GENERALIZED GEMINAL FUNCTIONAL THEORY

by

BEREND CHRISTOPHER RINDERSPACHER

(Under direction of Peter R. Schreiner)

ABSTRACT

In order for computational chemistry to become a viable alternative to experiment and become a true predictor for science and industry, high accuracy must be attainable in a time-frame that does not exceed the time spent on experiment. Current methods are either hampered by a lack of accuracy (e.g., Hartree-Fock theory) or, in the worst case, insurmountable computational efforts for anything but the simplest problems (i.e. full configuration interaction [FCI]). The working equation of quantum chemistry, the Schrödinger equation, relies on two-body operators only. This simplicity is deceiving in that the boundary conditions enforced by the nature of electrons, which are fermions, links these two-body operators in a very complex manner. The exact solution given by FCI results in a computational cost which is exponential in the size of the computed molecule. Due to the simple nature of the contracted Hamiltonian ($K = nh + \binom{n}{2}\frac{1}{r_{12}}$), which is a two-electron operator, the computational complexity of the Schrödinger equation should be bounded by $O(n^6)$. this work shows that the ground state of a molecule can be expressed very accurately in terms of a geminal g in the form of $\hat{A}g(1,2)f(3\dots)$, where $\hat{A}$ is the anti-symmetrizer and f is some $n-2$-electron function. Using this result, the generalized antisymmetric geminal product (GAGP) combines the ideas of geminal functional theory and Hartree-Fock theory to several geminals ($\psi = \prod_{i=1}^{n/2} g_i(2i-1, 2i)$). GAGP lays the groundwork for the highest accuracy at a cost that scales with $O(n^6)$. The GAGP approach was further tested on sample 4-electron systems, Li^-, Be, B^+, as well as LiH, BeH^+, He_2, H_3^- and their respective dissociated species, with the moderately large basis set cc-pVDZ. The calculations confirmed the theoretical results and rendered excellent accuracy.

INDEX WORDS: Geminal Functional Theory; GFT; Quantum Chemistry; Reduced Density Matrix.

GENERALIZED GEMINAL FUNCTIONAL THEORY

by

BEREND CHRISTOPHER RINDERSPACHER

Diplom-Chemiker, Georg-August-Universität Göttingen, Germany, 2000

A Dissertation Submitted to the Graduate Faculty of the University of Georgia in Partial
Fulfillment of the Requirement for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2005

GENERALIZED GEMINAL FUNCTIONAL THEORY

by

BEREND CHRISTOPHER RINDERSPACHER

Major Professor:	Peter R. Schreiner
Committee:	Henry F. Schaefer
	Nigel G. Adams

Electronic Version Approved:

Maureen Grasso

Dean of the Graduate School

The University of Georgia

May 2005

ACKNOWLEDGMENTS

I would like to thank everybody whose help was instrumental to this work. I thank Dr. Peter R. Schreiner for his continuing support and advice even through such adversity as a separation by the Atlantic ocean as well as providing me with all the hardware necessary to do our computational work. I thank Dr. Henry F. Schaefer for welcoming me as one of his group in Peter's absence as well as all the paper work he and Linda S. Rowe, to whom I am also indebted, did for me. I thank Drs. Wesley Allen and Paul von Ragué-Schleyer and everybody else who showed up for group meetings including the presenters for all the enlightening insights elucidated in those seminars. Of course, I am as well grateful to Drs. Lucia M. Babcock and Nigel G. Adams, who have served so faithfully on my advisory committee. I would also like to thank my fellow group members Derek M. Wolfe, L. Ellen Downs, Matthias Prall, Alexander Wittkopp for a wonderful time and most invigorating discussions. Among Dr. Schaefer's group I thank Steven Wheeler, Joe Kenny, Edward Valeev, and Joe Larkin for their input and friendship (this obviously also goes for the aforementioned).

I thank the US Department of Energy and the Krell Institute for supporting me through the Computational Graduate Student Fellowship – I recommend any graduate student in computational science apply.

I apologize if I have forgotten anybody. It is not a personal slight, but my inadequacy.

Contents

CHAPTER 1

INTRODUCTION

When quantum mechanics was introduced in the early 20th century, only the most simple problems like the hydrogen atom could be tackled. With the emergence of computers of ever-increasing power in the seventies of the previous century, it became feasible to solve quantum mechanics problems on a more practical level. Yet the challenge remains to improve results systematically instead of using the dumb brute force method of waiting for bigger and faster computers.

Quantum physics is essentially ruled by a single eigenvalue problem, and all current methods attempt in one way or another to solve this one equation – the Schrödinger equation as introduced by Austrian physicist Erwin Schrödinger. The active operator in this equation is the Hamiltonian named after the Irish astronomer Sir William Rowan Hamilton. The Hamiltonian incorporates all physical interactions of which the inter-electronic contributions and the fermion nature of the electrons are the culprits of complexity. For not explicitly interacting or bosonic particles, the equations may be disentangled and a closed solution derived. Neither of those conditions is true for electrons and hence most efforts are directed at computing the deviation from non-interacting solutions. Such methods include the average potential Hartree-Fock *ansatz*,[1,2] configuration interaction,[3,4] coupled cluster theory and, coming from a different angle, density functional theory. All of these will be introduced in more detail later. Configuration interaction and coupled cluster theory can be systematically improved; but to do so increases computational complexity systematically as well. Despite these problems, quantum chemistry has contributed in many ways to the general field of chemistry, like Hückel theory,[5] molecular-orbital theory, or the Woodward-Hoffman rules; sometimes even superseding experiment.[6,7]

1. Aim of Research

The purpose of this research was to develop a conceptually and computationally simple method of exceptional accuracy. The nature of the Hamiltonian suggests a certain simplicity of its eigensolutions; yet, it has proven to be quite elusive in explicit terms. Generalized anti-symmetrized geminal products, as introduced later, incorporate much of this simplicity.

CHAPTER 2

BACKGROUND

1. Foundations of Quantum Chemistry

In the 1920s quantum mechanics was developed, explaining reality in terms of operators and eigenvalue problems.[5] Operators are linear mappings from space X into space Y, often from a space of functions into the same space. Quantum mechanical spaces are usually Hilbert spaces, i.e., they are endowed with a scalar product and its corresponding norm and metric, e.g., $\mathcal{L}_2(X, \mu)$, the space of square integrable functions on the measurable space (X, μ), with scalar product $\langle\psi|\phi\rangle = \int \psi^* \phi d\mu(X)$. The hermitian operators represent measurable quantities while the functions represent the physical state description.

1.1. The Hamilton Operator

The Hamilton operator, which is named after the Irish astronomer and mathematician Sir William Rowan Hamilton, is the fundamental operator of quantum mechanics. It describes all the energetic physics of reality. The associated eigenvalue equation (1) is the Schrödinger equation named after Austrian physicist Erwin Schrödinger.[8]

$$E\Psi = \mathbf{H}\Psi = [\mathbf{T} + \mathbf{V}]\,\Psi \tag{1}$$

$\mathbf{H}$ is hermitian, i.e., $\left\langle\psi\middle|\mathbf{H}\phi\right\rangle = \left\langle\phi\middle|\mathbf{H}\psi\right\rangle^*$, and composed of a kinetic component $\mathbf{T}$ and a (scalar) potential component $\mathbf{V}$. In this description, we have omitted the anti-symmetrization conditions for fermions, like electrons, as well as the symmetrization conditions for bosons, like the helium nucleus. Instead, it is assumed that $\mathbf{H}$, which is fully symmetric, is applied to functions which are anti-symmetric for fermions, i.e., $f(x_1, x_2) = -f(x_2, x_1)$, and symmetric for bosons, i.e., $f(x_3, x_4) = f(x_4, x_3)$.

$$\mathbf{T} = \sum_i \frac{1}{2m_i}\left(\mathbf{p}_i - \frac{q_i}{c}\mathbf{A}_i\right)^2 = \sum_i \frac{1}{2m_i}\mathbf{p}_i^2 + \sum_i \frac{q_i}{2cm_i}\mathbf{A}_i^2 - \sum_i \frac{q_i}{2cm_i}\left\{\mathbf{p}_i, \mathbf{A}_i\right\} \tag{2}$$

$$\mathbf{p}_i = -\mathcal{I} \cdot \hbar \cdot \begin{pmatrix} \frac{\partial}{\partial x_i} \\ \frac{\partial}{\partial y_i} \\ \frac{\partial}{\partial z_i} \end{pmatrix} \tag{3}$$

The general description of $\mathbf{T}$ can be found in equation (2).[2,5,9] Here, $\mathbf{A}$ is a vector potential, that incorporates, for instance, magnetic fields; $\hbar = h/2\pi$ is comprised of the fundamental Planck's constant; q_i, m_i are the charge and mass, respectively, of particle i; c is the speed of light; and $\{A, B\} = AB + BA$ denotes the anti-commutator.

The scalar potential $\mathbf{V}$ usually contains the particle interactions. Due to their mass the nuclei move very slowly compared to electrons. Therefore, Born and Oppenheimer introduced the approximation of stationary nuclei with respect to electrons.[2] The Born-Oppenheimer approximation simplifies the solution of the Schrödinger equation to a solution for electrons only. We will restrict ourselves to the Born-Oppenheimer approximation for the remainder of the dissertation unless explicitly noted otherwise.

$$\mathbf{V} = \sum_{i>j} \frac{q_i q_j}{r_{ij}}, \; r_{ij} = \left|\vec{x}_i - \vec{x}_j\right| \tag{4}$$

In this description, we have neglected the fundamental property of spin, for which there is no counterpart in classical mechanics. This property was discovered by Stern and Gerlach in 1922 in an experiment in which they noticed that a stream of electrons is split into two separate streams when subjected to a magnetic field.[9,10] The intrinsic property was dubbed "spin" and electrons were assigned either α or β spin depending on to which stream of electrons they belonged. The interactions between spins and angular momenta are usually neglected in terms of operators but are retained by employing wave functions of explicit spins.

1.2. Second Quantization

In the previous section, the Hamiltonian was defined in terms of derivative and multiplicative operators, but sometimes it is more convenient or simpler to discuss the Hamiltonian in terms of operators on tensor subspaces. Tensor spaces are best introduced via an example pertinent to our problem of solving the Schrödinger equation.[11] Assume two square-Lebesgue integrable functions $f, g \in V = \mathcal{L}_2(X)$, then $f(x) \cdot g(y) =: f \otimes g$ is a tensor from the tensor product $T^2(V) = V \otimes V = \text{span}\{f \otimes g | f, g \in V\}$. A tensor space is then the direct vector sum $\bigoplus_{i=0}^{\infty} T^i(V)$, where T^0 is the field k of the (Hilbert) vector space $T^1(V) = V$. Multiplication and addition in this algebra reflect the properties of the functional space, i.e., if $t = f + g \otimes h$ and $t' = i \otimes i - h \otimes f \otimes j$, then $t + t' = f + g \otimes h + i \otimes i - h \otimes f \otimes j$ and $t \otimes t' = f \otimes i \otimes i - f \otimes h \otimes f \otimes j + g \otimes h \otimes i \otimes i - g \otimes h \otimes h \otimes f \otimes j$. It is worth noting that $T^p(\mathcal{L}_2(X))$ is isomorphic to $\mathcal{L}_2(X^p)$.

The set of symmetric and anti-symmetric tensors are respective subalgebras $S(V)$ and $A(V)$ of $T(V)$. We can furthermore define a multiplication $\wedge$ on $S(V)/A(V)$ similar to $\otimes$ by applying their respective projectors $\hat{A}/\hat{S}$ to the tensor product $\otimes$, i.e., $t \wedge t' = \hat{A}\,(t \otimes t')$. The projectors are constructed from the projections in the generating spaces. Given $T^p(V)$ there is a group Γ_p of permutations on $\{1 \ldots p\}$. With each permutation $\sigma \in \Gamma_p$ on the indices, we identify an operator $\hat{P}_\sigma$ on $T^p(V)$ such that $\hat{P}_\sigma f_1 \otimes \cdots \otimes f_p = f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(p)}$. Then $\hat{A}_p = \frac{1}{|\Gamma_p|} \sum_{\sigma \in \Gamma_p} \operatorname{sgn}(\sigma) \hat{P}_\sigma$ and $\hat{S}_p = \frac{1}{|\Gamma_p|} \sum_{\sigma \in \Gamma_p} \hat{P}_\sigma$, where sgn is -1 if σ has even order and 1 if σ has odd order. These operators can be extended to $T(V)$ by the identity on all components outside $T^p(V)$ and so $\hat{A} = \sum_p \hat{A}_p$ and $\hat{S} = \sum_p \hat{S}_p$.

With this tensor algebra defined it is easy to define the so-called annihilation and creation operators of second quantization. If $f \in V$, then we will define the creation operator $\hat{a}_f^\dagger : A(V) \to A(V)$ such that $t \in A^i(V) \mapsto \sqrt{i+1} \cdot f \wedge t$ and the annihilation operator $\hat{a}_f : T(V) \to T(V)$ restricted to $A(V)$ such that $t = v \otimes t' \mapsto \sqrt{i} \cdot f^*(t_1)t'$, where $v \in V$, $t' \in T(V)$, and $f^* \in V^* : f^*(x) = \langle f|x \rangle \, \forall x \in V$. The space $A(V)$ is also commonly known as Fock space; the definition for $S(V)$ is analogous. $\hat{a}_f^\dagger$ and $\hat{a}_f$ are adjoints to one another, i.e., $\langle f \left| \hat{a}_g \right| h \rangle = \left\langle h \left| \hat{a}_g^\dagger \right| f \right\rangle^*$. In the following I will use $\hat{O}$ for operators on $T(V)/A(V)/S(V)$ and $\mathbf{O}$ for operators on $\mathcal{L}_2(X^p)$ or more generally $\mathcal{L}_2^\infty(X) = \bigoplus_{p=0}^\infty \mathcal{L}_2(X^p)$. The annihilation and creation operators follow commutation relations as shown in equations (5) to (7) on the following page. Of particular note is that $\hat{a}_f^\dagger \hat{a}_f$ is a projection operator on the space of symmetric/anti-symmetric tensors containing f in each product.

$$\hat{a}_p^\dagger \hat{a}_q^\dagger = (\delta_{pq} - 1)\hat{a}_q^\dagger \hat{a}_p^\dagger \Leftrightarrow \{\hat{a}_p^\dagger, \hat{a}_q^\dagger\} = \delta_{pq}\hat{a}_q^\dagger \hat{a}_p^\dagger = 0 \tag{5}$$

$$\hat{a}_p^\dagger \hat{a}_q = \delta_{pq} - \hat{a}_q \hat{a}_p^\dagger \Leftrightarrow \{\hat{a}_p^\dagger, \hat{a}_q\} = \delta_{pq} \tag{6}$$

$$\hat{a}_p \hat{a}_q = (\delta_{pq} - 1)\hat{a}_q \hat{a}_p \Leftrightarrow \{\hat{a}_p, \hat{a}_q\} = \delta_{pq}\hat{a}_q \hat{a}_p = 0 \tag{7}$$

With these relations it is possible to express operators $\hat{O}$ on $T(V)$ as a sum of products of creation operators followed by a product of annihilation operators. So it is possible to describe the (Born-Oppenheimer) Hamiltonian operator in a simple manner as in equation (8).

$$\hat{H} = \sum_{p,q \in B(V)} \langle p|\mathbf{h}_1 q\rangle \, \hat{a}_p^\dagger \hat{a}_q + \frac{1}{2} \sum_{p,q,r,s \in B(V)} \left\langle p \otimes q \middle| \frac{1}{r_{12}} \cdot r \otimes s \right\rangle \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s \tag{8}$$

$$\mathbf{h}_1 = \left(\mathbf{p}_1 - \frac{q_e}{c}\mathbf{A}_1\right)^2 + \sum_{\text{nuclei } I} \frac{q_e q_I}{r_{1I}} \tag{9}$$

$B(V)$ denotes an orthonormal basis of V, and q_e is the charge of an electron. The simplicity immediately catches the eye as this formulation is the Hamiltonian for a (Born-Oppenheimer) system independent of electron count. The number of electrons has been instead buried in the wave function to which the Hamiltonian is applied. Finally, any operator $\mathbf{O}$ on $\mathcal{L}_2^\infty(X^p)$ can be identified with an operator $\hat{O}$ on $T(V)$ and thereby on $A(V)/S(V)$ by the isomorphism $\imath$ from $\mathcal{L}_2^\infty(X)$ to $T(V)$ by $\hat{O} = \imath \circ \mathbf{O} \circ \imath^{-1}$. I will use these different expressions of essentially the same operator depending on which property needs highlighting or which is simpler to describe explicitly.

2. Conventional Computational Methods

Despite the elegance of second quantization, solving the Schrödinger equation remains analytically impossible for anything but the simplest cases, like the hydrogen atom. Therefore, limited by the computing power available during each era, many different techniques have been developed to provide high accuracy at relatively low cost.

2.1. Basis Sets

As mentioned earlier, the solution of the Schrödinger equation resides in the Hilbert space of Lebesgue-square-integrable functions $\mathcal{L}_2$, which is unfortunate since $\mathcal{L}_2$ is infinitely dimensional and, therefore, intractable. Hence, actual computations have to rely on truncations of $\mathcal{L}_2$ to finite dimensions and much research has gone into the development and discovery of good finite basis sets, which span the finite space. As a basis set is expanded, the quality of the approximation cannot decrease, as the space spanned by the smaller basis set is contained within the space spanned by the larger basis set. Therefore, series of basis sets have been developed for which the solution is systematically improved.

The first basis sets were comprised of Slater-type orbitals[12] (STOs, see equation (10)), named after Bohr's interpretation of electrons on orbits around the nuclei. These are based on the solutions to the hydrogen problem, which can be solved analytically (in spherical coordinates see equation (11)).

$$f(r,\theta,\phi) = r^{n-1} \cdot \mathrm{e}^{-\zeta r} \cdot Y_l^m(\theta,\phi) \tag{10}$$

$$f_{n,l,m}(r,\theta,\phi) = R_{nl}(r) \cdot Y_l^m(\theta,\phi) \tag{11}$$

$$R_{nl}(r) = \left(\frac{Zr}{a_0 n}\right)^l \mathrm{e}^{-\frac{Zr}{2a_0 n}} P_{n-l-1}\left(\frac{Zr}{a_0 n}\right) \tag{12}$$

$$Y_l^m(\theta, \phi) = \mathrm{e}^{im\phi} L_l^m(\cos\theta) \frac{1}{2\pi} \tag{13}$$

Here, P_ν are the Laguerre polynomials; and Y_l^m represent the angular momentum, and L_l^m are the Legendre polynomials.

To the detriment of the quantum community, the integrals using either one of these functions are extremely complicated and require much computational time. So, quantum theorists have looked to Gaussian type orbitals[13] (GTOs, see equation (14)). The necessary integrals are so simple that using 4 or 5 GTOs for each STO is no computational disadvantage. Hence, the idea of bundling functions as linear combinations of GTOs with appropriate coefficients and exponents into contractions emerged. Using a contraction reduces the dimension of the generated Hilbert space while retaining the simplicity of Gaussian functions and an accurate description.

$$f = x^l \cdot y^m \cdot z^n \cdot \mathrm{e}^{-\zeta\left(x^2+y^2+z^2\right)} \tag{14}$$

The sum $L = l + m + n$ is interpreted as angular momentum, so functions with $L = 0$ can be called s-functions as in the hydrogen case.

Generally, each atom in a given chemical problem is assigned a set of functions, possibly contractions. Often the contractions are fitted to experimental data or STOs. The union of these sets span the finite-dimensional Hilbert space. Not all basis sets perform equally well for different problems, which has resulted in an abundance of different basis sets and a need for consistent nomenclature.

One common nomenclature for basis sets describes them according to the number of contractions, e.g., DZ (double zeta) for a basis set with 2 GTOs in the contraction. Since the core electrons, which are close to the nucleus, need fewer functions than the valence electrons, more functions are assigned to the valence electrons, which is denoted with SV or V for "split valence." The prefix of "p" denotes that additional polarization functions have been added, usually higher-angular-momentum functions. A further prefix of "cc-" stands for correlation-consistent basis sets. These basis sets incorporate effects of explicitly correlating electrons instead of mere average effects, as we will discuss later.[14,15] Therefore, the basis set "cc-pVQZ" is a correlation consistent basis set with polarization functions, a split valence shell, and 4 functions per contraction (quadruple zeta).

2.2. Hartree-Fock Formalism

To this day the Hartree-Fock formalism (HF) is one of the most important quantum methods; be it in its own right or as a point of departure for higher-level calculations.[1] At the heart of HF is the iterative improvement of a guess at the solution of the Schrödinger equation using a single determinant as introduced by Slater[2] (see equation (15)). Using this guess a very simple expression can be derived for the energy of such an N-electron function (see equation (16) on the following page excluding the internuclear contributions).

$$|\chi_1 \cdots \chi_N| = \frac{1}{N!} \cdot \begin{vmatrix} \chi_1(\vec{x}_1) & \cdots & \chi_N(\vec{x}_1) \\ \vdots & \ddots & \vdots \\ \chi_1(\vec{x}_N) & \cdots & \chi_N(\vec{x}_N) \end{vmatrix} \tag{15}$$

$$E_{HF} = \big\langle |\chi_1 \cdots \chi_N| \big| \mathbf{H} |\chi_1 \cdots \chi_N| \big\rangle = \sum_{i=1}^{N} \big\langle \chi_i \big| \mathbf{h}_1 \chi_i \big\rangle + \sum_{i>j} \left\langle |\chi_i \chi_j| \left| \frac{1}{r_{12}} |\chi_i \chi_j| \right. \right\rangle \tag{16}$$

$$\mathbf{h}_1 = \frac{1}{2m_1} \mathbf{p}_1^2 + \sum_{\text{nuclei } A} \frac{q_A}{R_{1A}} \tag{17}$$

Here χ_i represent spin-functions.

Variation of the one-electron functions χ_i to first order leads to an eigenvalue problem of the Fock operator $\mathbf{f}_1$, which renders a new set of one-electrons functions for the next iteration. The HF equations (18) are often interpreted to subject an electron to the average field of electrons $N - 1$ without explicit correlation. The solutions to equation (18) are ordered with respect to their eigenvalues from lowest to highest. Then they are reinserted into the equations to generate a new set of eigenfunctions until convergence of the energy has been achieved. As can be seen in equation (19), the eigenvalues do not just sum up to give the HF energy because the inter-electronic contributions are involved in each eigenvalue. Thus they are counted twice and must be subtracted.

(18)

$$\mathbf{f}_1 \chi_j(\vec{x}_1) = \mathbf{h}_1 \chi_j(\vec{x}_1) + \sum_{i=1}^{N} \int \frac{|\chi_i(\vec{x}_2)|^2}{r_{12}} d\vec{x}_2 \chi_j(\vec{x}_1) - \sum_{i=1}^{N} \int \frac{\chi_i(\vec{x}_2)^* \chi_j(\vec{x}_2)}{r_{12}} d\vec{x}_2 \chi_i(\vec{x}_1) = \epsilon_j \chi_j(\vec{x}_1)$$

$$E_{HF} = \sum_{i=1}^{N} \epsilon_j - \sum_{i,j:1 \to N} \int_{-\infty}^{\infty} \frac{|\chi_i(\vec{x}_1) \chi_j(\vec{x}_2)|^2}{r_{12}} d\vec{x}_1 d\vec{x}_2 \tag{19}$$

The above equations can be subjected to various constraints and frequently are. A very common restriction is that each spatial function is assigned an α- as well as a β-spin-function. This is called restricted HF (RHF); merely requiring that there be equal numbers of α- and

β-electrons is restricted open-shell HF (ROHF); no constraints is appropriately referred to as unrestricted HF (UHF).

2.3. Configuration Interaction

Building on HF theory, configuration interaction (CI) deals in expanding the trial space of functions from the subset of Slater determinants to linear subspaces.[3,4,16] CI derives its name from the fact that it incorporates the interactions between multiple Slater determinants. Using the language of second quantization, it is possible to describe a systematically increasing series of subspaces.

$$\hat{C}_0 = c_0 \tag{20}$$

$$\hat{C}_1 = \sum_{a \in V, i \in V} c_i^a \cdot \hat{a}_a^\dagger \hat{a}_i \tag{21}$$

$$\hat{C}_2 = \sum_{a > b \in V, i > j \in V} c_{ab}^{ij} \cdot \hat{a}_a^\dagger \hat{a}_b^\dagger \hat{a}_i \hat{a}_j \tag{22}$$

$$\vdots$$

$$\hat{C}_m = \sum_{\substack{a_1,\ldots,a_m \in V, \\ i_1,\ldots,i_m \in V}} c_{a_1,\ldots,a_m}^{i_1,\ldots,i_m} \cdot \hat{a}_{a_1}^\dagger \cdots \hat{a}_{a_m}^\dagger \hat{a}_{i_1} \cdots \hat{a}_{i_m} \tag{23}$$

$$|\Psi\rangle = \sum_{m=0}^{n} \hat{C}_m \, |\Phi_0\rangle \tag{24}$$

Here, a, b, and a_k denote functions from the basis from V which are not occupied (used) in the HF reference Φ_0, while i, j, i_k are occupied in the HF reference. In this form it is obvious that all configuration operators with $m > N$, the number of electrons, just map Φ_0 to 0. The order m of the CI operator is called the excitation level, referring to the fact that occupied spin-functions are being replaced by unoccupied spin-functions which have greater HF expectation values and lead to higher energies for the corresponding Slater determinant. The energy of such a function $E_{CI}^{n} = \left\langle \Psi \middle| \hat{H} \middle| \Psi \right\rangle$, with maximum excitation n, can be minimized with respect to the coefficients. Variation of the coefficients is equivalent to solving the eigenvalue problem for $\hat{H}$ restricted to the subspace of $T^{N}(V)$ spanned by the set $\{\hat{a}_{a_1}^{\dagger} \cdots \hat{a}_{a_m}^{\dagger} \hat{a}_{i_1} \cdots \hat{a}_{i_m} \left|\Phi_0\right\rangle | m \leq n\}$. The degree of excitation is usually denoted by appending the excitation levels by their initial, i.e., CIS for single excitations, CISD for single and double excitations, etc. When $n = N$, all of $T^{N}(V)$ is spanned and, hence, the method is called full CI (FCI). Since HF is considered to be an averaged electron-interaction model, the energy difference between FCI and HF must be due to correlation of the electrons and is therefore called electron correlation (EC) energy. Compared to HF, the computational cost rises considerably when adding higher excitation levels. For order m the resultant dimensionality of the eigenvalue problem is augmented by $\binom{\dim V - N}{m} \cdot \binom{N}{m}$. In the case of FCI, the dimensionality is $\binom{\dim V}{N}$, which is prohibitive in anything but the simplest cases. The ultimate goal of all methods is to reproduce the FCI properties without the associated FCI computational cost.

2.4. Perturbation Theory

Another way to increase accuracy starting from the HF solution, is perturbation theory (PT). In Rayleigh-Schrödinger PT,[17] we start from an operator $\hat{H}_0$ for which the solution is known and slowly vary a potential $\hat{W}$ which results in the final operator. We will only consider the nondegenerate case, as it exemplifies the general method, while degeneracy adds nothing but complexity.

$$\hat{H} = \hat{H}_0 + \lambda \hat{W}, \lambda \in [0,1] \tag{25}$$

$$E_n = \sum_m \lambda^m E_n^{(m)} \tag{26}$$

$$|n\rangle = \sum_m \lambda^m \left|n^{(m)}\right\rangle \tag{27}$$

Let $|n\rangle$ be the nth eigensolution of $\hat{H}$ with corresponding eigenvalue E_n. Furthermore, let $\left|n^{(0)}\right\rangle$ is the solution to the unperturbed problem in $\hat{H}_0$. The unknowns $|n\rangle$ and E_n can be expanded as polynomials of the perturbation variable λ', which continuously turns the perturbation on or off. Using that $\left\langle n \middle| \hat{H} \middle| n \right\rangle = E_n$, we can arrange a new polynomial in λ as in equation (30).

$$\hat{H}_0 + \lambda \hat{W} |n\rangle = E_n |n\rangle \tag{28}$$

$$\Leftrightarrow \sum_{m=1}^{\infty} \lambda^m \left(\hat{H}_0 \left|n^{(m)}\right\rangle + \hat{W} \left|n^{(m-1)}\right\rangle \right) = \sum_{m=1}^{\infty} \lambda^m \sum_{i+j=m} E_n^{(i)} \left|n^{(j)}\right\rangle \tag{29}$$

$$\Leftrightarrow \sum_{m=1}^{\infty} \lambda^m \left(\hat{H}_0 \left|n^{(m)}\right\rangle + \hat{W} \left|n^{(m-1)}\right\rangle - \sum_{i+j=m} E_n^{(i)} \left|n^{(j)}\right\rangle \right) = 0 \tag{30}$$

The polynomial in λ of equation (30) is zero for all λ, which implies that all coefficients must be zero as well. These conditions allow rearrangement for solutions of $\left|n^{(m)}\right\rangle$ and $E_n^{(m)}$ in an iterative manner by inverting $\hat{H}_0 - E_n^{(0)}$ on the space generated by the zero-order solutions excepting state n. By projecting both sides onto $\left|n^{(0)}\right\rangle$ in equation (32), an expression for $E_n^{(m)}$ is derived which, upon reinsertion, eliminates any components of $\left|n^{(0)}\right\rangle$ from (32). It is therefore permissible to invert $\hat{H}_0 - E_n^{(0)}$.

$$\hat{H}_0 \left|n^{(m)}\right\rangle + \hat{W} \left|n^{(m-1)}\right\rangle - \sum_{i \leq m} E_n^{(m-i)} \left|n^{(i)}\right\rangle = 0 \tag{31}$$

$$\Leftrightarrow \left(\hat{H}_0 - E_n^{(0)}\right) \left|n^{(m)}\right\rangle = \sum_{i<m} E_n^{(m-i)} \left|n^{(i)}\right\rangle - \hat{W} \left|n^{(m-1)}\right\rangle \tag{32}$$

$$E_n^{(m)} = \left\langle n^{(0)} \middle| \hat{W} \middle| n^{(m-1)} \right\rangle \tag{33}$$

$$\left|n^{(m)}\right\rangle = \left(\hat{H}_0 - E_n^{(0)}\right)^{-1} \left(\sum_{i<m} E_n^{(m-i)} \left|n^{(i)}\right\rangle - \hat{W} \left|n^{(m-1)}\right\rangle\right) \tag{34}$$

Since $\hat{H}_0 - E_n^{(0)}$ essentially operates on the FCI space, the inversion will take exponential computational time if no simplifications can be introduced. For HF, this is indeed the case. The unperturbed Hamiltonian $\hat{H}_0$ for this purpose is $\hat{f}_1$ and $\hat{W} = \hat{h}_1 - \hat{f}_1 + \frac{1}{2} \sum_{p,q,r,s \in B(V)} \left\langle p \otimes q \middle| \frac{1}{r_{12}} \cdot r \otimes s \right\rangle \hat{a}_p^\dagger \hat{a}_q^\dagger \hat{a}_r \hat{a}_s$, to which the HF energies are actually the first order perturbations.[a] The zeroth order solutions in this case are sums of eigenvalues of $\mathbf{f}_1$ ($\sum_{i=1}^n \epsilon_i$); as noted earlier, these are different from the actual HF energy. PT on HF with this partitioning of the Hamiltonian is named after Møller and Plesset[2,9,17,18] and dubbed MP(m) for the solution of mth order. $\hat{W}$ excites states $\left|n^{(m-1)}\right\rangle$ by no more than two, on the other hand $\left(\hat{H}_0 - E_n^{(0)}\right)^{-1}$ excites by at least one. Therefore starting from the HF function, $\left|n^{(m)}\right\rangle$ cannot be comprised of excitations higher than m.

A review of PT reveals the continuous mapping of corresponding eigenvalues, i.e., the third HF state is mapped to the third eigenstate of the full problem, which makes it safe

---

[a] see section 2.2.

for computing excited states. Furthermore, a single set of equations allows the computation of any solution for a given value of λ. However, PT does not always converge and often oscillates from order to order.[17]

2.5. Coupled-Cluster Theory

In the late 1950s, physicists Coester and Kümmel developed coupled-cluster theory (CC) for the treatment of nuclear matter.[19,20] Only six years later Ĉíẑek extended CC to the electron correlation problem, and it has enjoyed intensive research ever since.[21–25]

CC utilizes a very simple *ansatz* of an exponential operator. The degree of the cluster operator $\hat{T}$ is usually truncated to avoid the complexity of FCI, which in turn truncates the exponential expansions as terms go to zero. As with CI, appending the degrees of excitation, as in singles or double with 'S' or 'D', respectively, clarifies in which approximation the theory was applied.

$$|\Psi\rangle = \mathrm{e}^{\hat{T}}\,|\Psi_0\rangle \tag{35}$$

$$\mathrm{e}^{\hat{T}} = \sum_{k=0}^{\infty} \frac{1}{k!}\hat{T}^k \tag{36}$$

$$\hat{T} = \sum_{k=1}^{\infty} \hat{T}_k = \sum_{k=1}^{\infty} \sum_{\substack{\{i_1,\dots,i_k\}\subset B(V)\\ \{j_1,\dots,j_k\}\subset B(V)}} t^{j_1\cdots j_k}_{i_1\dots i_k}\hat{a}^{\dagger}_{i_1}\cdots\hat{a}^{\dagger}_{i_k}\hat{a}_{j_1}\cdots\hat{a}_{j_k} \tag{37}$$

The great advantage of CC is the fact that with relatively few degrees of freedom a large number of determinants are reached. So, for instance in CCS, double and triple excitations are incorporated as $\hat{T}_1^2$ and $\hat{T}_1^3$, respectively. Each order k of the cluster operator adds $\binom{\dim V}{k}^2$ degrees of freedom. Starting from a single Slater determinant Φ_0 as a reference, the problem is simplified by the fact that only occupied orbitals need to be considered for annihilation, which leaves $\binom{\dim V-N}{k}\cdot\binom{N}{k}$ degrees of freedom. As with PT, when order N is reached, CC reduces to FCI. In the following, I will restrict myself to the case of a single determinant reference.

$$E_{CC} = \left\langle \Phi_0 \middle| \hat{H}\mathrm{e}^{\hat{T}} \middle| \phi_0 \right\rangle = \left\langle \Phi_0 \middle| \hat{H}\left(1+\hat{T}_1+\hat{T}_2+\frac{1}{2}\hat{T}_1^2\right) \middle| \Phi_0 \right\rangle \tag{38}$$

$$\hat{\hat{H}} = \mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}} = \sum_n^{\infty}\frac{1}{n!}[\hat{H},\hat{T}]_n = \sum_n^{4}\frac{1}{n!}[\hat{H},\hat{T}]_n \tag{39}$$

$$[\hat{H},\hat{T}]_0 = \hat{H},\;\; [\hat{H},\hat{T}]_{n+1} = [[\hat{H},\hat{T}]_n,\hat{T}] \tag{40}$$

$$E_{CC} = \left\langle \Phi_0 \middle| \hat{\hat{H}} \middle| \Phi_0 \right\rangle \tag{41}$$

There are two equivalent ways of describing the energy of the CC system. The first as described in equation (38) is justified because if $\mathrm{e}^{\hat{T}}\left|\Phi_0\right\rangle$ is a solution to the Hamiltonian, then $\left\langle \Phi_0 \middle| \hat{H}\mathrm{e}^{\hat{T}} \middle| \Phi_0 \right\rangle = E\cdot\left\langle \Phi_0 \middle| \mathrm{e}^{\hat{T}} \middle| \Phi_0 \right\rangle = E$. This avoids the necessity of normalizing $\mathrm{e}^{\hat{T}}\left|\Phi_0\right\rangle$ and is called intermediate normalization. By a similar argument, the second formulation (equation (41)) utilizes the Hausdorff expansion (equation (39)) to simplify the infinite sum. Despite the fact that in general $\mathrm{e}^{-\hat{T}} \neq \mathrm{e}^{\hat{T}^*}$, it is admissible to apply $\mathrm{e}^{-\hat{T}}$ since $E\left\langle \Phi_0 \middle| \mathrm{e}^{-\hat{T}}\mathrm{e}^{\hat{T}} \middle| \Phi_0 \right\rangle = E$ if $E\cdot\mathrm{e}^{\hat{T}}\left|\Phi_0\right\rangle = \hat{H}\mathrm{e}^{\hat{T}}\left|\Phi_0\right\rangle$. In order to determine the cluster amplitudes $t^{j_1,\cdots}_{i_1,\ldots}$, it suffices to notice the validity of equation (42).

$$\hat{\hat{H}}\left|\Phi_0\right\rangle = E_{CC}\left|\Phi_0\right\rangle \perp \hat{a}^{\dagger}_{i_1}\cdots\hat{a}^{\dagger}_{i_k}\hat{a}_{j_1}\cdots\hat{a}_{j_k}\left|\Phi_0\right\rangle \forall\{i_1,\ldots,i_k,j_1,\ldots,j_k\}\subset B(V) \tag{42}$$

The unfortunate side effect of intermediate normalization is that CC is no longer variational, i.e., there is no guarantee that the CC energy is an upper bound to the ground state energy. On the other hand CC retains size-consistency as well as size-extensivity found in FCI. A computational method is size-consistent if for any two non-interacting systems A and B the energy solution is merely the sum of each contribution ($E_{AB} = E_A + E_B$). Size-extensivity describes the fact that the energy of a method scales linearly with the number of electrons.

2.6. Density Functional Theory

Hohenberg and Kohn took an entirely different approach to solving for the ground state of a given Hamiltonian.[26] Instead of focusing on wave functions, they proved that there exists a functional which maps an electron density to an energy, for which the ground state density is the global minimum with the corresponding ground state energy; hence the name density functional theory. Despite the mathematical simplicity, this functional is not known. Through the years, many attempts with varying degrees of success have been made to develop functionals which display the required properties. For conceptual convenience, the energy is usually divided into different parts.

$$E = E_T + E_V + E_J + E_{XC} \tag{43}$$

- E_T is the kinetic contribution to the energy.
- E_V is the nucleus-electron contribution to the energy (nuclear Coulomb term).
- E_J is the uncorrelated electron-electron Coulomb interaction.
- $E_{XC} = E_X + E_C$ is the exchange-correlation term which incorporates corrections for the fermion nature of electrons.

The terms E_T, E_V, and E_J were quickly approximated reasonably well. In equations (44) to (46) and for the remainder of this chapter, $\rho(\vec{r})$ denotes the electron density and R_a denote the positions of the nuclei.

$$E_T = \frac{3}{10}\left(6\pi^2\right)^{\frac{2}{3}} \int \rho_\alpha^{\frac{5}{3}}(\vec{r})dr + \frac{1}{8}\int \rho_\alpha^{-1}(\vec{r})\left|\nabla\rho_\alpha(\vec{r})\right|^2 dr \tag{44}$$

$$E_V = -\sum_a^m q_a \int \frac{\rho(\vec{r})}{|r - R_a|}dr \tag{45}$$

$$E_J = \frac{1}{2}\int\int \frac{\rho(\vec{r}_1)\rho(\vec{r}_2)}{|\vec{r}_1 - \vec{r}_2|}d\vec{r}_1 d\vec{r}_2 \tag{46}$$

An example of an exchange functional (E_X^D) was developed by Dirac,[27] in which the electrons are viewed as a gas. A vastly popular exchange functional (E_X^B) was introduced by Becke[28] in 1988, building on Dirac's success. It incorporates a semi-empirical factor $b = 0.0042$ exemplifying how difficult *a priori* development of density functionals is. An early simple correlation functional (E_C^W), which is also parameterized, was reported by Wigner[29] in 1938.

$$E_X^D = -\frac{3}{2}\left(\frac{3}{4\pi}\right)^{\frac{3}{4}} \int \rho^{\frac{4}{3}}(\vec{r})dr \tag{47}$$

$$E_X^B = E_X^D - b\int \rho_\alpha^{4/3}\frac{\left(\frac{|\nabla\rho_\alpha|}{\rho_\alpha^{4/3}}\right)^2}{1 + 6\beta\frac{|\nabla\rho_\alpha|}{\rho_\alpha^{4/3}}\sinh^{-1}\frac{|\nabla\rho_\alpha|}{\rho_\alpha^{4/3}}}d\vec{r} \tag{48}$$

$$E_C^W = -4a\int \frac{\rho_\alpha(\vec{r})\rho_\beta(\vec{r})}{\rho(\vec{r})}\cdot\frac{1}{1 + d\rho^{-\frac{1}{3}}(\vec{r})}\,d\vec{r} \tag{49}$$

DFT methods like B3LYP[28,30] enjoy considerable popularity among organic chemists despite some spectacular failures,[31,32] e.g., acetylene, and the lack of systematic improvement as is highlighted by the unintuitive and complex nature of Becke's exchange functional. DFT is also routinely employed in condensed matter physics.

3. Reduced Density Matrices and Wave Functions

In order for geminal functional theory to reduce the complexity of solving the Schrödinger equation, it is necessary to find a simpler expression of the energy. Although $\mathbf{H}$ is comprised of two-particle operators only, it remains an n-electron operator. Therefore, reducing the Hamiltonian to a two-particle operator is a key to simplifying the energy expression. Given a wave function, the energy can be expressed using a two-particle operator as follows:

$$\begin{aligned} E =& \int \psi(1,\ldots,n)\mathbf{H}\psi(1,\ldots,n)d1\cdots dn \\ =&\sum_i \int \psi(1,\ldots,n)\mathbf{h}_i\psi(1,\ldots,n)d1\cdots dn+ \\ & \sum_{i,j} \int \psi(1,\ldots,n)\frac{1}{r_{ij}}\psi(1,\ldots,n)d1\cdots dn \qquad (50) \\ =&n\int \psi(1,\ldots,n)\mathbf{h}_1\psi(1,\ldots,n)d1\cdots dn+ \\ & \frac{n(n-1)}{2}\int \psi(1,\ldots,n)\frac{1}{r_{12}}\cdot\psi(1,\ldots,n)d1\cdots dn \qquad (51) \end{aligned}$$

In equations (50) and (51), we have merely transposed the electrons such that electrons i and j are in the positions of electrons 1 and 2. Since both wave functions undergo transposition, the sign changes cancel. All one-electron interactions of electron i are contained in $\mathbf{h}_i$ (see equation (17) on page 11), and all integrals involving $\mathbf{h}_i$ are the same; as are the two-electron integrals. This realization leads to equation (51).

Equation (51) on the preceding page suggests the following definitions and leads to the simpler, yet equivalent equation (54) for the energy. The goal of geminal functional theory is to exploit these equations.

$$D^{(m)} = \int \psi(1', \ldots, m', m+1, \ldots n) \cdot \psi(1, \ldots, m, m+1, \ldots n)^* \cdot d(m+1) \cdots dn \tag{52}$$

$$\mathbf{K} = \frac{n}{2}(\mathbf{h}_1 + \mathbf{h}_2) + \frac{n(n-1)}{2}\frac{1}{r_{12}} \tag{53}$$

$$E = \operatorname{tr}\left(\mathbf{D}^{(2)}\mathbf{K}\right) \tag{54}$$

The energy and wave functions are intricately connected to reduced density matrices. Equation (52) defines the integration kernel for the reduced density matrix of mth order (m-RDM) for an n-electron wave function. In the following, plain O denotes the integration kernel of a linear operator $\mathbf{O}$ such that

$$\mathbf{O}f \mapsto \int O(x_1 \ldots x_n, x'_1 x'_n) f(x'_1 \ldots x'_n \ldots x_N) dx'_1 \ldots x'_n.$$

As equation (54) highlights, reduced density matrices are the second part in simplifying the energy expression for geminal functional theory. Depending on the intended emphasis, there are two normalization conventions:

a) $\operatorname{tr}\left(\mathbf{D}^{(m)}\right) = 1$,

b) $\operatorname{tr}\left(\mathbf{D}^{(m)}\right) = \frac{n!}{m!}$.

To avoid confusion, the reduced density matrix will be denoted as ρ_m for the latter case.

While the full wave function space scales with $\binom{k}{n}$, the reduced density matrix scales with $\binom{k}{m}$ – a handsome reduction in complexity. If all matrices corresponded to some wave function of N electrons, finding the ground state and its energy would be trivial indeed; but, they are not. To compound matters, it is not true that an RDM for N electrons is also an RDM for some wave function of N' electrons, where $N \neq N'$. If there exists a wave function such that a given matrix is that function's RDM, that matrix is called N-representable and the wave function is one of its N-representations. Direct verification of the N-representability of a matrix is currently intractable for anything but the simplest cases. Instead, necessary and sufficient conditions are sought that are easily verifiable. Positive-semi-definiteness is one such necessary condition. By construction, RDMs are positive, semi-definite operators, i.e., for any x, $\left\langle x \middle| \hat{D}^{(m)} \middle| x \right\rangle \geq 0$. Using second quantization, we find that $D^{(m)}_{i_1 \ldots i_m, j_1 \ldots j_m} := \left\langle i_1 \ldots i_m \middle| \hat{D}^{(m)} \middle| j_1 \ldots j_m \right\rangle = \left\langle \Psi \middle| \hat{a}^\dagger_{i_1} \cdots \hat{a}^\dagger_{i_m} \hat{a}_{j_1} \cdots \hat{a}_{j_m} \middle| \Psi \right\rangle$.

$$Q^{(2)}_{ij,ab} = \left(\langle \Psi | \, \hat{a}^\dagger_i \hat{a}_j \right) \cdot \left(\hat{a}^\dagger_a \hat{a}_b \, | \Psi \rangle \right) = \delta_{ja} D^{(1)}_{i,b} - D^{(2)}_{ia,jb} \tag{55}$$

$$\begin{cases} P^{(2)}_{ij,ab} = \left(\langle \Psi | \, \hat{a}_i \hat{a}_j \right) \cdot \left(\hat{a}^\dagger_a \hat{a}^\dagger_b \, | \Psi \rangle \right) \\ \qquad = \delta_{jb}\delta_{ia} - \delta_{ib}\delta_{ja} + \delta_{ja}\hat{D}^{(1)}_{i,b} - \delta_{ia}\hat{D}^{(1)}_{j,b} - \delta_{jb}\hat{D}^{(1)}_{i,a} + \delta_{ib}\hat{D}^{(1)}_{j,a} + \hat{D}^{(2)}_{ab,ij} \end{cases} \tag{56}$$

$$\hat{D}^{(2)} \geq 0, \;\; \hat{Q}^{(2)} \geq 0, \;\; \hat{P}^{(2)} \geq 0 \tag{57}$$

It is immediately evident that we will arrive at another positive, semi-definite matrix via permuting the creation and annihilation operators such that $\hat{a}_{j_k}$ is only replaced by $\hat{a}^\dagger_{i_{m+1-k}}$ and vice versa. These operators can be related to the m-RDM using the commutation relationships of creation and annihilation operators outlined in equations (5) to (7) on page 7. For the 2-RDM, there are only two distinct variations (equations (55) and (56)).

The RDMs play a larger role than just defining the energy, since it turns out that a wave function can also be described by the eigenfunction of its reduced density matrix:[33,34]

$$\mathbf{D}_{\psi}^{(p)} d_i^{(p)} = \lambda_i d_i^{(p)} \tag{58}$$

$$\mathbf{D}_{\psi}^{(p)} d_i^{(q)} = \lambda_i d_i^{(q)} \tag{59}$$

$$\psi = \sum c_i d_i^{(p)}(1,\ldots,p) d_i^{(q)}(p+1,\ldots,p+q) \tag{60}$$

where $p+q$ are the number of electrons and $\lambda_i = |c_i|^2$. These functions are called natural p/q-functions. In particular, natural functions appear in pairs, which for two-electron functions (i.e. geminals) leads to the simple form:

$$g(1,2) = \sum \xi_i \left| d_{2i-1} d_{2i} \right| \tag{61}$$

In general the transposition of a pair of electrons leads to

$$\begin{aligned}\psi(1,\ldots,n) &= \sum_i \xi_i g_i(1,2) d_i^{(n-2)}(3,\ldots,n) \\ &= \sum_i \xi_i g_i(3,4) d_i^{(n-2)}(1,2,5,\ldots,n).\end{aligned} \tag{62}$$

where $\xi_i^2 = \lambda_i$. Therefore, any wave function $(n > 1)$ can be expressed in terms of its natural geminals. Furthermore, if p is odd and $p < q$, then

$$\int d_i^{(p)}(p+1\ldots 2p) d_i^{(q)}(p+1\ldots p+q) dp \ldots d(2p) = 0 \tag{63}$$

There is no equivalent for even p.[35]

4. Anti-Symmetrized Geminal Products (AGP)

In 1953, geminals were first introduced as primary variables in quantum chemistry via strongly orthogonal anti-symmetric geminal powers (SOAGP).[36–38] A strongly orthogonal geminal is such that

$$\int g_1(1,2)g_2(2,3)d2 = 0. \tag{64}$$

while the weaker condition of orthogonality requires $\int g_1(1,2)g_2(1,2)d1d2 = 0$. Here it is convenient to expand the tensor definition. We can expand our notion of $T^p(V)$ to a space $T^p_q(V) = \bigotimes_{i=1}^p V \otimes \bigotimes_{i=1}^q V^*$. A Hilbert space V is isomorphic to its dual space V^* ($V \cong V^*$) via $f \in V \mapsto f^* : f^*(x) = \langle f|x\rangle \forall x \in V$. Therefore, any tensor in $T^p_q(V) \cong T^{p+q}(V)$ is also a linear operator generated by $\hat{t}^p_q = \bigotimes_{i=1}^p a_i \otimes \bigotimes_{i=1}^q b_i^*$ from $T^q(V)$ to $T^p(V)$ by $\bigotimes_{i=1}^q v_i \mapsto \prod_{i=1}^q b^*_{q-i}(v_i) \bigotimes_{i=1}^p a_i$. Then the strong orthonormality condition reduces to a matrix equation for linear operators on $T^1(V)$ (see section 1.2).

$$\hat{g}_i\hat{g}_j^* = 0 \tag{65}$$

Equation (65) implies that the image spaces of g_i and g_j are disjoint, i.e. $\mathrm{Im}(g_i)\cap\mathrm{Im}(g_j) = \emptyset$. Thus, they can be described within the same basis such that both have the form given by equation (61) on the preceding page.[39]

4.1. Wave Functions of Strongly Orthogonal Geminals

In general, a wave function might be constructed as the anti-symmetric product of geminals. Adding the restraint of strong orthogonality simplifies the problem of overlap.

$$\Psi_{SOAGP} = \mathcal{A}_n \prod g_i(2i-1, 2i) \tag{66}$$

In order to determine the energy and other properties of an SOAGP function, the second-order reduced density matrix (equation (67)) is necessary.

$$D^{(2)}(\psi_{SOAGP}; 12, 1'2') = \sum b(ijkl) \cdot |\alpha_i(1)\alpha_j(2)| \cdot |\alpha_k(1')\alpha_l(2')|^* \tag{67}$$

where α_i are the natural spin-orbitals of the geminals g_j and

- $\exists n : \{i, j\} = \{k, l\} = \{2n-1, 2n\} =: \sigma \Rightarrow b(\sigma\sigma) = c_i \lambda_\sigma$
- $\sigma \neq \tau \exists i : \hat{\sigma} g_i^* \neq 0 \neq \hat{\tau} g_i^* \Rightarrow b(\sigma\tau) = c_i \xi_\sigma \xi_\tau^*$
- $\sigma \neq \tau, i \in \sigma, j \in \tau, \hat{\tau} g_m^* \neq 0, \hat{\sigma} g_n^* \neq 0, n \neq m \Rightarrow b(ijij) = c_{ij} \lambda_\sigma \lambda_\tau$
- $c_i = \prod\limits_{j \neq i} \sum\limits_{\hat{\sigma} g_j^* \neq 0} \lambda_\sigma / c, c_{ij} = \prod\limits_{k \neq i, k \neq j} \sum\limits_{\hat{\sigma} g_k^* \neq 0} \lambda_\sigma / c, c = 5 \prod\limits_{j} \sum\limits_{\hat{\sigma} g_j^* \neq 0} \lambda_\sigma$ and in all other cases $b = 0$.
- where $\sigma = \{2i-1, 2i\} \Rightarrow \hat{\sigma} = |\alpha_{2i-1}\alpha_{2i}|$

This second-order reduced density matrix for SOAGP has a block diagonal form – one dense block for each geminal and one diagonal block for the mixing between geminals:

$$\mathbf{D}^{(2)}_{SOAGP} = \begin{pmatrix} \boxed{b(\sigma_1\tau_1)} & & & & \\ & \ddots & & & \\ & & \boxed{b(\sigma_l\tau_l)} & & \\ & & & b(1_1 1_2 1_1 1_2) & \\ & & & & \ddots \end{pmatrix} \tag{68}$$

where $i_j \in \sigma : \hat{\sigma} g_j^* \neq 0$ and $\hat{\sigma}_i g_i^* \neq 0$. The natural orbitals of this density matrix are the natural orbitals of the generating geminals.

Optimizing the energy for SOAGP poses the interesting problem of partitioning of the geminal spaces. There are three variables for optimization:

a) basis functions ϕ_i,

b) geminal coefficients x_i of the basis function ϕ_i,

c) partitioning of ϕ among the $N/2$ geminals comprising the SOAGP wave function.

The simultaneous optimization of the latter two points can be done by solving coupled eigenvalue equations. These equations can be derived from imposing variational conditions on the energy.[37]

$$\epsilon_\mu c_{\mu j} = \sum_i c_{\mu i} H_{ij}(\mu) \tag{69}$$

$$H_{ij}(\mu) = \left\langle \phi_{2i-1} \wedge \phi_{2i} \middle| \hat{K} \middle| \phi_{2j-1} \wedge \phi_{2j} \right\rangle + \delta_{ij} \sum_{\nu \neq \mu} \left\langle \phi_{\mu 2i-1} \wedge \phi_{\mu 2i} \middle| \hat{K} \middle| g_\nu \right\rangle \tag{70}$$

where $a \wedge b$ indicates the anti-symmetric tensor product of a and b. Similarly, equations can be derived for the optimization of the basis functions.

SOAGP has been applied to systems like LiH, BH[40] or NH.[41] In these calculations, it was found that the majority of the energy lowering was due to explicit correlation found in the individual geminals, while the inter-geminal interactions displayed positive as well as negative character. Due to the strong orthogonality condition, the inter-geminal correlation is insufficiently described. For n electrons, there are $n(n-2)/4$ inter-geminal correlations while there are only $n/2$ intra-geminal interactions. As may be expected, the percentage of correlation energy retrieved decreases with the number of electrons.

4.2. Singlet-Type Strongly Orthogonal Geminals

Recently, an alternative scheme based on singlet-type strongly orthogonal geminals (SSG) was proposed.[38] In this scheme, the wave function is split into geminal subspaces depending on the number of spin-up or spin-down electrons, while the wave function is "filled up" with one Slater determinant.

$$\psi_{SSG} = \mathbf{A}_n \left[\prod_{i=1}^{n_\alpha} g_i(2i-1, 2i) \right] \cdot \left| \phi_k(2n_\alpha + 1) \cdots \phi_{k+n_\beta-n_\alpha}(n_\alpha + n_\beta) \right| \tag{71}$$

$$\begin{aligned} ND^{(2)}_{SSG} =& s \sum_{i=1}^{n_\alpha} c_i g_i(1,2) g_i(1'2') + g \sum_{k,l} s_{kl} \left| \phi_k(1)\phi_l(2) \right| \cdot \left| \phi_k(1')\phi_l(2') \right| + \\ & \sum_{\substack{k,i \\ j \in \sigma : \hat{\sigma} g_i^* \neq 0}} c_i s_k |g_{i\sigma}|^2 \left| \phi_k(1) g_{ij}(2) \right| \cdot \left| \phi_k(1') g_{ij}(2') \right| \end{aligned} \tag{72}$$

where $\Delta n = n_\beta - n_\alpha$, $s_{kl} = \prod_{i \neq k, i \neq l} |\phi_i|^2$, $s_k = \prod_{j \neq k} |\phi_j|^2$, $s = \prod_j |\phi_j|^2$, $c_i = \prod_{j \neq i} |g_j|^2$, $g = \prod |g_i|^2$, and $N = (n_\alpha + \Delta n(\Delta n - 1)/2 + 2n_\alpha \Delta n) sg$. Hence, the SSG wave function is the anti-symmetrized product of a Slater determinant with an SOAGP wave function. Here the geminal subspaces are optimized by comparing the energies of a given geminal primitive $|g_{i,2j-1}g_{i,2j}|$ in geminal i and the same primitive in geminal j.

The SSG approximation has been used to study a group of diatomics from the G2/97 test set[42,43] as well as the potential energy surface (PES) of carbon monoxide. Except for non-covalently bonded, e.g., highly polarized, molecules, SSG is superior to HF in describing geometry and often comparable to coupled cluster in the singles and doubles approximation (CCSD). When comparing harmonic frequencies, SSG performs even better, *generally* being comparable to CCSD.

The PES of CO for SSG is only slightly better than HF but qualitatively correct in that it is smooth and describes dissociation towards infinity. Although unrestricted MP2[17] and CCSD(T) recover much more of the correlation energy than SSG, the associated curves are not as smooth as for SSG. The equilibrium distance found with SSG is extremely close to experiment ($r_e^{SSG} = 1.126$ Å, $r_e^{expt.} = 1.1283$ Å).[38]

4.3. Perturbation Theory on SOAGP

Recently, SOAGP has been used as the reference state in PT.[44,45] As discussed in section 2.4, the full Hamiltonian is split into two parts of which $\hat{H}_0$ can be solved for a state (equations (73) and (74)). The resulting conditions for first order (equations (74) to (77) on pages 28–29) describe an easily solved linear system of equations.

$$\hat{H} = \hat{H}_0 + \hat{W} \tag{73}$$

$$(\hat{H}_0 - E_0)\left|\Psi^{(0)}\right\rangle = 0 \tag{74}$$

$$\Psi^{(i)} = \sum_i d_i \Psi_i, \langle \Psi_0 | \Psi_i \rangle = \delta_{i0} \tag{75}$$

$$\left\langle \Psi_j \left| \hat{W} \right| \Psi^{(0)} \right\rangle + \sum_i d_i \left\langle \Psi_j \left| (\hat{H}_0 - E_0) \right| \Psi_i \right\rangle = 0 \, \forall j \tag{76}$$

$$E^{(2)} = \left\langle \Psi_0 \left| \hat{W} \right| \Psi^{(1)} \right\rangle, E^{(1)} = \left\langle \Psi_0 \left| \hat{W} \right| \Psi_0 \right\rangle \tag{77}$$

In order to define $\hat{H}_0$ such that Ψ_{SOAGP} is an eigenfunction, $\hat{H}_0$ was split into a sum of operators $\hat{H}_K$. These operators $\hat{H}_K$ are defined so that $\hat{H}_K g_K = E^0_K g_K$ (equation (79)). Then, the zeroth order energy is merely $\sum E^0_K$ and $\hat{W} = \hat{H} - \hat{H}_0$.

$$\hat{h}^{\text{eff}}_{\mu\nu} = \hat{h}_{\mu\nu} + \sum_{\lambda\sigma} \hat{D}^{(1)}_{SOAGP} \left\langle \nu \wedge \sigma \left| \frac{1}{r_{12}} \right| \mu \wedge \lambda \right\rangle \tag{78}$$

$$\hat{H}_K = \sum_{\mu,\nu \text{ of } g_k} \hat{h}^{\text{eff}}_{\mu\nu} \hat{a}^\dagger_\mu \hat{a}_\nu + \frac{1}{2} \sum_{\mu\nu\lambda\sigma \text{ of } g_K} \left\langle \mu\nu \left| \frac{1}{r_{12}} \right| \lambda\sigma \right\rangle \hat{a}^\dagger_\lambda \hat{a}^\dagger_\mu \hat{a}_\sigma \hat{a}_\nu \tag{79}$$

where, given a, a is *of* g_K, if $\hat{g}_K \hat{g}^*_K a \neq 0$, and $\hat{a}^\dagger_\alpha, \hat{a}_\alpha$ represent creation and annihilation operators of α, respectively, in the second quantization formalism.

PESs for fluorine, HF and water using this *ansatz* showed promise.[45] In all cases the perturbative approach improved the accuracy considerably at a small increase of computational cost. Especially interesting is the possibility of linear scaling.

4.4. Anti-Symmetrized Geminal Product[46]

Since only the eigenfunctions of a second order reduced density matrix are necessary, it is natural to construct a wave function from *just* such functions. Given only one geminal g a simple function for a geminal functional theory can be derived.

$$\psi_{AGP} = g^N = \mathcal{A}_N \prod_{i=1}^{N/2} g(2i-1, 2i) \tag{80}$$

Since the generating geminals of an SOAGP wave function can be described in the same basis, it is tempting to assume that summing the geminals leads to an AGP generating geminal for the SOAGP wave function; that is not the case.

$$\Psi_{SOAGP} \propto \Psi_{AGP}(g) \not\Leftarrow g = \sum_i g_i \tag{81}$$

The AGP function for the summation includes contributions from g_i^N and other mixed higher order products, which are not present in the SOAGP function.

Given such a function, the second order density matrix has the following form:[35]

$$\rho_2(\psi_{AGP}; 12, 1'2') = \sum b(ijkl) \cdot |\alpha_i(1)\alpha_j(2)| \cdot |\alpha_k(1')\alpha_l(2')|^* \tag{82}$$

where α_i are the natural spin-orbitals of g and

- $\exists n : \{i, j\} = \{k, l\} = \{2n-1, 2n\} =: \sigma \Rightarrow b(\sigma\sigma) = 2c\lambda_\sigma a_{m-1}(\bar{\sigma})$
- $\sigma \neq \tau \Rightarrow b(\sigma\tau) = 2c\xi_\sigma\xi_\tau^* a_{m-1}(\bar{\sigma}\bar{\tau})$
- $\sigma \neq \tau, i \in \sigma, j \in \tau \Rightarrow b(ijij) = 2c\lambda_\sigma\lambda_\tau a_{m-2}(\bar{\sigma}\bar{\tau})$
- $\prod(1 + \lambda_\sigma t) = \sum a_m t^m; a_{m-1}(\bar{\sigma}) = \frac{\partial a_m}{\partial \lambda_\sigma}; a_{m-2}(\bar{\sigma}\bar{\tau}) = \frac{\partial^2 a_m}{\partial \lambda_\sigma \partial \lambda_\tau}, c = 1/a_m$

This matrix has a very simple block diagonal structure of only two blocks — one of which is diagonal.

$$\rho_2(\psi_{AGP}) = \left(\begin{array}{c|cc} b(\sigma\tau) & & \\ \hline & b(1313) & \\ & & \ddots \end{array}\right) \tag{83}$$

Seeing this simple form raises the question whether the natural geminals of the wave function are the same as the determinants in equation (61) on page 23. This is definitely not the case since $|d_1 d_3|$ does not appear in the definition of the generating geminal, but is a natural geminal of the AGP wave function. On the other hand, the natural orbitals of the geminal and the AGP wave function do indeed coincide.

$$(n-1)\hat{\rho}^{(1)} = \sum_{\substack{\sigma \\ i\in\sigma}} |i\rangle\langle i|\, c\lambda_\sigma a_{m-1}(\bar{\sigma}) + \sum_{\sigma>\tau}\sum_{\substack{i\in\sigma \\ j\in\tau}} \left(|i\rangle\langle i| + |j\rangle\langle j|\right) c\lambda_\sigma\lambda_\tau a_{m-2}(\bar{\sigma}\bar{\tau}) \tag{84}$$

$$= \sum_{\substack{\sigma \\ i\in\sigma}} |i\rangle\langle i|\, c\lambda_\sigma a_{m-1}(\bar{\sigma}) + \sum_{\sigma\neq\tau}\sum_{\substack{i\in\sigma \\ j\in\tau}} |i\rangle\langle i|\, c\lambda_\sigma\lambda_\tau a_{m-2}(\bar{\sigma}\bar{\tau})$$

$$= \sum_{\substack{\sigma \\ i\in\sigma}} |i\rangle\langle i|\, c\lambda_\sigma (2m-1) a_{m-1}(\bar{\sigma}) \tag{85}$$

Equation (85) implies that the first order reduced density matrix is degenerate for every eigenvalue. The final trace normalizes to $n(n-1)$ as expected. Any Slater determinant, e.g., the HF solution, can be described by a very simple geminal.

$$|\psi_1 \cdots \psi_n| \propto \Psi_{AGP}(g) \Leftarrow g = \sum_{i=1}^{n/2} |\psi_{2i-1}\psi_{2i}| \tag{86}$$

Since all natural orbitals are degenerate, any collection of pairs of spin-orbitals will produce a geminal which gives rise to the Slater determinant, so there is no unique geminal which represents the Slater determinant.

As equation (82) on the page before demonstrates, the exact wave function is no longer explicitly needed nor is an approximate wave function used. Instead, the generating geminal determines the wave function as well as the reduced density matrices and thereby the energy of an AGP wave function.

5. Geminal Functional Theory (GFT)

It has been shown that any evenly degenerate $\mathbf{D}^{(1)}$ can be generated by an AGP function.[35,47] So in the spirit of density functional theory a functional can be defined by[48]

$$E\left[\mathbf{D}^{(1)}\right] = \inf_{\mathbf{D}^{(2)} \to \mathbf{D}^{(1)}} \operatorname{tr}\left(\mathbf{D}^{(2)}\mathbf{K}\right) \tag{87}$$

where all N-representable $\mathbf{D}^{(2)}$ which reduce to $\mathbf{D}^{(1)}$ are considered. Also, a Hamiltonian of even number of electrons which is invariant under time-reversal (i.e. most chemically interesting Hamiltonians) has been shown[49] to have a ground state with evenly degenerate $\mathbf{D}^{(1)}$. It is therefore possible to restrict the search based on a geminal. Disregarding the N-representability leads to a lower bound on the FCI energy while restricting the searched second order reduced density matrices beyond N-representability gives a true upper bound on the energy. A functional is thus reduced to the choice of subset of all $\mathbf{D}^{(2)}$. One such functional would be the restriction to AGP second order reduced densities only, as are those of SOAGP and SSG. Alternatively, the 2-RDMs derived from AGP may be used as points of departure for energy minimizations which do not enforce full N-representability, but instead merely enforce the necessary positive, semi-definiteness conditions (see section 3, equations (55) to (57) on page 22). This method has been employed with disappointing errors on small systems.[50–53]

CHAPTER 3

FROM QUANTUM CHEMISTRY TO COMPUTATIONAL CHEMISTRY

1. The PSI 3.0 Program Package

The PSI3 program package is an extensive quantum chemistry package. It offers various implementations of HF, CC, CI, and more. It is primarily being developed by Drs. T. Daniel Crawford at Virginia Polyechnical Institute and State University, C. David Sherrill and Edward F. Valeev at Georgia Institute of Technology, and Rollin A. King at Bethel University. The source code and documentation is available online under `http://www.psicode.org/`. For the purpose of implementing generalized anti-symmetrized geminal products (GAGP) it was only necessary to utilize the overlap-, nuclear attraction-, and electron-repulsion-integrals. The following functions are employed in the GAGP implementation:

- `psi_start()`: Start the PSI session initializing the input- and output-files and various other small tasks.
- `psi_stop()`: Reverses the processes of `psi_start()`.
- `ip_cwk_add()`: Add a structural name to the working keyword tree.
- `psio_init()`: Initialize the Input/Output data structures.
- `psio_done()`: Free all I/O data structures.
- `chkpt_init()`: Initialize the checkpoint files.

- `chkpt_rd_enuc()`: Read the nuclear repulsion from the checkpoint file.
- `chkpt_rd_nso()`: Read the number of symmetry-adapted orbitals from the checkpoint file.
- `chkpt_close()`: Close the checkpoint file(s).
- `tstart()`: Start the timing.
- `tstop()`: Stop timing.
- `iwl_rdone()`: Read the one-electron integrals with labels.
- `iwl_rdtwo()`: Read the two-electron integrals with labels.

The integrals were computed using `cints` included in PSI3. Unlike previous chapters in which merely algebraic properties were discussed, it is helpful to understand how the integrals are computed. As mentioned in chapter 2, section 2.1, all integrals are between Gaussian functions.

$$\begin{aligned}
&(88) \qquad \langle p_1^m \,|\mathbf{v}|\, p_2^n\rangle = \int v(x)\cdot(x-x_1)^m \mathrm{e}^{-\zeta_1(x-x_1)^2}(x-x_2)^n \mathrm{e}^{-\zeta_2(x-x_2)^2}dx \\
&\qquad\qquad = \int v(x)(x-x_1)^m(x-x_2)^n \mathrm{e}^{-(\zeta_1+\zeta_2)\left(x-\frac{\zeta_1x_1+\zeta_2x_2}{\zeta_1+\zeta_2}\right)^2}\cdot \mathrm{e}^{\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}(x_1-x_2)^2}dx \\
&(89) \qquad\qquad = \mathrm{e}^{\frac{\zeta_1\zeta_2}{\zeta_1+\zeta_2}(x_1-x_2)^2}\sum_{i=0}^{m+n} a_i \int v(x)x^i \mathrm{e}^{-\bar{\zeta}(x-\bar{x})^2}dx \\
&(90) \qquad p_i^j = (x-x_i)^j \mathrm{e}^{-\zeta_i(x-x_i)}
\end{aligned}$$

As equation (89) indicates all integrals can be derived from simple Gaussian integrals over a potential v. As seen earlier there are only three potentials to be considered:

a) $v\left(\vec{x}\right) = 1$

b) $v\left(\vec{x}\right) = \frac{1}{|\vec{x}|}$

c) $v\left(\vec{x}_1, \vec{x}_2\right) = \frac{1}{|\vec{x}_1 - \vec{x}_2|}$

The first case is easily solved using partial differentiation; $\int x^{2n} \mathrm{e}^{-\zeta x^2} dx = \prod_{i=1}^{n} \frac{2i-1}{2\zeta} \int \mathrm{e}^{-\zeta x^2} dx = \prod_{i=1}^{n} \frac{2i-1}{2\zeta} \sqrt{\pi/\zeta}$, odd polynomials integrate to zero.

2. Minimization Techniques

In the previous discussions optimizations of one kind or another were at the heart of the problem. Finding the ground state, in particular, is a minimization problem. There are several general optimization schemes depending on the amount of available information. They all have in common that they may not yield a (correct) result unless the starting point is sufficiently close to the desired solution. Of course, without knowing the solution it is usually not possible to know whether the guess is suitably close.

2.1. Newton-Raphson Optimizations

The preferred optimization technique of non-linear problems is the Newton-Raphson optimization. When it does converge, it is guaranteed to converge quadratically, i.e., $|e_{n+1}| \leq C\,|e_n|^2$ for some positive constant C and successive errors e_i. This means that if two digits are correct at one iteration roughly 4 will be correct the next and 8 the iteration thereafter.

$$0 \approx \nabla f\left(\vec{x}_i\right) + \nabla\nabla^* f\left(\vec{x}_i\right) \cdot \left(\vec{x}_{i+1} - \vec{x}\right) \Rightarrow \vec{x}_{i+1} = \vec{x}_i - \left(\nabla\nabla^*\right)^{-1} f\left(\vec{x}_i\right) \nabla f\left(\vec{x}_i\right) \tag{91}$$

The disadvantage of a Newton-Raphson algorithm is that the first and second derivatives are needed (equation (91)). Many problems do not allow for an easy or efficient (analytical) evaluation of the second derivative. To alleviate this problem quasi-Newton-Raphson methods have been developed, of which the BFGS algorithm, named after Broyden,[54] Fletcher,[55] Goldfarb,[56] and Shanno,[57] is the most popular. In this algorithm the second derivatives are approximated using previous iterate values.

$$\Delta\hat{M}_{k+1} = \frac{1}{\Delta\vec{x}^*_{k+1}\Delta f_{k+1}}\Delta f_{k+1}\Delta f^*_{k+1} + \frac{1}{\Delta\vec{x}^*_{k+1}\hat{M}_k\Delta\vec{x}_{k+1}}\hat{M}_k\Delta\vec{x}_{k+1}\Delta\vec{x}^*_{k+1}\hat{M}_k \tag{92}$$

$$\Delta f_{k+1} = \nabla f\left(\vec{x}_{k+1}\right) - \nabla f\left(\vec{x}_k\right) \tag{93}$$

$$\Delta\vec{x}_{k+1} = -\hat{M}_k\nabla f(\vec{x}_k) = \vec{x}_{k+1} - \vec{x}_k \tag{94}$$

$$\hat{M}_{k+1} = \hat{M}_k - \Delta\hat{M}_{k+1} \tag{95}$$

$$\hat{M}_k \approx \nabla\nabla^* f\left(\vec{x}_k\right) \tag{96}$$

$$\vec{x}_{k+1} = \vec{x}_k - \hat{M}_k^{-1}\nabla f(\vec{x}_k) \tag{97}$$

From equation (92) it is immediately evident, that for poorly conditioned systems the correction to the Hessian blows up, when $\Delta\vec{x}^*_k\Delta f_{k+1} \approx 0$ or $\Delta\vec{x}^*_k\hat{M}_k\Delta\vec{x}_k \approx 0$. In particular, if the function f is invariant under scaling, i.e., $f(s\vec{x}) = f(\vec{x})$, then $\nabla f\left(\vec{x}_k\right) \perp \vec{x}_k \forall k$. Hence, $\nabla\nabla^* f(\vec{x}_k)\vec{x}_k = 0$ and neither method is applicable without modification. Usually this is known *a priori*, so that the Hessian can be unitarily rotated and the coordinate along $\vec{x}_k$ eliminated. In such a case another danger is the increase of $\vec{x}_k$, as the step size is always perpendicular to $\vec{x}_k$ and, thence, $\left|\vec{x}_{k+1}\right|^2 = \left|\Delta\vec{x}_{k+1}\right|^2 + \left|\vec{x}_k\right|^2$. The increasing norm of $\vec{x}$ introduces numerical errors, and trying to curtail those errors by renormalization renders the approximate Hessian useless (see appendix A on page 73 for an implementation in `gagp_nr.cc`).

FIGURE 1. Newton Raphson of λ_σ for AGP with the generating geminal $g = \sum \lambda_\sigma |\alpha_{2\sigma-1}\alpha_{2\sigma}|$ for Be.

Research for improved convergence of quasi-Newton-Raphson methods is ongoing despite the durability of the BFGS formula over 30 years.[58]

2.1.1. *An Example for Newton-Raphson Optimizations.* AGP is highly non-linearly dependent on the eigenvalues of the 1-RDM of the generating geminal. As figure 1 shows the convergence is almost instantaneous.

2.2. Lagrange-Multiplier Methods

Lagrange multipliers were introduced expressly to deal with difficult boundary conditions. Again the use of tensors allows for an elegant discussion of Lagrange multipliers. Let $X = \{f : \mathbb{R}^n \to \mathbb{R}\}$ denote the ring of functionals on n variables. Inadvertently, the functional F

to be considered for optimization is in X. Then, let X_n be the n-dimensional vector space with coefficients in X. The beauty of tensors now allows us to define a new multiplication on $T(X_n)$ via the equivalence of $T^p_q(X_n)$ and $T^{p+q}(X_n)$. The multiplication is generated by $t \cdot t' = x^*_p(x'_1)\left(\bigotimes_{i=1}^{p-1} x_i\right) \otimes \left(\bigotimes_{i=2}^{q} x'_i\right) \in T^{p+q-2}(X_n)$ for $t = \bigotimes_{i=1}^{p} x_i \in T^p(X_n)$ and $t' = \bigotimes_{i=1}^{q} x'_i \in T^q(X_n)$. Essentially, this extends the multiplication of vectors with matrices and vectors to three-dimensional and higher matrices. Furthermore, define on $T(X_n)$ a differentiation operator $\hat{\partial}$ such that $\hat{\partial}(ft) = \left(\sum_{i=1}^{n} \frac{\partial f}{\partial e_i}\tilde{e}_i\right) \otimes t$ where $e_i, \tilde{e}_i$ are corresponding orthonormal basis vectors of $\mathbb{R}^n$ and X_n, respectively, $f \in X$, and $t = \bigotimes_{i=1}^{q} \tilde{e}_{j_i} \in T^q(X_n)$.

Now, given a set of boundary conditions, a functional may be considered an implicit coordinate definition. The m boundary conditions are collected in an m-dimensional vector on $T(X_n)$, with usually each $B_i \in T^0(X_n)$. A new functional $\mathfrak{L} \in X$ is defined reusing the original F and burying the boundary conditions B_i in the new functional explicitly.

$$\mathfrak{L} = F + \sum_{i=1}^{m} \mu_i B_i = F + \vec{B}^*\vec{\mu} \tag{98}$$

$$\hat{\partial}\mathfrak{L} = \hat{\partial}F + \sum_{i=1}^{m} \mu_i \hat{\partial} B_i = \hat{\partial}F + \hat{\partial}\vec{B}^*\vec{\mu} \tag{99}$$

The m Lagrange multipliers μ_i are to be determined during optimization and may as such be considered composite functions. They are chosen such that the derivative $\hat{\partial}\mathfrak{L}$ along a direction $d_i \in X_n$ is zero by solving the linear system in equation (100) on the following page. These directions are collected in a vector $\vec{D} = (d_1 \cdots d_m)^*$, which is, upon closer inspection, equivalent to an $m \times n$ matrix. As equation (101) on the next page suggests, choosing directions seems superfluous, if $\vec{D}$ as a matrix is invertible; but $\vec{D}$ as a matrix can only be invertible if m equals the dimension n of $\hat{\partial}F$. This is generally not the case and also highly uninteresting since all coordinates are fixed by the boundary conditions. A good choice of directions is $\hat{\partial} B_i$ since it eliminates optimizing along the constraint direction and results in a hermitian, positive semi-definite $\hat{B}_0$. If $\hat{B}$ is not invertible then the boundary conditions are not linearly independent. If this is the case, boundary conditions may be eliminated for the current iteration.

$$\hat{B}\vec{\mu} = \begin{pmatrix} d_1 \cdot \hat{\partial} B_1 & \cdots & d_1 \cdot \hat{\partial} B_m \\ \vdots & & \vdots \\ d_m \cdot \hat{\partial} B_1 & \cdots & d_m \cdot \hat{\partial} B_m \end{pmatrix} \vec{\mu} = \begin{pmatrix} d_1 \cdot \hat{\partial} F \\ \vdots \\ d_m \cdot \hat{\partial} F \end{pmatrix} = \vec{D}\hat{\partial} F \tag{100}$$

$$(\vec{D}\hat{\partial}\vec{B}^*)\vec{\mu} = \vec{D}\hat{\partial} F \tag{101}$$

$$\hat{B}_0 = \begin{pmatrix} \hat{\partial} B_1 \cdot \hat{\partial} B_1 & \cdots & \hat{\partial} B_m \cdot \hat{\partial} B_1 \\ \vdots & & \vdots \\ \hat{\partial} B_1 \cdot \hat{\partial} B_m & \cdots & \hat{\partial} B_m \cdot \hat{\partial} B_m \end{pmatrix} = \hat{\partial}\vec{B}\hat{\partial}\vec{B}^* \tag{102}$$

When the Lagrangian system is solved, any number of optimization techniques may be employed to find solutions to the reduced functional. Yet, care must be taken, when second order effects are considered. Assuming the multipliers to be constant in second derivatives may not lead to satisfactory results as the second derivatives do not reflect constraining the functional to the boundary conditions via the Lagrange multipliers. Slow convergence or even non-convergence may be the results; on the other hand, equation (106) on the following page demonstrates how unwieldy second derivatives are even if the first derivative with respect to $\vec{\mu}$ is ignored in the first instance. For an implementation in `gagp_nr.cc` see appendix A on page 73. Luckily, in most cases only a single constraint need be considered, usually a normality condition, which greatly simplifies matters.

$$\mathfrak{L} = F + \vec{B}^{*} \cdot \left(\hat{B}_0^{-1}\hat{\partial}\vec{B}\hat{\partial}F\right) = F + \vec{B}^{*}\vec{\mu} \tag{103}$$

$$\hat{\partial}\mathfrak{L} = \hat{\partial}F + \hat{\partial}\vec{B}^{*}\left(\hat{B}_0^{-1}\hat{\partial}\vec{B}\hat{\partial}F\right) \tag{104}$$

$$\hat{\partial}^2\mathfrak{L} = \hat{\partial}^2F + \hat{\partial}\vec{B}^{*} \otimes \hat{\partial}\vec{\mu} + \hat{\partial}^2\vec{B}^{*} \cdot \vec{\mu} \tag{105}$$

$$\hat{\partial}\vec{\mu} = \hat{B}_0^{-1}\hat{\partial}\vec{B}\hat{\partial}^2F + \hat{\partial}\hat{B}_0^{-1}\hat{\partial}\vec{B}\hat{\partial}F + \hat{\partial}\hat{B}_0^{-1}\hat{\partial}^2\vec{B}\hat{\partial}F \tag{106}$$

2.3. The Steepest Descent Method

A simple and intuitive method, which works very well with Lagrange multipliers, is the idea of following the gradient to a valley of the functional. As a minimum is approached the gradient diminishes and round-off errors may overwhelm the gradient. Steepest descent also approaches the minimum at best linearly and convergence is especially slow in the direct vicinity of the minimum. A variation that does marginally better is a steepest descent method scaled by λ_k, such that particularly flat surfaces can be traversed faster, while very steep surfaces may be traversed slower to avoid repeatedly missing the minimum or failing to converge entirely.

$$\vec{x}_{k+1} = \vec{x}_k - \lambda_k\hat{\partial}F(\vec{x}_k) \tag{107}$$

2.4. Generating Set Searches

In many cases derivatives are too difficult or even impossible to compute. The goal of direct search methods is to persevere in the face of insufficient information, like the lack of derivatives or extremely complex surfaces. Generating set search methods encompass a large set of conventional methods, like finite-difference algorithms, which approximate the derivative.

For a very general generating set search (GSS),[59] nine constant parameters define the starting conditions and control the algorithm:

- A functional $f : \mathbb{R}^n \to \mathbb{R}$.
- A starting point $x_0 \in \mathbb{R}$.
- A step-length convergence tolerance $\Delta_{tol} > 0$.
- An initial step-length control value $\Delta_0 > \Delta_{tol}$.
- An upper-bound contraction parameter θ_{max}.
- A continuous convergence-control function $\rho : [0, \infty) \to \mathbb{R}$ such that $\rho(t)$ decreases with increasing t and $\lim \rho(t)/t = 0$.
- Upper and lower bounds $\beta_{max} \geq \beta_{min} > 0$ on the lengths of the vectors in any generating set.
- A lower bound κ_{min} on the cosine measure of any two vectors in a generating set.

With these parameters in place, a simple algorithm may be defined. First, a set of directions $D_k = G_k \cup H_k$ is defined, which is comprised of a generating set G_k, which positively spans $\mathbb{R}$, i.e., $\forall v \in \mathbb{R}^n \forall g \in G_k \exists \lambda_g \geq 0 : v = \sum_{g \in G_k} \lambda_g g$, and a supplemental set H_k, the importance of which will become clear in the example introduced later. The generating set is subject to the constraints $\beta_{min} \leq ||d|| \leq \beta_{max} \forall d \in G_k$ and, for any two directions $d_1, d_2 \in D_k$, they span an angle of at least κ_{min}, $\cos(d_1, d_2) = \frac{d_1 \cdot d_2}{||d_1|| \cdot ||d_2||} \geq \kappa_{min}$. The supplemental set H_k only needs to comply with the simple requirement that $\beta_{min} < ||d|| \forall d \in H_k$.

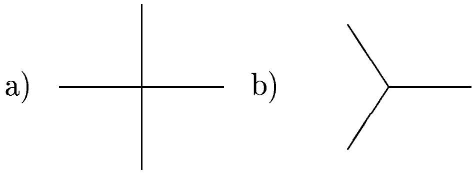

FIGURE 2. a) Compass search directions; b) Simplex search directions.

Next, the actual search step is chosen. If there exists a direction $d_k \in D_k$ such that $f(x_k + \Delta_k d_k) < f(x_k) - \rho(\Delta_k)$, then set $x_{k+1} = x_k + \Delta_k d_k$ and expand the step-length via $\Delta_{k+1} = \phi_k \Delta_k$, where $\phi_k \geq 1$. Otherwise, if for all directions $d \in D_k$ $f(x_k + \Delta_k d) > f(x_k) - \rho(\Delta_k)$, then the iterate remains unchanged, $x_{k+1} = x_k$, while the step-length is contracted, $\Delta_{k+1} = \theta_k \Delta_k$, where $0 < \theta_k < \theta_{max}$. If the step-length falls beneath the tolerance threshold, $\Delta_{k+1} < \Delta_{tol}$, then the algorithm is terminated.

An obvious problem is to choose a direction, if several directions lead to a decrease of f. The convergence control function ρ alleviates this problem by rejecting search directions with insufficient decrease of f; but care must be taken that ρ does not decrease too slowly leading to slow convergence. Even so, there may be more directions than one which satisfy the sufficient decrease condition. There are two viable approaches to this problem:

a) Compute $f(x_k + \Delta_k d)$ for all $d \in D_k$ and choose the direction with the largest decrease.
b) In order to save computation time, take the first direction which satisfies the sufficient-decrease condition.

In the second case, the order of computation is no longer irrelevant which brings us to choosing G_k and H_k.

The two simplest cases for choosing a generating set are the "compass" directions and the n-dimensional simplex (see Fig. 2). For compass search, the generating set consists of positive and negative unit vectors $\pm\vec{e}_i$, resulting in $2n$ directions. Walking along the compass directions leads to a rectangular grid or lattice. The n-dimensional simplex is more complicated to explain, but conceptually simple and, more importantly, computationally cheaper. A simplex is a set of $n+1$ vectors $\{\vec{x}_i\}$ in $\mathbb{R}^n$ for which all pairs of vectors have the same angle and $\sum \vec{x}_i = \vec{0}$. These vectors may be generated iteratively as the space is expanded from $\mathbb{R} \to \mathbb{R}^n$. Starting from $\vec{x}_1^{(1)} := \vec{e}_1$ and $\vec{x}_2^{(1)} := -\vec{e}_1$, and proceeding from the set $\{\vec{x}_i^{(m)}\}_{i=1}^{m+1}$ to $m+1$, we set $\vec{x}_i^{(m+1)} := \frac{\sqrt{m(m+2)}}{m+1}\vec{x}_i^{(m)} - \frac{1}{m+1}\vec{e}_{m+1}$ and $\vec{x}_{m+1}^{(m+2)} := \vec{e}_{m+1}$. Then, $\vec{x}_i^{(m)} \cdot \vec{x}_j^{(m)} = -\frac{1}{m}$ if $i \neq j$, and 1 otherwise.

2.5. An Example for GSS

As for Newton-Raphson, AGP optimization will serve as an example. Since the derivatives consume so much computational time, a good direct search method is desirable. For the generating set, we start with the $n+1$-dimensional simplex and proceed through the GSS algorithm with the following parameters:

- Start from
  - a) the UHF generating geminal.
  - b) a random HF geminal.
- $\Delta_{tol} = 10^{-16}$.
- $\Delta_0 = 10^{-4}$.
- $\theta_{max} = \sqrt{2}$.
- $\rho = 0$.
- $\beta_{max} = \beta_{min} = 1$.

- $\kappa_{min} = \frac{1}{n+1}$.

At the end of each iteration, the entire simplex is rotated in the plane between the first vector and the step-direction vector such that in the new set the first vector is the step-direction vector. The set of directions is augmented by one additional direction if at least one vector failed to yield a decrease. In that case, an approximate of the derivative is computed from the computed values. This additional direction makes up the set $\mathcal{H}_k$, which serves just this purpose. A step in the direction of the approximate gradient is computed and compared to the last successful step. If it is superior, the last computation is chosen. If no direction results in a decrease, the simplex is rotated to represent the approximated derivative in addition to decreasing the step-length with $\theta_k = 1/\sqrt{2}$. Consecutive rotations incorporate some higher-order information as the simplex is not generally invariant under rotations.

FIGURE 3. Convergence for Be starting from a UHF guess. The baseline is the HF energy.

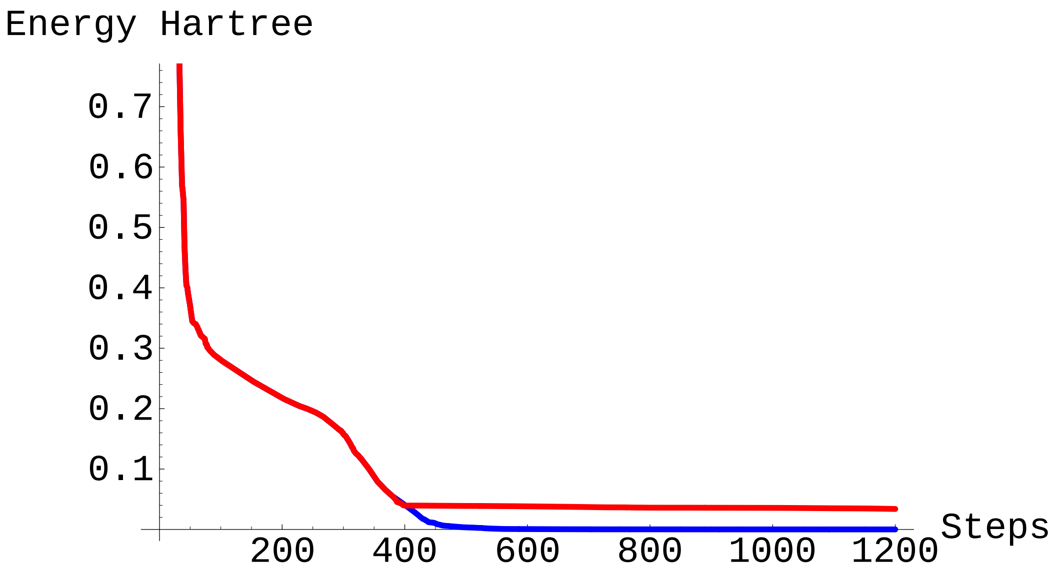

FIGURE 4. Convergence for Be starting from a dumb guess. The red is for restarted calculation every 400 steps. The blue line is a continuous run. The baseline is the HF energy.

Figures 3 and 4 report the convergence results. It is interesting to note that restarting the algorithm is actually preferable. This may be attributed to the complexity of the surface, which forces tight turns and small steps, or an insufficiently aggressive expansion strategy ϕ_k.

3. Linear Algebra in C++

As is apparent from the foregoing chapters, linear algebra plays a major role in quantum chemistry. The object-oriented programming language C++ has the flexibility to describe the mathematical objects while retaining the optimization opportunities of procedural languages, like its parent C. In plain C, programming linear algebra (LA) is hampered by the repetitive programming and renaming of algorithms whenever a new object type is introduced. C++ reduces the renaming complexity and a lot of reimplementation via overloading objects. Overloading allows the redefinition of an object with different parameters, such that for instance `multiply(x,y)` can be used independently of whether `x` and `y` are of type `double`, or `int`, or a mix thereof. Another advantage over C is the easy introduction of operators like "*" for multiplication. Adding these together allows the development of a library such that the user can essentially write programs in terms of mathematical operations.

Hiding the specifics of type-dependent implementations is only the first layer of at least three that may be introduced. If nothing else, the commonly available and standardized BLAS routines may be wrapped by this layer; but much more can be done. There are two important aspects to optimization: storage management and abstract properties. For instance, a block of memory allocated for a matrix may be interpreted as a vector of column vectors or a vector of row vectors. Choosing to have a single block of memory versus allocating each vector individually is a storage management decision that is hidden from the first layer of complexity. On the other hand, row versus column matrices is an abstraction decision which is also hidden from the first layer of complexity. In the previous sections, we have seen more examples of abstract properties of matrices, e.g., symmetry. Such properties allow new storage types which take advantage of them, roughly half the memory need be stored for a symmetric matrix without loss of information. Therefore the second layer of complexity encompasses choice of storage types and abstractions for the input. This allows for dynamic decisions by the implementer to choose the fastest or most memory-conserving algorithm. As an example, multiplying two column matrices is fastest if the resulting matrix is also a column matrix, for two row matrices a resultant row matrix is fastest, but for multiplying a row matrix with a column matrix neither storage-type has an advantage over the other. Hence, in the third layer this constraint is relaxed as the overall flow of the program may be fastest if certain types are maintained, e.g., adding a row matrix to the product of a row matrix with a column matrix. In the final layers, control of intermediates is left to the user, since many may be reused and allocating and freeing memory uses valuable computational time. Each layer need not be reimplemented but instead makes use of calling the next higher level.

The layered system affords maximum user-friendliness while maintaining high flexibility and optimizability. Good programming practice calls for well maintained and documented code and debugability. The latter is achieved by requiring functions to have integrity checks which can be turned on or off with the appropriate compilation switch and are linked to informative error messages and call trace. The integrity checks also serve the purpose of secure programming. Memory management is accomplished using reference counted objects, which destroy themselves automatically in the C++ language when they are no longer used. Given this framework, it is therefore easy to extend the library with new functionality (see appendix A on page 73 for implementations).

CHAPTER 4

THE GROUND STATE

For the vast majority of problems in quantum chemistry, the ground state of the Hamiltonian needs to be found; for transition state searches or potential energy surfaces in general, the ground state of each geometry chiefly contributes to the dynamical physics; for excitations in spectroscopy, the excitation energy as a relative energy demands the ground state energy. As the general explicit solution to the ground state still eludes the quantum community, it is necessary to investigate which properties are deducible without having the exact solution. Given these properties we can improve the approximations to the ground state by requiring them.

To this end, we will return to tensor products and I will rehash some of their properties. Every tensor t in $G \otimes F$ can be expressed as $\sum g_i \otimes f_i$. Furthermore, the assignment is unique if the g_i are linearly independent (or vice versa). Since $T^m(V)$ is a tensor product, for instance $T^2(V) \otimes T^{m-2}(V)$, every tensor of $A^m(V)$, which is a subspace of $T^m(V)$, may be described in such a way. In chapter 2, section 3, we were introduced to this notion for the eigenvectors of RDMs. In that case, the set of vectors g_i was set to the set of eigenvectors of an RDM and the vectors f_i were eigenvectors of the complementary RDM.

Given a set of orthonormal vectors $g_i \in A^2(V)$, corresponding homomorphisms $\hat{G}_i$ (equation (108) on the next page) may be defined from $A^m(V)$ to $A^{m-2}(V)$. If V has dimension $\dim V = n$, then $A^m(V)$ has dimension $\dim A^m(V) = \binom{n}{m}$, while $A^{m-2}(V)$ has dimension

$\dim A^{m-2}(V) = \binom{n}{m-2} = \dim A^m(V) \cdot \frac{m(m-1)}{(n-m+2)(n-m+1)}$. Therefore, the kernel or null-space $N(\hat{G}_i)$ has a dimension of at least $\frac{(n+1)(n-2m)}{(n-m+2)(n-m+1)} \cdot \dim A^m(V)$. Thus, the hermitian operator $\hat{G}_i^* \hat{G}_i$ has a kernel of the same dimension. Furthermore, $\sum \hat{G}_i^* \hat{G}_i$ is the identity on $A^m(V)$. Hence, the intersection of all kernels must contain only the null-vector.

$$\hat{G}_i : A^m(V) \to A^{m-2}, t = \sum g_i \otimes f_i \mapsto f_i \tag{108}$$

Equipped with these operators, the Hamiltonian may be described in yet another way. The Hamiltonian can be directly related to the eigenvalues of $\mathbf{K}$, if we choose the set of eigenvectors k_i of $\mathbf{K}$ as g_i. In the following, we will assume the eigenvalues ordered from lowest to highest.

$$|\Psi\rangle = \sum |k_i\rangle \otimes \hat{K}_i |\Psi\rangle \Rightarrow 1 = \sum |k_i\rangle \otimes \hat{K}_i \tag{109}$$

$$E = \left\langle \Psi \middle| \hat{K} \middle| \Psi \right\rangle = \sum \epsilon_i \left\langle \Psi \middle| \hat{K}_i^* \hat{K}_i \middle| \Psi \right\rangle \tag{110}$$

$$\Rightarrow \hat{H} = \sum \epsilon_i \hat{K}_i^* \hat{K}_i = \sum \epsilon_i |k_i\rangle \otimes \hat{K}_i \tag{111}$$

$$\mathbf{K} k_i = \epsilon_i k_i \tag{112}$$

In this form, the properties of $\hat{K}_i^* \hat{K}_i$ and $|k_i\rangle \otimes \hat{K}_i$ dictate the energy. For an eigenstate of $\hat{H}$ it is thus necessary to fulfill equation (114).

$$\hat{H}\Psi = \sum \epsilon_i |k_i\rangle \otimes \hat{K}_i |\Psi\rangle = E \sum |k_i\rangle \otimes \hat{K}_i |\Psi\rangle \tag{113}$$

$$\Leftrightarrow E \left\langle \Psi \middle| \hat{K}_i^* \hat{K}_i \middle| \Psi \right\rangle = \epsilon_i \forall i \tag{114}$$

Taking a basis e_i for $A^m(V)$, the energy can further be described by equation (116) on the following page. Since the energy is invariant under unitary transformations, let the set

$\{e_i\}$ consist of the eigenvectors of $\hat{K}_0^*\hat{K}_0$.

$$\Psi = \sum \psi_i e_i, \sum |\psi_i|^2 = 1 \tag{115}$$

$$E = \sum \psi_i^* \psi_j \epsilon_k \left\langle e_i \left| \hat{K}_k^* \hat{K}_k \right| e_j \right\rangle \tag{116}$$

The energy is over-defined by the coefficients ψ_i, if $\binom{\dim V}{2} < \binom{\dim V}{m}$. Then, it is possible to vary the energy while constraining contributions. For instance, it is possible to maximize the contribution of ϵ_0 while retaining all other contributions.

$$\delta E = \epsilon_0 \delta \left\langle \Psi \left| \hat{K}_0^* \hat{K}_0 \right| \Psi \right\rangle \tag{117}$$

$$= \epsilon_0 \sum |\psi_i|^2 \left\langle e_i \left| \hat{K}_0^* \hat{K}_0 \right| e_i \right\rangle \tag{118}$$

Therefore, the ϵ_0-contribution must be maximal for the ground state with respect to constraining the other contributions. The same is true for all other contributions as well. Since e_i are eigenvectors of $\hat{K}_0^*\hat{K}_0$, we arrive at equation (118). As mentioned earlier, at least $\frac{(n+1)(n-2m)}{(n-m+2)(n-m+1)} \cdot \dim A^m(V)$ of these summands must be 0. For convenience, let the eigenvalues of $\hat{K}_i^*\hat{K}_i$ be in descending order. Obviously, the contribution is largest if the largest eigenvalue of $\hat{K}_i^*\hat{K}_i$ is the only contributor, i.e., $|\psi_0|^2 = \sum |\psi_i|^2$ and therefore $\delta E = \epsilon_0 \left\langle e_0 \left| \hat{K}_0^* \hat{K}_0 \right| e_0 \right\rangle \delta |\psi_0|^2$. It is therefore reasonable to start the search for the ground state from the eigenvectors of $\hat{K}_i^*\hat{K}_i$.

In the general case where g_i is a geminal, the eigenvectors of $\hat{G}_i^*\hat{G}_i$ are products of the form $g_i \wedge f$ where $f \in A^{m-2}(V)$. This can be deduced from the definition of $\hat{G}_i$. Let $e_j^{(i)}$ be a unit eigenvector of $\hat{G}_i^*\hat{G}_i$ with eigenvalue $\lambda_{j,i}^2$ and let $f_j^{(i)} = \hat{G}_i e_j^{(i)} / \|\hat{G}_i e_j^{(i)}\|$, then $\left| \left\langle g_i \otimes f_j^{(i)} | e_j^{(i)} \right\rangle \right|^2 = \left\langle e_i \left| \hat{G}_i^* \hat{G}_i \right| e_i \right\rangle = \lambda_{j,i}^2$. Again, let the eigenvalues be in descending order.

Starting at $j = 0$, we arrive at equations (119) to (123).

$$\begin{align}
\lambda_{0,i}^2 &= \left|\left\langle g_i \otimes f_0^{(i)} | e_0^{(i)} \right\rangle\right|^2 \tag{119}\\
&= \left|\left\langle g_i \wedge f_0^{(i)} | e_0^{(i)} \right\rangle\right|^2 \tag{120}\\
&\leq \left\|e_0^{(i)}\right\|^2 \cdot \left\|g_i \wedge f_0^{(i)}\right\|^2 \tag{121}\\
&= \left\langle g_i \otimes f_0^{(i)} | g_i \wedge f_0^{(i)} \right\rangle \tag{122}\\
&\leq \left\langle g_i \wedge f_0^{(i)} \left| \hat{G}_i^* \hat{G}_i \right| g_i \wedge f_0^{(i)} \right\rangle / \left\|g_i \wedge f_0^{(i)}\right\|^2 \tag{123}
\end{align}$$

Since, $\lambda_{0,i}^2$ is the largest eigenvalue,

$$\left\langle g_i \wedge f_0^{(i)} \left| \hat{G}_i^* \hat{G}_i \right| g_i \wedge f_0^{(i)} \right\rangle / \left\|g_i \wedge f_0^{(i)}\right\|^2 \leq \lambda_{0,i}^2. \tag{124}$$

Hence, $e_0^{(i)} = g_i \wedge f_0^{(i)} / \left\|g_i \wedge f_0^{(i)}\right\|$, and g_i and $f_0^{(i)}$ are natural states of $e_0^{(i)}$.

We can proceed in the same manner for all other eigenvectors by noting that $\left\langle e_k^{(i)} | e_j^{(i)} \right\rangle = \left\langle g_i \otimes f_k^{(i)} | e_j \right\rangle / \|g_i \wedge f_k^{(i)}\| = 0$ whenever $k < j$. Therefore, $\lambda_{j,i}^2$ is maximal with respect to the space perpendicular to all $f_k^{(i)}$, which allows the extension of the argument for $e_0^{(i)}$ to all $e_j^{(i)}$. So, we see that all eigenvectors of $\hat{G}_i^* \hat{G}_i$ with non-zero eigenvalues are of the form $g_i \wedge f_j^{(i)}$ for a given orthonormal set $\{g_i\}$.

Having asserted the structure of $\hat{G}_i$, it is only natural to consider the relationship between different $\hat{G}_i$. To this end, first consider a permutation of an anti-symmetric eigenfunction $g \wedge f = \mathbf{A}_m(g \otimes f)$ of $\hat{G}$. For a permutation σ of m indices with operation $\hat{P}_\sigma$ on $T^m(V)$, the product $\hat{A}_m \hat{P}_\sigma = \operatorname{sgn} \sigma \cdot \hat{A}_m$, where sgn is -1 if σ has even order and 1 if σ has odd order. Therefore, $\mathbf{A}_m \mathbf{P}_\sigma (g \otimes f) = \operatorname{sgn} \sigma \mathbf{A}_m (g \otimes f)$. Furthermore, any linear combination of permutations of $g \otimes f$ are mapped to a mere multiple of $g \wedge f$. Now let $g \wedge f = \sum c_i g_i \otimes f_i =$

$C\sum_\sigma \operatorname{sgn}\sigma\cdot\mathbf{P}_\sigma(g\otimes f)$, where C is a constant and $g=g_0$, $f=f_0$ and g_i/f_i are natural states of $g\wedge f$. Then for any non-zero $c_ig_i\otimes f_i=\sum_\sigma p_\sigma\mathbf{P}_\sigma(g\otimes f)$ for some set of constants p_σ. Thus, for any non-zero $g_i\otimes f_i$, $g_i\wedge f_i=C_ig\wedge f$ with constant prefactor C_i. Therefore, any two pairs of natural states anti-symmetrize to the same normalized wave function.

Returning to the ground state, let g_i be the natural geminals of the ground state. Then the ground state is defined by the corresponding $\hat{G}_i$ and their eigenvectors (see equation (125)).

$$\hat{H}=\sum_i\left\langle g_i\left|\hat{K}\right|g_i\right\rangle\hat{G}_i^*\hat{G}_i+\sum_{i>j}\left\langle g_i\left|\hat{K}\right|g_j\right\rangle\left(\hat{G}_i^*\hat{G}_j+\hat{G}_j^*\hat{G}_i\right)\tag{125}$$

$$\hat{H}'=\sum_i\left\langle g_i\left|\hat{K}\right|g_i\right\rangle\hat{G}_i^*\hat{G}_i\tag{126}$$

Since the natural states are perpendicular, the ground state has no cross-contributions. Therefore, the ground state is an eigenvector of $\hat{H}'$ in equation (126).

$$\hat{H}\left|\Psi_{gs}\right\rangle=\sum_i\left\langle g_i\left|\hat{K}\right|g_i\right\rangle c_i\hat{G}_i^*\hat{G}_i\left|g_i\wedge f_i\right\rangle=E\left|\Psi_{gs}\right\rangle\tag{127}$$

$$\Psi=\sum c_ig_if_i\tag{128}$$

An obvious solution to $\hat{H}'$ is a solution that is an eigenvector to all $\hat{G}_i^*\hat{G}_j$. Then, the ground state would have to be proportional to some $g\wedge f$. Conversely, the Hamiltonian may be described in terms of $\hat{F}_i:A^m(V)\to A^2(V)$.

$$\hat{H}=\sum\hat{F}_i^*\hat{K}\hat{F}_i\tag{129}$$

Again, the simplest solution to this operator is a simultaneous solution to all $\hat{F}_i^*\hat{K}\hat{F}_i$, such that $\Psi_{gs}=f\wedge g$ for some geminal g and correpsonding $m-2$-electron function.

The preceding discussion shows that the ground state is more dependent on the underlying properties of N-representability than it is dependent on the values of $\mathbf{K}$. The search for the ground state reduces to the search for a set of geminals such that their associated $\hat{G}_i^*\hat{G}_i$ or $\hat{F}_i^*\hat{K}\hat{F}_i$ have simultaneous eigenvectors. Good candidates are the natural geminals of simple products $g \wedge f$. If $g \wedge f$ can be confirmed as the ground state, a powerful necessary condition has been found greatly reducing the complexity of solving the Schrödinger equation.

CHAPTER 5

REDUCED DENSITY MATRICES RECONSIDERED

As the discussion in section 3 of chapter 2 highlights, the structure of the m-RDMs and the 2-RDM, in particular, are of utmost importance. The fact that 1- and $(n-1)$-states are very well defined with respect to each other raises the question what further relationships exist between the natural m-states. The relationships between the natural orbitals (1-states) and natural geminals (2-states) especially warrant exploration since they determine the energy. Restrictions on the natural geminals reduce the set of 2-RDMs which are N-representable and therefore the computational cost.

Let $\{\alpha_i\}$ be the set of natural orbitals of an N-electron wave function. Then we can write the 2-RDM in terms of the natural orbitals (equation (130)).

$$D^{(2)}(12;1'2') = \sum_{i>j,k>l} d_{ij,kl} \left|\alpha_i\alpha_j\right|(12) \left|\alpha_k\alpha_l\right|(1'2') \tag{130}$$

$$D^{(2)}(12;1'2') = D_1^{(2)} + D_2^{(2)} \tag{131}$$

$$D_2^{(2)} = \sum_{i\neq k,j} d_{ij,kj} \left|\alpha_i\alpha_j\right|(12) \left|\alpha_k\alpha_j\right|(1'2') \tag{132}$$

$D_1^{(2)}$ is positive semi-definite and reduces to $D^{(1)}$ and its natural geminals share their natural orbitals with $D_1^{(2)}$. If $D_2^{(2)}$ is to be small compared to $D_1^{(2)}$, as semi-definiteness implies, i.e., $D^{(2)} \approx D_1^{(2)}$, then this condition implies that the natural geminals must all share their natural orbitals with the generating wave function. This circumstance cuts down greatly on the number of N-representable 2-RDMs. Furthermore, the 2-RDM is parameterized by the 1-RDM, i.e., it is possible to choose a 1-RDM, all of which are N-representable, and restrict calculations to positive semi-definite 2-RDMs of the form of $D^{(2)}_1$. Conversely, given a geminal with a set of natural orbitals, it is possible to construct 2-RDMs which reduce to 1-RDMs with the same set of natural orbitals. Hence, geminals parameterize a set of 1- and 2-RDMs. How do the P- and Q-conditions affect the 2-RDM in this representation? The P-matrix is essentially a correction to the diagonal of the 2-RDM. Equation (133) shows that the P-condition lowers the values on the diagonal with respect to $D^{(2)}$. This implies that the diagonal of $D^{(2)}$ must exceed the minimum necessary for positive semi-definiteness given the offdiagonal values of the 2-RDM. The P-condition gives a lower bound on $D^{(2)}$. The Q-condition, on the other hand, imposes an upper bound on $D^{(2)}$.

$$P^{(2)}(12;1'2') = -I(12;1'2') + D^{(2)}(12;1'2') + \sum_{i>j} \left(d^{(1)}_{i,i} + d^{(1)}_{j,j} \right) |\alpha_i \alpha_j|(12)|\alpha_i \alpha_j|(1'2') \tag{133}$$

$$Q^{(2)}(12;1'2') = \sum_{i,j} d^{(1)}_{i,i} \alpha_j(1)\alpha_j(1')\alpha_i(2)\alpha_i(2') - D^{(2)}(1'2;12') \tag{134}$$

In order to characterize the 2-RDM more in terms of its 1-states, let β_i be the respective natural $n-1$-states relative to the natural orbitals α_i, then $\beta_i(2\ldots n) = \sum_j \xi_{ij}\alpha_j(2)\beta_{ij}(3\ldots n)$ and we get

$$\begin{aligned} -\psi =& \mathbf{P}_{12}\psi = \mathbf{P}_{12} \sum_i \xi_i \alpha_i(1)\beta_i(2\ldots n) \\ =& \mathbf{P}_{12} \sum_{ij} \xi_i \xi_{ij} \alpha_i(1)\alpha_j(2)\beta_{ij}(3\ldots n) \end{aligned}$$

$$= \sum_{ij} \xi_i \xi_{ij} \alpha_i(2) \alpha_j(1) \beta_{ij}(3 \ldots n)$$

$$= \sum_{ij} -\xi_i \xi_{ij} \alpha_i(1) \alpha_j(2) \beta_{ij}(3 \ldots n)$$

Therefore, $\psi = \sum_{i>j} \xi'_{ij} |\alpha_i \alpha_j|(12) \beta_{ij}(3 \ldots n)$, where $\xi'_{ij} = \xi_i \xi_{ij} \sqrt{2}$. It follows then, that $D^{n-2} = \sum_{i>j} |\xi'_i|^2 \beta_{ij}(3 \ldots n) \beta_{ij}(3' \ldots n')^*$ due to $|\alpha_i \alpha_j| \perp |\alpha_k \alpha_l|$ for $\{i, j\} \neq \{k, l\}$. If $\{\gamma_i\}$ is a set of orthonormal $n-2$-electron functions, $\beta_{ij} = \sum \gamma_{ijk} \gamma_k$, and $\{\eta_i = \sum_{jk} \xi'_{jk} \gamma_{jki} |\alpha_j \alpha_k|\}$ are geminals, then $f = \sum_i \eta_i(12) \gamma_i(3 \ldots n)$ and $D^{(2)} = \sum_i \eta_i(12) \eta_i(1'2')^*$. The definition of β_{ij} implies that for a given i, that $\{\beta_{ij}\}_j$ is an orthonormal set, which may be chosen as $\{\gamma_j\}_j$. We see that the 1-RDM can thus be reconstructed from a positive definite matrix of geminals which share the 1-states of the full system as 1-states for each 2-state. Therefore, given a 1-RDM, an estimate of the ground state energy may be found by varying over all positive semi-definite matrices which are produced by η_i which share their 1-states with the 1-RDM and reproduce the 1-RDM. This set includes all AGP and SOAGP 2-RDMs.

CHAPTER 6

AUFBAU-ANSATZ FOR GENERALIZED GFT

Encouraged by the previous chapter, applying the *aufbau* principle of Hartree-Fock theory generalizes the AGP *ansatz.*

$$\Psi_{Aufbau} = \mathbf{A_n} \prod_{i=1}^{n/2} g_i(2i-1, 2i) \tag{135}$$

where g_i are geminals. Some work has been done on functions for which the g_i are strongly orthogonal.[38] Such functions are inexpensively computed and have been shown to describe bonding well.[38] When coupled with perturbation theory, an accuracy *en par* with CCSD can be attained.[60] In general, SOAGP is expected to produce very good results, when electron pairs interact very weakly. While Hartree-Fock functions are AGP functions as well as anti-symmetrized products of strongly orthogonal geminals (SOAGP), this is generally not the case for SOAGP nor for the generalized AGP (GAGP) *ansatz.*

An interpretation of the meaning of the geminals may be derived from ordering them with respect to $\left\langle g_i \left| \hat{K}\hat{D}^{(2)}_{Aufbau} \right| g_i \right\rangle$. In keeping with Hartree-Fock theory, the lowest energy geminals can be determined to describe the core electrons while increasing energy terms describe increased mixing with the valence electrons and therefore explain, for instance, bonding.

The most general case does not allow for a direct description of the 2-RDM but it is possible to build the 2-RDM iteratively as follows: Let ψ_p be anti-symmetric and let $\mathbf{D_p^{(m)}}$

be its m-RDM with integration kernel $D_p^{(m)}$.

$$\psi_{p+2} = \mathbf{A_{p+2}} g(1,2)\psi_p(3,\dots,p+2) \tag{136}$$

$$\overline{\psi}_{p+2} = g(1,2)\psi_p(3,\dots,p+2) \tag{137}$$

Since $(p+2)(p+1)\mathbf{A_{p+2}} = \left(1 - \sum_{i=2}^{p+2} \mathbf{P_{1i}}\right)\left(1 - \sum_{j=3}^{p+2} \mathbf{P_{2j}}\right)\mathbf{A_p^{(2)}}$, where $\mathbf{P_{ij}}$ is the permutator of particles i and j and $\mathbf{A_l^{(k)}}$ is the anti-symmetrizer of l particles starting at $k+1$,

$$\begin{aligned}(p+2)(p+1)A_{p+2}\overline{\psi}_{p+2} &= \left(1 - \sum_{i=2}^{p+2} \mathbf{P_{1i}}\right)\left(1 - \sum_{j=3}^{p+2} \mathbf{P_{2j}}\right)\overline{\psi}_{p+2} \\ &= \left(2 - 2\sum_{i=3}^{p+2}[\mathbf{P_{1i}} + \mathbf{P_{2i}}] + \sum_{i \neq j > 2} \mathbf{P_{1i}}\mathbf{P_{2j}}\right)\overline{\psi}_{p+2}.\end{aligned} \tag{138}$$

where we have used the fact that g and ψ_p are antisymmetric with respect to elementary permutation of their particles. Hence, there are six contributions to the 2-RDM (Eqn. (139)-(145) on the following page).

$$D_{p+2}^{(2)} = F_1 + (F_2 + F_2^*) + (F_3 + F_3^*) + F_4 + (F_5 + F_5^*) + F_6 \tag{139}$$

$$\mathbf{F_1} = \mathbf{g(12)g(1'2')}\|\boldsymbol{\psi_p}\|^{\mathbf{2}} \tag{140}$$

$$F_2 = -p\int g(12)g^*(1'3)D_p^{(1)}(3;2') - g(12)g^*(2'3)D_p^{(1)}(3;1')d3 \tag{141}$$

$$F_3 = \binom{p}{2}\int g(12)g^*(34)D_p^{(2)}(34;1'2')d3d4 \tag{142}$$

$$\begin{aligned}
F_4 =& 2\binom{p}{2}\int g(13)g^*(2'4)D_p^{(2)}(24;1'3) - g(23)g^*(2'4)D_p^{(2)}(14;1'3)d3d4+\\
&2\binom{p}{2}\int g(13)g^*(1'4)D_p^{(2)}(24;2'3) - g(23)g^*(1'4)D_p^{(2)}(14;2'3)d3d4+\\
&p\left(D_g^{(1)}(1,1')D_p^{(1)}(2;2') + D_g^{(1)}(2,2')D_p^{(1)}(1;1')\right) -\\
&p\left(D_g^{(1)}(2,1')D_p^{(1)}(1;2') + D_g^{(1)}(1,2')D_p^{(1)}(2;1')\right)
\end{aligned} \tag{143}$$

$$\begin{aligned}
F_5 =& 2\binom{p}{2}\int D_g^{(1)}(1;4)D_p^{(2)}(24;1'2') - D_g^{(1)}(2;4)D_p^{(2)}(14;1'2')d4\\
&- 6\binom{p}{3}\int g(13)g^*(45)D^{(3)}(245;1'2'3)d3d4d5\\
&+ 6\binom{p}{3}\int g(23)g^*(45)D^{(3)}(145;1'2'3)d3d4d5
\end{aligned} \tag{144}$$

$$\begin{aligned}
F_6 =& \binom{\mathbf{p}}{\mathbf{2}}\mathbf{D_p^{(2)}(12;1'2')\|g\|^2}+\\
&2(p-2)\binom{p}{2}\int D_g^{(1)}(4;5)D_p^{(3)}(125;1'2'4)d3d4d5+\\
&\binom{p-2}{2}\binom{p}{2}\int g(34)g^*(56)D_p^{(4)}(1256;1'2'34)d3d4d5d6
\end{aligned} \tag{145}$$

Equation (140) on the page before contains the geminal's direct contribution to the 2-RDM. The cofactor is large for a single geminal on the diagonal. The computation is of the order $O(k^4)$, in the size of the basis set (k), for setting up the product of two geminals represented as vectors in the vector space acted on by $D^{(2)}$. In equation (141) on the preceding page the connections between those parts of $D_p^{(1)}$ which are not strongly orthogonal to g are projected out. On the other hand, equation (142) on the page before reinserts the gg^* components already in $D_p^{(2)}$ and weighs them accordingly. If gg^* is already a non-zero component in $D_p^{(2)}$, it will not be diminished. The computation reduces to simple matrix-matrix and matrix-vector multiplications, so the order of computation is $O(k^4)$ for the setup of the matrix elements. Equations (143) on the preceding page and (144) on the page before add crossproducts missed by equations (141) and (142) on page 58. Unlike before, all the entries of the density matrix need to be considered and the order of computation is $O(k^6)$. $D_p^{(2)}$ is retained and weighed accordingly in equation (145) on the page before. This guarantees that large contributions of previous iterations will remain large in the next iteration.

Equations (144) on the preceding page and (145) on the page before introduce problematic higher-order contributions. Since $g(34) = \sum_i \xi_i \alpha_i(3)\beta_i(4)$,

$$\int g(34)g(35)d3 = \sum_i |\xi_i|^2 \beta_i(4)\beta_i(5).$$

Also,

$$D^{(2)}(12;34) = \sum_i \int D^{(3)}(125;346)\beta_i(5)\beta_i(6)d5d6.$$

Due to $\sum_i |\xi_i|^2 = 1$ and $|\xi_i|^2 \leq 1/2$, we can conclude

$$\begin{aligned} \frac{1}{2}D_p^{(2)}(12;1'2') &\geq \int g(34')^* g(3'4')D_p^{(3)}(123;1'2'3')d3d3'd4' \\ &= \int g(34')^* g(3'4')\psi_p(1234\ldots p)\psi_p(1'2'3'4\ldots p)^* d3'd4'd3\ldots dp. \end{aligned} \tag{146}$$

By a similar argument:

$$\begin{aligned} \frac{1}{p}D_p^{(2)}(12;1'2') &\geq \int g(34)^* g(3'4')D_p^{(4)}(1234;1'2'3'4')d3d4d3'd4' \\ &= \int g(34)^* g(3'4')\psi_p(12345\ldots p)\psi_p(1'2'3'4'5\ldots p)^* d3d4d3'd4'd5\ldots dp. \end{aligned} \tag{147}$$

If $g_{k+1} \perp g_i \forall i \leq k$, then the overlap is proportional to $\int g_i(1,2)g_j(3,4)g_{k+1}(2,3)d2d3$ for $i < j < k+1$ because the g_l are pairwise orthogonal and $\psi_p = \mathbf{A_{2k}} \prod_{i=1}^k g_i = \sum_\pi x_\pi \prod g_i(\pi_{2i-1}\pi_{2i})$, where π is a permutation of $2k$ particles and x_π is a prefactor. Since $\int g_i(1,2)g_j(2,1)d1d2 = 0$,

$\int g_i(1,2)g_j(2,3)d2$ is composed of small contributions of variable sign, which can be added as corrections for equations (146) on the preceding page and (147) on the page before, e.g., by fitting to $D_p^{(2)}$. In zeroth order approximation, they may also be ignored since cancellation is very likely. In general the necessary positive semi-definiteness of $D^{(2)}$, $P^{(2)}$, and $Q^{(2)}$ may be enforced to correct for omissions, for instance, by enforcing the structure of $D_1^{(2)}$ of equation (131) on page 54 in chapter 2, section 3. There is also a necessary condition on the trace of the final 2-RDM.

$$\begin{aligned}\int |\psi_{p+2}|^2 d1 \ldots dp+2 =& \operatorname{tr} \hat{D}^{(2)}_{p+2}\\ =& 2\|g\|^2\|\psi_p\|^2 - 2p \operatorname{tr}\left(\hat{D}^{(1)}_g \hat{D}^{(1)}_p\right) + \\ & \binom{p}{2}\int g(1,2)D^{(2)}_p(3,4;1,2)g(3,4)d1d2d3d4\end{aligned} \tag{148}$$

Since there are $n/2$ steps of iteration and equation (143) on page 59 is the computational bottleneck, the overall order of computational complexity is $O(\frac{n}{2}k^6)$, where k is the number of basis functions and n is the number of electrons.

We may cast the same problem using the langugage of second quantization to derive more insight into the higher-order contributions. We recall that the matrix elements of the 2-RDM are $\left\langle \Psi \left| \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l \right| \Psi \right\rangle$. Identifying each generating geminal with an operator such that $|g_i\rangle = \hat{g}_i^\dagger |1\rangle = \sum g_{kl}^{(i)*} \hat{a}_k^\dagger \hat{a}_l^\dagger |1\rangle$, these matrix elements are $\left\langle 1 \left| \hat{g}_1 \cdots \hat{g}_{n/2} \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_k \hat{a}_l \hat{g}_{n/2}^\dagger \cdots \hat{g}_1^\dagger \right| 1 \right\rangle$. Using the commutation relations for creation and annihilation operators, the operators may be passed through the geminal operators (see equation (152)).

$$\hat{a}_\nu \hat{g}_i^\dagger = \hat{g}_i^\dagger \hat{a}_\nu + \hat{g}_i^{(\nu)\dagger} \tag{149}$$

$$\hat{g}_i^{(\nu)\dagger} = \sum_{j<\nu} g_{\nu j}^{(i)*} \hat{a}_j^\dagger - \sum_{j>\nu} g_{j\nu}^{(i)*} \hat{a}_j^\dagger \tag{150}$$

$$\hat{a}_\nu \prod \hat{g}_i^\dagger = \prod \hat{g}_i^\dagger \hat{a}_\nu + \sum \hat{g}_\nu^{(i)\dagger} \prod_{j\neq i} \hat{g}_j^\dagger \tag{151}$$

$$\begin{aligned}\hat{a}_\nu \hat{a}_\mu \prod \hat{g}_i^\dagger =& \prod \hat{g}_i^\dagger \hat{a}_\nu \hat{a}_\mu + \\ & \sum \hat{g}_i^{(\nu)\dagger} \prod_{j\neq i} \hat{g}_i^\dagger \hat{a}_\mu + \sum \hat{g}_i^{(\mu)\dagger} \prod_{j\neq i} \hat{g}_i^\dagger \hat{a}_\nu + \\ & \sum g_{\nu\mu}^{(i)*} \prod_{j\neq i} \hat{g}_j^\dagger - \sum \hat{g}_i^{(\nu)\dagger} \hat{g}_j^{(\mu)\dagger} \prod_{k\notin\{i,j\}} \hat{g}_k^\dagger\end{aligned} \tag{152}$$

Applying equation (152) to the vacuum $|1\rangle$ eliminates the annihilation operators leaving only creation operators (see equation (153))

$$\begin{aligned}\hat{a}_\nu \hat{a}_\mu \left| \bigwedge g_i \right\rangle &= \hat{a}_\nu \hat{a}_\mu \prod \hat{g}_i^\dagger \left|1\right\rangle \\ &= \sum g_{\nu\mu}^{(i)*} \prod_{j\neq i} \hat{g}_i^\dagger \left|1\right\rangle - \sum \hat{g}_i^{(\nu)\dagger} \hat{g}_j^{(\mu)\dagger} \prod_{k\notin\{i,j\}} \hat{g}_k^\dagger \left|1\right\rangle\end{aligned} \tag{153}$$

The expansion of the matrix elements in second quantization shows that no more than two geminals are directly involved in the 2-RDM. The first and last terms of equation (154) on the next page are derived from positive, semi-definite operators, while the second and third term are derived from a negative, semi-definite operator.

(154)

$$\begin{aligned}\left\langle \bigwedge g_i \left| \hat{a}_r^\dagger \hat{a}_s^\dagger \hat{a}_k \hat{a}_l \right| \bigwedge g_i \right\rangle = &\left\langle 1 \left| \left(\sum_i \left(\prod_{j \neq i} \hat{g}_j \right) g_{rs}^{(i)} \right) \left(\sum_i g_{kl}^{(i)*} \prod_{j \neq i} \hat{g}_j^\dagger \right) \right| 1 \right\rangle - \\ &\left\langle 1 \left| \left(\sum_i \left(\prod_{j \neq i} \hat{g}_j \right) g_{rs}^{(i)} \right) \left(\sum_{i \neq j} \hat{g}_i^{(k)\dagger} \hat{g}_j^{(l)\dagger} \prod_{\kappa \notin \{i,j\}} \hat{g}_\kappa^\dagger \right) \right| 1 \right\rangle - \\ &\left\langle 1 \left| \left(\sum_{i \neq j} \left(\prod_{\kappa \notin \{i,j\}} \hat{g}_\kappa \right) \hat{g}_j^{(s)} \hat{g}_i^{(r)} \right) \left(\sum_i g_{kl}^{(i)*} \prod_{j \neq i} \hat{g}_j^\dagger \right) \right| 1 \right\rangle + \\ &\left\langle 1 \left| \left(\sum_{i \neq j} \left(\prod_{\kappa \notin \{i,j\}} \hat{g}_\kappa \right) \hat{g}_j^{(s)} \hat{g}_i^{(r)} \right) \left(\sum_{i \neq j} \hat{g}_i^{(k)\dagger} \hat{g}_j^{(l)\dagger} \prod_{\kappa \notin \{i,j\}} \hat{g}_\kappa^\dagger \right) \right| 1 \right\rangle\end{aligned}$$

The first term incorporates only components of 2-RDMs of smaller products of the g_i. The last term collects second through sixth order RDMs of smaller products of g_i. The middle terms hold corrections to second, third, and fourth order of the 2-RDM. Since the second-order corrections are already taken care of in the iterative scheme, the majority of the third- and fourth-order corrections are expected to be related to retaining the D-, P- and Q-positivity conditions.

1. Choice of $\mathbf{g_i}$

There are several constraints that can be loosened progressively going from an AGP wave function to a generalized AGP function:

a) Release of ξ_i only, i.e.,

$$g_j = \sum \xi_i^{(j)} \left| \alpha_{2i-1} \alpha_{2i} \right|$$

b) Releasing the ordering of the natural orbitals.

$$g_i = \sum \xi_j^{(i)} \left| \beta_{2j-1}^{(i)} \beta_{2j}^{(i)} \right|$$

$$\left\{ \beta_j^{(i)} \right\}_j = \left\{ \beta_l^{(k)} \right\}_l \forall i, k$$

c) Increasing the number of used geminals to l geminals.

$$\Psi = \mathbf{A_n} \prod_{i=1}^{l} \Psi_{\mathrm{AGP}}^{(m_i)}[g_i] \left(M_i + 1 \ldots M_{i+1} \right)$$

$$\sum_{1 \leq j < i} m_j = M_i, \ \sum_{i=1}^{l} m_i = n$$

In the first case, the 2-RDM retains the form for an AGP function as well as its natural orbitals expanding the set of N-representable 2-RDMs for geminal functional theory.[48] This approach also encompasses all SOAGP functions via the constraint $\xi_\nu^{(i)^*} \xi_\nu^{(j)} = \delta_{ij} |\xi_\nu^{(i)}|^2$.

The second case potentially encompasses more off-diagonal entries but does not necessarily retain the natural orbitals. This should lead to simpler expressions for the 2-RDM with respect to the troubles encountered in equations (144) and (145) on page 59. Since $\psi = \sum_\pi x_\pi \prod_{i=1}^{n/2} g(\pi_{2i-1}\pi_{2i})$ where π is a permutation of n particles, the results of chapter 5 on page 54 ($D^{(2)} \approx D_1^{(2)}$) justify this approach due to $g_i(12) \sum_{\pi'} \prod_{j \neq i} g_j(\pi'_{2j-1}\pi'_{2j})$ being a summand of ψ, with π' a permutation in $n-2$ particles.

In the third case a wide variety of density matrices are included covering the former two cases. One choice of geminals are the eigengeminals of $\mathbf{K}$ with the lowest eigenvalues. Equation (140) on page 58 suggests a large coefficient for these geminals and hence a low energy. Another consideration may be geminals with extended orthogonality conditions, for instance $\hat{g}_i \hat{g}_j \hat{g}_k = 0$ whenever the three geminals have different labels. This last condition considerably cuts down on the contributions for consideration in third and fourth order RDMs.

CHAPTER 7

4-ELECTRON SYSTEMS WITH GAGP

In order to verify the applicability of GAGP, several 4-electron systems were investigated. Four-electron systems do not suffer from approximations to higher-order density matrices, i.e., setting the higher-order contributions to zero in equations (144) to (145) on page 59 is not an approximation. Thus, we can examine two conjectures: for one, we can distinguish between GAGP and AGP; secondly, we will see whether it is fair to say that the ground state is a GAGP function.

The FCI and HF energies were converged to 10^{-8} Hartree using the PSI3 program suite.[61] The energy of $g_1 \wedge g_2$ is a bilinear functional in each geminal of $E : A^2(V) \times A^2(V) \to \mathbb{R}$ according to $(x, y) \mapsto \left\langle x \wedge g \middle| \hat{H} \middle| y \wedge g \right\rangle$, where g is either g_1 or g_2, because $\wedge$ is linear. The same is true of the overlap $\left\langle x \wedge g \middle| \hat{O} \middle| y \wedge g \right\rangle = \langle x \wedge g | y \wedge g \rangle$. It follows that the eigenstates of the corresponding homomorphism $\hat{H}'$ of $A^2(V)$ are exact solutions subject to the stationarity of g and unit overlap. Therefore, in the case of GAGP, the energy was optimized as follows:

a) Start the calculation from the HF guess via $g_1 = |\phi_1 \phi_2|$ and $g_2 = |\phi_3 \phi_4|$.

b) Compute $\hat{H}_1'$ and the overlap matrix $\hat{O}_1$, with $g = g_1$ in above definitions of the operators.

c) Solve the eigenvalue problem for $\hat{O}_1$ and transform $\hat{H}_1'$ by $\hat{O}_1^{-\frac{1}{2}}$.

d) Solve the eigenvalue problem for $\hat{O}_1^{-\frac{1}{2}} \hat{H}_1' \hat{O}_1^{-\frac{1}{2}}$ and set g_2 to the lowest eigenvalue-solution. That eigenvalue is the energy of $g_1 \wedge g_2$.

e) Repeat b through d for $g = g_2$.

f) Repeat b through e until convergence, i.e., the change of energy from one iteration to the next is below 10^{-8}.

This approach neglects the dynamic interactions between the geminals but guarantees decreasing energies and, while convergence for each geminal is cubic, the overall convergence falls short of such an optimistic estimate. Rather, it is linear at best.

As table 1 shows, the GAGP functions recover the electron correlation almost exactly. As we progress from Li^- to B^+ we see a decrease of overlap of the two generating geminals demonstrating the departure from a pure AGP function and we see the first conjecture confirmed. This may be attributed to the core electrons being bound stronger to the nucleus resulting in a separation of core and valence electrons.

For the dissociated species of LiH, the percentage of retrieved correlation is comparatively low. This is due to numerical inaccuracies of the convergence method. The absolute error of electron correlation using GAGP is on the order of 10^{-4} for bound LiH, while it is 10^{-5} for dissociated LiH. In bound LiH, the electron residing on the hydrogen atom correlates strongly with the lithium electrons, which leads to a large overall correlation energy of $3.109065 \cdot 10^{-2}$ Hartree. In dissociated LiH, we see very little correlation, a mere $2.1699 \cdot 10^{-4}$ Hartree, in the first place, probably due to the disruption of relocating just one electron from lithium to hydrogen. For all other hydrogen abstractions, there is considerable electron correlation which makes LiH and BeH^+ rather unique in this series.

Contrary to the trend in atoms, with increased charge the AGP character increases for LiH and BeH^+. Protonating Li^- and Be results in more postive charge, which attracts the

TABLE 1. FCI, GAGP, HF energies and the percentage of electron correlation (EC) recovered by GAGP of various 4-electron atoms and molecules using the cc-pVDZ basis set. The energies were converged to 10^{-8} Hartree.

Species	E_{FCI}	E_{GAGP}	E_{HF}	% EC GAGP	$\langle g_1 \vert g_2 \rangle$
Li^{-}	-7.44785514	-7.44779627	-7.41681880	99.81	0.983
Be	-14.61740950	-14.61732472	-14.57233761	99.81	0.964
B^{+}	-24.29384952	-24.29376828	-24.23456235	99.86	0.689
LiH[a]	-8.01477504	-8.01467655	-7.98368439	99.68	0.901
$Li\cdots H$[b]	-7.93191592	-7.93190893	-7.93169893	96.77	0.687
BeH^{+}[a]	-14.88484983	-14.88453406	-14.84960071	99.10	0.999
$(Be\cdots H)^{+}$[b]	-14.77529442	-14.77527156	-14.77472034	96.02	0.867
He_2 at eq.[a]	-5.77519593	-5.77519072	-5.71032168	99.99	$3.14 \cdot 10^{-6}$
He_2 at ∞[b]	-5.77518966	-5.77518966	-5.71032095	100	$3.83 \cdot 10^{-13}$
HeH^{-}[a]	-3.35750027	-3.35749691	-3.30402998	99.99	$9.56 \cdot 10^{-6}$
$(He\cdots H)^{-}$[b]	-3.35745161	-3.35745161	-3.30398420	100	$3.38 \cdot 10^{-13}$
linear H_3^{-}[a]	-1.64393530	-1.63425811	-1.57545074	85.87	0.056
lin. $(H_2\cdots H)^{-}$[b]	n/a	-1.59624834	-1.53310194	n/a	$7.5 \cdot 10^{-15}$

[a] CCSD geometry

[b] Distance was 10^9 Bohr.

electrons more through Coulombic interaction, but unlike in the atoms, the charge is less localized, which apparently reduces the separation of electron pairs. So, going from LiH to BeH^{+}, at either distance, results in a marked increase of AGP character.

As expected, the helium dimer is very weakly bound and shows, both at "infinite" distance as well as at equilibrium distance, decidedly non-AGP character. It is not surprising that GAGP does well for the helium dimers, since each helium atom can be described by a geminal exactly. The very small interactions between the two at long distances essentially turn the problem into an SOAGP problem. Although the absolute energies show very little variation, the large nuclear repulsions are offset by strong electronic interactions, which does not *a priori* imply the validity of an SOAGP approach. Yet again, the divergence from AGP is solidified. While ambiguity of the generating geminals may be blamed for the departure from AGP in the case of the atoms, i.e., there might be a g such that $g \wedge g = g_1 \wedge g_2$, there is no room for such an argument in the case of the non-interacting helium atoms. The helium is decidedly not a mere AGP function, but a SOAGP function.

No FCI energy could be found for "infinitely" separated H_3^-, but the sheer amount of electron correlation predicted by GAGP and the minute overlap found suggest 100% recovery. The sub-par performance for bound H_3^- may be attributed to insufficient convergence since the calculation took more than 300 steps while other calculations finished in under 100 steps.

With the exception of linear H_3^-, the GAGP function recovers the electron-correlation energy quantitatively. This datum is strong evidence for the second conjecture, that the ground state is a GAGP function for 4-electron systems. Extending to 2^n-electron systems is natural as each follow the same mathematics. Given a $2m$ electron Hamiltonian, an operator $\mathbf{K}_m$ analogous to $\mathbf{K} =: \mathbf{K}_2$ may be constructed, and thus a product of m-electron functions should result in the correct answer. Then again those functions can be broken down into further products, if m is divisible by two. Thence, we arrive at the important result that GAGP describes 2^n-electron systems exactly.

CHAPTER 8

CONCLUSIONS

Generalized AGP uncovers a new side to quantum chemistry. Extending the idea of antisymmetric geminal products, the *aufbau ansatz* allows for a flexible construction of N-representable 2-RDMs, which allow for high-quality calculations of energies and related properties. As the complete wave function is not explicitly used, the computational effort remains polynomial (roughly $O(n^7)$), if the basis set size is linearly dependent on the number of electrons. The numerical examples of chapter 7 have confirmed the conjectures of chapter 4. For 4-electron systems, the electron-correlation energy is quantitatively recovered. Hence it is reasonable to assume that the ground state is of the form in equation (155).

$$\Psi = g(1,2) \wedge \phi(3, \ldots, n) \tag{155}$$

Although GAGP suffers from approximations if there are more than 4 electrons, they may be introduced to compute the 2-RDM nonetheless. The P-, Q- and D-positivity conditions must be enforced for any approximation to be successful. In general GAGP does not collapse to AGP justifying the additional computational effort. Similar to traditional Hartree-Fock theory, the generating geminals may be interpreted like molecular orbitals yielding additional insight, e.g., into the nature of bonds. As the separated-helium-dimer example demonstrates, each geminal represents the non-interacting helium atoms, nicely illustrating the fact that GAGP, as well as AGP, are inherently multi-determinantal. Therefore, problems associated with symmetry breaking and size-extensivity are naturally avoided. The investigation of GAGP and 2-RDMs has revealed further methods with possible restrictions that allow the computation of systems beyond four electrons. A matter that has not been investigated is the use of domain decomposition techniques like finite or spectral elements. In such approaches only $D_1^{(2)}$ or even the diagonal of the 2-RDM and continuity conditions on the 1-RDM are needed, which simplifies the N-representability problem as well as the computation of these entities for GAGP virtually eliminating the need for higher-order RDMs. Thus, domain decomposition methods are worthwile for future research on the subject of GAGP.

APPENDIX A

CD Contents

On the CD there is a copy of the `C++` template library for linear algebra with sparse implementations. Also there is a library for RHF, UHF, AGP functions, GAGP functions and their respective density matrices. All code is replete with documentation found in the directory `2dft.iso/html/index.html`. After copying the code, it can be compiled by executing `make` in its root directory. The file `Makefile` contains the `PSI` variable, which needs to be set to the location of PSI 3.0. Furthermore, various options may be set in `Makefile` for specific optimizations, etc. Since no integrals are computed by the programs a PSI 3.0 installation has to be present. The `input.dat` file must contain all necessary molecular specifications as well as the basis set. Running `input` and `cints` from the PSI suite of programs prepares the job for executing any of the commands on the CD.

Publication History

[1] B. C. Rinderspacher and P. R. Schreiner, *An* Aufbau-Ansatz *for Geminal Functional Theory*, submitted.

[2] B. C. Rinderspacher and P. R. Schreiner, *Geminal Functional Theory*, in *Electronic Encyclopedia of Computational Chemistry*, edited by P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner, John Wiley & Sons, New York, 2005 Accepted.

[3] B. C. Rinderspacher and P. R. Schreiner, *An* Aufbau-Ansatz *for Geminal Functional Theory*, Poster at the Handy Conference at Cambridge University, 2004.

[4] A. de Meijere, A. F. Khlebnikov, S. I. Kozhushkov, K. Miyazawa, D. Frank, P. R. Schreiner, B. C. Rinderspacher, D. S. Yufit, and J. A. K. Howard, *A Convergent Route to Enantiomerically Pure Higher [$n-2$]Triangulanedimethanol Derivatives and [n]Triangulanes ($n \geq 7$)*, in *Angew. Chem.-Int. Edit.*, 2004, **43**(47), pp. 6391 – 6391.

[5] B. C. Rinderspacher and P. R. Schreiner, *Structure-Property Relationships of Prototypical Chiral Compounds: Case Studies*, in *J. Phys. Chem. A*, 2004, **108**(15), pp. 2867–2870.

[6] B. C. Rinderspacher and P. R. Schreiner, *Optimizing the Energy for Geminal Functional Theory*, Poster at the CSGF Conference, 2004.

[7] P. R. Schreiner, A. A. Fokin, O. Lauenstein, Y. Okamoto, T. Wakita, B. C. Rinderspacher, G. H. Robinson, J. K. Vohs, and C. F. Campana, *Pseudotetrahedral Polyhaloadamantanes as Chirality Probes: Synthesis, Separation, and Absolute Configuration*, in *J. Am. Chem. Soc.*, 2002, **124**(45), pp. 13348–13349.

[8] A. de Meijere, A. F. Khlebnikov, S. I. Kozhushkov, R. R. Kostikov, P. R. Schreiner, A. Wittkopp, B. C. Rinderspacher, H. Menzel, D. S. Yufit, and J. A. K. Howard, *The First Enantiomerically Pure*

[n]Triangulanes and Analogues: σ-[n]Helicenes with Remarkable Features, in *Chem.-Eur. J.*, 2002, **8**(4), pp. 828–842.

[9] B. C. Rinderspacher and P. R. Schreiner, *Basis-Set Dependence of ORD- and CD-Spectra*, Poster at CSGF Conference, 2002.

[10] S. I. Kozhushkov, R. R. Kostikov, A. P. Molchanov, R. Boese, J. Benet-Buchholz, P. R. Schreiner, C. Rinderspacher, I. Ghiviriga, and A. de Meijere, *Tetracyclopropylmethane: A Unique Hydrocarbon with S-4 Symmetry*, in *Angew. Chem.-Int. Edit.*, 2001, **40**(1), pp. 180–183.

[11] B. C. Rinderspacher and P. R. Schreiner, *Basis Set Dependence of ORD- and CD-Spectra.*, in *Abstr. Pap. Am. Chem. Soc.*, 2001, **222**, pp. U405–U405.

[12] B. C. Rinderspacher, *Quantum Chemistry of Chiral Systems*, Master's thesis, Georg-August-Universität Göttingen, 2000.

Bibliography

[1] D. R. Hartree, *Proc. Camb. Phil. Soc.*, 1928, **24**, 89–110.

[2] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Publications Inc., New York, 1996.

[3] J. Simons, *J. Phys. Chem.*, 1991, **95**, 1017–1029.

[4] F. Neese, *J. Chem. Phys.*, 2003, **119**, 9428–9443.

[5] J. Reinhold, *Quantentheorie der Moleküle*, B. G. Teubner, Stuttgart, 1994.

[6] S. J. Kim, P. R. Schreiner, P. V. Schleyer, and H. F. Schaefer, *J. Phys. Chem.*, 1993, **97**, 12232–12238.

[7] P. R. Schreiner, S. J. Kim, H. F. Schaefer, and P. V. Schleyer, *J. Chem. Phys.*, 1993, **99**, 3716–3720.

[8] E. Schrödinger, *Ann. Physik*, 1926, **79**, 361–376.

[9] J. J. Sakurai, *Modern Quantum Mechanics*, Addison-Wesley Publishing Company, 1994, revised ed.

[10] A. P. French and E. F. Taylor, *An Introduction to Quantum Physics*, W. W. Norton, New York, 1978.

[11] T. Yokonuma, *Tensor Spaces and Exterior Algebra*, American Mathematical Society, Providence, R.I., 1992.

[12] W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Willey & Sons, New York, 1986.

[13] M. J. Frisch, J. A. Pople, and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.

[14] T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[15] T. H. Dunnning, K. A. Peterson, and D. Woon, *Encyclopedia of Computational Chemistry*, John Wiley & Sons, New York, 1998 p. 88.

[16] L. Adamowicz and J. P. Malrieu, *J. Chem. Phys.*, 1996, **105**, 9240–9247.

[17] D. Cremer, in *Encyclopedia of Computational Chemistry*, eds. P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner, John Wiley & Sons, New York, 1998 pp. 1706–1735.

[18] C. M. ller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618.

[19] F. Coester, *Nucl. Phys.*, 1958, **7**, 421–424.

[20] F. Coester and H. Kummel, *Nucl. Phys.*, 1960, **17**, 477–485.

[21] J. Cizek, *J. Chem. Phys.*, 1966, **45**, 4256–&.

[22] D. Mukherjee, R. K. Moitra, and A. Mukhopadhyay, *Mol. Phys.*, 1975, **30**, 1861–1888.

[23] I. Lindgren, *Int. J. Quantum Chem.*, 1978, pp. 33–58.

[24] J. P. Malrieu, P. Durand, and J. P. Daudey, *J. Phys. A-Math. Gen.*, 1985, **18**, 809–826.

[25] U. S. Mahapatra, B. Datta, B. Bandyopadhyay, and D. Mukherjee, *Adv. Quant. Chem.*, 1998, **30**, 163–193.

[26] P. Hohenberg and W. Kohn, *Phys. Rev. B*, 1964, **79**, 361–376.

[27] P. A. M. Dirac, *Proc. Camb. Phil. Soc.*, 1930, **26**, 376–385.

[28] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

[29] E. Wigner, *Trans. Faraday Soc.*, 1938, **34**, 678–685.

[30] P. M. W. Gill, in *Encyclopedia of Computational Chemistry*, eds. P. von Ragué Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schaefer, and P. R. Schreiner, John Wiley & Sons, New York, 1998 pp. 678–689.

[31] T. G. Wright, *J. Chem. Phys.*, 1996, **105**, 7579–7582.

[32] Z. L. Cai, K. Sendt, and J. R. Reimers, *J. Chem. Phys.*, 2002, **117**, 5543–5549.

[33] E. Schmidt, *Math. Annalen*, 1907, **63**, 443.

[34] B. C. Carlson and J. M. Keller, *Phys. Rev.*, 1961, **121**, 659.

[35] A. J. Coleman, *J. Math. Phys.*, 1965, **6**, 1425–1431.

[36] A. C. Hurley, J. Lennard-Jones, and J. A. Pople, *Proc. R. Soc. London Ser. A*, 1953, **220**, 446–455.

[37] D. M. Silver, E. L. Mehler, and K. Ruedenberg, *J. Chem. Phys.*, 1970, **52**, 1174–1180.

[38] V. A. Rassolov, *J. Chem. Phys.*, 2002, **117**, 5978–5987.

[39] T. Arai, *J. Chem. Phys.*, 1960, **33**, 95–98.

[40] D. M. Silver, K. Ruedenberg, and E. L. Mehler, *J. Chem. Phys.*, 1970, **52**, 1181–1205.

[41] D. M. Silver, K. Ruedenberg, and E. L. Mehler, *J. Chem. Phys.*, 1970, **52**, 1206–1227.

[42] L. A. Curtiss, P. C. Redfern, K. Raghavachari, and J. A. Pople, *J. Chem. Phys.*, 1998, **109**, 42–55.

[43] L. A. Curtiss, K. Raghavachari, P. C. Redfern, and J. A. Pople, *J. Chem. Phys.*, 1997, **106**, 1063–1079.

[44] E. Rosta and P. R. Surjan, *Int. J. Quantum Chem.*, 2000, **80**, 96–104.

[45] E. Rosta and P. R. Surjan, *J. Chem. Phys.*, 2002, **116**, 878–890.

[46] A. J. Coleman and V. I. Yukalov, *Reduced Density Matrices: Coulson's Challenge*, Springer, Berlin, 2000.

[47] A. J. Coleman, *Int. J. Quantum Chem.*, 1997, **63**, 23–30.

[48] D. A. Mazziotti, *J. Chem. Phys.*, 2000, **112**, 10125–10130.

[49] D. W. Smith, *Phys. Rev.*, 1966, **147**, 896.

[50] T. Juhasz and D. A. Mazziotti, *J. Chem. Phys.*, 2004, **121**, 1201–1205.

[51] T. Juhasz and D. A. Mazziotti, *J. Chem. Phys.*, 2005, **122**.

[52] G. Gidofalvi and D. A. Mazziotti, *Chem. Phys. Lett.*, 2004, **398**, 434–439.

[53] G. Gidofalvi and D. A. Mazziotti, *J. Chem. Phys.*, 2005, **122**.

[54] C. G. Broyden, *Math. Comput.*, 1970, **24**, 365–&.

[55] R. Fletcher, *Comput. J.*, 1970, **13**, 317–&.

[56] D. Goldfarb, *Math. Comput.*, 1970, **24**, 23–&.

[57] D. F. Shanno, *SIAM J. Numer. Anal.*, 1970, **7**, 366–&.

[58] J. M. Bofill, *Int. J. Quantum Chem.*, 2003, **94**, 324–332.

[59] T. G. Kolda, R. M. Lewis, and V. Torczon, *SIAM Rev.*, 2003, **45**, 385–482.

[60] V. A. Rassolov, F. Xu, and S. Garashchuk, *J. Chem. Phys.*, 2004, **120**, 10385–10394.

[61] T. D. Crawford, C. D. Sherrill, E. F. Valeev, J. T. Fermann, R. A. King, M. L. Leininger, S. T. Brown, C. L. Janssen, E. T. Seidl, J. P. Kenny, and W. D. Allen, *PSI 3.2*, 2003.