic aspects of the method. Nowadays, a more elegant and powerful formulation of the DMRG method has begun to be well established, and it will be described in later chapters of this thesis. However, most of the literature published in physics and chemistry up to recent days was based on the algorithmic formulation,

thus making it hard to omit providing such a historical perspective.

The structure of this chapter is as follows: we will first illustrate the dimensionality problem of the Full Configuration Interaction (FCI) method in Sec. 2.1, and the renormalization procedure as applied by Wilson will be briefly outlined in Sec. 2.2. Section. 2.3 will introduce the DMRG background together with the discussion of the origin of the truncation procedure present in the DMRG method. These considerations will lead us to presentation of the DMRG algorithm for two–body Hamiltonians (as used in quantum chemistry) in Sec. 2.4. In that section, we will also discuss the DMRG expansion procedure and the construction of the Hamiltonian matrix, as well as the decimation step of the DMRG algorithm.

2.1 The dimensionality problem. Renormalization

Let us assume that we would like to calculate energy or properties of a system described by a general two–body Hamiltonian,

$$\hat{H} = \sum_{pq} t_{pq} \hat{p}^\dagger \hat{q} + \frac{1}{4} \sum_{pqrs} v_{pqsr} \hat{p}^\dagger \hat{q}^\dagger \hat{r} \hat{s} \tag{2.1}$$

in which the operator $\hat{p}^\dagger$ creates an $\uparrow$ or $\downarrow$ electron on the spin orbital p, and t_{pq} and v_{pqsr} are integrals characteristic of the system of interest. One of the simplest methods of finding the energy of a system is to express its Hamiltonian matrix in a basis and try to diagonalize such a matrix. A natural basis for such a problem is the basis in which the states are constructed in the following form

$$|1\rangle \otimes |2\rangle \otimes \cdots \otimes |n\rangle, \tag{2.2}$$

where p is a state of a single spin orbital $|\uparrow\rangle$ or $|\downarrow\rangle$ and the symbol $\otimes$ is used to denote a direct product. Hence, the size of the space of interest grows exponentially, like 2^n,[1] with the number of orbitals, n. Consequently, even for small orbital spaces full configuration interaction (FCI)

[1] Of course, if one does not consider a space with all possible numbers of electrons, but a space with a fixed number of electron and a particular spin state, then the dimension is $\frac{2S+1}{N+1}\binom{n+1}{\frac{1}{2}N-S}\binom{n+1}{n-\frac{1}{2}N-S}$, where n is the number of orbitals, N is the number of electrons, and S is the total spin eigenvalue.

calculations are beyond reach, and the largest "state of the art calculations" performed to date involved up to $O(20)$ orbitals.

One can ask now, if it is possible to efficiently reduce the size of the basis for the problem in a controlled way without causing a drastic loss in accuracy of the results. Such a reduction is called *numerical renormalization*, and it may be done very efficiently for certain classes of Hamiltonians. Such problems must reveal a scale separation, and a long–wavelength low–energy scale has to dominate the physical properties of such systems. Thus, short length physical properties may be integrated out and summed into qualitative modifications of the long–range behavior. To be more specific, one can start with some microscopic Hamiltonian and progressively integrate out the unimportant degrees of freedom using successive renormalization group transformations. Hence, one obtains a simplified "renormalized" effective Hamiltonian that essentially captures the physics of the system under investigation. These observations were the essence of the numerical renormalization group proposed by Wilson [60]. Wilson mapped the Kondo problem [69] onto a 1D lattice, and proposed the following RG procedure:

1. Take a subsystem containing A orbitals that is small enough that the Hamiltonian of this subsystem H_A can be diagonalized exactly.

2. Diagonalize exactly the Hamiltonian H_A and save the M lowest eigenvalues and their corresponding eigenvectors. Note that $1 \leq M \leq \dim(A)$, in which $\dim(A)$ means the the dimension of the space spanned over the A orbitals.

3. Transform the Hamiltonian H_A and other operators needed to construct the full Hamiltonian to the new truncated basis made up of the M lowest eigenvectors, namely $\tilde{H}_A = U_A^\dagger H_A U_A$ and $\tilde{O}_A = U_A^\dagger O_A U_A$, where U_A is a matrix, whose columns are the eigenvectors corresponding to the M lowest eigenvectors of the H_A.

4. Add a single orbital to $\tilde{H}_A$ and construct the interaction between the added orbital and the already existing transformed Hamiltonian. In this step, a Hamiltonian H_{A+1} is formed for a subsystem containing $A+1$ orbitals.

5. Go to point 2 and substitute H_{A+1} for H_A.

The success of the above scheme for the Kondo problem gave hope that the numerical RG scheme may also be useful for treating other 1D many–body problems such as the Hubbard or Heisenberg models. However, the results obtained for such models did not turn out to be satisfactory. The numerical RG scheme had the underlying assumption that the low–energy eigenstates obtained from the system of size A are important in making up the low–energy states of a system of size $A+1$. Consequently, when the procedure was applied to one–dimensional models such as Hubbard or Heisenberg model with no intrinsic energy scale separations, the results became unsatisfactory after just a few iterations.

The reason for the RG failure in many–body problems was not obvious; it was not until 1992 that White and Noack [70] identified the reasons for the breakdown of the RG procedure by considering a simple well–known one–body problem, the particle in box. They concluded that the ground state wave function for a particle in a larger box cannot be built out of the ground state wave functions corresponding to the two smaller boxes. In particular, the ground state wave function built out of those for the smaller boxes has a node in the center of the larger box, while the true ground state wave function of the larger box has the largest amplitude there, as illustrated in Fig. 2.1. Consequently, White and Noack realized that the decision regarding which states should be retained for a finite size subsystem has to based on the analysis not only of the subsystem itself, but also of the subsystem embedded in a larger environment trying to describe the thermodynamic limit system that is containing the subsystem of interest. This realization is the key point underlying the success of the density matrix renormalization group (DMRG) method.

2.2 DMRG truncation

We have learned that it is not enough to consider only the subsystem itself, and that it is necessary to try to mimic the whole system with a subsystem as a part embedded in the whole system. Thus, the whole system under consideration can be separated into the system S and the environment E. During the process of expansion, both of these parts can be enlarged by a single orbital. From

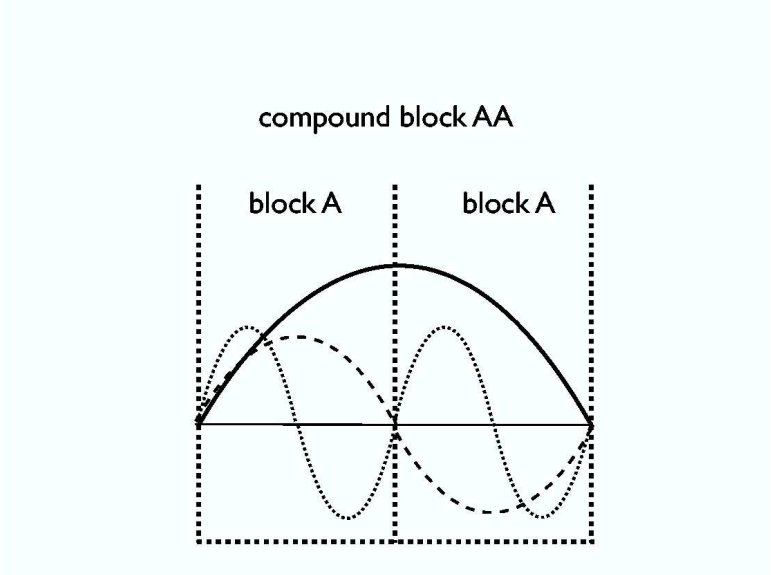

Figure 2.1: Lowest states (dashed lines) for the "building unit" A boxes. Lowest state (solid line) for the composite AA box. Note that the nodes of the dashed lines are placed in the middle of the composite box where the solid line has its maximum amplitude.

now on we will call the whole physical system, which consists of the system plus the environment, a superblock, while system or environment will be referred to as a block. Thus, the wave function of the expanded by one orbital system S has the following form

$$|\phi_i^S\rangle = \sum_{m^S=1}^{M^S} \sum_{\sigma^S=1}^{N_{\rm orb}} c_{m^S\sigma^S} |m^S\rangle |\sigma^S\rangle, \tag{2.3}$$

where m^S labels M^S states of the system before expanding it by one orbital, and σ^S labels the possible $N_{\rm orb}$ states of a single added orbital. Analogous procedure can be performed for the environment block,

$$|\phi_j^E\rangle = \sum_{m^E=1}^{M^E} \sum_{\sigma^E=1}^{N_{\rm orb}} c_{\sigma^E m^E} |\sigma^E\rangle |m^E\rangle. \tag{2.4}$$

Consequently, the wave function of the superblock obtained during the diagonalization procedure will have the following form

$$|\Psi\rangle = \sum_{i=1}^{N^S} \sum_{j=1}^{N^E} c_{ij} |\phi_i^S\rangle |\phi_j^E\rangle, \tag{2.5}$$

where the expanded system and the environment have orthonormal product bases with dimensions $N^S = M^S N_{\text{orb}}$ and $N^E = N_{\text{orb}} M^E$, respectively. Analogous to Wilson's renormalization procedure, one has to develop a truncation scheme to reduce the dimension of the expanded system from N^S to some smaller number M^S. However, in contrast to Wilson's approach, one does not consider only the Hamiltonian (or wave function) of the system itself; rather, the wave function for the whole superblock has to be constructed and considered. Thus, a new truncation scheme has to be proposed that is different from the original Wilson scheme. To justify the truncation present in the DMRG method, I will follow the three arguments listed in the excellent review [71] by Schollwöck.

1. *Optimization of the wave function by minimization of quadratic norm,* $|||\Psi\rangle - |\tilde{\Psi}\rangle||^2$ *between the expanded* $|\Psi\rangle$ *and truncated* $|\tilde{\Psi}\rangle$ *wave function.* An argument analogous to this one was originally presented by White in the early papers [9, 72].

   The wave function of the superblock is given by Eq. (2.5). Since one would like to truncate the N^S system states to some smaller number $M^S < N^S$, one can create M^S orthonormal states $|\alpha\rangle = \sum_i u_{\alpha i} |\phi_i^E\rangle$ and express the total truncated wave function as,

$$|\tilde{\Psi}\rangle = \sum_{\alpha=1}^{M^S} \sum_{j=1}^{N^E} a_{\alpha j} |\alpha\rangle |\phi_j^S\rangle. \tag{2.6}$$

   Since we would like to minimize the quadratic metric, $|||\Psi\rangle - |\tilde{\Psi}\rangle||^2 = \langle\Psi|\Psi\rangle - \langle\Psi|\tilde{\Psi}\rangle - \langle\tilde{\Psi}|\Psi\rangle + \langle\tilde{\Psi}|\tilde{\Psi}\rangle$, and assuming that we have only real coefficients, one tries to minimize

$$|||\Psi\rangle - |\tilde{\Psi}\rangle||^2 = 1 - 2\sum_{\alpha i j} c_{ij} a_{\alpha j} u_{\alpha i} + \sum_{\alpha j} a_{\alpha j}^2 \tag{2.7}$$

   with respect to coefficients $a_{\alpha j}$ and $u_{\alpha i}$. Thus, Eq. (2.7) is stationary in $a_{\alpha j}$ if we have $\sum_i c_{ij} u_{\alpha i} = a_{\alpha j}$. This result implies that

$$|||\Psi\rangle - |\tilde{\Psi}\rangle||^2 = 1 - \sum_{\alpha i i'} u_{\alpha i} \rho_{i i'} u_{\alpha i'}, \quad \rho_{i i'} = \sum_j c_{ij} c_{i' j} \tag{2.8}$$

must be stationary for the global minimum with respect to the distance (the quadratic norm), and that one has to find the $u_{\alpha i}$ coefficients that minimize the distance. Using the Rayleigh–Ritz variational principle, one finds that Eq. (2.8) is stationary, if $|\alpha\rangle$ are the eigenvectors of the density matrix ρ. Hence, the global minimum is found by choosing $|\alpha\rangle$ to be the M^S eigenvectors $|w_\alpha\rangle$ that correspond to the largest w_α eigenvalues of the density matrix. Since all the eigenvalues of the density matrix are non–negative, we obtain

$$\||\Psi\rangle - |\tilde{\Psi}\rangle\|^2 = 1 - \sum_{\alpha=1}^{M^S} w_\alpha = \epsilon_p, \tag{2.9}$$

where ϵ_p is a *truncated weight.*

2. *Optimization of expectation values* [73]. This argument can be helpful to understand the origin of the density matrix used in DMRG and its difference with the n-body reduced density matrix commonly used in electronic structure calculations. The DMRG reduced density matrix is different due to the fact that the summation only run over the block basis states and not over all electron coordinates except the coordinates for n electrons (as it is for the n-body reduced density matrix). An excellent reference explaining the origin and the properties of the RDM used in DMRG can be found in Ref. [74]. The wave function of a superblock (see Eq. (2.5)) can be represented as a pure state $|\Psi\rangle$. In statistical physics, the physical state of the system is described by a reduced density matrix $\hat{\rho}$,

$$\hat{\rho} = \mathrm{Tr}_E |\Psi\rangle\langle\Psi|, \tag{2.10}$$

where one traces out the basis states from the environment block. Consequently, one obtains

$$\langle \phi_i^S | \hat{\rho} | \phi_{i'}^S \rangle = \sum_j c_{ij} c_{i'j}^*. \tag{2.11}$$

Such a matrix has N^S eigenvalues w_α and orthonormal eigenvectors $\hat{\rho}|w_\alpha\rangle = w_\alpha|w_\alpha\rangle$, where $w_\alpha \geq 0$ and $\sum_\alpha w_\alpha = 1$. Let us demonstrate now that the ground state is best described if one retains the M^S eigenvectors that correspond to the largest eigenvalues. Let us consider

some operator $\hat{A}$ that is acting on the system. The expectation value of $\hat{A}$

$$\langle \hat{A} \rangle = \frac{\langle \Psi | \hat{A} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \mathrm{Tr}_S \hat{\rho} \hat{A} \equiv \epsilon_A \tag{2.12}$$

can be expressed in the density matrix eigenbasis yielding

$$\langle \hat{A} \rangle = \sum_{\alpha=1}^{N^S} w_\alpha \langle w_\alpha | \hat{A} | w_\alpha \rangle. \tag{2.13}$$

Next, as assumed earlier, one can truncate the space up to the M^S dominant eigenvectors $|w_\alpha\rangle$ corresponding to the largest eigenvalues w_α, and one finds

$$\langle \hat{A} \rangle_{\mathrm{approx}} = \sum_{\alpha=1}^{M^S} w_\alpha \langle w_\alpha | \hat{A} | w_\alpha \rangle \tag{2.14}$$

with the error estimated by

$$|\langle \hat{A} \rangle_{\mathrm{approx}} - \langle \hat{A} \rangle| \leq \left(\sum_{\alpha > M^S}^{N^S} w_\alpha \right) c_A \equiv \epsilon_\rho \epsilon_A, \tag{2.15}$$

where the *truncated weight* is defined as

$$\epsilon_\rho = 1 - \sum_{\alpha=1}^{M^S} w_\alpha. \tag{2.16}$$

Such an estimate holds in particular for energies. For an expectation value of a specific operator $\langle \hat{A} \rangle$ the estimate could be tightened and a more efficient truncation procedure can be designed. However, for an arbitrary bounded operator acting on the system, the above prescription is optimal. Since we retain M^S eigenvectors corresponding to the largest eigenvalues, the rapid decay of the density matrix eigenvalues is crucial for the performance of the truncation procedure. It should be also mentioned that the *truncation weight* defined in Eq. (2.16) is equal to the absolute error made by the truncation procedure if the description of the environment is exact. One has to keep in mind that not only the *truncation error*, but

also the *environmental error* should be considered. Consequently, the errors for calculated observables in DMRG are usually larger than the truncation error itself, even though an iterative procedure is employed to eliminate the environmental error.

3. *Optimization of entanglement* [75, 76, 77, 78, 79]. Here, the DMRG wave function is considered from the quantum information theory perspective. The DMRG wave function is *entangled*, which means that it cannot be written as a simple product of one system and one environment state. Following *Schmidt decomposition* [80] of the wave function $|\Psi\rangle$, one can study bipartite entanglement, and such a procedure will be illustrated below.

   In the DMRG superblock wave function, the matrix of the coefficients c_{ij} is $(N^S \times N^E)$–dimensional, and we can assume that $N^S \neq N^E$. Singular Value Decomposition allows to write $\mathbf{c} = \mathbf{U}\mathbf{D}\mathbf{V}^\mathbf{T}$, where $\mathbf{U}$ is an $(N^S \times N^E)$–dimensional matrix with orthonormal columns, $\mathbf{D}$ is an $(N^E \times N^E)$–dimensional matrix with non–negative entries $D_{\alpha\alpha} = \sqrt{w_\alpha}$, and $\mathbf{V}^\mathbf{T}$ is a $(N^E \times N^E)$–dimensional unitary matrix. Consequently, the superblock wave function $|\Psi\rangle$ can be written as

$$|\Psi\rangle = \sum_{i=1}^{N^S}\sum_{\alpha=1}^{N^E}\sum_{j=1}^{N^E} U_{i\alpha}\sqrt{w_\alpha}V^T_{\alpha j}|\phi^S_i\rangle|\phi^E_j\rangle = \sum_{\alpha=1}^{N^E}\sqrt{w_\alpha}\left(\sum_{i=1}^{N^S}U_{i\alpha}|\phi^S_i\rangle\right)\left(\sum_{j=1}^{N^E}V^T_{\alpha j}|\phi^E_j\rangle\right). \quad (2.17)$$

   The orthonormality properties of $\mathbf{U}$ and $\mathbf{V}^\mathbf{T}$ ensure that $|w^S_\alpha\rangle = \sum_i U_{i\alpha}|\phi^S_i\rangle$ and $|w^E_\alpha\rangle = \sum_j V_{j\alpha}|\phi^E_j\rangle$. Using such orthonormal bases for the two blocks, the Schmidt decomposition of the superblock is

$$|\Psi\rangle = \sum_{\alpha=1}^{N_{\mathrm{Schmidt}}}\sqrt{w_\alpha}|w^S_\alpha\rangle|w^E_\alpha\rangle, \quad (2.18)$$

   in which $N_{\mathrm{Schmidt}} \leq \min(N^S, N^E)$ non–zero coefficients have replaced the c_{ij} coefficients of the original Eq. (2.5).

   Let us now try to explain why the number of non–zero coefficients is equal to the minimum of the system and the environment dimensions. If one constructs the density matrices of

the system and the environment

$$\hat{\rho}^S = \sum_{\alpha}^{N_{\text{Schmidt}}} w_\alpha |w_\alpha^S\rangle\langle w_\alpha^S|, \quad \hat{\rho}^E = \sum_{\alpha}^{N_{\text{Schmidt}}} w_\alpha |w_\alpha^E\rangle\langle w_\alpha^E| \tag{2.19}$$

after tracing out the environment and system states, respectively, they have the same number of non–zero eigenvalues which is equal to the smaller dimension $\min(N^S, N^E)$ and the same eigenvalue spectrum. One can then use the obtained eigenvalues w_α to measure the entanglement between the system and environment using the von Neumann entropy

$$S_{\text{vonNeumann}} = -\text{Tr}\hat{\rho}\ln_2\hat{\rho} = -\sum_{\alpha}^{N_{\text{Schmidt}}} w_\alpha \ln_2 w_\alpha. \tag{2.20}$$

Consequently, the entropy, and thus the entanglement, is maximized if one retains the M^S largest eigenvalues with the corresponding eigenvectors since the function $-x\ln_2 x$ grows monotonically for $0 < x \leq 1/e$.

The three arguments presented above give us confidence in the particular scheme of truncation used in DMRG. In the DMRG method, the "renormalization of the degrees of freedom", which is the core idea of the RG methods, is executed by the truncation of eigenvalues for the reduced density matrix. Consequently, calling the DMRG algorithm a renormalization group method is justified if one considers iterative truncation of the degrees of freedom leading to effective Hamiltonians. Nevertheless, DMRG is not providing an infrared or ultraviolet energy cutoff that was considered essential for the traditional RG methods.

2.3 The DMRG algorithm

In the previous section, the structure of the DMRG wave function was explained and the arguments shedding light on the truncation procedure were presented. In this section, the focus will be put on practical usage of that procedure and its employment in the algorithmic scheme behind the DMRG method. The algorithm presented here will be separate from tedious implementation details that will be outlined in the later sections of this chapter. We will also focus here on the

scheme (finite–system DMRG) that is used for calculations in quantum chemistry, rather than on the traditional scheme (infinite–system DMRG) used for calculation of one–dimensional lattice systems in solid state physics.

Note that from now on we will no longer use the names *system* and *environment* and will use *left* and *right* instead. Although the names "system" and "environment" are historically justified, they do not have any meaning for a reader who is unfamiliar with the historical development of the DMRG method. Consequently, left and right refer to the present geometric arrangement, and can be more easily intuitively understood without referring to any background information. Let us assume that all needed quantities are already accessible, and that the DMRG algorithm was initialized earlier. The initialization procedure is a detail from the implementation point of view, and a brief prescription for starting the algorithm will be given in the later sections. Without loss of generality, we assume that the orbital space is spanned by m orbitals. In the following, we will use left (with small l) for the space before blocking or expansion, while the expanded space will be denoted by Left (with capital L), and an analogous notation will be used for the right block.

1. *Blocking.* The left block is built out of n orbitals, and the Hilbert space describing such an orbital space in the left block is spanned by M^L states $\{|m_n^L\rangle\}$. We assume that we know and have access to the Hamiltonian $H_n^{(l)}$ and the matrices of operators $\langle l_\mu|\hat{O}|l_\lambda\rangle$ expressed in the basis spanned by M^L states. Similarly, the right orbital space contains $m-n-2$ orbitals, and that the corresponding Hilbert space is described by M^R states. We also assume that the matrices of the Hamiltonian $H_{m-n-2}^{(r)}$ and all other operators are expressed in such a basis.

   The left orbital block is enlarged by one orbital, thus forming a Left block containing $n+1$ orbitals, the Hilbert space is spanned by $N^L = M^L N_{\mathrm{orb}}$ states expressed as $|\phi_i^L\rangle = \sum_{m^L=1}^{M^L}\sum_{\sigma^L=1}^{N_{\mathrm{orb}}} c_{m^L\sigma^L}|m^L\rangle|\sigma^L\rangle$, where $i = 1,\ldots,N^S$ or simply $\{|m_n^L\rangle|\sigma\rangle\} \equiv \{|m_n^L\sigma\rangle\}$. Analogously, the right orbital space is also enlarged by one orbital forming the new Right space containing $m-n-1$ orbitals. We shall assume real eigenvectors without loss of generality. Next, the new Left Hamiltonian, $H_{n+1}^{(L)}$ and the new Left operators are expressed in the new Left block basis. Similarly, the operators in the new Right block are formed.

2. *Diagonalization.* The Hamiltonian matrix is a very large matrix, and thus to diagonalize it, one uses a matrix diagonalization procedure such as the Davidson [81, 82] or Lanczos [83, 84, 85] algorithm. In most common usage, the ground state is the target state although is also possible to target some of the lowest excited states. The dimension of the diagonalization space is $N^L N^R$. In such a space, the wave function of the superblock is found after diagonalization of the Hamiltonian matrix of the superblock that never has to be constructed explicitly.

3. *Construction of the new optimal basis.* One constructs the density matrix for the Left block, $\hat{\rho} = \mathrm{Tr}_E |\Psi\rangle\langle\Psi|$, by tracing out the states from the Right block. Next, such a density matrix is diagonalized and the eigenvalues w_α, and corresponding eigenvectors $|w_\alpha\rangle$, are sorted in descending eigenvalue order. The M^L eigenvectors corresponding to the largest eigenvalues are chosen to create the new basis vectors in the Left block which contains $n+1$ orbitals.

4. *Transformation/Decimation.* One constructs the transformation matrix O_L, which is a rectangular $(N^L \times M^L)$–dimensional matrix, with the matrix elements being given by $\langle m_n^L \sigma | m_{n+1}^L \rangle$, which are the eigenvectors of the reduced density matrix. Next, the Left Hamiltonian and all the matrices of the operators from Left block of dimension $M^L N_{\mathrm{orb}}$ are expressed in the new many-body basis consisting of the most important states of dimension M^L. This is done by truncating the basis by the help of the transformation matrices

$$A^{\mathrm{truncated}} = O_L^T A O_L. \tag{2.21}$$

5. *Saving.* After the transformation, the Hamiltonian matrix $H^{(l)}_{(n+1)}$ and all operator matrices are saved on the disk, and one starts again from point 1.

The above procedure is illustrated in Fig. 2.2.

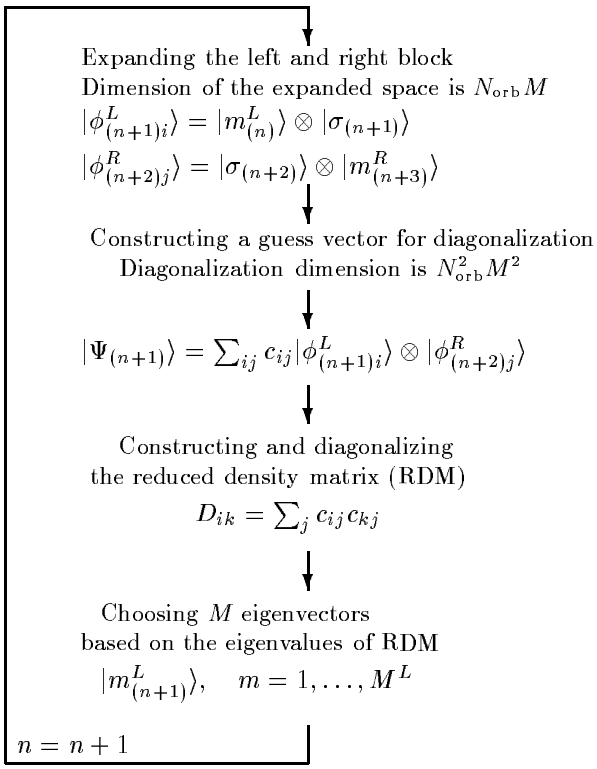

Figure 2.2: Operations performed during a single step of the DMRG algorithm. We assume that there are $n + 1$ orbitals in the expanded Left space and $m - n - 1$ orbitals in the expanded Right space.

Note that during the blocking procedure one orbital was added to the left block, thus expanding the number of orbitals in this block from n to $n+1$. During the subsequent blocking step one more orbital is added to the left block. This continues until the Right block reaches a dimension that can be exactly described by M^R states. This growth of the Left block and shrinkage of the Right block is called a *Left–to–Right sweep*. When the Right block reaches some minimum size and becomes exact, the growth direction is reversed and the *Right–to-Left sweep* starts. During this sweep the Right block is growing while the Left block is shrinking. To achieve convergence, we sweep several times from one side to the other. Here, we summarize briefly how the matrix representation of the operators needed for the construction of the superblock Hamiltonian is handled during subsequent sweeps.

1. *Left–To–Right.* The number of orbitals in the Left block is increased by one orbital. In this new basis, the Hamiltonian matrix as well as the matrices of the operators from the Left block are built explicitly. The Hamiltonian and the operators from the Right block , which were produced in the previous sweep, are read from the disk. In the Left–to–Right phase of the algorithm, all newly obtained matrices and the Hamiltonian from the Left block are saved on the disk for use in the next Right–to–Left sweep .

2. Orbitals are added until the Left block reaches the size $n = m - k$, where m is the total number of orbitals, and k is the number of orbitals possible for a space described exactly by M^R orbitals.

3. *Right-To-Left.* In the last step, the Right block has k orbitals. We start to increase the number of orbitals in the Right block and decrease the size of the Left block. In the Right–to–Left part of the algorithm, all the operators and the Hamiltonian from the Right block are stored, and the matrices needed in the Left block are those from the previous Left–to–Right sweep.

4. Orbitals are added until the Right block reaches the size $n = m - k$, and then the algorithm returns to performing a Left–to-Right sweep.

During the sweeping stage of the algorithm, the energy should converge to a nearly constant value. This occurs due to the improvements of the basis states that are used to obtain the Hamiltonian matrix elements. In the DMRG method, we start from a very poor basis in the Right block, and consequently the sweeping stage of the algorithm helps to recover from the poor starting point and eliminates the environmental error. The number of sweeps needed to achieve convergence may vary from a few to several dozens, depending on the M value and the complexity of the system of interest. Usually, the quality of DMRG calculation cannot be judged simply based on the results for a single M value, and one needs to carry out and judge the quality of the calculation for various M values. A pictorial description of the algorithm is shown in Fig. 2.3

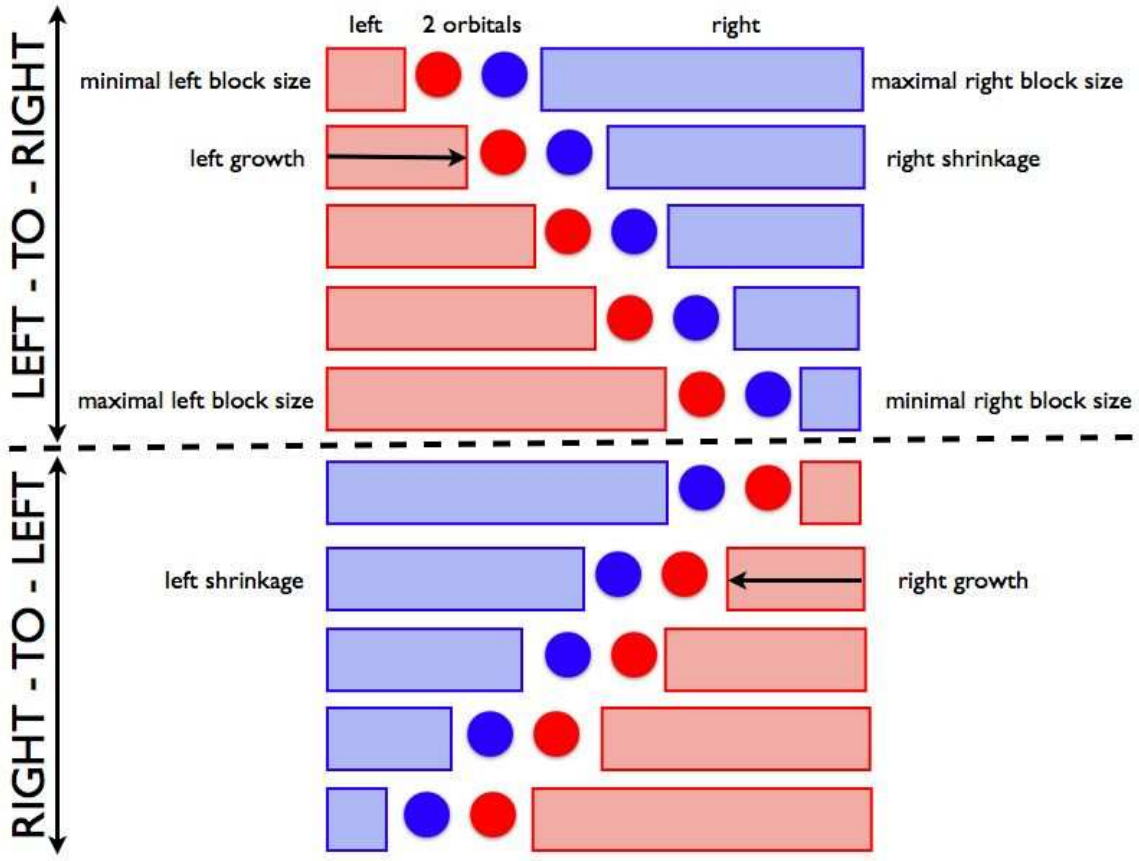

Figure 2.3: A schematic illustration of a the sweeping stage of a DMRG run.

2.4 Implementation details

In the previous sections, the DMRG method was discussed in general terms. We will now focus on some implementation problems that are important for understanding the performance of the DMRG method. We will start by explaining the initialization procedure needed to start the DMRG calculation. Next, we will briefly discuss the blocking step and the explicit matrix representation of the second quantized operators used in DMRG. We will then explain how the starting vector for the diagonalization procedure is found, and finally, how the diagonalization of the sparse superblock matrix can be done in an efficient manner.

2.4.1 Initialization

In contrast with other methods in quantum chemistry, in DMRG one works with the matrix representation of the second–quantized operators. This means that every operator used to construct the Hamiltonian matrix needs to be initially expressed in the basis containing determinants in the occupation number representation. To initialize the matrices of the second quantized operators for the left part is simple since one starts by taking only few orbitals (usually only one) into the left orbital space. In such a small orbital space, all the determinants spanning the complete space are known, and one can express all needed operators in such a basis. Here, we give an example of a matrix form of the $a^\dagger_\uparrow$, $a^\dagger_\downarrow$ and $a^\dagger_\uparrow a_\downarrow$ operators

$$\mathbf{a}^\dagger_\uparrow = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \mathbf{a}^\dagger_\downarrow = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \quad \mathbf{a}^\dagger_\uparrow \mathbf{a}_\downarrow = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \tag{2.22}$$

in the basis for a single orbital with four possible occupancies $\{|\ \rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$ that are the row and column basis states. These operators fulfill the local anticommutation relation, $\boldsymbol{a^\dagger_\sigma a_\sigma + a_\sigma a^\dagger_\sigma = I\delta_{\sigma\sigma'}}$, where I is the 4×4 identity matrix. Using these matrices one expresses the Hamiltonian operator and other operators from the left block. In the case of the Right block,

the situation is more complicated because one cannot use the exact basis, and one has to take only a small subset of determinants with a number of electrons and spins that sum up with the states from the Left block to give a proper total number of electrons and proper total S_z. Usually the determinants chosen for the initial Right block basis have the lowest energy orbitals occupied, and are thus closest to the Hartree–Fock determinant. The matrix elements between these previously chosen states are calculated to obtain the approximate operator matrices in the Right block. Note too that due to the incompleteness of the basis in the Right block, one cannot use the resolution of the identity to calculate a multiple label operator (e.g. $a_i^\dagger a_j$) as a product of simple a_i and a_j operators. The many–index operators need to be calculated explicitly. The initialization procedure that preserves the spin– and spatial–symmetry of the problem will be discussed in the next chapter. The potential difficulties and convergence problems that may result from a poor guess of states in the Right block will be addressed in the chapter dealing with local minima in the DMRG method.

2.4.2 Blocking

In this step, a full basis of one orbital, O_L, consisting of four possible occupancies of a single orbital, $\{|\quad\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$, is added to the optimally contracted basis states obtained in the previous step for the left block. One constructs new states by combining the quantum numbers and restricting them to be consistent with the proper number of electrons and proper S_z

$$\begin{aligned} L \otimes O_L \quad &= \{|L_1\rangle, |L_2\rangle, |L_3\rangle, \ldots, |L_n\rangle\} \otimes \{|\quad\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\} \\ &= \{|L'_1\rangle, |L'_2\rangle, |L'_3\rangle, \ldots, |L'_{4n}\rangle\}. \end{aligned} \tag{2.23}$$

Next, we express the operators in this newly created product basis. Basically, one encounters two cases that need to be handled. The first one is when some of the indices of such many–index operator belong to the orbitals present in unexpanded block. One then takes the direct product of the operator matrices a_i from the left block with the matrices that belong to the added orbital a_j. In this way, one constructs all the matrix representation of operators in the expanded Left

block explicitly in the space of all states spanned by $L^{'} = L \otimes O_L$. We will use $\hat{O}$ to denote an operator that acts only on the left block, and $\hat{a}$ to denote an operator that acts only on the basis of the added orbital. The direct product of such operators is then

$$\widetilde{a_i a_j} = a_i \otimes a_j, \tag{2.24}$$

in which i identifies an orbital in the unexpanded left block and j labels the added orbital. The above expression can be generalized as,

$$\langle l_\mu o_\nu | \hat{O}\hat{a} | o_\kappa l_\lambda \rangle = \langle l_\mu | \langle o_\nu | \hat{O}\hat{a} | o_\kappa \rangle | l_\lambda \rangle = \langle l_\mu | \hat{O} | l_\lambda \rangle \langle o_\nu | \hat{a} | o_\kappa \rangle \sigma(o_\nu, \hat{O}), \tag{2.25}$$

in which $\sigma(o_\nu, \hat{O})$ denotes the parity factors responsible for the sign changes associated with the commutation rules of the second quantized operators. The symbols l_μ and l_λ label the basis states from the left block before expansion, while o_ν and o_κ label the states of the added orbital.

The second case that needs to be considered is when all the indices belong either to the unexpanded block or to the added orbital. It is important to notice that operators belonging to the left block are many–index operators, and they have to be expanded in this form and cannot be recreated from products of one–index operator due to the truncation of the resolution of identity that is performed during the decimation procedure. Consequently, an expansion of a many–index operator (from 1– to 3–indices) that acts only on the left unexpanded states and does not have any labels that act on the added orbital can be expressed as

$$\widetilde{\widetilde{a_k a_l}} = \widetilde{a_k a_l} \otimes \delta, \tag{2.26}$$

if k and l are restricted to the unexpanded block. Expressing the above equation in more general terms gives

$$\langle l_\mu o_\nu | \hat{O} | o_\kappa l_\lambda \rangle = \langle l_\mu | \langle o_\nu | \hat{O} | o_\kappa \rangle | l_\lambda \rangle = \delta_{o_\nu o_\kappa} \langle l_\mu | \hat{O} | l_\lambda \rangle \sigma(o_\nu, \hat{O}), \tag{2.27}$$

where $\sigma(o_\nu, \hat{O})$ are parity factors. Similarly, the expansion of the operators acting on the added

orbital can be expressed as

$$\langle l_\mu o_\nu|\hat{a}|o_\kappa l_\lambda\rangle = \langle l_\mu|\langle o_\nu|\hat{a}|o_\kappa\rangle|l_\lambda\rangle = \langle o_\nu|\hat{a}|o_\kappa\rangle\delta_{l_\mu l_\lambda}, \tag{2.28}$$

where one does not need to take into account the parity factor since $\langle o_\nu|\hat{a}|o_\kappa\rangle$ is just a matrix element, and it can be evaluated inside the bracket that later produces the $\delta_{l_\mu l_\lambda}$ function.

2.4.3 Diagonalization step

The key to understanding the performance and computational scaling of the DMRG algorithm lies in the diagonalization of the superblock Hamiltonian. There exist numerical methods that allow one to diagonalize large sparse matrices, and such methods rely on iterative multiplications of the superblock Hamiltonian matrix and the guess vector. The most commonly used among such methods are Davidson [81] and Lanczos [83] sparse matrix diagonalization methods. These methods allow one to calculate a few eigenvalues that are at the edge of the full spectrum.

In DMRG, the superblock matrix has the number of rows and columns equal to $(N_{\rm orb}^2 M^2)$. This dimensionality can be explained if one notices that the superblock Hamiltonian H_{LR} can be built as a direct product of the following form

$$H_{LR} = H_L \otimes \mathbf{1}_R + I_L \otimes I_R + \mathbf{1}_L \otimes H_R, \tag{2.29}$$

where H_L and H_R are the $N_{\rm orb}M$ dimensional Hamiltonian matrices in the expanded Left and Right blocks, respectively. The $I_L \otimes I_R$ term describes the interaction between the Left and Right blocks, i.e., the terms in the superblock Hamiltonian that have mixed indices belonging to both Left and Right block. The Left and Right Hamiltonians are matrices that are stored on disk and updated at every expansion step, which is then followed by the decimation step.

One can ask why one does not store the Hamiltonian matrix elements in the form $\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_l \hat{a}_k$ from the Left and Right blocks. To store such matrix elements would require $O(M^2 k^4)$ of the disk space since all the four index operators have to be stored because the truncation does not preserve the resolution of identity, and the 4–index operators cannot be recreated out of 1–index ones.

Consequently, we would need to keep k^4 operators that are $M \times M$–dimensional; this explains the above disk storage requirement. However, if one prepares the Left or Right Hamiltonian matrix that has all the 4-index operators summed up with integrals, and after every expansion such a Left or Right Hamiltonian matrix is updated, we then avoid the storage of the 4–index operators. The only matrices of operators that need to be stored are then the ones required to construct the interaction term between the Left and Right blocks. Since the Hamiltonian matrix element has 4–indices, it can be constructed from two 2–index operators or a single 1–index operator with 3–index operator. However, one has to notice that the matrix elements of the Hamiltonian are multiplied with integrals. Consequently, the following *complementary operators* [19] can be used to "assemble" the interaction term between the Left and Right blocks:

$$P^R_{ij} = \sum_{kl \in L} v_{ijkl} a_k a_l \tag{2.30}$$

$$Q^R_{ij} = \sum_{kl \in L} x_{ijkl} a^\dagger_k a_l, \quad x_{ijkl} = v_{ijkl} - v_{jikl} - v_{ijlk} + v_{jilk} \tag{2.31}$$

$$R^R_i = \sum_{jkl \in L} w_{ijkl} a^\dagger_j a_k a_l, \quad w_{ijkl} = v_{ijkl} - v_{jikl} \tag{2.32}$$

$$S^R_i = \sum_{j \in L} t_{ij} a_j \ . \tag{2.33}$$

$$\tag{2.34}$$

The interaction term then can be expressed as

$$I_L \otimes I_R = \sum_{i \in L} a^\dagger_i S^R_i + \sum_{ij \in L} (a^\dagger_i a^\dagger_j P^R_{ij} + a^\dagger_i a_j Q^R_{ij}) + \sum_{i \in L} a^\dagger_i R^R_i + \sum_{i \in L} a^\dagger_i R^L_i + \text{adjoint.} \tag{2.35}$$

Hence, one requires to store only 2– and 1–index operators on the disk.

Let us discuss now the multiplication of the superblock Hamiltonian by a vector. If the Hamiltonian matrix was constructed explicitly and multiplied by a vector, then the cost of such a multiplication would scale as (column size)$\times$(row size), making it proportional to $O(M^4)$. However, in DMRG, it is only $O(M^3)$. Such a lower scaling is achieved by taking advantage of the direct product structure of the superblock Hamiltonian when the multiplication by a vector is performed.

Consider an operator that has indices belonging to the Left block and Right block, e.g. $I = I_L \otimes I_R$. If we express such an operator in the product basis of the superblock $\{|L_\mu R_\kappa\rangle\}_{\mu=1,\ldots,N_{\mathrm{orb}}M;\kappa=1,\ldots,N_{\mathrm{orb}}M}$, we obtain

$$\langle R_\lambda L_\nu|I|L_\mu R_\kappa\rangle = \langle R_\lambda|\langle L_\nu|I_L I_R|L_\mu\rangle|R_\kappa\rangle = \langle L_\nu|I_L|L_\mu\rangle\langle R_\lambda|I_R|R_\kappa\rangle\sigma(I_R, L_\mu). \tag{2.36}$$

Then, one can express the multiplication by a vector as

$$\sum_\mu \sum_\kappa \langle R_\lambda L_\nu|I|L_\mu R_\kappa\rangle c_{\mu\kappa} = \sum_\mu \sum_\kappa \langle L_\nu|I_L|L_\mu\rangle\langle R_\lambda|I_R|R_\kappa\rangle\sigma(I_R, L_\mu)c_{\mu\kappa}, \tag{2.37}$$

where one can perform the multiplication by the vector in two steps, thus making it more efficient. The first multiplication performed would be

$$\sum_\kappa \langle R_\lambda|I_R|R_\kappa\rangle c_{\mu\kappa} = B^R_{\lambda\mu}, \tag{2.38}$$

and the intermediate matrix $B^R_{\lambda\mu}$ produced can be used in a subsequent multiplication

$$\sum_\mu \langle L_\nu|I_L|L_\mu\rangle B^R_{\lambda\mu}\sigma(I_R, L_\mu) = \tilde{c}_{\nu\lambda}. \tag{2.39}$$

The matrices of the Hamiltonians of the Left and Right block can be multiplied by a vector in an analogous manner. Consequently, for a single interaction term the multiplication of the superblock Hamiltonian matrix by a vector is changed to two matrix multiplications, with scaling proportional to $O(M^3)$. There are of order $O(k^2)$ operators that are required to construct the interaction term between the Left and Right block. Hence, the overall scaling of a single multiplication of the superblock Hamiltonian matrix by a vector is proportional to $O(M^3k^2)$.

2.4.4 The Guess vector

We learned previously that the cost of a single multiplication of the superblock Hamiltonian matrix by a vector is proportional to $O(M^3k^2)$. However, due to the iterative nature of the

diagonalization method used for large sparse matrices, one may need hundreds of iterations – multiplications of the Hamiltonian matrix by a vector, if one starts from a random vector. Thus, it is important to provide a good starting vector for the diagonalization procedure since such a starting point can lower the number of iterations drastically and speed up the the algorithm, even up to one order of magnitude.

In the finite DMRG algorithm, the total Hilbert space does not change since it is simply spanned by N_{orb}^k states, where N_{orb} is the number of possible states for one orbital and k is the total number of orbitals. However, the effective Hilbert space of dimension $N_{\mathrm{orb}}^2 M^2$ in which we perform the calculation changes at every step of the sweep. In the case of exact FCI calculations, there is a unique mapping between the wave functions produced at each step of a sweep; in this case the transformation of the wave functions between steps would be exact. If one considers the incomplete space calculations, there is no no exact transformation that can transform the wave function exactly between steps of sweep. Nevertheless, the DMRG calculations are based on the assumption that truncation still gives an almost exact wave function. Hence, one can expect that the incomplete transformation of such a wave function will still result in almost complete overlap of the wave function between the steps of a sweep, consequently leading to a huge reduction of the number of iterations in the diagonalization procedure due to use of a very good starting vector.

The method of transforming the wave function from step to step of a sweep was presented by White [86]. Let us assume that we have n orbitals in the Left block and $k-n-2$ orbitals in the Right block, where k is the total number of orbitals. We previously obtained a wave function $\langle m_n \sigma_{n+1} \sigma_{n+2} m_{k-n-2} | \Psi \rangle$ at the nth step, and we would like to transform it to the $(n+1)$th step to provide a guess vector $\langle m_{n+1} \sigma_{n+2} \sigma_{n+3} m_{k-n-3} | \Psi \rangle$ for the diagonalization procedure. From the current nth step of the Left–to–Right sweep, we also know the transformation matrices O_L that we will denote now as $\langle m_n \sigma_{n+1} | m_{n+1} \rangle$, thus indicating that they truncate the $N_{\mathrm{orb}} M$–dimensional basis back to M states. We also assume that from the previous Right–to–Left sweep we know the O_R transformation matrix $\langle m_{k-n-3} \sigma_{n+3} | m_{k-n-2} \rangle$. We can now perform the first incomplete

basis transformation that transforms the Left block and the added orbital to a new Left block

$$\langle m_{n+1}\sigma_{n+2}m_{k-n-2}|\Psi\rangle = \sum_{m_n\sigma_{n+1}} \langle m_{n+1}|m_n\sigma_{n+1}\rangle\langle m_n\sigma_{n+1}\sigma_{n+2}m_{k-n-2}|\Psi\rangle. \tag{2.40}$$

The next transformation uses the result of the first one and expands the current Right block into a Right block with one orbital less plus a single orbital, namely

$$\langle m_{n+1}\sigma_{n+2}\sigma_{n+3}m_{k-n-3}|\Psi\rangle = \sum_{m_{k-n-2}} \langle m_{k-n-3}\sigma_{n+3}|m_{k-n-2}\rangle\langle m_{n+1}\sigma_{n+2}m_{k-n-2}|\Psi\rangle. \tag{2.41}$$

Consequently, the new predicted wave function vector is produced, and it can be used as a starting vector in the new $(n+1)$th step of a sweep. The cost of producing such a starting vector is $O(M^3)$ since it involves two matrix multiplications. The iteration number needed for the diagonalization procedure is usually reduced to less than 10 iterations, in comparison with ≈ 100 when no starting vector prediction is used.

2.4.5 Computational cost

Finally, we are prepared to consider the computational scaling of the DMRG algorithm. Let us first discuss the disk storage required. As explained in an earlier section, due to the efficient summation of the integrals with 3– and 4–index operators, we need to store only 1– and 2–index operators. Consequently, the disk storage for a single step scales as $O(M^2k^2)$. However, we need to store these 1– and 2–index operators for every step within a sweep, and thus the final disk storage is $O(M^2k^3)$. To be able to carry out the calculation efficiently for every step, it is good to have some of the operators in the direct access memory. The optimal solution is to have $O(M^2k^2)$ memory available to fit the complete operator array from a single step. However, sometimes this may become difficult, and it is possible then to make more use of disk, where the sequential batches of operators are read from the disk to the memory. In such a case, it is possible to perform DMRG calculations when as few as four $O(M^2)$ operators can be fitted in memory, thus limiting the memory requirements to $O(M^2)$. Nevertheless, to decrease the memory usage and replace it

with disk requires a data structure that allows us to minimize number of disk reads and writes. We list the memory and disk usage in Table 2.1.

Table 2.1: Disk and memory usage in DMRG

Disk	Memory
$O(M^2k^3)$	$O(M^2k^2)$

Let us now consider the operation count. The three most important steps are: blocking, diagonalization of the Hamiltonian matrix, and decimation. The scaling of a single blocking step is proportional to $O(M^2k^3)$. Such a scaling can be explained by noticing that all the matrices in the blocking step have M^2 elements, and to expand some of the complementary operators, we need to have three loops over the orbital indices, thus yielding $O(k^3)$. The next important step in the DMRG algorithm, diagonalization of the Hamiltonian matrix, scales as $O(M^3k^2)$ since the multiplication of the Hamiltonian matrix by a vector can be factorized into $O(k^2)$ multiplications of $M \times M$–dimensional matrices. The decimation step is entirely based on matrix multiplication that scales as $O(M^3)$ and has to be performed for all the $O(k^2)$ operators that are needed for the next blocking step and diagonalization. Hence, the cost of a single decimation step is $O(M^3k^2)$. All the operation counts given above were calculated for a single step within a sweep; however, we have $O(k)$ steps within a sweep, and consequently all the scaling has to be multiplied by k. All the operations counts are reported in Table 2.2.

Table 2.2: Operation count in DMRG

Renormalization transform	Steps	Total
blocking $O(M^2k^3)$	$O(k)$	$O(M^2k^4)$+$O(M^3k^3)$
Davidson $O(M^3k^2)$		
decimation $O(M^3k^2)$		

Finally, we note that usually the number of states kept M is larger than the number of orbitals k; thus, the term $O(M^3k^3)$ prevails in the scaling. Conversely, it is also possible for some systems to achieve almost constant M when the number of orbitals k is growing. In such situations M^2 or M^3 can be treated as a prefactor making the scaling proportional to $O(k^4)$.

Chapter 3

Symmetry– and spin–adapted DMRG

After discussing the basic considerations of the DMRG algorithmic structure, we are prepared to learn about the inclusion of basic symmetries in the DMRG algorithm. Symmetries that are preserved by the DMRG algorithm can be exploited to reduce storage and computation time. Moreover, because of the decomposition of the Hilbert space into many symmetry sectors, the lowest-lying states from those sectors are accessible to the diagonalization procedure in the same way as the ground state. Let us also mention that aside from pure efficiency reasons, the inclusion of symmetries is also physically motivated. There are many interesting systems, such as molecular magnets or active centers of metalloproteins, that are very multiconfigurational and for which the different spin states are very close–lying. Thus, the DMRG method is a natural candidate to handle the multiconfigurational nature of these system; however, the DMRG method must also be able to target their spin and spatial symmetries correctly. This chapter will present results for the DMRG method with spin and spatial symmetry adaptation. Our aim is to explain in a detailed manner the modification of our current version of the DMRG algorithm in comparison to the traditional DMRG implementation. We stress that our algorithm does not change the overall $O(M^3k^3)$ scaling of the method, and that it is relatively simple to introduce into a general DMRG

code. In Sec. 3.1, we present a brief summary of symmetry issues in DMRG. In Sec. 3.2, we will then discuss the standard DMRG algorithm in order to introduce necessary terminology that will be used in Sec. 3.3 to describe our version of the spin–adapted DMRG algorithm in detail. All the newly introduced steps will be listed and compared to the previously existing DMRG algorithm. In Sec. 3.4, we will present a brief summary of previous spin–adaptation schemes and discuss the advantages and drawbacks of possible modifications. In Sec. 3.5, we will then show results of calculations performed using our DMRG code for a model system, HNCO, and for methylene CH_2, calculating several states for both systems. We compare our results with the FCI results for these systems. Finally, Sec. 3.6 presents conclusions in the context of further development of the DMRG method.

3.1 Symmetries and conserved quantum numbers in DMRG

3.1.1 Abelian symmetries

In the DMRG method the orbital subspace is divided into four subspaces $(l), (l_s), (r_s)$, and (r) corresponding to right block, right site, left site and left block, respectively; see Fig. 3.1. Such a division is reflected in the wave function form

$$|\Psi\rangle = \sum_{\mu' i \nu' j} c_{\mu' i \nu' j} |\tilde{L}_{\mu'}\rangle |i\rangle |j\rangle |\tilde{R}_{\nu'}\rangle = \sum_{\mu\nu} c_{\mu\nu} |L_\mu\rangle |R_\nu\rangle, \tag{3.1}$$

in which $\tilde{L}_{\mu'}$, i, j, and $\tilde{R}_{\nu'}$ denote states characteristic of the subspaces $(l), (l_s), (r_s)$, and (r), respectively. The L_μ and R_ν states have composite μ and ν indices which result from the direct product between (l) and (l_s) vector spaces and similarly between (r_s) and (r). The structure of the above wave function gives rise to a natural Abelian symmetry blocking within the DMRG scheme since the total number of electrons N_{tot}, the total number of α electrons N_α, and the total spatial symmetry of the wave function ir, is created by the sum of the electrons in the Left and Right blocks (in the case of N_{tot} and N_α) or the product of the symmetry labels from the Left $|L_\mu\rangle$ and Right $|R_\nu\rangle$ states. Rephrasing this argument, if one knows the total number

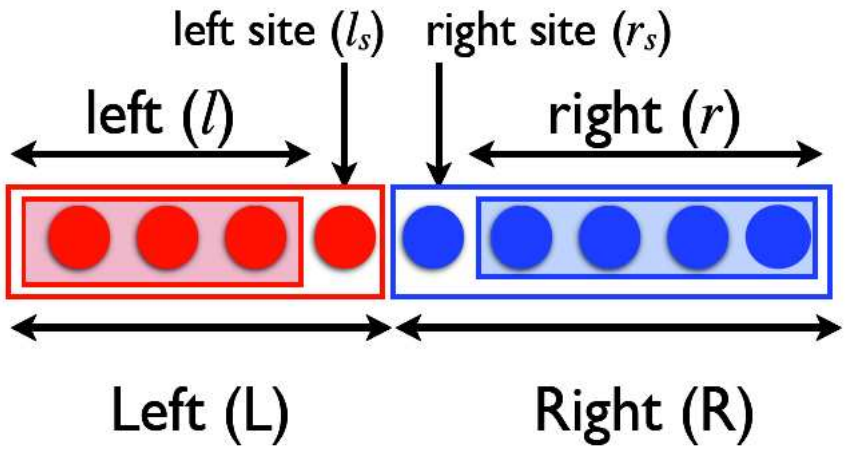

Figure 3.1: Orbital spaces in DMRG. The Left (Right) orbital space is obtained as a result of the direct product between the left (right) space and the l_s (r_s) site.

of electrons, the total number of α electrons, and the symmetry of the $|L_\mu^{T_L}\rangle$ state and of the total wave function, then one can uniquely predetermine the appropriate label of the Right state $|R_\nu^{T-T_L}\rangle$. In particular

$$|\Psi^T\rangle = \sum_{T_L} \sum_{\mu\nu} c_{\mu\nu}^{T_L} |L_\mu^{T_L}\rangle |R_\nu^{T-T_L}\rangle, \quad T = N_{\text{tot}}, N_\alpha, ir, \tag{3.2}$$

where N_{tot} is the total number of electrons, N_α is the total number of α electrons and ir is an index representing the irreducible representation of the wave function. The T_L label denotes the quantities for the Left block and $T_R = T - T_L$ those for the Right block. Thus, due to this one–to–one correspondence of the labels between the Left and Right parts, inclusion of Abelian symmetries in the DMRG code is relatively straightforward. It means that the vectors $|L_\mu\rangle$ and $|R_\nu\rangle$ can be blocked according to the specific T_L and $T - T_L$. Such a block will be later referred to as a partition. Thus, all operators used in DMRG can be expressed in a matrix representation as dense blocks of nonzero matrix elements with all other matrix elements equal to zero. The blocking of the reduced density matrix is an important consequence of the above statement. Since the transformation of the basis into the eigenbasis of the reduced density matrix present in DMRG

preserves the block structure of the operators, only the dense blocks of nonzero elements have to be stored for all the operators contributing to the Hamiltonian.

3.1.2 Non-Abelian symmetries

The total spin $\hat{S}^2$ is another conserved symmetry. However, knowing the spin symmetry label of a state in the Left part and the total spin of the wave function does not uniquely determine the spin symmetry of the state in the Right part which will couple with the Left state to the appropriate spin symmetry of the total wave function,

$$|\Psi^S\rangle = \sum_{S_L,S_R} \sum_{\mu\nu} c^S_{\mu\nu} |L^{S_L}_\mu\rangle |R^{S_R}_\nu\rangle, \tag{3.3}$$

where all the S_R values such that $S = |S_L + S_R|, \ldots, |S_L - S_R|$ are possible. Thus, there is no one–to–one correspondence between the spin symmetry states in the Left and Right parts. The above fact creates an additional complication in that the $\hat{S}^2$ operator does not commute with the reduced density matrix calculated in DMRG. Thus, a straightforward diagonalization will lead to a mixed basis. The spin–adaptation of the DMRG algorithm is a complicated procedure, and the first rigorous algorithm for including SU(2) spin symmetry, a non-Abelian symmetry, was designed by McCulloch and Gulácsi [87, 88, 89]. This was done for the Kondo Lattice model [90] frequently studied in the solid state physics community. The McCulloch and Gulácsi's implementation of spin symmetry was rather complicated, and was based on the Clebsch-Gordan transformation and the elimination of quantum numbers by the use of the Wigner-Eckart theorem. In our implementation, we ensure that complete multiplets are included in the wave function expansion (see Eq. (3.3)), while the diagonalization procedure targets pure spin states of interest (for details see Sec. 3.3). Such a scheme allows us to avoid the complication of the Clebsch-Gordan transformation at the price of a less efficient computational procedure. To the best of our knowledge, this is the first full spin–adaptation of the DMRG method developed for a general two–body Hamiltonian and applied to chemical systems.

3.2 The DMRG algorithm

The object here is to introduce only a basic description and the notation necessary for understanding the modifications of the spin-adapted version of the DMRG algorithm described in the next section. The first description of the DMRG algorithm was given by White [72]; and for an excellent and detailed later description from a more chemical perspective, see Chan and Head-Gordon [22].

In the DMRG method one separates the spatial orbital space into four orbital subspaces which we will denote as a left block (l), left site (l_s), right site (r_s) and right block (r), respectively, as shown in Fig. 3.1. We will later call Left the bigger expanded vector space created from a direct product of (l) with (l_s), and analogously we will call Right the vector space resulting from a direct product between (r_s) and (r). To make the notation more precise, by "left" we mean the smaller orbital space before taking the direct product with the site subspace to yield the final expanded Left subspace. The many-body basis states are constructed iteratively in the Left and Right parts of the orbital subspace as

$$|L_\mu^{T_L}\rangle = \sum_i d_{i\mu}^{T_L} |\Psi_i^{T_L}\rangle, \quad |R_\nu^{T_R}\rangle = \sum_j d_{j\nu}^{T_R} |\Psi_j^{T_R}\rangle, \tag{3.4}$$

where $|\Psi_i^{T_L}\rangle$ and $|\Psi_j^{T_R}\rangle$ are all possible Slater determinants in the Left and Right orbital blocks, respectively. The superscripts T_L and T_R refer to all possible partitions created by considering a particular assignment of electrons between the Left and Right subspaces, as well as the symmetry labels of the orbitals.

The main goal of the DMRG method is to iteratively find the most compact many-body basis, or $d_{i\mu}^{T_L}$ and $d_{j\nu}^{T_R}$ coefficients, which for a given number of μ and ν states will optimize the wave function of the whole system. In the first step of the algorithm, the left subspace contains only one orbital and the basis states $|\tilde{L}_\mu^{T_l}\rangle$ can be expressed trivially as, $\{|\ \ \rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$, the four possible occupations of one spatial orbital. All the remaining orbitals, except for the two site orbitals, are in the right subspace where the basis states $|\tilde{R}_{\nu'}^{T_r}\rangle$ need to be approximated by a set of chosen Slater determinants. The iterative optimization employs a blocking step that simply

moves one spatial orbital with all possible occupancies from the Left subset to the Right subset, or vice versa

$$|L_\mu^{T_l}\rangle = |\tilde{L}_{\mu'}^{T_l}\rangle| \quad \rangle, \tag{3.5}$$

$$|L_\mu^{T_l+1\alpha}\rangle = |\tilde{L}_{\mu'}^{T_l}\rangle| \uparrow\rangle, \tag{3.6}$$

$$|L_\mu^{T_l+1\beta}\rangle = |\tilde{L}_{\mu'}^{T_l}\rangle| \downarrow\rangle, \tag{3.7}$$

$$|L_\mu^{T_l+1\alpha+1\beta}\rangle = |\tilde{L}_{\mu'}^{T_l}\rangle| \uparrow\downarrow\rangle. \tag{3.8}$$

Thus, by a sequence of blocking steps all possible Slater determinants can be generated and introduced into the many-body basis. The total wave function of the system is then obtained by Davidson diagonalization of the Hamiltonian matrix expressed in the product basis of the Left and Right blocks. The total wave function can be written as

$$|\Phi_{\text{tot}}\rangle = \sum_{\mu\nu}\sum_{T_L} c_{\mu\nu}^{T_L}|L_\mu^{T_L}\rangle|R_\nu^{T_R}\rangle, \tag{3.9}$$

where $T_L + T_R = T$.

The next step, decimation, can be considered the most essential part of the method. In this step one chooses M basis states ($|\tilde{L}_{\gamma'}^{T_L}\rangle$, where $\gamma' = \{1,\ldots,M\}$), from the set $\{|L_\mu^{T_L}\rangle\}$, where $\mu = \{1,\ldots,4M\}$. The wave function then can be expressed in the truncated many-body basis as

$$|\Phi_{\text{truncated}}\rangle = \sum_{\gamma'=1}^{M}\sum_{\nu'=1}^{M}\sum_{T_L} c_{\gamma'\nu'}^{T_L}|\tilde{L}_{\gamma'}^{T_L}\rangle|\tilde{R}_{\nu'}^{T_R}\rangle. \tag{3.10}$$

In the following steps, one keeps moving the orbitals from the Right to the Left block until only one orbital remains in the Right subspace. Then, one starts the procedure from the beginning and moves the orbitals from the Left to the Right subspace.

A procedure defined in such a way allows one to construct M many-body basis states that are linear combinations of all possible determinants created in a given Hilbert space. The DMRG many-body basis can be compared to the configuration state functions (CSFs) used in the con-

tracted multireference configuration interaction (MRCI) approaches. However, one has to keep in mind that in MRCI methods the CSFs are contracted over determinants which have the same orbital excitation pattern. In DMRG, the contraction runs over all possible determinants within partitions T_L and T_R in the Left and Right orbital subspaces, respectively. These contractions are independent of the excitation level of the determinants from a chosen reference since there is no reference determinant in the DMRG scheme. Thus, the method is general and equally well suited for treating single and multireference cases. Realizing the above facts, we may call DMRG a general, truncated, contracted, iterative configuration interaction method with an ansatz for the wave function given by Eq. (3.9) and the truncation of the basis defined in Eq. (3.10). The above description of the DMRG algorithm explains the construction of the DMRG many-body basis, as well as the form of the final wave function. In practice, however, one never constructs the explicit form of the many-body basis in terms of the primitive basis of Slater determinants (except perhaps for the analysis of the DMRG wave function in terms of Slater determinants that was recently reported by Moritz and Reiher [31]). The information about the basis is hidden in the matrix representation of the second–quantized operators used for the on–the–fly construction of the Hamiltonian matrix required by the Davidson diagonalization procedure.

3.3 The symmetry- and spin-adapted DMRG algorithm

We will first present a general description of our spin–adaptation scheme which will later be followed by a more detailed account of the changes introduced into the standard DMRG method. The scheme for a single step of the spin–adaptation procedure may be found in Fig. 3.2. Our spin–adaptation procedure is based on two essential conditions. The first is that the basis states span the complete manifold of the $\hat{S}^2$ operator. It means that all states $\{|L_\lambda, S_L, i\rangle\}_{i=-S_L,\ldots,S_L}$ in the Left block (and similarly in the Right block) that can be produced by acting with the $\hat{S}_L^-$ operator on a state $|L_\lambda, S_L, S_L\rangle$ with the highest S_L^z, need to be present in the basis. An immediate consequence of this first condition is that if we select a state $|L_\lambda, S_L, i\rangle$ to be included in the basis, we also have to include all the remaining states $|L_\lambda, S_L, j\rangle$ for $j = -S_L, \ldots, S_L$, and

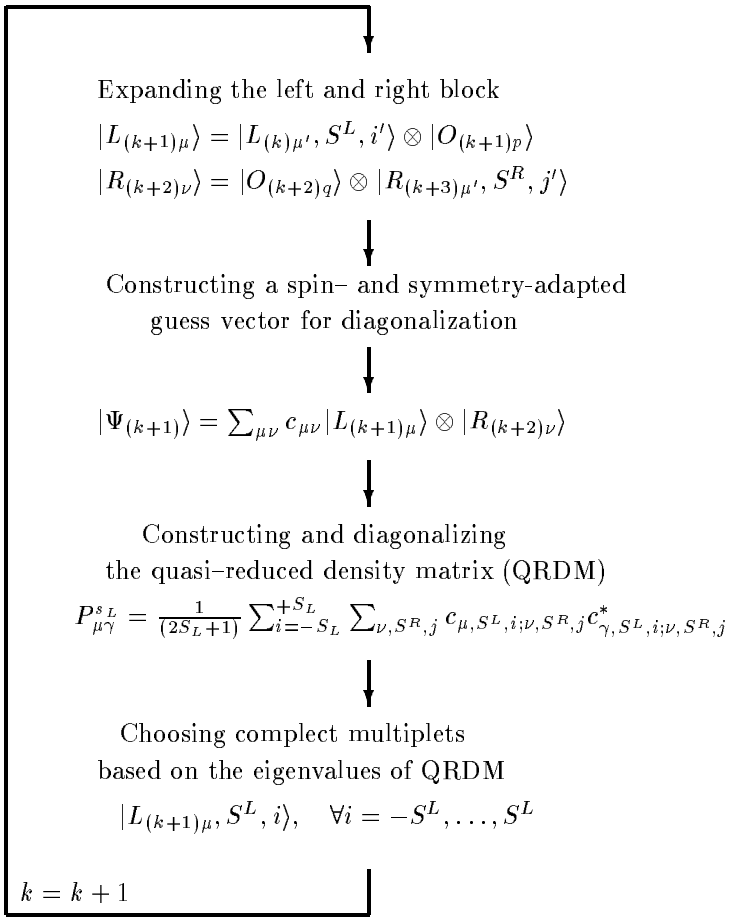

Figure 3.2: A single step of the DMRG spin–adaptation procedure in the Left–to–Right sweep. The $|L_{(k)\mu'}, S^L, i'\rangle$ states are the spin–adapted old states coming from the last $(k-1)$-th step in the Left–to–Right sweep. The μ' index labels the M basis states. The p index is used to label the four possible occupancies of the site orbital. The S^L index labels the multiplets and is a $\hat{S}_L^2$ eigenvalue, while i is the S^z component of a given S^L multiplet.

$j \neq i$. Thus, we can only choose or discard complete multiplets from the basis. We will make use of this first condition during the construction of the basis states for the Right part in the initial steps of the algorithm when the basis for the Right block has to be approximated. In these steps we can start from the Slater determinants, which are not the $\hat{S}_R^2$ eigenfunctions, but they span the complete manifold as long as one can generate them by iteratively flipping of the α electrons starting from the highest S_R^z determinant. Once more, we take advantage of the first condition during the expansion procedure where the obtained states are not $\hat{S}_L^2$ (or $\hat{S}_R^2$) eigenfunctions, but they still span the complete manifold.

The second essential condition for the spin-adaptation scheme is the construction of the eigenvectors of the modified reduced density matrix in such a way that they are also eigenfunctions of the $\hat{S}_L^2$ (or $\hat{S}_R^2$) operator. Additionally, the blocks of different multiplicities in the density matrix are decoupled. This condition is crucial for imposing the commutation condition between the reduced density matrix and the $\hat{S}_L^2$ (or $\hat{S}_R^2$) operator.

In order to introduce spatial and spin symmetry into the DMRG code, we had to modify some parts of the DMRG code significantly. We will describe now algorithmically the essence of the changes and some of the possible improvements which they may introduce.

3.3.1 The initialization procedure

Even though the optimal many-body basis for the Right orbital subspace is not known during the first steps of the algorithm, the matrix representation of the Hamiltonian has still to be created. Thus, one needs to calculate an approximate representation of the necessary creation and annihilation operators and various auxiliary operators [19] in the basis prepared for the Right block. The quality of the guessed basis can affect the rate of convergence of the DMRG calculation [91]. It is therefore desirable to include the most important Slater determinants into the basis used to construct the matrix representation of the operators. One also needs to take care of incorporating proper symmetries; otherwise the convergence will be impaired since during later steps of the algorithm the wave function will need to recover from the spin or spatial symmetry

contamination introduced during the initial stages. A detailed explanation and an explicit example of the terminology used in this section can be found in Appendices A and B.

1. *Preparing the manifold of the Slater determinants in order to span $\hat{S}^2$ eigenfunctions in the right orbital subspace.* For every possible partition (where by partition we mean a particular N_α, N_β number) in the Right block, we look for the irreducible representations that can couple with the states from the Left block to yield the required symmetry for the overall wave function. Only Slater determinants belonging to these irreducible representations need to be created. In every possible partition, we look for single determinants that upon action with S_r^+ will return zero and are the S_r^2 eigenfunctions. Such determinants will be the high spin determinants from ($N_\alpha \neq 0$, $N_\beta = 0$) partitions, and also the determinants with maximal number of paired electrons ($N_\alpha \neq 0$, $N_\beta \neq 0$, where $N_\alpha \geq N_\beta$). We take one such lowest energy determinant for every symmetry block basing our choice of determinants on the orbital energies in the right orbital subspace obtained from a previous Hartree–Fock calculation for the whole system.

   Next, we look for partitions with ($N_\alpha \neq 0$, $N_\beta \neq 0$). We check whether for these partitions we are able to construct a determinant with unpaired α and β electrons that is not an eigenfunction of S_r^2, and upon action of S_r^+ will not return zero, but rather will produce a determinant (or linear combination of determinants) with a number of α electrons that does not have a partner in the Left block. Such partitions that can result from the action of S_r^+ on the existing partitions and do not have a partner in the Left block will be called superfluous partitions. We take a single high–spin determinant that is an S_r^2 eigenfunction in every symmetry block from the superfluous partitions. Then, one flips the α electrons in all the previously obtained determinants; consequently, the determinants belonging to the lower S^z components of the multiplets in all partitions are produced. Hence, even though the superfluous partitions do not contribute to the total wave function of the system, they still need to be considered if we are to obtain a proper manifold of Slater determinants spanning the $\hat{S}^2$ adapted space in the partitions that may lead to the superfluous ones after acting with a ladder operator. In the basis of Slater determinants defined above, we construct the

$\hat{S}_r^- = \sum_{i \in \text{right}} \hat{S}_i^-$ operator as well as the $\hat{S}_r^2$ operator.

2. *Preparing a compact basis for the right orbital subspace.* For the partitions with $N_\beta = 0$ or with the maximal number of paired electrons, where acting with $\hat{S}_r^+$ gives zero, the $\hat{S}_r^2$ matrix representation is trivially diagonal within spatial symmetry blocks. However, in the partitions leading to the superfluous partitions and the lower $\hat{S}^z$ partitions, the $\hat{S}_r^2$ matrix still has to be diagonalized. Nevertheless, it is not computationally demanding since one diagonalizes the $\hat{S}_r^2$ matrix for every irreducible representation separately, and one chooses only a very limited number of Slater determinants in the step 1 of the procedure. Next, one transforms the $\hat{S}_r^-$ matrix to the $\hat{S}_r^2$ eigenbasis. Subsequently, one acts on a single $\hat{S}_r^2$ eigenvector obtained that corresponds to the highest S^z component of a multiplet from the existing partitions and from the partitions leading to the superfluous ones with the $\hat{S}_r^-$ matrix producing the lower S^z components of the basis. This procedure leads to a spin-adapted basis in the right orbital subspace. In such a basis we have at least one S^z state of a multiplet per symmetry block, and every such state is a linear combination of the most important determinants in a given partition. Such a restriction (one state per symmetry block) is not necessary in principle (provided that one applies numerical noise [22] or a perturbative correction [50] to recover the missing partitions during the initial stages of the DMRG algorithm). Moreover, to keep one state per multiplet and per symmetry block for the larger examples may become too demanding. In such cases, provided that the numerical noise is used to recover the missing partitions, the multiplets coming from the superfluous partitions and from some symmetry blocks can be omitted during the initialization procedure.

3. *Transformation of the operators to the $\hat{S}_r^2$ adapted basis.* One calculates on–the–fly, when needed, the approximate matrix representation of creation, annihilation, and auxiliary operators in the Slater determinant basis. Such operators are then transformed on–the–fly to the spin–adapted basis by the help of the basis transformation matrices described above that are also produced only when required.

Our experience indicates that such a procedure allows one to reach convergence quickly and to minimize the number of states needed to be included during the initialization procedure due to the spatial symmetry constraints. The rapid convergence can be attributed to the inclusion of the most important determinants in the basis and to targeting a proper spin state from the start.

3.3.2 The blocking step

There is no change to the usual blocking procedure present in the DMRG algorithm due to the spin–adaptation since the spatial orbital (site orbital), which is moved to the Left orbital subspace, has all four possible occupancies (all the components for expressing the $\hat{S}_L^2$ adapted eigenstates are present and the complete manifold can be spanned); i.e., even though the states $|L_\mu\rangle = |\tilde{L}_{\mu'}\rangle \otimes |l_{s\gamma}\rangle$, where $\mu' = 1, \ldots, M$ and $\gamma = 1, \ldots, 4$, created by the direct product are not the eigenstates of $\hat{S}_L^2$, they span the whole manifold needed for the $\hat{S}_L^2$ eigenstates, and we would be able to obtain the $\hat{S}_L^2$ adapted states by a basis transformation, but this is not necessary.
In order to take care of the Abelian spatial symmetry, we assign the spatial symmetry labels to the created matrix blocks (the labels will follow from the product of the old label with the site label), yielding the matrix representation of the operators blocked according to both the particle number and to the irreducible representations of the symmetry group.

3.3.3 Diagonalization of the Hamiltonian matrix

In order to minimize the number of iterations in the Davidson diagonalization in the initial steps of the DMRG algorithm and to possibly calculate excited states corresponding to different multiplicities and irreducible representations, one has to provide a starting vector targeting the state of interest for the diagonalization procedure. We choose a small number of states (from 100 to 1000) corresponding to the largest elements of our preconditioner, which in our case is expressed as $H^0 = \sum_{i \in \text{Left} \oplus \text{Right}} f_{ii} \hat{a}_i^\dagger \hat{a}_i$, where f_{ii} are the orbital energies taken from the Fock matrix. The choice of the preconditioner elements is designed not to break the degeneracies present in H^0, and complete multiplets of all degenerate states are included. In such a basis of chosen states, we construct a small Hamiltonian matrix $\hat{H}_{\text{small}}$ and a small $\hat{S}^2_{\text{small}}$ matrix. Next, we diagonalize

exactly the $\hat{S}^2_{\rm small}$ matrix obtaining the eigenvectors which are later used to transform the small Hamiltonian matrix $\hat{H}_{\rm small}$ to the $\hat{S}^2_{\rm small}$ eigenbasis. By comparing the eigenvalues of the $\hat{S}^2_{\rm small}$ matrix with the required multiplicity, we can easily choose as a starting vector an eigenvector of the $\hat{H}_{\rm small}$ matrix corresponding to the targeted multiplicity. Next, the obtained vector is expressed in the large basis. This procedure allows us to target easily different spin states corresponding to different spatial symmetries. If one of the targeted states corresponds to the same symmetry and multiplicity as the other, then it can be still obtained by orthogonalization to the previously produced state. Such a procedure for obtaining the starting vector is used only during the initial steps. At later stages of the algorithm, during the sweeping stage, the starting vector is transformed from one step to the other [92] and there is no need for the above starting procedure.

In the Davidson diagonalization procedure, the step of vector multiplication by the auxiliary operators [19, 22] is performed in a block fashion only for those spatial symmetry partitions which can couple to the overall spatial symmetry of the wave function.

3.3.4 Construction of the density matrix in the spin-adapted basis

As mentioned previously, Abelian symmetry operators commute with the reduced density matrix. This does not hold, however, for an operator corresponding to the total spin $\hat{S}^2$. The construction of the density matrix needs to be reformulated in order to include the extra constraints resulting in the conservation of the total spin. Let us compare a general spin–adapted DMRG scheme with the previously described non-spin–adapted DMRG algorithm. The basis vectors in the Left block can be written as eigenstates of the total spin:

$$|L_\mu, S^L, i\rangle, \quad i = -S_L, \ldots, S_L, \tag{3.11}$$

where $\hat{S}^2_L|L_\mu, S^L, i\rangle = S^L(S^L+1)|L_\mu, S^L, i\rangle$. The total wave function can be expressed in the form,

$$|\Phi_{\rm tot}\rangle = \sum_{\mu, S^L, i} \sum_{\nu, S^R, j} c_{\mu, S^L, i; \nu, S^R, j} |L_\mu, S^L, i\rangle |R_\nu, S^R, j\rangle. \tag{3.12}$$

The coefficients $c_{\mu S^L,i;\nu,S^R,j}$ are Clebsch-Gordan coefficients and are no longer independent of each other. The truncated wave function can be written as

$$|\Phi_{\text{truncated}}\rangle = \sum_{\mu',S^L,i'} \sum_{\nu,S^R_\nu,j'} c_{\mu',S^L,i';\nu,S^R,j'} |L_{\mu'},S^L,i'\rangle |R_\nu,S^R,j'\rangle. \tag{3.13}$$

We try again to minimize the metric $\||\Phi_{\text{tot}}\rangle - |\Phi_{\text{truncated}}\rangle\|^2$ with respect to $c_{\mu',S^L,i';\nu,S^R,j'}$ and $s^\mu_{\mu'} = \langle L_{\mu'},S^L,i'|L_\mu,S^L,i\rangle$, with a constraint that $\langle L_{\mu'},S^L,i'|L_{\gamma'},S^L,k'\rangle = \delta_{\mu'\gamma'}\delta_{i'k'}$ and a constraint essential for spin eigenfunctions $\langle L_{\mu'},S^L,i'|S^2|L_{\mu'},S^L,i'\rangle = S^L$. One can prove then that $|L_{\mu'},S^L,i'\rangle$ are the eigenvectors of a quasi–density matrix [87]

$$P^{s_L}_{\mu\gamma} = \frac{1}{(2S_L+1)} \sum_{i=-S_L}^{+S_L} \sum_{\nu,S^R,j} c_{\mu,S^L,i;\nu,S^R,j} c^*_{\gamma,S^L,i;\nu,S^R,j}. \tag{3.14}$$

The quasi–density matrix is diagonal with respect to an operator corresponding to $\hat{S}^2_L$ and is derived from the original reduced density matrix by neglecting the off–diagonal elements that would mix $\hat{S}^2_L$ eigenstates. In addition, the blocks corresponding to various S^z are summed to yield the averaged density matrix.

The algorithm that provides a quasi–density matrix may be considered an essential part of the scheme since it gives the transformation matrices which will transform the matrix representations of the DMRG operators to a new spin-adapted basis.

The above discussion applies to the case when the states in the Left and Right block are the eigenstates of $\hat{S}^2_L$ and $\hat{S}^2_R$ respectively. In our case, the states in the Left and Right subspace span the complete manifold required for the total spin eigenstates, but are not yet the total spin eigenstates. The spin–adapted basis is prepared by acting on every $\hat{S}^2_L$ eigenvector with the highest S^z_L value from the high spin partitions and the $\hat{S}^2_L$ eigenvectors from the partitions leading to the superfluous partitions with the $\hat{S}^-_L$ matrix and producing the basis states in the lower partitions by consecutive laddering. Next, the reduced density matrix is expressed in the spin–adapted basis

$$\mathbf{D}^{\mathbf{S^L;i,ir}}_{\text{new}} = (\mathbf{U}^{\mathbf{S^L,i,ir}})^\dagger \mathbf{D}^{\mathbf{i,ir}} \mathbf{U}^{\mathbf{S^L,i,ir}}, \tag{3.15}$$

where $\mathbf{U}^{\mathbf{S^L,i,ir}}$ are all vectors of given multiplicity S^L and symmetry ir in the S^z_L partition i, and $\mathbf{D}^{\mathbf{i,ir}}$ is a sub-block of the density matrix of a given symmetry in the S^z_L partition i. Blocks of different multiplicities are decoupled because separate multiplications are performed for every block of different S^z_L and symmetry. Even though the new density matrix is created independently for every different S^z_L partition, it cannot be diagonalized independently in every such partition since the states from different S^z_L partitions may contribute to the same multiplet S^L. To prevent losing the S^z_L components of a multiplet, we sum up the new density matrices for different S^z_L blocks belonging to the same multiplet and obtain the quasi–density matrix

$$\mathbf{D}^{\mathbf{S^L;ir}}_{\mathbf{quasi}} = \frac{1}{2S^L+1}\sum_i \mathbf{D}^{\mathbf{S^L;i,ir}}_{\mathbf{new}}. \tag{3.16}$$

The quasi–density matrix is averaged over all S^z blocks contributing to a given multiplet by dividing the resulting matrix by the multiplicity. Next, the averaged quasi–density matrix is diagonalized and the obtained vectors $\mathbf{V}^{\mathbf{S^L,ir}}$ are multiplied by the previously obtained $\mathbf{U}^{\mathbf{S^L,i,ir}}$ vectors

$$\mathbf{T}^{\mathbf{S^L,i,ir}} = \mathbf{U}^{\mathbf{S^L,i,ir}}\mathbf{V}^{\mathbf{S^L,ir}}, \tag{3.17}$$

in order to create basis states in every S^z_L block. These states will later be used to transform the operators to a new truncated basis. Additionally, we assign the same eigenvalue $v^{S^L,ir}$ for every created state in a given multiplet. Due to the averaging over the number of S^z_L partitions, such a procedure ensures that every state within a multiplet gets the average weight defined above. The commutation of the basis states with $\hat{S}^-_L$ and $\hat{S}^2_L$ is ensured by Eq. (3.17)

3.3.5 Choice of the basis states

As in the usual DMRG algorithm, one truncates the basis trying to retain only the most important states. However, in the spin-adapted version we cannot choose eigenvalues from every S^z partition independently, and we have to choose all S^z components of a multiplet. We choose the largest eigenvalues from the quasi–density matrix spectrum, automatically providing a spin–adapted basis due to the earlier construction of the RDM ensuring that all S^z components of a multiplet

correspond to the same eigenvalue. If we discard a state $|L_\xi, S^L, n\rangle$, we have to discard all the rest of its S^z components contributing to the multiplet. The eigenvectors selected according to the above criteria are stored and later used for the transformation of the operator matrices in the following steps.

3.4 Differences with previous algorithms

One of the first methods attempting to target spin states was introduced by Legeza and Solyom [93] and later applied by Legeza *et al.* [26] to a quantum chemical example of LiF ionic–neutral curve crossing. In this method, the spin–reversal operators were used to flip the spins along the z direction. In the case of the k-DMRG used by Legeza, the starting wave function in the Left orbital subspace was constructed to contain basis states with $N_{\rm up}$ and $N_{\rm down}$ quantum numbers, providing that both m_s and $-m_s$ were present in the basis. During the iterations such states belonged to the same density matrix eigenvalue; therefore, they both were retained if chosen. However, this approach did not provide a full adaptation of the $\hat{S}^2$ symmetry, and components of the singlet and quintet states still could have been mixed. Even though in many applications quintet and singlet state are usually well separated, this may not be true for systems involving transition metals and other more complicated examples.

In the work of Moritz *et al.* [94], different spin states of CsH were targeted at large nuclear separations where spin contamination was becoming a problem. The targeting of a particular state was done by the introduction of a level shift parameter,

$$H_{\rm shift} = H + aS_-S_+ = H + a(S^2 - S^{z\,2} - S^z), \tag{3.18}$$

where $H_{\rm shift}$ is a modified Hamiltonian matrix and a is a positive constant. The modified Hamiltonian is used to target high spin ($S = S^z$) states, and the level shift moves spin states with S higher than the target to a higher energy. Such a technique may work well for the lowest state of a given multiplicity but can be problematic for excited states (or mixing in of excited states). Since $[H, S_-S_+] = 0$, the Hamiltonian matrix and the shifted Hamiltonian matrix have a simultaneously

diagonalizing basis. The energy can be expressed as

$$E_{\text{shift}} = E + a(S'(S'+1) - S^2 - S), \tag{3.19}$$

where E is the eigenvalue of the unmodified Hamiltonian matrix and $S' \geq S+1$ is the multiplicity. The modification of the Hamiltonian matrix described above can lead to problems with the convergence of the Davidson diagonalization, if the parameter a is chosen in such a way that it destroys the diagonally dominant character of the matrix (which normally is the most appropriate to be handled by Davidson procedure). Another problem, which can still occur, is mixing of the total spin states due to encountering numerically near–degenerate states.

A rigorous spin–adaptation procedure was proposed by McCulloch and Gulácsi [87, 88, 89] and tested on a Kondo Lattice example [90], a model Hamiltonian widely used in solid state physics. In their algorithm, the Clebsch–Gordan transformation is used whenever the Left and Right blocks are joined together, and an inverse transformation has to be used to write the final wave function as a product of the Left and Right blocks. The calculated density matrix is the quasi-density matrix given by Eq. (3.16) that is also used in our algorithm. Similarly, they trace over the S^z components of a multiplet and label the spin–averaged quasi-density matrix by multiplets instead of states. However, in order to achieve computational gain in storage of the matrix representation of operators, McCulloch and Gulácsi use the fact that for systems with $SU(2)$ symmetry, the matrix elements of the Hamiltonian are independent of the S^z projection of the spin state. Capitalizing on this, the highest possible partition ($S^z = S$) is chosen, since it gives the smallest number of terms in the Clebsch-Gordan transformation. The rest of the S^z projections is then obtained using the Wigner-Eckart theorem. For a singlet wave function, the computational gain in storage and speed is significant in comparison with our algorithm, since in the McCulloch and Gulácsi algorithm the number of basis states is the same if one keeps m states (or m multiplets) with total spin. (The lower projections are obtained by the Wigner-Eckart theorem and the states from these projections do not need to be stored and transformed). However, the speed advantages for intermediate total spin values are much reduced because there are many combination of states from the Left and Right subspace that yield the desired target state multiplicity (for

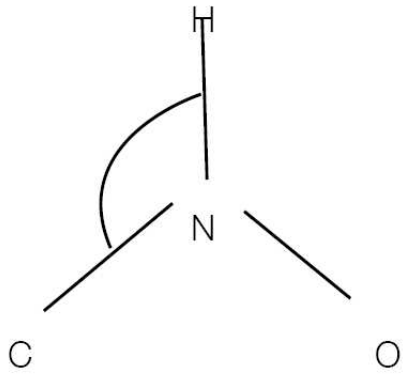

Figure 3.3: The HNCO molecule.

singlets only the Left and Right states with the same S and opposite S^z values will combine to the total singlet). In addition, efficient implementation of the Clebsch-Gordan transformation and the Wigner-Eckart theorem is complicated, and there is a certain time overhead (around 20%) related to this transformation. It is also impossible to switch easily between spin–adapted and non–spin–adapted versions of the algorithm, although that may be advantageous for systems in which the spin contamination may not have any significance due to the large separation of states. Additionally, with our spin-adaptation we can easily compare similar calculations, and hence we can more easily establish the efficiency of the algorithm.

3.5 Numerical results

In order to test the robustness of our implementation, we have carried out several tests on a model system - the HNCO molecule displaying C_s symmetry (see Fig. 3.3). The parameter varied was the C–N–H angle. We performed the calculations in the STO-3G basis set and dropped the three lowest orbitals. We stress that our calculations were not done to investigate any chemical effect, but were performed to test the effectiveness of the targeting of the spin and spatial symmetry states by our method. The above model system has only 13 orbitals (after dropping three core

orbitals) and 16 electrons, which allows us to converge the number of kept states M close to the FCI limit and allows comparisons with FCI calculations that are only possible for a small orbital space. We performed all the FCI calculation using the DALTON [95] calculation package.

A singlet state of A_1 symmetry was calculated for three different values of M, the number of kept eigenvectors. The differences between our results and the FCI values are presented in Table 3.1. It can be seen that especially for small values of M the energy is converging faster

Table 3.1: The energy error in mH between FCI and DMRG for various M vales for the lowest singlet state of A_1 symmetry state of HNCO.

Ang[deg]	$M = 150$	$M = 350$	$M = 500$
110.0	7.427	2.653	0.893
115.0	11.045	2.786	0.960
120.0	13.102	2.665	0.991
125.0	12.765	2.729	0.934
127.0	11.844	2.681	0.943
128.0	11.828	2.616	0.933
130.0	11.657	2.577	0.959
135.0	11.546	2.557	0.940
140.0	11.223	2.453	0.936
145.0	10.672	2.330	0.862
150.0	9.618	2.092	0.743
155.0	8.690	1.904	0.666
160.0	7.785	1.575	0.559

to the FCI energy for some values of the varied angle than it does for the others. Since the occupation numbers vary along the curve, different numbers of eigenvectors may be needed to recover completely the non-dynamic correlation contribution at different angles. The distortion of parallel behavior for the curve can be remedied by introducing a constant discarded weight through the curve instead of the fixing the number of eigenvectors. The previous discussion can be supported by the fact that the discarded weight (weight of the rejected eigenvalues) was highest for the points for which the difference with the FCI energy was the largest.

In principle, our code allows one to target spin states with every possible value of S and S^z, We carried out the calculation for a low–spin triplet state ($S^z = 0$) and for the high–spin triplet ($S^z = 1$) of A_1 symmetry. We performed similar calculations for the quintet state targeting both

$S^z = 0$ and $S^z = 2$ states. In both calculations, $M = 350$ was used. The differences between the results obtained and the FCI values are shown in Table 3.2. It can be seen that the high–spin

Table 3.2: The energy difference in mH between FCI and DMRG for $M = 350$ for the low–spin A_1 triplet, high–spin A_1 triplet, low–spin A_1 quintet and high–spin A_1 quintet for HNCO.

Ang[deg]	triplet $S^z = 0$	triplet $S^z = 1$	quintet $S^z = 0$	quintet $S^z = 2$
110.0	2.413	2.106	1.688	0.696
115.0	2.436	2.000	1.690	0.663
120.0	2.231	2.025	1.696	0.683
125.0	2.149	1.952	1.582	0.645
127.0	2.172	1.965	1.640	0.686
128.0	2.183	1.949	1.565	0.638
130.0	2.156	1.921	1.595	0.647
135.0	2.109	1.849	1.408	0.707
140.0	2.044	1.816	1.461	0.717
145.0	1.948	1.753	1.486	0.791
150.0	1.939	1.776	1.573	0.732
155.0	2.047	1.768	1.705	0.836
160.0	3.024	2.579	1.664	0.858

results are much closer to the FCI result than the low–spin results. At first, such a difference may seem surprising since for the FCI results there is no difference between the high–spin and the low–spin results for a particular multiplicity. The differences in the DMRG scheme can be explained by noting that the number of S^z components contributing to a multiplet is smaller for the high–spin case since there are more superfluous partitions in the Left or Right subspace. In our version of the spin-adapted DMRG method, we have to keep all the respective S^z components contributing to a multiplet. Thus, with the same number of states kept, one can have more multiplets, and therefore more accurate result for the high–spin case (since there is a smaller number of contributing S^z components than in the corresponding low–spin case). Similarly, one can see that the differences between FCI and the DMRG results are smallest for the quintet, somewhat larger for the triplet, and largest for a singlet state for a given M. A similar argument concerning the number of kept multiplets while retaining the same number of states M can be also made in this case.

Finally, let us compare a spin–adapted DMRG result with the traditional non–spin–adapted

scheme. A plot of the singlet–triplet spin curve crossing for A_1 symmetry is seen in Fig. 3.4. The DMRG results for both the spin–adapted and non–spin–adapted cases were obtained for

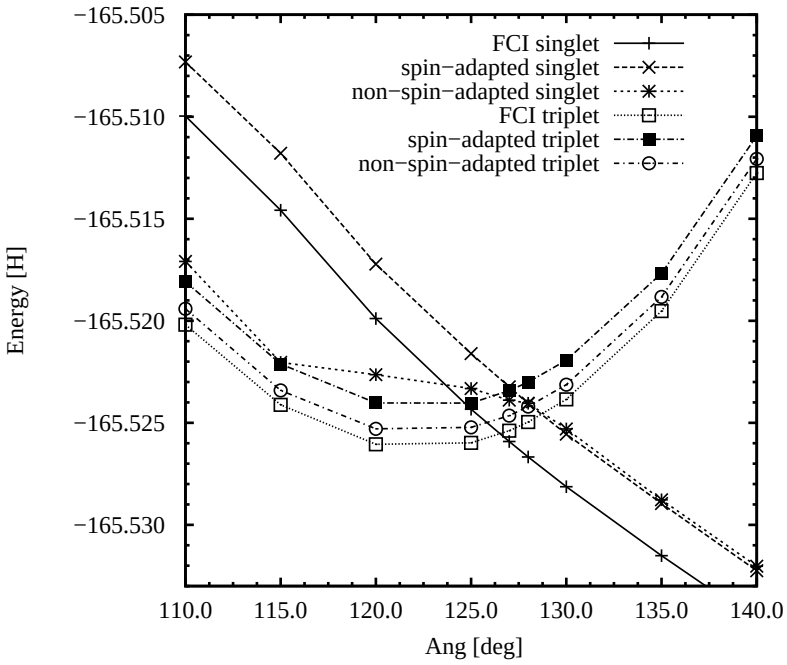

Figure 3.4: The singlet–triplet curve crossing between two states of the A_1 symmetry for the HNCO molecule. The DMRG curves are calculated with $M = 350$. The parameter varied is the C–N–H angle.

$M = 350$ states. The triplet curves are the high–spin ones, since in the case of non–spin–adapted DMRG, we could not converge the low–spin triplet state due to the singlet–triplet mixing. It can be seen that the spin–adapted DMRG curves are parallel to the FCI curves for all the analyzed multiplicities. The position of the curve crossing is predicted correctly to within one degree, in comparison with the FCI result. The high–spin triplet curve for the non–spin–adapted DMRG is also in good agreement with the FCI triplet result (such a result is of course expected since for the high-spin state mixing with the singlet state is impossible). The non–spin–adapted singlet curve first trails the triplet state since the triplet curve is lower in energy and the Davidson diagonalization usually converges to the lowest state. In the region of the crossing, because of strong singlet–triplet mixing, the non–spin–adapted DMRG gives multiplicity of two and fails

to illustrate the curve crossing correctly. Finally, the non–spin–adapted result recovers from the strong contamination at higher angle values and starts trailing the singlet state. The multiplicities for the non–spin–adapted singlet curve for different angles are shown in Table 3.3.

Table 3.3: The multiplicity of the non–spin–adapted singlet of the A_1 symmetry state for HNCO as a function of the angle Ang between the C and H atoms.

Ang[deg]	110	115	120	125	127	128	130	135	140	150	155	160
Multiplicity	2.99	3.00	2.96	2.28	1.99	1.67	1.00	1.00	1.00	1.00	1.00	1.00

We note the surprising fact that the non–spin–adapted singlet DMRG curve is higher than the spin–adapted one for the angles larger than 125 degrees. This can be understood by realizing that in the traditional formulation of the DMRG method, one minimizes iteratively the norm $|||\Phi_{\rm tot}\rangle - |\Phi_{\rm truncated}\rangle||^2$ between the true ground state and the truncated wave function and not the total energy. Thus, the contamination of the truncated wave function does not need to cause the energy lowering. The energy difference between the non–spin–adapted and spin–adapted version will be smaller for increasing values of M due to the smaller difference between $|||\Phi_{\rm tot}\rangle - |\Phi_{\rm truncated}\rangle||^2$. Thus, in the case of well separated states for high values of M, one can expect a negligible difference between the spin–adapted and non–spin–adapted DMRG method.

Finally, we present results from the spin-adapted DMRG method for methylene CH_2 obtained using cc-pVDZ basis set [96] for a singlet and quintet state of A_1 symmetry and for a triplet state of A_2 symmetry. The differences between FCI and the spin–adapted DMRG and the non-spin–adapted DMRG for singlet of A_1 symmetry for $M = 180$ and $M = 250$ are presented in Table 3.4. It can be seen that for smaller value of M, the spin–adapted results are closer to the FCI energy than are the non–spin–adapted ones. The non-spin–adapted DMRG is only closer to the FCI result at the $R = 4.00$ Å point, because at this stretched geometry it effectively converges to a quintet state yielding the multiplicity of 4.95. For M=250 the difference between FCI and the spin–adapted DMRG is very similar to the difference obtained using the non–spin–adapted DMRG. For the shorter distances, the non–spin–adapted results are slightly closer to FCI, while for the equilibrium and stretched bond distances the spin–adapted results have smaller difference

Table 3.4: The difference between the FCI energy and the DMRG energy as a function of CH stretch distance R for the CH_2 molecule, the singlet state of the A_1 symmetry in the cc-pVDZ basis set.

R[Å]	spin–adapted M=180	non–spin–adapted M=180	spin–adapted M=250	non–spin–adapted M=250
0.90	1.129	1.279	0.602	0.587
0.95	1.178	1.348	0.622	0.619
1.00	1.216	1.412	0.640	0.631
1.08	1.321	1.490	0.664	0.674
1.12	1.408	1.500	0.695	0.719
1.14	1.389	1.531	0.709	0.714
1.20	1.476	1.607	0.760	0.779
1.30	1.484	1.663	0.750	0.775
1.40	1.506	1.663	0.733	0.760
1.60	1.444	1.603	0.656	0.679
1.80	1.256	1.510	0.596	0.611
2.00	1.143	1.383	0.528	0.545
2.50	0.995	1.048	0.405	0.345
3.00	0.638	0.895	0.265	0.221
4.00	0.467	0.632	0.161	0.148

with FCI. At $R = 4.00$ Å the non-spin–adapted DMRG gives a multiplicity of 2.63.

The methylene example is presented to illustrate the fact that with the spin–adapted DMRG one is able to obtain potential energy curves that are reasonably parallel to the FCI curve, even for small values of M. In DMRG, in contrast to most other quantum chemistry methods, the curves are represented very well at the stretched limit since the automatic selection of the many–body states always chooses a proper state needed for the description of a particular region in the curve. However, at equilibrium and at distances around $1.14 - 1.50$ Å, the differences between FCI and DMRG result becomes the largest. This is most likely because the DMRG method has difficulty modeling the dynamic correlation contribution when M is not large enough. A plot of the energy differences between DMRG and FCI as a function of the CH bond length for a singlet state of methylene is seen in Fig. 3.5. For the smaller value of $M = 250$, the error ranges from 0.760 mH at $R = 1.20$ Å to 0.161 mH for the stretched $R = 4.00$ Å geometry. Similar behavior can be seen for a larger value of $M = 500$, where the differences are the largest (0.115 mH) at $R = 1.20$ Å and become gradually smaller, approaching 0.018 mH at $R = 4.00$ Å. It can be seen that for both M

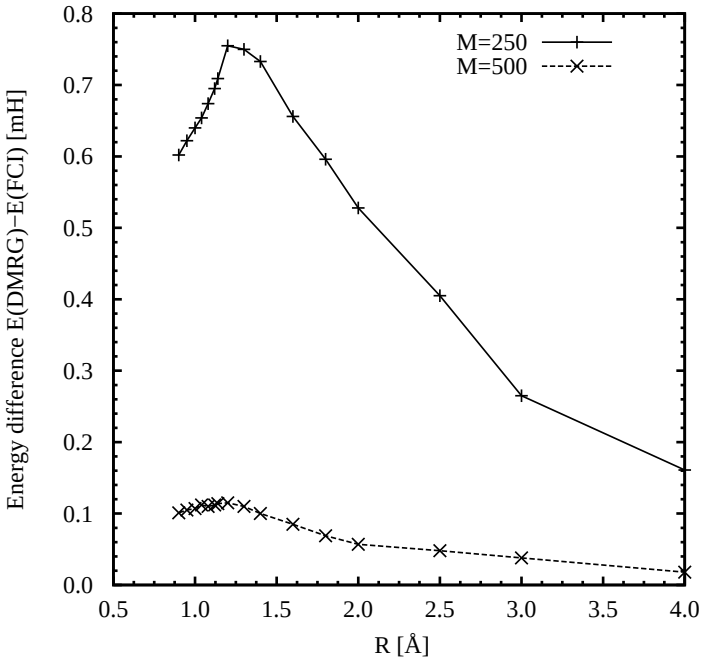

Figure 3.5: The difference between the FCI and the spin–adapted DMRG energy for two M values as a function of the CH stretch distance R for the singlet state of the A_1 symmetry for CH_2.

values, the DMRG results are within one mH of the FCI result, yielding a potential energy curve reasonably parallel to the FCI result.

To support the previously obtained conclusions, Fig. 3.6 presents the energy differences from the

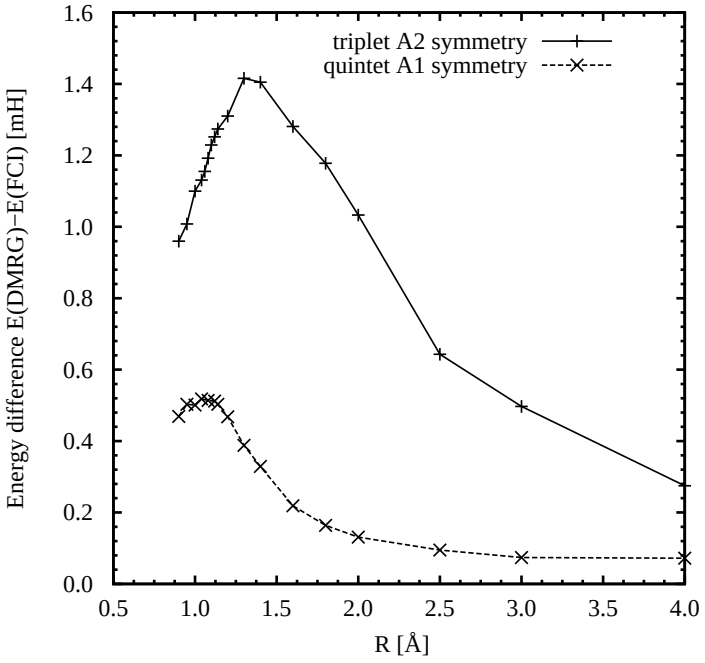

Figure 3.6: The difference between the FCI and the spin–adapted DMRG energy for the low–spin triplet and the low–spin quintet of the A_1 symmetry, $M = 250$, as a function of the CH stretch distance R for CH_2.

FCI energy for the spin–adapted DMRG calculations for the low–spin triplet of A_2 symmetry and the low–spin quintet of A_1 symmetry, both for $M = 250$. Both curves stay relatively parallel to the FCI curve presenting a similar behavior to the singlet state and having the stretched limit points closer to the FCI result.

3.6 Conclusions

In order to formulate our spin and spatial symmetry adapted DMRG scheme, we have introduced several modifications to the original DMRG procedure. The key assumptions of the proposed

spin–adaptation procedure are: i) spanning the complete manifold of states needed for acting with $\hat{S}^2$ in the Left and Right orbital subspace, and ii) ensuring that the reduced density matrix (quasi reduced density matrix), obtained in the DMRG procedure for the Left or Right orbital block, commutes with the $\hat{S}^2$ operator within this orbital block. A detailed account of the changes introduced was presented potentially allowing other interested researchers to follow the developed procedure and compare it easily with the previously used DMRG scheme.

The spin and spatial symmetry adaption procedure introduced herein has been compared with previous schemes designed to target required spin symmetries. To the best of our knowledge, this is the first application of the fully spin-adapted DMRG algorithm to a chemical example. The first spin-adaptation scheme was introduced by McCulloch and Gulácsi and applied to the Kondo lattice model [90] studied in solid state physics. Our implementation is more costly for the calculation of singlet state than theirs, but it allows one to switch conveniently between spin–adapted and non–spin–adapted versions.

We presented results for the HNCO model system targeting the singlet, triplet and quintet states of A_1 symmetry. We have listed the multiplicities for the spin–adapted and non–spin–adapted calculations, showing that the non–spin–adapted DMRG method failed to model the proper point of spin curve crossing and yielded incorrect multiplicities. The advantage of spin–adaptation is easier convergence to the state of interest, yielding a relatively parallel curve to the FCI one. The problems with either collapsing to a different spin state than targeted or lack of convergence is removed. Next, the potential energy curve for CH_2 was calculated and the energies for the lowest singlet and quintet states of A_1 symmetry and the lowest triplet state of A_2 symmetry were obtained as a function of the CH bond length. It was shown that the differences between the DMRG and FCI result are smallest at the stretched limit for both the spin–adapted and non–spin–adapted cases, and the both schemes converge to the proper limit. It was also shown that the spin–adapted DMRG does not necessarily result in higher energies than the non-spin adapted version, since in the DMRG method one minimizes the norm $\||\Phi_{\text{tot}}\rangle - |\Phi_{\text{truncated}}\rangle\|^2$ between the true ground state and the truncated wave functions and not the total energy. The only case, in which the non–spin–adapted solution seemed to work comparably to the spin–adapted

one is the case with no curve crossing for higher multiplicity states, where the non–spin–adapted method seemed to give lower energies than the spin–adapted one, and the resulting potential energy curves were parallel to the FCI curve.

Due to automatic selection of the many-body states, DMRG is a method naturally well suited for recovering non-dynamic correlation. Thus, DMRG can be used to describe active spaces, which are larger then the ones treated currently with the FCI method (e.g. as a part of the CASSCF method). Many chemically interesting systems such as molecular magnets or transition metal complexes have highly multireference character and present significant complications in their spin-states. These systems will require large active spaces, which can be treated using the DMRG method. We hope that our development of the spin-adapted version of the DMRG algorithm will be useful in future applications to describe such highly multireference systems.

Chapter 4

Modern ansatz

In the earlier chapters, we focused on the algorithmic description of the DMRG method. Although DMRG has already been proven to be useful empirically, there still remains an important step to make – to establish rigorously the physical basis of the algorithm. The connection between DMRG and matrix product states (MPS) [97, 98, 99, 100] was first noticed by Östlund and Rommer [34, 35], who identified the thermodynamic limit of DMRG with a position independent matrix product wave function. In recent years, a coherent theoretical picture of the DMRG wave function was further generalized by Verstraete [40] finally emerging in a multi–dimensional formulation of the DMRG method [41, 42, 43]. We will attempt to discuss the DMRG wave function from such a perspective. Up to now, DMRG was mostly described as an algorithm; consequently, there were no explicit equations describing the ansatz of the wave function, thus diminishing the understanding of the method in the quantum chemistry community. A description involving explanation of the DMRG wave function in terms of MPS was given earlier by Hachmann *et al.* [33].

The structure of this chapter is: in Sec. 4.1, we compare the one– and two–site DMRG algorithm paying special attention to the theoretical DMRG wave function structure rather than the usual algorithmic description. Next, we discuss the variational character of the convergence through the sweep for the one–site algorithm. In Sec. 4.2, we focus on the problem of getting

trapped in local minima and ways of avoiding it. Numerical results are presented in Sec. 4.3. Finally, we present our conclusions in Sec. 4.4.

4.1 Matrix-product states. The DMRG wave function structure

In the DMRG method, the orbital space is divided into two parts that we call Left and Right, as shown in Fig. 3.1. We denote by Left (with capital L) the bigger orbital space created by a direct product of the smaller left (with small l) space with a single orbital s. Analogous notation is used for the Right block. Let us assume that the single orbital has n_o possible occupancies and that there are k orbitals in the Left space and l orbitals in the Right space. The total number of orbitals is $k + l = p$. The full dimensions of the Left and Right spaces are $(n_o)^k$ and $(n_o)^l$, respectively. It is then clear, that due to the exponential growth of the dimension of the spaces, one is not able to carry out calculations with the full dimensions of the Left and Right spaces, even for orbital spaces of moderate sizes. Let us assume then that one is able to perform calculations involving approximately M basis states in the left orbital space, and that the full dimension of the smaller (not expanded) orbital space (left) is described by M states $\{|l_\mu^{(k-1)}\rangle\}_{\mu=1,\ldots,M}$, where the superscript $(k-1)$ denotes the number of the last orbital contained in the left subspace. One can then express all the $n_o M$ states $\{|L_{\mu'}^{(k)}\rangle\}_{\mu'=1,\ldots,n_o M}$ needed for the description of the bigger, expanded Left space as

$$|L_{\mu'}^{(k)}\rangle = |l_\mu^{(k-1)}; s^k\rangle = |l_\mu^{k-1}\rangle \otimes |s^k\rangle, \tag{4.1}$$

where $L_{\mu'}^{(k)}$ is a composite index created from the $l_\mu^{(k-1)}$ and s^k indices. The index s^k labels the basis states of the k-th single orbital in the chain. One can define a truncation matrix $(A^k)_{l_\nu s; l_\mu}$ [101] that truncates the expanded basis of the dimension $n_o M$ back to the M "most important" basis states

$$\langle l_\nu^{(k-1)} s^k | l_\mu^{(k)}\rangle = (A^k)_{l_\nu s; l_\mu}, \tag{4.2}$$

where the superscript k labeling the A matrix means that such a matrix was constructed at the k-th step of a sweep (or the k-th orbital in a chain). We note that the $(A^k)_{l_\nu s; l_\mu}$ matrix has M rows and $n_o M$ columns. One can change the notation and build a separate matrix for every possible occupancy of the k–site orbital defining a set of matrices $(A[s^k])_{l_\nu; l_\mu}$ [102] that are $M \times M$ dimensional. Using this newly introduced notation, we are ready to express the projected "most important" basis states in the Left block as

$$|l_\nu^{(k)}\rangle = \sum_\mu \sum_{s^k} (A[s^k])_{l_\mu; l_\nu} |l_\mu^{(k-1)}\rangle \otimes |s^{(k)}\rangle, \tag{4.3}$$

where the action of the A matrix is traditionally referred to in the operational definition of the DMRG algorithm as a truncation procedure.

The next step of the procedure (subsequent expansion and then a truncation) will lead to the product of the A matrices,

$$\begin{aligned}
|L_{\gamma'}^{(k+1)}\rangle &= |l_\nu^{(k)}\rangle \otimes |s^{(k+1)}\rangle, \\
|l_\gamma^{(k+1)}\rangle &= \sum_\mu \sum_{s^k} \sum_{s^{(k+1)}} \sum_\nu (A[s^k])_{l_\mu; l_\nu} (A[s^{k+1}])_{l_\nu; l_\gamma} |l_\mu^{(k-1)}\rangle \otimes |s^{(k)}\rangle \otimes |s^{(k+1)}\rangle, \\
|l_\gamma^{(k+1)}\rangle &= \sum_\mu \sum_{s^k} \sum_{s^{k+1}} (A[s^k] A[s^{k+1}])_{l_\mu; l_\gamma} |l_\mu^{(k-1)}\rangle \otimes |s^{(k)}\rangle \otimes |s^{(k+1)}\rangle,
\end{aligned} \tag{4.4}$$

and the explicit multiplications between them. At every step, when the left orbital space is expanded by one orbital (reaching $n_o M$ states), it has to be truncated back to the M most important states since as assumed earlier one can afford only a calculation with M states in the left space. Thus, one can use recursion and show that the states used to build the left subspace containing n orbitals,where $k < n$, can be expressed as

$$|l_\lambda^{(n)}\rangle = \sum_\mu \sum_{s^k, \ldots, s^n} (A[s^k] \ldots A[s^n])_{l_\mu; l_\lambda} |l_\mu^{(k-1)}\rangle \otimes |s^{(k)} \ldots s^{(n)}\rangle, \tag{4.5}$$

where we have used the previous assumption that the space containing $k-1$ orbitals has the full dimension of M and is described by $\{|l_\mu^{(k-1)}\rangle\}_{\mu=1,\ldots,M}$ states. Since such a space is not truncated

as $(n_o)^{(k-1)} \approx M$, we can express the states in this subspace as $|l_\mu^{(k-1)}\rangle = |l_s^{(1)} \dots l_s^{(k-1)}\rangle$ and write Eq. (4.5) in a more general form as

$$|l_\lambda^{(n)}\rangle = \sum_{s^1,\dots,s^n} (A[s^k]\dots A[s^n])_{s^1\dots s^{k-1};l_\lambda} |s^{(1)}\dots s^{(n)}\rangle. \tag{4.6}$$

It is clear that after truncation in the Left block generated by the DMRG procedure, the states in this block are expressed as products of $M \times M$ matrices. Hence, one calls these states *matrix product states*. To be more precise, they are called position dependent matrix product states as the A operators depend on the orbital (site) label k that indicates their "position" in the the orbital chain. Analogous expansion in terms of matrix products can be generated for all the states in the Right block. The A truncation matrix for the Right block is then defined as,

$$\langle r_\kappa^k | s^k r_\lambda^{(k+1)}\rangle = (A^k)_{r_\kappa;sr_\lambda} \equiv (A[s^k])_{r_\kappa;r_\lambda}. \tag{4.7}$$

Let us now express the DMRG wave function as a product of states from the expanded Left and Right blocks

$$|\Psi\rangle = \sum_{\mu'\nu'} c_{L_{\mu'}^n R_{\nu'}^{n+1}} |L_{\mu'}^n\rangle |R_{\nu'}^{n+1}\rangle, \tag{4.8}$$

where we have assumed that we have n orbitals in the Left block and the remaining orbitals are in the Right block. We can rewrite the above formula using explicitly the attached orbital labels obtaining the following form of the wave function

$$|\Psi\rangle = \sum_{\mu\nu} \sum_{s^n s^{n+2}} c_{l_\mu^{n-1} s^n s^{n+1} r_\nu^{n+2}} |l_\mu^{n-1}\rangle |s^n\rangle |s^{n+1}\rangle |r_\nu^{n+2}\rangle. \tag{4.9}$$

Hence, we can express the total wave function generated in the two–site DMRG method as

$$\begin{aligned} |\Psi\rangle \;=\; & \sum_{\mu\nu} \sum_{s^1,\dots,s^n} \sum_{s^{n+1},\dots,s^p} c_{l_\mu^{n-1} s^n s^{n+1} r_\nu^{n+2}} (A[s^k]\dots A[s^{n-1}])_{s^1\dots s^{k-1};l_\mu} \times \\ & (A[s^{n+2}]\dots A[s^{p-k}])_{r_\nu;s^{p-k+1}\dots s^p} |s^{(1)}\dots s^{(n)}\rangle |s^{(n+1)}\dots s^{(p)}\rangle \end{aligned} \tag{4.10}$$

which is a product of two single orbitals and two $M \times M$ matrix–products states.

Note that the wave function form given in Eq. (4.10) is not a product of only $M \times M$ matrices, as it would be in the case of the true matrix–product state since the $c_{l_\mu^{n-1} s^n s^{n+1} r_\nu^{n+2}}$ matrix is $(n_o M) \times (n_o M)$ dimensional. However, one can express the vector coefficient matrix as a set of n_o^2 matrices of dimension $M \times M$, namely $(\tilde{c}[s^n s^{n+1}])_{l_\mu^{n-1} r_\nu^{n+2}}$. Hence, Eq. (4.10) can be rewritten as a linear combination of products of $M \times M$ matrices. We also notice that one can place the matrix of the eigenvector coefficients in the middle of the Eq. (4.10), resulting in a *center matrix formulation*

$$\begin{aligned} |\Psi\rangle &= \sum_{\mu\nu} \sum_{s^1,\ldots,s^n} \sum_{s^{n+1},\ldots,s^p} (A[s^k] \ldots A[s^{n-1}])_{s^1 \ldots s^{k-1}; l_\mu} (\tilde{c}[s^n s^{n+1}])_{l_\mu^{n-1} r_\nu^{n+2}} \times \\ &\quad (A[s^{n+2}] \ldots A[s^{p-k}])_{r_\nu; s^{p-k+1} \ldots s^p} |s^{(1)} \ldots s^{(n)}\rangle |s^{(n+1)} \ldots s^{(p)}\rangle. \end{aligned} \tag{4.11}$$

Thus, the whole DMRG procedure can be visualized as a sequence of optimization steps of the wave function of Eq. (4.10). During these optimization steps performed every time one orbital is attached to the left and right orbital subspace, the optimized $c_{l_\mu^{n-1} s^n s^{n+1} r_\nu^{n+2}}$ coefficients are obtained by Davidson diagonalization and these coefficients are then used to obtain improved ansatz matrices A for truncating the expanded block. The states from the reduced block are used again in the expansion procedure in which a new $(n+1)$-th orbital is added to the left block and new wave function coefficients $c_{l_\mu^n s^{n+1} s^{n+2} r_\nu^{n+3}}$ are obtained. Thus, a self-consistent procedure is defined leading to the optimal wave function defined in Eq. (4.10).

Finally, let us derive the conditions such that the truncation matrices A will produce an orthonormal set of states $\langle l_\mu^k | l_\nu^k \rangle = \delta_{\mu,\nu}$. Using Eq. (4.2) we can obtain

$$\begin{aligned} \langle l_\mu^{(k)} | l_\nu^{(k)} \rangle &= \sum_{\lambda,\kappa} \sum_{s^k, \tilde{s}^k} (A^\dagger[\tilde{s}^k])_{l_\mu; l_\kappa} (A[s^k])_{l_\lambda; l_\nu} \langle l_\kappa^{(k-1)} | l_\lambda^{(k-1)} \rangle \langle \tilde{s}^{(k)} | s^{(k)} \rangle \\ &= \sum_{s^k} \sum_{\lambda} (A^\dagger[s^k])_{l_\mu; l_\lambda} (A[s^k])_{l_\lambda; l_\nu}. \end{aligned} \tag{4.12}$$

Hence, we arrive at the matrix form $\sum_{s^k} \mathbf{A}[s^k]^T \mathbf{A}[s^k] = \mathbf{1}$ that is appropriate for the A matrices from the Left block. One can check that the A matrices from the Right block fulfill the condition

$\sum_{s^m} \mathbf{A}[s^m]\mathbf{A}[s^m]^T = \mathbf{1}$. Although this ansatz for the wave function is uniquely defined, there is freedom in defining the procedure for obtaining the A matrices [103]. In the current implementations of the DMRG method, the A matrices have been built out of the eigenvectors of the reduced density matrix (RDM), defined as

$$D_{L_{\mu'}L_{\nu'}} = \sum_{\lambda'} c_{L_{\mu'}R_{\lambda'}} c_{L_{\nu'}R_{\lambda'}}, \tag{4.13}$$

where the sum runs over all $R_{\lambda'}$ states in the Right orbital subspace. Thus, the A matrices are defined to minimize the metric, $\||\Psi^T_{\mathrm{DMRG}}\rangle - |\Psi^E_{\mathrm{DMRG}}\rangle\|^2$ between the expanded and truncated DMRG wave functions. However, in principle the A matrices also can be built in different ways (e.g., one can minimize the energy and define the A matrices by the condition $\frac{\partial\langle\Psi|\hat{H}|\Psi\rangle}{\partial A} = 0$), providing only that they fulfill Eq. (4.12) and somehow truncate the space.

Finally, we note that the wave function of Eq. (4.10) has the form $\mathbf{L} \bullet \bullet \mathbf{R}$, where the orbital unit was added to both the left and right subspaces. By the orbital unit, we mean a set of orbitals that in the existing implementations of the DMRG algorithm was restricted to only one (as it also was in the example above). In principle, it can be an arbitrary number of orbitals for which we are able to know all the states exactly, i.e., an orbital space in which we are able to perform FCI calculation. However, one can also imagine an alternative scheme, in which the orbital unit (or in our case a single orbital) is added only to one of the subspaces, i.e., $\mathbf{L} \bullet \mathbf{R}$. We will now explicitly discuss these two cases.

4.1.1 Two–site algorithm

First, we shall discuss a scheme in which the orbital unit is added to both blocks during the expansion procedure, namely, (Left=left+orbital unit) and (Right=right+orbital unit), or simply $\mathbf{L} \bullet \bullet \mathbf{R}$. We will use the two–site algorithm in which the orbital unit is simply one orbital as a special example of this case. For the two–unit algorithm, the dimensions of the extended Right and Left spaces are both same, and equal to $n_o M$ states. Thus, the dimension of the Hamiltonian matrix entering the Davidson diagonalization is $(n_o^2 M^2) \times (n_o^2 M^2)$. Next, the vector obtained

from Davidson diagonalization is used to create the reduced density matrix. From Eq. (4.13) we see that in the construction of the RDM for the Left block, the summation runs over all the states belonging to the Right block, and the dimension of the RDM is $(n_o M) \times (n_o M)$ (for the Right RDM the summation runs over the states from the Left block and the Right RDM dimension is also $(n_o M) \times (n_o M)$). Since the number of non–zero eigenvalues after the diagonalization of the RDM has to be equal to the minimum of the Left and Right RDM ranks, and they are both equal, the number of non–zero eigenvalues is typically equal to $n_o M$. The original role of the right orbital unit was to prevent the calculation from getting stuck into a local minimum since it was observed that the **L • •R** scheme is much less prone to getting stuck in a local minimum than is the **L • R** scheme. The direct product between the left (right) block and the single orbital helps to create a many–body basis, as was already explained earlier.

The many–body basis production is achieved in an iterative manner since during the initial stages of the algorithm the basis states in the Right block are only approximated. One performs a direct product of subsequent orbital units with the left and right blocks followed by the construction of the reduced density matrix and a selection of many–body states. After the number of orbitals in the Left block reaches $p - s - 1$, where p is the total number of orbitals and s is the number of orbitals in the orbital unit, one reverses the roles of the Left and Right block, and the algorithm performs a Right–to–Left sweep. Thus, the main task of the sweeping stage is to lead the many–body basis to convergence where the energy, wave function coefficients c, and the truncation matrices A stop changing.

4.1.2 One–site algorithm

The ansatz for the two–site DMRG wave function was written in Eq. (4.10); one can write a similar ansatz for the one–site DMRG,

$$
\begin{aligned}
|\Psi\rangle \;=\; & \sum_{\mu\nu} \sum_{s^1,\ldots,s^n} \sum_{s^{n+1},\ldots,s^p} c_{l_\mu^{n-1} s^n r_\nu^{n+1}} \left(A[s^k] \ldots A[s^{n-1}]\right)_{s^1 \ldots s^{k-1}; l_\mu} \times \\
& \left(A[s^{n+1}] \ldots A[s^{p-k}]\right)_{r_\nu; s^{p-k+1} \ldots s^p} |s^{(1)} \ldots s^{(n)}\rangle |s^{(n+1)} \ldots s^{(p)}\rangle, \qquad (4.14)
\end{aligned}
$$

in which the $c_{l_\mu^{n-1} s^n r_\nu^{n+1}}$ is not $(n_o M) \times (n_o M)$ dimensional as in the two–site case, but only $(n_o M) \times M$ dimensional. Hence, we can express it as a set of n_o matrices of dimension $M \times M$ written as $(\tilde{c}[s^n])_{l_\mu^{n-1} r_\nu^{n+1}}$. Consequently, even though the algorithmic modification between the two– and one–site DMRG is minor, there is a difference between the two– and one–site DMRG wave function parametrizations, and one can expect different convergence characteristics of the energy for the two cases.

For the one–site algorithm, only one of the orbital spaces is expanded by a single orbital (or a single orbital unit in a general case). If the left space is expanded, the dimensions of the orbital spaces are $n_o M$ and M for Left and Right, respectively. Thus, the dimension of the Hamiltonian matrix in the Davidson diagonalization is $(n_o M^2) \times (n_o M^2)$, which is n_o^2 times smaller than in the case of the two–unit algorithm. Next, after creating the reduced density matrix, we notice that the dimensions of the Left and Right density matrices are $(n_o M) \times (n_o M)$ and $M \times M$, respectively. Hence, the Left reduced density matrix can have only M nonzero eigenvalues. In such a case, the choice of the M eigenvalues becomes trivial, and we have to ensure only that we choose M eigenvalues that are above numerical zero. Thus, there is no truncation error when the $n_o M$ dimensional product basis, which we compactly denote as $|l_\mu^{(k)}\rangle_{\mu=1,\ldots,M} \otimes |s^{(k+1)}\rangle_{s=1,\ldots,n_o}$, is decimated to the M–dimensional DMRG basis $|l_\nu^{(k+1)}\rangle_{\mu=1,\ldots,M}$, because the reduced density matrix (traced over M states from the Right block) can have only M non–zero eigenvalues. Consequently, the variational wave function in the basis $|l_\mu^{(k+1)}\rangle_{\mu=1,\ldots,M} \otimes |s^{(k+2)}\rangle_{s=1,\ldots,n_o} \otimes |r_\lambda^{(k+3)}\rangle_{\lambda=1,\ldots,M}$ can be represented *exactly* in the $M \times M$ basis $|l_\nu^{(k+2)}\rangle_{\nu=1,\ldots,M} \otimes |r_\lambda^{(k+3)}\rangle_{\lambda=1,\ldots,M}$. It follows then from the variational theorem that the energy calculated using the one–site algorithm can only remain constant or be lower than the energy from the previous step.

Thus, the energy within a sweep converges monotonically, and at convergence, the wave function between the steps of a sweep does not differ. In contrast to the above for the two–site algorithm, the spanned space changes between steps due to truncation of the basis vectors during the decimation procedure, such that the space does not contain the previous wave function. Consequently, when the orbital unit is added to both the left and right blocks, the energy lowers

approximately in the middle of a sweep, but then increases as the sweep is continued. This is true even at convergence, and hence the wave functions at different steps of a sweep are different. The variational property of the one–site algorithm was noticed earlier by Chan and Head–Gordon [22]. Another very beautiful description of the variational character of the one–site algorithm in terms of matrix product states (MPS) can be found in a paper by Verstraete *et al.* [40].

In the one–site scheme, there is no formal truncation error, $tre = 1 - \sum_i^M w_i$, where w_i are the eigenvalues of the DMRG reduced density matrix. Hence, the extrapolation formula proposed by Legeza *et al.* [25], $ln\frac{E-E_{\rm FCI}}{E_{\rm FCI}} = a\ln(tre) + b$, where $E_{\rm FCI}$, a and b are determined from the fit, cannot be used. However, a theoretical convergence formula found by Chan *et al.* [104] can be applied, $|\delta E| \sim exp[-\kappa(\ln M)^2]$, where κ is a parameter and M is the number of states kept.

We shall summarize the distinct features of the $\mathbf{L} \bullet \mathbf{R}$ scheme. The energy decreases or stays constant at every step of the algorithm. The cost of the Davidson diagonalization per step is n_o^2 times smaller than the two–site algorithm. Additionally, for the one–site algorithm it is easier to choose M retained states from the reduced density matrix spectrum since there are only M non–zero eigenvalues. The merit of the one–site algorithm lies in yielding at the convergence a single wave function through a sweep. Thus, in the final converged sweep, one can produce the two–body density matrix corresponding to a single wave function (providing only minimal storage conditions that we will describe later). Although it may seem that the one–site algorithm is preferable to the two–site algorithm, the next section will describe problems that appear when the one–site scheme is used.

4.1.3 Properties of the DMRG ansatz

In Chapter 3, we called DMRG a general, truncated, contracted, iterative configuration interaction method. Such a name highlights some properties of the DMRG ansatz. After providing the arguments that put DMRG on a firm algorithmic and theoretical footing, we are prepared to summarize the features of the DMRG ansatz.

- *Variational character*

  In DMRG, the energy is calculated as an expectation value of the DMRG wave function

with Hamiltonian, and hence DMRG results provide an upper bound to the true energy. In the operational definition of the DMRG method, this is achieved during the Davidson diagonalization. For an increasing number of kept states M, due to the increase of the diagonalization space, the energy approaches from above the true ground state energy that can be obtained in the given orbital basis. Hence, the method is systematically improvable with the increasing number M of states that are kept.

- *Lack of bias towards any reference*

  In contrast with traditional quantum chemistry methods, in DMRG one does not define a space from which the electrons are promoted. Hence, there is no determinant or set of determinants that can be treated as a reference. All possible determinants that can be created in a given orbital space are automatically included in the DMRG wave function during subsequent blocking steps. Consequently, DMRG is equally suited for treating single or multireference cases, due to the flexibility of the ansatz that can describe dynamic as well as non–dynamic correlation. Nevertheless, although treating the dynamic correlation does not pose any issues formally, in practice treating dynamic correlation with DMRG is inefficient and computationally expensive since it requires huge values of M. The wave function ansatz converges much faster for non–dynamic correlation than for dynamic.

- *Efficient contraction. Compactness*

  In the FCI method, the number of parameters used in the wave function expansion grows exponentially with the size of the orbital space. Conversely, in DMRG the number of parameters required in the wave function expansion scales as $O(M^2k)$ since k matrices of dimension $M \times M$ are needed to write the wave function expansion. The columns of the truncation matrices A are built from the eigenvectors of the reduced density matrix produced in DMRG. Every such state is a basis vector for DMRG, and every such basis state can be expressed in terms of Slater determinants if one wants to investigate the structure of the contraction performed in the DMRG method. In DMRG, the contraction runs over all possible determinants within partitions in the Left and Right orbital subspaces. These contractions are independent of the excitation level of the determinants from a chosen ref-

erence since there is no reference determinant in the DMRG scheme. Consequently, the DMRG contraction will be superior to the contracted basis produced with the configuration state functions (CSFs). One can ask the question, how many basis states M does one need to describe a system of interest? The answer, of course, depends on the system, but in general keeping a moderate number of states $O(100-500)$ for appropriate class of systems allows one to describe qualitatively correctly the chemistry of the system. However, a larger number of states may be needed to attain quantitative accuracy.

- *Suitability as a black–box method*

  In the DMRG method, relatively little interaction with the user is required, besides specifying the number of states M. The assessment of the quality of the results is also quite easy since one can always conduct calculations for larger number of states M. If the difference between such runs is not significant, one can expect that the calculation is relatively converged. If there is a substantial difference, then one needs to increase the number of kept basis states M to achieve a result that can be trusted. It is important to stress that all the levels of possible accuracy from $M = 1$ (that is equivalent to Hartree–Fock if starting with molecular orbitals) to values of M equivalent to those used in FCI calculations, are accessible within a single implementation by simply changing the parameter M. Moreover, the results for several M values can be extrapolated to the FCI limit within a given orbital basis.

- *Size–consistency*

  A method is size–consistent if the wave function for two non–interacting subsystems separates into a product of the wave functions corresponding to the subsystems. The DMRG wave function is size–consistent if the DMRG calculation is done in a local basis and the orbitals belonging to the two subsystems are not intermixed, but are ordered $a_1, \ldots, a_n, b_{n+1}, \ldots, b_{n+m}$, where a_n denotes n–th orbital in system A, and b_m denotes m–th orbital in system B. Knowing that the DMRG wave function illustrates the entanglement between the A and B parts of the superblock, it can be written as $|\Psi\rangle = \sum_\nu |\Psi_\nu^A\rangle|\Psi_\nu^B\rangle$. However, for infinite separation between the two subsystems A and B, the two wave functions

are not entangled, and we can write $|\Psi\rangle = |\Psi^A\rangle|\Psi^B\rangle$, as is required for a size–consistent method. Using the DMRG ansatz, we can show that

$$\begin{aligned}
|\Psi\rangle &= \sum_{\mu\nu} \sum_{s^1,\ldots,s^n} \sum_{s^{n+1},\ldots,s^p} (A[s^k]\ldots A[s^{n-1}])_{s^1\ldots s^{k-1};l_\mu} c[s^n]_{l_\mu r_\nu} \times \\
&\quad (B[s^{n+1}]\ldots B[s^{n+m-k}])_{r_\nu;s^{n+m-k+1}\ldots s^p} |s^{(1)}\ldots s^{(n)}\rangle |s^{(n+1)}\ldots s^{(n+m)}\rangle \\
&= \sum_{\nu} \Bigg(\sum_{s^1,\ldots,s^n} \sum_{\mu} (A[s^k]\ldots A[s^{n-1}])_{s^1\ldots s^{k-1};l_\mu} c[s^n]_{l_\mu r_\nu} |s^{(1)}\ldots s^{(n)}\rangle \times \\
&\quad (B[s^{n+1}]\ldots B[s^{n+m-k}])_{r_\nu;s^{n+m-k+1}\ldots s^p} |s^{(n+1)}\ldots s^{(n+m)}\rangle \Bigg), \qquad (4.15)
\end{aligned}$$

and consequently one can see that for a finite separations the index ν is greater than 1, while for the infinite separation, where the two subsystems are not entangled, $\nu = 1$.

- *Suitable for efficient parallelization*

  Many of the operations performed in the DMRG code, such as blocking or decimation, can be performed locally, because different matrix representations of the second quantized operators can be distributed among available processors and almost no communication between the processors is needed. The multiplication of the superblock Hamiltonian matrix by a vector, which is the most important step for the overall performance, can also be performed in the local processors without communication, provided that the vector is sent to the local processor and collected after the multiplication. The parallel DMRG performance scales linearly with the number of processors.

- *Symmetry*

  As was presented in Chapter 3, it is possible to preserve the spin and spatial symmetry in DMRG calculations. The possible basis states can be sorted according to the spin and spatial symmetry, thus allowing the reduced density matrix to be blocked accordingly and preventing different spin and spatial symmetries from being mixed.

- *Generality*

  The DMRG method can easily be generalized to deal with excited states. Since in DMRG

the state is obtained as a result of the Davidson diagonalization, the excited states can be obtained in the same way as they are produced in the CI type of schemes. Of course further improvements are also possible, and one has been presented by Dorando *et al.* [45]. Similarly, ionized or open shell states can be obtained within the model. Gradients and properties can also be calculated easily once the two–body density matrix is available. A scheme for obtaining the two–body density matrix will be presented in Chapter 6.

4.2 Local minima

The DMRG optimization procedure may visit many local energy minima in the space spanned by the products of the basis states of two (or one) orbitals and two matrix–product states from the left and right spaces. In principle, there is of course no guarantee of reaching the global minimum within such a space, and the energy minimization can get stuck in a meta–stable state far from the true energy minimum. In practice, however, this problem may be overcome quite efficiently.

Let us explain how the local energy minima translate to the structure of the reduced density matrix produced in DMRG, and eventually to the structure of the DMRG wave function. The reduced density matrix obtained in DMRG can be blocked

$$D^{T_L}_{\kappa\lambda} = \sum_{\nu} c^{T_L}_{\kappa\nu} (c^{T_L}_{\lambda\nu})^*, \quad T_L = N, N_\alpha, \mathrm{ir} \tag{4.16}$$

according to the T_L label that includes the particle number N, the number of α electrons N_α, and symmetry label ir [46]. We have simplified the notation and replaced the $L_{\mu'}$ label by μ (and analogously for the Right states). After the diagonalization of the RDM, the eigenvectors from different T_L blocks (later called partitions) are retained. Every different distribution of the eigenvectors between partitions corresponds to a different minimum that can be reached by the minimization procedure. Thus, getting stuck in a local minimum corresponds to a situation in which reaching some of the distributions of the eigenvectors between the partitions is no longer possible. Note that these minima do not correspond to different electronic states; a number of these minima is present for every electronic state surface.

The one–site algorithm is very prone to getting stuck in a local minimum since after the initialization procedure, during the sweeping stage, the number of eigenvalues per partition stops changing and reaching a different distribution of the eigenvectors between blocks stops being possible. Moreover, since for the one–site algorithm the energy can only decrease, if a local minimum exist on the path of the energy minimization then the one–site procedure will converge to it. The only energy decrease, which remains, comes from optimizing coefficients within the eigenvectors.

Another example of getting stuck in a local minimum, which occurs for both the one–site and two–site algorithms, is encountered during the initial guess. The most common example of such a situation happens when a block representing a certain symmetry or particle number is lost and cannot be restored. Thus, if a symmetry block (or state) in the Right or Left subspace is missing, it cannot be expressed in the total wave function. Such a symmetry block (or state) is not represented in the RDM, and as a result the transformation matrices built from the RDM eigenvectors will not possess the relevant states. Since one of the subspaces does not have the corresponding states to couple to the states of the other subspace, it may be difficult to restore the relevant symmetry blocks (states) due to the self-consistent nature of the sweeping stage in the DMRG algorithm. Hence, it is extremely important to apply the perturbative procedure [50] described below or numerical noise [22] to the reduced density matrix created during the initial steps of the DMRG run. Numerical noise was used by White in order to recover missing partitions for the model lattice systems [9, 72] and later also for chemical examples [19, 20]. The noise was also used in all calculations done in Chan's group [22, 104, 29, 28, 32, 33, 45] since due to the size of the systems treated, the initial guess in the right space is far too expensive to account for all the needed partitions. Application of numerical noise or the perturbation effectively restores the states missed during the initial steps when a guess of the states spanning the Right block is performed, thus making quality of the guess in the Right subspace almost inessential. Another more tedious approach may be to keep a certain number of states in every symmetry and particle block. However, if the symmetry group of the molecule is degenerate, we have no ability to account for all the possible representations coming from the projection of the higher group into

its Abelian subgroup. Neither perturbative correction nor random noise destroys the proper symmetry of the wave function obtained in DMRG, if the perturbed density matrix is blocked and diagonalized within appropriate symmetry and spin sectors. Thus, the perturbative correction or the numerical noise should always be applied to prevent losing states and getting stuck in a local minimum. It has to be stressed that if these corrections are not applied during the initial stages, one observes that the calculations depend strongly on the initial guess in the right block. Moreover, the difference between different orbital orderings is enhanced due to missing different partitions for different orderings, leading to the observation that the energy difference between different orderings is quite large, which is not the case if the noise or perturbation is used.

A perturbative solution, which is more elegant than the earlier application of numerical noise, was proposed by White [50]. We follow the argument presented in that paper in order to derive a correction to the reduced density matrix that will help the calculation to avoid getting stuck in local minima. The Hamiltonian matrix in the whole (Left and Right) orbital space can be written in the shortest form as,

$$H = \sum_i t_i O_L^i O_R^i, \tag{4.17}$$

where the matrices O_L^i act only on the Left orbital subspace, and O_R^i on the Right orbital subspace. The above formula is a short representation of the total Hamiltonian expressed as: $H_{LR} = H_L \otimes \mathbf{1}_R + I_L \otimes I_R + \mathbf{1}_L \otimes H_R$. In an ideal case, we may imagine that the Right block is complete and has all possible states. The total wave function after basis transformations of the Left and Right subspaces to the density matrix eigenstates will have the form

$$|\Psi\rangle = \sum_\mu c_\mu |L_\mu\rangle |R_\mu\rangle, \tag{4.18}$$

where $|L_\mu\rangle$ and $|R_\mu\rangle$ are the eigenstates of the Left and Right reduced density matrices. However, in a realistic case, we may not have all the states μ from the Right block that may couple to the states from the Left block. If we label by μ the states that are present in the Right block, and by μ' the states that are absent, we may express the projector $P = \sum_\mu |R_\mu\rangle\langle R_\mu|$ and its orthogonal complement $Q = 1 - P = \sum_{\mu'} |R_{\mu'}\rangle\langle R_{\mu'}|$. The unperturbed ground state energy E_0 and the

unperturbed density matrix D_0 are obtained by using the incomplete Right basis with missing μ' states. We can express the perturbation to the Hamiltonian as an operator that allows the μ and μ' states to couple, which was impossible when using only a zero order Hamiltonian. By such a defined perturbation, the missing states from the Right block have a chance to become accounted for during the later steps of the algorithm. We define the perturbation as

$$H' = \sum_i t_i O_L^i (Q O_R^i P + P O_R^i Q), \tag{4.19}$$

where $H' = H - H_0$. The first order perturbative correction to the wave function due to H' is

$$|\Psi'\rangle = \sum_i t_i (E_0 - H_0)^{-1} O_L^i Q O_R^i |\Psi\rangle. \tag{4.20}$$

Next, we can approximate the $|\Psi'\rangle$ as

$$|\Psi'\rangle \approx \sum_\mu c_\mu \sum_i \frac{t_i}{\epsilon_i} O_L^i Q O_R^i |L_\mu\rangle |R_\mu\rangle, \tag{4.21}$$

where we have used $\epsilon_i = E_0 - E_i$. We now can express our new wave function as $|\Psi\rangle = |\Psi_0\rangle + |\Psi'\rangle$, and the density matrix that can be created using the above wave function as $D_{\text{new}} = |\Psi_0\rangle\langle\Psi_0| + |\Psi'\rangle\langle\Psi_0| + |\Psi_0\rangle\langle\Psi'| + |\Psi'\rangle\langle\Psi'|$. The two middle terms from D_{new} will vanish because $Q|\Psi_0\rangle = 0$, and the only non–zero correction to the density matrix coming from the first order wave function is equal to

$$\Delta D = \sum_{\mu\nu} c_\mu c_\nu \sum_{ii'} \frac{t_i}{\epsilon_i} \frac{t_{i'}}{\epsilon_{i'}} O_L^i |L_\mu\rangle\langle L_\nu| O_L^{i\dagger} M_{\nu i' i \mu}, \tag{4.22}$$

where $M_{\nu i' i \mu} = \langle R_\nu | O_R^{i'\dagger} Q O_R^i | R_\mu \rangle$. One can ignore the off–diagonal term of the introduced correction and approximate it by $M_{\nu i' i \mu} \approx \delta_{\mu\nu}\delta_{ii'} b_\mu$. Using the above assumption, we obtain

$$\Delta D = \sum_\mu c_\mu^2 \sum_i b_i \frac{|t_i|^2}{\epsilon_i^2} O_L^i |L_\mu\rangle\langle L_\mu| O_L^{i\dagger}. \tag{4.23}$$

We can then replace $b_i \frac{|t_i|^2}{\epsilon_i^2}$ by a_i obtaining

$$\Delta D = \sum_i a_i O_L^i D_0 O_L^{i\dagger}, \tag{4.24}$$

where we have used the fact that $D_0 = \sum_\mu c_\mu^2 |L_\mu\rangle\langle L_\mu|$ is a density matrix determined in the usual way. In practice, since we do not know b_i and ϵ_i^2, one takes a_i to be a small constant independent of i, yielding the final perturbative correction formula

$$\Delta D = \sum_i a O_L^i D_0 O_L^{i\dagger}. \tag{4.25}$$

As we discussed earlier, the perturbative correction must be applied to the one–site scheme to prevent it from getting stuck in a local minimum. However, to make the one–site algorithm stable, the final converged energy should also stay reasonably independent of the initial value of the used perturbative parameter a and of the scheme for switching the perturbation to zero. If the energy was reasonably independent of the starting parameters and the switching algorithm, it would indicate that we have reached a global energy minimum. We will show later that this is quite complicated in the case of the one–site algorithm. Thus, it may seem that a more promising alternative for avoiding the local minima is to converge the DMRG calculation while using the two–site procedure, which usually does not get stuck in local minima easily, and switching near convergence to the one–site algorithm to calculate the two–body density matrix. In such a case, however, we have to perform some additional sweeps with the one–site algorithm (in order to get the stable converged energy at all steps of the sweep) after converging the usual two–site DMRG calculation.

4.3 Numerical results

4.3.1 Energy profile

We have tested the energy convergence profile for the two– and one–site algorithms on methylene, CH_2, using the cc-pVDZ basis set [96]. The calculation involved 24 orbitals and 8 electrons.

Unless mentioned otherwise, all of the calculations presented in this section, are performed using the spin– and symmetry adapted version of the DMRG algorithm [46]. For the two–site algorithm, Fig. 4.1 shows that after the initial stages of the run, where the energy is decreasing everywhere, it starts to be the lowest at a particular step of the sweep. Moreover, Fig. 4.1 shows the energy

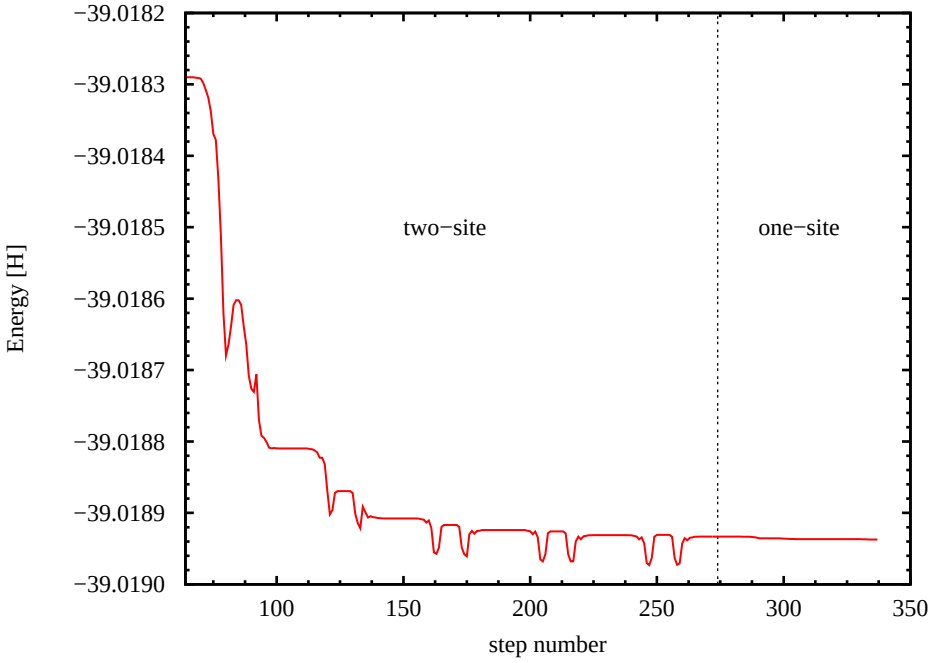

Figure 4.1: Energies from the sweeping stage of the two–site algorithm for the lowest A_1 singlet state of CH_2. The number of states kept is $M = 250$. The geometry used was $R = 1.20$ Å and the HCH angle was 104.308 degree. The calculation was done in the cc-pVDZ basis set.

behavior that is obtained on the switching from the two–site algorithm to the one–site algorithm. We see that the final energy obtained is very close to the energy plateau of the two–site sweep. For the one–site algorithm, Fig. 4.2 shows that the energy converges to a single value. Note that the converged energy is higher for the one–site (−39.018163 H) algorithm than for the two–site one (−39.018972 H). Such a result is expected, since for the same M the dimension of the diagonalization space is smaller in the case of the one–site algorithm. For the one–site algorithm, we have used as a convergence criterion the requirement that the difference between the greatest

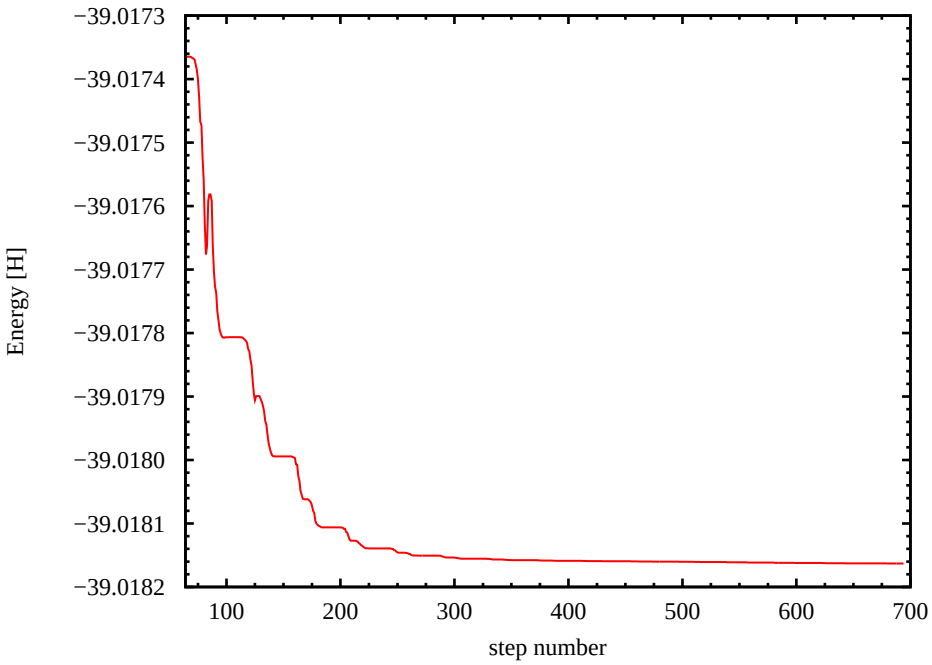

Figure 4.2: Energies from the sweeping stage of the one–site algorithm for the A_1 singlet state of CH_2. The number of states kept is $M = 250$. The geometry used was $R = 1.20$ Å and the HCH angle was 104.308 degree. The calculation was done in the cc-pVDZ basis set.

and lowest energy point within a sweep must stay below 10^{-7} H. For the two–site algorithm, we cannot compare the energies within a sweep, but we can compare the lowest energy points from subsequent sweeps, assuming it converged when the difference stays below 10^{-7} H. One can see from Figs. 4.1 and 4.2 that to achieve the desired convergence, the number of sweeps needed for the two–site algorithm was smaller than for the one–site case. We needed six sweeps for the two–site run and seventeen for the one-site one. However, the difference in the timings between these two runs reveals exactly the opposite trend requiring 1.8 times more time for the two–site case than for the one–site one. This can be understood, if one analyses the differences in the number of iterations performed at every Davidson diagonalization step in the one– and two–site cases. The number of Davidson iterations for the two–site algorithm are shown in Fig. 4.3, while Fig. 4.4

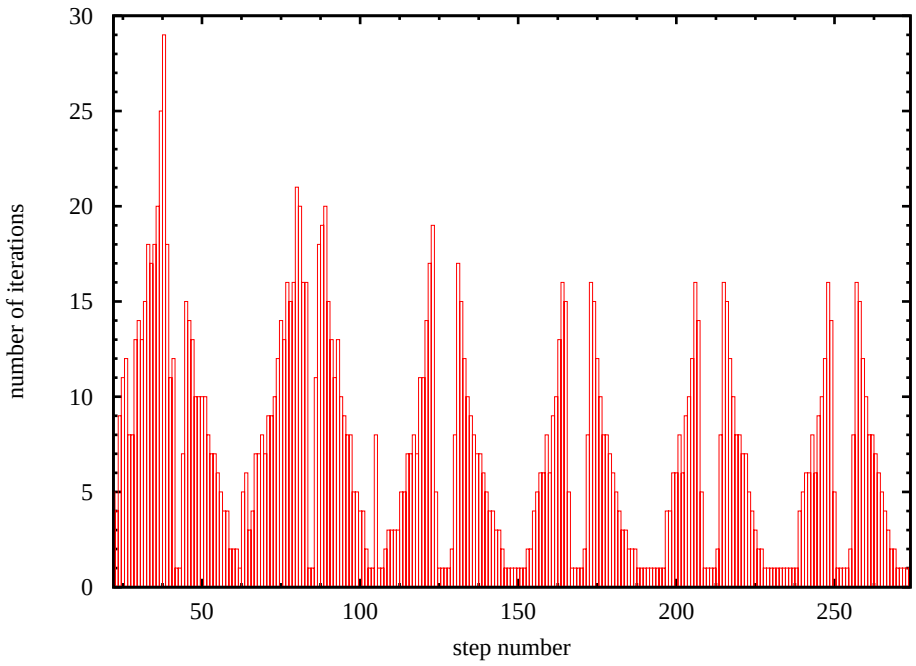

Figure 4.3: Number of iterations during the Davidson diagonalization in the sweeping stage of the two–site algorithm for the A_1 singlet state of CH_2. The convergence for the residual was 10^{-6}.

illustrates the numbers of iterations in the case of one–site DMRG. The convergence chosen in the Davidson diagonalization for the residual was 10^{-6}. In the one–site algorithm, when the

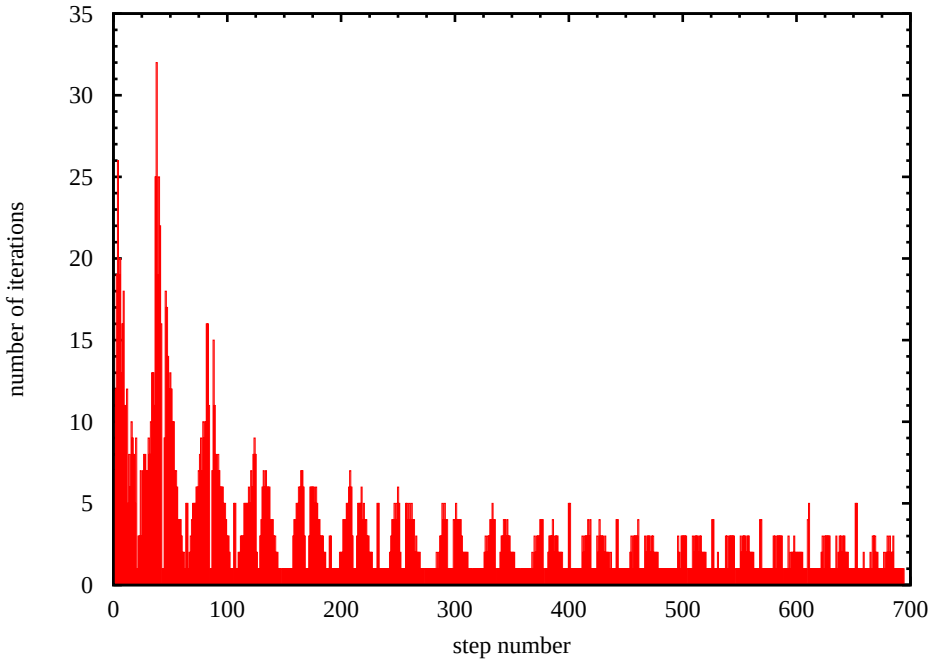

Figure 4.4: Number of iterations during the Davidson diagonalization from the sweeping stage of the one–site algorithm for the A_1 singlet state of CH_2. The convergence for the residual was 10^{-6}.

perturbation is set to zero, the wave function transformation between steps within a sweep is exact, except for very few steps. Thus, the guess vector for the Davidson diagonalization is very good, making the sweeps leading to the high convergence relatively cheap. The other factor, which makes the one–site run relatively cheaper, is the smaller dimension (n_o^2 times smaller number of matrix elements, where n_o is the number of states for the single orbital) of the Hamiltonian matrix than for the two–site run. In the current calculations, the perturbation was set to zero after the 130-th step of the one–site DMRG run. It is possible to see from Fig. 4.4 that the number of iterations is lowered past that step, finally reaching only five iterations in the most expensive step within a sweep, while the rest of the steps require only one or two iterations. We also have used the perturbative correction for the two–site run to recover missing quantum numbers. In this case the number of iterations also decreases beyond the 106-th step when the perturbation is set to zero; however, this decrease is not as significant as in the case of the one–site algorithm.

For the two–site case, as one can expect, the difference between the lowest and highest energy values within a sweep decrease with the increasing number of states kept. To emphasize this point Table 4.1 presents the energy differences between the lowest and highest energy step from

Table 4.1: The energy difference in H between the highest and lowest energy step within the last sweep of the calculations for several M values. The calculated state is the singlet state of the A_1 symmetry for the CH_2 molecule in the cc-pVDZ basis set.

M	one–site	two—site
120	9.23×10^{-8}	2.05×10^{-4}
180	7.91×10^{-8}	7.10×10^{-5}
250	4.35×10^{-8}	4.51×10^{-5}
350	9.54×10^{-8}	1.13×10^{-5}
500	6.48×10^{-8}	2.01×10^{-6}

the last sweep during both one– and two–site calculations for several values of M. In the case of the one–site run, the increasing number of eigenvectors does not influence the difference. This is understandable since the energy within a sweep is essentially converged to a single number up to the chosen convergence criterion, in this case 10^{-7} H of the energy difference. Hence, the energy changes only at the 8-th digit after the decimal point. In the case of the two–site algorithm, the

energy lowering is present only in the middle of a sweep, and for the smaller values of M the difference between the lowest and the highest energy point in a sweep can be quite significant. However, for calculations with higher number of states, the difference should become gradually smaller, as seen in Table 4.1.

4.3.2 Perturbative correction. Smoothness of potential energy surfaces.

This section analyzes the influence of the perturbation on the one– and two–site calculations. In the two–site algorithm the perturbation is used to recover the missing quantum numbers during the initialization procedure. It is very important to use the perturbation if the basis for the Right subspace during the initial guess does not contain all possible particle numbers or irreducible representations that can couple to the Left block. If the perturbation or random noise [22] is not used, then it is not unlikely that the two–site energy will get stuck in a local minimum far from the correct value causing, for example, a lack of smoothness of potential energy surfaces. Here we present a potential energy surface for methylene (see Fig. 4.5) calculated without the use of perturbative correction or numerical noise, where the non–smoothness of the curve is caused by a poor guess of the Slater determinants in the right block resulting in missing symmetry blocks able to couple to the Left block. We base our criterion for choosing the Slater determinants during the initial steps on the energy of the orbitals. Since we take only a small number of states it may happen that some symmetry states able to couple to the Left block are not qualified based on the energy criterion. The first severe jump visible in the curve is the result of such a case. However, when the perturbative correction (p=0.001) was applied, even though the guess in the right block remained poor, the curve became smooth and the missing symmetry blocks were recovered. Hence, once the perturbative correction or the noise is used, the initial guess of the Slater determinants in the right part is less important.

In the case of the one–site algorithm, the loss of quantum numbers during initialization is handled, as in the two–site case, by addition of the perturbative correction. However, in the later sweeps the perturbation helps to avoid the local minima. Our experience is that for both the one– and two–site schemes the most effective schedule is to chose the magnitude of the perturbative

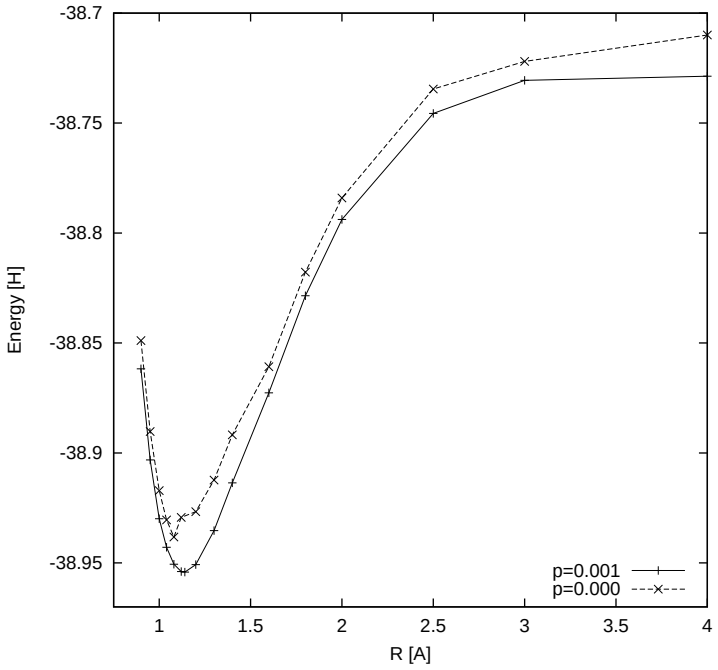

Figure 4.5: The potential energy surface for the A_1 singlet state of a CH_2 molecule as a function of the CH stretch. For $p = 0.000$ no perturbation was used, and for the other case a perturbation of $p = 0.001$ magnitude was applied. The HCH angle was 104.308 degree. The calculation was done in the DZ basis set.

parameter in such way that the energy drops at every point of the sweep. It is important to start lowering the perturbation when the energy is still dropping, because it leads to the lowest final energy. The best initial magnitude of the perturbation parameter a is between $10^{-2} - 10^{-3}$, as was found earlier by White [50] for one–dimensional models in solid state physics. We lower the magnitude of the perturbation parameter by one order of magnitude if the difference between the lowest and highest energy within a sweep, or in case of the two–site run between the lowest energies in the two subsequent sweeps, is approximately $1 - 3$ mH. Finally, the perturbation is set to zero if due to previous lowering it reaches a value of $10^{-4} - 10^{-5}$. After setting it to zero, one performs sweeps until the desired convergence is achieved.

It is important to examine how the perturbative correction influences the one– and two–site DMRG runs. We have also tested a schedule, which we label as "two–site/one–site", where we first do a two–site calculation, and then at convergence switch to the one–site algorithm. The energy values obtained for different M and different magnitudes of the perturbative correction are listed in Tables 4.2–4.5. For methylene, Table 4.2 shows that the results calculated without a perturbative correction are around 20 mH higher for the two–site and two–site/one–site algorithms. For the one–site case, the result without perturbation is 70 mH higher. For a particular number of states M, the lowest energy is obtained from the two–site procedure, and a slightly higher energy from obtained in the two–site/one–site approach. The pure one–site algorithm gives much higher energies. The difference in the final energies for different perturbation magnitudes is smaller, nearly negligible, in the case of the two–site and two–site/one–site runs compared to the one–site run.

For the H_8 and H_{12} linear chains of hydrogen atoms, Tables 4.3 and 4.4, respectively, the perturbative correction accidentally does not play any role for the two–site and two–site/one–site calculations. However, in case of the one–site calculations it seems to be absolutely essential to use it. Additionally, the results of the one–site calculations depend strongly on the magnitude of the perturbation used, differing in the most extreme case by up to 3.5 mH between different perturbation magnitudes.

Finally, for the N_2 example the results show tendencies similar to those in the methylene case.

Table 4.2: Energy obtained for different perturbation magnitudes a and different M values for CH_2 molecule in the DZ basis set. The calculation involved 13 orbitals and 12 electrons after dropping one core orbital. The HCH angle was 104.308, and the CH distance was $R = 1.20$ Å. The calculations were converged to 10^{-7} H; thus, energy differences below 10^{-7} H are listed as zero. The unperturbed, $a = 0$, energy is listed as well as the the lowest energy when $a \neq 0$. The remaining energies for $a \neq 0$ are listed relative to the the lowest perturbed energy, $E(a \neq 0) - E_{\text{lowest}}(a \neq 0)$, in H.

M/a	0.0	0.01	0.001	0.0001
two–site				
80	-38.926763	0.0	-38.945941	1.55×10^{-4}
100	-38.926814	-38.946167	1.98×10^{-6}	0.0
150	-38.926829	-38.946473	0.0	0.0
200	-38.926829	-38.946516	0.0	0.0
two–site/one–site				
80	-38.926715	6.35×10^{-6}	-38.945843	1.41×10^{-5}
100	-38.926799	4.92×10^{-6}	-38.946110	8.18×10^{-6}
150	-38.926828	-38.946455	0.0	0.0
200	-38.926829	-38.946510	0.0	0.0
one–site				
80	-38.870263	-38.9444512	2.63×10^{-3}	3.42×10^{-3}
100	-38.870286	-38.945287	6.55×10^{-5}	3.94×10^{-4}
150	-38.879685	2.69×10^{-5}	-38.946146	4.74×10^{-4}
200	-38.881046	-38.946443	0.0	3.54×10^{-6}
250	-38.934704	-38.946527	0.0	0.0

Table 4.3: Energies obtained for different perturbation magnitudes a and different M values for a linear chain H_8 in the DZ basis set. The calculation involved 16 orbitals and 8 electrons. The distance between the hydrogen atoms was $R = 1.50$ Å. The convergence, and the energy listing is same as in Table 4.2

M/a	0.0	0.01	0.001	0.0001
two–site				
50	-4.207030	-4.207036	0.0	0.0
100	-4.213468	4.00×10^{-5}	-4.213465	4.19×10^{-6}
150	-4.215900	0.0	0.0	-4.215910
200	-4.216741	2.45×10^{-6}	2.45×10^{-6}	-4.216744
two–site/one–site				
50	-4.202961	0.0	-4.202969	0.0
100	-4.212293	1.22×10^{-6}	0.0	-4.212290
150	-4.215298	0.0	-4.215277	6.70×10^{-6}
200	-4.216360	1.11×10^{-6}	1.46×10^{-6}	-4.216360
one–site				
50	-4.030163	-4.170801	6.87×10^{-6}	3.47×10^{-3}
100	-4.031622	3.58×10^{-3}	-4.203052	3.52×10^{-3}
150	-4.031624	2.83×10^{-4}	-4.211310	8.37×10^{-4}
200	-4.031624	-4.215176	1.31×10^{-4}	7.79×10^{-4}
250	-4.031624	1.25×10^{-4}	-4.216120	3.75×10^{-4}

Table 4.4: Energies obtained for different perturbation magnitudes a and different M values for a linear chain H_{12} in the DZ basis set. The calculation involved 24 orbitals and 12 electrons. The distance between hydrogen atoms was $R = 1.50$ Å. The convergence, and the energy listing is same as in Table 4.2

M/a	0.0	0.01	0.001	0.0001
two–site				
150	-6.300424	6.75×10^{-5}	8.80×10^{-5}	-6.300642
200	-6.305927	-6.306055	1.61×10^{-5}	2.14×10^{-4}
250	-6.310282	-6.310186	2.48×10^{-5}	4.97×10^{-5}
300	-6.313563	8.78×10^{-6}	-6.313610	0.0
two–site/one–site				
150	-6.296823	1.46×10^{-5}	3.31×10^{-5}	-6.296865
200	-6.302916	-6.303000	9.78×10^{-6}	2.09×10^{-5}
250	-6.307709	1.69×10^{-5}	5.22×10^{-5}	-6.307656
300	-6.311220	4.19×10^{-6}	-6.311271	0.0
one–site				
150	-6.027855	2.08×10^{-3}	1.02×10^{-2}	-6.277350
200	-6.027864	-6.286299	2.94×10^{-3}	1.53×10^{-3}
250	-6.028392	-6.294225	1.28×10^{-3}	4.71×10^{-4}
300	-6.038685	-6.300464	2.87×10^{-3}	2.79×10^{-3}
350	-6.039478	-6.303848	2.89×10^{-3}	2.81×10^{-3}
400	-6.039479	-6.306899	2.06×10^{-3}	3.22×10^{-3}

Table 4.5: Energies obtained for different perturbation magnitudes a and different M values for N_2 molecule in the cc-pVDZ basis set. The calculation involved 26 orbitals and 10 electrons. The distance between nitrogen atoms atoms was $R = 1.10$ Å. The convergence, and the energy listing is same as in Table 4.2

M/a	0.0	0.01	0.001	0.0001
two–site				
200	-109.070082	-109.264927	4.53×10^{-5}	1.28×10^{-4}
300	-109.075608	-109.268477	1.89×10^{-5}	2.30×10^{-4}
350	-109.086326	2.11×10^{-4}	-109.269702	1.95×10^{-4}
two–site/one–site				
200	-109.070081	-109.263976	4.15×10^{-5}	8.67×10^{-5}
300	-109.075607	6.93×10^{-6}	-109.267848	1.61×10^{-4}
350	-109.086326	1.90×10^{-4}	-109.269134	1.23×10^{-4}
one–site				
200	-109.063032	-109.252121	3.77×10^{-3}	4.54×10^{-3}
300	-109.071515	-109.264458	1.17×10^{-3}	4.30×10^{-3}
350	-109.081434	4.84×10^{-4}	-109.264628	2.85×10^{-3}

However, the results calculated without a perturbative correction yield an energy that is more than 0.1 H higher in comparison with cases in which the perturbative correction is used. Additionally, the difference in the number of states needed to converge to the same energy between the two–site (or two–site/one–site) calculation and one–site calculation is the most pronounced for this example.

From the results presented, one sees that the one–site run results depend much more strongly on the perturbation magnitude than do the two–site run results. Moreover, the one–site algorithm by itself appears to be incapable of reaching the lower energies obtained from the two–site/one–site run. At final stages solutions of both such runs satisfy the one–site MPS ansatz. Thus we may conclude that the lower solution is accessible to the two–site/one–site scheme, but remains hard to find for the pure one–site scheme alone. Our tests have shown that the one–site algorithm energy does not depend only on the magnitude of the initial perturbation, but also on the way the perturbation is switched off to zero. Such a feature, especially for low M values, may make the results obtained from the one–site DMRG hard to reproduce, and clearly more work is needed to design a robust scheme. Thus, the one–site algorithm is not as stable in reaching the energy minimum as is the two–site algorithm. Since we think that it is desirable to have a method that

does not depend strongly on the initial perturbation value, initially we choose to perform the two–site calculations and at convergence to switch to the one–site procedure to obtain the density matrix coming from a single wave function. Such a schedule of course is more expensive than the two–site calculation alone, but the additional expense is not very significant since the fact that no perturbative correction is used in the one site algorithm makes the number of Davidson iterations extremely low.

4.4 Conclusions

Since the DMRG method is a flexible, not biased towards any reference, and size–consistent approach that allows one to recover non–dynamic correlation (as previously noted by Hachmann *et al.* [45]), our motivation was to produce two–body (and possibly higher–body) density matrices since they will be useful for future development involving the DMRG method. To produce a two–body density matrix at a certain step within a sweep would require one to increase the disk storage and to use 4–index operators. Hence, it is advantageous to produce different elements of the two–body density matrix at different steps of a sweep and to store only 1– and 2–index operators without an increase of the required storage. However, in such a scheme, to obtain the two–body density matrix that avoids possible N–representability problem, it is important that the wave function stays the same at every step of a sweep. Consequently, the ansatz underlying the DMRG wave function (matrix product states) was discussed, showing that it leads to implications concerning the energy and wave function convergence between sweeps.

The one–unit scheme $\mathbf{L} \bullet \mathbf{R}$ (parametrized by n_o of $A[q^k]$ and $c[q^k]$ $M \times M$–dimensional matrices where k labels the position in the chain and q the number of states n_o in the orbital unit) leads to monotonic convergence of the energy during a sweep and a variational optimum. Finally, the same wave function is obtained between steps at convergence, thus making the one–unit algorithm an obvious candidate for producing the N–representable two–body matrix. However, it was demonstrated by numerical examples (CH_2, hydrogen chains, and N_2) that the one–unit scheme (e.g. one–site) gets trapped easily in a local minimum and cannot be easily converged to the

optimal solution even though the perturbative correction [50] is used to "allow the minimization to travel between the minima". Thus, the relation between the structure of the wave function and the local minima was explained, leading to a conclusion that the DMRG energy optimization travels in a complicated landscape, and that a scheme allowing the optimization to avoid being trapped in local minima is crucial.

The other type of DMRG ansatz, the two–unit ansatz $\mathbf{L} \bullet \bullet \mathbf{R}$ (e.g. two–site), was found to be much more stable in reaching the minimum after the initial perturbative correction was applied to account for the symmetries missing during the initial guess. Additionally, the results obtained did not depend significantly on the initial perturbation strength. The two–unit algorithm (parametrized by n_o of $A[q^k]$ and n_o^2 of $c[q^k q^{k+1}]$ $M \times M$–dimensional matrices) does not lead to energy decrease at every step and at convergence does not produce the same wave function at every step, due to the conditions imposed on the construction of the A matrices. However, the energy obtained from the two–unit (e.g. two–site) run is lower than that obtained from the one–unit algorithm, due to the greater number (by a factor of n_o) of c matrices that parametrize the wave function.

Consequently, since our major motivation was to produce an N–representable two–body density matrix, we decided to use the two–site algorithm to reach convergence and get to a proper minimum and then use the operator matrices obtained to start the one–site procedure in order to get an N–representable two–body density matrix. The numerical examples presented have proven that such a two–site/one–site scheme provides a near optimal solution for the one–site procedure, and allows one to avoid any dependence on the initial perturbation value. However, such a scheme has the flaw that the energy obtained during the one–site part of the run is higher than the energy obtained from the two–site part, due to the different number of c matrices parametrizing the wave function for the two– and one–site parts of the run.

It is possible, however, to use an approach that will provide the same variational freedom for both parts of the run, and will allow us to obtain a single wave function at convergence, as well as to avoid being trapped in local minima. A simple modification to consider in future is to use a two–unit/one–unit algorithm where the unit in the first part of the run will be built out of one

orbital ((**L**•)(•**R**)), while the second part of the run will be built out of two orbitals ((**L** • •)**R**). Consequently, the parametrization will use the same number of $c[q^k q^{k+1}]$ matrices for both parts of the run, and thus has the same expense of the Davidson diagonalization step, but different numbers of A matrices, $A[q^k]$ and $A[q^k q^{k+1}]$, respectively. Hence, one can expect that the energy will not increase during the one–unit part of the run.

Chapter 5

Generalization to higher dimensions

Let us take a closer look at the one–dimensional matrix product state (MPS) ansatz. We can write the wave function without prior contraction of the indices,

$$|\Psi\rangle = \sum_{s^1,\ldots,s^n} \sum_{i_1,i_2,\ldots,i} (A[s^1])_{1 i_1} (A[s^2])_{i_1 i_2} \ldots (A[s^n])_{i_n 1} |s^1 s^2, \ldots, s^n\rangle, \tag{5.1}$$

where we have written all the A matrices except the first and last ones, as 3–tensors. The first and the last A matrix are only 2–tensors, since there is no need to contract them to the left or the right, respectively. Let us write the A matrices as if they were sites on an imaginary lattice where the indices of the uncontracted A matrices correspond to the bonds of the lattice that connect the sites. Consequently, contraction of the indices of the A matrices corresponds to renormalization or reduction of the number of degrees of freedom in the vector spaces that span "over the bonds of the lattice" – imaginary degrees of freedom that provide flexibility to the ansatz.

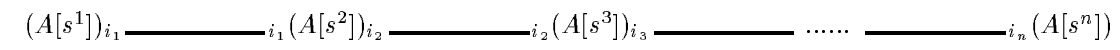

Figure 5.1: Imaginary (computational) lattice and the associated tensors A for the one–dimensional MPS ansatz. Four tensors $(A[s^p])_{i_k i_{k+1}}$, $s = \{0, \uparrow, \downarrow, \uparrow\downarrow\}$, are associated with each site p of the lattice. Each tensor has bonds i_k, i_{k+1} corresponding to the bonds connecting the site p with the neighboring sites.

5.1 Projected Entangled Pair States

For each A matrix we need two indices to describe the 1D lattice; one can ask now whether we could increase the number of indices and describe a multi–dimensional lattice? The answer is yes! Such a multi–dimensional formulation of the DMRG method is called projected entangled pair states (PEPS), and it was introduced in 2004 by Verstraete and Cirac [41, 42, 43]. An interesting application of PEPS to the Shastry-Sutherland antiferromagnet [105] was presented by Isacsson and Syljuåsen [106]. If we increase the rank of the A tensors (or simply the number of indices), we will be able to describe a 2–dimensional lattice. Such an open boundary 3×3–dimensional lattice is shown in Fig. 5.2. At every site p, we have two $A[s^p]$ tensors corresponding to two possible occupancies $s = \{\uparrow, \downarrow\}$. Such A tensors have a rank determined by the number of bonds in the imaginary lattice. For example, the internal sites have the four indices u for up, d for down, l for left, and r for right that correspond to the imaginary bonds on the computational two–dimensional lattice. The outer sites have only two or three indices. Every such index has a dimension D. Hence, at every internal site, we have two ($s = \{\uparrow, \downarrow\}$) D–dimensional rank–4 tensors $(A[s^p])^{u,d}_{l,r}$.

Note that we have introduced the term *imaginary* or *computational lattice* that is used to visualize the A tensors and the common indices between them that are later contracted. However, it is important to realize that the computational lattice is only auxiliary and does not need to coincide with the real lattice of spins. Sites on both of the lattices should be in the same spots, which ensures that the dimension of the Hilbert space is same for both imaginary and real lattices. However, bonds do not need to be the same for both lattices since the bonds in the imaginary lattice illustrate the contraction structure, and such contractions in principle may be performed differently. Hence, bonds in the imaginary lattice can build a honeycomb lattice as shown in

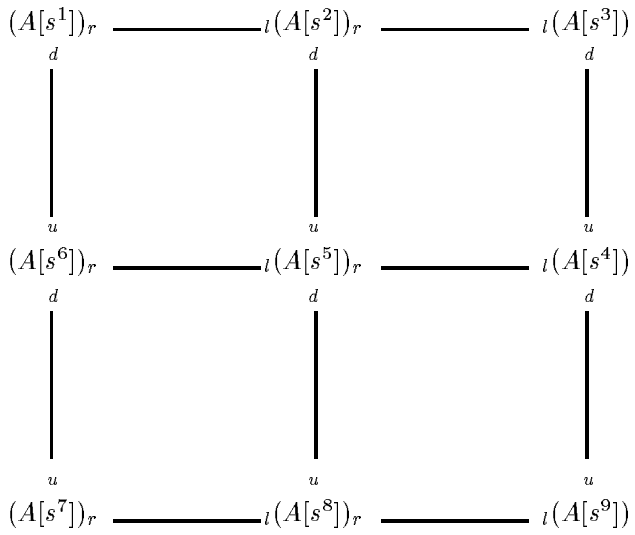

Figure 5.2: Imaginary 3×3 (computational) lattice and the associated tensors A for the two–dimensional PEPS ansatz. Two tensors $(A[s^p])^{ud}_{lr}$, $s = \{\uparrow, \downarrow\}$, are associated with each site p of the lattice. Each tensor has bonds u, d corresponding to the bonds connecting in the vertical direction (u for up, and d for down), and l, r corresponding to the horizontal direction (l for left, and r for right) that connect the site p with the neighboring sites.

Fig. 5.3, or a griddle with some bonds absent, as in Fig. 5.4. The two–dimensional lattices,

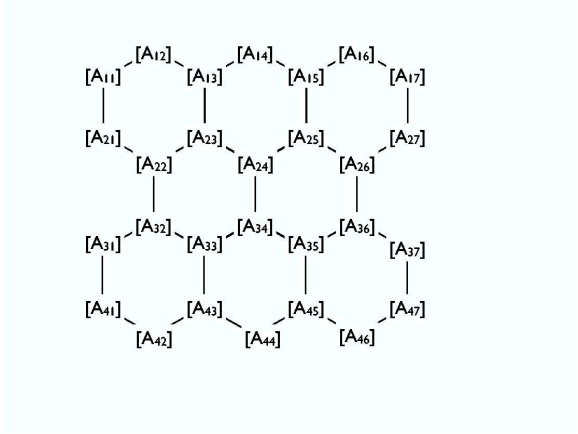

Figure 5.3: Honeycomb type of computational lattice. Internal A tensors are 3–dimensional, while external tensors are only 2–dimensional.

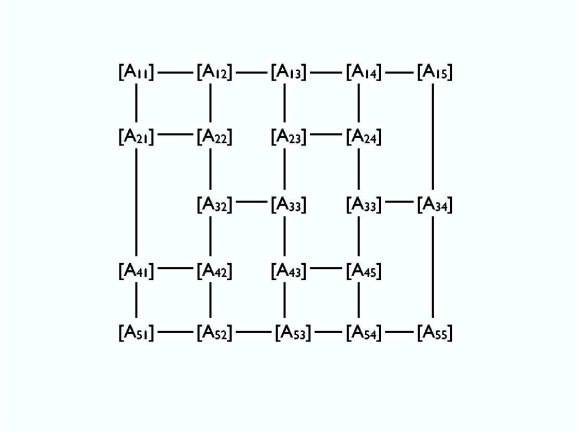

Figure 5.4: Griddle type of computational lattice. As for the honeycomb lattice, each internal A tensor is 3–dimensional; however, the scheme of the connections between rows and columns is different.

with a tensor associated with every spin, are a graphical representation of a class of states called *projected entangled pair states* (PEPS). A few remarks are now appropriate. This chapter should be considered as a presentation of a direction that allowed the solid state physics community to generalize the 1D DMRG ansatz to multiple dimensions. However, is it important to realize that although the PEPS ansatz deals remarkably well with bosons, it is still not clear how PEPS can be

used to handle fermions besides half–filling. Development in the multi–dimensional formulations of DMRG is still in its infancy, and certainly some time will be needed before a feasible scheme for using PEPS in quantum chemistry can be worked out. Consequently, this chapter is intended to be a theoretical presentation of the PEPS ansatz, rather than a practical scheme that may be very shortly used for calculations.

5.2 Minimization of energy

If one wants to minimize the energy, one needs to calculate the expectation value of the Hamiltonian with the wave function

$$E = \frac{\langle\Psi|H|\Psi\rangle}{\langle\Psi|\Psi\rangle}. \tag{5.2}$$

Consequently, one has to be able to evaluate the two types of overlaps $\langle\Psi|\Psi\rangle$ and $\langle\Psi|\Phi\rangle$, where we define $|\Phi\rangle = H|\Psi\rangle$, and additionally we assume the use of PEPS to construct the trial wave function. The previously described A tensors are used in the following way to construct a trial wave function on a lattice containing k sites,

$$|\Psi\rangle = \sum_{s^1=\uparrow,\downarrow} \cdots \sum_{s^k=\uparrow,\downarrow} \mathrm{Tr}(A[s^1]\ldots A[s^n])|s^1\ldots s^k\rangle, \tag{5.3}$$

where the trace (symbol Tr( )) means that the repeating indices are contracted. Such a contraction can be shown explicitly when using our small 3×3 lattice

$$\begin{aligned}\mathrm{Tr}(A[s^1]A[s^2]\ldots A[s^9]) = \sum_{\substack{i_1,i_2,i_4,i_5,i_7,i_8\\ j_1,j_2,j_3,j_4,j_5,j_6}} & A[s^1]^{j_1}_{i_1} A[s^2]^{j_2}_{i_1,i_2} A[s^3]^{j_3}_{i_2} \times \\ & A[s^6]^{j_1,j_6}_{i_5} A[s^5]^{j_2,j_5}_{i_5,i_4} A[s^4]^{j_3,j_4}_{i_4} \times \\ & A[s^7]^{j_6}_{i_7} A[s^8]^{j_5}_{i_7,i_8} A[s^9]^{j_4}_{i_8}.\end{aligned} \tag{5.4}$$

The summation in Eq. (5.4) runs over the repeating indices, and every such index can have D values. Different values of D may be needed to illustrate the entanglement between the sites of the lattice. The lowest limit $D = 1$ will give a result equivalent to the local mean–field method,

while for larger values of D some additional entanglement between the sites of the lattice will be illustrated.

Now, that we know the expression for the wave function, which is a projected entangled pair state (PEPS), let us discuss how the norm $N = \langle\Psi|\Psi\rangle$ is calculated. We have already learned how the ket $|\Psi\rangle$ is expressed, and the bra $\langle\Psi|$ is just a Hermitian conjugate of the ket. However, the summation over the repeating indices (or renormalization of the degrees of freedom) should not be taken until the expressions for both bra and ket are created. Let us illustrate this procedure while taking the 3×3 lattice for bra and ket,

$$N = \langle\Psi|\Psi\rangle = \sum_{s'^1,\ldots,s'^9} \sum_{s^1,\ldots,s^9} \sum_{\substack{i'_1,i'_2,i'_4,i'_5,i'_7,i'_8 \\ j'_1,j'_2,j'_3,j'_4,j'_5,j'_6}} \sum_{\substack{i_1,i_2,i_4,i_5,i_7,i_8 \\ j_1,j_2,j_3,j_4,j_5,j_6}} \langle s'^1 \ldots s'^9|$$
$$A^*[s'^1]^{j'_1}_{i'_1} A^*[s'^2]^{j'_2}_{i'_1,i'_2} A^*[s'^3]^{j'_3}_{i'_2} A^*[s'^6]^{j'_1,j'_6}_{i'_5} A^*[s'^5]^{j'_2,j'_5}_{i'_5,i'_4} A^*[s'^4]^{j'_3,j'_4}_{i'_4} A^*[s'^7]^{j'_6}_{i'_7} A^*[s'^8]^{j'_5}_{i'_7,i'_8} A^*[s'^9]^{j'_4}_{i'_8} \times$$
$$A[s^1]^{j_1}_{i_1} A[s^2]^{j_2}_{i_1,i_2} A[s^3]^{j_3}_{i_2} A[s^6]^{j_1,j_6}_{i_5} A[s^5]^{j_2,j_5}_{i_5,i_4} A[s^4]^{j_3,j_4}_{i_4} A[s^7]^{j_6}_{i_7} A[s^8]^{j_5}_{i_7,i_8} A[s^9]^{j_4}_{i_8} |s^1 \ldots s^9\rangle. \quad (5.5)$$

The summations within the bra and ket cannot be performed prior to evaluating the $\langle\Psi|\Psi\rangle$ integral since we have to be able to consider all possible combinations of the products between the i and j indices for the bra and ket tensors A^* and A, respectively. Considering all possible combinations of the i and j indices is not different than taking the direct product between the two tensors, so we define

$$[E_p]^{\tilde{u},\tilde{d}}_{\tilde{l},\tilde{r}} = [E_p]^{(u'u),(d',d)}_{(l'l),(r'r)} = \sum_{s^p} A^*[s^p]^{u',d'}_{l',r'} \otimes A[s^p]^{u,d}_{l,r}, \quad (5.6)$$

in which $\tilde{l}$, $\tilde{r}$, $\tilde{u}$, and $\tilde{d}$ are composite indices, each of them changing from 1 to $D \times D = D^2$ since they are built from the indices of the A tensors.

During the last considerations, our attention might have been distracted from the true purpose of this procedure. We want to minimize the expectation value $E = \langle H\rangle$ with respect to the A tensors; thus, we have to find a procedure in which the A tensors will be updated iteratively. When either the elements of A tensors or the obtained energy stop changing, we will stop the iterative procedure because such A tensors of dimension D will be optimal in the Hilbert space under consideration. To update a single tensor $A[s^p]$ at site p, we single it out from all the

other tensors, while for all the remaining tensors we construct the D^2–dimensional E tensors. Consequently, if we single out site 5, we can write the norm from Eq. (5.5) as

$$\langle\Psi|\Psi\rangle = \sum_{s'^5}\sum_{s'^5}\sum_{\substack{i'_1,i'_2,i'_4,i'_5,i'_7,i'_8\\ j'_1,j'_2,j'_3,j'_4,j'_5,j'_6}}\sum_{\substack{i_1,i_2,i_4,i_5,i_7,i_8\\ j_1,j_2,j_3,j_4,j_5,j_6}}\langle s'^5|[E_1]^{(j'_1j_1)}_{(i'_1i_1)}[E_2]^{(j'_2j_2)}_{(i'_1i_1),(i'_2i_2)}[E_3]^{(j'_3j_3)}_{(i'_2i_2),(i'_3i_3)}$$
$$[E_6]^{(j'_3i_3),(j'_6j_6)}_{(i'_5i_5)}A^*[s'^5]^{j'_2,j'_5}_{i'_5,i'_4}A[s^5]^{j_2,j_5}_{i_5,i_4}[E_4]^{(j'_3i_3),(j'_4j_4)}_{(i'_4i_4)}[E_7]^{(j'_6j_6)}_{(i'_7i_7)}[E_8]^{(j'_5i_5)}_{(i'_7i_7),(i'_8i_8)}[E_9]^{(j'_4i_4)}_{(i'_8i_8)}|s^5\rangle. \quad (5.7)$$

In the next step, one contracts all indices except those corresponding to site 5, yielding

$$\langle\Psi|\Psi\rangle = \sum_{s'^5}\sum_{s'^5}\sum_{\substack{i'_4,i'_5\\ j'_2,j'_5}}\sum_{\substack{i_4,i_5\\ j_2,j_5}}A^*[s'^5]^{j'_2,j'_5}_{i'_5,i'_4}\delta_{s'^5,s^5}[N_5]^{(j'_2j_2),(j'_5j_5)}_{(i'_5i_5),(i'_4i_4)}A[s^5]^{j_2,j_5}_{i_5,i_4}. \quad (5.8)$$

Writing the above formula in a simplified form gives

$$\langle\Psi|\Psi\rangle = A_5^\dagger N_5^{\rm eff}A_5, \quad N_5^{\rm eff} = \delta_{s'_5,s_5}N_5. \quad (5.9)$$

We can see that A_5 is a vector of length $2D^4$ (since s^5 has two possible values and we have D^4 possible elements of the tensor), while $N_5^{\rm eff}$ has $4D^8$ elements (since we have $\delta_{s'_5,s_5}$ with 2×2 possible outcomes and D^8 elements of the tensor). Consequently, the evaluation of the norm can be performed efficiently if we write $N_5^{\rm eff}$ as a $2D^4\times 2D^4$–dimensional matrix that is multiplied by vectors A_5 of length $2D^4$.

A new A tensor can be found during subsequent iterations of the algorithm, when the numerator $\langle\Psi|H|\Psi\rangle$ of the expectation value $E=\langle H\rangle$ is evaluated. The Hamiltonian matrix is a sum of terms $\hat{H}=\sum_{n=1,M}\hat{H}^{(n)}$. Each term $\hat{H}^{(n)}$ can be expressed as a product of the second quantized operators corresponding to different sites i, $\hat{H}^{(n)}=\prod_i\hat{O}_i^{(n)}$. One can define now a tensor E analogous to the one defined previously, but also with the information about the second quantized operator,

$$E_i^{(n)} = \sum_{s^i,s'^i}A^*[s'^i]\otimes A[s^i]\langle s'|\hat{O}_i^{(n)}|s\rangle. \quad (5.10)$$

The expression $\langle\Psi|\hat{H}^{(n)}|\Psi\rangle$ can be evaluated in same way as was $\langle\Psi|\Psi\rangle$, if we use $E_i^{(n)}$ tensors

instead of the E_i. As previously, we choose a single site k that we leave uncontracted, and we create the $E_i^{(n)}$ tensors, and then sum over all repeating indices for all other remaining sites. Finally, we arrive at the expression

$$\langle\Psi|\hat{H}^{(n)}|\Psi\rangle = A_k^\dagger H_k^{(n)} A_k. \tag{5.11}$$

Summing up all the contributions for $n = 1$ to $n = M$, then one obtains the full effective Hamiltonian for site k

$$\langle\Psi|\hat{H}|\Psi\rangle = A_k^\dagger H_k^{\text{eff}} A_k. \tag{5.12}$$

As a result, the expectation value to be minimized can be now written as

$$E = \frac{A_k^\dagger H_k^{\text{eff}} A_k}{A_k^\dagger N_k^{\text{eff}} A_k}, \tag{5.13}$$

and consequently transformed into a generalized eigenvalue equation

$$H_k^{\text{eff}} A_k = \lambda N_k^{\text{eff}} A_k. \tag{5.14}$$

Using this equation, one can obtain the new A_k tensor corresponding to a lower energy. One can then sweep from $1 \leq k \leq N$ over all possible sites in the lattice until the energy stops changing.

Up to now, we have given a somewhat naive illustrative description of the procedure, the simplification coming from the fact that we have been using the 3×3 lattice, and we did not pay too much attention to the construction of the E_i tensors. Let us consider now the problems that have to be handled in real calculations. However, we stress that these problems are not technical details of the calculations, but are the essential operations that make the whole calculation possible.
We have learned that to evaluate either the overlap of the wave function with itself $\langle\Psi|\Psi\rangle$ or with some other function $\langle\Psi|\Phi\rangle$, we need to construct and then sum over the repeating indices of the E_i tensors. Every E_i tensor was prepared using summation over possible site occupancies in bra and ket, while the direct product was taken between the virtual indices forming finally the composite indices. Consequently, we can imagine a computational lattice of the E_i tensors as illustrated by

Fig. 5.5. To calculate the expectation value, one needs to contract the indices of the E_i tensors.

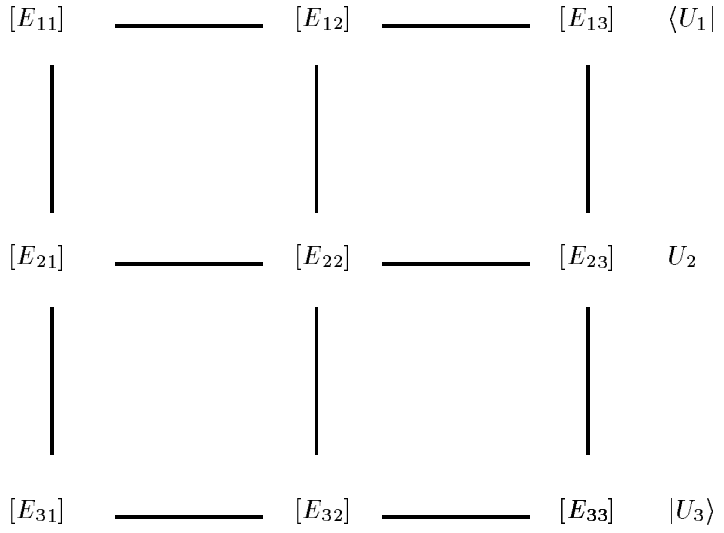

Figure 5.5: Structure of the contractions in $\langle\Psi|\Psi\rangle$. The first and last rows can be interpreted as MPS $\langle U_1|$ and $|U_3\rangle$, while the middle row U_2 is MPO. The contraction between all tensors is equal to $\langle U_1|U_2|U_3\rangle$.

This can be done in a row fashion. The first and last row contraction can be written as matrix product state (MPS), while the middle rows are the matrix product operators (MPO),

$$|U_1\rangle = \sum_{\tilde{d}_1,\ldots,\tilde{d}_L=1}^{D^2} \mathrm{Tr}([E_{11}]^{\tilde{d}_1\cdots}[E_{1L}]^{\tilde{d}_L})|\tilde{d}_1,\ldots,\tilde{d}_L\rangle, \tag{5.15}$$

$$U_k = \sum_{\substack{\tilde{d}_1,\ldots,\tilde{d}_L=1\\ \tilde{u}_1,\ldots,\tilde{u}_L=1}}^{D^2} \mathrm{Tr}([E_{k1}]^{\tilde{u}_1\tilde{d}_1\cdots}[E_{kL}]^{\tilde{u}_L\tilde{d}_L})|\tilde{u}_1,\ldots,\tilde{u}_L\rangle\langle\tilde{d}_1,\ldots,\tilde{d}_L|, \tag{5.16}$$

$$\langle U_N| = \sum_{\tilde{u}_1,\ldots,\tilde{u}_L=1}^{D^2} \mathrm{Tr}([E_{L1}]^{\tilde{u}_1\cdots}[E_{LL}]^{\tilde{u}_L})\langle\tilde{u}_1,\ldots,\tilde{u}_L|, \tag{5.17}$$

in which $\tilde{u}$ and $\tilde{d}$ are the composite indices (u,u') and (d,d'), respectively. Thus, a scalar product or expectation value can be expressed in terms of these MPS and MPO as

$$\langle\Psi|\Psi\rangle = \langle U_L|U_{L-1}\ldots U_2|U_1\rangle. \tag{5.18}$$

However, the evaluation of the above expression poses serious problems. With every contraction of two rows, the virtual dimension increases by a factor of D^2. For example, if one contracts the first and second rows $|U_2\rangle = U_2|U_1\rangle$, one obtains an MPS that is analogous to the the $|U_1\rangle$, but has all the E tensors replaced by

$$[E_{1,q}]^{d_q}_{l,r} \rightarrow \sum_{d_q=1}^{D^2} [E_{1,q}]^{d_q}_{l,r} \otimes [E_{2,q}]^{d'_q,d_q}_{l',r'}, \tag{5.19}$$

and consequently it has a greater number of left and right indices. The $|U_1\rangle$ was represented by a set of ND^2 $D^2 \times D^2$ dimensional matrices, while the new $|U_2\rangle$ MPS is represented by ND^2 $D^4 \times D^4$ dimensional matrices.

In general, we can write that every new row, which is a product of a contraction of two rows, can be written as a contraction of some F tensors

$$|U_k^{\text{new}}\rangle = \sum_{\tilde{d}_1,\ldots,\tilde{d}_L=1}^{D^2} \text{Tr}([F_{11}]^{\tilde{d}_1\cdots}[F_{1L}]^{\tilde{d}_L})|\tilde{d}_1,\ldots,\tilde{d}_L\rangle, \tag{5.20}$$

where the F tensors are defined as

$$[F_{1,q}]^{d_q}_{(ll'),(rr')} = \sum_{d_q=1}^{D^2} [E_{1,q}]^{d_q}_{l,r} \otimes [E_{2,q}]^{d'_q,d_q}_{l',r'}. \tag{5.21}$$

Due to the proliferation of the left and right dimensions with every multiplication of an MPS by MPO, the calculation soon becomes intractable. Nevertheless, one can work out a procedure that is similar in spirit to 1D DMRG, and one can approximate the MPS from Eq. (5.20) parametrized by a set of $D^4 \times D^4$ matrices by another MPS parametrized by a set of $D_f \times D_f$ matrices, such that $D_f < D^4$,

$$|\tilde{U}_k^{\text{new}}\rangle = \sum_{\tilde{d}_1,\ldots,\tilde{d}_L=1}^{D^2} \text{Tr}([\tilde{F}_{11}]^{\tilde{d}_1\cdots}[\tilde{F}_{1L}]^{\tilde{d}_L})|\tilde{d}_1,\ldots,\tilde{d}_L\rangle. \tag{5.22}$$

As in DMRG, we require that the norm $K = ||U_k^{\text{new}} - \tilde{U}_k^{\text{new}}||^2$ tends to zero. Such a minimization is done in an iterative way. One starts with a guess for the $\tilde{F}$ matrices that are present in the MPS, and then optimizes a single $[\tilde{F}_{mn}]^d_{lr}$ tensor while keeping all the others fixed. The optimal

$[\tilde{F}_{mn}]^d_{lr}$ is then found by solving a set of linear equations. In the next step, one moves to the next tensor and proceeds in this vein until the desired degree of convergence is reached. The norm can be written as

$$K = \langle U_k^{\text{new}} | U_k^{\text{new}} \rangle + \langle \tilde{U}_k^{\text{new}} | \tilde{U}_k^{\text{new}} \rangle + 2\text{Re}\langle U_k^{\text{new}} | \tilde{U}_k^{\text{new}} \rangle, \tag{5.23}$$

and to evaluate it we need to find the $D_f^2 \times D_f^2$ matrices

$$G_{mn} = \sum_u [\tilde{F}^*_{mn}]^u_{l'r'} \otimes [\tilde{F}_{mn}]^u_{lr}, \tag{5.24}$$

the $D_f D^4 \times D_f D^4$ matrices

$$H_{mn} = \sum_u [\tilde{F}^*_{mn}]^u_{l'r'} \otimes [F_{mn}]^u_{lr}, \tag{5.25}$$

and the $D^8 \times D^8$ matrices

$$J_{mn} = \sum_u [F^*_{mn}]^u_{l'r'} \otimes [F_{mn}]^u_{lr}. \tag{5.26}$$

Using the resulting matrices, the norm can be written as

$$K = \prod_{mn} G_{mn} - 2\text{Re} \prod_{mn} H_{mn} + \prod_{mn} J_{mn}. \tag{5.27}$$

We would like now to minimize the norm with respect to a given tensor $[\tilde{F}^*_{pq}]^u_{l'r'}$ that is located at site pq. To do this, we need to solve a set of equations

$$\frac{\partial K}{\partial [\tilde{F}^*_{mn}]^u_{l'r'}} = \Big[\prod_{\substack{m<p \\ n<q}} G_{mn} \Big]_{l'l} [\tilde{F}_{pq}]^u_{lr} \Big[\prod_{\substack{m>p \\ n>q}} G_{mn} \Big]_{r'r} - \Big[\prod_{\substack{m<p \\ n<q}} H_{mn} \Big]_{l'l} [F_{pq}]^u_{lr} \Big[\prod_{\substack{m>p \\ n>q}} H_{mn} \Big]_{r'r} = 0. \tag{5.28}$$

One can treat the $[\tilde{F}_{pq}]^u_{l'r'}$ and $[F_{pq}]^u_{l'r'}$ tensors as vectors, and consequently one is able to write

$$G\tilde{\mathbf{F}}^u_{pq} = H\mathbf{F}^u_{pq}. \tag{5.29}$$

Solution of this equation will provide the new $[\tilde{F}_{pq}]^u_{l'r'}$ tensors that for a given number of indices D_f are closest, in the sense of the norm, to the original MPS.

Let us list the most important steps of the procedure. One sweeps over all the lattice sites k from $1 \leq k \leq n$ during the minimization of the expectation value of the Hamiltonian, and one obtains new $A[s^k]$ tensors. The difficult step in calculating the expectation value is the evaluation of the N_k^{eff} and H_k^{eff} tensors, due to the contractions between the rows of the lattice. When two rows of the computational lattice are contracted, then the resulting matrix product state $|U_k^{\text{new}}\rangle$ has the number of indices multiplied by D^2 causing the exponential growth of the dimensionality. Consequently, a procedure similar to the original DMRG is introduced to find an approximate matrix product state $|\tilde{U}_k^{\text{new}}\rangle$ with a smaller dimension that has the smallest distance $K = \|U_k^{\text{new}} - \tilde{U}_k^{\text{new}}\|^2$ from the original $|U_k\rangle$.

A simple computational scheme, in which one sweeps along the rows of the computational lattice to achieve convergence, is listed below.

do *p=1, N*

do *q=1, N*

1N. contract the rows starting from $\langle U_1|$ up to the $p-1$ row

2N. contract the rows starting from $|U_N\rangle$ up to the $p+1$ row

3N. every time a new MPS is created, solve the set of linear equations Eq. (5.29) to reduce the dimension of the tensors

4N. with the resulting MPS $\langle U_{k-1}|$ from the row contractions contract the $[E_{pl}]^d_{(l'l),(r'r)}$ tensors for columns $1 \leq l \leq q-1$

5N. with the resulting MPS $|U_{k+1}\rangle$ from the row contractions contract tensors for columns $q+1 \leq l \leq N$

6N. the resulting tensor has 4 composite indices and can be used to determine N_{pq}^{eff}

do *n=1, M*

repeat operations **1N-6N** for term $H_{pq}^{(n)}$ in the Hamiltonian using tensors $[E_{pl}^{(n)}]$ that correspond to the operators present in the Hamiltonian.

$H_{pq}^{\text{eff}} = H_{pq}^{\text{eff}} + H_{pq}^{(n)}$

end do

solve the eigenvalue equation Eq. (5.14) to obtain λ^{new} and the tensor $[A_{pq}]$

if the difference $|\lambda^{\text{old}} - \lambda^{\text{new}}| \leq \epsilon_{\text{conv}}$ **then exit**

end do

end do

One has to stress, however, that to achieve a speedup in a practical implementation, one will not construct every time a new MPS that correspond to every row. One can save time and store appropriate tensors and perform only few contractions.

5.3 Time–evolution

The techniques described above can be used efficiently in simulation of the time–evolution. One can start with a PEPS $|\Psi_A^0\rangle$ with virtual dimension D at time zero as a starting point. Subsequently, such a PEPS is evolved by a time–evolution operator $U = e^{iH\delta t}$ to produce another PEPS $|\Psi_B\rangle$,

$$|\Psi_B\rangle = U|\Psi_A^0\rangle \tag{5.30}$$

that has the virtual dimension D_B greater than the original dimension D. Exactly the same technique that was used in the previous section to truncate the dimension is used now to find a new PEPS $|\tilde{\Psi}_B\rangle$ with dimension D that has a minimal distance to $|\Psi_B\rangle$. Such a truncated $|\tilde{\Psi}_B\rangle$ then becomes a starting point for the next time–evolution step.

The complication of the above prescription lies in the operator U that causes an exponential increase in the complexity of the virtual dimension of the PEPS. To overcome this difficulty, one introduces the Suzuki–Trotter decomposition to approximate U by product of some operators, each of which increases the virtual dimension only by a constant factor. The Suzuki–Trotter decomposition expansion to first order can be written as

$$\begin{aligned} e^{(A+B)} &= \lim_{n\to\infty} (e^{\frac{A}{n}} e^{\frac{B}{n}})^n \\ &= \lim_{n\to\infty} (1 + \frac{A}{n} + \frac{B}{n})(1 + \frac{A}{n} + \frac{B}{n}) + \dots \end{aligned} \tag{5.31}$$

or equivalently

$$e^{\delta(A+B)} = \lim_{\delta \to 0}(e^{\delta A} e^{\delta B} + O(\delta^2)). \tag{5.32}$$

For many models, the Hamiltonian matrix can then be decomposed into a horizontal–even, a horizontal–odd, a vertical–even and a vertical–odd part,

$$H = H_{he} + H_{ho} + H_{ve} + H_{vo}, \tag{5.33}$$

and the time evolution operator can be re-expressed as

$$U = e^{iH\delta t} \approx e^{iH_{he}\delta t} e^{iH_{ho}\delta t} e^{iH_{ve}\delta t} e^{iH_{ve}\delta t}. \tag{5.34}$$

Consequently, after every such action of the evolution operator the virtual dimension increases only by a constant factor, and such an MPS produced in this way can be then truncated and approximated by some other MPS that becomes the starting state for the evolution in the next step.

Chapter 6

Density matrices in DMRG

One can hope that a procedure combining the DMRG method in the active space with the optimization of the orbitals can be developed. Such a procedure, analogous to CASSCF [107], could provide a reference wave function to which the dynamic correlation can be added, preferably by some type of CC ansatz (as an example of such a procedure, see [48, 49, 47]). The above modifications, if successful, could allow the DMRG method to enter further into the realm of inorganic chemistry applications and target active spaces previously inaccessible for the FCI part of calculations.

To achieve any of these aims, it is of prime importance to be able to calculate one– and two–body density matrices since they are required for orbital optimization, as well as for the calculation of molecular gradients. The one–body density matrix can be calculated easily within the DMRG scheme, since all the required matrix representations of second quantized operators are readily available at every step of the algorithm. However, in order to calculate the two–body density matrix at the lowest energy step within a sweep, we would need to severely increase the disk storage. Alternatively, to avoid this increase in disk requirements, one can produce different two–body density matrix elements at different steps of the DMRG sweep. Such a procedure, however, can lead to a non-N-representable two–body density matrix, especially for a lower number of basis states.

In this chapter, we consider an alternative DMRG scheme that allows us to obtain an N–representable two–body density matrix. A key part of our scheme is the one–site algorithm previously proposed by White [50]. In the one–site algorithm, the energy decreases at every step of the sweep. Hence, the same wave function is obtained at the convergence at every step of a sweep, and the two–body density matrix can be assembled from the needed pieces during the last sweep yielding the N-representable two–body density matrix.

6.1 The two–body reduced density matrix

The two–body reduced density matrix is a 4–index quantity that has the form

$$\Gamma_{ijkl} = \langle \Psi | a_i^\dagger a_j^\dagger a_l a_k | \Psi \rangle, \tag{6.1}$$

where $|\Psi\rangle$ is the DMRG wave function and the i, j, k, and l are the spin orbital labels. In the one– or two–site DMRG algorithm one does not need to store the 4–index quantities that are matrix representations of the second–quantized operators due to summing the 3– and 4–index operators with the integrals [19, 22]. The DMRG disk storage requirements are $O(M^2K^2)$, where M is the number of kept many–body basis states and K is the total number of orbitals resulting from storing only one– and two–index operators. For the usual two–site DMRG algorithm, if one wanted to calculate the two–body reduced density matrix at the lowest energy point in a sweep, it would require storing K^4 operators of size M^2, which obviously cannot be afforded easily. We would like to have a procedure that will let us obtain the two–body reduced density matrix as a byproduct of the DMRG run without any increase in the current storage requirements. To achieve this, we note that the four orbital labels i, j, k, and l in Γ_{ijkl} can be distributed between the Left and Right orbital subspaces. Thus, as we will explain later in detail, different elements of the two–body density matrix can be produced at different steps of a sweep without increasing the current storage. One has to stress, however, that the way of building the two–body density matrix is not a unique one [108] if one produces different elements at the different step of a sweep. Nevertheless, for the one–site algorithm, the result should not depend significantly on the way the

two–body density matrix was constructed, and the differences should only come from numerical artifacts.

We can take advantage of the one–site DMRG algorithm, in which the energy converges monotonically in every sweep at every step, and at convergence the energy at different steps of a sweep remains sufficiently constant. Hence, the wave function between the steps is also the same up to convergence, and a two–body density matrix corresponding to a single wave function can be obtained. In the case of the two–site algorithm, the energy changes between the steps of a sweep, and at the final sweep the wave function still differs between the steps, and the two–body density matrix elements produced at different steps will come from different wave functions. (However, for completeness we note that for larger values of M the energy and the wave function within a sweep for the two–site algorithm should not change very significantly.) Thus, for the one–site algorithm we have a procedure for creating a proper two–body density matrix for which the N-representability problem is avoided. The N–representability aspect of such a two–body density matrix does not depend on the number of states, M, used in the DMRG calculation. Let us list in detail how this two–body density matrix is constructed.

1. $\forall_{ijkl}$ if $i \neq j \neq k \neq l$ where the i, j, k, and l labels refer to the site orbital labels rather than spin orbitals. One can always find a place in a sweep where two of the labels are in the Left block and two other labels are in the Right block. (One may need to permute the labels of Γ and consider an appropriate sign as a result of this permutation.) One can express the DMRG wave function as a product of the many–body states from the Left and Right subspaces obtaining the following formulas for a case in which two of the labels are present in the Left block and the remaining ones are in the Right block

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big(\langle L_\mu|a_i^\dagger a_j^\dagger|L_\kappa\rangle\langle R_\nu|a_l a_k|R_\lambda\rangle\Big), \tag{6.2}$$

$$\begin{aligned}\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big((-1)^{q_\nu}\delta_{jl}\langle L_\mu|a_i^\dagger|L_\kappa\rangle\langle R_\nu|a_k|R_\lambda\rangle + \\ -\langle L_\mu|a_i^\dagger a_l|L_\kappa\rangle\langle R_\nu|a_j^\dagger a_k|R_\lambda\rangle\Big),\end{aligned} \tag{6.3}$$

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big(\langle L_\mu|a_l a_k|L_\kappa\rangle\langle R_\nu|a_i^\dagger a_j^\dagger|R_\lambda\rangle\Big), \tag{6.4}$$

in which q_ν corresponds to the number of electrons in the $\langle R_\nu|$ state.

2. $\forall_{ijkl}$ such that $i = k$ and $i \neq j \neq l$ or $i \neq j = l$, the same rule as in the first case will apply. The only exception will be present for the elements involving the edge of the chain. As an example of such operators we may list $a^\dagger_{1\alpha}a^\dagger_{1\beta}a_{2\alpha}$, where two labels correspond to the first orbital that is the edge orbital. The edge orbitals are the orbitals creating the smallest space to which we add a single orbital in the first step of the DMRG sweep. For cases involving labels coming from the edge orbitals and the first orbital adjacent to the edge, one needs to construct 3-index operator. These are, however, very few operators present only at the very first step of the sweep. These operators do not need to be stored and can be constructed on-the-fly using

$$\langle l_\mu|\langle s_\kappa|\hat{O}_1\hat{O}_2|s_\lambda\rangle|l_\nu\rangle = (-1)^{q_\kappa}\langle l_\mu|\hat{O}_1|l_\nu\rangle\langle s_\kappa|\hat{O}_2|s_\lambda\rangle, \tag{6.5}$$

in which q_κ refers to the number of electrons in the $\langle s_\kappa|$ state, $\hat{O}_2$ is a 2-index operator in the left block basis before expansion, and $\hat{O}_1$ is a site operator in the site basis. The above multiplication can be done because the resolution of identity is exact between the attached site orbital and the left block.

3. $\forall_{ijkl}$ when three labels are same and one is different. Here one uses the fact that for a single orbital we know matrix representation of a 3–index operators explicitly. The following formulas can be used to produce the appropriate matrix elements of the two–body density matrix

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big((-1)^{q_\nu}\langle L_\mu|a^\dagger_i|L_\kappa\rangle\langle R_\nu|a^\dagger_j a_l a_k|R_\lambda\rangle\Big),\ j = l = k,\ i \neq j \tag{6.6}$$

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big((-1)^{q_\nu}\langle L_\mu|a^\dagger_i a^\dagger_j a_l|L_\kappa\rangle\langle R_\nu|a_k|R_\lambda\rangle\Big),\ i = j = l,\ k \neq l. \tag{6.7}$$

An exception occurs for the edge orbitals since it will require creating 4-index operators (similar rules as in the previous case will apply).

4. $\forall_{ijkl}$ if $i = j = k = l$. One can create these matrix elements when the Left block is expanded by orbital i, because then we know the exact representation of the 4–index operator for a single orbital. We can expand it very easily on–the–fly to the full Left basis without storing of the

operator explicitly using

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big(\langle L_\mu|a_i^\dagger a_j^\dagger a_l a_k|L_\kappa\rangle\delta_{\nu\lambda}\Big),\ i=j=k=l, \tag{6.8}$$

$$\Gamma_{ijkl} = \sum_{\mu\nu}\sum_{\lambda\kappa} c_{\mu\nu}c_{\lambda\kappa}\Big(\delta_{\mu\kappa}\langle R_\nu|a_i^\dagger a_j^\dagger a_l a_k|R_\lambda\rangle\Big),\ i=j=k=l. \tag{6.9}$$

Hence, to obtain a two–body density matrix corresponding to a single wave function and to avoid storing the 3– and 4–index quantities, it is sufficient to construct different Γ elements at different steps of a sweep of the one–site algorithm. Using the above procedure, one can always produce the two–body density matrix during the converged last sweep, provided that we have access to the 2–index operators. The above statement can be generalized, and one can produce a n–body reduced density matrix provided that in the block, which is coupled to the expanded block, we have access to the n–label operators, and that in the expanded block we produce on–the–fly the n–index operators by performing the direct product of (n-1)– or (n-2)–index operators with the site orbital. For example, one can construct a 3–body reduced density matrix if one stores the 3–labels operators in the Right block and produces on–the–fly in the Left block all needed 3–index operators by coupling of two– and one–index operators with the expansion orbital.

6.1.1 Computational cost

Let us discuss the computational cost of producing the two–body density matrix. In order to produce Γ_{ijkl} elements, the matrix representation of the second–quantized operators from the Left and Right blocks has to be multiplied by wave function coefficients. The required multiplication has the following form

$$\Gamma_{ijkl} = \sum_{\mu'\nu'}\sum_{\mu\nu} c_{\mu\nu}\langle L_\mu|\hat{O}^L|L_{\mu'}\rangle\langle R_\nu|\hat{O}^R|R_{\nu'}\rangle c_{\mu'\nu'}, \tag{6.10}$$

where $\hat{O}^L$ and $\hat{O}^R$ denote the second quantized operators in the Left and Right blocks, respectively. Such a multiplication can be performed in two steps. In the first step, the $\langle R_\nu|\hat{O}^R|R_{\nu'}\rangle$ matrix

elements are multiplied by a vector

$$X^{ik}_{\nu\mu'} = \sum_{\nu'} \langle R_\nu | \hat{O}^R | R_{\nu'} \rangle c_{\mu'\nu'}, \tag{6.11}$$

and the $X^{ik}_{\nu\mu'}$ intermediate term is created. We have explicitly written the i, and k labels that represent orbitals (the assignment of the orbital labels may be different, but the general procedure of the multiplication remains still unchanged). To prepare the $X^{ik}_{\nu\mu'}$ intermediates $O(M^3k^2)$ operations are required. Next, a similar intermediate can be formed by performing a multiplication of $\langle L_\mu | \hat{O}^L | L_{\mu'} \rangle$ by a vector,

$$Y^{jl}_{\nu\mu'} = \sum_{\mu} c_{\mu\nu} \langle L_\mu | \hat{O}^L | L_{\mu'} \rangle. \tag{6.12}$$

As in the previous multiplication the cost of this one scales as $O(M^3k^2)$ since k^2 operators are multiplied, and the cost of single such multiplication scales as $O(M^3)$. The final, third operation is to multiply the two intermediates together

$$\Gamma_{ijkl} = \sum_{\mu'} \sum_{\nu} Y^{jl}_{\nu\mu'} X^{ik}_{\nu\mu'}. \tag{6.13}$$

In this multiplication, one performs $O(M^2)$ operations for k^4 operators; thus, the scaling is proportional to $O(M^2k^4)$. Consequently, to obtain the two–body density matrix the number of operations required is $O(M^2k^4) + O(M^3k^2)$. Note that for the two–body matrix we have k^4 and k^2 factors in the scaling, for the three–body density matrix (since it is a 6–index quantity) they become k^6 and k^3. Hence, the cost of preparing the three–body density matrix is proportional to $O(M^2k^6) + O(M^3k^3)$.

6.1.2 Numerical results

Let us discuss the quality of the two–body density matrix built from different matrix elements obtained at different steps within a sweep for the one– and two–site algorithm. It is hard to find an objective criterion that will allow us to compare the quality of the results, but we can calculate the energy using this two–body density matrix and compare it to the lowest energy point

Table 6.1: The absolute value of the energy difference in H between the energy calculated from the two–body density matrix and the DMRG run, $\Delta E = |E(2bdm) - E(DMRG)|$, for several M values. The calculated state is the singlet state of the A_1 symmetry for the CH_2 molecule in the cc-pVDZ basis set.

M	one–site	two–site
120	1.2×10^{-7}	3.3×10^{-4}
180	1.3×10^{-7}	2.1×10^{-4}
250	5.7×10^{-8}	1.1×10^{-4}
350	6.1×10^{-8}	9.8×10^{-5}
500	4.5×10^{-8}	8.7×10^{-5}

in the final sweep. The deviation from the lowest energy value is a measure of the accuracy of the approximation used in the construction of the two–body density matrix. One has to stress, however, that the way of building the two–body density matrix is not a unique one. Thus, our results concerning the two–body density matrix may have been slightly different if the way of constructing the two–body density matrix had been chosen differently. However, the general trends should stay the same. Our results are presented in Table 6.1. One can easily see that for smaller numbers of states M, we do not obtain good results from the two–site algorithm, and the deviation between the energy obtained from the two–body density matrix and DMRG calculated energy is significant. However, when M increases the energy difference decreases as expected. The increase of the number of states is, however, not a desired remedy since for large systems one may not be able to increase the number of states due to computer memory and time constraints. In the one–site DMRG, the energy difference between the DMRG and the two–body density matrix energy is independent of M and depends only on the convergence. Thus, the one–site DMRG is a good way of tackling the problem of obtaining a two–body density matrix corresponding to a single wave function. The best practical way to perform such a calculation is **i)** first perform the two–site calculations and then when the desired convergence is reached **ii)** switch to the one–site procedure and calculate the two–body density matrix at the last converged sweep of the one–site run.

Chapter 7

DMRG-SCF

7.1 Introduction

The description of electron correlation in molecular systems requires us to take into account non–dynamic as well as dynamic correlation. Non–dynamic correlation can be understood in a mathematical sense as the effect of several large coefficients in the wave function expansion

$$|\Psi\rangle = \sum_i c_i |\phi_i\rangle, \tag{7.1}$$

where $|\phi_i\rangle$ are all possible Slater determinants in a given Hilbert space. It can be understood intuitively that capturing the non–dynamic correlation is crucial for the qualitatively correct description of the wave function. Hence, the non–dynamic correlation contributions can be contained in a reference wave function that qualitatively describes the chemistry of the system. The small coefficients present in the wave function expansion can be associated with dynamic correlation. The dynamic correlation is mostly associated with quantitative improvement of the reference wave function. Such an improvement is made by accounting for excitations out of the reference state to the high energy virtual orbitals.

Electronic wave functions are often dominated by several electronic configurations thus reveal-

ing a serious contribution of the non–dynamic correlation. Such situations are usually encountered in the description of reaction processes in which chemical bonds are broken, but they also arise in ground state systems that are described by several resonance structures at the equilibrium geometry. It is desirable to be able to produce a qualitatively correct wave function that will illustrate the properties of the systems dominated by non–dynamic correlation. Such a wave function may also become a reference (orbital generator) for a description that will account for the missing dynamic correlation contribution. The most intuitive solution to the problem of recovering the non–dynamic correlation is to carry out a full configuration interaction (FCI) calculations in the orbital space (active space) responsible for creating the most important electronic configurations, while simultaneously optimizing the core and virtual orbitals that have double and zero occupancy, respectively. Such an approach known as CASSCF was developed by Roos *et al.* [107, 109] and was used to perform calculations for systems with several important electronic configurations. The obvious limitation of the CASSCF procedure is the size of the accessible active space since the FCI calculations have to be carried out in that active space causing the number of configurations to grow exponentially. To overcome this problem, a restricted active space self–consistent field (RASSCF) [110, 111, 112] approach was developed that further divides the active space into smaller subspaces, thereby trying to decrease the huge dimension of the active space. However, even RASSCF allowed one to access only a few applications in the field of inorganic chemistry due to the intrinsic high dimensionality of the problem and the difficulty in choosing the appropriate RAS spaces.

In recent years the density matrix renormalization group [9, 72] (DMRG) has been investigated in quantum chemistry resulting in many interesting papers [113, 114, 10, 14, 19, 20, 22, 29, 28, 115, 32, 33, 45, 116, 25, 26, 21, 117, 30, 94, 91, 31, 118, 119, 120, 46]. It was shown that DMRG is a multiconfigurational method with a polynomial scaling which provides a reference wave function able to recover the non-dynamic correlation contribution. The goal of this chapter is to take advantage of the strength of the DMRG method in recovering the non-dynamic correlation and its efficient treatment of strongly correlated systems in order to use it for calculations in the active space, while the orbitals outside the active space are optimized in the same fashion as in

the usual CASSCF procedure. The orbital optimization with DMRG in the active space will be called the DMRG-SCF method. One can hope that the DMRG-SCF procedure will enlarge the realm of applicability of the current CASSCF procedure due to the possibility of treating larger active spaces. Thus, DMRG-SCF only requires the user to chose an active space, and by using DMRG in the active space it automatically selects the appropriate contracted basis for the Hilbert space. Hence, the DMRG-SCF active space is incomplete, but unlike the RASSCF method, the reduction of the dimensionality in the active space does not require any user input. DMRG-SCF is also systematically improvable and can reach the FCI limit. Additionally, DMRG-SCF is also size-consistent and size–extensive in contrast with RASSCF active space methods.

The structure of this chapter is as follows: in Sec. 7.2, we present a short description of the symmetry– and spin–adapted DMRG method that is used in the active space, thus allowing to target different symmetry– and spin–states. Section 7.3 then discusses the method of obtaining the one– and two–body density matrices as a byproduct of the DMRG run. We take advantage of the variational character of the convergence through the sweep for the one–site algorithm to obtain the two–body matrix. In Sec. 7.4, we focus on the discussion of the orbital rotation scheme, first presenting a general sketch of the equations used. The second part of Sec. 7.4 will then address considerations characteristic for DMRG relating to the incomplete treatment of the active space. Illustrative numerical examples are presented in Sec. 7.5, and our conclusions are presented in Sec. 7.6.

7.2 The DMRG method in the active space

General algorithmic aspects of the DMRG method can be found in an excellent description by Chan and Head–Gordon [22]. Here we aim to give only very basic information that is essential to understand the truncation of the basis states present in DMRG method. Additionally, we mention very briefly the use of symmetries in DMRG since we use the symmetry– and spin–adapted DMRG method [46] in the active space.

In DMRG we split the orbital space into two orbital subspaces, Left and Right. The Left

subspace is built out of a smaller left subspace plus a single orbital, Left=left+orbital. Analogous notation is used for the Right subspace. The DMRG wave function that is spanned in these two subspaces has the following form,

$$\begin{aligned} |\Psi\rangle &= \sum_{\mu'\nu'} c_{\mu'\nu'} |L_{\mu'}\rangle |R_{\nu'}\rangle, \\ |\Psi\rangle &= \sum_{\mu\nu} \sum_{s^l s^r} c_{l_\mu s^l s^r r_\nu} |l_\mu\rangle |s^l\rangle |s^r\rangle |r_\nu\rangle, \end{aligned} \tag{7.2}$$

where s^l and s^r label the n_o possible occupancies of the single orbital (or orbital unit) attached to the left and right subspaces, and l_μ and r_ν mark the basis states in the left and right subspace, respectively. Such a form of the wave function can be expanded and rewritten in terms of the matrices that are responsible for the sequence of renormalization operations, namely,

$$\begin{aligned} |\Psi\rangle = \sum_{\mu\nu} \sum_{s^1,\ldots,s^n} \sum_{s^{n+1},\ldots,s^p} & (A[s^k]\ldots A[s^{n-1}])_{s^1\ldots s^{k-1};l_\mu} (\tilde{c}[s^n s^{n+1}])_{l_\mu r_\nu} \times \\ & (A[s^{n+2}]\ldots A[s^{p-k}])_{r_\nu;s^{p-k+1}\ldots s^p} |s^{(1)}\ldots s^{(n)}\rangle |s^{(n+1)}\ldots s^{(p)}\rangle, \end{aligned} \tag{7.3}$$

where $k < n$. The $(A[s^q])_{l_\gamma,l_\lambda}$ are $M \times M$ dimensional matrices used to parametrize the wave function. The labels $s^q = 1,\ldots,n_o$, indicate the occupancies of every orbital (or orbital unit) at the qth step of a sweep. The labels $\gamma = 1,\ldots,M$ and $\lambda = 1,\ldots,M$ correspond to the basis states in either left or right subspace. Hence, the DMRG wave function is a linear combination of products of $M \times M$–dimensional matrices. Consequently, instead of depending on an exponential number of parameters, as in the case of the FCI method, the DMRG parametrization depends only on the $O(M^2 k n_o)$ parameters since all the information about the wave function can be obtained while producing and storing k matrices that are $M \times M$–dimensional, where k is the number of orbitals in the chain, n_o is the number of possible occupancies of a single orbital, and M is the row (column) dimension of the A matrix. Thus, such a parametrization will converge towards the FCI energy for increasing values of M. Hence, if an efficient procedure for choosing the most important basis states for the description of the wave function is established, one can truncate effectively the row and column dimensions of the A matrix to the M most important states. One

of the most effective means of truncation is to produce the A matrices as the eigenvectors of the many–body reduced density matrix for the Left subspace due to summation over the states from the Right subspace,

$$\begin{aligned} D_{\mu'\nu'} &= \sum_{\lambda'} c_{\mu'\lambda'} c_{\nu'\lambda'} \\ D_{\mu'\nu'} A_{\nu'\kappa} &= a_\kappa A_{\mu'\kappa}, \quad \kappa = 1, \ldots, M, \\ A_{\mu'\kappa} &= A_{\mu s^n;\kappa} = (A[s^n])_{\mu\kappa}, \end{aligned} \tag{7.4}$$

where the A matrix is a matrix constructed from M eigenvectors corresponding to a_κ largest eigenvalues. An analogous procedure has to be applied for the Right subspace.

The A matrices also have to be constrained to take advantage of spatial and spin symmetry. A separate A matrix is constructed for every spatial symmetry block. Hence, every A matrix has an additional label that represents the irreducible representation. To take spin symmetry into account, one has to impose on the Left (Right) reduced density matrices from Eq. (7.4) that they commute with the $\hat{S}_L^2$ ($\hat{S}_R^2$) operator. Hence, the A matrix for the Left block is built out of the $\hat{S}_L^2$ eigenvectors since the reduced density matrix and $\hat{S}_L^2$ have a simultaneously diagonalizing basis. Consequently, the truncation procedure present in DMRG allows to choose automatically the basis states that are most important for the description of the space, thus providing the possibility of treating larger active spaces. Additionally, the symmetry– and spin–adapted DMRG method, which is used in the active space, is able to resolve spin states. Thus, a spin–adapted DMRG method, which can deal with large active spaces, may provide a robust tool for inorganic chemistry applications.

7.3 Construction of the one– and two–body density matrix in the DMRG method

In the DMRG–SCF method, the active space is described by DMRG. Thus, to perform orbital rotations in the DMRG scheme, one needs to construct the one–, $\Gamma_i^j = \langle\Psi|a_i^\dagger a_j|\Psi\rangle$, and the

two–body density matrix, $\Gamma_{ij}^{lk} = \langle\Psi|a_i^\dagger a_j^\dagger a_k a_l|\Psi\rangle$, where $|\Psi\rangle$ is the DMRG wave function and i, j, k, and l are the spin orbitals labels. A detailed description of the two–body density matrix construction within the DMRG run can be found in Chapter 6.

In the ordinary DMRG calculations, at every step of a sweep one stores $O(k^2M^2)$ elements on the disk, where k is the number of orbitals and M is the number of basis states kept. Such a small size of disk storage is needed because after previous summation with the integrals, only the 2–index operators are required to set up the Hamiltonian matrix [19, 22]. Consequently, the one–body density matrix can be constructed at every step of a sweep since it requires only 2-index operators from the Left and Right block. However, in the case of the two–body density matrix the situation is more complicated due to the truncation of the basis performed in the DMRG method. Thus, in the left (right) unexpanded orbital subspace, one cannot construct the 4–index operator (e.g. $a_{i_1}^\dagger a_{i_2}^\dagger a_{i_3} a_{i_4}$) needed for producing the two–body density matrix as a product of the available 2–index operators, since the resolution of the identity is truncated within the left (right) space. Hence, to construct the two–body density matrix at a particular step of a sweep, it would be necessary to store 4–index operators from the left and right blocks resulting in a severe disk storage increase from $O(k^2M^2)$ to $O(k^4M^2)$ at every step of a sweep.

To avoid this increased disk storage, instead of calculating the density matrix at one particular step in a sweep, most elements of the two-body density are constructed at the steps for which two labels of the density matrix are in the Left block and the other two labels are in the Right block. This omits certain elements in which the same label occurs multiple times (three or four times), but they conveniently may be calculated at the site corresponding to the repeated label, as the required operators can be constructed on the fly. The essential feature of this algorithm is the fact that different elements of the two-body density matrix are calculated at different steps of a sweep. This would create an N-representability problem unless the wave function at different steps in the sweep is exactly the same. This is not the case for the two–site algorithm.

For this reason, the monotonically convergent one–site DMRG algorithm is used to obtain the same wave function at every step of the sweep. Consequently, when using the one–site algorithm, different elements of the two–body density matrix obtained at different steps of a sweep still come

from a single wave function. However, the benefits of the one–site algorithm do not come without an additional price, and such a scheme is prone to get stuck in local minima far from the ground state wave function. To avoid such problems, we use the traditional two–site DMRG, which is much less prone to get stuck in a local minimum, and only at convergence, we do switch to the one–site procedure, in order to create a proper density matrix that corresponds to a single wave function. Such a scheme will later be denoted as a two–site/one–site scheme [38]. The drawback of such a scheme is that the final energy obtained, which comes from the one–site part of the run, is higher than the energy obtained from the two–site part of the run. At every step of a sweep, the disk storage required to produce such a density matrix is $O(k^2M^2)$; this is the usual DMRG storage requirement, and the two–body density matrix is a by-product of a regular DMRG run.

7.4 Solution of orbital optimization equations in DMRG-SCF

The basic algorithm that we use to solve the orbital optimization equations has been developed such that it can be used in an analogous fashion to solve for t-amplitudes in an internally contracted multireference coupled cluster (CC) theory, including the use of Brueckner orbitals. The multireference CC methodology is still under development, but it has guided the orbital optimization scheme that is also used in the MCSCF, CASSCF and DMRG procedures, as implemented in a Waterloo version of ACES2 [121]. For this reason, the scheme is somewhat different than typical schemes that are used for orbital optimization in CASSCF calculations [122]. The first part of this section is independent of DMRG and only requires knowledge of the reference one- and two-particle reduced density matrices in a spatial orbital basis that could be derived from a DMRG, CASSCF (or internally contracted MRCC) calculation. In the second part of this section, the focus is on DMRG specific issues, and aspects of both stability and efficiency will be addressed.

Stability issues become quite important if the DMRG basis is incomplete, such that the DMRG results might be somewhat removed from the full CI result in a given orbital subspace. In future

applications, it is anticipated that such "low M" DMRG calculations might be very useful when starting to explore an unknown chemical system. In particular, low M DMRG results could be used to generate an approximate set of active orbitals and to decide on an appropriate active space for a range of geometries. In addition, low M DMRG calculations could be used to perform a preliminary geometry optimization that can be refined afterwards. In practice, it is important therefore to achieve smooth convergence, particularly when the DMRG calculation itself is not very accurate due to low M.

7.4.1 General (DMRG-independent) considerations

The orbital rotation amplitudes are obtained from a coupled cluster like parametrization using single excitations only. It should be noted that the parameterization is non-unitary, and as a result the equations for the t_1–amplitudes are exactly fourth-order, and no approximations are needed in this respect. This can be contrasted with the traditional unitary parameterizations of CASSCF calculations that are typically truncated at second– or third–order. To establish some useful notation, we distinguish virtual orbitals, labeled a, b, active orbitals labeled m, n, core orbitals labeled i', j' , and finally the combination of active and core (or occupied) orbitals that are labeled i, j. Atomic orbitals are labeled with Greek indices, e.g. γ, δ, while generic orthogonal MO orbitals carry labels p, q. The coupled cluster orbital rotation equations for the OV (occupied-virtual) and CA (core-active) sectors read

$$\begin{aligned}
\langle R|\hat{E}^i_a e^{-\hat{T}}\hat{H}e^{\hat{T}}|R\rangle &= g^i_a(t) = 0,\\
\langle R|\hat{E}^{i'}_a e^{-\hat{S}}e^{-\hat{T}}\hat{H}e^{\hat{T}}e^{\hat{S}}|R\rangle &= \langle R|\hat{E}^{i'}_a e^{-\hat{S}}\hat{\bar{H}}e^{\hat{S}}|R\rangle = \bar{g}^{i'}_m(s),\\
\bar{g}^{i'}_m(s) &= 0,
\end{aligned} \tag{7.5}$$

where $|R\rangle$ is the reference wave function and $\hat{E}^p_q = \hat{p}^\dagger_\alpha \hat{q}_\alpha + \hat{p}^\dagger_\beta \hat{q}_\beta$ are unitary group generators. In case of DMRG, $|R\rangle$ is $|\Psi_{\mathrm{DMRG}}\rangle$ in the active space. The operator $\hat{T} = \sum_{i,a} t^i_a \hat{E}^a_i$ generates occupied to virtual excitations (OV-rotations), while $\hat{S} = \sum_{i',m} s^{i'}_m \hat{E}^m_{i'}$ generates excitations from core to active (CA-rotations). Analogously to the Fock space coupled cluster equations, the

equations for $\hat{T}$ and $\hat{S}$ can be solved sequentially, and the equations for $\hat{T}$ are independent of $\hat{S}$. The detailed equations only depend on the one- and two-particle reduced density matrices of the (DMRG) reference state in a spatial orbital basis. The use of spatial orbitals, which are the same for α and β spin, really means that the orbitals are optimized for a state-averaged ensemble that comprises a complete multiplet of eigenstates of the $\hat{S}^2$ operator. Hence, Eq. (7.5) implicitly averages over all components of a multiplet, as this results in the same spatial density matrices as for one member of the ensemble (that is used in practice). This is easily shown, if $|R\rangle$ is a rigorous spin eigenstate, when using the spin-ladder operators that commute with the generators of the unitary group.

Since the initial guess orbitals can be quite poor in the beginning of the orbital optimization procedure, the magnitudes of the $\hat{S}$ and $\hat{T}$ amplitudes are monitored not to exceed a threshold value using an "Augmented Hessian" solver [123]. In practice, this means that rather than solving $g_a^i(t) = 0$, we solve for

$$g_a^i(t) - \lambda(\mathbf{g}^\dagger(0)\mathbf{t})t_a^i = 0, \tag{7.6}$$

where λ is an integer with a value from the set $0, 1, 2, 4, 8, \ldots$. The actual chosen value of λ is the first (lowest) value from the sequence for which the norm of the vector of t-amplitudes does not exceed a threshold value (0.2 in the current version of the program). In the initial cycles of the calculation, λ can take on somewhat larger values (e.g., 8 or so), while as the orbitals become more reasonable the parameter approaches the value 0. To solve the non-linear equations for t, a linear quasi-Newton equation is solved to obtain the correction to the t-amplitudes, an approach that is also based on an augmented Hessian. This correction to the linear equation is easily rationalized, and it also clarifies ideas regarding the analogous correction term in the non-linear equation, Eq. (7.6). If the linear (Newton) equation of interest is $\mathbf{Ax} + \mathbf{y} = \mathbf{0}$, then this equation is replaced by the eigenvalue equation

$$\begin{pmatrix} \mathbf{A} & \mathbf{y} \\ \mu\mathbf{y}^\dagger & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{1} \end{pmatrix} = \epsilon \begin{pmatrix} \mathbf{x} \\ \mathbf{1} \end{pmatrix}$$

or simply

$$\mathbf{A}\mathbf{x} + \mathbf{y} - \mu(\mathbf{y}^\dagger\mathbf{x})\mathbf{x} = 0$$
$$\mu(\mathbf{y}^\dagger\mathbf{x}) = \epsilon \tag{7.7}$$

A parameter μ (from a set $0, 1, 2, 4, \ldots$) is used to solve for the correction vector $\mathbf{x}$, such that it remains reasonably small. Then the t–amplitudes are updated as $\mathbf{t}^{(n+1)} = \mathbf{t}^{(n)} + \mathbf{x}$, while the new residual is calculated using Eq. (7.6). Hence, a double layer of Augmented Hessian ideas is used to create a stable solution for the initial t-amplitudes that enter the orbital rotation step. A similar strategy is used to solve for the s–amplitudes that correspond to the core–active rotations.

From the $\hat{T}$ and $\hat{S}$ amplitudes and the original orbitals $|i\rangle$, $|m\rangle$ a new orthonormal set of orbitals $|\tilde{i}\rangle$, $|\tilde{\tilde{m}}\rangle$ is obtained as follows,

$$|\tilde{i}\rangle = |i\rangle + \sum_{a,\tilde{i}} t_a^{\tilde{i}} |a\rangle \tag{7.8}$$
$$|\tilde{\tilde{m}}\rangle = |m\rangle - \sum_{\tilde{i}'} s_m^{\tilde{i}'} |\tilde{i}'\rangle \tag{7.9}$$

These new active orbitals, with both the OV and CA-rotations mixed in, are orthonormalized using Löwdin orthonormalization. Subsequently, the new core orbitals $|\tilde{i}'\rangle$ are orthonormalized to the active orbitals and then subjected to Löwdin orthogonalization. The virtual space is then orthogonalized to the new occupied space and put in orthonormal form using the Löwdin procedure.

At this point, we have shown how from the input one- and two-particle reduced density matrices and the Hamiltonian integrals one can obtain an improved set of orbitals. We have also assumed that the method is invariant under orbital rotations within the core, active, and virtual spaces, the same as for the CASSCF case. For DMRG, the precise definition of the active orbitals is relevant, and the method is not invariant under the orbital rotations in the active space (this will be addressed in the next subsection). A weakness of the direct application of the above orbital rotation scheme is a lack of coupling between the many-body reference state and the optimized

orbitals. Without introducing such a coupling one can easily get oscillations: one set of orbitals A gives rise to a state $|R_1\rangle$ for which the orbitals are rotated to a set B. If in this orbital set one obtains a state $|R_2\rangle$ for which the orbital rotation procedure would yield orbitals close to the original set A, one cannot (or would only very slowly) reach convergence by a straightforward iterative procedure. In principle, one can design a procedure in which both the orbital and electronic state parameters are optimized simultaneously, and this has been done very effectively in CASSCF calculations. However, our ultimate goal is to include dynamic correlation in the form of a coupled cluster ansatz, and it is likely that a coupling scheme would be quite tedious to implement for such methods.

Partly for reasons of generality, a DIIS scheme [124, 125] is used both to accelerate convergence and to introduce stability when there is a danger of oscillations. The DIIS scheme is very similar to the scheme used in the Brueckner CC equations [126, 127, 128, 129]. The new orbitals are related to a set of reference orbitals by a unitary operator that can be parameterized as $U = e^{\Lambda}$, $\Lambda^{\dagger} = -\Lambda$. From the known matrix U (in a reference MO basis), one can easily calculate $\Lambda = ln(U)$, using the series expansion for the natural logarithm. The DIIS extrapolation is used for the operator Λ and for the one-particle reduced density matrix of the reference state (in the AO basis). The inclusion of this latter matrix is responsible for the coupling between orbital-rotation parameters and the many-body state.

The DIIS extrapolation yields a new operator Λ^{DIIS}, and a new set of orbitals is obtained as $U^{\mathrm{new}} = e^{\Lambda^{\mathrm{DIIS}}}$. Because extrapolation is performed on the antisymmetric amplitudes, they always define an orthonormal orbital basis. If the Λ^{DIIS} parameters are too large, the reference orbitals and the DIIS procedure are restarted using the current (best) set of orbitals as a new set of reference orbitals. Such a restart avoids some potential alignment/ordering issues of the orbitals and a possible ambiguity of signs that might affect the DIIS procedure.

The above discussion summarizes our current orbital optimization strategy. Such a strategy is also applicable to CASSCF and internally contracted Brueckner MRCC or to the state-selective equation of motion method. All of these approaches are invariant to rotations of the orbitals in the active space. The following subsection addresses some additional issues that are peculiar to

DMRG.

7.4.2 DMRG-specific considerations

For a full CASSCF calculation, the precise definition of the active orbitals does not matter because FCI is used in the active space, and consequently the wave function is invariant under a rotation of orbitals in the core, active and virtual subspaces separately. However, because the DMRG wave function is incomplete in the active space, the results do depend (slightly) on the precise definition of the active orbitals, and the DMRG result also depends on the ordering of these active orbitals. This is a somewhat delicate issue in DMRG, and our current strategy can probably be improved, in the sense that for a particular M basis states in DMRG, we cannot claim to reach the lowest possible energy, in particular regarding orbital ordering.

Our current goal is to converge smoothly to a local minimum that is reasonably close to the optimal (unknown) result. Natural orbitals (as available in the current cycle) are used to define the active orbitals, and the ordering of the orbitals can be specified on input (as defined by the symmetry of the orbitals and the rank of the occupation numbers in a particular symmetry block). Alternatively, by default the natural orbitals are ordered from most strongly to most weakly occupied. This works well if all occupation numbers are fairly different, but if some of them are nearly degenerate, it is better to specify the ordering on input. Moreover, if natural occupation numbers are nearly degenerate within a symmetry block, we do not yet have a robust procedure, particularly if these nearly degenerate natural orbitals fluctuate during the iterative procedure. In general, as the natural orbitals and their occupation numbers change during the convergence procedure, this strategy can easily lead to oscillations, particularly if the DMRG results depend strongly on the nature of the natural orbitals (this most often happens when a low value of M is used).

In a more sophisticated convergence procedure that is currently used, at some point a reference set of natural orbitals in the active space is defined, as well as a reference DMRG-ordering of these orbitals. In subsequent cycles, the overlap between the reference active orbitals $|m_0\rangle$ (in the AO basis) and current (natural) orbitals $|\tilde{n}\rangle$ is obtained, $X_{mn} = \langle m_0|\tilde{n}\rangle$, and an orbital rotation is

defined such that the current orbitals have maximum overlap with the reference $|m_0\rangle$ orbitals. Since the orbital subspaces spanned by these respective orbitals are slightly different, the overlap in the rotated basis is not exactly diagonal, but it is typically close to diagonal. Here, we use precisely the same procedure as we use in diabatization schemes in vibronic coupling applications [130, 131], and such a procedure requires only an overlap matrix as an input. This procedure amounts to a locking of the natural orbitals, such that they change in a minimal fashion as required by the orbital optimization procedure. This also allows us to maintain a well-defined ordering of the orbitals, at least starting from the point when a reference set of active orbitals is defined. As in the DIIS procedure, the locking of the active orbitals is restarted when the overlap between the reference and true natural orbitals is too far from being diagonal. Hence, at convergence, the DMRG orbitals and ordering are not precisely the natural orbitals, as at some point in the procedure they are locked to a reference set of orbitals that is obtained earlier in the procedure. This means that it will be hard or impossible to exactly reproduce our results for low M calculation using another DMRG-SCF code without knowing the details of the locking procedure. In this sense, this version of DMRG-SCF is not a model chemistry. In principle, the same procedure can be used for MCSCF calculations that are also not invariant under orbital rotations. Using such a procedure one would not obtain an exact minimum, however.

It may be of some interest to reflect on an alternative strategy to this somewhat elusive problem of the definition and ordering of orbitals in DMRG. In principle, one can also minimize the energy with respect to orbital rotations and ordering in the active space. As M in DMRG reaches the value that almost gives the FCI energy, all of these orbital optimization parameters would become redundant; and this indicates that this orbital optimization problem can easily become nearly singular as the quality of the DMRG calculation improves. Moreover, there will be a continuous transition between orbital orderings that could be alternately described by orbital rotations. It is not unlikely that different orbital orderings would define different local minima in the overall energy landscape. We also know that there exist many different close lying minima corresponding to different partitionings of the M basis states over the different sectors with a particular number of electrons, spin and spatial symmetry. All of these considerations together are sufficiently

discouraging that we do not fancy tackling the problem of active orbital optimization/ordering in this way. Locking the orbitals to achieve smooth convergence to a somewhat arbitrary point on the energy landscape close to the optimal solution is all we can hope to achieve.

A final issue to address is entirely specific to DMRG. As one is close to convergence, it is advantageous to start the DMRG calculation from the previous set of operators and wave functions at each step in a sweep. Since the DMRG-operators almost do not change when the orbitals are close to convergence, the operator matrix elements are an excellent starting point. In addition, in the later stages of the calculation it is advantageous to lock the number of basis states in each partition (symmetry, number of electrons, spin state) to avoid oscillations. As for the precise definition of the active orbitals, restarting DMRG and locking the number of states per partition implies we do not have full control over the final digits of the energy, and our final results do depend (very slightly) on arbitrary decisions in the implementation, e.g., when to lock, or the initial starting point of the orbitals or guess. In the end, the calculation is path dependent, but there is no chemical significance to the arbitrariness in the final energies. The arbitrariness exists because one wants to achieve a smooth convergence and minimize the number of steps that is needed to get a converged result. We introduce the following terminology: we will call a "macroiteration" a complete DMRG run and a calculation of the orbital rotation parameters afterwards. The step and sweep refers to the typical DMRG calculation that is performed in the active space. In our current implementation, the DMRG-SCF equations are converged in three stages:

1. *Initial macroiterations.* The DMRG calculation is started from scratch, the perturbative correction [50, 38] is applied in conjunction with the two-site DMRG algorithm. Convergence criteria are fairly crude, and in the final step the switch is made to the one-site DMRG algorithm that is used without the perturbative correction since the starting point is provided by the two–site algorithm. In the one–site part of the run, a one- and (nearly) N-representable two-particle density matrix is calculated.

2. *Intermediate macroiterations.* The calculation is restarted from the converged DMRG operators obtained during the previous macroiteration. The two-site DMRG is used and a

10 times smaller perturbative correction [50, 38] is applied to achieve slightly more robust convergence. Again, the density matrices are obtained at convergence after switching to the one–site DMRG algorithm without the perturbative correction.

3. *Improved/final macroiterations.* Only the monotonically convergent one–site algorithm is used without perturbative correction. The DMRG calculation is restarted from the previous macroiteration DMRG result. Using the one-site algorithm means that the number of DMRG basis states per partition is locked. Convergence threshold of both the Davidson procedure and the sweeping algorithm is increased in these stages of the calculation. In the final stage of the calculation (very close to convergence), the DMRG algorithm typically performs only one or two sweeps to reach convergence, and one can converge to essentially arbitrary (or machine) precision.

7.5 Numerical results

7.5.1 Comparison with CASSCF

We have carried out a CASSCF and DMRG-SCF calculation for the HN_3 system, illustrated by Fig. 7.1, targeting a singlet state of A_1 symmetry. We treat the HN_3 molecule as a model system for a strongly correlated small system since it has a significant number of occupation numbers very different from 0 or 2. There are 12 orbitals and 10 electrons placed in the active space, and the calculations are done using the cc-pVTZ [132] basis set. Figure 7.2 shows the energy plotted as a function of the NN bond stretch, while the NH bond and the angles are held fixed.

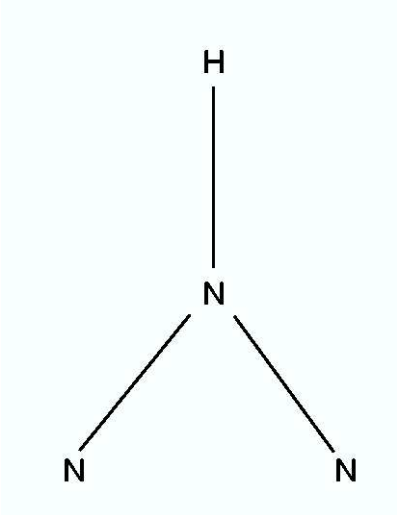

Figure 7.1: The HN_3 model system. The NN bonds will be stretched. The NNN angle is 118 deg and the HN bond length is 1.035 $\AA$.

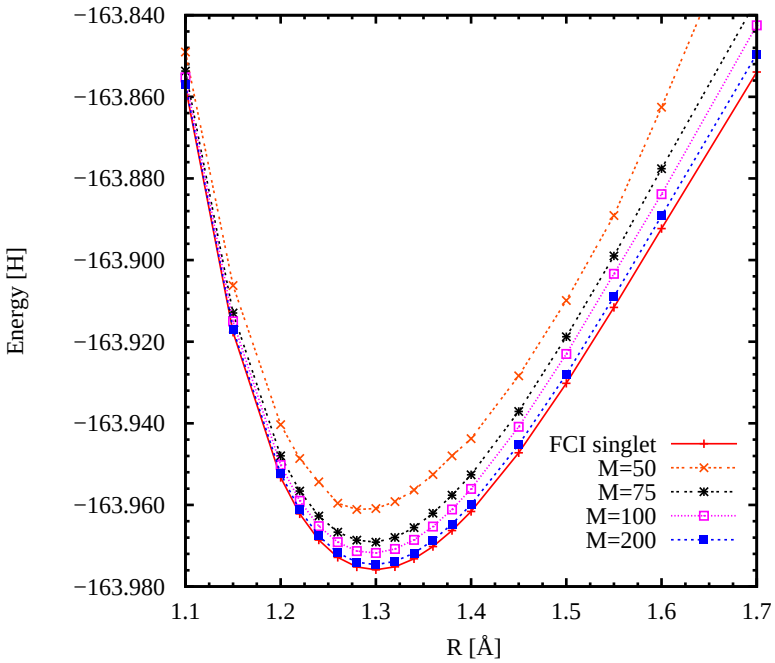

Figure 7.2: The DMRG-SCF energy for a singlet state of A_1 symmetry of a HN_3 model system as a function of NN bond stretch. The cc-pVTZ [132] basis set was used and a (12, 10) active space was employed for the active space calculations.

One can see that the energy minimum is placed correctly even for very low M calculations, and that for subsequent larger values of M the curve gets closer to the CASSCF curve. Additionally, the energy differences are plotted in Fig. 7.3 to show that the curves obtained remain smooth even for small M. Only the lowest M curve ($M = 50$) is not smooth, and there are a few jumps when the energy is compared vs. CASSCF energy. However, one has to stress that it is likely that this small number of basis states ($M = 50$) is simply not large enough to illustrate the complexity of the system since there is a very significant improvement for $M = 75$, and starting from there the differences with the CASSCF present a similar qualitative behavior. Thus, as expected, there is a minimal number of states needed to illustrate the basic complexity of the system, and only the results for M values larger than such a minimal value can be qualitatively trusted as descriptive for the system of interest. Finally, one can see that the energy differences from the CASSCF are

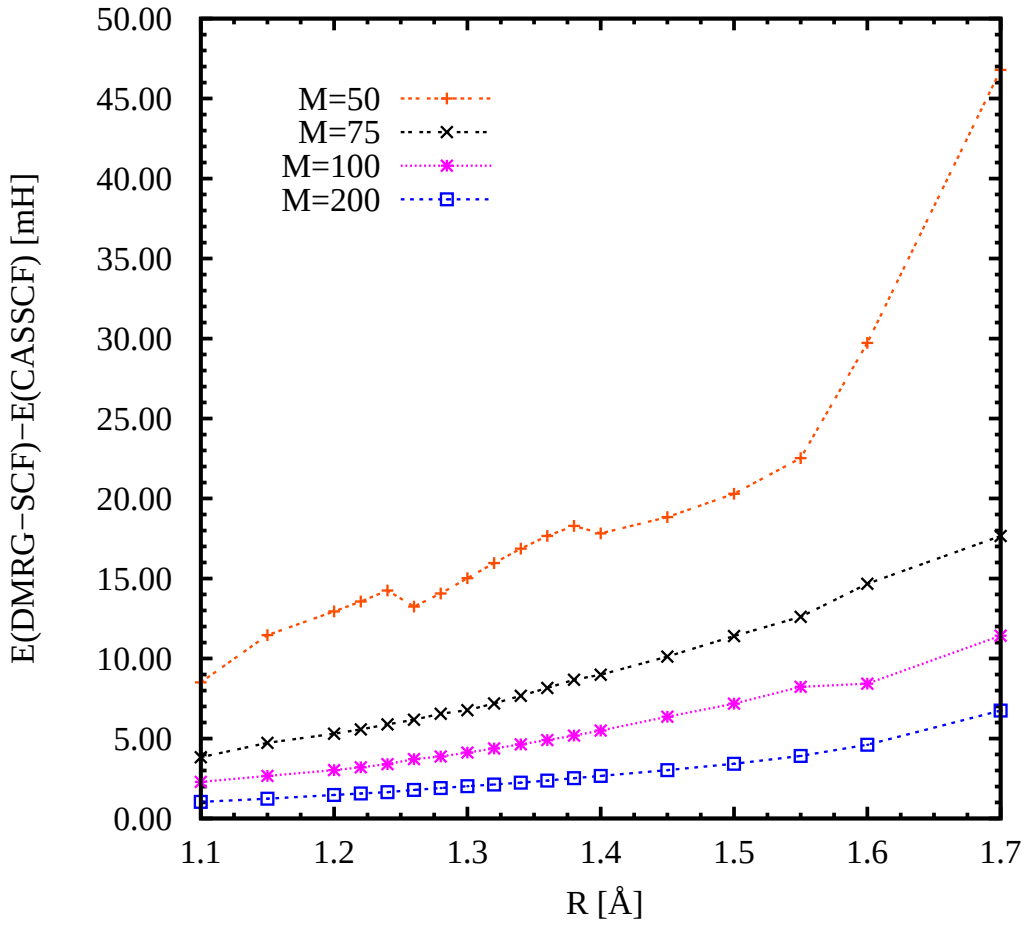

Figure 7.3: The energy difference between DMRG-SCF and CASSCF for a singlet state of A_1 symmetry of a HN_3 model system as a function of NN bond stretch. The cc-pVTZ [132] basis set was used and (12, 10) active space was employed for the active space calculations.

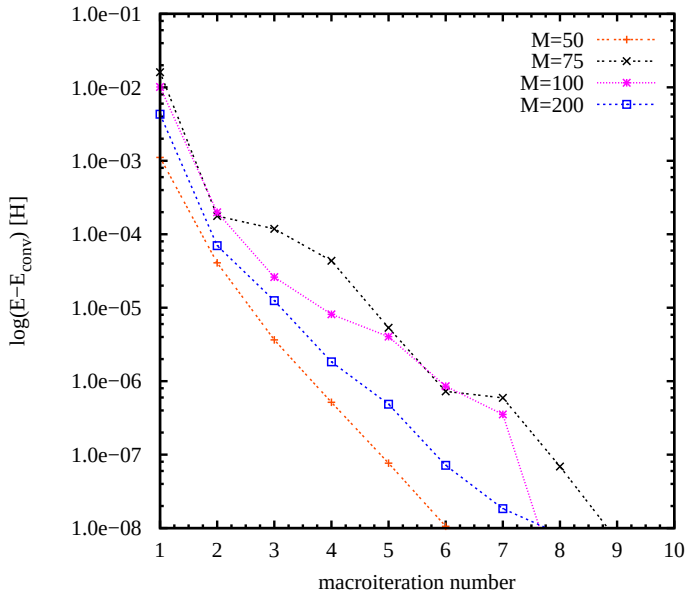

Figure 7.4: The logarithm of the difference between the final converged energy and the energy obtained during the subsequent DMRG-SCF macroiterations for a singlet state of A_1 symmetry of a HN_3 model system. The cc-pVTZ [132] basis set was used and a (12, 10) active space was employed for the active space calculations.

approaching the exact result in a very systematic manner, allowing extrapolation of the results for several different M values to the CASSCF limit.

7.5.2 Energy convergence profile

For the HN_3 example, Fig. 7.4 shows the energies obtained at every macroiteration for several M values. The desired convergence for the t_1 amplitudes was 10^{-5}. One can see that for all M values used in the simulation, the energy after certain macroiteration does not change significantly, and it is possible therefore to save a lot of time while restarting the macroiteration from the matrices obtained in the previous macroiteration. To obtain high convergence of the energy (e.g. 10^{-7} H), it is important to lock the number of kept eigenvectors per partition. This is achieved only using the one–site DMRG algorithm that gives a fully converged two-body density matrix. In addition, the natural orbitals are locked using the diabatic scheme described in Sec. 7.4. The

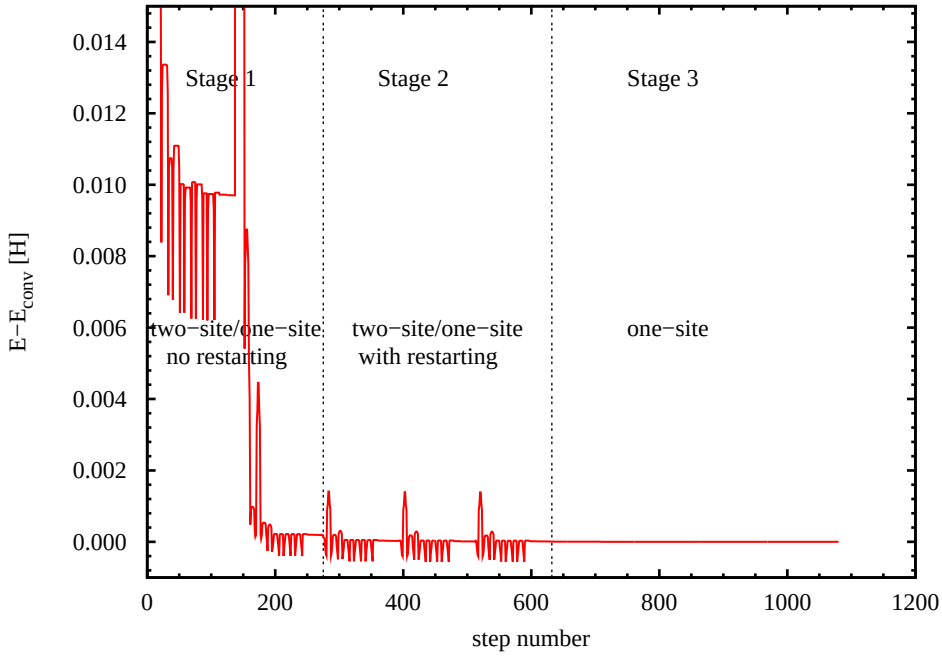

Figure 7.5: The energy difference between the converged final energy and the energies obtained at the subsequent DMRG steps during a single DMRG-SCF macroiteration for the A_1 singlet state of HN_3 at $R = 1.28$ $\AA$ in the cc-pVTZ [132] basis set, and a (12, 10) active space was employed for the active space calculations. The three stages of the DMRG-SCF run are illustrated.

one–site algorithm has the additional advantage that it is about four times cheaper than the two–site algorithm. The choice of when to switch between the subsequent stages of the algorithm, like the two–site/one–site without restarting, the two–site/one–site with restarting, and finally the one–site is of course open to experimentation. We know that our current procedure of switching between the subsequent stages clearly requires more tuning. The primary reason why we choose to switch between the different versions of the DMRG algorithm during different DMRG–SCF macroiterations is the cost reduction of the DMRG run.

To illustrate the problem, the energy convergence profile for a full DMRG–SCF calculation for HN_3 is shown in Fig. 7.5, where all the subsequent stages of the run are marked. The energy change between the beginning of the first sweep and the final step of the first macroiteration, see stage 1 of the run (two–site/one–site without restarting), is usually very significant and reaches 2.6

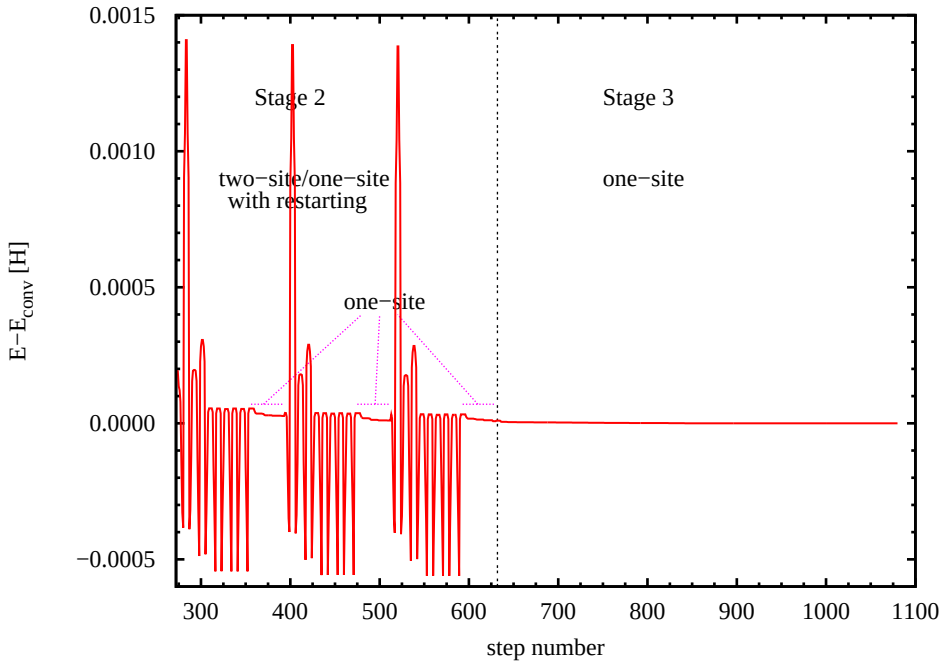

Figure 7.6: The energy difference between the converged final energy and the DMRG energies obtained at the stage 2 and stage 3 during a single DMRG-SCF macroiteration for the A_1 singlet state of HN_3 at $R = 1.28$ $\AA$ in the cc-pVTZ [132] basis set was used and a (12, 10) active space was employed for the active space calculations.

H in the current example. Once the t_1 rotation amplitudes are small, and the orbital minimization enters stage 2, the energy change during the run is small since every macroiteration is restarted from the matrices obtained during the previous macroiteration.

We present an enlargement of stage 2 and stage 3 in Fig. 7.6. It is possible to see that the energy obtained during the one–site part of the two–site/one–site run is higher than the energy obtained in the two–site part of the run. However, since we produce the two–body density matrix in the one–site part of the run, the one–site energy is considered as the final energy. Hence, to achieve a high convergence, we use only the one–site algorithm without a perturbation [50, 38], because such a scheme is less costly. The final stage 3 of the calculations takes advantage of only the one–site algorithm; one even resigns from the use of the perturbative corrections since at this stage the eigenvectors in the partitions are distributed almost optimally, due to restarting from

Table 7.1: The energy in H for the $(H_2)_{10}$ hydrogen chain as a function of the symmetric stretch of the bond distance. The DZ basis set and a (20, 20) active space was used.

	$M = 100$		$M = 200$		$M = 300$	
$R[\AA]$	two–site	one–site	two–site	one–site	two–site	one–site
0.70	-10.236952	-10.236781	-10.242937	-10.241664	-10.244932	-10.243940
0.90	-10.925346	-10.922076	-10.933774	-10.931448	-10.938354	-10.936379
0.95	-10.955546	-10.951696	-10.966742	-10.964268	-10.972012	-10.969831
1.00	-10.956176	-10.952117	-10.969949	-10.967170	-10.975782	-10.973391
1.05	-10.934885	-10.930493	-10.947184	-10.946881	-10.957699	-10.954923
1.10	-10.898006	-10.894938	-10.914202	-10.910766	-10.923164	-10.920071
1.30	-10.650756	-10.646707	-10.682331	-10.679208	-10.698105	-10.691530
1.50	-10.349893	-10.346140	-10.402966	-10.400826	-10.421936	-10.427312

the previous iteration (as illustrated in the stage 3 in Fig. 7.6). Hence, the previously obtained energy from stage 2 is improved, and the result can be converged, if needed, to very high accuracy.

7.5.3 Symmetric stretch of a hydrogen chain

We have analyzed a symmetric stretch for a singlet state of A_1 symmetry of a $(H_2)_{10}$ hydrogen chain in the DZ [133, 134] basis set. All the bonds starting from 0.70 $\AA$ were stretched reaching the subsequent lengths 0.90, 0.95, 1.00, 1.05, 1.10, 1.30, and 1.50 $\AA$. The calculation employed 20 orbitals and 20 electrons in the active space, which is beyond the capabilities of ordinary CASSCF due to the size of the FCI calculations within the active space. Such a system is a good example of a highly multireference system, where the stretched geometry has all the occupation numbers in the active space significantly different from 0 or 2. The number of one–body density matrix eigenvalues n between $1.95 \geq n \geq 0.05$ is 3 for 0.70 $\AA$, 6 for 0.90–1.00 $\AA$, 8 for 1.05–1.10 $\AA$, 13 for 1.30 $\AA$, and 20 for 1.50 $\AA$.

The energies obtained for different values of the bond distance are listed in Table 7.1. We report both the one–site energy and the two–site energy since the one–site algorithm was used to obtain the two–body density matrix free from the N–representability problem, while the two–site algorithm was used before we switched to the one–site procedure. Our calculation was pronounced converged if either the rotation amplitudes first reached a value of 10^{-5}, or the energy was changing

less than 10^{-6} H. The greatest number of macroiterations needed during this calculation was 13 for the $M = 300$, $R = 0.70$ $\AA$ point, but most of the other points took around 8 macroiterations.

7.5.4 Cr_2Mn_2 example

Finally, we consider the Cr_2Mn_2 metal cluster as an example of where we would like to push the application of the DMRG-SCF method. Here, we present only very preliminary results since the current algorithm still requires more tuning. The atoms were put on the corners of a square in such a way that same atoms are on the diagonal, and the length of the square edge is 1.497 $\AA$. A singlet state of A_1 symmetry was calculated using the Ahlrichs-VDZ [1, 2] basis set. All the $3d$–orbitals from chromium and from manganese were taken to the active space; this yielded 24 orbitals and 22 electrons in the active space. The calculation was carried out for $M = 150$. This calculation is certainly quite far from the FCI limit, and a much larger number of states will be needed to approach that limit. However, here we only want to illustrate that the low M calculations can be used to perform an initial orbital optimization.

Figure 7.7 illustrates how the energy drops in the subsequent macroiterations. During the first 17 macroiterations, the new DMRG-SCF macroiteration is not restarted from the previous DMRG result since the optimized electronic state may be changing during the early macroiterations when the orbitals are very far from the optimal ones. Thus, by not restarting one tries to avoid locking the electronic state too early. The restarting starts at the 18th macroiteration, and the two–site/one–site scheme is restarted from the previous matrices. The last three macroiterations (29-31) are performed solely within the one–site algorithm scheme since they yield only a tiny energy improvement. The difference between the energy obtained in the initial and final macroiteration is approximately 5 H indicating that the starting orbitals were of very poor quality. This calculation illustrates the stability of the algorithm since it allows the result to be converged to the 10^{-6} H. Such a high convergence clearly is not needed during the initial calculation of orbitals, but may be useful for geometry optimization. Thus, the DMRG-SCF method is a stable and relatively robust tool allowing one to reach convergence even for low M calculations and difficult systems such as transition metal clusters.

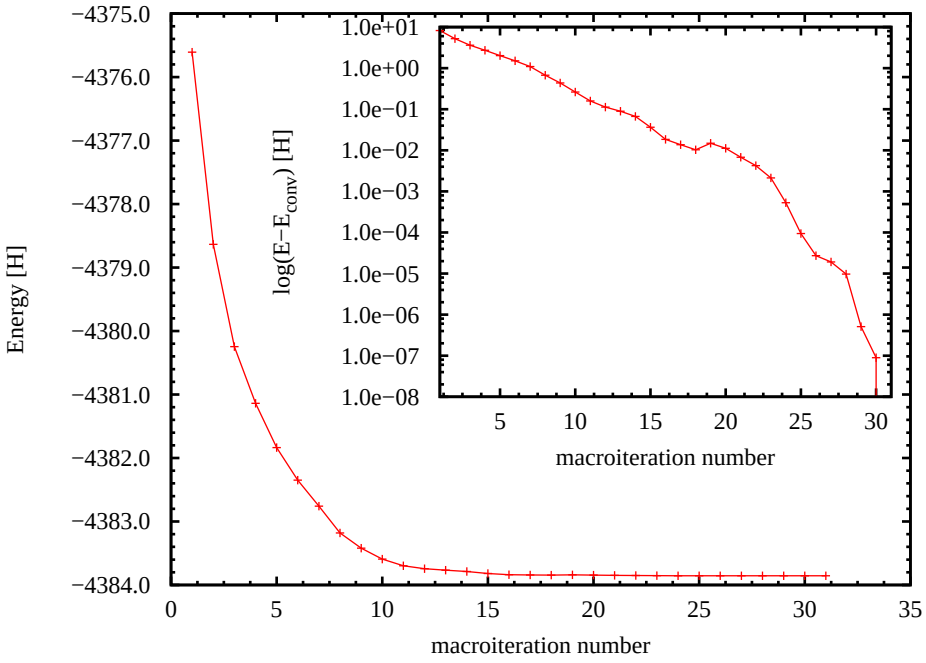

Figure 7.7: The energy in H obtained during the DMRG-SCF macroiterations with $M = 150$ for a singlet state of A_1 symmetry of Cr_2Mn_2 system in the Ahlrichs-VDZ [1, 2] basis set with 24 orbitals and 22 electrons in the active space. The inset contains logarithm of the difference between the final energy and the energy obtained during subsequent iterations.

7.6 Conclusions

We have implemented a DMRG–SCF method that is an orbital optimization scheme with the DMRG method used in the active space. Such a scheme allows one to treat active spaces larger than those currently accessible to the CASSCF method, since CASSCF is limited by the scaling of the full CI calculation used for the description within the active space. Our scheme scales like $O(M^3k^3)$ in the active space, where M is the number of basis states used in DMRG, and k is the number of the active space orbitals. Hence, the scaling of the method is polynomial in the active space. Outside the active space, the scaling is proportional to $O(p^5)$ of the 4-index transformation, where p is the number of orbitals in the AO basis set. In the active space, we use the spin– and symmetry–adapted DMRG scheme that has the same cost as the traditional DMRG scheme, but the reduced density matrix obtained in this scheme commutes with $\hat{S}^2$, hence providing the possibility of resolving the spin states. The only "information" required from DMRG by the orbital rotation part is the one– and two–body density matrix. The two–body density matrix is calculated using the one–site scheme that allows one to obtain a two–body density matrix free from the N–representability problem. The orbital rotation amplitudes are obtained from a coupled cluster like parameterization using only single excitations; thus, the parameterization is non-unitary, and as a result the equations for the t_1–amplitudes are exactly fourth-order. Our scheme is analogous to the CASSCF method without explicit coupling of the orbital rotation coefficients to the DMRG wave function. Consequently, it would be advantageous to introduce such an explicit coupling between vector coefficients and orbital rotation parameters, which is commonly employed in CASSCF calculations [122], to reduce the number of required macroiterations and to improve the stability of the convergence procedure.

DMRG-SCF is a method in which the active space is incomplete, and thus it is not invariant under the active–active rotation. Consequently, the low M calculations, where active–active rotations are not negligible, were discussed and the specific methods to account for it were mentioned. However, for high M values, where the DMRG energy approaches the FCI result, the DMRG-SCF is not impaired by the existence of active–active rotations since they become negligible. To test the robustness of the approach, a comparison between DMRG-SCF results for various M

values and CASSCF results for the HN_3 molecule was made. It was concluded that even the low M calculation have qualitatively correct features. Additionally, to illustrate the stability of the convergence, the exact scheme that we use to lead the calculation towards convergence (namely the two–site/one–site without restarting, the two–site/one–site with restarting, and finally the one–site scheme) was discussed using the HN_3 example with a (12, 10) active space and using Cr_2Mn_2 with a (24, 22) active space. Clearly, other variants of the convergence scheme are also possible and improvements of the existing variant will be considered in the future. Finally, a calculation for a symmetric stretch of $(H_2)_{10}$ hydrogen chain with 20 orbitals and 20 electrons in the active space was presented.

One can hope that the DMRG–SCF scheme will be used as an orbital generator in multireference approaches when high quantitative accuracy is required from the calculations. If one aims to get a qualitative accuracy, the DMRG–SCF scheme can be used by itself, and it can provide results for cases in which the active spaces are beyond the reach of the FCI method. It was demonstrated earlier by Chan's group that DMRG performs extremely well for 1D systems [33] and is able to treat even 100 electrons in 100 active orbitals. Consequently, one can expect that the DMRG–SCF method will provide an excellent way to perform calculations and get qualitatively correct results for chains of metal oxides or polycarbons with multiple bonds. Such applications are well beyond the traditional CASSCF method.

The application of the DMRG–SCF method to 3D systems poses a bit more difficulties since usually one cannot expect that the wave function for 3D systems can benefit significantly from the DMRG local ansatz that is very efficient for 1D systems. However, one can expect that this problem may be remedied since multidimensional DMRG schemes are currently under development [41, 42, 43]. Nevertheless, even though the current DMRG is not as effective in 3D as in 1D, it still can be used to treat in a systematic, reliable way challenging 3D systems with up to $O(40)$ orbitals in the active space (e.g. molecular magnets, and iron–sulphur active complexes in proteins).

Chapter 8

Towards inclusion of dynamic correlation

In the previous chapters, the distinction between dynamic and non–dynamic correlation was made, and we have learned that dynamic correlation is responsible for the quantitative improvement of the wave function and energy, while the qualitative features of the wave function are illustrated when considering the non–dynamic correlation. It was also explained that non–dynamic correlation is connected to the multireference nature of the wave function associated with the electron correlation in nearly degenerate valence orbitals, while dynamic correlation can be associated with energy excitations into the virtual orbitals. The inclusion of the non–dynamic correlation effects is necessary when considering bond breaking, radicals or transition states. Although it may seem that the correction coming from the dynamic correlation is relatively small, it usually has to be included to achieve good agreement with experiments. One also has to stress that most calculations are carried in a way that provides first a multireference wave function (e.g. from CASSCF), and then dynamic correlation is added on top of it. However, the character of the multireference wave function can change if it is obtained in the presence of dynamic correlation in the first place. Consequently, the two types of the correlation are intermixed, and often cannot be considered separately.

The success of the DMRG method lies in recovering the most important basis states from the full possible Hilbert space for the description of the targeted state. These states have the largest coefficients in the wave function expansion. The efficiency of the truncation procedure hinges on the assumption that the density matrix for the system of interest has a rapidly decaying spectrum [104]; thus, a few vectors corresponding to the largest eigenvalues suffice to represent the system very accurately. Although such a flexible DMRG truncation is a blessing if one wants to include the states responsible for the non–dynamic correlation and multireference nature, it is a curse if one wants to include the dynamic correlation since it requires an enormous number of states corresponding to the small eigenvalues of the reduced density matrix produced in DMRG.

Consequently, a compromise has to be made, and the DMRG method may be the best choice if one wants to obtain a reference state using a DMRG-SCF procedure. Afterwards, the dynamic correlation can be added to this reference. Obviously, one of the best candidates to bring correlation into the DMRG–SCF reference wave function is the coupled cluster method; examples of such an approach may be found in Refs. [47, 48, 49]. However, due to the complexity of such an approach, it is far beyond the scope of this thesis to try to familiarize the reader with that branch of development. Instead, we will try to describe possible employment of the DMRG-SCF wave function in n–electron valence state perturbation theory (NEVPT) [51, 52, 53], and the possible use of the DMRG or DMRG–SCF wave function in density functional calculations (DFT) with the split two–electron integrals in the short– and long–range part of the Coulomb operator [54, 55, 56].

8.1 NEVPT with DMRG–SCF wave function

Multireference perturbation theory (MRPT) is used to describe the correlation energy in systems that are not described properly by a single determinant wave function. The MRPT methods belong mostly to two cases. In the first, an effective Hamiltonian is constructed in the model space, and then it is diagonalized [135, 136]. In the second, the zero order Hamiltonian is diagonalized in a chosen space, then a perturbation is applied to that zero order variational wave function. Complete active space perturbation theory (CASPT2) [137, 138] is one of the most well known

methods that belongs to the second family. In CASPT2, the zero order wave function, which is obtained from the FCI calculations in the active space, is perturbed by applying the excitation operators and using a Møller–Plesset type of monoelectronic zero order Hamiltonian. CASPT2 can suffer from the intruder state problem that can cause the perturbation series to diverge. Such a problem arises when the eigenvalue of the zero order Hamiltonian and the energy of the wave function from the outer space are very close, thus causing the denominator to be close to zero. Although such difficulties can be removed by introducing denominator shifts, this is a rather *ad hoc* way of solving the problem that is caused by lack of balance in the treatment of the interaction with the zero order Hamiltonian of the CAS space and the outer determinantal space. One remedy involves introducing Dyall's Hamiltonian [139] as a zero order Hamiltonian that consists of the complete active space Hamiltonian and the Møller–Plesset zero order Hamiltonian in the inactive and secondary spaces. Dyall's Hamiltonian, marked later as H^D, has the reference as an eigenfunction without the necessity of projection; it naturally incorporates denominator shifts that appear in terms of active space Fock operators, and it does not give rise to intruder states. We discuss here only the strongly contracted variant of NEVPT2; however a partially contracted as well as uncontracted variants are also possible.

The wave function of interest $|\Psi_m\rangle$ is approximated by a zero order wave function $|\Psi_m^{(0)}\rangle$ that can be obtained from CASCI or CASSCF calculations,

$$P_{\mathrm{CAS}} H P_{\mathrm{CAS}} |\Psi_m^{(0)}\rangle = E_m^{(0)} |\Psi_m^{(0)}\rangle, \tag{8.1}$$

where P_{CAS} is s projector onto the CAS space S. The CASCI wave function has all the possible occupation patterns in the active space, but the same pattern of occupation in the inactive space, and thus it can be written as an antisymmetrized product of the inactive and active parts

$$|\Psi_m^{(0)}\rangle = |\Phi_c \Psi_m\rangle. \tag{8.2}$$

The perturbed wave functions have different occupation pattern in the inactive and active space. They are created from $|\Psi_m^{(0)}\rangle$ when k electrons are excited from the core to the active space, or

when k electrons are excited from the active to the virtual space. Consequently, the number of active electrons changes as $-2 \leq k \leq 2$, and thus the perturbed wave functions have the form $|\Phi_l^{(-k)}\Psi_\mu^{(k)}\rangle$. Every perturbed wave functions belongs to the $S_l^{(k)}$ spaces, where k is the number of electrons promoted or removed from the active space, and l labels the occupation pattern of the inactive orbitals. Different versions of the NEVPT theory are defined depending on the number of perturbed functions chosen from the $S_l^{(k)}$ subspaces. In the strongly contracted version of the NEVPT, we chose only one perturbed function from each subspace $S_l^{(k)}$. Such a function is obtained from the the action of the perturbation on the zero order wave function,

$$P_{S_l^{(k)}} H |\Psi_m^{(0)}\rangle = V_l^{(k)} |\Psi_m^{(0)}\rangle, \tag{8.3}$$

$$|\Phi_l^{(-k)}\Psi_\mu^{(k)}\rangle = V_l^{(k)} |\Psi_m^{(0)}\rangle, \tag{8.4}$$

where $P_{S_l^{(k)}}$ denotes a projector onto $S_l^{(k)}$.

The possible $S_l^{(k)}$ spaces can be assigned to eight classes that differ by the number of electrons excited between the active, core, and virtual orbitals. A pictorial representation of these excitation classes is given in Fig. 8.1. The eight possible classes of perturbation operators are listed below

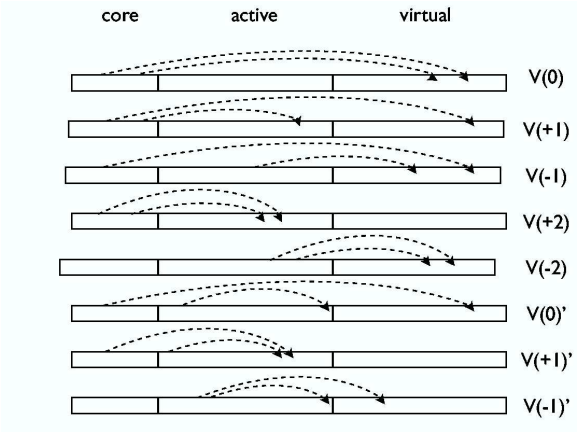

Figure 8.1: Eight possible types of excitations between the core, active, and virtual space

$$\hat{V}^{(0)}_{ij,rs} = \langle rs||ji\rangle a^\dagger_r a^\dagger_s a_i a_j, \tag{8.5}$$

$$\hat{V}^{(+1)}_{ij,r} = \sum_a \langle ra||ji\rangle a^\dagger_r a^\dagger_a a_i a_j, \tag{8.6}$$

$$\hat{V}^{(-1)}_{i,rs} = \sum_a \langle rs||ai\rangle a^\dagger_r a^\dagger_s a_i a_a, \tag{8.7}$$

$$\hat{V}^{(+2)}_{ij} = \sum_{a<b} \langle ab||ji\rangle a^\dagger_a a^\dagger_b a_i a_j, \tag{8.8}$$

$$\hat{V}^{(-2)}_{rs} = \sum_{a<b} \langle rs||ba\rangle a^\dagger_r a^\dagger_s a_a a_b, \tag{8.9}$$

$$\hat{V}^{(0)'}_{i,r} = \sum_a \sum_b \langle ra||ib\rangle a^\dagger_r a^\dagger_a a_b a_i + \sum_j \langle rj||ij\rangle a^\dagger_r a^\dagger_j a_j a_i + \langle r|h|i\rangle a^\dagger_r a_i, \tag{8.10}$$

$$\hat{V}^{(+1)'}_{i} = \sum_{a<b} \sum_c \langle ab||ic\rangle a^\dagger_a a^\dagger_b a_c a_i + \sum_j \sum_a \langle aj||ij\rangle a^\dagger_a a^\dagger_j a_j a_i + \sum_a \langle a|h|i\rangle a^\dagger_a a_i, \tag{8.11}$$

$$\hat{V}^{(-1)'}_{r} = \sum_a \sum_{b<c} \langle ra||cb\rangle a^\dagger_r a^\dagger_a a_b a_c + \sum_a \sum_j \langle rj||aj\rangle a^\dagger_r a^\dagger_j a_j a_a + \sum_a \langle r|h|a\rangle a^\dagger_r a_a, \tag{8.12}$$

with $i, j, \ldots, a, b, \ldots$ and $r, s, \ldots$ denoting the core, active, and virtual indices, respectively. Every term $\hat{V}^k_l$ can be written as a product of two terms where

(i1) the first is a simple product of creation/annihilation operators with inactive indices, $\hat{A}^k_i$.

(i2) the second is a sum of terms obtained by multiplying an integral by a product of creation/annihilation operators with only active indices, $\hat{V}^k_l$.

$$|\Phi^{(-k)}_l \Psi^{(k)}_l\rangle = \hat{A}^k_l \hat{V}^k_l |\Phi_c \Psi_m\rangle = |(\hat{A}^k_l \Phi_c)(\hat{V}^k_l \Psi_m)\rangle \tag{8.13}$$

The norm of the above wave function can be calculated noting that $\langle \hat{A}^k_l \Phi_c | \hat{A}^k_l \Phi_c\rangle = 1$, and thus one only needs to evaluate N^k_l as

$$N^k_l = \langle \Phi^{(-k)}_l \Psi^{(k)}_l | \Phi^{(-k)}_l \Psi^{(k)}_l\rangle = \langle \Psi_m | (\hat{V}^k_l)^\dagger \hat{V}^k_l | \Psi_m\rangle. \tag{8.14}$$

The $\hat{A}_i^k$ and ${}^a\hat{V}_l^k$ operators for all the possible classes are listed below

$$\hat{A}^{(0)}_{ij,rs} = a_r^\dagger a_s^\dagger a_i a_j, \qquad \hat{V}^{(0)}_{ij,rs} = \langle rs||ji\rangle, \tag{8.15}$$

$$\hat{A}^{(+1)}_{ij,r} = a_r^\dagger a_i a_j, \qquad \hat{V}^{(+1)}_{ij,r} = \sum_a \langle ra||ji\rangle a_a^\dagger, \tag{8.16}$$

$$\hat{A}^{(-1)}_{i,rs} = a_r^\dagger a_s^\dagger a_i, \qquad \hat{V}^{(-1)}_{i,rs} = \sum_a \langle rs||ai\rangle a_a, \tag{8.17}$$

$$\hat{A}^{(+2)}_{ij} = a_i a_j, \qquad \hat{V}^{(+2)}_{ij} = \sum_{a<b} \langle ab||ji\rangle a_a^\dagger a_b^\dagger, \tag{8.18}$$

$$\hat{A}^{(-2)}_{rs} = a_r^\dagger a_s^\dagger, \qquad \hat{V}^{(-2)}_{rs} = \sum_{a<b} \langle rs||ba\rangle a_a a_b, \tag{8.19}$$

$$\hat{A}^{(0_1)'}_{i,r} = a_r^\dagger a_i, \qquad \hat{V}^{(0_1)'}_{i,r} = \sum_a \sum_b \langle ra||ib\rangle a_a^\dagger a_b, \tag{8.20}$$

$$\hat{A}^{(0_2)'}_{ij,r} = a_r^\dagger a_j^\dagger a_j a_i, \qquad \hat{V}^{(0_2)'}_{ij,r} = \sum_j \langle rj||ij\rangle, \tag{8.21}$$

$$\hat{A}^{(0_3)'}_{i,r} = a_r^\dagger a_i, \qquad \hat{V}^{(0_3)'}_{i,r} = \langle r|h|i\rangle, \tag{8.22}$$

$$\hat{A}^{(+1_1)'}_{i} = a_i, \qquad \hat{V}^{(+1_1)'}_{i} = \sum_{a<b} \sum_c \langle ab||ic\rangle a_a^\dagger a_b^\dagger a_c, \tag{8.23}$$

$$\hat{A}^{(+1_2)'}_{ij} = a_j^\dagger a_j a_i, \qquad \hat{V}^{(+1_2)'}_{ij} = \sum_j \sum_a \langle aj||ij\rangle a_a^\dagger, \tag{8.24}$$

$$\hat{A}^{(+1_3)'}_{i} = a_i, \qquad \hat{V}^{(+1_3)'}_{i} = \sum_a \langle a|h|i\rangle a_a^\dagger, \tag{8.25}$$

$$\hat{A}^{(-1_1)'}_{r} = a_r^\dagger, \qquad \hat{V}^{(-1_1)'}_{r} = \sum_a \sum_{b<c} \langle ra||cb\rangle a_a^\dagger a_b a_c, \tag{8.26}$$

$$\hat{A}^{(-1_2)'}_{r,j} = a_r^\dagger a_j^\dagger a_j, \qquad \hat{V}^{(-1_2)'}_{r,j} = \sum_a \sum_j \langle rj||aj\rangle, \tag{8.27}$$

$$\hat{A}^{(-1_3)'}_{r} = a_r^\dagger, \qquad \hat{V}^{(-1_3)'}_{r} = \sum_a \langle r|h|a\rangle a_a. \tag{8.28}$$

Let us now discuss how the second order correction to the energy can be obtained. We know that once the energies $E_l^{(k)}$ associated with the perturbed functions $|\Psi_l^{(k)}\rangle$ are found, the zero order Hamiltonian can be written as

$$H_0 = \sum_{l,k} |\Psi_l^{(k)'}\rangle E_l^{(k)} \langle \Psi_l^{(k)'}| + \sum_m^{\text{CAS}} |\Psi_m^{(0)}\rangle E_m^{(0)} \langle \Psi_m^{(0)}|, \quad |\Psi_l^{(k)'}\rangle = \frac{1}{\sqrt{N_l^{(k)}}} |\Psi_l^{(k)}\rangle. \tag{8.29}$$

Note that the zero order Hamiltonian in the virtual space is only defined in the space spanned

by the perturbed wave functions $|\Psi_l^{(k)}\rangle$, and not in the complete Hilbert space. However, only these basis states are required to develop the perturbation theory. In the next step, the first order correction to the wave function can be found as

$$\begin{aligned} \Psi_m^{(1)} &= \sum_{kl} |\Psi_l^{(k)'}\rangle \frac{\langle \Psi_l^{(k)'} | H | \Psi_m^{(0)} \rangle}{E_m^{(0)} - E_l^{(k)}} = \sum_{kl} |\Psi_l^{(k)'}\rangle \frac{1}{\sqrt{N_l^{(k)}}} \frac{\langle \Psi_l^{(k)} | V_l^{(k)} | \Psi_m \rangle}{E_m^{(0)} - E_l^{(k)}} \\ &= \sum_{kl} |\Psi_l^{(k)'}\rangle \frac{\sqrt{N_l^{(k)}}}{E_m^{(0)} - E_l^{(k)}}, \end{aligned} \tag{8.30}$$

and consequently the second order energy can be written as

$$E_m^{(2)} = \sum_{kl} \frac{\langle \Psi_l^{(k)'} | H | \Psi_m^{(0)} \rangle^2}{E_m^{(0)} - E_l^{(k)}} = \sum_{kl} \frac{N_l^{(k)}}{E_m^{(0)} - E_l^{(k)}}. \tag{8.31}$$

We now will make use of the previously mentioned Dyall's Hamiltonian

$$H^D = H_i + H_v, \tag{8.32}$$

where the original Dyall's Hamiltonian is separated into a diagonal one–electron Hamiltonian for the inactive space,

$$H_i = \sum_i^{\text{core}} \epsilon_i a_i^\dagger a_i + \sum_r^{\text{virt}} \epsilon_r a_r^\dagger a_r + C, \tag{8.33}$$

and a two–electron Hamiltonian representing the active space

$$H_v = \sum_{ab}^{\text{act}} h_{ab} a_a^\dagger a_b + \frac{1}{2} \sum_{abcd}^{\text{act}} \langle ab|cd\rangle a_a^\dagger a_b^\dagger a_d a_c. \tag{8.34}$$

The h_{ab} integral is defined as

$$h_{ab} = \langle a|h + \sum_i (J_i - K_i)|b\rangle, \tag{8.35}$$

and the C constant

$$C = \sum_i^{\text{core}} \langle i|h|i\rangle + \frac{1}{2} \sum_{ij}^{\text{core}} \langle ij||ij\rangle - \sum_i^{\text{core}} \epsilon_i, \tag{8.36}$$

is introduced to make H^D equivalent to the full Hamiltonian in the CAS space, thus providing

the definition of zero order energy as

$$H^D \Psi_m^{(0)} = E_m^{(0)} \Psi_m^{(0)}. \tag{8.37}$$

The energies ϵ_i and ϵ_r appearing in the one–electron H_i are defined in the following way

$$\epsilon_i = -\langle a_i \Psi_m^{(0)} | H | a_i \Psi_m^{(0)} \rangle + E_m^{(0)}, \tag{8.38}$$

$$\epsilon_r = \langle a_r^\dagger \Psi_m^{(0)} | H | a_r^\dagger \Psi_m^{(0)} \rangle - E_m^{(0)}. \tag{8.39}$$

The energies of the basis functions $\Psi_l^{(k)}$ used in the perturbative expansion are defined using Dyall's Hamiltonian

$$E_l^{(k)} = \frac{1}{N_l^{(k)}} \langle \Psi_l^{(k)} | H^D | \Psi_l^{(k)} \rangle, \tag{8.40}$$

hence ensuring that the interaction between the active orbitals is taken properly into account. Additionally, it also ensures computational simplicity by requiring only the calculation of the expectation values between $|\Psi_m^{(0)}\rangle$ and certain operators. On the contrary, some other non–diagonal zero order Hamiltonians require iterative solving of a set of equations, which may be cumbersome and computationally demanding.

We can now rewrite Eq. (8.40) as

$$\begin{aligned} N_l^{(k)} E_l^{(k)} &= \langle \Psi_m^{(0)} | V_l^{(k)\dagger} H^D V_l^{(k)} | \Psi_m^{(0)} \rangle \\ &= \langle \Psi_m^{(0)} | V_l^{(k)\dagger} V_l^{(k)} H^D | \Psi_m^{(0)} \rangle + \langle \Psi_m^{(0)} | V_l^{(k)\dagger} [H^D, V_l^{(k)}] | \Psi_m^{(0)} \rangle, \end{aligned} \tag{8.41}$$

and using the previously defined $E_m^{(0)}$ we obtain

$$\begin{aligned} E_l^{(k)} &= E_m^{(0)} + \frac{1}{N_l^{(k)}} \langle \Psi_m^{(0)} | V_l^{(k)\dagger} [H^D, V_l^{(k)}] | \Psi_m^{(0)} \rangle \\ &= E_m^{(0)} + \Delta\epsilon_l + \frac{1}{N_l^{(k)}} \langle \Psi_m^{(0)} | V_l^{(k)\dagger} [H_v, V_l^{(k)}] | \Psi_m^{(0)} \rangle. \end{aligned} \tag{8.42}$$

The sum of the orbital energies of the electrons promoted into the virtual space minus the orbital

energies of the holes created in the core space is is denoted $\Delta\epsilon_l$. The correction to the energy coming from each of the eight subspaces $S_l^{(k)}$ will have the following general form

$$E(S_l^{(k)}) = \frac{N(S_l^{(k)})}{a\epsilon_x + b\epsilon_y + \frac{1}{N(S_l^{(k)})}\langle\Psi_m^{(0)}|V_l^{(k)\dagger}[H_v, V_l^{(k)}]|\Psi_m^{(0)}\rangle}, \tag{8.43}$$

which suggests that the accuracy of the norm $N(S_l^{(k)})$ will play a major role. The cumbersome terms containing the three– and four–body density matrix in $\langle\Psi_m^{(0)}|V_l^{(k)\dagger}[H_v, V_l^{(k)}]|\Psi_m^{(0)}\rangle$ are probably of lesser importance, and one can expect that their approximation by a suitable method may not have a major influence on the accuracy of the results. We list all the density matrices appearing in the numerators and the denominators for all of the eight possible subspaces in Table 8.1.

Table 8.1: The density matrices required to calculate the second order contribution to the NEVPT energy for all of the possible subspaces. 1-, 2-, 3- and 4-b mean one–, two–, three–, and four–body density matrix, respectively. No dm required means that the knowledge of the molecular integrals is required to calculate the correction.

subspace	norm	denominator
$S_{ij,rs}^{(0)}$	no dm required	no dm required
$S_{ij,r}^{(+1)}$	1-b	1-b, 2-b
$S_{i,rs}^{(-1)}$	1-b	1-b, 2-b
$S_{ij}^{(+2)}$	1-b, 2-b	1-b, 2-b, 3-b
$S_{rs}^{(-2)}$	1-b, 2-b	1-b, 2-b, 3-b
$S_{i,r}^{(0)'}$	1-b, 2-b	1-b, 2-b, 3-b
$S_i^{(+1)'}$	1-b, 2-b, 3-b	1-b, 2-b, 3-b, 4-b
$S_r^{(-1)'}$	1-b, 2-b, 3-b	1-b, 2-b, 3-b, 4-b

There are several ways of obtaining either the exact or the approximate three–body density matrix:

- calculation of the exact 3–body density matrix using the DMRG method. This involves storing the 3–index operators in the active space, thus requiring $O(M^2k^3)$ storage at every step of a sweep, where M is the number of basis states, and k is the number of active orbitals. The number of multiplications that is performed scales as $O(M^2k^6) + O(M^3k^3)$. Hence,

the overall cost of DMRG calculations in the active space is increased from the $O(M^2k^4) + O(M^3k^2)$ required to produce the two–body density matrix to $O(M^2k^6) + O(M^3k^3)$. For the exact discussion of the cost of preparing a many–body density matrix in DMRG, see Chapter 6.

- calculation of the exact terms involving the 3–body density matrix summed with the appropriate integrals using some DMRG tricks that will be explained later.

- approximation [140] of terms involving the higher order density matrices by *cumulant decomposition* [141, 142, 143].

One has the following options to proceed in order to calculate the second order correction to the energy. One can calculate the three–body density matrix exactly, and approximate only the four–body density matrix by cumulant decomposition. However, due to the severe increase in computational cost, it would be highly desirable to avoid calculation of the exact three–body density matrix. Of course, there also exists the most approximate option in which all three– and four–body density matrices appearing in both numerators and denominators are approximated using the cumulant decomposition.

Since one can expect from Eq. (8.43) that the accuracy of the numerators may be quite important to obtain the correct result, it may be advantageous to prepare only the norms using DMRG tricks (explained later), thus causing factorization. Additionally, the cost of the factorization needed for the evaluation of the norm is comparable to that for the preparation of the two–body density matrix. On the other hand, the cumbersome terms in the denominators containing the three– and four–body density matrix in $\langle \Psi_m^{(0)} | V_l^{(k)\dagger} [H_v, V_l^{(k)}] | \Psi_m^{(0)} \rangle$ are probably of lesser importance, and one can expect that their approximation by a suitable method may not have major influence on the accuracy of the results. One possible way [140] of approximation of the denominator terms with the higher order density matrices is *cumulant decomposition* [141, 142, 143].

To further validate the assumption that the approximate reconstruction of the elements involving the three– and four–body density matrices in the denominators should not affect the results considerably, one can use the data presented by Angeli *et al.* in Ref. [53]. For all internuclear

distances between 3.20–20.00 bohr, they list the energy contribution of the eight subspaces to the second order perturbation correction for the ground state of the Cr_2 molecule. The details of the analysis seen in Table 8.2 show that the greatest contribution for all distances comes from the $S^{(-1)}_{i,rs}$ subspace that requires only one– and two–body density matrices. The next greatest contribution comes from the $S^{(0)}_{ji,rs}$ subspace that also requires only one– and two–body density matrices. The subspaces $S^{(-2)}_{rs}$ and $S^{(0)'}_{i,r}$, which require three–body density matrix, yield similar approximately third largest contributions to energy. Consequently, the proper cumulant decomposition of the 3–body density matrix is important, but one can also expect that this can be done quite efficiently. The subspaces $S^{(1)'}_{i}$ and $S^{(-1)'}_{r}$, which require the four–body density matrix, make the smallest contribution to the energy from all subspaces. Hence, even if the cumulant decomposition of the four–body density matrix may not have very high accuracy, the total energy still should not be impaired.

Table 8.2: Percentage contribution to the second order correction to the energy for the ground state of Cr_2 coming from the various subspaces. The contributions were calculated based on the data presented in Ref. [53]

subspace	$R = 3.20$ (Bohr)	$R = 4.00$ (Bohr)	$R = 20.00$ (Bohr)
$S^{(0)}_{ij,rs}$	18.29	19.91	21.44
$S^{(+1)}_{ij,r}$	7.39	7.45	7.70
$S^{(-1)}_{i,rs}$	27.72	31.64	34.51
$S^{(+2)}_{ij}$	3.52	2.65	2.43
$S^{(-2)}_{rs}$	17.72	17.90	17.05
$S^{(0)'}_{i,r}$	19.90	17.65	16.85
$S^{(+1)'}_{i}$	0.55	0.18	0.00
$S^{(-1)'}_{r}$	4.89	2.60	0.00

We will focus now on discussion of the option explained above, namely, DMRG tricks to calculate the three–body density matrices in the numerators and approximation of the three– and four–body density matrices present in the denominators by cumulant expansion. Let us first discuss how the N^k_l norms, which be later needed for calculation of the second order correction to the energy, can be calculated for each of the eight possible subspaces. The detailed derivation

of these expressions can be found in Appendix C.

To calculate the norm of the perturbed functions created from the action of $V(0)$, $V(+1)$, $V(-1)$, $V(+2)$, $V(-2)$, and $V(0)'$ on the $|\Psi_m^{(0)}\rangle$,

$$N_{ij,rs}^{(0)} = \langle rs||ji\rangle^2 \tag{8.44}$$

$$N_{ij,r}^{(+1)} = \langle ra||ji\rangle^2 - \sum_a \sum_b \langle ra||ji\rangle\langle rb||ji\rangle\Gamma_a^b, \tag{8.45}$$

$$N_{i,rs}^{(-1)} = \sum_a \sum_b \langle rs||ai\rangle\langle rs||bi\rangle\Gamma_b^a, \tag{8.46}$$

$$N_{ij}^{(+2)} = \langle ab||ji\rangle^2 - \sum_{a<b}\langle ab||ji\rangle\Bigg(\sum_{a<d}\langle ad||ji\rangle\Gamma_b^d + \sum_{b<d}\langle bd||ji\rangle\Gamma_a^d + \tag{8.47}$$

$$+\sum_{c<a}\langle ca||ji\rangle\Gamma_b^c - \sum_{c<b}\langle cb||ji\rangle\Gamma_a^c + \sum_{c<d}\rangle\langle cd||ji\rangle\Gamma_{ba}^{cd}\Bigg),$$

$$N_{rs}^{(-2)} = \sum_{a<b}\sum_{c<d}\langle rs||ba\rangle\langle rs||dc\rangle\Gamma_{cd}^{ba}, \tag{8.48}$$

$$N_{i,r}^{(0)'} = \sum_{abc}\langle ra||ib\rangle\Bigg(\langle ra||ic\rangle\Gamma_c^b - \sum_d\langle rd||ic\rangle\Gamma_{ac}^{bd}\Bigg) + 2\sum_{cd}\langle rc||id\rangle\Gamma_d^c h_{ir}^{\text{eff}} + (h_{ir}^{\text{eff}})^2 \tag{8.49}$$

one only requires the one– and two–body density matrices that can be constructed in the DMRG–SCF procedure, as was shown in the previous chapters. We have introduced above the auxiliary notation,

$$h_{ir}^{\text{eff}} = \sum_{j\neq i}\langle rj||ij\rangle + \langle r|h|i\rangle. \tag{8.50}$$

The situation becomes more complicated if one wants to calculate the norms of the perturbed functions for the $V(+1)'$ and $V(-1)'$ subspaces since they require the three–body density matrix

involving the active space labels, namely

$$
\begin{aligned}
N_i^{(+1)'} &= \sum_{a<b}\sum_{c}\sum_{d<e}\sum_{f}\langle ab||ic\rangle\langle de||if\rangle(\delta_{ad}\delta_{be}\Gamma^c_f-\delta_{ad}\Gamma^{ce}_{bf}-\delta_{bd}\delta_{ae}\Gamma^c_f+ \\
&\quad +\delta_{bd}\Gamma^{ce}_{af}+\delta_{ae}\Gamma^{cd}_{bf}-\delta_{be}\Gamma^{cd}_{af}+\Gamma^{cde}_{baf}) \\
&\quad +\sum_{a}\sum_{d<e}\sum_{f}h^{\rm eff}_{ia}\langle de||if\rangle(\delta_{ad}\Gamma^e_f-\delta_{ae}\Gamma^d_f+\Gamma^{de}_{af})+ \\
&\quad +\sum_{a<b}\sum_{c}\sum_{d}h^{\rm eff}_{id}\langle ab||ic\rangle(\delta_{ad}\Gamma^c_b-\delta_{bd}\Gamma^c_a+\Gamma^{cd}_{ba})+\sum_{ad}h^{\rm eff}_{ia}h^{\rm eff}_{id}(\delta_{ad}-\Gamma^d_a), \qquad (8.51)\\
N_r^{(-1)'} &= \sum_{a}\sum_{b<c}\sum_{d}\sum_{e<f}\langle ra||cb\rangle\langle rd||ef\rangle(\delta_{ad}\Gamma^{cb}_{ef}-\Gamma^{cbd}_{aef}) \\
&\quad +\sum_{a}\sum_{d}\sum_{e<f}h^{\rm eff}_{ra}\langle rd||ef\rangle\Gamma^{ad}_{ef}+ \\
&\quad +\sum_{a}\sum_{b<c}\sum_{d}h^{\rm eff}_{rd}\langle ra||cb\rangle\Gamma^{cb}_{ad}+\sum_{a}\sum_{d}h^{\rm eff}_{ra}h^{\rm eff}_{rd}\Gamma^a_d. \qquad (8.52)
\end{aligned}
$$

Note, however, that the required term involving the three–body density matrix for the $N_i^{(+1)'}$ norm is of the form

$$
\begin{aligned}
N^{(0)'}(3) &= \sum_{a<b}\sum_{c}\sum_{d<e}\sum_{f}\langle ab||ic\rangle\langle de||if\rangle\langle\Psi^{(0)}_m|a^\dagger_c a^\dagger_d a^\dagger_e a_b a_a a_f|\Psi^{(0)}_m\rangle \qquad (8.53)\\
&= \sum_{\substack{\mu\nu\\ \mu'\nu'}}\sum_{\substack{a<b\\ c}}\sum_{\substack{d<e\\ f}}c_{\mu\nu}c_{\mu'\nu'}\langle ab||ic\rangle\langle de||if\rangle\langle L_\mu|a^\dagger_c a_b a_a|L_{\mu'}\rangle\langle R_\nu|a^\dagger_d a^\dagger_e a_f|R_{\nu'}\rangle+ \\
&\quad +\sum_{\substack{\mu\nu\\ \mu'\nu'}}\sum_{\substack{a<b\\ c}}\sum_{\substack{d<e\\ f}}c_{\mu\nu}c_{\mu'\nu'}\langle ab||ic\rangle\langle de||if\rangle\langle L_\mu|a^\dagger_d a_b a_a|L_{\mu'}\rangle\langle R_\nu|a^\dagger_c a^\dagger_e a_f|R_{\nu'}\rangle+ \\
&\quad +\sum_{\substack{\mu\nu\\ \mu'\nu'}}\sum_{\substack{a<b\\ c}}\sum_{\substack{d<e\\ f}}c_{\mu\nu}c_{\mu'\nu'}\langle ab||ic\rangle\langle de||if\rangle\langle L_\mu|a^\dagger_d a^\dagger_c a_b a_a|L_{\mu'}\rangle\langle R_\nu|a^\dagger_e a_f|R_{\nu'}\rangle+h.c.
\end{aligned}
$$

Consequently, for the first term of the above equation, one can prepare complementary operators such as

$$
\begin{aligned}
N^{(0)'}_{p1}(i) &= \sum_{\substack{a<b\\ c}}\langle ab||ic\rangle\langle L_\mu|a^\dagger_c a_b a_a|L_{\mu'}\rangle \qquad (8.54)\\
N^{(0)'}_{q1}(i) &= \sum_{\substack{d<e\\ f}}\langle de||if\rangle\langle R_\nu|a^\dagger_d a^\dagger_e a_f|R_{\nu'}\rangle, \qquad (8.55)
\end{aligned}
$$

and perform the needed multiplication by the wave function coefficients only once during the last converged sweep at the last step of the lattice. The cost of preparing a single operator of this form is proportional to $O(M^2k^3l)$, where M is the number of basis states in the DMRG calculation, k is the number of active orbitals, and l is the number of core orbitals. The cost of the multiplication is proportional to $O(M^3l)$ since one needs to perform the above listed multiplications for every core orbital. For the second term of Eq. (8.53), the situation becomes more complicated, and one has to prepare the following operators

$$N_{p2}^{(0)'}(i,c,d) = \sum_{a<b}\langle ab||ic\rangle\langle L_\mu|a_d^\dagger a_b a_a|L_{\mu'}\rangle \tag{8.56}$$

$$N_{q2}^{(0)'}(i,c,d) = \sum_{ef}\langle de||if\rangle\langle R_\nu|a_c^\dagger a_e^\dagger a_f|R_{\nu'}\rangle. \tag{8.57}$$

The cost of preparing such operators is proportional to $O(M^2k^4l)$, and is still comparable to the cost of the two–body density matrix evaluation. The storage requirement increases, and it is a bottleneck for such a procedure; however, one can assume that it is possible to process such operators in batches corresponding to the core label. The cost of the final multiplication, which has to be performed only once at the end of the chain, is $O(M^3k^2l)$. Note, however, that not all of the required elements of the three–body density matrix can be obtained in such a way, and one also has to consider another decomposition of operators between the Left and Right parts, namely

$$N_r^{(0)'}(i,d) = \sum_{\substack{a<b\\c}}\langle ab||ic\rangle\langle L_\mu|a_c^\dagger a_d^\dagger a_b a_a|L_{\mu'}\rangle \tag{8.58}$$

$$N_s^{(0)'}(i,d) = \sum_{ef}\langle de||if\rangle\langle R_\nu|a_e^\dagger a_f|R_{\nu'}\rangle. \tag{8.59}$$

Such operators are needed when there are four identical indices in the three–body density matrix or two pairs of identical indices. Since there are at least two identical labels in the $N_r^{(0)'}(i,d)$ operator, the cost of preparing such an operator does not increase, but the cost of storage does, as well as the cost of the multiplication which is now proportional to $O(M^3lk)$.

Similar reasoning can be applied to the calculation of the $N_r^{(-1)'}$ norm, and analogous complementary operators involving a virtual orbital r instead of core orbital can be prepared. The construction of the complementary operators and the subsequent multiplication by the zero order wave function does not exceed the cost of the construction of the three–body density matrix, which requires $O(M^2k^6) + O(M^3k^3)$ multiplications. The most expensive step that needs to be performed for the $N_r^{(-1)'}$ norm is proportional to $O(M^3k^2n)$, where n is the number of virtual orbitals. In cases in which $M < k^3$, the $O(M^2k^6)$ term dominates in the cost of the calculation of the three–body density matrix. In this case the DMRG decomposition trick is less expensive than the construction of the exact density matrix, providing that $Mn < k^4$ and $Ml < k^4$ where n is the number of virtual orbitals and l is the number of core orbitals. If $k^3 < M$, it is cheaper to construct the three–body density matrix instead of the auxiliary operators expressed above.

We have shown that in DMRG, due to the availability of the matrix representation of the second–quantized operators, one is able to perform the above explained factorization and obtain the terms that involve the three–body density matrix summed with integrals in a cheaper way than the cost of the exact construction of the 3–body density matrix prior to the summation with the integrals. However, the above construction worked only when certain assumptions regarding the relations between the number of states M and the number of orbitals were made. Since one cannot expect them to be true always, one can use the resolution of identity and approximate some the auxiliary operators, namely

$$\begin{aligned} N_{p2}^{(0)'}(i,c,d) &= \sum_{a<b}\langle ab||ic\rangle\langle L_\mu|a_d^\dagger a_b a_a|L_{\mu'}\rangle \qquad (8.60)\\ &\approx \sum_{\lambda}\sum_{a<b}\langle ab||ic\rangle\langle L_\mu|a_d^\dagger|L_\lambda\rangle\langle L_\lambda|a_b a_a|L_{\mu'}\rangle \\ &= \sum_{\lambda}\langle L_\mu|a_d^\dagger|L_\lambda\rangle N_{p2}^{(0)'}(i,c) \\ N_{q2}^{(0)'}(i,c,d) &= \sum_{ef}\langle de||if\rangle\langle R_\nu|a_c^\dagger a_e^\dagger a_f|R_{\nu'}\rangle \qquad (8.61)\\ &\approx \sum_{\kappa}\sum_{ef}\langle de||if\rangle\langle R_\nu|a_c^\dagger|R_\kappa\rangle\langle R_\kappa|a_e^\dagger a_f|R_{\nu'}\rangle \\ &= \sum_{\kappa}\langle R_\nu|a_c^\dagger|R_\kappa\rangle N_{q2}^{(0)'}(i,d). \end{aligned}$$

Consequently, preparing such operators is not more costly than preparation of the auxiliary operators used in the ordinary DMRG calculations. Of course one has to mention that for small M values such an approximation can give a very poor result. However, this is a controllable approximation since we know the error of the the truncation, and we can estimate whether it is reasonable to use such an approximation. Thus, using the resolution of identity allows us to approximate the numerators in a controlled way; additionally, it allows us to further reduce the cost in comparison to that for forming the auxiliary operators summed with integrals.

Let us show now the derivation of the last term of Eq. (8.42), which appears in the denominator, for some simple cases. For the $S^{(0)}_{ij,rs}$ subspace, we have $[H_v, V^{(0)}_{ij,rs}] = 0$ resulting in

$$E^{(k)}_l = E^{(0)}_m + \epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j \tag{8.62}$$

and consequently,

$$E^{(2)}_m(S^{(0)}_{ij,rs}) = \frac{N^{(0)}_{ij,rs}}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_j}. \tag{8.63}$$

The situation becomes more complicated if we start to consider the $S^{(+1)}_{ij,r}$ subspace. When calculating the commutator, we obtain

$$\begin{aligned}[H_v, V^{(+1)}_{ij,r}] \;=\; & -\sum_a \sum_{cd} \langle ra||ji\rangle h_{cd}\delta_{ad} a^\dagger_c a^\dagger_r a_i a_j + \\ & +\frac{1}{2}\sum_a \sum_{cdef} \langle ra||ji\rangle\langle cd|ef\rangle(\delta_{af} a^\dagger_c a^\dagger_d a_e a^\dagger_r a_i a_j - \delta_{ae} a^\dagger_c a^\dagger_d a_f a^\dagger_r a_i a_j).\end{aligned} \tag{8.64}$$

Subsequently, acting with $V^{(+1)\dagger}_{ij,r}$ leads to

$$\begin{aligned}&\langle\Psi^{(0)}_m|V^{(+1)\dagger}_{ij,r}[H^D, V^{(+1)}_{ij,r}]|\Psi^{(0)}_m\rangle = \sum_{acg} h_{cb}\langle ra||ji\rangle\langle rg||ji\rangle\langle\Psi^{(0)}_m|\delta_{gc} - a^\dagger_c a_g|\Psi^{(0)}_m\rangle + \\ &+\frac{1}{2}\sum_a\sum_{cde}\sum_g \langle cd|ea\rangle\langle ra||ij\rangle\langle rg||ji\rangle\langle\Psi^{(0)}_m|\delta_{gc}a^\dagger_d a_e - \delta_{gd}a^\dagger_c a_e + a^\dagger_c a^\dagger_d a_g a_e|\Psi^{(0)}_m\rangle + \\ &+\frac{1}{2}\sum_a\sum_{cdf}\sum_g \langle cd|af\rangle\langle ra||ij\rangle\langle rg||ji\rangle\langle\Psi^{(0)}_m|\delta_{gc}a^\dagger_d a_f - \delta_{gd}a^\dagger_c a_f + a^\dagger_c a^\dagger_d a_g a_f|\Psi^{(0)}_m\rangle,\end{aligned} \tag{8.65}$$

$$E_m^{(2)}(S_{ij,r}^{(+1)}) = \frac{N_{ij,r}^{(+1)}}{\epsilon_{ij,r}^{(+1)} + \epsilon_r - \epsilon_i - \epsilon_j}, \tag{8.66}$$

where

$$\epsilon_{ij,r}^{(+1)} = \frac{1}{N_{ij,r}^{(+1)}} \langle \Psi_m^{(0)} | V_{ij,r}^{(+1)\dagger} [H^D, V_{ij,r}^{(+1)}] | \Psi_m^{(0)} \rangle. \tag{8.67}$$

Similarly, one can develop all of the commutators appearing for the $S_{rs,i}^{(-1)}$ subspace and obtain the expressions for the energy,

$$\begin{aligned} [H_v, V_{rs,i}^{(-1)}] &= -\sum_a \sum_{cd} \langle rs||ai\rangle h_{cd} \delta_{ac} a_d a_r^\dagger a_s^\dagger a_i + \\ &+ \frac{1}{2} \sum_a \sum_{cdef} \langle rs||ai\rangle \langle cd|ef\rangle (\delta_{ac} a_d^\dagger a_f a_e a_r^\dagger a_s^\dagger a_j - \delta_{ad} a_c^\dagger a_f a_e a_r^\dagger a_s^\dagger a_i), \end{aligned} \tag{8.68}$$

$$\begin{aligned} &\langle \Psi_m^{(0)} | V_{rs,i}^{(-1)\dagger} [H_v, V_{rs,i}^{(-1)}] | \Psi_m^{(0)} \rangle = -\sum_{adg} h_{ab} \langle rs||ai\rangle \langle rs||gi\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle + \\ &-\frac{1}{2} \sum_a \sum_{def} \sum_g \langle ad|ef\rangle \langle rs||ai\rangle \langle rs||gi\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d^\dagger a_f a_e | \Psi_m^{(0)} \rangle + \\ &\frac{1}{2} \sum_a \sum_{cef} \sum_g \langle ca|ef\rangle \langle rs||ai\rangle \langle rs||gi\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_c^\dagger a_f a_e | \Psi_m^{(0)} \rangle, \end{aligned} \tag{8.69}$$

$$E_m^{(2)}(S_{rs,i}^{(-1)}) = \frac{N_{rs,i}^{(-1)}}{\epsilon_r + \epsilon_s - \epsilon_i - \epsilon_{rs,i}^{(+1)}}. \tag{8.70}$$

To calculate the second order correction to the energy for the $S_{ij,rs}^{(0)}$, $S_{ij,r}^{(+1)}$, and $S_{rs,i}^{(-1)}$ subspaces, one requires only one– and two–body density matrices that have the active space indices. These density matrices may be calculated in the earlier DMRG–SCF calculations. However, to calculate the corrections coming from the rest of the subspaces, the three–body density matrix is required. One can try to employ, as for the numerators, the trick with the factorization of the sums that appear in the terms involving the three–body density matrix and integrals. However, such a trick is not effective here, as will be illustrated for the $S_{ij}^{(+2)}$ subspace, and some other way of calculating the terms involving the three–body density matrix has to be found. For the $S_{ij}^{(+2)}$ subspace the

commutator developed has the form

$$
\begin{aligned}
[H_v, V_{ij}^{(+2)}] &= \sum_{a<b}\sum_{cd}\langle ab||ij\rangle h_{cd}(\delta_{ad}a_c^\dagger a_b^\dagger a_i a_j + \delta_{bd}a_c^\dagger a_a^\dagger a_i a_j) + \\
&+\frac{1}{2}\sum_{a<b}\sum_{cdef}\langle ab||ij\rangle\langle cd|ef\rangle((\delta_{ae}\delta_{bf} - \delta_{af}\delta_{be})a_c^\dagger a_d^\dagger a_i a_j - \delta_{ae}a_c^\dagger a_d^\dagger a_b^\dagger a_f a_i a_j + \\
&+\delta_{af}a_c^\dagger a_d^\dagger a_b^\dagger a_e a_i a_j + \delta_{eb}a_c^\dagger a_d^\dagger a_a^\dagger a_f a_i a_j - \delta_{bf}a_c^\dagger a_d^\dagger a_a^\dagger a_e a_i a_j),
\end{aligned}
\tag{8.71}
$$

and after substituting it into Eq. (8.42) we obtain

$$
\begin{aligned}
&\langle\Psi_m^{(0)}|V_{ij}^{(+2)\dagger}[H_v, V_{ij}^{(+2)}]|\Psi_m^{(0)}\rangle = \langle\Psi_m^{(0)}|\sum_{a<b}\sum_{c}\sum_{a'<b'}\langle ab||ij\rangle\langle a'b'||ij\rangle h_{ca} \times \\
&(\delta_{a'c}(\delta_{bb'} - a_b^\dagger a_{b'}) + \delta_{b'c}(\delta_{a'b} - a_b^\dagger a_{a'}) + \delta_{a'b}a_c^\dagger a_{b'} - \delta_{bb'}a_c^\dagger a_{a'} + a_c^\dagger a_b^\dagger a_{b'}a_{a'}) + \\
&+\sum_{a<b}\sum_{c}\sum_{a'<b'}\langle ab||ij\rangle\langle a'b'||ij\rangle h_{cb}(\delta_{a'c}(\delta_{ab'} - a_a^\dagger a_{b'}) + \delta_{b'c}(\delta_{a'a} - a_a^\dagger a_{a'}) + \\
&+\delta_{a'a}a_c^\dagger a_{b'} - \delta_{ab'}a_c^\dagger a_{a'} + a_c^\dagger a_a^\dagger a_{b'}a_{a'}) + \\
&+\frac{1}{2}\sum_{a<b}\sum_{cd}\sum_{a'<b'}\langle ab||ij\rangle\langle a'b'||ij\rangle(\langle cd|ab\rangle - \langle cd|ba\rangle)(\delta_{a'c}(\delta_{db'} - a_d^\dagger a_{b'}) + \\
&+\delta_{b'c}(\delta_{a'd} - a_d^\dagger a_{a'}) + \delta_{a'd}a_c^\dagger a_{b'} - \delta_{db'}a_c^\dagger a_{a'} + a_c^\dagger a_d^\dagger a_{b'}a_{a'}) + \\
&+\frac{1}{2}\sum_{a<b}\sum_{cdef}\sum_{a'<b'}\langle ab||ij\rangle\langle a'b'||ij\rangle\langle cd|ef\rangle(-\delta_{ae}a_{b'}a_{a'}a_c^\dagger a_d^\dagger a_b^\dagger a_f + \delta_{af}a_{b'}a_{a'}a_c^\dagger a_d^\dagger a_b^\dagger a_e + \\
&+\delta_{eb}a_{b'}a_{a'}a_c^\dagger a_d^\dagger a_a^\dagger a_f - \delta_{bf}a_{b'}a_{a'}a_c^\dagger a_d^\dagger a_a^\dagger a_e)|\Psi_m^{(0)}\rangle.
\end{aligned}
\tag{8.72}
$$

It is clear that the last term will require the three–body density matrix. Since the order of the creation and annihilation operators is reversed, one needs to bring it into the proper order obtaining

$$
\begin{aligned}
a_{b'}a_{a'}a_c^\dagger a_d^\dagger a_b^\dagger a_f &= (\delta_{a'c}\delta_{db'} + \delta_{b'c}\delta_{a'd})a_b^\dagger a_f - (\delta_{a'c}\delta_{bb'} + \delta_{b'c}\delta_{a'b})a_d^\dagger a_f + \\
&+(\delta_{a'd}\delta_{bb'} - \delta_{b'd}\delta_{a'b})a_c^\dagger a_f + \delta_{a'c}a_d^\dagger a_b^\dagger a_{b'}a_f + \delta_{b'c}a_d^\dagger a_b^\dagger a_{a'}a_f + \\
&-\delta_{a'd}a_c^\dagger a_b^\dagger a_{b'}a_f + \delta_{b'd}a_c^\dagger a_b^\dagger a_{a'}a_f + \delta_{a'b}a_c^\dagger a_d^\dagger a_{b'}a_f + \\
&-\delta_{bb'}a_c^\dagger a_d^\dagger a_{a'}a_f + a_c^\dagger a_d^\dagger a_b^\dagger a_{b'}a_{a'}a_f.
\end{aligned}
\tag{8.73}
$$

Consequently, the most problematic part is the evaluation of terms of the following type

$$\begin{aligned} C_{ij}^{(+2)}(3) &= \sum_{a<b}\sum_{cdef}\sum_{a'<b'} \langle ab||ij\rangle\langle a'b'||ij\rangle\langle cd|ef\rangle\delta_{ae}\langle\Psi_m^{(0)}|a_c^\dagger a_d^\dagger a_b^\dagger a_{b'}a_{a'}a_f|\Psi_m^{(0)}\rangle \qquad (8.74)\\ &= \sum_{a<b}\sum_{cdf}\sum_{a'<b'} \langle ab||ij\rangle\langle a'b'||ij\rangle\langle cd|af\rangle\langle\Psi_m^{(0)}|a_c^\dagger a_d^\dagger a_b^\dagger a_{b'}a_{a'}a_f|\Psi_m^{(0)}\rangle \end{aligned}$$

As for the previous occurrence of the three–body density matrix, one can prepare a suitable decomposition, namely

$$C_p^{(+2)}(a) = \sum_{cdf}\langle cd|af\rangle\langle L_\mu|a_c^\dagger a_d^\dagger a_f|L_{\mu'}\rangle \qquad (8.75)$$

$$C_q^{(+2)}(b,ij) = \sum_{a'<b'}\langle a'b'|ij\rangle\langle R_\nu|a_b^\dagger a_{b'}a_{a'}|R_{\nu'}\rangle. \qquad (8.76)$$

Consequently, the final multiplication takes a form

$$C_{ij}^{(+2)}(3) = \sum_{\substack{\mu\nu\\ \mu'\nu'}}\sum_a\sum_b c_{\mu\nu}c_{\mu'\nu'}C_q^{(+2)}(b,ij)\langle ab||ij\rangle C_p^{(+2)}(a) + h.c. \qquad (8.77)$$

The cost of preparing the complementary operators is $O(M^2k^4)$ and $O(M^2k^3l^2)$, where k denotes the number of active orbitals, and l is the number of core orbitals. The second complementary operator is more expensive to prepare than the standard operators that are kept in the DMRG calculation, and that cost is greater than the cost of the DMRG run in which the two–body density matrix is prepared. The storage required for the second complementary operator per step is proportional to $O(M^2kl^2)$, which is much greater than the original DMRG storage requirement. Finally, the cost of the multiplication is $O(M^3k^2l^2)$ which is also greater than the standard DMRG cost of $O(M^3k^3)$. Moreover, not all needed elements of the three–body density matrix can be obtained in such a way, and when there are four identical indices in the three–body density matrix or two pairs of identical indices, one also has to consider another decomposition of operators

between the Left and Right parts,

$$C_r^{(+2)}(a,a') = \sum_{cdf}\langle cd|af\rangle\langle L_\mu|a_c^\dagger a_d^\dagger a_f a_{a'}|L_{\mu'}\rangle \tag{8.78}$$

$$C_s^{(+2)}(ba',ij) = \sum_{b'}\langle a'b'|ij\rangle\langle R_\nu|a_b^\dagger a_{b'} a_{a'}|R_{\nu'}\rangle. \tag{8.79}$$

Even though assembling such operators does not cost more than in the previous case since there is one repeating index for the $C_r^{(+2)}(a,a')$ operator, the storage is severely increased. The cost of the multiplication by the vector is proportional to $O(M^3k^3l^2)$. Finally, the expression for the second order correction to to energy due to the inclusion of the $S_{ij}^{(+2)}$ subspace is

$$E_m^{(2)}(S_{ij}^{(+2)}) = \frac{N_{ij}^{(+2)}}{\epsilon_{ij}^{(+2)} - \epsilon_i - \epsilon_j}. \tag{8.80}$$

As in the previous case, very similar expressions are produced when one considers the commutator that has to be considered for the $S_{rs}^{(-2)}$ subspace. Detailed derivations for all the remaining subspaces are given in Appendix C. For the $S_{rs}^{(-2)}$ subspace, an analogous decomposition with the core orbitals replaced by the virtual orbitals has to be created for the terms involving the three–body density matrix. The total cost of the second type of multiplication is hence $O(M^3k^3n^2)$, which is probably comparable to (or even exceeds) the cost of assembling the three–body density matrix. Consequently, some other solution has to be found to take into account the terms involving the three–body density matrix for the $S_{ij}^{(+2)}$ and $S_{rs}^{(-2)}$ spaces.

The the $S_{i,r}^{(0)'}$ subspace also has contributions with the three–body density matrices. As for the $S_{ij}^{(+2)}$ and $S_{rs}^{(-2)}$ spaces, it is shown in Appendix C that the decomposition with the complementary operators is too expensive.

The remaining two subspaces $S_r^{(-1)'}$ and $S_j^{(+1)'}$ yield the most complicated expressions that involve the four–body density matrix. Although a suitable decompositions can be prepared, as discussed in Appendix C, the overall cost of preparing the terms summed with integrals that contain the four–body density matrix is probably roughly similar to that for the exact production of the three–body density matrix that we tried to avoid. Thus, we have to find some other way

of approximating the four–body density matrix that will not increase significantly the operation count that is required for the DMRG run with the two–body density matrix production.

From the previous considerations, it can be concluded that the calculations of the numerators in the expressions for the second order energy can be done quite easily and the needed elements can be prepared in DMRG without too much effort. However, the calculation of the energies appearing in the denominators presents a considerable challenge, due to the elements containing the three– and four–body density matrix that cannot be factorized in an advantageous manner. Although such a calculation with an explicit preparation of the elements that contain three– and four–body density matrices summed up with the integrals should still be cheaper than preparing the four–body density matrix itself, the overall cost of such procedure may be very high.

8.2 DMRG-DFT

The decay of the density matrix eigenvalues is very fast in one–dimensional models with a gap [71, 104], but it is much slower, most likely not exponential, for the two–body Hamiltonians that are considered in quantum chemistry. For molecules, the density matrix spectrum can have many small eigenvalues that cannot be neglected if one wants to achieve quantitative accuracy. Consequently, in the traditional one–dimensional DMRG, as used in quantum chemistry, to include the eigenvectors that correspond to the small eigenvalues in the basis, one needs to deal with huge values of M, thus making the whole procedure very expensive. Hence, it is crucial to be able to perform DMRG calculations with small M and to include the dynamic correlation using some other technique. The perturbation theory applied to the zero order DMRG–SCF wave function is one example of such a treatment.

In this section, we describe an unusual example in which the density functional theory (DFT) is used to recover the contribution coming from dynamic correlation. Frequently the dynamic correlation can be calculated with a moderate cost with DFT. However, in general, DFT faces serious problems when trying to treat non–dynamic correlation. On the other hand, the non–dynamic correlation can be very well accounted for by a method such as CASSCF (or DMRG-SCF

in our case). It may therefore be useful to try to develop a hybrid model that has the ability of the wave function method to treat non–dynamic correlation together with the efficiency of DFT in handling dynamic correlation. During previous years, some approaches (see for example Ref. [144, 145]) in this spirit were attempted; however, it was highly nontrivial (and rather unsuccessful) to adapt the functional in such way to avoid double counting of the dynamical correlation. The double counting problem is present since in these formulations the dynamic correlation in the active space was taken into account by the wave function method and by DFT, thus counting this effect twice. This problem can be eliminated if one uses the two–electron integral split into long– and short—range parts, as proposed by Savin [54, 55, 56]. The long–range part of the interaction is treated by the wave function method, while the short–part is treated by DFT.

Let us sketch briefly the main principles of such a hybrid method. The Coulomb interaction W_{ee} is separated into long– and short–range parts

$$W_{ee} = W_{ee}^{\mathrm{lr},\mu} + W_{ee}^{\mathrm{sr},\mu}, \tag{8.81}$$

with the long–range part given by

$$W_{ee}^{\mathrm{lr},\mu} = \frac{1}{2} \sum_{i \neq j} \frac{\mathrm{erf}(\mu r_{ij})}{r_{ij}}, \tag{8.82}$$

where erf is the standard error function, and μ is is a free parameter that can be varied between $[0, +\infty)$ that describes the long– and short–range separation. Consequently, the universal functional $F[n]$ can be written in terms of the long– and short–range contributions

$$F[n] = F^{\mathrm{lr},\mu}[n] + E_H^{\mathrm{sr},\mu}[n] + E_{xc}^{\mathrm{sr},\mu}[n], \tag{8.83}$$

$$F^{\mathrm{lr},\mu}[n] = \min_{\Psi \to n} \{ \langle \Psi | \hat{T} + \hat{W}_{ee}^{\mathrm{lr},\mu} | \Psi \rangle \}, \tag{8.84}$$

where $E_H^{\mathrm{sr},\mu}[n]$ and $E_{xc}^{\mathrm{sr},\mu}[n]$ mean the Hartree and the exchange and correlation energies connected to the short range–interaction. In a traditional DFT approach, the density is obtained from a fictitious non–interactive system. In this hybrid method, the density is obtained from a fictitious

long–range interacting system, while the Kohn–Sham DFT is recovered in the $\mu = 0$ limit, and the pure wave function method can be obtained if $\mu = +\infty$. Consequently, the ground state energy can be produced as

$$E_0 = \min_n \left\{ F[n] + \int d\mathbf{r} v_{ne}(\mathbf{r}) n(\mathbf{r}) \right\}, \tag{8.85}$$

and after plugging the expression for $F[n]$, one obtains

$$E_0 = \min_\Psi \left\{ \langle \Psi | \hat{T} + \hat{W}_{ee}^{\mathrm{lr},\mu} | \Psi \rangle + E_H^{\mathrm{sr},\mu}[n_\Psi] + E_{xc}^{\mathrm{sr},\mu}[n_\Psi] + \int d\mathbf{r} v_{ne}(\mathbf{r}) n(\mathbf{r}) \right\}, \tag{8.86}$$

where n_Ψ means the electron density coming from the wave function Ψ and $v_{ne}(\mathbf{r})$ is the local nuclear attraction potential. To minimize the wave function $\Phi^{\mathrm{lr},\mu}$ coming from the above equation, one needs to solve the following self–consistent equation

$$(\hat{T} + \hat{W}_{ee}^{\mathrm{lr},\mu} + \hat{V}^{\mathrm{sr},\mu}) | \Phi^{\mathrm{lr},\mu} \rangle = \epsilon^\mu | \Phi^{\mathrm{lr},\mu} \rangle, \tag{8.87}$$

where ϵ^μ is a Lagrange multiplier and $\hat{V}^{\mathrm{sr},\mu}$ is the effective local potential operator

$$\hat{V}^{\mathrm{sr},\mu} = \int d\mathbf{r} v^{\mathrm{sr},\mu}(\mathbf{r}) \hat{n}(\mathbf{r}), \tag{8.88}$$

where $\hat{n}(\mathbf{r})$ denotes the density operator and $v^{\mathrm{sr},\mu}$ is defined by

$$v^{\mathrm{sr},\mu}(\mathbf{r}) = v_{ne}(\mathbf{r}) + \frac{\delta E_H^{\mathrm{sr},\mu}}{\delta n}(\mathbf{r}) + \frac{\delta E_{xc}^{\mathrm{sr},\mu}}{\delta n}(\mathbf{r}). \tag{8.89}$$

From the above equation, it is clear that for $\mu = 0$ the long–range interaction $\hat{W}_{ee}^{\mathrm{lr},\mu}$ vanishes and the the Kohn–Sham equations are recovered. Conversely, for $\mu \to +\infty$ the short–range part of the interaction $\hat{W}_{ee}^{\mathrm{sr},\mu}$ vanishes, and the the usual Schrödinger equation is recovered. For all values of the μ parameter between 0 and $+\infty$, the method treats the long–range effect within the wave function approach, while the short–range effects are treated by DFT. It is of course necessary to note that the exact form of the short–range exchange and correlational functional $E_{xc}^{\mathrm{sr},\mu}$ is unknown, and in practice one uses an approximate form of the functional. The two

approximate short–range functionals that have been considered before are the short–range local–density approximation [56, 146] (SRLDA) and the short–range generalized gradient approximation (GGA) denoted PBEHSERI, as an extension of the PBE functional [147].

One can imagine that in such a scheme the wave function can be chosen to be DMRG. The DMRG calculation will need to be performed with the modified long–range integrals. Next, the resulting one–body density matrix can be taken, and based on it the density required for the DFT calculation can be obtained, consequently enabling one to solve Eq. (8.87).

Of course, such an alternative would involve performing DMRG calculations in the full space, and it would be effective for orbital spaces of rather small sizes. However, it is interesting to ask how the modification of the Hamiltonian influences the decay of the DMRG density matrix spectrum? If the modified integrals caused a faster decay of the spectrum, one could expect that the needed M value will become smaller, and one would be able to treat larger spaces. However, Eq. (8.87) can use any *ab initio* wave function as $|\Phi^{\mathrm{lr},\mu}\rangle$; thus, a similar treatment may be used for the DMRG CASPT2 (or NEVPT2) wave function. Such an option will allow one to treat much larger orbital spaces since the DMRG calculation is performed only in the active space.

Let us briefly mention the advantages of using such a hybrid method with DMRG. It could help to account for the dynamical correlation in the DMRG method, but it would not require one to produce the two–body density matrix since the density n_Ψ can be obtained using only the one–body density matrix, thus eliminating from the DMRG algorithm the $O(M^2k^4) + O(M^3k^2)$ step responsible for the construction of the two–body density matrix. Additionally, it could improve the efficiency of the DMRG calculations since the modified Hamiltonian with the long–range integrals should resemble a system with a faster–decaying density matrix spectrum. The size of the needed orbital basis set should be also smaller since DFT is proven to converge much better within a series of basis sets.

Nevertheless, there are also couple of disadvantages. The most obvious one is the presence of the empirical parameter μ that may be system dependent, and all that one can hope to achieve is to find a value of this parameter that will be appropriate for many classes of systems. Another disadvantage is the empirical character of the exchange and correlational functional $E_{xc}^{\mathrm{sr},\mu}$; however,

here one can expect that it can have a simple and well understood form since it only covers the effects of the dynamic correlation. Note too that there is no systematic way that allows one to improve the DFT part of the calculation.

The long chains of transition metal oxides can be targeted as a potential application of the DMRG-DFT hybrid method since the number of orbitals outside the active space may be too large to be handled by method such as CASPT2 itself. If a CASPT2 wave function with the DMRG treatment in the active space was used as the *ab initio* wave function in Eq. (8.87), then the size of the virtual space should be much smaller due to the faster convergence with the basis set. One can imagine that DMRG would be able to treat a chain of atoms with strong non–dynamic correlation contribution if the interactions between them will be restricted to a few neighbors. DFT can be very efficient in treating dynamic correlation effects coming from the possible ligand orbitals that will not belong to the active space. Such a treatment can be generalized later to include the embedding effects coming from the inclusion of the fragment under study into the whole crystal structure.

Appendix A

S^2 eigenstates in a product basis

Here we present an example that helps explain the terminology used and discusses the spin eigenstates in a product basis. Consider a system having five orbitals and four electrons (two alpha and two beta electrons). We assume for simplicity that all the orbitals are of A_1 symmetry and that two orbitals are present in the Left block and three orbitals are in the Right block. We have the following distributions of electrons between the Left and Right blocks.

Left	Right
$2\alpha 2\beta$	0
$1\alpha 2\beta$	1α
$2\alpha 1\beta$	1β
2β	2α
$1\alpha 1\beta$	$1\alpha 1\beta$
2α	2β
1α	$1\alpha 2\beta$
1β	$2\alpha 1\beta$
0	$2\alpha 2\beta$

(Note that the numbers in this example refer to the number of α or β electrons in the Left or Right part and not the the number of the orbital that is occupied. Hence $2\alpha 2\beta$ means that two α and two β electrons are present in the Left (or Right) block.

We consider only spin multiplets in the Right block since the Left block will be analogous. The first type of multiplet is generated by high spin determinants that consist of some doubly occupied orbitals and the remaining singly occupied α-spin orbitals. Partitions 0, 1α, 2α consist exclusively of such determinants with no doubly occupied orbitals and 0, 1, and 2 single α electrons, respectively. Likewise, the determinants in the $1\alpha 1\beta$, $2\alpha 1\beta$, and $2\alpha 2\beta$ partitions, all for maximal number of paired electrons, are the single determinant S_R^2 eigenstates. Each of these high–spin determinants generates a complete multiplet, and the laddering down will define states in the 1β, $1\alpha 1\beta$, 2β and $1\alpha 2\beta$ partitions. *Complete multiplets* generated from the high–spin determinants are the simplest type. These multiplets are typically used in the initialization procedure. If one considers the $2\alpha 1\beta$ partition, one can recognize that besides the $|\uparrow\downarrow \quad \uparrow \quad 0\rangle$ determinant, there are some determinants (e.g. $|\uparrow \quad \downarrow \quad \uparrow\rangle$) with unpaired β electrons. These determinants with unpaired β electrons are generated from a $|\uparrow \quad \uparrow \quad \uparrow\rangle$ determinant. However, such a partition 3α is not present in the Right block since it does not have a partner in the Left block. Thus, 3α is a superfluous partition. It can be used to generate lower S^z determinants in the $2\alpha 1\beta$ partition. The multiplet (quartet) obtained from action of S_R^- on 3α is *incomplete* since the 3α and 3β components are not present. Such incomplete multiplets are obtained from a single determinant in a superfluous partition by an action of S_R^-. Finally, there are two other possible linear combinations of ($|\uparrow \quad \uparrow \quad \downarrow\rangle$, $|\uparrow \quad \downarrow \quad \uparrow\rangle$, $|\downarrow \quad \uparrow \quad \uparrow\rangle$) that cannot be obtained by action of S_R^- on 3α. These combinations correspond to a doublet. Such a multiplet is complete, as it is contained in both $2\alpha 1\beta$ and $1\alpha 1\beta$ partition, but it is not generated by a single determinant.

Appendix B

The initialization procedure

We use the system described in the previous appendix to present an explicit example illustrating the construction of the determinantal space spanning the S^2_R manifold in the right block. For larger systems, for efficiency reasons the determinants coming from superfluous partitions should be omitted.

For every symmetry block able to couple to the Left block

1. Find all possible ($N_\alpha \neq 0$, $N_\beta = 0$) partitions. For every such partition, construct one lowest energy determinant. Find ($N_\alpha \neq 0$, $N_\beta \neq 0$, where $N_\alpha \geq N_\beta$) partitions that can have the highest S^z components of a multiplet. For every such partition, construct one lowest energy determinant with maximal number of paired electrons e.g., using the previous example:

$$
\begin{aligned}
0 \qquad & |0 \quad 0 \quad 0\rangle \\
1\alpha \qquad & |\uparrow \quad 0 \quad 0\rangle \\
2\alpha \qquad & |\uparrow \quad \uparrow \quad 0\rangle \\
1\alpha 1\beta \qquad & |\uparrow\downarrow \quad 0 \quad 0\rangle \\
2\alpha 1\beta \qquad & |\uparrow\downarrow \quad \uparrow \quad 0\rangle \\
2\alpha 2\beta \qquad & |\uparrow\downarrow \quad \uparrow\downarrow \quad 0\rangle.
\end{aligned}
$$

2. Find superfluous partitions that can have the highest S^z components of a multiplet. For every such partition, construct one lowest energy determinant.

   The superfluous partitions

   $$
   \begin{aligned}
   3\alpha \quad (\text{from} \quad 2\alpha 1\beta) \qquad & |\uparrow \quad \uparrow \quad \uparrow\rangle \\
   3\alpha 1\beta \quad (\text{from} \quad 2\alpha 2\beta) \qquad & |\uparrow\downarrow \quad \uparrow \quad \uparrow\rangle
   \end{aligned}
   $$

3. Flip the single spins in every previously obtained determinant until there are no more single alpha electrons to flip. (We will explicitly list only the 1α partition and the 3α superfluous partition since the remaining partitions follow analogous rules.)

   $$1\alpha \; |\uparrow \quad 0 \quad 0\rangle$$

   obtaining the 1β partition

   $$|\downarrow \quad 0 \quad 0\rangle$$

   From the 3α superfluous partition

   $$3\alpha \; |\uparrow \quad \uparrow \quad \uparrow\rangle$$

   obtaining the existing $2\alpha 1\beta$ partition

$$|\uparrow\ \uparrow\ \downarrow\rangle, |\uparrow\ \downarrow\ \uparrow\rangle, |\downarrow\ \uparrow\ \uparrow\rangle$$

obtaining the existing $1\alpha 2\beta$ partition

$$|\downarrow\ \downarrow\ \uparrow\rangle, |\downarrow\ \uparrow\ \downarrow\rangle, |\uparrow\ \downarrow\ \downarrow\rangle$$

obtaining the superfluous 3β partition

$$|\downarrow\ \downarrow\ \downarrow\rangle$$

4. In the obtained determinantal basis, construct and diagonalize the S_r^2 matrix in every partition except the superfluous ones. (The superfluous partitions are only used to produce a proper S_r^2 manifold.)

5. Construct the S_r^- matrix in every partition except the superfluous ones.

6. Transform the S_r^- matrix to the S_r^2 eigenbasis.

7. Act iteratively with $\hat{S}_r^-$ starting from the single obtained $\hat{S}_r^2$ eigenvector that is the highest S^z components of a multiplet from the existing partitions and from the partitions leading to the superfluous ones.

8. Save the basis vectors obtained.

9. Calculate on-the-fly the matrix representation of the operators in the right block in the previously obtained determinantal basis, and transform such obtained operators to the S_r^2 adapted basis using the transformation matrices obtained at the previous point.

Go to point 1 for the next symmetry block

Appendix C

Auxiliary formulas for NEVPT approach

The calculation of the norms for all the eight subspaces is summarized in Eqs. (C.1) to (C.10).

$$N_l^{(k)} = \langle \Psi_l^{(k)} | \Psi_l^{(k)} \rangle = \langle \Psi_m^{(0)} | V_l^{(k)\dagger} V_l^{(k)} | \Psi_m^{(0)} \rangle \tag{C.1}$$

$$N_{ij,rs}^{(0)} = \langle rs||ji\rangle^2 \tag{C.2}$$

$$\begin{aligned} N_{ij,r}^{(+1)} &= \sum_a \sum_b \langle ra||ji\rangle\langle rb||ji\rangle\langle \Psi_m^{(0)} | a_a a_b^\dagger | \Psi_m^{(0)} \rangle \\ &= \sum_a \sum_b \langle ra||ji\rangle\langle rb||ji\rangle\langle \Psi_m^{(0)} | \delta_{ab} - a_b^\dagger a_a | \Psi_m^{(0)} \rangle \\ &= \langle ra||ji\rangle^2 - \sum_a \sum_b \langle ra||ji\rangle\langle rb||ji\rangle \Gamma_a^b \end{aligned} \tag{C.3}$$

$$\begin{aligned} N^{(-1)}_{i,rs} &= \sum_a \sum_b \langle rs||ai\rangle\langle rs||bi\rangle\langle\Psi^{(0)}_m|a^\dagger_a a_b|\Psi^{(0)}_m\rangle \\ &= \sum_a \sum_b \langle rs||ai\rangle\langle rs||bi\rangle\Gamma^a_b \end{aligned} \tag{C.4}$$

$$\begin{aligned} N^{(+2)}_{ij} &= \sum_{a<b}\sum_{c<d}\langle ab||ji\rangle\langle cd||ji\rangle\langle\Psi^{(0)}_m|a_b a_a a^\dagger_c a^\dagger_d|\Psi^{(0)}_m\rangle \\ &= \sum_{a<b}\sum_{c<d}\langle ab||ji\rangle\langle cd||ji\rangle\langle\Psi^{(0)}_m|\delta_{ac}\delta_{bd} - \delta_{cb}\delta_{ad} - \delta_{ac}a^\dagger_d a_b + \\ &\quad +\delta_{cb}a^\dagger_d a_a + \delta_{da}a^\dagger_c a_b - \delta_{bd}a^\dagger_c a_a + a^\dagger_c a^\dagger_d a_b a_a|\Psi^{(0)}_m\rangle \\ &= \langle ab||ji\rangle^2 - \sum_{a<b}\sum_{a<d}\langle ab||ji\rangle\langle ad||ji\rangle\Gamma^d_b + \sum_{a<b}\sum_{b<d}\langle ab||ji\rangle\langle bd||ji\rangle\Gamma^d_a + \\ &\quad +\sum_{a<b}\sum_{c<a}\langle ab||ji\rangle\langle ca||ji\rangle\Gamma^c_b - \sum_{a<b}\sum_{c<b}\langle ab||ji\rangle\langle cb||ji\rangle\Gamma^c_a + \sum_{a<b}\sum_{c<d}\langle ab||ji\rangle\langle cd||ji\rangle\Gamma^{cd}_{ba} \end{aligned} \tag{C.5}$$

$$\begin{aligned} N^{(-2)}_{rs} &= \sum_{a<b}\sum_{c<d}\langle rs||ba\rangle\langle rs||dc\rangle\langle\Psi^{(0)}_m|a^\dagger_b a^\dagger_a a_c a_d|\Psi^{(0)}_m\rangle \\ &= \sum_{a<b}\sum_{c<d}\langle rs||ba\rangle\langle rs||dc\rangle\Gamma^{ba}_{cd} \end{aligned} \tag{C.6}$$

$$
\begin{aligned}
N_{i,r}^{(0)'} &= \langle\Phi_c\Psi_m^{(0)}|(\sum_a\sum_b\langle ra||ib\rangle a_i^\dagger a_b^\dagger a_a a_r + \sum_j\langle rj||ij\rangle a_i^\dagger a_j^\dagger a_j a_r + \langle r|h|i\rangle a_i^\dagger a_r) \times \qquad \text{(C.7)}\\
&\quad (\sum_c\sum_d\langle rc||id\rangle a_r^\dagger a_c^\dagger a_d a_i + \sum_k\langle rk||ik\rangle a_r^\dagger a_j^\dagger a_j a_i + \langle r|h|i\rangle a_r^\dagger a_i)|\Psi_m^{(0)}\Phi_c\rangle \times \\
&= \sum_{ab}\sum_{cd}\langle ra||ib\rangle\langle rc||id\rangle\langle\Psi_m^{(0)}|a_b^\dagger a_a a_c^\dagger a_d|\Psi_m^{(0)}\rangle + \sum_{cd}\sum_{j\neq i}\langle rj||ij\rangle\langle rc||id\rangle\langle\Psi_m^{(0)}|a_c^\dagger a_d|\Psi_m^{(0)}\rangle + \\
&\quad + \sum_{cd}\langle r|h|i\rangle\langle rc||id\rangle\langle\Psi_m^{(0)}|a_c^\dagger a_d|\Psi_m^{(0)}\rangle + \sum_{ab}\sum_{k\neq i}\langle ra||ib\rangle\langle rk||ik\rangle\langle\Psi_m^{(0)}|a_b^\dagger a_a|\Psi_m^{(0)}\rangle + \\
&\quad + \sum_{j\neq i}\sum_{k\neq i}\langle rj||ij\rangle\langle rk||ik\rangle + \sum_{k\neq i}\langle r|h|i\rangle\langle rk||ik\rangle + \sum_{ab}\langle ra||ib\rangle\langle r|h|i\rangle\langle\Psi_m^{(0)}|a_b^\dagger a_a|\Psi_m^{(0)}\rangle + \\
&\quad + \sum_{j\neq i}\langle rj||ij\rangle\langle r|h|i\rangle + \langle r|h|i\rangle^2 \\
&= \sum_{ab}\sum_{cd}\langle ra||ib\rangle\langle rc||id\rangle\langle\Psi_m^{(0)}|\delta_{ac}a_b^\dagger a_d - a_b^\dagger a_c^\dagger a_a a_d|\Psi_m^{(0)}\rangle + \\
&\quad + 2\sum_{cd}\langle rc||id\rangle\langle\Psi_m^{(0)}|a_c^\dagger a_d|\Psi_m^{(0)}\rangle\left(\sum_{j\neq i}\langle rj||ij\rangle + \langle r|h|i\rangle\right) + \\
&\quad + \sum_{k\neq i}\langle rk||ik\rangle\left(\sum_{j\neq i}\langle rj||ij\rangle + 2\langle r|h|i\rangle\right) + \langle r|h|i\rangle^2 \\
&= \sum_{abd}\langle ra||ib\rangle\left(\langle ra||id\rangle\Gamma_d^b - \sum_c\langle rc||id\rangle\Gamma_{ad}^{bc}\right) + 2\sum_{cd}\langle rc||id\rangle\Gamma_d^c\left(\sum_{j\neq i}\langle rj||ij\rangle + \langle r|h|i\rangle\right) + \\
&\quad + \sum_{k\neq i}\langle rk||ik\rangle\left(\sum_{j\neq i}\langle rj||ij\rangle + 2\langle r|h|i\rangle\right) + \langle r|h|i\rangle^2 \\
&= \sum_{abc}\langle ra||ib\rangle\left(\langle ra||ic\rangle\Gamma_c^b - \sum_d\langle rd||ic\rangle\Gamma_{ac}^{bd}\right) + 2\sum_{cd}\langle rc||id\rangle\Gamma_d^c h_{ir}^{\text{eff}} + (h_{ir}^{\text{eff}})^2
\end{aligned}
$$

$$
h_{ir}^{\text{eff}} = \sum_{j\neq i}\langle rj||ij\rangle + \langle r|h|i\rangle \qquad \text{(C.8)}
$$

$$
\begin{aligned}
N_i^{(+1)'} &= \langle \Phi_c \Psi_m^{(0)} | (\sum_{a<b} \sum_c \langle ab||ic\rangle a_i^\dagger a_c^\dagger a_b a_a + \sum_j \sum_a \langle aj||ij\rangle a_i^\dagger a_j^\dagger a_j a_a + \sum_a \langle a|h|i\rangle a_i^\dagger a_a) \times \qquad \text{(C.9)} \\
&\quad (\sum_{d<e} \sum_f \langle de||if\rangle a_d^\dagger a_e^\dagger a_f a_i + \sum_k \sum_d \langle dk||ik\rangle a_d^\dagger a_k^\dagger a_k a_i + \sum_d \langle d|h|i\rangle a_d^\dagger a_i) | \Psi_m^{(0)} \Phi_c \rangle \\
&= \sum_{a<b} \sum_c \sum_{d<e} \sum_f \langle ab||ic\rangle \langle de||if\rangle \langle \Psi_m^{(0)} | \delta_{ad}\delta_{be} a_c^\dagger a_f - \delta_{ad} a_c^\dagger a_e^\dagger a_b a_f - \delta_{bd}\delta_{ae} a_c^\dagger a_f + \\
&\quad + \delta_{bd} a_c^\dagger a_e^\dagger a_a a_f + \delta_{ae} a_c^\dagger a_d^\dagger a_b a_f - \delta_{be} a_c^\dagger a_d^\dagger a_a a_f + a_c^\dagger a_d^\dagger a_e^\dagger a_b a_a a_f | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{j\neq i} \sum_a \sum_{d<e} \sum_f \langle aj||ij\rangle \langle de||if\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_e^\dagger a_f - \delta_{ae} a_d^\dagger a_f + a_d^\dagger a_e^\dagger a_a a_f | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_a \sum_{d<e} \sum_f \langle a|h|i\rangle \langle de||if\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_e^\dagger a_f - \delta_{ae} a_d^\dagger a_f + a_d^\dagger a_e^\dagger a_a a_f | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{a<b} \sum_c \sum_{k\neq i} \sum_d \langle ab||ic\rangle \langle dk||ik\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_c^\dagger a_b - \delta_{bd} a_c^\dagger a_a + a_c^\dagger a_d^\dagger a_b a_a | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{ad} \sum_{j\neq i} \sum_{k\neq i} \langle aj||ij\rangle \langle dk||ik\rangle \langle \Psi_m^{(0)} | \delta_{ad} - a_d^\dagger a_a | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{ad} \sum_{k\neq i} \langle a|h|i\rangle \langle dk||ik\rangle \langle \Psi_m^{(0)} | \delta_{ad} - a_d^\dagger a_a | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{a<b} \sum_c \sum_d \langle ab||ic\rangle \langle d|h|i\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_c^\dagger a_b - \delta_{bd} a_c^\dagger a_a + a_c^\dagger a_d^\dagger a_b a_a | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{ad} \sum_{j\neq i} \langle d|h|i\rangle \langle aj||ij\rangle \langle \Psi_m^{(0)} | \delta_{ad} - a_d^\dagger a_a | \Psi_m^{(0)} \rangle + \sum_{ad} \langle a|h|i\rangle \langle d|h|i\rangle \langle \Psi_m^{(0)} | \delta_{ad} - a_d^\dagger a_a | \Psi_m^{(0)} \rangle \\
&= \sum_{a<b} \sum_c \sum_{d<e} \sum_f \langle ab||ic\rangle \langle de||if\rangle \langle \Psi_m^{(0)} | \delta_{ad}\delta_{be} a_c^\dagger a_f - \delta_{ad} a_c^\dagger a_e^\dagger a_b a_f - \delta_{bd}\delta_{ae} a_c^\dagger a_f + \\
&\quad + \delta_{bd} a_c^\dagger a_e^\dagger a_a a_f + \delta_{ae} a_c^\dagger a_d^\dagger a_b a_f - \delta_{be} a_c^\dagger a_d^\dagger a_a a_f + a_c^\dagger a_d^\dagger a_e^\dagger a_b a_a a_f | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_a \sum_{d<e} \sum_f h_{ia}^{\text{eff}} \langle de||if\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_e^\dagger a_f - \delta_{ae} a_d^\dagger a_f + a_d^\dagger a_e^\dagger a_a a_f | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{a<b} \sum_c \sum_d h_{id}^{\text{eff}} \langle ab||ic\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_c^\dagger a_b - \delta_{bd} a_c^\dagger a_a + a_c^\dagger a_d^\dagger a_b a_a | \Psi_m^{(0)} \rangle + \\
&\quad + \sum_{ad} h_{ia}^{\text{eff}} h_{id}^{\text{eff}} \langle \Psi_m^{(0)} | \delta_{ad} - a_d^\dagger a_a | \Psi_m^{(0)} \rangle \\
&= \sum_{a<b} \sum_c \sum_{d<e} \sum_f \langle ab||ic\rangle \langle de||if\rangle (\delta_{ad}\delta_{be} \Gamma_f^c - \delta_{ad} \Gamma_{bf}^{ce} - \delta_{bd}\delta_{ae} \Gamma_f^c + \\
&\quad + \delta_{bd} \Gamma_{af}^{ce} + \delta_{ae} \Gamma_{bf}^{cd} - \delta_{be} \Gamma_{af}^{cd} + \Gamma_{baf}^{cde}) + \sum_a \sum_{d<e} \sum_f h_{ia}^{\text{eff}} \langle de||if\rangle (\delta_{ad} \Gamma_f^e - \delta_{ae} \Gamma_f^d + \Gamma_{af}^{de}) + \\
&\quad + \sum_{a<b} \sum_c \sum_d h_{id}^{\text{eff}} \langle ab||ic\rangle (\delta_{ad} \Gamma_b^c - \delta_{bd} \Gamma_a^c + \Gamma_{ba}^{cd}) + \sum_{ad} h_{ia}^{\text{eff}} h_{id}^{\text{eff}} (\delta_{ad} - \Gamma_a^d)
\end{aligned}
$$

$$
\begin{aligned}
N_r^{(-1)'} &= \langle \Phi_c \Psi_m^{(0)} | (\sum_a \sum_{b<c} \langle ra||cb\rangle a_c^\dagger a_b^\dagger a_a a_r + \sum_a \sum_j \langle rj||aj\rangle a_a^\dagger a_j^\dagger a_j a_r + \\
&+ \sum_a \langle r|h|a\rangle a_a^\dagger a_r \Big(\sum_d \sum_{e<f} \langle rd||ef\rangle a_r^\dagger a_d^\dagger a_e a_f + \sum_d \sum_k \langle rk||dk\rangle a_r^\dagger a_k^\dagger a_k a_d + \sum_d \langle r|h|d\rangle a_r^\dagger a_d \Big) | \Psi_m^{(0)} \Phi_c \rangle \\
&= \sum_a \sum_{b<c} \sum_d \sum_{e<f} \langle ra||cb\rangle \langle rd||ef\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_c^\dagger a_b^\dagger a_e a_f - a_c^\dagger a_b^\dagger a_d^\dagger a_a a_e a_f | \Psi_m^{(0)} \rangle \\
&+ \sum_a \sum_j \sum_d \sum_{e<f} \langle rj||aj\rangle \langle rd||ef\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d^\dagger a_e a_f | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_d \sum_{e<f} \langle r|h|a\rangle \langle rd||ef\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d^\dagger a_e a_f | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_{b<c} \sum_d \sum_k \langle ra||cb\rangle \langle rk||dk\rangle \langle \Psi_m^{(0)} | a_c^\dagger a_b^\dagger a_a a_d | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_j \sum_d \sum_k \langle rj||aj\rangle \langle rk||dk\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_d \sum_k \langle r|h|a\rangle \langle rk||dk\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_{b<c} \sum_d \langle ra||cb\rangle \langle r|h|d\rangle \langle \Psi_m^{(0)} | a_c^\dagger a_b^\dagger a_a a_d | \Psi_m^{(0)} \rangle \\
&+ \sum_a \sum_j \sum_d \langle rj||aj\rangle \langle r|h|d\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_d \langle r|h|a\rangle \langle r|h|d\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle \\
&= \sum_a \sum_{b<c} \sum_d \sum_{e<f} \langle ra||cb\rangle \langle rd||ef\rangle \langle \Psi_m^{(0)} | \delta_{ad} a_c^\dagger a_b^\dagger a_e a_f - a_c^\dagger a_b^\dagger a_d^\dagger a_a a_e a_f | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_d \sum_{e<f} h_{ra}^{\text{eff}} \langle rd||ef\rangle \langle \Psi_m^{(0)} | a_a^\dagger a_d^\dagger a_e a_f | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_{b<c} \sum_d h_{rd}^{\text{eff}} \langle ra||cb\rangle \langle \Psi_m^{(0)} | a_c^\dagger a_b^\dagger a_a a_d | \Psi_m^{(0)} \rangle + \\
&+ \sum_a \sum_d h_{ra}^{\text{eff}} h_{rd}^{\text{eff}} \langle \Psi_m^{(0)} | a_a^\dagger a_d | \Psi_m^{(0)} \rangle \\
&= \sum_a \sum_{b<c} \sum_d \sum_{e<f} \langle ra||cb\rangle \langle rd||ef\rangle (\delta_{ad} \Gamma_{ef}^{cb} - \Gamma_{aef}^{cbd}) + \sum_a \sum_d \sum_{e<f} h_{ra}^{\text{eff}} \langle rd||ef\rangle \Gamma_{ef}^{ad} + \\
&+ \sum_a \sum_{b<c} \sum_d h_{rd}^{\text{eff}} \langle ra||cb\rangle \Gamma_{ad}^{cb} + \sum_a \sum_d h_{ra}^{\text{eff}} h_{rd}^{\text{eff}} \Gamma_d^a
\end{aligned}
\tag{C.10}
$$

The commutator that has to be considered for the $S_{rs}^{(-2)}$ subspace has the form

$$\begin{aligned}[H_v, V_{rs}^{(-2)}] &= \sum_{a<b}\sum_{cd} h_{cd}\langle rs||ba\rangle(\delta_{ac}a_r^\dagger a_s^\dagger a_b a_d - \delta_{bc}a_r^\dagger a_s^\dagger a_a a_d) + \\ &+\frac{1}{2}\sum_{a<b}\sum_{cdef}\langle rs||ba\rangle\langle cd|ef\rangle((\delta_{bd}\delta_{ac} - \delta_{bc}\delta_{da})a_r^\dagger a_s^\dagger a_f a_e + \\ &-\delta_{ac}a_r^\dagger a_s^\dagger a_d^\dagger a_b a_f a_e + \delta_{ad}a_r^\dagger a_s^\dagger a_c^\dagger a_b a_f a_e),\end{aligned} \tag{C.11}$$

with the final term and energy expression given by

$$\begin{aligned}&\langle\Psi_m^{(0)}|V_{rs}^{(-2)\dagger}[H_v, V_{rs}^{(-2)}]|\Psi_m^{(0)}\rangle = \langle\Psi_m^{(0)}|\sum_{a'<b'}\sum_{a<b}\sum_{cd} h_{cd}\langle rs||b'a'\rangle\langle rs||ba\rangle \times \\ &(\delta_{ac}a_{b'}^\dagger a_{a'}^\dagger a_b a_d - \delta_{bc}a_{b'}^\dagger a_{a'}^\dagger a_a a_d) + \frac{1}{2}\sum_{a'<b'}\sum_{a<b}\sum_{cdef}\langle rs||ba\rangle\langle rs||b'a'\rangle\langle cd|ef\rangle \times \\ &((\delta_{bd}\delta_{ac} - \delta_{bc}\delta_{da})a_{b'}^\dagger a_{a'}^\dagger a_f a_e - \delta_{ac}a_{b'}^\dagger a_{a'}^\dagger a_d^\dagger a_b a_f a_e + \delta_{ad}a_{b'}^\dagger a_{a'}^\dagger a_c^\dagger a_b a_f a_e)|\Psi_m^{(0)}\rangle,\end{aligned} \tag{C.12}$$

$$E_m^{(2)}(S_{rs}^{(-2)}) = \frac{N_{rs}^{(-2)}}{\epsilon_r + \epsilon_s - \epsilon_{ij}^{(+2)}}. \tag{C.13}$$

Let us consider the $S_{i,r}^{(0)'}$ subspace. The commutator needed has the form,

$$\begin{aligned}[H_v, V_{i,r}^{(0)}] &= \sum_{ab}\sum_{cd} h_{cd}\langle ra||ib\rangle(\delta_{ad}a_c^\dagger a_b a_r^\dagger a_i - \delta_{bc}a_a^\dagger a_d a_r^\dagger a_i) + \\ &+\frac{1}{2}\sum_{ab}\sum_{cdef}\langle ra||ib\rangle\langle cd|ef\rangle(\delta_{ae}a_c^\dagger a_d^\dagger a_f a_b a_r^\dagger a_i - \delta_{af}a_c^\dagger a_d^\dagger a_e a_b a_r^\dagger a_i + \\ &-\delta_{bc}a_a^\dagger a_d^\dagger a_f a_e a_r^\dagger a_i + \delta_{af}a_a^\dagger a_c^\dagger a_f a_e a_r^\dagger a_i),\end{aligned} \tag{C.14}$$

and consequently the term needed,

$$\langle\Psi_m^{(0)}|V_{i,r}^{(0)\dagger}[H_v,V_{i,r}^{(0)}]|\Psi_m^{(0)}\rangle = \langle\Psi_m^{(0)}|\sum_{ab}\sum_{cd} h_{cd}h_{ri}^{\text{eff}}\langle ra||ib\rangle(\delta_{ad}a_c^\dagger a_b - \delta_{bc}a_a^\dagger a_d) + \tag{C.15}$$
$$+\frac{1}{2}\sum_{ab}\sum_{cdef} h_{ri}^{\text{eff}}\langle ra||ib\rangle\langle cd|ef\rangle(\delta_{ae}a_c^\dagger a_d^\dagger a_f a_b - \delta_{af}a_c^\dagger a_d^\dagger a_e a_b - \delta_{bc}a_a^\dagger a_d^\dagger a_f a_e + \delta_{af}a_a^\dagger a_c^\dagger a_f a_e) +$$
$$+\sum_{ab}\sum_{a'b'}\sum_{cd} h_{cd}\langle ra||ib\rangle\langle ra'||ib'\rangle(\delta_{ad}(\delta_{a'c}a_{b'}^\dagger a_b - a_{b'}^\dagger a_c^\dagger a_{a'}a_b) - \delta_{bc}(\delta_{a'a}a_{b'}^\dagger a_d - a_{b'}^\dagger a_a^\dagger a_{a'}a_d)) +$$
$$+\frac{1}{2}\sum_{ab}\sum_{a'b'}\sum_{cdef}\langle ra||ib\rangle\langle ra'||ib'\rangle\langle cd|ef\rangle(\delta_{ae}a_{b'}^\dagger a_{a'}a_c^\dagger a_d^\dagger a_f a_b - \delta_{af}a_{b'}^\dagger a_{a'}a_c^\dagger a_d^\dagger a_e a_b +$$
$$-\delta_{bc}a_{b'}^\dagger a_{a'}a_a^\dagger a_d^\dagger a_f a_e + \delta_{af}a_{b'}^\dagger a_{a'}a_a^\dagger a_c^\dagger a_f a_e)|\Psi_m^{(0)}\rangle$$

has contributions with the three–body density matrices. Thus, the last terms from the previous equation, after the operators are ordered correctly, will have the form,

$$a_{b'}^\dagger a_{a'}a_c^\dagger a_d^\dagger a_f a_b = \delta_{a'c}a_{b'}^\dagger a_d^\dagger a_f a_b - \delta_{a'd}a_{b'}^\dagger a_c^\dagger a_f a_b + a_{b'}^\dagger a_c^\dagger a_d^\dagger a_{a'}a_f a_b. \tag{C.16}$$

One can see that a similar decomposition as was used in previous cases can to be done to calculate these terms. The cost of the multiplication by vector will thus be proportional to $O(M^3k^2ln)$ and to $O(M^3k^3ln)$. Thus,as in the previous two cases, it may be advantageous to seek some alternative methods of producing such terms. The energy expression for the $S_{i,r}^{(0)'}$ subspace reads

$$E_m^{(2)}(S_{i,r}^{(0)'}) = \frac{N_{i,r}^{(0)'}}{\epsilon_r - \epsilon_i + \epsilon_{i,r}^{(0)'}}. \tag{C.17}$$

If one creates the commutator expression for $S_j^{(+1)'}$, one obtains

$$
\begin{aligned}
[H_v, V_i^{(+1)'}] \;=\; & \sum_{ef}\sum_{a<b}\sum_{c} h_{ef}\langle ab||ic\rangle(\delta_{fa}a_e^\dagger a_b^\dagger a_c a_i - \delta_{bf}a_e^\dagger a_a^\dagger a_c a_i) + \\
& + \sum_{ef}\sum_{j}\sum_{a} h_{ef}\langle aj||ij\rangle\delta_{af}a_e^\dagger a_j^\dagger a_j a_i + \sum_{ef}\sum_{a} h_{ef}\langle a|h|i\rangle\delta_{af}a_e^\dagger a_i + \\
& + \frac{1}{2}\sum_{defg}\sum_{a<b}\sum_{c}\langle de|fg\rangle\langle ab||ic\rangle((\delta_{af}\delta_{gb} - \delta_{ga}\delta_{bf})a_d^\dagger a_e^\dagger a_c a_i - \delta_{af}a_d^\dagger a_e^\dagger a_b^\dagger a_g a_c a_i + \\
& + \delta_{bf}a_d^\dagger a_e^\dagger a_a^\dagger a_g a_c a_i + \delta_{ag}a_d^\dagger a_e^\dagger a_b^\dagger a_f a_c a_i - \delta_{bg}a_d^\dagger a_e^\dagger a_a^\dagger a_f a_c a_i - \delta_{cd}a_a^\dagger a_b^\dagger a_e^\dagger a_g a_c a_i + \\
& + \delta_{ce}a_a^\dagger a_b^\dagger a_d^\dagger a_g a_f a_i) + \sum_{defg}\sum_{j}\sum_{a}\langle de|fg\rangle\langle aj||ij\rangle(\delta_{af}a_d^\dagger a_e^\dagger a_g a_j^\dagger a_j a_i + \\
& + \delta_{ag}a_d^\dagger a_e^\dagger a_f a_j^\dagger a_j a_i) + \sum_{defg}\sum_{a}\langle de|fg\rangle\langle a|h|i\rangle(\delta_{af}a_d^\dagger a_e^\dagger a_g a_i - \delta_{ag}a_d^\dagger a_e^\dagger a_f a_i).
\end{aligned}
\tag{C.18}
$$

We can now calculate the $V_i^{(+1)'\dagger}[H_v, V_i^{(+1)'}]$ expression; however, due to the length of the expression we will list only the $V_i^{(+1)'_1\dagger}[H_v, V_i^{(+1)'}]$ term since it contains the four–body density matrices,

$$
\begin{aligned}
& \langle\Psi_m^{(0)}|V_i^{(+1)'_1}[H_v, V_i^{(+1)'}]|\Psi_m^{(0)}\rangle = \langle\Psi_m^{(0)}|\sum_{a'<b'}\sum_{c'}\langle a'b'||ic'\rangle\Bigg(\sum_{ef}\sum_{a<b}\sum_{c} h_{ef}\langle ab||ic\rangle \times \\
& (\delta_{fa}a_{c'}^\dagger a_{b'}a_{a'}a_e^\dagger a_b^\dagger a_c - \delta_{bf}a_{c'}^\dagger a_{b'}a_{a'}a_e^\dagger a_a^\dagger a_c) + \sum_{e}\sum_{a} h_{ea}h_{ia}^{\mathrm{eff}}a_{c'}^\dagger a_{b'}a_{a'}a_e^\dagger + \\
& + \frac{1}{2}\sum_{defg}\sum_{a<b}\sum_{c}\langle de|fg\rangle\langle ab||ic\rangle((\delta_{af}\delta_{gb} - \delta_{ga}\delta_{bf})a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_c - \delta_{af}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_b^\dagger a_g a_c + \\
& + \delta_{bf}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_a^\dagger a_g a_c + \delta_{ag}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_b^\dagger a_f a_c - \delta_{bg}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_a^\dagger a_f a_c + \\
& - \delta_{cd}a_{c'}^\dagger a_{b'}a_{a'}a_a^\dagger a_b^\dagger a_e^\dagger a_g a_c + \delta_{ce}a_{c'}^\dagger a_{b'}a_{a'}a_a^\dagger a_b^\dagger a_d^\dagger a_g a_f) + \\
& + \sum_{defg}\sum_{a} h_{ai}^{\mathrm{eff}}\langle de|fg\rangle(\delta_{af}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_g - \delta_{ag}a_{c'}^\dagger a_{b'}a_{a'}a_d^\dagger a_e^\dagger a_f)\Bigg)|\Psi_m^{(0)}\rangle
\end{aligned}
\tag{C.19}
$$

while the other terms $(V_i^{(+1)'_2\dagger} + V_i^{(+1)'_3\dagger})[H_v, V_i^{(+1)}]$ are simpler. The terms containing the four–body density matrix

$$
C_i^{(+1)'}(4) = \sum_{\substack{a'<b'\\ c'}}\sum_{\substack{a<b\\ c}}\sum_{deg}\langle a'b'||ic'\rangle\langle de|ag\rangle\langle ab||ic\rangle\langle\Psi_m^{(0)}|a_{c'}^\dagger a_d^\dagger a_e^\dagger a_b^\dagger a_{b'}a_{a'}a_g a_c|\Psi_m^{(0)}\rangle \tag{C.20}
$$

can be factorized as were the terms discussed earlier. Let us now show just one of the very expensive terms arising during such a decomposition, thus proving that the whole procedure is very expensive

$$C_p^{(+1)'}(a,g,b',a') = \sum_{de}\langle de|ag\rangle\langle L_\mu|a_d^\dagger a_e^\dagger a_{b'}a_{a'}|L_{\mu'}\rangle, \tag{C.21}$$

$$C_q^{(+1)'}(i,a,c',g) = \sum_{bc}\langle ab||ic\rangle\langle R_\nu|a_{c'}^\dagger a_b^\dagger a_g a_c|R_{\nu'}\rangle \tag{C.22}$$

before performing the multiplication with the wave function

$$C_i^{(+1)'}(4) = \sum_{\mu\nu}\sum_{\mu'\nu'}\sum_{ag}\sum_{\substack{a'<b'\\ c'}} c_{\mu\nu}c_{\mu'\nu'}\langle a'b'||ic'\rangle C_p^{(+1)'}(a,g,b',a')C_q^{(+1)'}(i,a,c',g). \tag{C.23}$$

The costs of preparing the $C_p^{(+1)'}(a,g,b',a')$ and $C_q^{(+1)'}(i,a,c',g)$ operators are $O(M^2k^6)$ and $O(M^2k^5l)$ respectively. The space needed for storage is $O(M^2k^4)$ and $O(M^2k^3l)$ elements at every step of a sweep. The multiplication, which should be done only once at the last step of the chain, requires $O(M^3k^5l)$ operations. Thus, it seems that to prepare the four–body density matrix summed with appropriate integrals is very expensive (even after the attempted factorization), and such a step should be avoided if at all possible.

If one considers the appropriate expressions in the $S_r^{(-1)'}$ subspace, one obtains similar expressions for the commutator as in the previously analyzed case. The terms appearing in the denominator also contain the four–body density matrix, and thus are similarly expensive. The cost of creating such operators and their storage as well as the cost of multiplication, are analogous to the previous one having the core label i replaced by the virtual label r. The contribution to the energy for $S_i^{(1)'}$ and$S_r^{(-1)'}$ subspaces are

$$E(S_i^{(1)'}) = -\frac{N_i^{(1)'}}{\epsilon_i^{(1)'} - \epsilon_i} \tag{C.24}$$

and

$$E(S_r^{(-1)'}) = -\frac{N_r^{(-1)'}}{\epsilon_r^{(-1)'} - \epsilon_r}, \tag{C.25}$$

respectively, where $\epsilon_i^{(1)'}$ and $\epsilon_r^{(1)'}$ are calculated using the above expressions.

Bibliography

[1] A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).

[2] A. Schäfer, C. Huber, and R. Ahlrichs, J. Chem. Phys. **100**, 5829 (1994).

[3] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Sctructure Theory.* (Dover Publications, Inc., Mineola, New York, 1996).

[4] Notes from the lectures given by J. Paldus in Nijmegen, Netherlands. Spring 1980.

[5] F. Coester, Nucl. Phys. **7**, 421 (1958).

[6] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[7] B. O. Roos (Ed), *Lecture Notes in Quantum Chemistry, European Summer School in Quantum Chemistry* (Springer-Verlag, Berlin, Heidelberg, 1994).

[8] D. Mukherjee and S. Pal, Adv. Quantum Chem **20**, 291 (1989).

[9] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[10] G. Fano, F. Ortolani, and L. Ziosi, J. Chem. Phys. **108**, 9246 (1998).

[11] R. Pariser and R. Parr, J. Chem. Phys. **21**, 466 (1953).

[12] R. Pariser and R. Parr, J. Chem. Phys. **21**, 767 (1953).

[13] J. A. Pople, Trans. Faraday Soc. **49**, 1375 (1953).

[14] G. L. Bendazzoli, S. Evangelisti, G. Fano, F. Ortolani, and L. Ziosi, J. Chem. Phys. **110**, 1277 (1999).

[15] J. Paldus, M. Takahashi, and R. W. H. Cho, Phys. Rev. B **30**, 4267 (1984).

[16] J. Paldus, J. Čížek and M. Takahashi, Phys. Rev. A **30**, 2193 (1984).

[17] R. Podeszwa and S. Stolarczyk, J. Chem. Phys. **116**, 480 (2002).

[18] Hubbard model Hamiltonian:

$$\hat{H} = -t \sum_{<ij>,\sigma} (c^{\dagger}_{i\sigma} c_{j\sigma} + c^{\dagger}_{j\sigma} c_{i\sigma}) + U \sum_{i} n_{i\uparrow} n_{i\downarrow},$$

where $c^{\dagger}_{i\sigma}$ creates an electron of spin $\sigma =\uparrow$ or $\downarrow$ on orbital i,$n_{i\sigma} = c^{\dagger}_{i\sigma} c_{i\sigma}$, t is the hopping matrix element between neighboring orbitals, $< ij >$ denotes the sum over pairs of the nearest neighbor orbitals, and U is the energy cost due to the Coulomb repulsion of two electrons on the same orbital.

[19] S. R. White and R. L. Martin, J. Chem. Phys. **110**, 4127 (1999).

[20] S. Daul, I. Ciofini, C. Daul, and S. R. White, Int. J. Quantum Chem. **79**, 331 (2000).

[21] A. O. Mitrushenkov, G. Fano, F. Ortolani, R. Linguerri, and P. Palmieri, J. Chem. Phys. **115**, 6815 (2001).

[22] G. K. -L. Chan and M. Head-Gordon, J. Chem. Phys. **116**, 4462 (2002).

[23] E. Cuthill and J. McKee, in Proceedings of the 24th National Conference of the ACM (1969).

[24] J. Liu and A. Sherman, SIAM (Soc. Ind. Appl. Math.) J. Numer. Anal. **13**, 198 (1975).

[25] Ö. Legeza, J. Röder, and B. A. Hess, Phys. Rev. B **67**, 125114 (2003).

[26] Ö. Legeza, J. Röder, and B. A. Hess, Mol. Phys. **101**, 2019 (2003).

[27] Ö. Legeza and J. Sólyom, Phys. Rev. B **68**, 195116 (2003).

[28] G. K. -L. Chan, J. Chem. Phys. **120**, 3172 (2004).

[29] G. K. -L. Chan and M. Head-Gordon, J. Chem. Phys. **118**, 8551 (2003).

[30] G. Moritz, B. A. Hess, and M. Reiher, J. Chem. Phys. **122**, 024107 (2005).

[31] G. Moritz and M. Reiher, J. Chem. Phys. **126**, 244109 (2007).

[32] G. K. -L. Chan, M. Kállay, and J. Gauss, J. Chem. Phys. **121**, 6110 (2004).

[33] J. Hachmann, W Cardoen, and G. K. -L. Chan, J. Chem. Phys. **125**, 144100 (2006).

[34] S. Östlund and S. Rommer, Phys. Rev. Lett. **75**, 3537 (1995).

[35] S. Rommer and S. Östlund, Phys. Rev. B **55**, 2164 (1997).

[36] G. K. -L. Chan, J. J. Dorando, D. Ghosh, J. Hachmann, E. Neuscamman, H. Wang, T. Yanai, arXiv:0711.1398v1 [cond-mat.str-el].

[37] D. Ghosh,T. Yanai, and G. K. -L. Chan, submitted .

[38] D. Zgid and M. Nooijen, submitted .

[39] G. Vidal, Phys. Rev. Lett. **93**, 040502 (2004).

[40] F. Verstraete, D. Porras, and J. I. Cirac, Phys. Rev. Lett. **93**, 22720 (2004).

[41] F. Verstraete and J. I. Cirac, cond-mat/0407066.

[42] F. Verstraete, M. Wolf, D. Pérez-García and J. I. Cirac, IJMPB **20**, 5152 (2006).

[43] V. Murg, F. Verstraete, and J. I. Cirac, Phys. Rev. A **75**, 033605 (2007).

[44] D. Zgid and M. Nooijen, submitted .

[45] J. J. Dorando, J. Hachmann, and G. K. -L. Chan, J. Chem. Phys. **127**, 084109 (2007).

[46] D. Zgid and M. Nooijen, J. Chem. Phys. **128**, 014107 (2008).

[47] S. R. White, J. Chem. Phys. **117**, 7472 (2002).

[48] T. Yanai and G. K. -L. Chan, J. Chem. Phys. **124**, 194106 (2006).

[49] T. Yanai and G. K. -L. Chan, J. Chem. Phys. **127**, 104107 (2007).

[50] S. R. White, Phys. Rev. B **72**, 180403 (2005).

[51] C. Angeli, R. Cimiraglia, S. Evangelisti, T. Leininger, and J. P. Malrieu, J. Chem. Phys. **114**, 10252 (2001).

[52] C. Angeli, R. Cimiraglia, , and J. P. Malrieu, Chem. Phys. Lett. **350**, 297 (2001).

[53] C. Angeli, R. Cimiraglia, and J. P. Malrieu, J. Chem. Phys. **117**, 9138 (2002).

[54] R. Pollet, A. Savin, T. Leininger, and H. Stoll, J. Chem. Phys. **116**, 1250 (2002).

[55] T. Leininger, H. Stoll, H-. J. Werner, and A. Savin , Chem. Phys. Lett. **275**, 151 (1997).

[56] A. Savin, in *Recent Developments and Applications of Modern Density Functional Theory,* edited by J. M. Seminario (Elsevier, Amsterdam, 1996).

[57] Heisenberg model Hamiltonian:

$$\hat{H} = J \sum_{<ij>} \mathbf{S_i} \cdot \mathbf{S_j},$$

where $\mathbf{S_i} \cdot \mathbf{S_j} = S_i^z S_j^z + \frac{1}{2}(S_i^+ S_j^- + S_i^- S_j^+)$.

[58] Hubbard Model $\rightarrow U/t \gg 1 \rightarrow$ t-J Model $\rightarrow$ half $-$ filling $\rightarrow$ Heisenberg Model.

[59] K. G. Wilson, Rev. Mod. Phys. **55**, 583 (1983).

[60] K. G. Wilson, Rev. Mod. Phys. **47**, 773 (1975).

[61] X. Wang, K. Hallberg, and F. Naef, Lecture Notes in Physics **528**, 173 (1999).

[62] T. D. Kuhner and S. R. White, Phys. Rev. B **60**, 335 (1999).

[63] S. Ramasesha and K. Tandon, Lecture Notes in Physics **528**, 247 (1999).

[64] W. Barford and R. J. Bursill, Synthetic Metals **119**, 251 (2001).

[65] A. Race, W. Barford, and R. J. Bursill, Synthetic Metals **119**, 625 (2001).

[66] S. Moukouri and L. G. Caron, Lecture Notes in Physics **528**, 329 (1999).

[67] E. Jeckelmann and S. R. White, Phys. Rev. B **57**, 6376 (1998).

[68] C. Zhang, E.Jeckelmann, and S. R. White, Phys. Rev. Lett. **80**, 2661 (1998).

[69] J. Kondo, *Solid State Physics* edited by F. Seitz, D. Turnbull, and Ehrenreich, vol. 23 (Academic, New York, 1969).

[70] S. R. White and R. M. Noack, Phys. Rev. Lett. **68**, 3487 (1992).

[71] U. Schollwöck, Rev. Mod. Phys. **77**, 259 (2005).

[72] S. R. White, Phys. Rev. B **48**, 10345 (1993).

[73] S. R. White, Phys. Rep. **301**, 187 (1998).

[74] R. P. Feynman, *Statistical Mechanics. A Set of Lectures* (W. A. Benjamin, Inc., 1972).

[75] J. Gaite, Mod. Phys. Lett. A **16**, 1109 (2001).

[76] J. Gaite, *Entangelment entropy and the density matrix renormalization group* (2003).

[77] A. Galindo and M. A. Martín-Delgano, Rev. Mod. Phys. **74**, 347 (2002).

[78] T. J. Osborne and M. A. Nielsen, Quantum Inf. Process. **1**, 45 (2002).

[79] J. I. Latorre, E. Rico, and G. Vidal, Quantum Inf. Comput. **4**, 48 (2004).

[80] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Qunatum Information* (Cambridge University, Cambridge, England, 2000).

[81] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).

[82] G. L. G. Sleijpen and H. A. van der Vorst, SIAM J. Matrix Anal. Appl. **17**, 401 (1996).

[83] C. Lanczos, J. Res. Nat. Bur. Standards **45**, 255 (1950).

[84] J. Cullum and R. A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations* Vols. I and II (1985).

[85] G. Golub and C. F. van Loan, *Martix Computations* (John Hopkins University, Baltimore, MD, 1996).

[86] S. R. White, Phys. Rev. Lett. **77**, 2863 (3633).

[87] I. P. McCulloch and M. Gulácsi, Aust. J. Phys. **53**, 597 (2000).

[88] I. P. McCulloch and M. Gulácsi, Philos. Mag. Lett. **81**, 447 (2001).

[89] I. P. McCulloch and M. Gulácsi, Europhys. Lett. **57**, 852 (2002).

[90] K. Yamada, *Electron Correlation in Metals* (Cambridge University Press, 2004).

[91] G. Moritz and M. Reiher, J. Chem. Phys. **124**, 034103 (2006).

[92] S. R. White, Phys. Rev. Lett. **77**, 3633 (1997).

[93] Ö. Legeza and J. Sólyom, Phys. Rev. B **56**, 14449 (1997).

[94] G. Moritz, A. Wolf, and M. Reiher, J. Chem. Phys. **123**, 184105 (2005).

[95] DALTON, *a molecular electronic structure program,* Release 2.0, written by T. Helgaker, H. J. Jensen, P. Joergensen, J. Olsen, K. Ruud, H. Aagren, A. A. Auer, K. L. Bak, V. Bakken, O. Christiansen, S. Coriani, P. Dahle, E. K. Dalskov, T. Enevoldsen, B. Fernandez, C. Haettig, K. Hald, A. Halkier, H. Heiberg, H. Hettema, D. Jonsson, S. Kirpekar, W. Klopper, R. Kobayashi, H. Koch, K. V. Mikkelsen, P. Norman, T. B. Pedersen, M. J. Packer, T. A. Ruden, P. Salek, A. Sanchez, T. Saue, S. P. A. Sauer, B. Schimmelpfennig, K. O. Sylvester-Hvid, P. R. Taylor, and O. Vahtras.

[96] T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[97] A. Klümper, A. Schadschneider, and J. Zittartz, J. Phys. A: Math. Gen. **24**, L955 (1991).

[98] M. Fannes, B. Nachtergaele, and R. F. Werner, Commun. Math. Phys. **144**, 443 (1992).

[99] B. Derrida, M. R. Evans, V. Hakim, and V. Pasquier , J. Phys. A: Math. Gen. **26**, 1493 (1993).

[100] A. Klümper, A. Schadschneider, and J. Zittartz, Z. Phys. B **87**, 281 (1992).

[101] Although in the physics literature the A matrices are called the projection matrices, we decided to call them truncation matrices since they are not idempotent ($A^2 \neq A$) as expected for usual projection matrix.

[102] Formally, we should write $(A^k[s^k])_{l_\nu;l_\mu}$ instead of $(A[s^k])_{l_\nu;l_\mu}$ to explicitly express the dependence of the A matrix on the k label of the orbital. However, since we label the states of the single orbital s^k with the position label k, we do not repeat this label for the A matrix.

[103] In the same way in the coupled cluster method the exponential ansatz for the wave function is uniquely chosen, but there is some freedom how the equations for the amplitudes may be defined.

[104] G. K. -L. Chan, P. W. Ayers, and E. S. Croot,, J. Stat. Phys. **109**, 289 (2002).

[105] B. S. Shastry and B. Sutherland, Physica B & C **108**, 1069 (1981).

[106] A. Isacsson and O. F. Syljuåsen, Phys. Rev. E **74**, 026701 (2006).

[107] B. O. Roos, Adv. Chem. Phys. **69**, 399 (1987).

[108] The Γ_{ijkl} for some i, j, k and l labels can be obtain multiple times within a sweep. Thus, one has a freedom either to take such an element at a particular step within a sweep or sum all the repeating elements for every combination of i, j, k, l labels and divide the resulting sum by the number of occurrences of a particular element within a sweep.

[109] B. O. Roos, P. R. Taylor, and P. E. M. Siegbahn, Chem. Phys. **48**, 157 (1980).

[110] J. Olsen, B. O. Roos, P. Jørgensen, and H. J. Aa. Jensen, J. Chem. Phys. **89**, 2185 (1988).

[111] H. J. Aa. Jensen, P. Jørgensen, T. Helgaker, and J. Olsen, Chem. Phys. Lett. **162**, 355 (1989).

[112] P. -Å. Malmqvist, A. Rendell, and B. O. Roos, J. Chem. Phys. **94**, 5477 (1990).

[113] Z. Shuai, J. L. Brédas, S. K. Pati, and S. Ramasesha, Proc. SPIE **293**, 3145 (1997).

[114] D. Yaron, E. E. Moore, Z. Shuai, and J. L. Brédas, J. Chem. Phys. **108**, 7451 (1998).

[115] G. K. -L. Chan and T. van Voorhis, J. Chem. Phys. **122**, 204101 (2005).

[116] J. Hachmann, J. J. Dorando, M. Avilés, and G. K. -L. Chan, J. Chem. Phys. **127**, 134309 (2007).

[117] A. O. Mitrushenkov, R. Linguerri, P. Palmieri, and G. Fano, J. Chem. Phys. **119**, 4148 (2003).

[118] H. Ma, C. Liu, and Y. Jiang, J. Chem. Phys. **120**, 9316 (2004).

[119] H. Ma, F. Cai, C. Liu, and Y. Jiang, J. Chem. Phys. **122**, 104909 (2005).

[120] H. Ma, C. Liu, and Y. Jiang, J. Chem. Phys. **123**, 084303 (2005).

[121] ACES II is a program product of the Quantum Theory Project, University of Florida. Authors: J. F. Stanton, J. Gauss, S.A. Perera, J. D. Watts, A. D. Yau, M. Nooijen, N. Oliphant, P. G. Szalay, W. J. Lauderdale, S. R. Gwaltney, S. Beck, A. Balková, D. E. Bernholdt, K. K. Baeck, P. Rozyczko, H. Sekino, C. Huber, J. Pittner, W. Cencek, D. Taylor, and R. J. Bartlett. Integral packages included are VMOL (J. Almlöf and P. R. Taylor); VPROPS (P. Taylor); ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen, J. Olsen, and P. R. Taylor); HONDO/GAMESS (M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. J. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery).

[122] H. J. Aa. Jensen, H. Ågren, and J. Olsen, *Modern Techniques in Computational Chemistry* (ESCOM Science Publishers B. V., 1990).

[123] B. H. Lengsfield III, J. Chem. Phys. **73**, 382 (1980).

[124] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).

[125] P. Pulay, J. Comput. Chem. **3**, 556 (1982).

[126] C. Hampel, K. A. Peterson, and H-. J Werner, Chem. Phys. Lett. **190**, 1 (1992).

[127] M. Nooijen and V. Lotrich, J. Chem. Phys. **113**, 494 (2000).

[128] M. Nooijen and V. Lotrich, Int. J. Quantum Chem. **547**, 253 (2001).

[129] M. Nooijen M and K. R. Shamasundar, Collect. Czech. Chem. Commun. **70**, 1082 (2005).

[130] M. Nooijen, Int. J. Quantum Chem. **95**, 768 (2003).

[131] J. Neugebauer J, E. J. Baerends, and M. Nooijen, J. Chem. Phys. **121**, 6155 (2004).

[132] T. H. Dunning Jr., J. Chem. Phys. **90**, 1007 (1989).

[133] T. H. Dunning Jr., J. Chem. Phys. **53**, 2823 (1970).

[134] T. H. Dunning Jr., J. Chem. Phys. **55**, 716 (1971).

[135] B. H. Brandow, Rev. Mod. Phys. **39**, 771 (1967).

[136] P. H. Durand and J. P. Malrieu, Adv. Chem. Phys. **67**, 1 (1987).

[137] K. Andersson, P. Malmqvist, B. O. Roos, A. J. Sadlej, and K. Wolinski, J. Phys. Chem. **94**, 5483 (1990).

[138] K. Andersson, P. Malmqvist, and B. O. Roos, J. Chem. Phys. **96**, 1218 (1992).

[139] K. G. Dyall, J. Chem. Phys. **102**, 4909 (1995).

[140] G. K. -L. Chan, private communication.

[141] F. Colmenero and C. Valdemoro, Phys. Rev. A **47**, 979 (1993).

[142] K. Yasuda and H. Nakatsuji, Phys. Rev. Lett. **76**, 1039 (1996).

[143] D. A. Mazziotti, Chem. Phys. Lett. **289**, 419 (1999).

[144] S. Gusarov, P. Å. Malmqvist, R. Lindth, and B. O. Roos, Theor. Chem. Acc. **112**, 84 (2000).

[145] J. Gräfenstein and D. Cremer, Chem. Phys. Lett. **316**, 569 (2000).

[146] J. Toulouse, A. Savin, and H.-J. Flad, Int. J. Quantum Chem. **100**, 1047 (2004).

[147] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).