UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Improving Wavefunction Efficiency by Tessellating Correlation Factors and Coupled State-Specific Optimization

Permalink

https://escholarship.org/uc/item/4j57f3q0

Author

Van Der Goetz, Beatrice Weston

Publication Date

2021

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Improving Wavefunction Efficiency by Tessellating Correlation Factors and Coupled State-Specific Optimization

by

Beatrice Van Der Goetz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eric Neuscamman, Chair

Professor Martin Head-Gordon

Professor James Demmel

Spring 2021

Abstract

Improving Wavefunction Efficiency by Tessellating Correlation Factors and Coupled State-Specific Optimization

by

Beatrice Van Der Goetz

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Eric Neuscamman, Chair

Rearranging chemical bonds is chemistry. Simulating chemical reactions is an expensive and complex process, and is necessary to understand the photochemical reactions that drive processes like chemical light-harvesting. The electronic many-body physics describing the bonds that participate in these processes becomes complicated and expensive even on modestly sized molecules, and computationally affordable approximations can qualitatively fail. One approach to this problem relies on devising compact and expressive wavefunction forms that are simple enough to be efficiently computed yet complex enough to capture the subtleties of many-electron physics.

Due to their relaxed integrability conditions, position-space Quantum Monte Carlo methods permit the use of flexible wavefunction components like Jastrow correlation factors that can be used to exactly express wavefunction cusps, which are otherwise difficult to describe. As many existing factors used in these calculations focus primarily on short-range or general correlation effects, we aim to augment the library of real-space correlation factors by developing one designed to handle the strong electronic correlation of bond-breaking. These factors do this by accounting for correlations between populations of electrons in different pockets of space using a set of indicator-like functions fashioned into a tessellation of Voronoi cells. These Voronoi cells can be automatically constructed around atomic coordinates or further tailored to the chemical system, as the math describing them is flexible enough to allow more subtle intra-atomic subdivisions and curved interfaces. In simple test systems, this factor is able to correctly cleave chemical bonds when applied to a single-determinant reference, and the resulting wavefunction is competitive with highly-accurate but expensive methods.

Just as wavefunctions can be limited by their functional form, so too they can be limited by the definition of their reference point. Many excited state methods are based on a linear response formalism, in which excited states are generated in the language of perturbations applied to a fixed ground state. When the wavefunction qualitatively changes upon electronic excitation – as they do in charge-transfer states and core excitations – these methods can fail to predict excitation energies with errors on the scale of several electron-volts. Optimizing molecular orbitals for individual excited states is one particularly efficient way to make the necessary zeroth-order changes to capture these states, but these state-specific

methods can suffer from instabilities in their optimization. One particular benefit of this state-specific approach is that these tailored orbitals naturally compress the wavefunction, but in certain cases, the ground state can interfere when minimal representations of the excited state are sought, ultimately causing variational collapse. We show that this collapse behavior occurs in two different state-specific approaches, and show how this arises from an inadequately-modeled avoided crossing, and argue that orbital-CI coupling plays a key role in its prevention.

Newer excited state methods have also seen the use of target functions, where functions of the energy such as the square-gradient magnitude are minimized in place of the energy in order to stabilize state-specific optimization by transforming saddle points on the energy surface into minima on the target function surface. While pilot implementations of square-gradient-based optimizers for the Excited-State Mean-Field (ESMF) wavefunction are able to obtain state-specific orbitals at low cost, these are still new and have yet to benefit from numerical accelerations, limiting their use. For instance, while stable, the quasi-Newton optimizer in the ESMF-GVP implementation uses no Hessian preconditioner, and while stable to variational collapse due to its inclusion of orbital-CI coupling, is much slower relative to the occasionally-unstable ESMF-SCF implementation. Using the exact Hessian and full Newton-Raphson optimization as a benchmark, we explore a variety of Hessian approximations, and find that an approximate diagonal Hessian can accelerate the ESMF-GVP square-gradient minimization to match the speed of the gradient-only ESMF-SCF at mean-field cost-scaling while resisting variational collapse.

for mom

Contents

Acknowledgements

It has taken nearly a decade and finding myself here feels all but impossible.
Many deserve acknowledgement, for their mentorship, support, and friendship.

Edward Perry,
  whose class in organic chemistry opened my eyes to its beauty;
James Vyvyan and Paul Deck,
  for introducing me to the world of research;
Daniel Crawford,
  for an incomparable introduction to quantum chemistry;
Jim Tanko and Hervè Marand,
  for wonderful lectures and the longest semester of my undergraduate career;
Bill Floyd,
  for precise and lucid lectures on Real Analysis;
Susan Bell, Maria Cristina-Mendoza, and Linda Carlson
  for all their help.

Eric Neuscamman,
  for personal support, dedication to the craft of research, insightful mentorship,
  and our many, many enlightening conversations;
Luning, Jacki, Sergio, Becky, Connie, Tarini, Scott, and Rachel,
  for being a wonderful research group;

Brad, Alex, Will, Jessalyn, Harvey, and Hazel,
  who I will never forget;
Alexia,
  for everything, and who I wish nothing but the best;
Kali and Taeko,
  for being there for me;

Dad and Julie,
  for their constant encouragement;
Erik, Brittany, and Lauren,
  for being the best siblings I could ever have,
Mom,
  for all your love and support,
  without which I would not be here.

List of Figures

List of Tables

1 Introduction

Characterizing chemical reactions is at the heart of quantum chemistry. Many stable molecules are modeled well-enough by existing theory, and quality thermodynamic information for these systems can be routinely obtained.[3] While the molecular-electronic problem is not completely solved, chemical dynamics asks much more of theoretical methodologies.[4] In contrast to thermodynamics, which principally considers the endpoint structures of a given reaction, chemical dynamics involves simulation and characterization of molecules during a reaction process. The heavily strained – even broken – chemical bonds that appear in these transformations provide a challenge for more conventional methods that are primarily tuned for equilibrium chemical systems. Accurately simulating chemical reactions for even small molecules can require quality potential energy surfaces that can most reliably be obtained by expensive multi-configurational theories.[5–11]

As chemical bonds are stretched or twisted, complex many-body quantum correlation effects start to appear. For instance, electronic repulsion separates electron pairs into a superposition of correlated spin states between the two product fragments. Mathematically, these open-shell states can be represented by a multi-configurational wavefunction that fully maintains physical symmetries. Due to the high cost of fully including this correlation in the general case, this straightforward approach is often sidestepped in routine calculations in favor of approximate methods that break symmetry[12–16] with the hope that it can be later recovered. In the first part of chapter 1, we review the standard mathematical framework of many-body quantum mechanicals, and discuss several classic examples of electronic correlation as it manifests in chemical bond-breaking.

While breaking chemical bonds is an extreme physical example, and simpler cases of bond-breaking can be described in an *ad hoc* framework, corrections of this form can be difficult to generalize to more complex systems. On the other hand, formally exact methods can have trouble representing or finding necessary wavefunction components at reasonable cost, and often scale exponentially with the size of the system.[17–20] Finding alternative functional forms that can efficiently describe electronic correlation at low cost-scaling thus remains a worthy goal. The second part of chapter 1 introduces the flexible formalism of real-space Monte Carlo methods, and in chapters 2 and 3, we explore a wavefunction factor designed to correlate total electron counts between spatial regions that takes advantage of this flexibility. This is done in the language of spatial Voronoi partitions, and we show how the inclusion of these correlation factors can accurately and affordably describe the correlation of bond-breaking processes.

Electronic excitations can also require careful theoretical treatment, as excited states are often constructed around reference wavefunctions tuned solely for the ground state. This ground-state reference is especially unsuitable when there is a substantial migration of charge upon excitation, and can confound even expensive and otherwise accurate methodologies that do not adjust this reference.[21] In time-dependent density function theory (TD-DFT),[22–24] the most popular excited state method, adapting this ground-state reference for the excited state is an involved process, and requires extension of the underlying linear-response theory[25–28] and functional approximations.[28,29] Wavefunction-based methods have made effort to address this with state-averaged[30] and state-specific[31,32] approaches, yet do so through a non-linear orbital optimization problem that comes with its own fundamental complex-

ities. While recent advances in ESMF theory[33–37] have greatly improved the stability of these excited-state orbital optimization procedures, its implementation has yet to benefit from numerical accelerations and careful stability analyses necessary for their routine use. In chapter 4, we identify an instability unique to this optimization process that appears in a surprisingly simple molecular example, and show that a coupled optimization algorithm that avoids these instabilities can be accelerated to compete with a cheap but less-stable self-consistent field approach. This numerical acceleration is implemented as a Quasi-Newton Hessian preconditioner and is derived from and compared to results from analytical Hessian expressions.

1.1 Quantum Mechanics

Classical electrons do not form a bound molecular system,[38] and so must instead be described by quantum mechanics using a many-body electronic wavefunction:

$$\Psi(\boldsymbol{r}_1, \dots \boldsymbol{r}_n), \tag{1.1}$$

For our purposes, this wavefunction can be understood as a direct precursor to a many-body density distribution:[39,40]

$$\rho(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) = |\Psi(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n)|^2 \tag{1.2}$$

Physical attributes of our state – energy, properties, interactions – can be obtained as expectation values of physical operators on the normalized wavefunction:

$$O[\Psi] = \int d\boldsymbol{r}_1 \cdots \int d\boldsymbol{r}_n \Psi(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) \mathbf{O} \Psi(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) \Big/ \int d\boldsymbol{r}_1 \dots d\boldsymbol{r}_n |\Psi(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n)|^2 \tag{1.3}$$

These physical operators are defined by direct analogy to classical formulae and are identical with the notable exception of the momentum,[40] which is now written as a differential operator:

$$\mathbf{p} = -i\hbar\nabla \tag{1.4}$$

As these operators are linear, expectation values can be written as a bilinear form in what is known as bra-ket notation:

$$O[\Psi] = \frac{\langle \Psi | \mathbf{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{1.5}$$

which is a unified notation to describe the abstract linear spaces used so often in quantum mechanics: the 'ket', $|\Psi\rangle$, can be thought of as a vector; the operators $\mathbf{O}$ as linear operators in this space; and the 'bra', $\langle\Psi|$ as linear functionals.[41,42]

In direct analogy with classical Hamiltonian mechanics, the Hamiltonian operator is used to obtain the electronic energy of the system:[39]

$$E = \frac{\langle \Psi | \mathbf{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \tag{1.6}$$

and is particularly important since it defines the system's dynamics, as in the time-dependent Schrödinger equation:

$$-i\frac{\partial}{\partial t} |\Psi(t)\rangle = \mathbf{H} |\Psi(t)\rangle \tag{1.7}$$

The time-dependence of a quantum state can be thus be described by an exponential propagator applied to some initial state:

$$|\Psi(t)\rangle = \exp\left(-i\mathbf{H}t\right)|\Psi(0)\rangle \tag{1.8}$$

Eigenfunctions of the Hamiltonian operator exhibit trivial time-dependence as an oscillating phase factor e^{-iEt} and their density distribution remains stationary over time.[41] Analysis of quantum systems thus often begins by identifying these eigenstates, as in the time-independent Schrödinger equation:

$$\mathbf{H}|\Psi\rangle = E|\Psi\rangle \tag{1.9}$$

These stationary states can also be framed as stationary points of the energy functional with respect to some parameterized description of the state:

$$\nabla_p E(p) = 0 \quad \leftrightarrow \quad \mathbf{H}|\Psi(p)\rangle = E|\Psi(p)\rangle \tag{1.10}$$

which is the foundation for variational algorithms that search for these states in a wide variety of methods.[19,41,43–45]

In atomic units, the Hamiltonian for molecular electronic systems is written:[43]

$$\mathbf{H}\left(\{\boldsymbol{r}_i\},\{\boldsymbol{R}_i\}\right) = -\sum_I \frac{{\nabla_I}^2}{2M_I} + \sum_{I<J}\frac{q_I q_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} - \sum_i \frac{{\nabla_i}^2}{2} - \sum_{iI}\frac{q_I}{|\boldsymbol{R}_I - \boldsymbol{r}_i|} + \sum_{i<j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1.11}$$

As the mass scales of electrons and nuclei differ by a factor of nearly 2000, the electronic and nuclear degrees of freedom are often separated. This is done according to the Born-Oppenheimer approximation[46] in which the nuclear positions are fixed parameters and the first term, the nuclear kinetic energy, is dropped:

$$\mathbf{H}_{BO}\left(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_i\}\right) = \sum_{I<J}\frac{q_I q_J}{|\boldsymbol{R}_I - \boldsymbol{R}_J|} - \sum_i \frac{{\nabla_i}^2}{2} - \sum_{iI}\frac{q_I}{|\boldsymbol{R}_I - \boldsymbol{r}_i|} + \sum_{i<j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} \tag{1.12}$$

Electronic energy of the system is obtained as a function of the fixed nuclear coordinates:

$$E(\{\boldsymbol{R}_i\}) = \langle\Psi(\{\boldsymbol{r}_i\})|\mathbf{H}_{BO}(\{\boldsymbol{r}_i\};\{\boldsymbol{R}_i\})|\Psi(\{\boldsymbol{r}_i\})\rangle \tag{1.13}$$

Though this approximation is accurate in a wide variety of molecular systems, and this electronic surface can serve as a potential energy surface on which nuclei move, breakdowns in this approximation can and do occur to great effect in chemical reaction processes.[4,5]

1.2 Electronic Wavefunctions

Electrons belong to a class of particles known as Fermions, and as Fermions, exclude other electrons from occupying the same quantum state.[47,48] This property is responsible for the volume-occupying nature of matter.[49] The electronic wavefunction is a function of both

spatial electronic positions and quantum spin state, and is the differential probability density of a particular spatial and spin configuration of electrons.

$$|\Psi(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n)|^2 = P\left(\left\{\begin{array}{c}\text{electron 1 located at } \boldsymbol{r}_1\\ \text{electron 2 located at } \boldsymbol{r}_2\\ \vdots\\ \text{electron n located at } \boldsymbol{r}_n\end{array}\right\}\right) \tag{1.14}$$

Individual electrons are physically indistinguishable, which is reflected as permutational symmetry of the wavefunction magnitude:

$$|\Psi(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n)|^2 = |\Psi(\boldsymbol{r}_{\sigma(1)},\ldots,\boldsymbol{r}_{\sigma(n)})|^2, \qquad \sigma \in S_n \tag{1.15}$$

The exclusive occupation of Fermionic quantum states is known as the Pauli exclusion principle,[47,48] and can be expressed by permutational antisymmetry of the bare wavefunction, as an antisymmetric wavefunction vanishes on particle coalescence, reflecting the fact that no two electrons can be found at the same point in space:

$$\Psi(\boldsymbol{r}_1,\boldsymbol{r}_1,\dots) = -\Psi(\boldsymbol{r}_1,\boldsymbol{r}_1,\dots) \quad \rightarrow \quad \Psi(\boldsymbol{r}_1,\boldsymbol{r}_1,\dots) = 0 \tag{1.16}$$

In electronic structure theory, numerically solving such differential equations in multiple dimensions relies on transforming function spaces into linear function spaces using integration. This framework can be thought of as a generalization of Fourier transforms, as their construction relies on parallel mathematical arguments. However, this quantization process can be involved,[50] and can be summarized as a way to construct, manipulate, and evaluate an approximate antisymmetric wavefunction using a vector space of linear operators. We assume that the wavefunction can be written as a linear combination of orthonormal basis states:

$$|\Psi\rangle = \sum_i c_i \,|D_i\rangle\,, \qquad \int d\boldsymbol{r}_1 \ldots d\boldsymbol{r}_n \; D_i(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n) D_j(\boldsymbol{r}_1,\ldots,\boldsymbol{r}_n) = \delta_{ij} \tag{1.17}$$

And after substituting this into the time-independent Schrödinger equation and integrating against each of these basis states, we arrive at a linear equation whose solution provides us with the linear parameters c_i.

$$\sum_j \boldsymbol{H}_{ij} c_j = E c_i \tag{1.18}$$

$$\boldsymbol{H}_{ij} = \langle D_i|\mathbf{H}|D_j\rangle \tag{1.19}$$

Antisymmetrized products of single-particle wavefunctions, themselves based on analytical single-electron solutions to simple atomic systems, are used to construct these basis states. First, we define a set of atomic orbitals: products of spherical gaussian functions and spherical harmonics, centered on each atom, that are variationally tuned to mimic approximate solutions to many-body electronic states in atomic systems:[51,52]

$$\phi_i(r,\Omega) = \exp(-\alpha r^2) Y_l^m(\Omega) \tag{1.20}$$

Next, molecular orbitals are written as orthonormal linear combinations of these atomic orbitals:

$$\chi_i(\boldsymbol{r}) = \sum_j C_{ji}\phi_j(\boldsymbol{r}), \qquad \int d\boldsymbol{r}\ \chi_i(\boldsymbol{r})\chi_j(\boldsymbol{r}) = \delta_{ij} \tag{1.21}$$

And finally, electronic configurations are written as antisymmetric products of these molecular orbitals:[50]

$$|D_i\rangle = |\chi_1^{(i)} \dots \chi_n^{(i)}\rangle = \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) \prod_j \chi_{\sigma(j)}^{(i)} \tag{1.22}$$

This is the basic component of a many-body electronic state, and is known as a Slater determinant:[53]

$$D_i(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) = \begin{vmatrix} \chi_1^{(i)}(\boldsymbol{r}_1) & \chi_2^{(i)}(\boldsymbol{r}_1) & \dots & \chi_n^{(i)}(\boldsymbol{r}_1) \\ \chi_1^{(i)}(\boldsymbol{r}_2) & \chi_2^{(i)}(\boldsymbol{r}_2) & \dots & \chi_n^{(i)}(\boldsymbol{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1^{(i)}(\boldsymbol{r}_n) & \chi_2^{(i)}(\boldsymbol{r}_n) & \dots & \chi_n^{(i)}(\boldsymbol{r}_n) \end{vmatrix} \tag{1.23}$$

Though we use finite basis sets in practice, in principle, one can continually expand the linear space to approach completeness – i.e. cover all of space – in which every (normalizable, antisymmetric) wavefunction can be expressed exactly in the infinite basis limit.[42,54] We will also note that we are effectively working with two levels of linear spaces: the space of molecular orbitals (equation 1.21), and the space of determinants (equation 1.17). The molecular orbital space is a product of a square-integrable function space over physical coordinates composed with a two-dimensional spin space, and roughly represents the state of a single electron in a many-body quantum systems (thus: single-particle orbital). Basis elements of the linear determinant space are defined by the molecular orbitals present within a particular Slater determinant, and its dimension is as large as the (combinatorial) number of ways molecular orbitals can be assigned to electronic particles. These represent many-body electronic states, and superposition of its components describe intra-state correlations between different electrons in the many-body system, but luckily, many electronic systems can be qualitatively represented with a single determinant.[53] We will refer to this many-body space as the linear determinant space or the configuration interaction (CI) space interchangeably, so named because different electronic configurations 'interact' through off-diagonal elements of the linear Hamiltonian in equation 1.19.

Physically, we are most interested in finding low-energy ground state solutions to the time-independent Schrödinger equation, as systems tend to equilibrate toward states with lower energies. For the electronic problem in many small molecules, the energy gap between the ground and excited states is often large enough at equilibrium geometries that only the ground state is populated under ambient conditions. Conveniently, fast algorithms are well-known for the numerical calculation of the eigenvalues and eigenvectors[55–57] that represent stationary electronic states. Inconveniently, the size of the Slater-determinant space grows combinatorially with the number of single-particle states, as it consists of all combinations of the electrons assigned to molecular orbitals. And while single-determinant wavefunctions are a sufficient functional form for many ground states in equilibrium molecular systems, it is often necessary to go further in determinant space to capture more complex systems or chemical reaction processes in full. In thses cases, methods in quantum chemistry are

built to avoid directly confronting this exponentially-large determinant space by finding approximate solutions in a smaller subspace, exploiting compact wavefunction design and numerical sparsity wherever possible.[17–20] Selective CI methods,[58,59] for instance, attempt to find and fill in only physically significant parts of this space, using the sparsity of the linear Hamiltonian (eq. 1.19) to craft a small but accurate approximation to the wavefunction.

As Fermions, electrons have a quantum of intrinsic magnetic moment known as spin[53,60] that is irreducibly quantum-mechanical in nature. The full physical state of each electron is described by a spatial component, discussed above, alongside a spin component, whose quantum state can be fully described in a two-dimensional spin space. These states represent the electron's magnetic moment either aligned with or against a designated spin-quantization axis, typically defined as the z-axis, in which these spin states, written here in atomic units, are eigenstates of a spin-projection operator:

$$|s_i\rangle \in \{|\uparrow\rangle, |\downarrow\rangle\}, \qquad \mathbf{S}_z |\uparrow\rangle = \frac{1}{2} |\uparrow\rangle, \qquad \mathbf{S}_z |\downarrow\rangle = -\frac{1}{2} |\downarrow\rangle \tag{1.24}$$

Spin projection operators transform in the same way as angular momentum operators, and spin components along the $x-$ and $y-$axes can be directly obtained by commutation relations obtained by symmetry arguments.[60] The total spin operator is obtained as the sum of squares of the spatial components:

$$\mathbf{S}^2 = \mathbf{S}_x^2 + \mathbf{S}_y^2 + \mathbf{S}_z^2 \tag{1.25}$$

As projected spin is additive, generalizing this to multi-electron systems is relatively straightforward, and algebra proceeds along the same lines as in the single-particle case.[53] When spin-orbit coupling is neglected, which it often can be with the exception of the core orbitals of heavy elements,[50,60] the Hamiltonian is symmetric with respect to both projected spin $\mathbf{S}_z$ and total spin $\mathbf{S}^2$, and the overall spin-state of a wavefunction only indirectly affects the energy based on Pauli exclusion effects. As pairing electrons of opposite spin allows two electrons to occupy low-energy spatial states, the singlet spin state is often (correctly) assumed to be the lowest energy state. Though spin-states can formally be taken as mixtures of these two components (as it is in Unrestricted Hartree-Fock[16] or Generalized Hartree-Fock[15]), spatial components are often tied to one of the two spin-projected states. These components are also often spin-paired to create the desired closed-shell singlet configuration, though this is not always appropriate.

1.3 Second Quantization

Linear determinant spaces can more conveniently be written as a linear space of operators that act to populate or depopulate molecular orbitals in a Slater determinant wavefunction.[53] This space is isomorphic to the determinant space, as each determinant can be represented as a string of creation operators applied to an unoccupied 'vacuum state'. Each of these creation operators can be thought to append a column and row to an existing Slater determinant wavefunction:

$$|D_i\rangle = |\chi_1 \chi_2 \dots \chi_n\rangle = \mathbf{1}^\dagger \mathbf{2}^\dagger \dots \mathbf{n}^\dagger |0\rangle \tag{1.26}$$

Creation operators are complemented by a set of destruction operators that perform the opposite operation, instead removing a column and row of the specified orbital from the

determinant matrix:

$$\mathbf{1}\left|\chi_1\chi_2\ldots\chi_n\right\rangle = \left|\chi_2\ldots\chi_n\right\rangle \tag{1.27}$$

These operators are nilpotent, in that applying an annihilation operator to a vacant state or a creation operator to a populated state returns the zero vector:[53]

$$\mathbf{1}\left|0\right\rangle = 0, \qquad \mathbf{1}^\dagger\left|\chi_1\right\rangle = 0 \tag{1.28}$$

and also obey anti-commutation rules that implement Fermi antisymmetry:

$$\{\mathbf{i}^\dagger,\mathbf{j}^\dagger\} = 0, \qquad \{\mathbf{i},\mathbf{j}\} = 0, \qquad \{\mathbf{i},\mathbf{j}^\dagger\} = \delta_{ij} \tag{1.29}$$

With these relationships in place, expectation values of vectors in this space can be instead evaluated by algebraic manipulation and expectation values of these operators.[61] The full Hamiltonian can be represented using creation and annihilation operators, and this Second-Quantized form is often convenient when deriving working equations for numerical algorithms:

$$\mathbf{H} = \sum_{ij} h_{ij}\mathbf{i}^\dagger\mathbf{j} + \frac{1}{2}\sum_{ijkl} g_{ijkl}\mathbf{i}^\dagger\mathbf{k}^\dagger\mathbf{l}\mathbf{j} \tag{1.30}$$

Here the coefficients are defined by integrals of one-body ($h(\boldsymbol{r})$) and two-body ($1/r_{12}$) components of the electronic Hamiltonian:[53]

$$h_{ij} = \int d\boldsymbol{r}\ \chi_i(\boldsymbol{r})h(\boldsymbol{r})\chi_j(\boldsymbol{r}), \qquad g_{ijkl} = \int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2\ \chi_i(\boldsymbol{r}_1)\chi_j(\boldsymbol{r}_1)\frac{1}{r_{12}}\chi_k(\boldsymbol{r}_2)\chi_l(\boldsymbol{r}_2) \tag{1.31}$$

Where $\mathbf{h}(\boldsymbol{r})$ represents the single-electron terms that appear in equation 1.12:

$$h(\boldsymbol{r}) = -\frac{\nabla^2_{\boldsymbol{r}}}{2} - \sum_I \frac{q_I}{|\boldsymbol{R}_I - \boldsymbol{r}|} \tag{1.32}$$

Spin can be straightforwardly included by multiplying these integrals by the inner product between spin-states of each molecular orbital pair:

$$\begin{aligned} h_{ij} &= \langle s_i|s_j\rangle \int d\boldsymbol{r}\ \chi_i(\boldsymbol{r})\mathbf{h}(\boldsymbol{r})\chi_j(\boldsymbol{r}) \\ g_{ijkl} = (ij|kl) = \langle s_i|s_j\rangle\langle s_k|s_l\rangle &\int d\boldsymbol{r}_1 \int d\boldsymbol{r}_2\ \chi_i(\boldsymbol{r}_1)\chi_j(\boldsymbol{r}_1)\frac{1}{r_{12}}\chi_k(\boldsymbol{r}_2)\chi_l(\boldsymbol{r}_2) \\ &\langle s_i|s_j\rangle = \delta_{ij} \end{aligned} \tag{1.33}$$

and the energy of an electronic wavefunction can be written in terms of expectation values of these these second-quantized operators:

$$E = \langle\Psi|\mathbf{H}|\Psi\rangle = \sum_{ij}\gamma^{(1)}_{ij}h_{ij} + \frac{1}{2}\sum_{ijkl}\gamma^{(2)}_{ijkl}g_{ijkl} \tag{1.34}$$

$$\gamma^{(1)}_{ij} = \langle\Psi|\mathbf{i}^\dagger\mathbf{j}|\Psi\rangle, \qquad \gamma^{(2)}_{ijkl} = \langle\Psi|\mathbf{i}^\dagger\mathbf{k}^\dagger\mathbf{l}\mathbf{j}|\Psi\rangle \tag{1.35}$$

These expectation values depend only on the wavefunction, and are known as the one-body reduced density matrix and two-body reduced density matrix, respectively.[62,63]

1.4 Mean-Field Theory

In Hartree-Fock theory, the minimum energy of a single-determinant wavefunction is found by varying the molecular orbital coefficients defined by C in equation 1.21.[43,64,65] These molecular orbitals are defined as linear combinations of a set of basis orbitals $\{\phi_i\}$:

$$\chi_i(C;\boldsymbol{r}) = \sum_j C_{ji}\phi_j(\boldsymbol{r}) \tag{1.36}$$

and the energy is minimized with the constraint that molecular orbitals form an orthonormal set using the following Lagrangian:

$$\mathcal{L}(C) = \frac{\langle\Psi(C)|\mathbf{H}|\Psi(C)\rangle}{\langle\Psi(C)|\Psi(C)\rangle} + \sum_{ij}\boldsymbol{\epsilon}_{ij}\left(\int d\boldsymbol{r}\ \chi_i(C;\boldsymbol{r})\chi_j(C;\boldsymbol{r}) - \delta_{ij}\right) \tag{1.37}$$

with Lagrange multipliers ϵ_{ij}. After differentiating and imposing the stationarity condition:

$$\frac{\partial\mathcal{L}}{\partial C_{ij}} = 0 \tag{1.38}$$

the single-determinant energy is invariant to rotations amongst the occupied molecular orbitals, and $\boldsymbol{\epsilon}$ can be diagonalized to give a matrix eigenvalue equation:

$$F(C)C = SC\boldsymbol{\epsilon} \tag{1.39}$$

The matrix $F(C)$ is known as the Fock matrix and is a one-body approximation to the full Hamiltonian operator. As the entries of this matrix depends on the solutions to the above eigenvalue equation, the above equation must be solved self-consistently.

Self-consistent field (SCF) methods such as Spin-restricted Hartree-Fock (RHF) reduce a linear algebra problem in an intractably large space to a self-consistent eigenvalue problem, and single-determinant algorithms of this flavor are widely used in Quantum Chemistry. Hartree-Fock is still an approximation, and the most common corrections come in the form of Density Functional Theory[22,66–75] (DFT), where the effective Fock operator is modified to reproduce accurate energies and densities without increasing the cost of optimization beyond solving this self-consistent eigenvalue equation. As the goal of Kohn-Sham DFT is to completely avoid any complexity beyond single-determinant wavefunctions, we mention it here because, despite its wide success and popularity, it can encounter many of the same theoretical issues that afflict Hartree-Fock.

Electronic states of equilibrium systems are often well-described by a single-determinant, but the Hartree-Fock wavefunction nonetheless misses some details. While having optimizing this wavefunction with respect to orbital parameters means that we have explored a truncated space, it is nowhere near the size of the exponential linear CI space. Inaccuracies in Hartree-Fock are often understood in terms of the Coulombic repulsion term in the Hamiltonian:

$$\frac{1}{r_{12}} \tag{1.40}$$

without which single-determinant wavefunctions would provide exact solutions within the single-particle orbital basis. This can be seen by looking only at the one-body components

in equation 1.31, where diagonalizing the h matrix would provide an orbital basis in which every determinant is an eigenstate.

This Coulombic repulsion term means that electrons correlate their positions to avoid each other when possible. In the sense that, absent this repulsion, a single-determinant wavefunction would be exact, Hartree-Fock wavefunctions are often understood to be uncorrelated. This is not strictly correct, as the antisymmetry imposed by Pauli exclusion is a strong form of exclusive correlation, as the two-body density cannot be written as a simple product of one-body densities:

$$\rho(\boldsymbol{r}_1, \boldsymbol{r}_2) = \int \cdots \int d\boldsymbol{r}_3 \ldots d\boldsymbol{r}_n \; |\Psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n)|^2 \neq \rho_1(\boldsymbol{r}_1)\rho_2(\boldsymbol{r}) \tag{1.41}$$

This muddies the waters a bit when talking about correlation, as strong repulsive effects can be represented by the exclusive correlation of a determinant wavefunction.[14,76] It is thus more correct to consider a single determinant wavefunction as a set of uncorrelated Fermions whose density may require additional pieces of correlation to fully account for electronic repulsive effects. That additional correlation is necessary is most clear if we consider the case of two opposite-spin electrons:[77,78] antisymmetry does not correlate these electrons, but these electrons still repel each other, and their two-body density should be correlated more than their simple product wavefunction would imply.

1.5 H_2 Dissociation

Chemical bonds are essential components of stable molecules, and their presence can be inferred by inspection of the wavefunction coefficients by identifying as constructive linear combinations of orbitals on neighboring atoms. However, accurately representing broken or half-broken chemical bonds can require modification of our simple mean-field framework. The simplest bonding configuration is the single chemical bond present in the H_2 molecule and consists of a spatially symmetric, normalized sum between atomic orbitals centered around each atom. We represent this in the $\{\mathbf{L}^\dagger, \mathbf{R}^\dagger\}$ basis, with each element representing a basis function centered at the left and right atoms, respectively:

$$\sigma^\dagger = \frac{1}{\sqrt{2}}\left(\mathbf{L}^\dagger + \mathbf{R}^\dagger\right), \qquad \sigma^{*\dagger} = \frac{1}{\sqrt{2}}\left(\mathbf{L}^\dagger - \mathbf{R}^\dagger\right) \tag{1.42}$$

At equilibrium geometries, an RHF-style single-determinant wavefunction that populates this bonding orbital with a pair of opposite-spin electrons is a good approximation to the true ground state, and is the electronic wavefunction at points A and B in figure 1.1. If we insist on maintaining both spin and spatial symmetry between these two orbitals, by symmetry, this configuration will be our wavefunction at any bondlength if we limit ourselves to a single determinant. As we stretch the chemical bond, we expect the energy to approach the energy of two non-interacting hydrogen atoms, each of which has an exact energy of -0.5 $\mathrm{E_h}$. However, this single configuration deviates strongly from this expected result, and the predicted dissociation energy for the molecule is too large by hundreds of $\mathrm{mE_h}$.[79]

To see why this happens, we analyze the σ-populated single-determinant wavefunction in more detail:

$$\sigma_\uparrow{}^\dagger \sigma_\downarrow{}^\dagger \left|0\right\rangle \tag{1.43}$$

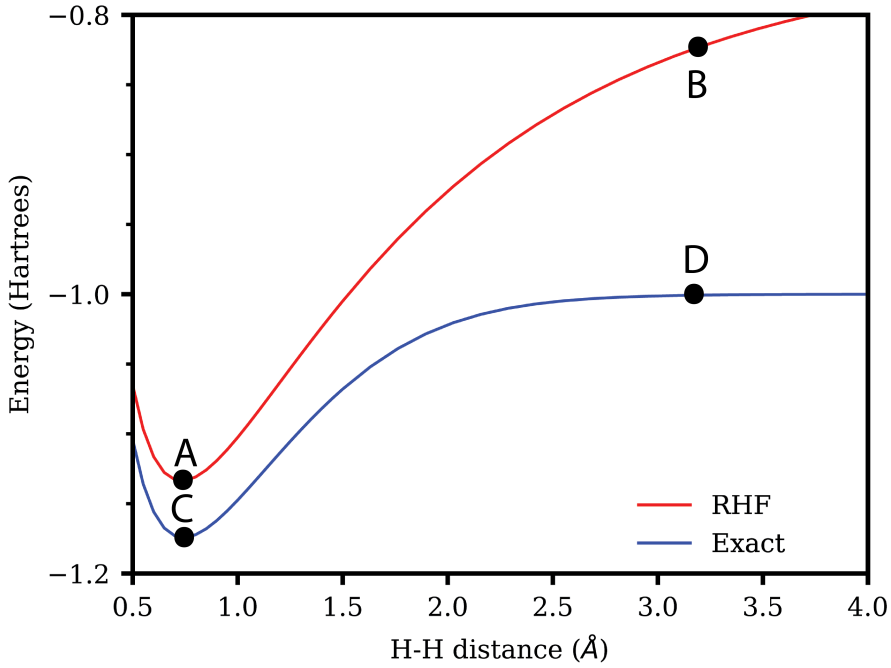

Figure 1.1: Dissociation of the dihydrogen molecule as a function of the distance between two hydrogen atoms. The upper curve, calculated by spin-restricted Hartree-Fock (RHF), has a wavefunction determined entirely by spin-symmetry, and is given by equation (1.43) at both points A and B. The lower curve corresponds to a near-exact result obtained by exhaustively diagonalizing the Hamiltonian in the full determinant space with a large single-particle basis (FCI in cc-pV5Z). This near-exact wavefunction at equilibrium (point C) is qualitatively captured by the single-determinant wavefunction (point A) while the near-exact wavefunction at stretched geometries (point D), given approximately by equation (1.48) is not.

We expand each of the $\sigma-$orbitals in the atomic orbital basis:

$$\frac{1}{\sqrt{2}}\left(\mathbf{L}_{\uparrow}{}^{\dagger}+\mathbf{R}_{\uparrow}{}^{\dagger}\right)\left(\mathbf{L}_{\downarrow}{}^{\dagger}+\mathbf{R}_{\downarrow}{}^{\dagger}\right)|0\rangle \tag{1.44}$$

After distributing the product, we arrive at an sum of four configurations with equal linear weights:

$$\frac{1}{\sqrt{2}}\left(\mathbf{L}_{\uparrow}{}^{\dagger}\mathbf{L}_{\downarrow}{}^{\dagger}+\mathbf{L}\uparrow^{\dagger}\mathbf{R}_{\downarrow}{}^{\dagger}+\mathbf{R}_{\uparrow}{}^{\dagger}\mathbf{L}_{\downarrow}{}^{\dagger}+\mathbf{R}_{\uparrow}{}^{\dagger}\mathbf{R}_{\downarrow}{}^{\dagger}\right)|0\rangle \tag{1.45}$$

At stretched bond distances, the overlap between ϕ_L and ϕ_R will vanish, and the energy can be written as a sum of the energies of each of the four configurations:

$$E=\frac{1}{\sqrt{2}}\Big[\langle L_{\uparrow}L_{\downarrow}|\mathbf{H}|L_{\uparrow}L_{\downarrow}\rangle+\langle L_{\uparrow}R_{\downarrow}|\mathbf{H}|L_{\uparrow}R_{\downarrow}\rangle+\langle R_{\uparrow}L_{\downarrow}|\mathbf{H}|R_{\uparrow}L_{\downarrow}\rangle+\langle R_{\uparrow}R_{\downarrow}|\mathbf{H}|R_{\uparrow}R_{\downarrow}\rangle\Big] \tag{1.46}$$

While the one-body contributions are the same for each of these terms, the first and last terms – those which assign two electrons to one atom and zero to the other – include an additional repulsion term between the two electrons. For instance:

$$\langle L_{\uparrow}L_{\downarrow}|\mathbf{H}|L_{\uparrow}L_{\downarrow}\rangle=2\int d\boldsymbol{r}\ \phi_L(\boldsymbol{r})\mathbf{h}(\boldsymbol{r})\phi_L(\boldsymbol{r})+\int\int d\boldsymbol{r}_1 d\boldsymbol{r}_2\ \phi_L(\boldsymbol{r}_1)\phi_L(\boldsymbol{r}_1)\frac{1}{r_{12}}\phi_L(\boldsymbol{r}_2)\phi_L(\boldsymbol{r}_2) \tag{1.47}$$

As the minimum energy state, the exact ground state will omit these two high-energy pieces and consist of only the two middle configurations (point D on fig. 1.1), and can also be written as a difference between the σ and σ^* configurations:

$$\frac{1}{\sqrt{2}}\left(\mathbf{L}_{\uparrow}{}^{\dagger}\mathbf{R}_{\downarrow}{}^{\dagger}+\mathbf{R}_{\uparrow}{}^{\dagger}\mathbf{L}_{\downarrow}{}^{\dagger}\right)|0\rangle=\left(\sigma_{\uparrow}{}^{\dagger}\sigma_{\downarrow}{}^{\dagger}-\sigma_{\uparrow}^{*\dagger}\sigma_{\downarrow}^{*\dagger}\right)|0\rangle \tag{1.48}$$

As long as we follow the symmetry of the Hamiltonian, it is not always easy to write this state using any less than these two determinant components. Because these components are unlinked and individually have the same energy, it is tempting to take only one to be an approximation to the ground state. If done, the wavefunction is no longer symmetric with regard to spin and is formally contaminated with incorrect spin-states.[16,80] While this can sometimes produce workable results, the effect of this can be unpredictable in more complex systems, and fixing spin symmetry once it has been broken is not always a cheap or simple task.[15,81,82] Breaking double bonds with a single-determinant wavefunction can also lead to an artificial breaking of spatial symmetries, which can be even more difficult to detect and fix.[83,84]

Wavefunctions that cannot be qualitatively described using a single-determinant are said to be strongly correlated.[12,13,85,86] Strong correlation can be understood as a strong statistical co-dependence of two electronic states that cannot be expressed in the confines of a single slater determinant. Looking at the exact dissociated H_2 ground state more closely, we can see that the electrons exactly avoid each other, as one electron is assigned to each atom in each configuration. To maintain the singlet spin-state, the symmetric combination of two configurations is necessary to correctly correlate the relative spin of the two electrons. While this correlation is simple to include in the H_2 molecule, straightforward multi-configurational treatments can quickly become unaffordable in larger chemical systems, and alternative and compact functional forms become necessary to express their extensive networks of correlation.

1.6 Density Matrices

While our example makes the presence of strong correlation relatively obvious, it relies on a qualitative understanding of correlation. A more general and mathematically-clear picture of strong correlation can be viewed from the perspective of density matrices,[62,63] defined in equation (1.35). We start with a multi-determinant wavefunction:

$$|\Psi\rangle = \sum_i c_i \, |D_i\rangle \tag{1.49}$$

First, we diagonalize the one-body density matrix to define a set of orbitals using the coefficients in the eigenvectors:

$$\gamma^{(1)} = U d U^T, \qquad \tilde{\mathbf{i}}^\dagger = \sum_j U_{ij} \mathbf{j}^\dagger \tag{1.50}$$

These are known as natural orbitals, and in this basis, the one-body density matrix is diagonal by construction. Writing the full expectation value out:

$$\langle \Psi | \tilde{\mathbf{k}}^\dagger \tilde{\mathbf{k}} | \Psi \rangle = \sum_{ij} c_i \, \langle D_i | \tilde{\mathbf{k}}^\dagger \tilde{\mathbf{k}} | D_j \rangle \, c_j = d_{kk} = \tilde{\gamma}^{(1)}_{kk} \tag{1.51}$$

When the configurations are written in the natural orbital basis – which can be done by substituting in the inverse of the above identity – we obtain:

$$\sum_{k, \chi_k \in |D_i\rangle} |c_i|^2 = d_{kk} \tag{1.52}$$

So that the eigenvalues of the one-body density matrix – the natural orbital occupation numbers – equal the total fractions of configurations that contain the corresponding natural orbital.

The natural orbital occupation numbers of single determinant wavefunctions are trivial, consisting of ones on the diagonal for occupied orbitals and zero everywhere else. As described above, strongly correlated wavefunctions deviate qualitatively from single-determinant wavefunctions, and this is reflected in their 1-RDMs as partially-occupied natural orbital occupations. In the H_2 example above, the density of the uncorrelated single-determinant configuration is given by:

$$\begin{array}{c} \\ \sigma_\uparrow \\ \sigma_\downarrow \\ \sigma^*_\uparrow \\ \sigma^*_\downarrow \end{array} \begin{array}{c} \begin{array}{cccc} \sigma_\uparrow & \sigma_\downarrow & \sigma^*_\uparrow & \sigma^*_\downarrow \end{array} \\ \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \end{array} \tag{1.53}$$

While the density of the strongly correlated multi-determinant wavefunction is given by:

$$\begin{array}{c} \\ \sigma_\uparrow \\ \sigma_\downarrow \\ \sigma^*_\uparrow \\ \sigma^*_\downarrow \end{array} \begin{array}{c} \begin{array}{cccc} \sigma_\uparrow & \sigma_\downarrow & \sigma^*_\uparrow & \sigma^*_\downarrow \end{array} \\ \begin{pmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 \end{pmatrix} \end{array} \tag{1.54}$$

In practice, strong correlation is often diagnosed using this type of analysis, as it is generally applicable and quantitative in nature. However, this is only most definitive near either end of the scale: ambiguity enters when occupation numbers take on intermediate values. While there is no universally-accepted standard for unambiguously identifying strong correlation in general, orbitals that are strongly partially-occupied can often predict qualitative failures in single-reference theory.

1.7 Twisted Ethene

Twisting a double-bond is another common motif that points to a strongly-correlated wavefunction. Drawing from the dissociation example, strong correlation can also be viewed through the lens of energetically degenerate configurations in the linear CI space:

$$\begin{bmatrix} \langle D_1|\mathbf{H}|D_1\rangle & \langle D_1|\mathbf{H}|D_2\rangle \\ \langle D_2|\mathbf{H}|D_1\rangle & \langle D_2|\mathbf{H}|D_2\rangle \end{bmatrix} \tag{1.55}$$

The degeneracy point is where:

$$\langle D_1|\mathbf{H}|D_1\rangle = \langle D_2|\mathbf{H}|D_2\rangle \tag{1.56}$$

and any nonzero amount of coupling swiftly rotates the optimal CI vector to place equal weights on each of the two configurations.

When twisting around a double-bond axis, we pay attention to valence π and π^* orbitals and the linear space of their doubly-occupied spin-restricted configurations in a minimal two-element p_z basis.[14] These occupation operators are applied to the closed-shell reference wavefunction ψ that populate the two carbon core-orbitals and five σ-bonding orbitals:

$$\pi^\dagger = \mathbf{p}_L^\dagger + \mathbf{p_R}^\dagger, \qquad \pi^{*\dagger} = \mathbf{p}_L^\dagger - \mathbf{p}_R^\dagger \tag{1.57}$$

$$\pi_\uparrow{}^\dagger \pi_\downarrow{}^\dagger \left|\psi\right\rangle, \qquad \pi_\uparrow^{*\dagger} \pi_\downarrow^{*\dagger} \left|\psi\right\rangle \tag{1.58}$$

Note that this linear space of closed-shell configurations $\{\pi, \pi^*\}$ is isomorphic to the dihydrogen example above. The π orbital, due to its bonding character, is lower in energy than the π^* orbital at equilibrium geometries. As the $HCCH$ dihedral angle is twisted to 90°, the energy of the π orbital increases and the π^* orbital decreases until they match. By symmetry, the potential curve from 90° to 180° is a reflection of the curve from 0° to 90° and the two orbitals switch in their bonding/antibonding character.

The coupling between these two configurations can be calculated:

$$\langle D_1|\mathbf{H}|D_2\rangle = (\pi\pi^*|\pi\pi^*) = \int d\boldsymbol{r}_1 d\boldsymbol{r}_2 \; \pi(\boldsymbol{r}_1)\pi^*(\boldsymbol{r}_1)\frac{1}{r_{12}}\pi(\boldsymbol{r}_2)\pi^*(\boldsymbol{r}_2) \tag{1.59}$$

As this integral is nonzero, these configurations have nonzero coupling, and at 90° the optimal linear coefficients are an even combination of bonding π and antibonding π^* configurations applied to a core reference state $|\psi\rangle$:

$$\frac{1}{\sqrt{2}}\left(\pi_\uparrow{}^\dagger \pi_\downarrow{}^\dagger - \pi_\uparrow^{*\dagger}\pi_\downarrow^{*\dagger}\right)\left|\psi\right\rangle \tag{1.60}$$

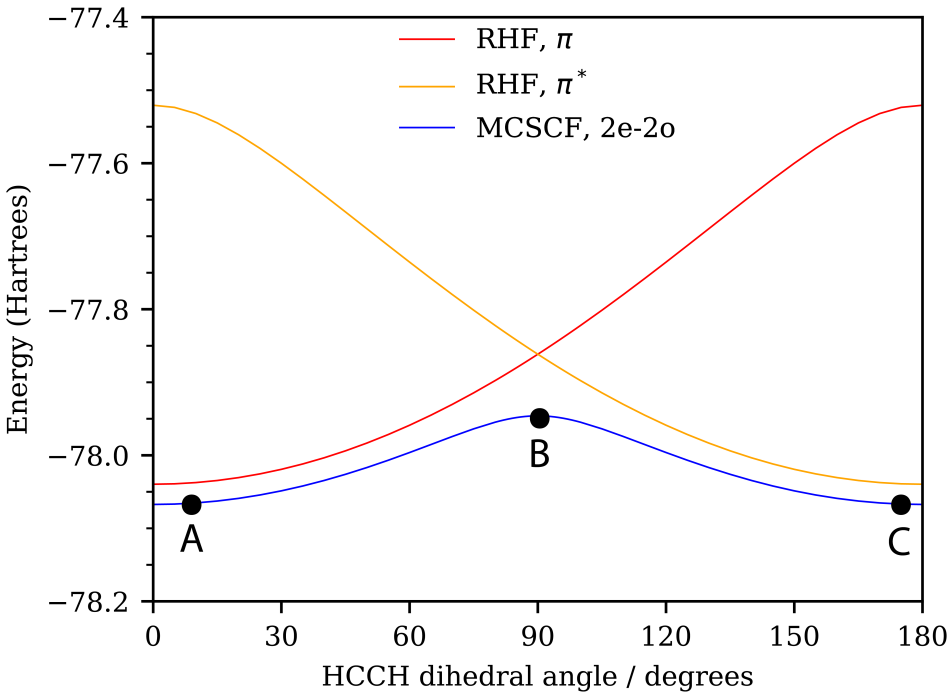

Figure 1.2: Plot of electronic energy as a function of HCCH dihedral angle in ethene. Single determinant wavefunctions with doubly occupied π and π^* orbitals (red and orange curves respectively, wavefunctions given in equation (1.58)) are good approximations to the equilibrium ground state wavefunction at points A and C respectively. These single-determinant states flip ordering and qualitative character at the D_{2d} symmetry point at 90°, and at this point (B), a multideterminant wavefunction (blue curve and wavefunction in equation (1.60)) is able to produce the correctly differentiable energy curve.

Just as in the dissociation case, when we rewrite this in the atomic orbital basis, the wavefunction is an open-shell singlet with electrons correlated between the two degenerate p-orbitals. Following the lowest-energy-single-determinant wavefunction in figure 1.2, we also see another hallmark failure of a single determinant wavefunction when treating strong correlation: a cusp at the degeneracy point.[53]

In a similar vein, charge transfer excited states[87–89] can require multiple determinants to accurately represent. In limiting cases, these states can be thought of as a long-range, single-electron excitation from a closed shell "donor" orbital to an unoccupied "acceptor" orbital. In order to be considered charge transfer, these donor and acceptor orbitals must be spatially separated. The similarity between this and the earlier two examples is no accident, and the H_2 ground state at dissociation can be viewed as a charge transfer process from a hydride anion to a hydrogen cation.

By similar logic, the spin symmetry of the Hamiltonian requires that the two configurations be degenerate, and unless there is no coupling between them, should be included with equal weight. However, while the linear response space of a single determinant (such as the CIS and ESMF wavefunctions, discussed more extensively in chapter 4) naturally accommodates these types of open-shell singles, they are not generally considered strongly correlated. Nonetheless, the wavefunctions commonly used to express them mathematically share the same open-shell diradical singlet character as dissociating H_2.

Though these are clear, limiting cases in which strong correlation appears, lines are not always so clearly drawn in larger molecular systems. Each of these examples presents two extremes, ultimately to point toward the physical and mathematical motifs characteristic of strong correlation. There is continuous transformation between the two extremes of a closed-shell single determinant reference and the strongly correlated multideterminant wavefunctions. This is most clear when looking at the two-configuration CI space vectors as a function of the energy difference:

$$\boldsymbol{H} = \begin{bmatrix} E_1 & H_{12} \\ H_{12} & E_2 \end{bmatrix} \tag{1.61}$$

After diagonalizing this matrix, we obtain the energies:

$$E_\pm = E_{av} \pm \frac{\sqrt{\Delta E^2 + 4H_{12}^2}}{2} \tag{1.62}$$

and normalized eigenvectors, which can be written in terms of a single angular parameter, θ:

$$\boldsymbol{v}_+ = \begin{bmatrix} \cos(\theta_+) \\ \sin(\theta_+) \end{bmatrix}, \qquad \boldsymbol{v}_- = \begin{bmatrix} -\sin(\theta_-) \\ \cos(\theta_-) \end{bmatrix} \tag{1.63}$$

which can be written directly in terms of the ratio between energy gap and coupling element:

$$\theta_\pm = \tan^{-1}\left(\frac{1}{\sqrt{x^2+1} \mp x}\right), \qquad x = \frac{\Delta E}{2H_{12}} \tag{1.64}$$

Assuming nonzero coupling, θ goes from 0° to 45° to 90° as ΔE goes from $-\infty$ to 0 to ∞.

Each of the two linear coefficients in this example correspond to single determinant components in a linear wavefunction. The two endpoints at $\theta = 0^\circ$ and $\theta = 45^\circ$ correspond to

the closed-shell singlet and open-shell singlet states, respectively, the latter of which contains equal parts of the two determinant components. The ratio x determines the angle, and at large magnitudes, gives single configuration wavefunctions as Hamiltonian eigenstates. By contrast, as x nears zero – which occurs as the energy difference ΔE approaches zero, and as the coupling element H_{12} increases without bound – the open-shell singlet becomes an eigenstate. In short, there a continuum between the two extremes of closed- and open-shell singlet, and energy difference and interstate coupling are the two components that directly determine where we are on this scale.

1.8 Strong Correlation Overview

These three examples are far from a complete account of how strong correlation manifests in wavefunctions. In summary, strong correlation is often marked by the following:

i) Stretched chemical bonds, twisted double bonds, and nonequilibrium geometries.

ii) Degenerate configurations in the linear determinant space.

iii) Non-analytic features such as derivative discontinuities in electronic energy curves.

iv) Fractionally occupied natural orbitals in the 1-RDM.

Aside from some limiting examples, strong correlation is not generally a binary quality of an electronic wavefunction – as natural orbital occupation numbers can vary continuously between zero and one. As a result, it can be difficult to clearly separate and diagnose the presence of strong correlation in more complex examples. In practice, orbital occupations in the 1-RDM that deviate strongly from fully occupied or fully vacant are considered strongly correlated. And although we have presented clear interpretations for the limiting cases of uncorrelated and strongly correlated wavefunctions, in more complex systems, it can be difficult to interpret intermediate occupations numbers in such a straightforward way.

Due to the contrasting manifestations of strong correlation and weak correlation, different theoretical approaches are best suited for each. Most methods in quantum chemistry primarily address weak correlation, as the most common use cases – e.g. organic molecules around equilibrium geometries – require no dedicated description of strong correlation. Density functional theory,[66,67] coupled cluster theory,[90,91] and Møller-Plesset perturbation theory[92] are all notable examples that rely on a single-determinant framework. In practice, DFT relies on heuristically constructing an accurate one-body Hamiltonian via the single-determinant Kohn-Sham approach;[67] Møller-Plesset is finite-order perturbation theory around a single determinant; and CCSD can be framed as a more sophisticated summation of perturbative energy diagrams that remains centered on a single determinant.

Strong correlation, by contrast, is most often treated using multideterminant methods. These approaches vary, and included truncated CI methods (MRCI,[53,93,94] SCI[58,95]); active space methods (CASSCF,[44,96–104] RASSCF[105–107]) and their perturbative extensions (CASPT2,[108,109] RASPT2,[110] NEVPT2[111–113]); stochastic methods (VMC,[114] DMC,[114] FCIQMC[115–117]); and others (MRCC,[118] Gutzwiller Factors,[119,120] RVB[121]). Many of these methods rely on a expanding the wavefunction into a limited subspace of the full CI space, and while the size of this linear expansion is manageably small in relatively simple systems,

in larger molecules, a qualitatively correct description of complex reaction processes can easily require linear determinant spaces so large as to be beyond computational reach.

Complete Active Space Self-Consistent Field (CASSCF) methods build these multideterminant spaces to include all configurations necessary to completely correlate a subset of orbitals, dubbed the "active space", and outside of this, orbitals are either fully occupied or fully unoccupied. An active space is specified by selecting a particular number of occupied and unoccupied orbitals, typically surrounding the HOMO-LUMO interface, and generated by taking all possible excitations between orbitals in this space. Configurations so generated form a basis of the multideterminant space used in the following multireference calculation. However, active space selection is not always straightforward, and often relies on practitioner intuition in practice, though methods have been proposed[122,123] to select orbitals in an automated and efficient way. Active spaces can be exhaustively checked, and can be considered to be large enough when increasing their size does not qualitatively affect the wavefunction, most easily judged by active space natural orbital occupation numbers.

CAS methods are the most exhaustive – and expensive – ways of constructing these spaces, but can guarantee that all correlation within the active space is accounted for. Variations like Restricted Active Space (RAS) methods break this active space into smaller subspaces to avoid correlating energetically distant orbitals in different subspaces.[105,108] In principle, combinatorially generating configurations from an active space is only one way to define a linear determinant space, and Selective Configuration Interaction (SCI)[124] or Density Matrix Renormalization Group (DMRG) do so respectively by perturbatively constructing this space or by approximating linear coefficients as a product of matrices.[45,125–128] This space can also be explored stochastically with Full Configuration Interaction Quantum Monte Carlo (FCIQMC).[115–117] Though these approaches expand what is possible, they ultimately end up succumbing to exponential cost-scaling, either directly through the size of the active space in CASSCF, or indirectly through statistical error in methods like FCIQMC.[117]

Though active spaces are combinatorial in size and suffer from exponential scaling, this cost-scaling depends only on the size of the active space, which can be tuned to capture only the strongest pieces of correlation. The size of the active space can be very sensitive to the molecular orbital basis, and so orbitals are often variationally optimized with respect to the energy in order to reduce the size of the active space. In Multiconfigurational Self-Consistent Field (MCSCF) theory,[53] the electronic wavefunction is written as

$$|\Psi\rangle = \sum_i c_i \exp\left(\mathbf{X}\right)|D_i\rangle\,, \qquad \mathbf{X} = \sum_{ia} X_i^a \left(\mathbf{a}^\dagger\mathbf{i} - \mathbf{i}^\dagger\mathbf{a}\right) \tag{1.65}$$

Electronic ground states are obtained by minimizing the energy with respect to both the linear parameters c_i and nonlinear parameters X that describe a rotation in the molecular orbital basis:

$$E_0 = \min_{c,X} \frac{\langle\Psi(c,X)|\mathbf{H}|\Psi(c,X)\rangle}{\langle\Psi|\Psi\rangle} \tag{1.66}$$

Due to the relative simplicity of the linear CASCI equations, optimization often proceeds in a two-step procedure:

i) For constant $\boldsymbol{X}$, obtain $\boldsymbol{c}$ that minimizes the energy, which can be written as a CASCI eigenvector equation:

$$\boldsymbol{H}\boldsymbol{c} = E_0\boldsymbol{c}, \qquad \boldsymbol{H}_{ij} = \langle D_i|\mathbf{H}|D_j\rangle \tag{1.67}$$

ii) For constant $\boldsymbol{c}$, locally expand the CASCI energy as a function of the orbital basis, and approach a nearby stationary point according to standard nonlinear optimization algorithms.

It seems natural and straightforward to separate the variables on whether they are or are not linear parameters, which does help in many cases.[53]

In contrast to single-determinant methods, symmetry-preserving open-shell excited states can be simply obtained by calculating higher roots of the CASCI equations. However, this comes with its own set of tradeoffs, as without modification, these excited states are obtained using orbitals optimal for the ground state. Simply selecting a higher root of the CASCI equations and searching for optimal excited-state orbitals is not an immediately stable proposition, as minimizations used to obtain ground states orbitals must now be recast as excited-state saddle-point searches.[129] As a result, the standard tactic is to instead find orbital stationary points of a linear combination of CASCI root energies in what is called state-averaged CASSCF:[30]

$$E_{avg} = \sum_i \alpha_i E_i \tag{1.68}$$

While more straightforward and stable, state-averaged orbitals straddle multiple different states, and when these states strongly differ, can inflate the size of active space necessary to accurately depict any particular state in the average. State-specific orbital optimization[31,32] will be discussed in more detail in chapter 4, and we will show that these can require careful modifications to be variationally stable.

1.9 Weak Correlation and Wavefunction Cusps

Coulombic repulsion between electrons is responsible for their correlation, and so far we have only discussed the regime of strong correlation, where this repulsion qualitatively impacts the structure of the electronic state. In contrast to strong correlation, weak electronic correlation manifests as a multitude of smaller corrections, and their cumulative effect results in small changes to the reference wavefunction, reflected mathematically by small deviations in natural orbital occupation numbers. Though weak correlation is more subtle, its effects can be significant,[77] as it is necessary to describe van der Waals interactions[72–74] and wavefunction cusps.[114,130,131,131–136] As strong and weak correlation effects dominate in opposing physical limits,[137–139] they are often most efficiently treated using different mathematical techniques. For instance, while weak correlation's multitude of corrections can be mostly corralled by perturbation theory[140] or coupled cluster theory,[90,91,141,142] the large corrections necessary for strong correlation often cannot be.

Capturing both strong and weak correlation together remains a serious challenge. Theories that exhaustively capture strong correlation, such as CASSCF or RASSCF,[44,143] scale exponentially with the size of the active space. Theories that try to instead sidestep this often still inherit exponential scaling nonetheless, such as FCIQMC which suffers exponentially growing noise due to the sign problem,[117] or DMRG which scales exponentially in all but one-dimensional systems.[45] Other approaches instead adopt approximations to avoid this scaling, such as Fixed Node Diffusion Monte Carlo[144,145] (FN-DMC) which fixes nodes to avoid exponential noise,[146] or Valence Bond theory[147,148] which concedes many internal

degrees of freedom. In addition, as these methods are designed to most completely capture the effects of strong correlation, their description of weak correlation can often be incomplete. Though this concern can be allayed by following up with perturbative corrections, doing so can be expensive (due to the large size of the wavefunction) and insufficient (due to low-orders of perturbation theory accessible) in complex cases.

To illustrate weak correlation more concretely and to cement nonlinear wavefunctions as simultaneously expressive and compact, we will consider a subtle feature of the exact electronic wavefunction that can be difficult to capture. The hydrogenic wavefunction contains a derivative discontinuity – a cusp – that exactly cancels the singular nuclear Coulomb potential upon electron-nuclear coalescence. In atomic units, the ground state wavefunction of the hydrogen atom is spherically symmetric and given by a normalized exponential that decays based purely on distance from the nucleus:

$$\psi(r) = Ne^{-r}, \qquad N^{-2} = \int |\psi(\boldsymbol{r})|^2 d\boldsymbol{r} \tag{1.69}$$

The Hamiltonian is the sum of a kinetic energy term (the radial Laplacian) and a potential energy term (the nuclear Coulombic potential).

$$\mathbf{H}\psi(r) = -\frac{1}{2}\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}Ne^{-r}\right) - \frac{Ne^{-r}}{r} \tag{1.70}$$

After going through the algebra while keeping the two terms separate, the singular Coulombic potential is exactly counter-balanced by an equivalent and opposite term that appears in the kinetic energy:[130]

$$\mathbf{H}\psi(r) = \left[\frac{Ne^{-r}}{r} - \frac{Ne^{-r}}{2}\right] - \frac{Ne^{-r}}{r} \tag{1.71}$$

By direct analogy, one expects that, in molecular systems, singular Coulombic terms in the Hamiltonian would cause wavefunction cusps to form at particle coalescence points. While the simple Hydrogen example directly illustrates the presence of this cusp for the electron-nucleus Coulomb potential, similar analysis can be applied to electron-electron coalescence. A thorough analysis of Coulombic operators in general along these lines brings us to the Kato cusp conditions, which formally quantifies wavefunction derivatives at these cusps.[130] These conditions vary based on atomic charge for nuclear-electron coalescence or relative spin for electron-electron coalescence.

The atomic orbital basis functions used in quantum chemistry are written as small combinations of products between radial gaussian functions and spherical harmonic angular functions.[52,53,149] For instance, a basis function for an atom at coordinates $\boldsymbol{R}_I$ is of the form:[53]

$$\phi_I(\boldsymbol{r}) = \sum_i c_i \exp\left(-\gamma_i|\boldsymbol{r} - \boldsymbol{R}_I|^2\right) Y_l^m(\Omega) \tag{1.72}$$

with spherical angle Ω signifying angular coordinates centered at the given single atomic position. These gaussian functions are tuned to mimic the low-energy orbitals surrounding each atom, and in routine calculations, are limited to accommodate only a few shells beyond the valence, with more added on for polarization effects or diffuse Rydberg states in

molecular ions.[53] Gaussian spherical harmonics are chosen as an (often worthwhile) compromise between accuracy and tractability over the exponential radial functions that serve as exact wavefunctions in single-electron atoms. This tradeoff is made because the integrals (eqs. (1.31)) required to transform the antisymmetric wavefunction to the linear determinant space can be written and evaluated analytically for radial gaussian functions but not for radial exponentials. However, in the gaussian basis, electron-nuclear cusp conditions can never be fulfilled since every radial gaussian has a vanishing derivative at its corresponding atomic origin.[53]

Further, while these electron-nuclear cusps could be described, in principle if not in practice, with exponential basis functions, this is not true of electron-electron cusps. The electron cusp condition applies whenever two electrons meet, which can happen at any point in space, and an accurate wavefunction description would be written explicitly in terms of interelectronic distances. However, evaluating the requisite integrals with wavefunction that include these terms is extremely difficultand limits the use of explicitly-correlated r_{12} and f_{12} methods[132–136,150] to simple systems.

Though methods based on analytic integration have difficulty directly addressing electronic cusps, stochastic position-space Monte Carlo methods, which randomly sample electronic positions, can natively accommodate correlation factors based on interelectronic distances. These Monte Carlo methods forgo the basis-function integrals in equation 1.31 in favor of stochastically sampling the many-body energy integral, which requires only evaluation of wavefunction amplitudes (and some related quantities) as a function of electronic positions. As inter-electronic distances are trivial to evaluate given a set of electronic coordinates, we introduce a set of factors written directly in terms of electron separation coordinates, that easily satisfy both cusp conditions exactly:

$$J_1(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n) = \prod_{iI} \exp\left(u_I(|\boldsymbol{r}_i - \boldsymbol{R}_I|)\right), \qquad \frac{\partial}{\partial R_i} u_I = -Z \tag{1.73}$$

$$J_2(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_n) = \prod_{i<j} \exp\left(u_2(|\boldsymbol{r}_i - \boldsymbol{r}_j|)\right), \qquad \frac{\partial}{\partial R_i} u_2 = \begin{cases} -1/2 & \text{same spin} \\ -1/4 & \text{opposite spin} \end{cases} \tag{1.74}$$

by imposing conditions[114,151,152] on the derivatives of u_I and u_2. In modern implementations,[114,153] the functions present in the exponent are parameterized as a minimal polynomial spline, and require as few as ten parameters for each function index.

Though we have recalled these factors to illustrate a concrete solution to the problem of electronic cusps, these factors are the beginning of a many-body expansion.[114,152,154] Much like these cusp-correcting Jastrows, higher-order components are multiplicative symmetric factors and are designed not to disrupt either wavefunction antisymmetry or the already-fulfilled cusp conditions.[155–157] Conveniently, Pauli exclusion ensures that components that depend on three or more electronic coordinates are much less important, as at least two of the electrons in close proximity will share a spin state. Higher-order terms thus involve nuclear coordinates, and much like molecular orbitals, are written in terms of atom-centered gaussian basis functions:

Exponential factors have also been explored in determinant space, where the nullpotency of creation and destruction operators help simplify and truncate terms present in the series expansion. The exponential cluster operator from coupled cluster is perhaps the most

well-known exponential operator in this context, and is famously effective at accurately capturing weak correlation.[90] The BCH expansion[158] undergirds the power of coupled cluster theory, and after several mild approximations, is used to show that infinite series of commutator expressions terminates at fourth order. Gutzwiller factors,[119,120] designed to capture insulator-to-conductor transitions in the Hubbard model,[138,139] are similar in form but different in function, and can be used to capture the strong correlation of localized electrons in a Wigner crystal.[137] Following the broad success of the cluster operator, more general correlation factors have been developed in the context of Valence Bond wavefunctions,[147,159–161] where the relative complexity of the AGP reference state encourages stochastic optimization methods.[162,163]

While second-quantized space is limited by the wavefunction basis, Gutzwiller factors in particular show that exponential factors can potentially describe strong correlation in an efficient way. Not only can position-space Monte Carlo naturally accommodate cusp-correcting Jastrow factors, but it also has the natural advantage of low cost-scaling, as it can skip the costly evaluation and storage of two-electron integrals. However, cusp-correcting Jastrow factors focus only on accurately capturing the intricate and close-range details surrounding wavefunction cusps and do not describe qualitative adjustments to the wavefunction necessary for strong correlation. In chapters 2 and 3, we show how to formulate a nonlinear factor that naturally describes the strong correlation of bond-breaking that is mathematically based on four-body position-space Jastrow factors and conceptually based on the projective action of Gutzwiller factors. The remainder of this chapter will be dedicated to reviewing the stochastic Monte Carlo Framework that these factors are posed in to provide a more complete context for their development.

1.10 Quantum Monte Carlo

Quantum Monte Carlo (QMC) is a broad umbrella term that encompasses stochastic approaches to solving the Schrödinger equation. While QMC includes algorithms like AFQMC[164] and FCIQMC,[115] which sample energy contributions from determinants in linear CI space, here we focus instead on two position-space QMC methods in particular. Variational Monte Carlo (VMC) and Diffusion Monte Carlo (DMC)[114,144,165] are of particular interest to us because they can naturally accommodate nonlinear modifications to the wavefunction based on electronic coordinates. The nonlinear Jastrow factors discussed later in chapters 2 and 3 are incompatible with the integral transformations described in equation (1.31) yet can capture wavefunction details that would otherwise require lengthy determinant expansions. In this section, we review the theory behind VMC and DMC in preparation of this later discussion, where we show how a nonlinear factor can capture the effects of strong correlation when applied to a single-determinant wavefunction.

In the context of VMC, the energy integral is evaluated directly using randomly-sampled electronic positions:[166]

$$E = \int |\Psi|^2 \frac{\mathbf{H}\Psi}{\Psi} = \left\langle \frac{\mathbf{H}\Psi}{\Psi} \right\rangle_{|\Psi|^2} \tag{1.75}$$

The energy is evaluated by calculating the statistical average of the local energy:

$$\frac{\int d\boldsymbol{x}\ \Psi(\boldsymbol{x})\mathbf{H}\Psi(\boldsymbol{x})}{\int d\boldsymbol{x}\ |\Psi(\boldsymbol{x})|^2} = \frac{\int d\boldsymbol{x}\ |\Psi(\boldsymbol{x})|^2 \frac{\mathbf{H}\Psi(\boldsymbol{x})}{\Psi(\boldsymbol{x})}}{\int d\boldsymbol{x}\ |\Psi(\boldsymbol{x})|^2} = \frac{1}{N}\sum_{\boldsymbol{x}_i\sim|\Psi|^2}^{N} \frac{\mathbf{H}\Psi(\boldsymbol{x}_i)}{\Psi(\boldsymbol{x}_i)} \tag{1.76}$$

using electronic positions sampled from the probability density of a given wavefunction Ψ.

Drawing samples from the wavefunction's probability density is not a trivial task, however, as this distribution is, in general, a many-dimensional object with unknown normalization. These samples are often generated by memoryless random walks in position space, known as Markov chains.[167,168] In a Markov chain, configurations are sequentially and stochastically generated according to a transition probability function that depends only on the current state. So if we start at some initial configuration x_1, our next configuration x_2 would be sampled according to a transition probability distribution $T(x_1 \to x_2)$ that depends only on x_1 and x_2. As a result, Markov chains are often understood as a discrete walker object probabilistically navigating a set of pair-wise connections between states. Markov chains are constructed purely by defining this transition probability function, and can be tuned so that the frequency at which a particular state is encountered follows a stable, prescribed probability distribution.

In order to sample this stationary probability distribution, Markov chains must meet a balance condition.[169] This balance condition essentially states that the probability of encountering a certain state must be equal the net probability flow into that state:

$$P(x_1) = \sum_x P(x)T(x \to x_1) \tag{1.77}$$

A stronger condition – used in many Monte Carlo algorithms due to its relative simplicity – is the detailed balance condition,[170] where the net probability flow between any two states is zero:

$$P(x_1)P(x_1 \to x_2) = P(x_2)P(x_2 \to x_1) \tag{1.78}$$

From this the balance condition can be derived by summing over all states x_i:

$$\sum_{x_i} P(x_i)P(x_i \to x) = \sum_x P(x)P(x \to x_i) = P(x) \tag{1.79}$$

As we want the sampled distribution to be $|\Psi(\boldsymbol{r})|^2$, imposing the detailed-balance condition gives us an equation for the relative transition probability:

$$\frac{|\Psi(\boldsymbol{r}_i)|^2}{|\Psi(\boldsymbol{r}_j)|^2} = \frac{T(\boldsymbol{r}_j \to \boldsymbol{r}_i)}{T(\boldsymbol{r}_i \to \boldsymbol{r}_j)} \tag{1.80}$$

Often, the transition probability is further broken up into two components whose product gives the net transition probability.[171] These two components are known the proposal (T) and acceptance (A) distribution, and respectively signify the probability of proposing a step to a particular state and the probability that that step is taken.

$$\frac{P(\boldsymbol{r}_j \to \boldsymbol{r}_i)}{P(\boldsymbol{r}_i \to \boldsymbol{r}_j)} = \frac{T(\boldsymbol{r}_j \to \boldsymbol{r}_i)A(\boldsymbol{r}_j \to \boldsymbol{r}_i)}{T(\boldsymbol{r}_i \to \boldsymbol{r}_j)A(\boldsymbol{r}_i \to \boldsymbol{r}_j)} \tag{1.81}$$

In stochastic algorithms, more samples reduce the statistical error of the final result, and it is often wise to make these distributions as simple as possible to reduce computational costs. Steps are often proposed randomly according to a symmetric distribution that depends only on the distance between two positions so that the proposal distribution terms cancel in the above ratio. As a result, we are left to define the acceptance probability to satisfy equation (1.81), which can be most simply done by setting:

$$A(\boldsymbol{r}_j \to \boldsymbol{r}_i) = \begin{cases} \frac{|\Psi(\boldsymbol{r}_i)|^2}{|\Psi(\boldsymbol{r}_j)|^2} & : \quad \frac{|\Psi(\boldsymbol{r}_i)|^2}{|\Psi(\boldsymbol{r}_j)|^2} \leq 1 \\ \\ 1 & : \quad \text{otherwise} \end{cases} \tag{1.82}$$

Generating Markov chains using these proposal and acceptance distributions is known as the Metropolis-Hastings algorithm,[171] and is summarized below.

i) Starting at a position $\boldsymbol{r}_i$, generate a proposed position $\boldsymbol{p}$ by sampling a symmetric proposal distribution $R(\boldsymbol{r}_i \to \boldsymbol{p})$.

ii) Calculate the relative probabilities of each of these positions by using the ratio of stationary probability distribution function evaluated at the original and proposed points: $\pi = \frac{P(\boldsymbol{p})}{P(\boldsymbol{r}_i)}$.

iii) Generate a random number from the uniform distribution $\rho \sim U(0,1)$ and accept the proposed step if $\rho < \pi$, in which case $\boldsymbol{r}_{i+1} = \boldsymbol{p}$

iv) Otherwise, reject the proposed step, in which case $\boldsymbol{r}_{i+1} = \boldsymbol{r}_i$.

v) Repeat

This sampling procedure is not without limitations. As we are sampling a many dimensional Cartesian space, there are an infinite number of states for us to consider, and a common proposal distribution is a gaussian centered at the current position. This creates statistical dependence between successive samples, as the proposal distribution is short-range, that goes against the presumption of the statistical independence of samples when calculating expectation values. This dependence can be accounted for by analyzing the autocorrelation of the measured quantity (such as energy) between successive samples, defined as the normalized covariance between pairs of time-staggered samples:[172–174]

$$C(t) = \frac{1}{N}\sum_i C(x_i, x_{i+t}) = \frac{\sum\limits_i \langle (\overline{x} - x_i)\,(\overline{x} - x_{i+t})\rangle}{\sum\limits_i \langle (\overline{x} - x_i)^2\rangle} \tag{1.83}$$

And finding the appropriate t for which this autocorrelation sufficiently small to claim that staggered samples are statistically independent. Empirically, autocorrelation can often be written as a decaying exponential function that depends on the number of samples between these staggered samples (t) with an autocorrelation timescale τ_A:

$$C(t) \approx e^{-t/\tau_A} \tag{1.84}$$

Because this autocorrelation function is itself subject to statistical issues, a blocking analysis[175] is often done instead where segments of successive samples are averaged together to form a reduced sequence of blocked samples.

$$s_j(n_b) = \frac{1}{n_b} \sum_i^{n_b} x_{j*n_b+i} \tag{1.85}$$

In this analyis, block sizes are increased until the blocked samples are statistically independent and can be used to calculate measured quantities.

The Central Limit Theorem[176] states that the distribution from which we draw our estimated energy expectation value $\overline{E}$:

$$\overline{E} = \frac{1}{N} \sum_i E_i \tag{1.86}$$

converges to a normal gaussian, centered at the true energy value with a variance that is inversely proportional to the number of samples: $\frac{\sigma^2}{N}$. Our final expectation value can be viewed as a random variable sampled from a gaussian distribution that narrows as we collect more samples. The error of our expectation values is the standard deviation of this distribution, and is given by $\frac{\sigma}{\sqrt{N}}$. Here σ^2 is the wavefunction variance:

$$\sigma^2 = \langle \Psi | (\mathbf{H} - E)^2 | \Psi \rangle \tag{1.87}$$

which is a measure of the quality of our wavefunction, and indeed improving our wavefunction also reduces the statistical error. According to the zero-variance principle, as we approach a Hamiltonian eigenstate, the local energy ratio $\frac{\mathbf{H}\Psi}{\Psi}$ becomes a constant distribution and the variance goes to zero.[177]

To calculate the cost-scaling of Monte Carlo Algorithms, we write the total cost as the product of the total number samples by the cost per sample. In order to maintain a constant absolute error, the number of samples should scale to match the variance of our energy distribution. Wavefunction variance is an extensive quantity, and scales linearly with the size of the system n, giving the error expression:

$$\epsilon = \sqrt{\frac{n\sigma^2}{N}} \tag{1.88}$$

And so to maintain a constant absolute error, the number of samples must also scale linearly with the size of the system:

$$N \propto n \tag{1.89}$$

Though the number of samples necessary to reduce this error to chemical accuracy (approximately 1 $\mathrm{mE_h}$ or $0.02eV$) can run into the millions, multiple Markov chains can be run independently. This allows QMC to take full advantage of massively parallel computer architectures and has been successfully used on systems in excess of a thousand electrons.[178–180]

While the cost per sample depends on the form of the wavefunction, the Multi-Slater Jastrow wavefunction is a popular choice:[114]

$$\Psi(\boldsymbol{x}) = \exp(J_1(\boldsymbol{x})) \exp(J_2(\boldsymbol{x})) \left[\sum c_i D_i(\boldsymbol{x}) \right] \tag{1.90}$$

The costly step here is the numerical evaluation of the determinant components at particle ensemble positions, $D_i(\boldsymbol{x})$, as the dimension of the determinant matrix scales linearly with the number of electrons. As the determinant of a matrix is equal to the product of its eigenvalues, this is straightforwardly done by diagonalizing the numerical matrix (or, equivalently, LU decomposition) at each sampled position at a cost-scaling of $O(n^3)$. Combining these two factors gives a total cost scaling of $O(n^4)$, so long as the number of determinant components in the wavefunction do not meaningfully scale with the system size.

1.11 Variational Monte Carlo

In Variational Monte Carlo, the electronic ground state is found as the variational energy minimum with respect to a set of wavefunction parameters:

$$\min_{\boldsymbol{p}} E(\boldsymbol{p}) = E_0 \tag{1.91}$$

$$E(\boldsymbol{p}) = \frac{\langle \Psi(\boldsymbol{p}) | \mathbf{H} | \Psi(\boldsymbol{p}) \rangle}{\langle \Psi(\boldsymbol{p}) | \Psi(\boldsymbol{p}) \rangle} \tag{1.92}$$

Many approaches to ground-state wavefunction optimization in VMC can be summarized as non-stochastic algorithms that use stochastically-calculated energy derivatives.[181,182] Energy derivatives are obtained by analytically differentiating the energy integral and sampling the differentiated integrand using Monte Carlo. A wide variety of minimization approaches have been proposed, including stochastic reconfiguration,[183–186] gradient descent,[187–190] and hybrid optimization methods.[191,192]

The most commonly used and widely-known approach in VMC is known as the Linear Method,[189,193,194] where the wavefunction is expanded into its parameters' tangent space:

$$\Psi_{lin}(\boldsymbol{p}; \Delta\boldsymbol{p}) = \Delta p_0 \Psi(\boldsymbol{p}) + \sum_i \Delta p_i \frac{\partial \Psi}{\partial p_i}(\boldsymbol{p}) \tag{1.93}$$

The parameter step $\Delta\boldsymbol{p}$ is calculated by imposing a stationarity condition on this wavefunction's energy:

$$\nabla_{\Delta\boldsymbol{p}} \langle \Psi_{lin} | \mathbf{H} | \Psi_{lin} \rangle - E_{lin} \nabla_{\Delta\boldsymbol{p}} \langle \Psi_{lin} | \Psi_{lin} \rangle = 0 \tag{1.94}$$

and can be directly written as an eigenvalue problem where matrix elements are expectation values of wavefunction parameter derivatives:

$$\boldsymbol{H}\Delta\boldsymbol{p} = \boldsymbol{S}\Delta\boldsymbol{p}\ E_{lin} \tag{1.95}$$

$$\psi^i = \begin{cases} \Psi(\boldsymbol{p}) & i = 0 \\ \frac{\partial \Psi}{\partial \boldsymbol{p}_i} & i \neq 0 \end{cases} \tag{1.96}$$

$$\boldsymbol{H}_{ij} = \langle \psi^i | \mathbf{H} | \psi^j \rangle \tag{1.97}$$

$$\boldsymbol{S}_{ij} = \langle \psi^i | \psi^j \rangle \tag{1.98}$$

The lowest-energy eigenvector $\Delta\boldsymbol{p}$ is determined, and the parameters are updated accordingly:

$$\boldsymbol{p}^{(i+1)} = \boldsymbol{p}^{(i)} + \Delta\boldsymbol{p}^{(i)} \tag{1.99}$$

These matrix elements are stochastically calculated in the same way as the energy in equations (1.76), with local derivative ratios taking the place of the local energy:

$$\langle \psi^i | \mathbf{H} | \psi^j \rangle = \int d\boldsymbol{x} |\Psi(\boldsymbol{x})|^2 \frac{\psi^i}{\Psi} \frac{\mathbf{H}\psi^j}{\Psi} \tag{1.100}$$

$$= \frac{1}{N} \sum_{\boldsymbol{x}_k \sim |\Psi|^2}^{N} \frac{\psi^i(\boldsymbol{x}_k)}{\Psi(\boldsymbol{x}_k)} \frac{\mathbf{H}\psi^j(\boldsymbol{x}_k)}{\Psi(\boldsymbol{x}_k)} \tag{1.101}$$

Once obtained, the eigenvalue problem in eq (1.95) is solved to provide the next proposed step and is repeated until the stationary condition is approximately met.

As stochastically-sampled quantities, Linear Method matrices – and thus the proposed parameter updates – contain statistical error. In order to prevent this error from setting an absolute limit on the step accuracy, correlated sampling is employed to instead sample energy differences across the proposed step. Correlated sampling in this context simply means that we sample the two distributions present in the energy difference using the same set of position-space samples:[167,195–197]

$$E(\boldsymbol{q}) - E(\boldsymbol{p}) = \frac{1}{N} \sum_{\boldsymbol{x} \sim |\Psi|^2} E(\boldsymbol{q}; \boldsymbol{x}) - E(\boldsymbol{p}; \boldsymbol{x}) \tag{1.102}$$

According to the central limit theorem, the variance of this estimated difference is proportional to the variance of the difference distribution:

$$\sigma^2[E(\boldsymbol{q})_i - E(\boldsymbol{p})_i] = \sigma^2[E(\boldsymbol{q})] + \sigma^2[E(\boldsymbol{p})] - 2\mathrm{cov}[E(\boldsymbol{q}, \boldsymbol{p})] \tag{1.103}$$

The key to correlated sampling is the strong dependence of these local energy distributions on each other; so long as the parameter step is small, the covariance term should be large.[195] For instance, if we write:

$$E(\boldsymbol{q}) = E(\boldsymbol{p}) + \sum_i \Delta \boldsymbol{p} \frac{\partial E}{\partial p_i} \tag{1.104}$$

we find that the variance of the energy difference is proportional to the size of the step taken:

$$\sigma^2[E(\boldsymbol{q}) - E(\boldsymbol{p})] = 2\sigma^2[\Delta \boldsymbol{p} \cdot \nabla_p E] \tag{1.105}$$

Sampling these Linear Method matrix elements can be complicated by the fact that these derivative ratios can be singular at particular points. While the expectation values are finite – as singular derivative ratios are paired with nodes in the wavefunction probability distribution – the variance of this distribution can nevertheless be infinite.[198] This will generally happen when nodes of the tangent space functions do not match those of the zeroth order wavefunction $\Psi(\boldsymbol{p})$, and troublesome derivatives in this regard include orbital parameters and linear CI coefficients. This can be remedied by instead picking samples from a nodeless guiding wavefunction[199,200] that is designed to remove the nodes in the wavefunction probability distribution by adding a positive shift.

In theory, the VMC engine is agnostic to the functional form it is minimizing, as sampling and optimization methods depend only on evaluation of the integrand and its derivatives.

While ground electronic states are a commonly-sought endpoint, and can be straightforwardly obtained by variationally minimizing the energy, excited states are not so simple. Energy minimization is not the only way to obtain Hamiltonian eigenstates, and we can instead minimize functions of the energy such as the variance:[33,201,202]

$$\sigma^2[\Psi] = \frac{\langle\Psi|(E-\mathbf{H})^2|\Psi\rangle}{\langle\Psi|\Psi\rangle} \tag{1.106}$$

Variance is notable in particular because it is always non-negative and attains a minimum value of zero at each eigenstate. Variance-minimum wavefunctions are not necessarily superior to one obtained through energy minimization,[202] but as all eigenstates are variance minima, the Variance can be used to target excited states.

In VMC, the square Hamiltonian operator present in the variance expression can be conveniently calculated by stochastically resolving an identity between the two powers:

$$\langle\Psi|\mathbf{H}^2|\Psi\rangle \approx \int d\boldsymbol{x}\, \langle\Psi|\mathbf{H}|\boldsymbol{x}\rangle\, \langle\boldsymbol{x}|\mathbf{H}\Psi\rangle \approx \sum_{x\in|\Psi|^2} \left(\frac{\mathbf{H}\Psi(x)}{\Psi(x)}\right)^2 \tag{1.107}$$

However, the energy variance does not favor any excited state in particular, and variance minimization will often simply converge to the eigenstate closest the initial guess. To this end, a variety of state-selective variational principles – Hamiltonian-based functionals designed to target particular excited states – have been developed.[31,33,203] As will be discussed further in chapter 4, state-specific optimizations in more conventional second-quantized spaces can encounter subtle difficulties, and VMC offers a parallel alternative with high flexibility.

In contrast to second-quantized methods, where basis components are analytically integrated to simplify expressions, VMC requires only that a wavefunction be locally evaluable. This relaxed condition allows the use of explicit correlation factors which are written in terms of separation coordinates usually impermissible in analytic integrals. While this exposes VMC to stochastic error which can require a large number of samples to overcome, this error is well-understood, controllable, and scales only linearly with the size of the system. Cusp-correcting Jastrow factors in particular can be evaluated at trivial cost and can be used to satisfy wavefunction cusp conditions which are otherwise difficult to satisfy. However, VMC wavefunctions are still limited by the boundaries of the functional form used for the wavefunction, and in the next section, we will describe a method which can go beyond these boundaries.

1.12 Diffusion Monte Carlo

While VMC and Jastrow factors are effective at neatly wrapping up bundles of dynamic correlation in a small correlation factor, VMC is still limited by the functional form of the trial wavefunction. This limitation can be overcome by projector Monte Carlo methods, which use imaginary time-propagation operators as projection operators, effectively projecting out excited state components from an initial wavefunction guess. Diffusion Monte Carlo[114,144,145,166] is one such method, which seeks to sample the exact ground state by reweighting trial wavefunction components according to a well-approximated imaginary-time propagation operator

The formal solution to the Time-Dependent Schrödinger equation can be written as an exponential operator known as the propagator applied to an initial state:

$$|\Psi(t)\rangle = e^{-i\mathbf{H}t} |\Psi(0)\rangle \tag{1.108}$$

If we expand the wavefunction in the complete set of eigenstates of $\mathbf{H}$, this time-dependence can be isolated to an oscillating phase factor:

$$|\Psi(t)\rangle = \sum_k c_k e^{-iE_k t} |\Psi_k\rangle \tag{1.109}$$

If we take t as an imaginary number, the "time-dependence" of the wavefunction instead reweights each eigenstate by an exponential factor:

$$|\Psi(-it)\rangle = \sum_k c_k e^{-E_k t} |\Psi_k\rangle \tag{1.110}$$

After rescaling to cancel the ground-state exponential factor and taking the infinite-time limit, we recover the ground state wavefunction:

$$\lim_{t\to\infty} e^{E_0 t}\Psi(-it) = \lim_{t\to\infty} \sum_k c_k e^{(E_0-E_k)t} |\Psi_k\rangle = c_0 |\Psi_0\rangle \tag{1.111}$$

Excited states are reweighted by an exponentially-decaying factor and are duly projected out of the initial wavefunction.

The idea behind Diffusion Monte Carlo[114,144,165] is to statistically evaluate this propagator in position space to obtain and average local energy samples from the ground state distribution. To set up this sampling procedure, we start with a mixed expectation value of the energy that assumes we're already sampling the ground state.

$$E_0 = \frac{\langle\Psi|\mathbf{H}|\Psi_0\rangle}{\langle\Psi|\Psi_0\rangle} = \lim_{t\to\infty} \frac{\langle\Psi|\mathbf{H}\exp(-\mathbf{H}t)|\Psi\rangle}{\langle\Psi|\exp(-\mathbf{H}t)|\Psi\rangle} \tag{1.112}$$

Once projected, we should have an eigenstate of the Hamiltonian, and acting the Hamiltonian on this state should provide an accurate estimate of the energy. To sample the ground state wavefunction, we split the exponential operator into smaller time-slices as in path-integral formulations and resolve stochastically-evaluated identity operators between each time-slice.

$$\langle\Psi|\mathbf{H}\exp(-\mathbf{H}t)|\Psi\rangle = \langle\Psi|\mathbf{H}|\prod_{i=0}^{n}\int\left[d\boldsymbol{r}^{(i)}\exp(-\mathbf{H}t/n)|\boldsymbol{r}^{(i)}\rangle\langle\boldsymbol{r}^{(i)}|\right]|\Psi\rangle \tag{1.113}$$

Here we sample pathways of coordinates from the iterated integration:

$$\left\{\boldsymbol{r}^{(1)},\ldots\boldsymbol{r}^{(n)}\right\} \tag{1.114}$$

with each successive component obtained by sampling the propagator using a weighted Markov chain process.

$$\langle\boldsymbol{r}^{(i+1)}|\exp(-\mathbf{H}t/n)|\boldsymbol{r}^{(i)}\rangle \tag{1.115}$$

As the Hamiltonian depends both on electrons' position and momentum, values for these propagator elements require additional approximation to obtain. To evaluate the propagator, we symmetrically break the operator into kinetic and potential energy components using the Trotter decomposition formula:[204,205]

$$e^{-\mathbf{H}t} \approx e^{-\mathbf{V}t/2} e^{-\mathbf{T}t} e^{-\mathbf{V}t/2} \tag{1.116}$$

which is exact up to second order in t. The flanking potential energy pieces are evaluated at their respective positions, after which a Fourier Transform is used to obtain an analytic expression for the propagator elements:

$$\exp\left(-\frac{V(\boldsymbol{r}^{(i+1)}) + V(\boldsymbol{r}^{(i)})}{2}\right) \int d\boldsymbol{p} \exp\left(\frac{-\boldsymbol{p}^2\tau}{2}\right) \langle \boldsymbol{r}^{(i+1)}|\boldsymbol{p}\rangle \langle \boldsymbol{p}|\boldsymbol{r}^{(i)}\rangle \tag{1.117}$$

Importance sampling, discussed in the previous section, is performed in DMC just as it is in VMC, and for similar reason. Particle configurations are sampled from a trial wavefunction distribution according to the metropolis algorithm as in VMC. Doing this requires that we modify the propagator elements, which turn out similar to equation (1.113), only with the Hamiltonian being substituted out for the local energy. In the full path integral expression, we effectively sample the VMC wavefunction and reweigh local energies by products of propagator weights. This can become statistically unstable, so weights larger than one are cloned into multiple Monte Carlo walkers and simulated independently. Walkers with weight below a tolerance are deleted, and walker population can be adjusted to maintain an equilibrium population for convenience.[166]

The projector formalism seeks the ground state of the Hamiltonian in position space and has no knowledge of antisymmetry conditions. If nothing is done to prevent it, stochastically simulating the above path sampling in the DMC integral will experience interference from the nodeless, bosonic ground state. This manifests as a statistical error that scales exponentially with the size of the system and is referred to as the sign problem.[206] Sign problems in general are endemic to projector Monte Carlo formulations, and can also be broadly understood either as statistical interference with phase-shifted wavefunction samples or as high relative error encountered when sampling oscillatory integrands.

To avoid statistical issues, and to ensure that we converge to a propery antisymmetric wavefunction, we use the Fixed Node Approximation,[144,207,208] in which the nodes of the sampled distribution are constrained to match those of a trial wavefunction. This can be straightforwardly done by adding an infinite potential barrier conicident with the guiding wavefunction's nodal surface:

$$\mathbf{H}_{DMC} = \mathbf{H} + \mathbf{V}_{FN}, \qquad \mathbf{V}_{FN}(\boldsymbol{x}) = \begin{cases} 0 & : \ \Psi(\boldsymbol{x}) \geq 0 \\ \infty & : \ \Psi(\boldsymbol{x}) < 0 \end{cases} \tag{1.118}$$

In the sampling procedure, this fixed node potential can be implemented by simply deleting or rejecting the move of any walker that crosses a node of the trial wavefunction. DMC is already variational to within statistical precision,[209] and adding this always-positive potential piece maintains this variationality.

At first glance, this appears to disrupt the stochastic sampling quite seriously by barricading half of space from being sampled at all. The tiling property[210,211] assures us that

these spaces – related by permutational symmetry – are simply reflections of each other, and sampling either the positive or negative pocket exclusively does not affect our final result. There is the more subtle concern that these nodal pockets may be topologically disconnected, leaving us with inaccessible pockets that are sampled. The positive and negative nodal pockets in free particle systems are fully connected[210] giving a total of two equivalent nodal pockets, and properly correlated wavefunctions often share this property.[212]

DMC thus requires an antisymmetric guiding wavefunction to be used for importance-sampling as well as to define a nodal surface constraint. However, the quality of the trial wavefunction amplitudes limit only the statistical error of the energy, while the quality of its nodal surface affects the variational energy estimate. For instance, in the Slater-Jastrow wavefunction, the always-positive, symmetric Jastrow factor leaves the nodes of the wavefunction untouched while significantly reducing the statistical error in DMC. By contrast, the determinant expansion wholly defines the nodal surface of the wavefunction, and orbital and linear determinant coefficients can be changed to manipulate the wavefunction's nodal surface.

Despite its simple definition, the antisymmetric nodal surface is a complicated and difficult-to-visualize $3N-1$ dimensional surface in position space,[210,211,213–216] as it is defined by a single constraint:

$$\{x : \Psi(\boldsymbol{x}) = 0\} \tag{1.119}$$

It should be emphasized that antisymmetric nodal surfaces are qualitatively different from the nodes of molecular orbitals, and the two are not generally coincident. Antisymmetric nodes are most obviously associated with electron-electron coalescence, this condition defines only a $3N-3$ dimensional surface A full description of the nodal surface thus includes two more dimensions that branch off the coalescence surface, and a straightforward characterization of these nodes beyond the linear multideterminant expansion remains elusive.

Research has been put toward directly finding nodal surfaces that better match those of the true ground state.[217] As above, it is not clear how to judge the quality of these nodal surfaces simply by inspection of the wavefunction parameters. While some procedures that directly target this nodal surface have been put forward, nodal surfaces are often assumed to improve with a wavefunction's variational energy, on the ground that an exact wavefunction will have an exact nodal surface. While an orbitally-optimized VMC wavefunction will often have a variationally better nodal surface, this need not always be the case.

However, variational changes made to a nodal surface often does improve DMC results.[218] However, as the fermionic components wholly determine the nodal surface of the wavefunction, only orbital basis parameters can change the structure of the nodal surface. Though, being symmetric, Jastrows don't affect the DMC energy directly, they can nonetheless be crucial to obtain an accurate nodal surface. This is because Jastrow parameters can couple to the orbital parameters, and can straightforwardly allow the orbital optimization to focus its efforts on correcting the nodal surface rather than making ambiguous tradeoffs when correcting for wavefunction magnitudes.

To summarize, VMC stochastically evaluates energy functions and derivatives and is compatible with nonlinear position-space factors that can easily express wavefunction cusps and short-range correlations. DMC is a highly accurate projector Monte Carlo method that effectively reweights VMC wavefunction within a fixed-node approximation to sample the

Hamiltonian ground state variationally. While both have to manage idiosyncracies of statistical sampling, both scale quartically ($O(N^4)$) with the size of the system with standard trial wavefunctions, and are exceedingly compatible with massively-parallel computer architectures. In the next two chapters we show how to design a nonlinear factor to handle the strong correlation of bond-breaking without explicit reference to the linear determinant space.

2 Number-Counting Jastrow Factors

2.1 Abstract

We demonstrate that 4-body real space Jastrow factors are, with the right type of Jastrow basis function, capable of performing successful wave function stenciling to remove unwanted ionic terms from an overabundant Fermionic reference without unduly modifying the remaining components. In addition to greatly improving size consistency (restoring it exactly in the case of a geminal power), real-space wave function stenciling is, unlike its Hilbert space predecessors, immediately compatible with diffusion Monte Carlo, allowing it to be used in the pursuit of compact, strongly correlated trial functions with reliable nodal surfaces. We demonstrate the efficacy of this approach in the context of a double bond dissociation by using it to extract a qualitatively correct nodal surface despite being paired with a restricted Slater determinant, that, due to ionic term errors, produces a ground state with a qualitatively incorrect nodal surface when used in the absence of the Jastrow.

2.2 Introduction

Linear wave functions in quantum chemistry are fundamentally limited by their inability to compactly describe a strongly correlated electronic state, a difficulty that arises directly from the factorial growth of Hilbert space in the quantum many-body problem. The field of quantum chemistry has therefore long pursued sophisticated nonlinear forms for its approximate wave functions.[45,90,97,98,219] It remains a challenging task to design an effective ansatz that is simultaneously size-consistent (gives the same energy for independent systems whether treated individually or together) and variational (gives an upper-bound to the true energy) with a cost that scales polynomially with the system size. Wave function stenciling, a generalization of Gutzwiller's approach,[119] where a nonlinear correlation factor removes unsuitable terms from an overabundant Fermionic expansion, has shown to achieve these three properties,[162,220,221] and so appears to be a promising ansatz design paradigm.

The Hilbert space Jastrow Antisymmetric Geminal Power (JAGP) is a characteristic example of this stenciling approach, and while it has proven effective at capturing strong correlation during bond dissociations,[222,223] it is much less effective for capturing the full range of dynamic correlation effects. One way to understand this difficulty is to consider that its number-operator based Jastrow factor, which is central to its stenciling strategy,[220] can also be written as a very limited coupled cluster doubles operator.[222] This incomplete reproduction of the doubles operator only partially recovers dynamic correlation, and although the Hilbert space Jastrow factor is effective at making large changes to the wave function, it is unable to include the multitude of small changes required by dynamic correlation. By contrast, more traditional real-space Jastrow factors, especially when used in tandem with diffusion Monte Carlo (DMC),[114] are renowned for their ability to capture dynamic correlation. This pairing has been employed as a reliable substitute for the "gold standard" coupled cluster with singles, doubles, and perturbative triples (CCSD(T))[90] in cases where the latter's high cost scaling makes it untenable.[142] In addition, real space Variational Monte Carlo (VMC) approaches remain the only tractable ab initio method capable of performing the difficult geometry optimization encountered in extended conjugated systems,[161,224] further

motivating the development of improvements to existing real space Jastrow factors.

At present, we find that Jastrow factors largely play opposing roles in real space and Hilbert space. Can we develop a real space Jastrow factor that can deliver both the large changes required for stenciling and the small changes required for dynamic correlation? The best current option for encoding static correlation effects in real space is through a costly determinantal expansion; could we instead construct an affordable Jastrow factor that describes these effects which is compatible with DMC in strongly correlated regimes? In this chapter, we present a real space Jastrow factor that does this, we explain why previous Jastrow formulations were unable to capture these details, and we demonstrate the efficacy of this stenciling strategy when the stenciled wave function is a single Slater determinant. In addition, we offer some thoughts on the conditions required for the action of these stenciling Jastrows in real space, and we speculate on how these Jastrow factors might be extended to larger and more complex molecular systems.

The essential challenge we confront in constructing a real space stencil is this: we desire a multiplicative factor that is not only capable of making large changes to the wave function, but also one that variationally favors these large changes. Making large changes to a wave function through a smooth, well-behaved, multiplicative Jastrow factor requires that the factor take on a large degree of curvature, and unless carefully controlled, this curvature will tend to raise the kinetic energy of the wave function. This will lead the variational principle to deem term deletion energetically unfavorable, even in cases where it should not be.

To address this challenge, we present a new form of four-body Jastrow factor that is better-suited to hiding its curvature in regions of low wave function value, where it will minimally raise the kinetic energy. These Jastrows factors are known as number counting Jastrow factors, since they can be thought to count electrons within local regions of space, and broadly mimic the mechanism by which Hilbert space Jastrows achieve stenciling. Combined with traditional two-body Jastrows, these number counting Jastrow factors are able to correct a single Slater determinant for both static and dynamic correlation effects within a real space framework that is both compatible with diffusion Monte Carlo approaches and whose complexity is explicitly polynomial.

2.3 Mimicking Hilbert Space Jastrows

Let us begin by reviewing Hilbert Space Jastrow Factors (HSJFs), which may be written in terms of a matrix $\boldsymbol{F}$ and second-quantized number operators $\hat{n}_i$ within an orthonormal (and typically local) one particle basis,

$$e^{\hat{J}_{HS}} = \exp\left(\sum_{ij} F_{ij}\mathbf{n}_i\mathbf{n}_j\right) \tag{2.1}$$

Note that these can be thought of as four-body e-e-N-N Jastrow factors, as the indices i and j run over orbitals that are localized at or near the nuclei while the results of operating with the number operators tell us about the positions of up to two different electrons.

As number operators are idempotent and overall constant factors irrelevant, $\boldsymbol{F}$ can be

chosen such that the HSJF contains any number of Gaussian factors

$$\hat{\Gamma}_{W,\,\mathrm{HS}} = \exp\left(-\xi\left(N - \sum_{p\in W}\mathbf{n}_p\right)^2\right) \tag{2.2}$$

for use in wave function stenciling. Application of one of these factors to a Fermionic wave function effectively reweights each configuration in that wave function's expansion within this particular orbital basis according to a Gaussian distribution in the total occupancy of an orbital subset W. Provided that the "projection strength" ξ is sufficiently large, such a Gaussian factor acts as a stencil, removing any configuration in which the set of orbitals W contains an electron count differing from N. Given two or more molecular fragments, this effect can be used to eliminate any configurations in which a fragment possesses an unphysical charge (an "ionic configuration"), which turns out to be sufficient for restoring size consistency to a geminal power.[162] Crucially, this factor does nothing to components of the wave function which do not deviate from the prescribed pattern of subsystem electron counts, preventing the HSJF from raising the kinetic energy of configurations that survive the stencil.

Unfortunately, a direct translation of the HSJF into real space is problematic for QMC methods due to the nonlocal nature of a number operator's real space form,

$$\mathbf{n}_p = \sum_i \phi_p(\boldsymbol{r}_i)\int \phi_p^*(\tilde{\boldsymbol{r}}_i)d\tilde{\boldsymbol{r}}_i. \tag{2.3}$$

Efficient stochastic integrations of a wave function in real space hinge on the ability to evaluate local wave function values $\Psi(\boldsymbol{r})$, which is complicated by the number operators' nonlocality. Instead, we will seek a local function Q, associated with a region W, enclosed within a Jastrow factor of similar Gaussian form

$$\Gamma_{W,\,\mathrm{RS}} = \exp\left(-\xi\left(N - \sum_i Q(\boldsymbol{r}_i)\right)^2\right) \tag{2.4}$$

that permits efficient local evaluation and, thanks to the sum over all the electron positions r_i, maintains the bosonic symmetry required by the Jastrow factor to keep the overall wave function correctly antisymmetric.

In order to mimic the effects of a HSJF, we therefore desire that each real space Gaussian component approximate the effects of its Hilbert space counterpart as closely as possible at any sampled position of the electrons; thus we want

$$\langle\boldsymbol{r}|\Gamma_{W,\,\mathrm{RS}}|\Psi\rangle \simeq \langle\boldsymbol{r}|\hat{\Gamma}_{W,\,\mathrm{HS}}|\Psi\rangle, \tag{2.5}$$

where $|\Psi\rangle$ is the Fermionic wave function that is to undergo stenciling. When basis orbitals in W are spatially separated from others in the system — an ideal that is often approached in the localized physics of strong correlation — it is sufficient to choose Q as a step function:

$$Q(\boldsymbol{r}_i) \to \begin{cases} 1 & \boldsymbol{r}_i \in R_W \\ 0 & \boldsymbol{r}_i \notin R_W \end{cases} \tag{2.6}$$

in which R_W is a region exclusively supporting the orbitals in W. To preserve smooth wave function derivatives and allow for a gradual approach to step like behavior in cases where orbital subsets are partially overlapping in space, we relax the step discontinuity at the boundaries of R_W by employing an analytical approximation to the Heaviside function (see Section 2.5).

So long as the smoothed form of Q rapidly approaches 0 as one moves away from the boundary of R_W, the Jastrow factor of Equation (2.4) retains the ability to precisely control the electron count on a subsystem that is spatially well-separated from other subsystems, as there is in this case ample room in between for the 1-to-0 switch to occur. Thus, as with a HSJF, the real space form presented here can fully eliminate ionic terms between well-separated subsystems, allowing it to restore exact size consistency to geminal powers and to aid in the repair of restricted Slater determinants. The key question now becomes whether we can construct functional forms for Q that permit useful demarcations of spatial regions while also ensuring that the curvature they introduce can be hidden in regions where its contribution to the kinetic energy, through the term

$$-\frac{1}{2}\int dr\ \langle\Psi|\boldsymbol{r}\rangle\langle\boldsymbol{r}|\big(\nabla^2\Gamma_{W,\,\mathrm{RS}}+2\nabla\Gamma_{W,\,\mathrm{RS}}\cdot\nabla\big)|\Psi\rangle, \tag{2.7}$$

is mitigated by small local wave function amplitudes $\langle\Psi|\boldsymbol{r}\rangle$. This is of course trivial when demarcating a region around a well-separated fragment, but becomes less so during dissociation events, where partial stenciling becomes beneficial long before the well-separated limit is reached.

2.4 Existing 4-body Jastrows

Before detailing our proposed form for a stenciling-friendly 4-body Jastrow factor, it is instructive to consider why existing 4-body forms are ill-suited for this task. Begin by considering a previously used form[121] for 4-body Jastrows that closely mirrors that of a HSJF:

$$e^{J_4}=\exp\left(\sum_{ijIJ}\Phi_I(\boldsymbol{r}_i)F_{IJ}\Phi_J(\boldsymbol{r}_j)+\sum_{iI}G_I\Phi_I(\boldsymbol{r}_i)\right). \tag{2.8}$$

By diagonalizing $\boldsymbol{F}$, choosing $\boldsymbol{G}$ appropriately, and ignoring changes to wave function normalization, one may convert this Jastrow into a product of Gaussians,

$$e^{J_4}\rightarrow\prod_J\exp\left(-\xi_J\left(N_J-\sum_{iI}U_{IJ}\Phi_I(\mathbf{r}_i)\right)^2\right), \tag{2.9}$$

in which $\boldsymbol{U}$ is the unitary matrix that diagonalizes $\boldsymbol{F}$. Written this way, we may immediately identify the linear combination $\sum_I U_{IJ}\Phi_I$ as one possible form for the counting function Q discussed in the previous section.

We may evaluate the suitability of 4-body Jastrows of the type given in equation (2.8) for use in HSJF-style stenciling by asking how easily these linear combinations can approximate a step function over a given region, and how much control they have over their curvature.

By considering the task of controlling the electron count on a single atom well-separated from the remainder of whatever system is being modeled, the above analysis makes plain that the two common forms for the basis functions Φ_I, atom-centered Gaussians[155] and symmetric polynomials,[152] are not effective for wave function stenciling in Hilbert Space. In the same way that one requires many Fourier components to converge to a square wave, small Gaussian expansions or low-order polynomial expansions are unable to faithfully approximate the switching behavior required for our Jastrow basis functions. Indeed, Gaussian functional forms contain significant curvature at and about the atom's center where the wave function is large in magnitude, and thus cannot engage in the curvature hiding necessary to avoid a rise in kinetic energy when ξ_J is large, i.e. in the strong stenciling regime. Although it is true that in the infinite basis set limit, a complete set of functions (such as the Gaussian spherical harmonics) can represent any smooth function, they will converge to the nearly steplike behavior required by Q only very slowly and so will retain appreciable curvature near the center of the counting region unless the Jastrow basis is made extremely large. In practice, therefore, the functional forms within previously-studied 4-body Jastrows were inappropriate for stenciling, and so, during optimization, the variational principle did not explore their ability to eliminate ionic terms, as doing so would have led to large, curvature-induced increases in the kinetic energy. Ultimately, as can be seen in Sorella's carbon dimer results,[225] the price for using a Jastrow basis that cannot easily represent a step function is, in the context of the JAGP, a size consistency error stemming from the inadequate suppression of ionic terms.

2.5 Mathematical formulation

We investigate the efficacy of real space number-counting Jastrow factors (NCJFs) that can be written in the same general structure as existing four-body Jastrows,

$$e^{J_{NC}} = \exp\left(\sum_{ijIJ} C_I(\boldsymbol{r}_i) F_{IJ} C_J(\boldsymbol{r}_j) + \sum_{iI} G_I C_I(\boldsymbol{r}_i)\right). \tag{2.10}$$

As discussed above, the key characteristic of NCJFs will lie in the choice of basis functions C_I, for which we select a form that can, to a certain degree, act as local real space approximations to Hilbert Space number operators. In the limit of disjoint orbital subspaces, bosonic step functions in real space can exactly reproduce the effects of a sum of Hilbert space number operators and can thus serve as a conceptual starting point for our basis functions. Although we will soften this step-function extreme by employing smooth functions, we will retain the spirit of spatially localized curvature so as to facilitate the curvature hiding that NCJFs require in order to effect strong stenciling without unphysically affecting the kinetic energy.

This goal in mind, we propose "counting" basis functions of the form

$$C(\boldsymbol{r}) = S(g(\boldsymbol{r})), \tag{2.11}$$

where the Fermi-Dirac-like function

$$S(x) = \frac{1}{1 + e^{-\beta x}} \tag{2.12}$$

plays the role of an analytic approximation to the Heaviside step function. The value of S asymptotically switches from zero to one as its argument traverses the origin, with β (which is *not* related to physical temperature) determining both the slope at the origin as well as the effective width of the switching region in which S meaningfully differs from zero or one and displays non-negligible curvature. The interior function $g(r)$ is a scalar-valued function of a real-space coordinate whose nodal surface defines the boundary, or switching surface, of the region within which electrons are to be counted. The volume for which g takes on positive values (negative values) is called the interior (exterior) of the counting region, since composition with the switching function S ensures that C asymptotically evaluates to one (zero) inside this region. We will refer to counting functions by the geometry of their switching surface, and in the present study we investigate both planar

$$g_P(\boldsymbol{r}) = (\boldsymbol{r} - \boldsymbol{c}) \cdot \hat{k} \tag{2.13}$$

and ellipsoidal

$$g_E(\boldsymbol{r}) = (\boldsymbol{r} - \boldsymbol{c})^T \boldsymbol{T} (\boldsymbol{r} - \boldsymbol{c}) - 1 \tag{2.14}$$

counting regions. The nodal surface of g_P is a plane centered at $\boldsymbol{c}$ and normal to the unit vector $\hat{k}$, while the nodal surface of g_E is an ellipsoid with center $\boldsymbol{c}$ and axes defined by the eigenvectors and eigenvalues of $\boldsymbol{T}$. Together with S, these counting regions provide us with a set of Jastrow basis functions whose only curvature appears at the edges of their counting regions, making it more amenable to being hidden in regions of low wave function magnitude. This positioning of curvature should be compared to the more traditional forms given in Section 2.4, which display significant curvature at their centers. Note that variational optimization of a NCJF may also involve parameters internal to g: in the case of the ellipsoidal counting function, this corresponds to optimizing the position ($\boldsymbol{c}$) as well as the shape and orientation ($\boldsymbol{T}$) of the switching surface.

In addition to this stenciling-friendly curvature, arithmetic operations between these counting basis functions correspond to set operations between their interior volumes, which gives their sums and products a somewhat intuitive meaning. For example, consider the large-β limit of these counting functions,

$$\lim_{\beta \to \infty} C(\boldsymbol{r}) = \begin{cases} 1 & \boldsymbol{r} \in A \\ 0 & \boldsymbol{r} \notin A \end{cases} \tag{2.15}$$

in which these functions revert to actual step functions. Spatial regions' complements now occur simply as

$$1 - C_A(\boldsymbol{r}) = C_{A^c}(\boldsymbol{r}), \tag{2.16}$$

while intersections

$$C_A(\boldsymbol{r}) C_B(\boldsymbol{r}) = C_{A \cap B}(\boldsymbol{r}) \tag{2.17}$$

and unions

$$C_A(\boldsymbol{r}) + C_B(\boldsymbol{r}) - C_{A \cap B}(\boldsymbol{r}) = C_{A \cup B}(\boldsymbol{r}) \tag{2.18}$$

arise from products and sums of counting functions. In this way, the quadratic form in equation (2.10) offers the possibility for the full set of first-order topological operations to arise naturally during the variational minimization of a NCJF, raising interesting questions as to whether adjacent regions will merge or produce cutouts from one another in pursuit of optimal stencils.

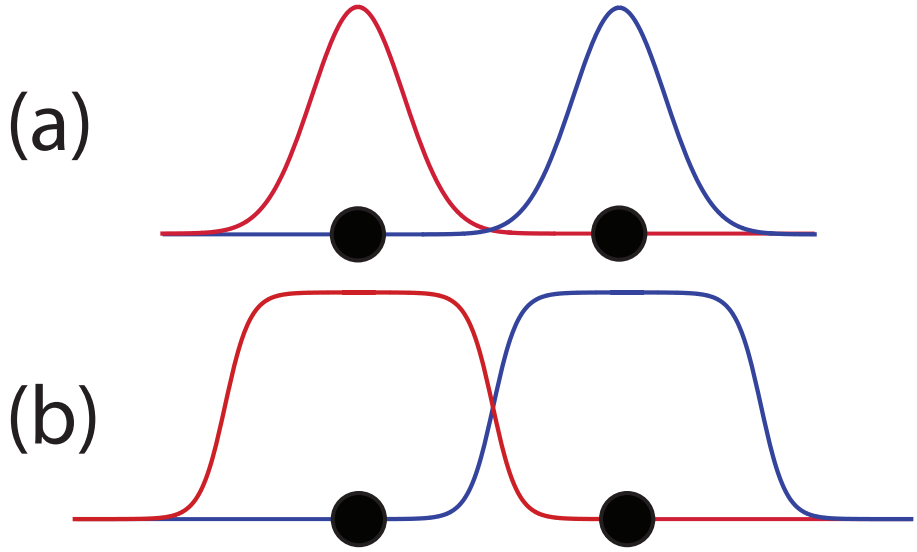

Figure 2.1: One-dimensional schematic for Jastrow basis functions in a diatomic molecule. Filled circles indicate atomic positions and colored lines represent Jastrow basis functions. With a Gaussian basis (a), curvature is concentrated around atomic positions, forcing the variational principle to balance ionic term deletion with the kinetic energy introduced by the Jastrow basis functions at the atomic center. With the counting function basis in (b), curvature is concentrated in regions away from atomic positions, minimizing this kinetic energy penalty.

2.6 Results

We have prepared a pilot implementation supporting planar and ellipsoidal NCJFs within a development version of QMCPACK.[226] NCJFs, as well as spline-based, cusp-correcting e-e and e-n two-body Jastrows and the molecular orbital coefficients were optimized with respect to energy using the variational Monte Carlo (VMC) linear method.[181,182,227,228] The Hamiltonian is taken as the non-relativistic electronic Hamiltonian under the Born-Oppenheimer approximation, with effective core potentials[229] used to replace carbon atoms' core electrons. RHF solutions are taken as the reference configurations at each geometry in the cc-pVTZ basis[52] and are generated by GAMESS.[230] Multireference configuration interaction calculations with the Davidson correction (MRCISD+Q[93,94]) were performed with MOLPRO[231–241] also in the cc-pVTZ basis.

In our results, we will distinguish wave functions based on the types of Jastrow factors employed, whether or not the molecular orbitals were re-optimized in the presence of the Jastrow, and, where applicable, whether the molecular orbitals are symmetric (SA) or have broken symmetry (SB). The presence of counting Jastrows will be denoted by C, traditional spline-based e-e and e-n Jastrows by T, and orbital re-optimization by the prefix "oo-". In all cases, JS stands for Jastrow-Slater. For example, a Jastrow-Slater wave function with both traditional and counting Jastrows whose orbitals were re-optimized starting from a broken symmetry orbital guess would be denoted as oo-CTJS-SB. Finally, DMC results will be denoted by adding DMC to the name of the wave function that fixes the nodal surface.

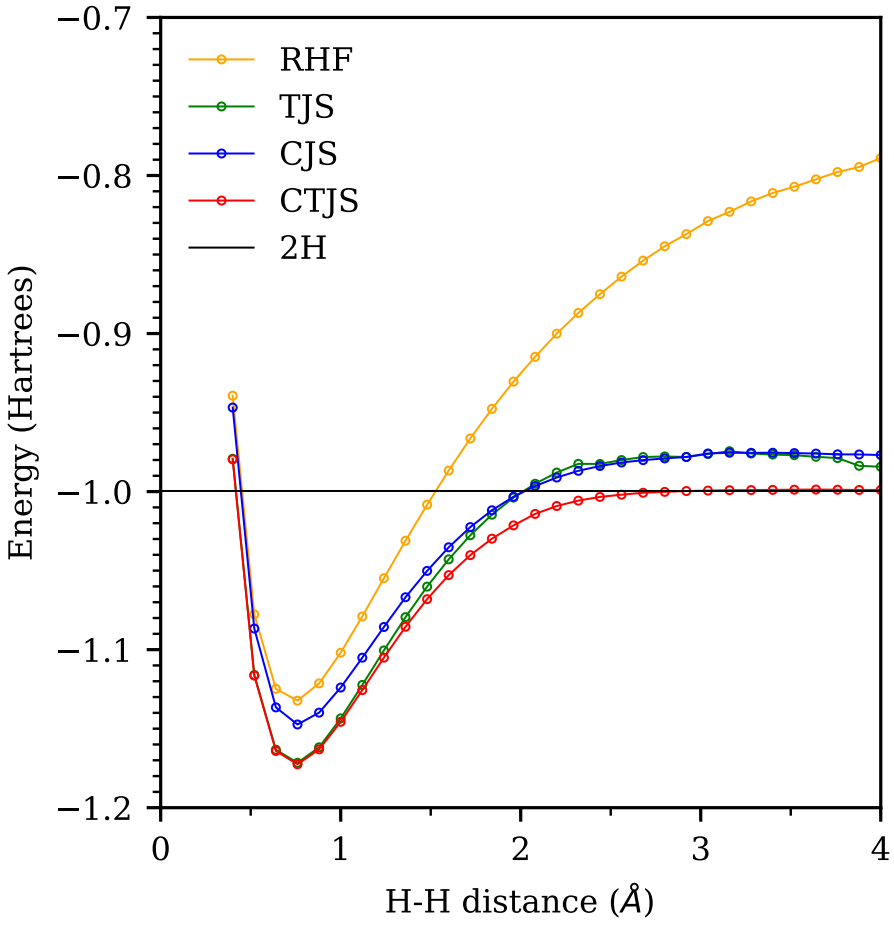

Figure 2.2: VMC energies for the dissociation of H_2 using a cc-pVTZ orbital basis. The solid black line is twice the VMC energy of a single H atom in the same cc-pVTZ basis.

2.7 Hydrogen Molecule

As a minimally correlated wave function, a single restricted Slater determinant is insufficient to describe electron correlation at stretched molecular geometries, which can lead to large size-consistency errors during molecular fragmentation. To correct this, we apply a simple NCJF with a basis consisting of two anti-aligned planar counting functions whose switching surfaces are set to bisect the H-H bond. The NCJF matrix parameters F_{IJ} and G_I are initially set to zero (so that the overall Jastrow factor is initially unity), after which both the matrix parameters (F_{IJ}, G_I) and basis function parameters ($\beta_I, \hat{k}_I, \boldsymbol{c}_I$) are optimized. Figures 2.2 and 2.3 show that NCJFs paired with either cusp-correcting Jastrows (CTJS), orbital optimization (oo-CJS), or both (oo-CTJS) prove far more effective at recovering size-consistency than when only two-body Jastrows (TJS) are used, even if assisted by orbital re-optimization (oo-TJS). At a separation of 4 Å, for example, we find that oo-CTJS is size-consistent to within 0.4 mE_h , while the smallest size-consistency error achievable without NCJFs is over 14 mE_h .

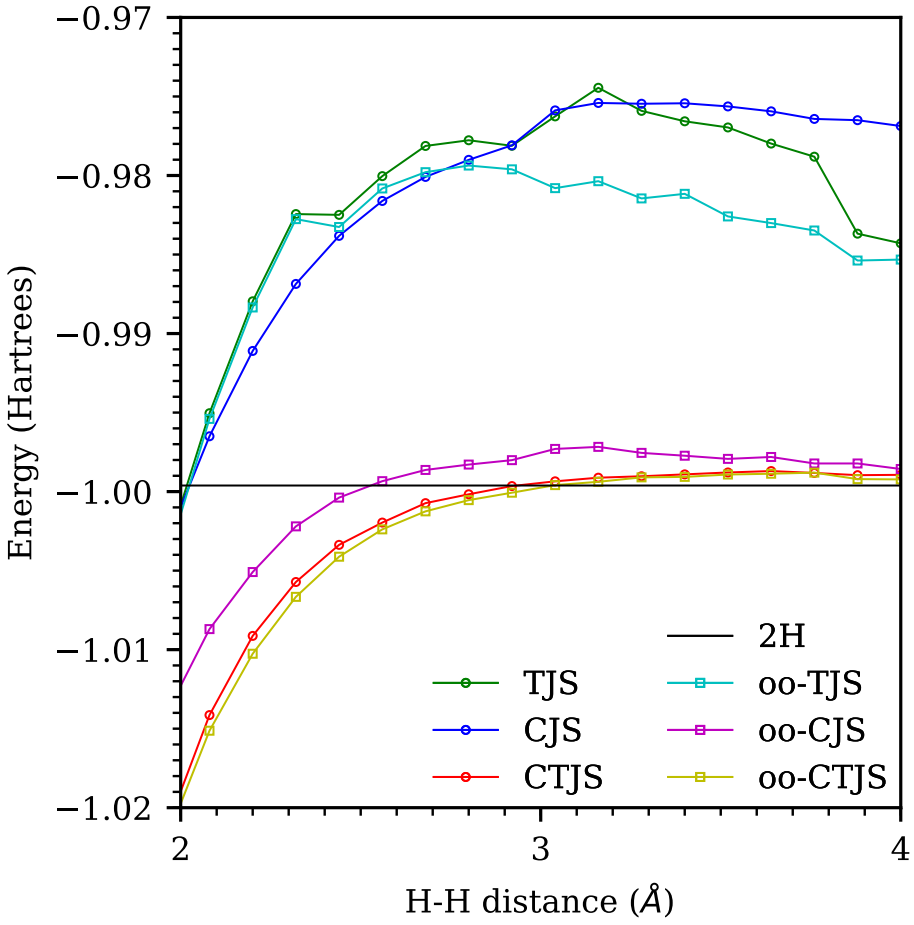

Figure 2.3: VMC energies for the dissociation of H_2 using a cc-pVTZ orbital basis, now focusing on stretched geometries. The solid black line is twice the VMC energy of a single H atom in the same cc-pVTZ basis.

2.8 Ethene

A variationally optimized Slater-Jastrow ansatz is often taken as a guiding function for diffusion Monte Carlo calculations, but the appearance of symmetry-broken minimum-energy solutions to the RHF equations at stretched geometries — which do not possess the correct nodal structure required by DMC — means that we cannot naively take minimum-energy configurations without issue. For instance, when stretching the C=C bond in ethene past 2.5 Å, an RHF solution with broken-symmetry orbitals sees its RHF energy drop below that of the symmetric-orbital solution. However, the nodal surface of this broken-symmetry solution is incorrect, and so when used in DMC it gives an energy that is 40 mE_h or more above that of a DMC based on the symmetric-orbital RHF solution (see Figure 2.3). In more complicated systems, such effects can be more pronounced, and it would be highly desirable to be able to predict beforehand which nodal surface is most appropriate. Given a sufficiently flexible trial function to optimize, VMC can in principle produce the correct nodal surface by selecting the VMC wave function with the lowest energy.

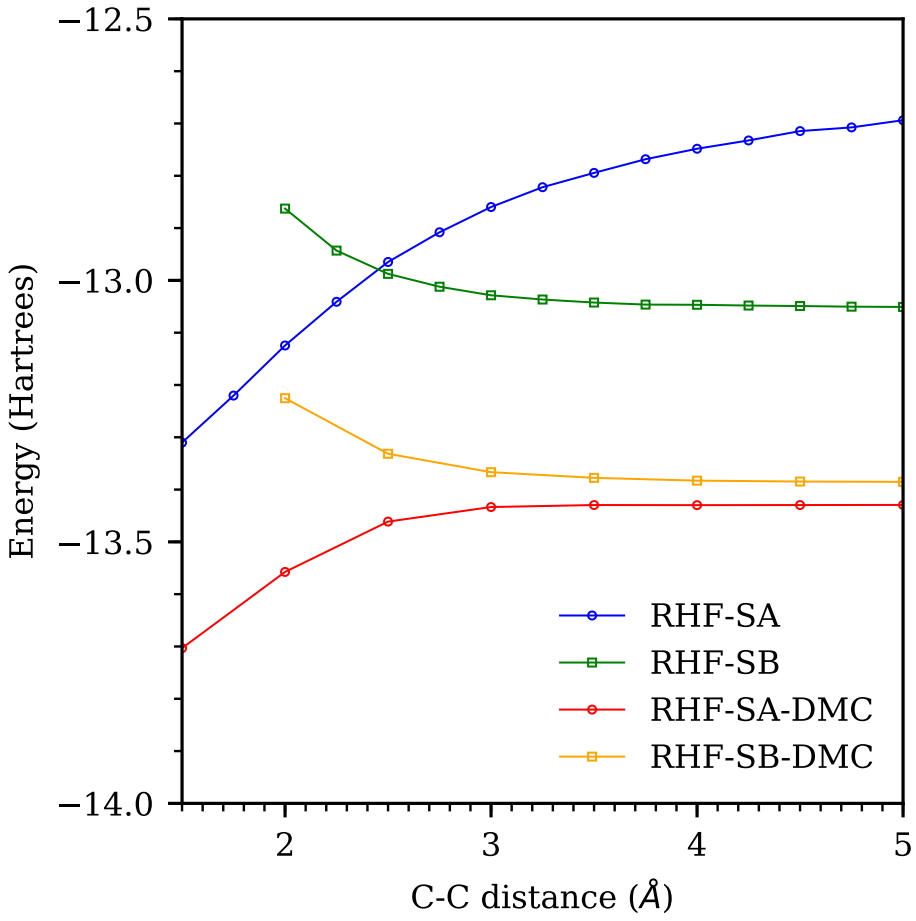

Figure 2.4: VMC and DMC energies of symmetry-adapted (SA) and symmetry-broken (SB) RHF solutions. The RHF minimum at stretched geometries corresponds to the symmetry-broken configuration, which is not the best DMC guiding function.

However, this approach will only be reliable if the trial function is flexible enough, and in the case of ethene, traditional Jastrow-Slater is not, even under orbital re-optimization, as can be seen in Figure 2.5. Although multiconfigurational expansions can be used in lieu of a single reference Fermionic function in order to achieve the flexibility needed to describe the strong correlation responsible for flipping the energy ordering of these two states, the complexity and thus cost of such an expansion must grow exponentially with the number of correlated bonds. The cost of a stenciling approach using NCJFs — assuming a constant number of counting basis functions per fragment — will by comparison scale only quadratically with fragment number, and so it would be quite useful if stenciling were able to capture a sufficient amount of strong correlation to produce the correct energy ordering of states at the VMC level.

Using the same planar NCJFs as in the hydrogen case (except now the planes bisect the C=C bond), we apply NCJFs, orbital optimization, and traditional two-body Jastrows to a single Slater determinant that is either a symmetry-adapted (SA) or symmetry-broken (SB) RHF solution. (Note that SA vs SB orbitals did not interconvert under orbital optimization

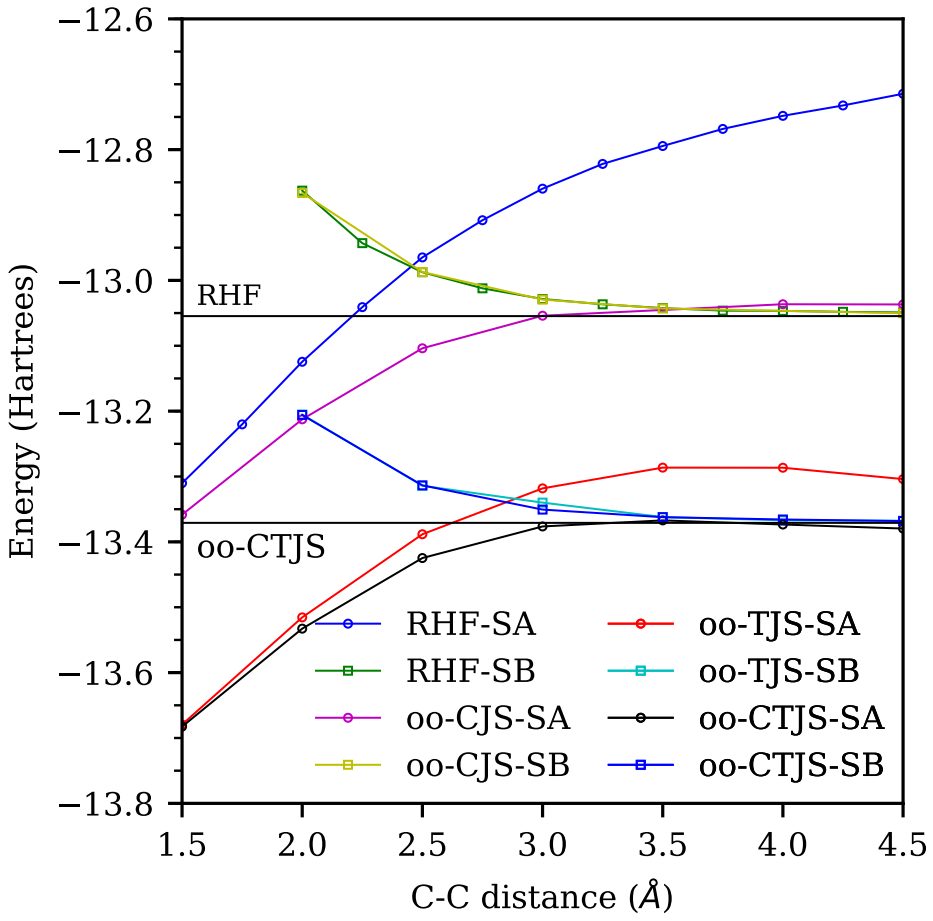

Figure 2.5: VMC energies with orbitals optimized together with Jastrow variables. The solid black lines are twice the VMC energy of a single CH_2 fragment, with no Jastrow factor (RHF) or with a cusp-correcting Jastrow (TJS), as indicated, and provide a reference for size-consistency.

and appear to represent two separate minima on the optimization surface.) In the most flexible case, oo-CTJS, VMC is now correctly able to predict that the SA energy lies below that of the SB energy, a prediction that fails to materialize if the NCJF is omitted (see Figure 2.5). Upon using the oo-CTJS-SA and oo-CTJS-SB wave functions to fix the DMC nodal surface, we find that the lower-energy VMC state now corresponds to the lower energy DMC result, and that the lower energy DMC result is in close agreement with MRCISD+Q (see figure 2.6). Thus, in the case of the ethene double bond dissociation at least, the NCJFs' ability to suppress spurious ionic terms within a single Slater determinant is sufficient to produce a qualitatively correct state ordering and nodal surface without resorting to multi-determinantal expansions.

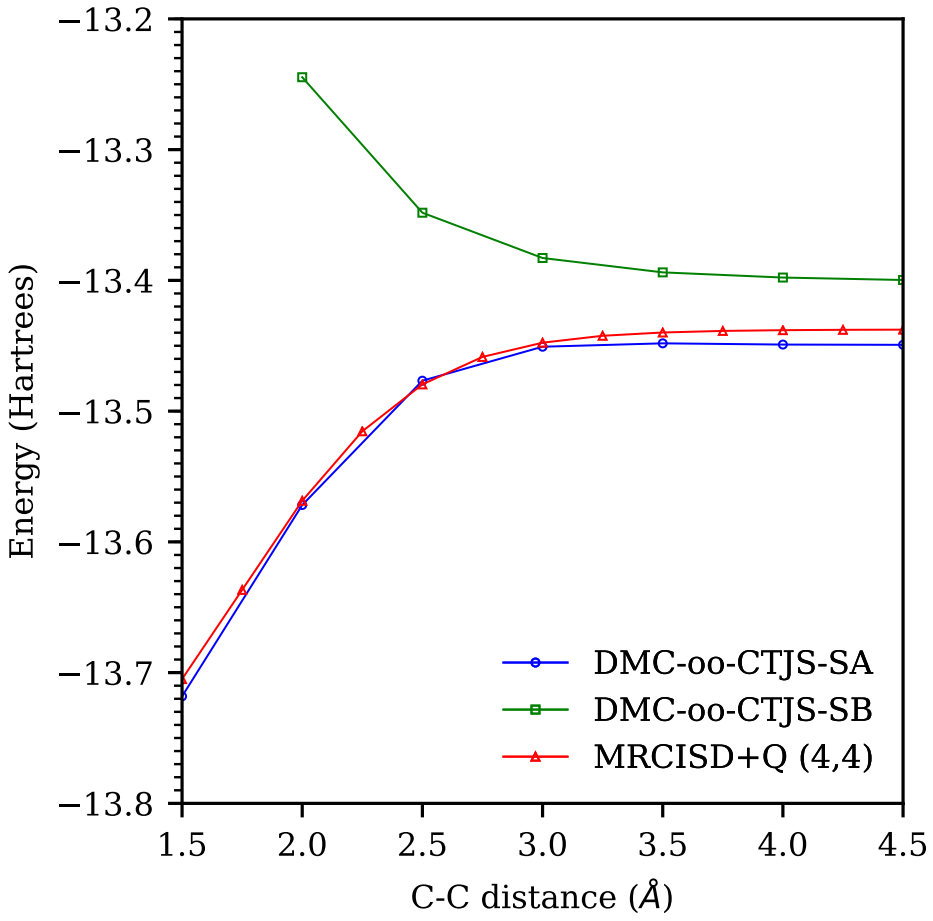

Figure 2.6: DMC energies based on the oo-CTJS-SA and oo-CTJS-SB nodal surfaces, along with results for (4e,4o)-MRCISD+Q.

2.9 Curvature Hiding

Using ellipsoidal basis functions, we will demonstrate the size-consistency problem encountered with basis functions used previously in four-body Jastrow factors. As noted earlier, the main problem associated with taking atom-centered Gaussian functions as the Jastrow basis lies in their inability to effectively hide their curvature in regions where wave function values are small. We reproduce this effect in ethene at a stretched geometry (4.5 Å) by changing our Jastrow basis from C=C bond-bisecting planes to a set of ellipsoidal counting functions in which we scan over an axis scaling parameter L while keeping one edge orthogonally bisecting the C=C bond axis. NCJF parameters ($\boldsymbol{F}$, $\boldsymbol{G}$, β) are fixed at values optimal for planar counting functions, which allows us to reproduce the behavior of the planar basis as we increase the axis scale (at large values of L). This gives us a suitable baseline to compare the ellipsoidal and planar Jastrow basis at different geometries, and, as we expect, VMC energies match reasonably well when each ellipsoidal counting function encompasses an entire CH_2 fragment.

However, as we shrink the ellipsoidal switching surface to only partially encompass each

fragment, Figure 2.6 shows that the overall energy increases, which can be explained by the fact that the curvature at the edge of the ellipsoidal counting region is now cutting through a region with appreciable wave function magnitude. For Jastrow basis functions like these too-small ellipses, these energetic penalties prevent the variational principle from increasing the value of elements of the Jastrow matrix $\boldsymbol{F}$, precluding any stenciling behavior. When instead the basis functions allow for curvature hiding, such as when the ellipsoidal regions are large enough to fully encompass the CH_2 fragment, the variational principle is again free to restore size-consistency by increasing the magnitude of $\boldsymbol{F}$ elements, deleting spurious ionic terms without these energy penalties.

This example demonstrates the necessity of hiding curvature present in the Jastrow basis within regions of low wave function value. The conventional choice of atom-centered Gaussians as the Jastrow basis place this curvature directly on top of atomic positions, making ionic term deletion variationally unfavorable, as discussed in section 2.2.

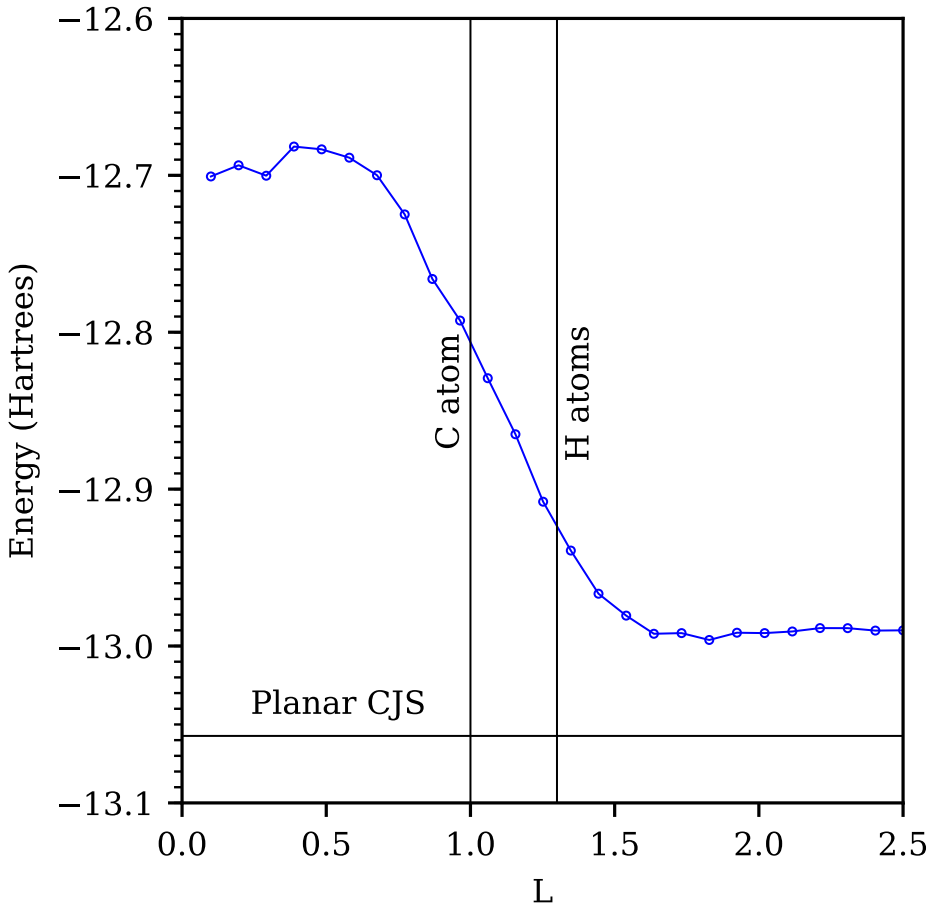

Figure 2.7: CJS energies using ellipsoidal counting functions of variable size, placed according to Figure 2.8. Vertical lines indicate values of L at which the ellipsoidal counting functions' switching surfaces cut directly across an atomic nucleus. The horizontal black line indicates the VMC energy of planar counting regions, which should serve as a lower bound for the energy in the $L \to \infty$ limit.

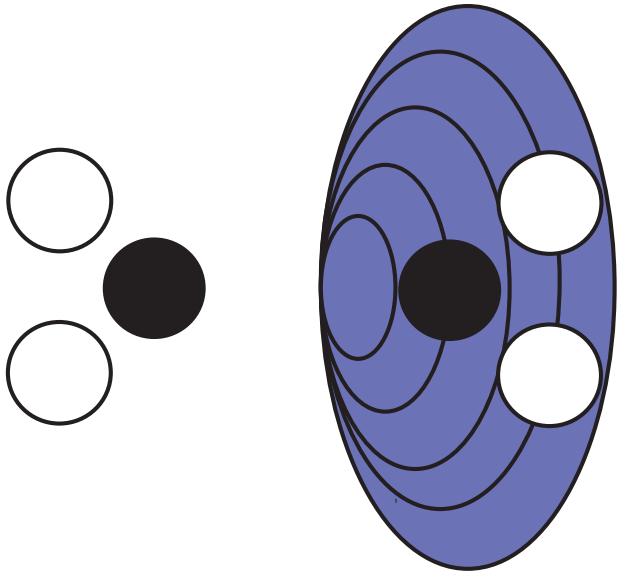

Figure 2.8: Two-dimensional schematic for counting function setup in Figure 6. Black circles represent carbon atom positions, white circles represent hydrogen atom positions, and blue ellipses represent ellipsoidal counting functions. Multiple ellipsoidal counting functions are shown superimposed with larger ellipses corresponding to larger values of the scaling parameter L. The ellipse intersecting the carbon nucleus (the second smallest ellipse) corresponds to $L = 1$ and the ellipse intersecting the hydrogen nuclei (the second largest ellipse) corresponds to $L \approx 1.3$.

2.10 Basis Construction Schemes

While we have shown that the combination of suitably chosen Jastrow basis functions with a general 4-body Jastrow factor form can successfully include strong correlation effects through the stenciling of ionic terms, a generally applicable method requires some rule or prescription for how the Jastrow counting functions are to be chosen for an arbitrary molecule. Let us discuss and discard two options based on the current planes and ellipses before motivating future work with some observations on the properties that a general NCJF basis should satisfy.

First, one might choose to place planar counting functions so as to bisect each bond in a molecule. Such an approach would prepare the ansatz for suppressing unwanted ionic terms in any given bond, but the infinite extent of the planes would clearly not in general satisfy the requirement that the counting function's curvature be hidden in regions of small wave function magnitude. What if the plane from one bond intersects a far-away atom?

Second, one could consider using atom-centered ellipsoids for the counting regions, hoping to take advantage of set operations to generate unions of ellipsoidal counting regions where necessary to encompass an overall fragment. While this scheme sounds more promising, the data presented in Figure 2.9 show that in practice, such set operations do not work out cleanly. The trouble in this case is due to the fact that the optimal switching functions are much smoother than sharp step functions (which would have dire kinetic energy conse-

quences), and so clean set operations to create a union of neighboring counting regions are not achievable within the chosen 4-body form of the overall NCJF. While the symmetry of ethene still allows for the elimination of ionic terms via an $\boldsymbol{F}$ that suppresses terms in which the left-hand and right-hand fragments' not-quite-correctly-unioned counting regions give differing electron counts, the imperfections in the union create residual Jastrow curvature in between the C and H atoms where the wave function magnitude is not small. This residual curvature increases the kinetic energy of the neutral terms that survive the stenciling process, and it appears from the results in Figure 2.9 that this effect is large enough that the variational principle instead chooses to eschew strong suppression of ionic terms.

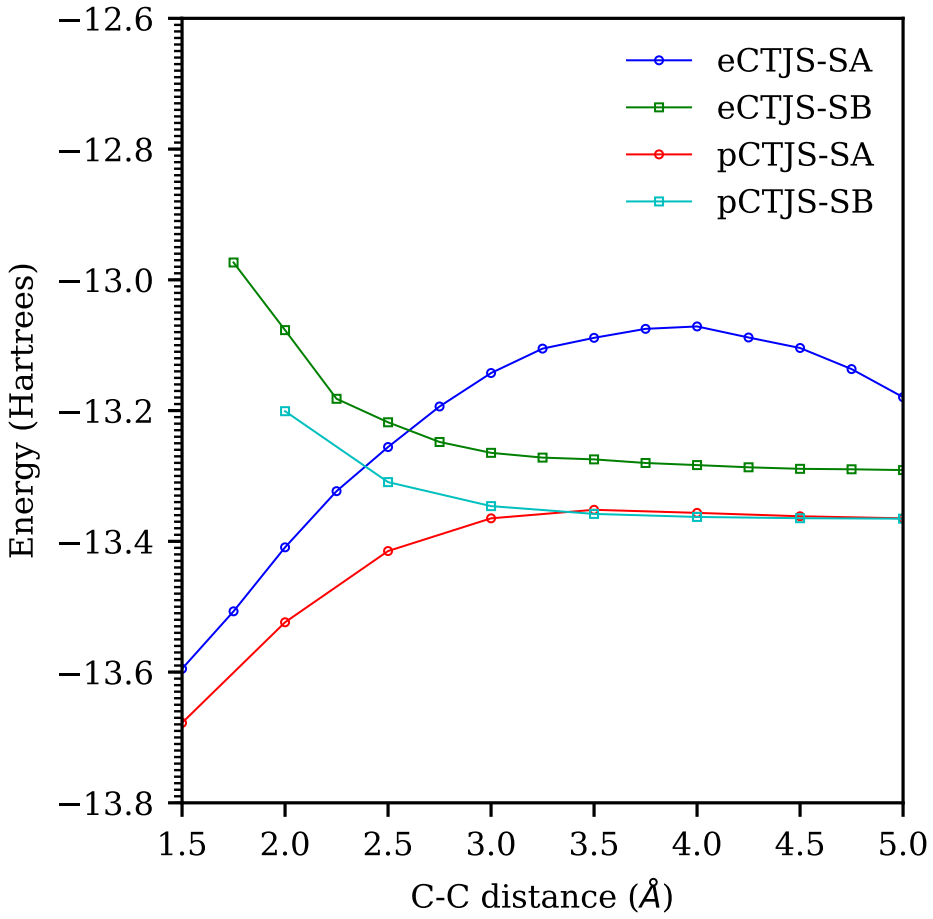

Figure 2.9: VMC energies when using atom-centered ellipsoids (eCTJS) vs C=C bond bisecting planes (pCTJS) for the counting regions, applied to both symmetric and symmetry-broken RHF determinants.

The difficulties in the above two schemes highlight the properties that should be sought in future for general-purpose NCJF basis functions. First, the function should be finite in spatial extent, so that when used for stenciling in one region they do not unduly affect the kinetic energy in distant parts of the molecule. Second, the functions must be capable of clean set operations so that they can combine when necessary to form a counting region

around a group of atoms. Finally, they must remain efficiently evaluable for a randomly chosen configuration of the electrons so as not to disrupt the algorithmic requirements of VMC.

We conclude by outlining a simple scheme that fits our above criteria. First, along each internuclear axis, we place a pair of axis-bisecting, anti-aligned planar counting functions. Second, for each atom, we multiply together all planar counting function which contain that atom, and take each product formed in this way as our counting function basis. These atom-centered, Voronoi-like counting functions cleanly partition all of real space into disjoint, local cells, and aside from minor deviations around their edges, behave like their constituent planar counting functions under set operations. Finally, a naive implementation of this scheme has cost which scales only quadratically with the number of atoms in the system, and does not meaningfully change cost scaling of Monte Carlo sampling when paired with a standard Fermionic ansatz.

2.11 Conclusions

We have demonstrated that 4-body real space Jastrow factors are, with a suitable choice of Jastrow basis functions, capable of performing strong wave function stenciling, in which a multiplicative Jastrow factor makes a large change to the wave function by deleting unphysical configurations from a simple but overabundant Fermionic reference. In particular, these Jastrow factors are capable of eliminating ionic terms between well-separated molecular fragments, which restores exact size consistency to the geminal power and greatly improves the situation for restricted Slater determinants, bringing real space Jastrows in line with the size-consistency-restoring properties already enjoyed by Hilbert space Jastrows. Unlike their Hilbert space brethren, the real space Jastrows presented here are compatible with diffusion Monte Carlo, which creates exciting possibilities for generating qualitatively correct nodal surfaces in strongly correlated regimes with a variational Monte Carlo approach that is both polynomial cost and size consistent. Indeed, our preliminary results show that, when equipped with these stenciling-capable Jastrow factors, the variational minimization of a single reference Jastrow-Slater trial function produces a qualitatively correct nodal surface during the double bond dissociation of ethene, which in turn leads diffusion Monte Carlo to produce an accurate potential energy curve. As every step in this process has a polynomially scaling cost, it will be very exciting in future to test the efficacy of this combination in larger and more strongly correlated settings.

The key development allowing for effective stenciling was the introduction of a new form of 4-body Jastrow basis function, in which a smoothed indicator function is used to check whether or not each electron is within a given region of space. These basis functions thereby allow the overall Jastrow factor to count and control how many electrons are in a given region, which in turn allows for the suppression of unwanted ionic configurations. Unlike previously explored basis function forms, these counting functions have no curvature except at the boundary of their spatial region, allowing them to participate in strong stenciling so long as the boundaries are arranged so as to hide their curvature in regions of small wave function magnitude. In contrast, Gaussian-type basis functions have significant curvature at their centers, leading to kinetic energy changes that prevent effective stenciling. The most pressing priority in the future development of these number counting Jastrow factors is

to formulate them in a way that permits for black-box treatments of arbitrary molecules in which the variational principle can decide automatically how to demarcate important regions in which to count and control electron number. Although the planar and ellipsoidal forms used in this study do not appear to support this black-box ideal, research into promising alternatives is underway.

3 Automated Spatial Tessellations for Correlation Factors

3.1 Abstract

We introduce a basis of counting functions that, by cleanly tessellating three dimensional space, allows real space number counting Jastrow factors to be straightforwardly applied to general molecular situations. By exerting direct control over electron populations in local regions of space and encoding pairwise correlations between these populations, these Jastrow factors allow even very simple reference wave functions to adopt nodal surfaces well suited to many strongly correlated settings. Being trivially compatible with traditional Jastrow factors and diffusion Monte Carlo and having the same cubic per-sample cost scaling as a single determinant trial function, these Jastrow factors thus offer a powerful new route to the simultaneous capture of weak and strong electron correlation effects in a wide variety of molecular and materials settings. In multiple strongly correlated molecular examples, we show that even when paired with the simplest possible single determinant reference, these Jastrow factors allow quantum Monte Carlo to out-perform coupled cluster theory and approach the accuracy of traditional multi-reference methods.

3.2 Introduction

Characterizing novel strongly-correlated systems requires an accurate description of the electronic wavefunction. Traditionally, this is achieved by adding correlations to a minimally-correlated mean-field reference.[43,53] It is often useful to divide these correlations into two distinct classes – strong (static) and weak (dynamic)[242] – based on their relative magnitudes and impact on the accuracy of the reference wavefunction. This dichotomy is not well-defined, but it is useful in limiting cases, as each class describes effects that arise in different physical limits. As different types of these correlation effects take on distinct mathematical forms, they are often treated using distinct theoretical methods. A quantitative method must consider both to a high accuracy, a task that sharply grows in difficulty with the complexity and size of the system.

These corrections introduce statistical correlations between individual electrons within a many-body reference wavefunction. Static or strong correlations are those that require large, qualitative corrections to a mean-field state to effectively describe, and manifest in chemical systems as stretched or broken bonds. By contrast, weak or dynamic correlations are those that give rise to more subtle effects like particle coalescence cusps[114,130,151] and van der Waals interactions,[243] and are often described by perturbative or explicitly correlated[53,136] methods. Without additional approximations, combining methods individually well-suited for different classes of correlations can lead to prohibitive computational costs that are much greater than the cost of the components. As a result, theoretical methods in quantum chemistry still struggle to capture important contributions from both types of correlations simultaneously in complex systems while remaining computationally tractable.[244–247]

In Fock space, static correlation is often treated using active-space methods,[44,105,248] and involves optimizing the energy of a wavefunction restricted to a subspace of configurations generated from orbitals within a limited energy window. In this subspace, the number

of configurations scales combinatorially with the number of electrons and orbitals, and in cases where highly degenerate patterns of orbitals emerge, even the size of this active space becomes impractical. Although great progress has been made in addressing this challenge by DMRG,[125–128,246,249] selective CI,[95,124,125,250–253] and FCIQMC,[115–117] it remains difficult to go beyond 40 active orbitals due to the exponential asymptotic scaling that these methods all possess.

Dynamic correlation is typically captured using perturbation theory.[43] These corrections are vital for accuracy, as they capture details such as wavefunction cusps[130] and van der Waals correlations.[243] However, the cost of including these corrections scales both with the system size and the size of the already-complex active-space wavefunction, and so in practice, adding dynamic correlation effects on top of aggressive static correlation treatments remains quite challenging. For example, in approaches like CASPT2,[108] the cost of perturbatively adding dynamic correlation to a CAS reference scales as the sixth power of the size of the active space. As a result, it remains interesting to look for approaches that incorporate static and dynamic correlation at low-order polynomial cost, even if they may be more approximate than other post-CAS methods.

By contrast, real-space wavefunctions offer a number of powerful advantages when attempting to capture both types of correlation simultaneously. For instance, the Slater-Jastrow wavefunction[114] is natively attuned to wavefunction cusps and can exactly express them using a small set of variational parameters. The Jastrow factor, a symmetric many-body multiplicative factor, is responsible for this compact description, as electron-electron cusps are naturally described using the relative particle coordinates available to its explicit position-space representation. This is in stark contrast to Fock space, where cusps may not have an exact representation and where an equivalent factor would require a lengthy enumeration of corrections.[53,136] This choice between real-space and Fock space wavefunctions matters, as concise descriptions of particle correlations may appear only within a particular representation. Those acquainted with real-space Monte Carlo[114] techniques might argue that projector-based methods such as Diffusion Monte Carlo (DMC) obviate the explicit formulation of every detail of correlation, a powerful advantage not enjoyed by many advanced Fock space methods. However, these methods rely on the accuracy of the wavefunction's nodal surface, which, though determined exclusively by the wavefunction's antisymmetric components, is nonetheless coupled to the symmetric Jastrow factor.

There have been efforts to expand the scope of Jastrow factors beyond particle cusp by writing three-body and four-body correlation terms using higher-order functions of inter-particle coordinates[154,254–256] or a standard atomic orbital basis.[121,156,157,257–261] Although the hydrogenic functions used in this expansion have proven successful building blocks for these sophisticated Jastrow factors, they have some formal shortcomings that makes the exploration of alternatives worthwhile, especially in the context of strong correlation. For example, enacting local projections on electron number, which can be useful when treating static correlation, requires a very large basis of hydrogenic functions due to the fact that the basis elements have non-zero curvature near their centers. As number projections can help break chemical bonds,[163] a Jastrow basis that cleanly accommodates them hold promise as a way to compactly encode some static correlation in a way that complements existing Jastrow factors.

In contrast to most standard methods, where a wavefunction is constructed through a

hierarchy of additive corrections, projection factors remove high-energy components already present in a reference state. Number projections are particularly suited to stenciling out these high-energy components from compact functional forms which contain an overabundance of spurious ionic configurations, such as the antisymmeterized geminal power (AGP) wavefunction.[147] Application of a Gutzwiller-style[120] number projection factor has been shown to successfully restore size consistency to the AGP,[162] a feat which would otherwise require an exponential number of individual additive corrections.

We have recently shown[262] that Jastrow factors with a basis of sigmoidal "counting functions" – real space functions which seek to mimic Fock space number operators – successfully perform this number projection when applied to single-determinant reference wavefunctions. Not only are these factors trivially compatible with existing Jastrow formulations that help treat dynamic correlation, but their ability to projectively encode static correlations may also help mitigate the size of multi-reference determinantal expansions needed to achieve accurate predictions. However, the real-space counting functions we used previously were constructed in an *ad hoc* way, and were not straightforwardly extensible to general chemical systems.

In this chapter, we present a Number counting Jastrow Factor that is capable of number projection using a basis of automatically-constructed counting functions. These counting functions partition space into disjoint regions and tally the total electron count within each, and pairs of the resulting populations are correlated in an exponential Jastrow factor. The functional form chosen for these counting functions provides them both with the flexibility to adopt complex, quadratic shapes and the simple formal properties that permit the creation of near-arbitrary spatial divisions. We provide two straightforward and automatic generation schemes for these counting functions which can be used to partition space into either spherical sections or polyhedral cells. These Jastrow factors were tested by applying them to single-determinant reference wavefunctions in several strongly correlated systems and exhibit surprising accuracy despite the simplicity of the reference state. In the future, these Jastrows may be combined with more sophisticated Fermionic components such as multi-Slater expansions or geminal power wavefunctions, but in this study we restrict ourselves to pairing them with a single determinant in order to explore just how far they can take us towards polynomial-cost static correlation on their own.

3.3 Review

Number counting Jastrow Factors, or counting Jastrows, are symmetric many-body factors defined in terms of region populations N_I and linear coefficients $\boldsymbol{F}_{IJ}, \boldsymbol{G}_I$:

$$e^{J_C} = \exp\left(\sum_{IJ} \boldsymbol{F}_{IJ} N_I N_J + \sum_I \boldsymbol{G}_I N_I\right) \tag{3.1}$$

Each region population N_I estimates the total electron population in a spatial region $\mathcal{R}_I$ for each particle configuration. Counting Jastrow factors are able to perform particle-number projections between these regions which can be directly seen by transforming the region

Table 3.1: Key for mathematical notation.

Symbol	Description
n_e	Number of electrons
n_C	Number of counting functions
$i, j, \ldots$	Particle index
$I, J, \ldots$	Counting region index
$\alpha, \beta, \ldots$	Atomic index
$a, b, \ldots$	Molecular orbital index
$\boldsymbol{r}_i$	Single particle position, $\in \mathbb{R}^3$
$\{\boldsymbol{r}_i\}$	Particle configuration, $\in \mathbb{R}^{3n_e}$
N_I	Region population
$\boldsymbol{F}_{IJ}, \boldsymbol{G}_I$	Population correlation coefficients
C_I	Counting function
$\mathcal{R}_I$	Counting region, $\subset \mathbb{R}^3$
$\partial\mathcal{R}_I$	Boundary of $\mathcal{R}_I$
$\mathcal{V}_I$	Voronoi region, $\subset \mathbb{R}^3$
$C_{I/J}$	Pairwise counting function
$\mathcal{R}_{I/J}$	Pairwise counting region
$\mathcal{V}_{I/J}$	Pairwise Voronoi region
g_I	Anchor gaussian
$\boldsymbol{n}$	Normal vector
$\boldsymbol{\mu}$	Mean vector
$\rho_I(N)$	Population density of $\mathcal{R}_I$

populations $\{N_I\}$ to the basis that diagonalizes $\boldsymbol{F}$:

$$e^{J_C} = \exp\left(\sum_I D_{II}\left(P_I - \tilde{N}_I\right)^2\right), \quad \boldsymbol{F} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^T \tag{3.2}$$

When D_{II} is negative, particle configurations are suppressed when the transformed region populations $\tilde{N}_I$ deviate from the prescribed population P_I, similar to performing a Hilbert space number projection.[162,163] Region populations are calculated using a set of real-space single-particle counting functions $\{C_I\}$ which are designed to behave like indicator functions over the counting regions $\{\mathcal{R}_I\}$:

$$C_I(\boldsymbol{r}) \approx \begin{cases} 1 & \boldsymbol{r} \in \mathcal{R}_I \\ 0 & \boldsymbol{r} \notin \mathcal{R}_I \end{cases} \tag{3.3}$$

As these functions are contained within the Jastrow factor, we must continuously approximate this discrete switch at the counting region boundary to prevent singularities from appearing in the wavefunction gradient. Each population is determined by summing the counting function values evaluated at each particle coordinate $\boldsymbol{r}_i$:

$$N_I = \sum_i C_I(\boldsymbol{r}_i) \tag{3.4}$$

The primary focus of this section is to detail a simple and flexible functional form for these counting functions with well-characterized counting regions that can be automatically constructed in a fully *ab initio* way.

We had previously[262] used simple counting functions of the form:

$$C(\boldsymbol{r}) = \frac{1}{1 + \exp\left(\alpha f(\boldsymbol{r})\right)} \tag{3.5}$$

which is a sigmoidal function with an inflection point at the zeros of f, at which it attains a maximum slope proportional to α. In the infinite slope limit, we can directly relate this to equation 3.3:

$$\lim_{\alpha\to\infty} \frac{1}{1 + \exp\left(\alpha f(\boldsymbol{r})\right)} = \begin{cases} 1 & f(\boldsymbol{r}) < 0 \\ 1/2 & f(\boldsymbol{r}) = 0 \\ 0 & f(\boldsymbol{r}) > 0 \end{cases} \tag{3.6}$$

and we find that the counting regions are totally determined by the sign-structure of the function f:

$$\mathcal{R} = \{\boldsymbol{r} : f(\boldsymbol{r}) < 0\} \tag{3.7}$$

The counting region boundary – which we will refer to as the switching surface, denoted $\partial\mathcal{R}$ – is the set of zeros of f or, equivalently, when the counting function attains the intermediate value of one-half:

$$\partial\mathcal{R} = \{\boldsymbol{r} : f(\boldsymbol{r}) = 0\} = \{\boldsymbol{r} : C(\boldsymbol{r}) = 1/2\} \tag{3.8}$$

By implicitly defining the switching surface through the zeros of f, we avoid committing it to any particular shape or topology. This approach is general and flexible, and mirrors techniques used in level set methods used to describe and propagate interfaces in fluid dynamics, computational geometry, and material science.[263]

With this analysis in place, we start by looking at switching surfaces that emerge from simple functional forms of f. When f is a linear form:

$$f_l(\boldsymbol{r}) = -\boldsymbol{n} \cdot (\boldsymbol{r} - \boldsymbol{\mu}), \quad C_l = \frac{1}{1 + \exp\left(-\boldsymbol{n} \cdot (\boldsymbol{r} - \boldsymbol{\mu})\right)} \tag{3.9}$$

The switching surface is a plane centered at $\boldsymbol{\mu}$ and normal to $\boldsymbol{n}$. Likewise, when f is a quadratic form:

$$\begin{aligned} f_q(\boldsymbol{r}) &= (\boldsymbol{r} - \boldsymbol{\mu})^T \boldsymbol{A} (\boldsymbol{r} - \boldsymbol{\mu}) - K, \\ C_q &= \frac{1}{1 + \exp\left(K\left((\boldsymbol{r} - \boldsymbol{\mu})^T \boldsymbol{A} (\boldsymbol{r} - \boldsymbol{\mu}) - 1\right)\right)} \end{aligned} \tag{3.10}$$

the switching surface adopts a spherical, ellipsoid, paraboloid, or hyperboloid shape, depending on the sign-structure of the eigenvalues of $\boldsymbol{A}$, centered at the point $\boldsymbol{\mu}$. Labeled examples of each of these counting functions are illustrated in figures 3.1 and 3.2, and we will often refer to counting functions by the shape of their switching surface.

We have shown that a pair of properly placed planar counting functions equip the counting Jastrow with the ability to project out ionic terms in a molecular dissociation process.[262] However, this relies on an *ad hoc* placement of the counting functions across the bond-axis, and immediately runs into problems if naïvely generalized. The gradient and Laplacian of the counting Jastrow exponent J_C attain extreme values near the switching surface, and when the switching surface overlaps with the wavefunction reference orbitals, these terms can easily heighten the kinetic energy of the wavefunction:

$$\begin{aligned} &\langle e^{J_C}\psi | \mathbf{T} | e^{J_C}\psi \rangle \\ &\quad = -\frac{1}{2} \int |e^{J_C}\psi|^2 \Big[\nabla^2 J_C + |\nabla J|^2 + \frac{2\nabla J_C \cdot \nabla \psi}{\psi} \\ &\qquad\qquad + \frac{\nabla^2 \psi}{\psi} \Big] d\,\{\boldsymbol{r}_i\} \end{aligned} \tag{3.11}$$

Kinetic energy introduced by these terms is not necessarily unphysical, as the electronic ground state performs a balancing act between kinetic and potential energy in which the counting Jastrow may freely participate. However, simply bisecting every chemical bond with a pair of planar counting functions in even trivial chemical systems can accidentally place the switching surface on top of a distant atom, as shown in figure 3.3. In many of these cases, it is easy to imagine modifying the counting region boundary to just avoid the distant nuclear center while making only a small change to the counting region itself, ultimately lowering the kinetic energy of the wavefunction with little to no tradeoff in potential energy. As a result, the number projecting action of the counting Jastrow is limited by the uncontrolled overlap between these planar switching surfaces and distant reference orbitals in this hypothetical scheme.

Thinking instead that finite, enclosed boundaries won't encounter this problem, we attempted to generalize this basis by placing a single spherical counting function around each atom, like in figure 3.4. However, as shown in figures 3.5 and 3.6, when stretching ethene

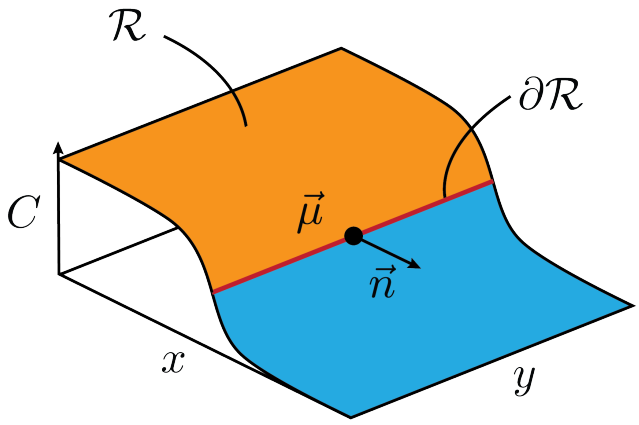

Figure 3.1: A graph of a three-dimensional planar counting function (equation 3.9) projected into the x-y plane. The counting region $\mathcal{R}$ is indicated in orange and the boundary $\partial\mathcal{R}$ as a red line. The boundary is the plane normal to $\boldsymbol{n}$ that intersects the point $\boldsymbol{\mu}$ and is the space where the counting function value equals 1/2.

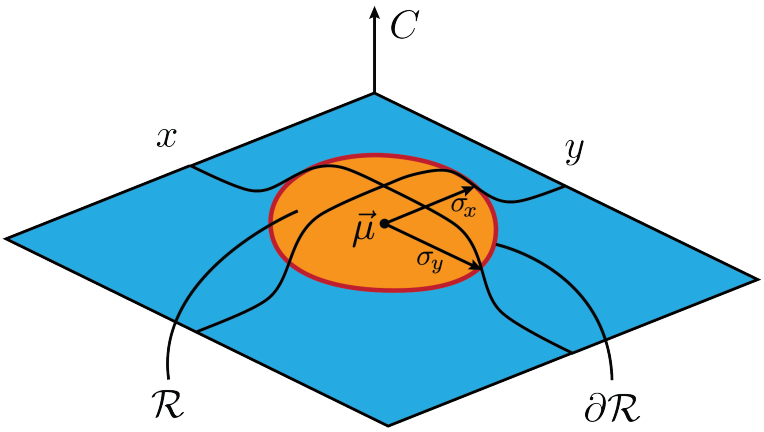

Figure 3.2: A graph of a three-dimensional ellipsoidal counting function projected into the x-y plane. The counting region $\mathcal{R}$ is indicated in orange and the boundary $\partial\mathcal{R}$ is indicated as a red line. The boundary is the ellipse centered at $\boldsymbol{\mu}$ with axis length and directions given by the eigenvalues and eigenvectors (respectively) of $\boldsymbol{A}^{-1/2}$ and is the space where the counting function value equals 1/2.

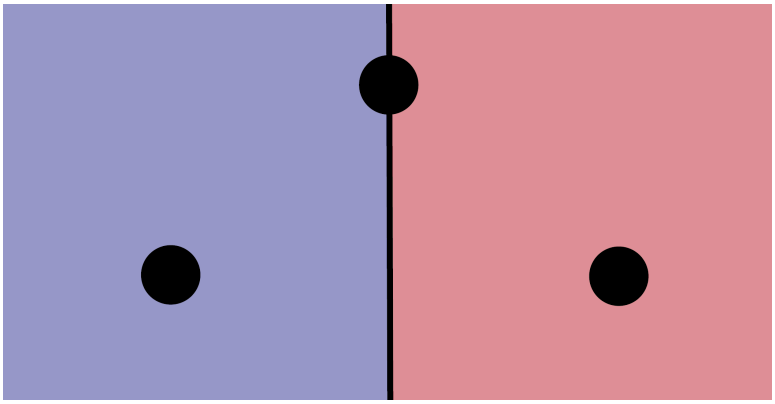

Figure 3.3: In a naïve generalization to a simple two-fragment system, we place a pair of planar counting functions bisecting each bond axis back-to-back. Shown is an isosceles arrangement of atoms (black circles) with a pair of planar counting functions bisecting the base of the triangle. The counting regions are indicated by blue and red regions and the switching surface by the black line. The switching surface intersects the atom at the top vertex, and introduces a kinetic energy cost that limits particle-number projections between electronic populations of the atoms at the base.

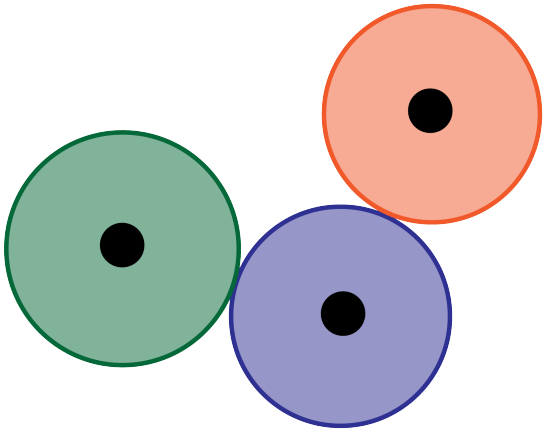

Figure 3.4: An illustration of a set of ellipsoidal regions (indicated by the shaded regions) centered at atomic positions indicated by the black circles. Spherically packing counting functions always produces an incomplete tiling of space, leaving significant curvature at the void boundary. Due to the curved boundary of the counting regions, the interface between neighboring counting regions is very narrow, making it impossible for particles to travel between them without encountering this void curvature and incurring a kinetic energy cost to prospective number projections. As the curvature from neighboring regions does not cancel out in simple linear combinations, attempting number projections between multi-atom fragments using linear combinations of counting functions will encounter this kinetic energy cost as well.

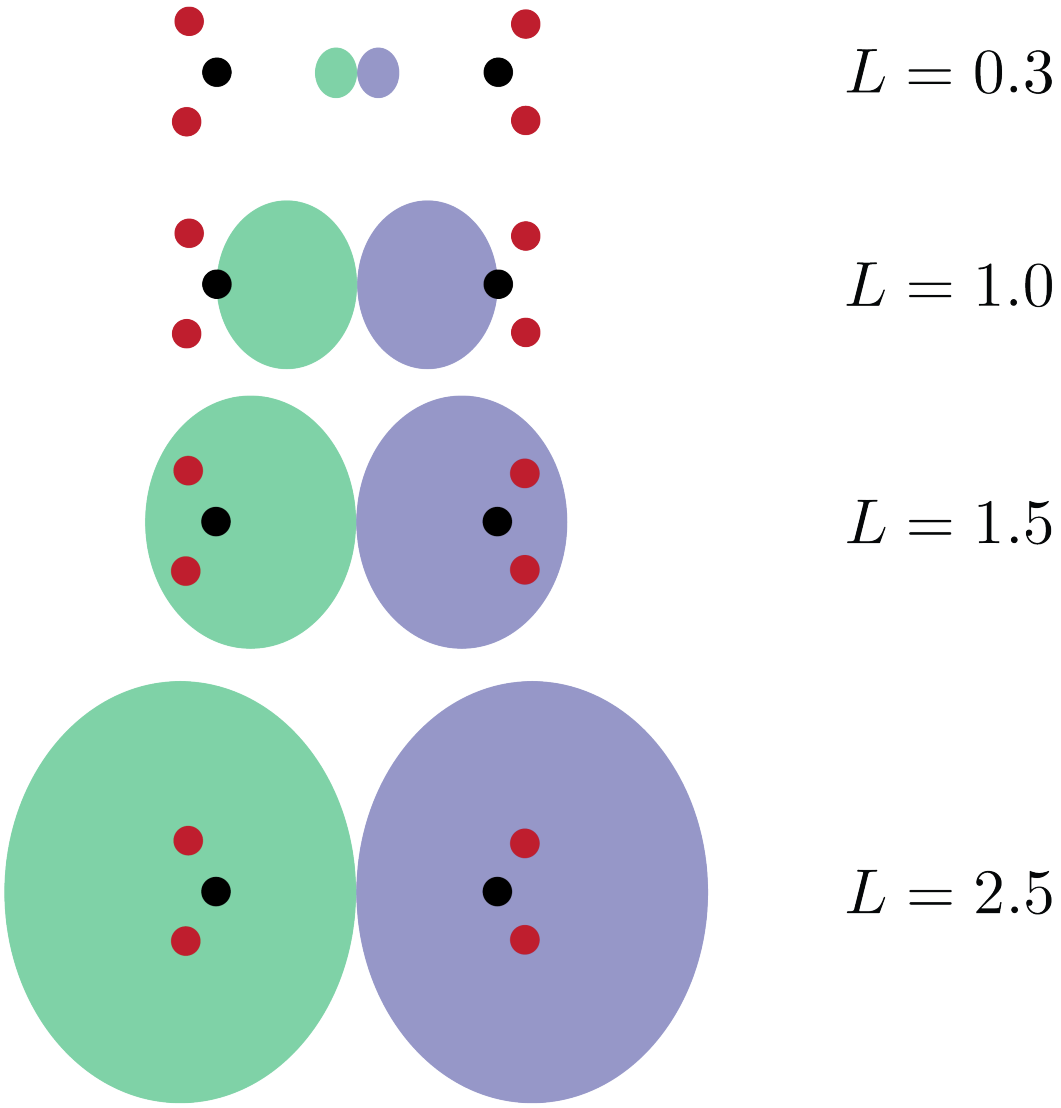

Figure 3.5: Schematic of the counting regions as a function of the scanning coordinate used in figure 3.6. Single-point VMC energy calculations were performed for ethene with a distance of 5 Å between carbon atoms as a function of the counting region scale factor L, indicated in the figure, where the pair of ellipsoidal counting regions were set to touch at the bond midpoint. The number projection was performed using $\boldsymbol{F}$ matrix parameters optimal for a pair of bond-bisecting planar counting functions.

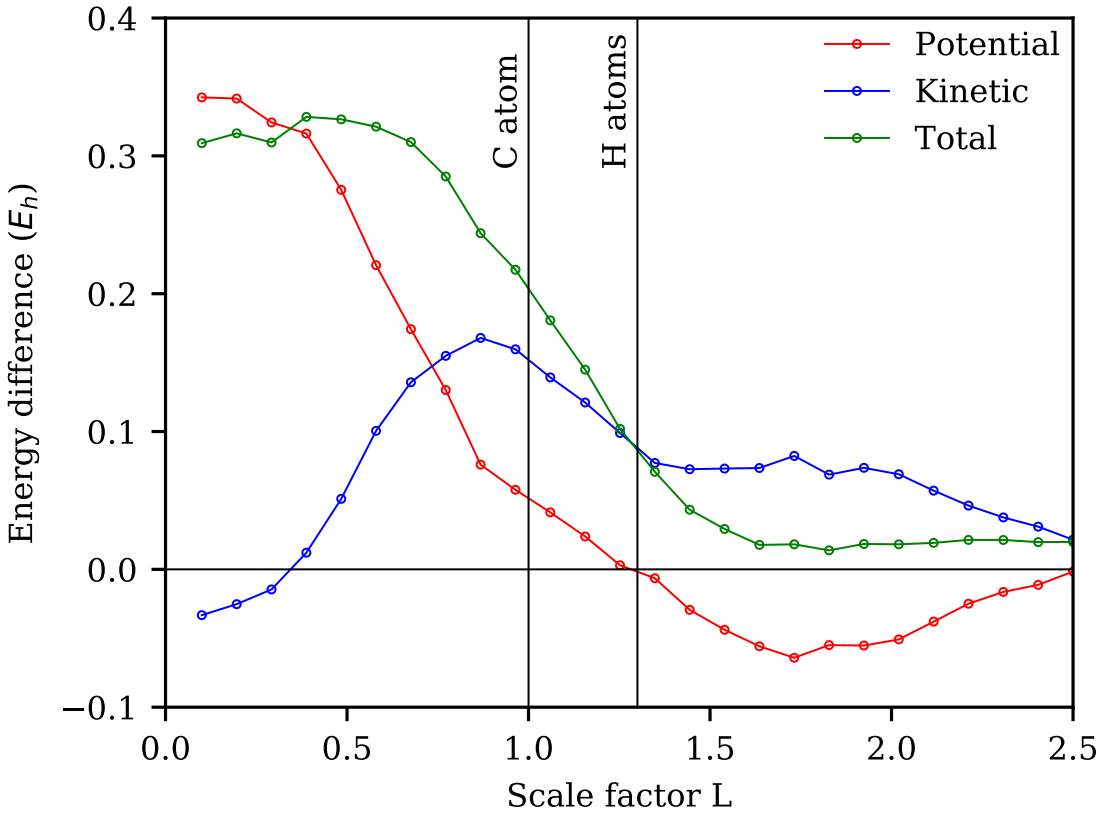

Figure 3.6: Plot of potential, kinetic, and total VMC energies for ethene using a CJS wavefunction with a counting function basis of scaled ellipses, shown in figure 3.5 relative to VMC energies of a CJS wavefunction using a pair of bond-bisecting counting functions. Scale factor values that correspond to an intersection of the ellipsoidal switching surfaces with atomic centers are indicated by the labeled vertical lines. As the ellipsoidal regions completely encompass each fragment, the projecting action of the ellipsoid counting functions approaches that of planar counting functions. At no point do the atom-centered ellipsoidal counting functions improve the total variational energy of the CJS wavefunction relative to the planar CJS benchmark indicated by the horizontal line.

symmetrically across the central double bond, we were unable to find an ellipsoidal counting function basis which improved the variational energy of a CJS wavefunction relative to one that used a pair of planes to bisect the chemical bond. Of course, the fundamental mathematical reason for this failure is the same: the switching surface of the ellipsoidal regions overlap with reference orbitals and limit the scope of variationally favorable number projections. In this case, the overlap originates from the incomplete spatial packing of the spherical switching surfaces which produce void regions that permeate a significant fraction of space, greatly restricting number projections from the counting Jastrow no matter how these ellipsoidal regions are placed.

3.4 Normalized Counting Functions

The collective pathologies of these schemes suggest that an improved counting function basis will have the following attributes:

1. *Localizability*: regions permit finite boundaries.
2. *Completeness*: regions completely tile space.
3. *Clean Additivity*: summation seamlessly combines counting functions.

In this study, we look at counting functions of the form:

$$C_I(\boldsymbol{r}) = \frac{g_I(\boldsymbol{r})}{\sum_J g_J(\boldsymbol{r})} \tag{3.12}$$

with the three-dimensional gaussian functions g_I:

$$\begin{aligned} g_I(\boldsymbol{r}) &= \exp\left(\boldsymbol{r}^T \boldsymbol{A}_I \boldsymbol{r} - 2\boldsymbol{B}_I^T \boldsymbol{r} + K_I\right) \\ &= \exp\left((\boldsymbol{r} - \boldsymbol{\mu}_I)^T \boldsymbol{A}_I (\boldsymbol{r} - \boldsymbol{\mu}_I) + \tilde{K}_I\right) \end{aligned} \tag{3.13}$$

which fulfills all three of these conditions. Note that normalized gaussian functions like this are widely used in statistical classification algorithms such as quadratic discriminant analysis[264] and gaussian mixture models,[265] which should not be too surprising, as the fundamental goal of our counting functions is to classify particles according to their location. We will refer to counting functions of this form as normalized counting functions built from the anchor gaussians g_I.

The relationship between the gaussian parameters $\{\boldsymbol{A}_I, \boldsymbol{B}_I, K_I\}$ and the counting regions they produce is not immediately clear, but upon further investigation, we find that the behavior of a normalized counting function can be modeled by what we will call a 'pair counting function' throughout most points in space. Pair counting functions are those formed from distinct pairs of anchor gaussians and are mathematically identical to the counting functions discussed in equation 3.10:

$$C_{I/J}(\boldsymbol{r}) = \frac{g_I(\boldsymbol{r})}{g_I(\boldsymbol{r}) + g_J(\boldsymbol{r})} = \frac{1}{1 + g_J(\boldsymbol{r})/g_I(\boldsymbol{r})} \tag{3.14}$$

where the first element of the compound index I/J indicates the index of the gaussian in the numerator. These pair counting functions approximate normalized counting functions for most of space in the sense that, for $I \neq J$:

1. $C_I(\boldsymbol{r}) \approx C_{I/J}(\boldsymbol{r})$ when $g_I(\boldsymbol{r})$ and $g_J(\boldsymbol{r})$ are the largest two anchor gaussians.
2. $C_I(\boldsymbol{r}) < C_{I/J}(\boldsymbol{r})$ everywhere.

The space where the first property does not hold for any pair of anchor gaussian is limited to neighborhoods about space where three anchor gaussians are equal:

$$\{\boldsymbol{r} : g_I(\boldsymbol{r}) = g_J(\boldsymbol{r}) = g_K(\boldsymbol{r})\} \tag{3.15}$$

As this set is defined by two independent constraints, it defines a one-dimensional path in space. More precisely, since normalized counting functions are continuous, this condition holds, to within a uniform convergence factor, except at a small neighborhood about these paths. The size of these neighborhoods shrink as we increase the gradient of the counting function at the switching surface – which is easily done by uniformly scaling the variance of all anchor gaussians – and vanishes as the gradient diverges. These pair counting functions are still good approximations to the boundary even at modest gradient values, and we can treat the switching surface of these normalized counting functions as a patchwork of quadratic surfaces formed by the pair counting functions, joined together at edges given by the paths in equation 3.15.

We define the counting region $\mathcal{R}_I$ as the region in which the normalized counting function is greater than one-half:

$$\mathcal{R}_I = \{\boldsymbol{r} : C_I(\boldsymbol{r}) > 1/2\} \tag{3.16}$$

since the second property holds for all $J \neq I$, it follows that:

$$C_I(\boldsymbol{r}) < \min_{J \neq I} C_{I/J}(\boldsymbol{r}) \tag{3.17}$$

so in order for a point to be contained in $\mathcal{R}_I$, it must also be contained in every pair counting region $\mathcal{R}_{I/J}$. As the region boundaries coincide with the pair region boundaries to within the aforementioned convergence factor, we conclude that $\mathcal{R}_I$ is well approximated as the intersection of the pair counting regions:

$$\mathcal{R}_I \approx \bigcap_{I \neq J} \mathcal{R}_{I/J} \tag{3.18}$$

to within this convergence factor. As a result, the counting regions of these normalized counting functions can be understood in terms of these pair counting functions, and we can use them study their boundaries facet by facet.

When this analysis is accurate, which it is nearly everywhere in space, these normalized counting functions are localizable, complete, and cleanly additive, as defined above. First, these normalized counting functions are localizable simply because they can be bounded into finite domains, either by a single ellipsoidal boundary or by multiple planar boundaries.

Secondly, due to the normalization condition, the tiling is complete, as the entire set of counting functions account completely for each particle at every point in space:

$$\sum_I C_I(\boldsymbol{r}_i) = 1 \tag{3.19}$$

Finally, these are cleanly additive, since the pair counting functions along the shared boundary between adjacent regions sum to one:

$$C_{I/J}(\boldsymbol{r}) = 1 - C_{J/I}(\boldsymbol{r}) \tag{3.20}$$

and the contracted counting function, given by the sum:

$$C_{I+J} = C_I + C_J \tag{3.21}$$

has no seam along the shared switching surface and acts as though its counting region is the union of the components' counting regions (to within some small convergence factor):

$$\mathcal{R}_{I+J} \approx \mathcal{R}_I \bigcup \mathcal{R}_J \tag{3.22}$$

These contractions can be exactly represented in the linear Jastrow coefficients $\boldsymbol{F}$ and $\boldsymbol{G}$ featured in equation 3.3, allowing variational methods to freely adjust them to form clean region combinations – a feat that was not always possible in the previous basis formulation. Thus, this normalized-gaussian functional form is able to perform number projections between composite regions without creating curvature in the interior of the contracted counting functions and unduly raising the kinetic energy of the wavefunction in the process.

Anchor gaussians are so named because their mean positions serve as central anchoring points for the normalized counting functions, and roughly embody their geometric centers, as will become more evident in the partitioning schemes that follow. The normalization condition in equation 3.12 used to generate these counting regions is at the heart of their convenient formal properties that allow them to automatically form regions with highly flexible and well-characterized quadratic boundaries. However, this normalization condition induces redundant wavefunction representations in the Jastrow parameter space, which manifest linear dependencies in the wavefunction tangent space and cause ill-conditioned numerics in variational optimization algorithms. In Appendix A, we trace the cause of these issues and describe how to remove this undesired behavior in the context of the Linear Method.

3.5 Classical Voronoi Partitioning

In a classical Voronoi tessellation,[266] space is divided into disjoint regions according to its distance from a set of Voronoi points $\{\boldsymbol{v}_I\}$. Each Voronoi point maps to a subset of space $\mathcal{V}_I$ which is defined as the region that is closer to that Voronoi point than any other:

$$\mathcal{V}_I = \{\boldsymbol{r} : \|\boldsymbol{r} - \boldsymbol{v}_I\| < \|\boldsymbol{r} - \boldsymbol{v}_J\|, \text{ for all } I \neq J\} \tag{3.23}$$

Each of these Voronoi regions can be written as the intersection of Voronoi regions built solely from pairs of Voronoi points:

$$\mathcal{V}_I = \bigcap_{J \neq I} \mathcal{V}_{I/J}, \quad \mathcal{V}_{I/J} = \{\boldsymbol{r} : \|\boldsymbol{r} - \boldsymbol{v}_I\| < \|\boldsymbol{r} - \boldsymbol{v}_J\|\} \tag{3.24}$$

The normalized counting functions described in section 3.4 can be parameterized to divide space in a nearly identical way.

When the quadratic parameters $\boldsymbol{A}_I$ for each anchor gaussian in a set of normalized counting functions are equivalent, the counting regions are equivalent to (within some convergence factor) Voronoi regions generated from the mean gaussian positions. For simplicity, we will look at the case where these matrices are all equal to some scalar multiple of the identity matrix:

$$\boldsymbol{A}_I \equiv \alpha \boldsymbol{I} \tag{3.25}$$

With this restriction on $\boldsymbol{A}_I$, these pair counting functions can be manipulated to adopt the general form of a planar counting function given in equation 3.9:

$$C_{I/J} = \frac{1}{1 + \exp\left[-4\alpha\left(\boldsymbol{r} - \left(\frac{\boldsymbol{\mu}_I + \boldsymbol{\mu}_J}{2}\right)\right) \cdot \left(\frac{\boldsymbol{\mu}_J - \boldsymbol{\mu}_I}{2}\right)\right]} \tag{3.26}$$

with:

$$\boldsymbol{\mu} = \frac{\boldsymbol{\mu}_J + \boldsymbol{\mu}_I}{2}, \quad \boldsymbol{n} = \frac{\boldsymbol{\mu}_J - \boldsymbol{\mu}_I}{2} \tag{3.27}$$

As a result, the pair counting region $\mathcal{R}_{I/J}$ is identical to the Voronoi regions $\mathcal{V}_{I/J}$ defined above, and following section 3.4, the counting regions can be written as the intersection of the pair counting regions (to within some convergence factor):

$$\mathcal{R}_I \approx \bigcap_{J \neq I} \mathcal{R}_{I/J} \tag{3.28}$$

which echoes the Voronoi condition given in equation 3.23.

One immediate application of this scheme is to place anchor gaussians such that their means coincide with each atomic center, which generates a set of atom-centered Voronoi counting regions. The sum of normalized counting function values over electronic coordinates generated in this way roughly corresponds to an atomic population analysis performed on an individual particle configuration. As a result, these partitions are able to identify and remove interatomic ionic terms such as those that appear in dissociation processes, with the added benefit of a systematic and automated generation scheme, unlike the simpler counting function basis used previously.

3.6 Spherical Voronoi Partitioning

Spherical coordinates are a natural choice in chemical systems due to the strong isotropic, attraction between electrons and nuclei. Likewise, counting regions that partition space into spherical sectors organically describe correlations between electrons near the same atomic center. One might imagine approximating this partition using the scheme described in section 3.5 by surrounding an atomic center with a set of anchor gaussians placed at the vertices of a closed polyhedron. An increasingly fine angular mesh of the resulting flat-faced Voronoi regions converge to the desired spherical partition, which could be recombined to form any angular shape using the property of pointwise additivity described in section 3.4. Though technically exact in the infinite limit, rough constructions like this can be unwieldy

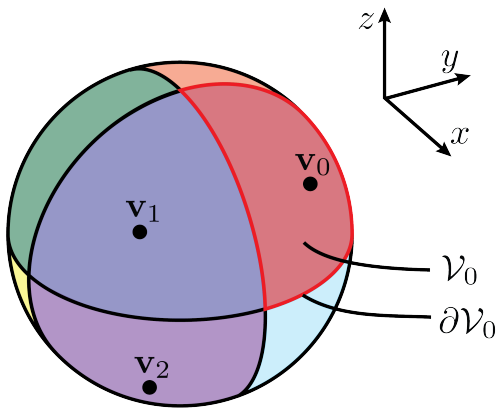

Figure 3.7: A spherical Voronoi diagram, in which the surface of a sphere is partitioned into angular sections based on the surface arc-length from a set of Voronoi points, indicated by labeled black circles. A planar division between these points in three dimensions is sufficient to produce these boundaries, and figure 3.8 shows how these sectors can be further subdivided using spherical boundaries.

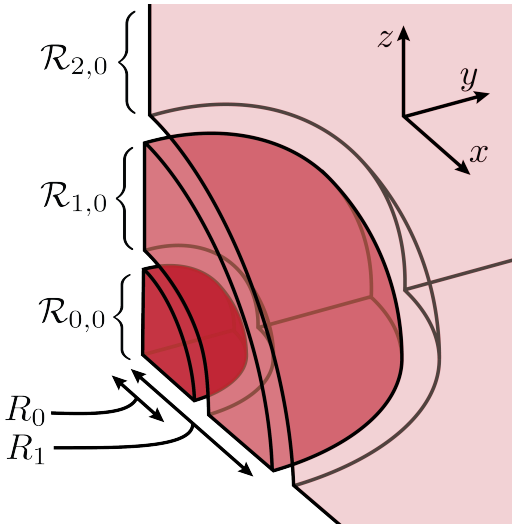

Figure 3.8: The angular sector in figure 3.7 subdivided into three radial shells with radial boundaries at R_0 and R_1. Gaps between regions are shown to highlight the location of radial boundaries. In this partitioning scheme, the other spherical regions in figure 3.7 are radially subdivided in exactly the same way.

and imprecise, increasing either computational cost or introducing an unknown convergence error in cases where exact spherical boundaries would be a better choice.

As discussed in the previous section, classical Voronoi tessellations are a natural way to partition space based on distances from a set of predetermined points. This Voronoi partitioning scheme may be easily adapted to partition a spherical surface in a directly analogous way, as depicted in figure 3.7. Formally, these partitions are defined by restricting our attention to the surface of a sphere and subdividing the surface regions according to the Voronoi condition (equation 3.23) with the distance measured by great-arcs on the spherical surface.[266] Fortunately, we can generate this partition using the classical Voronoi scheme described in the previous section by placing each anchor gaussian on the surface of this sphere, as the planar surfaces intersect the center of the sphere and produce the correct geodesic boundaries. As this generates only an angular partition, we will extend this scheme into three dimensions by further subdividing this partition into radial shells. Luckily, the normalized counting functions described in section 3.4 can easily accommodate the curved radial boundaries through the quadratic parameters $\boldsymbol{A}_I$ present in the anchor gaussians.

An algorithm to generate these spherical Voronoi partitions complete with explicit formulae for anchor gaussian parameters is given in appendix B, which we summarize here. First, we place N_Ω anchor gaussians with identical quadratic parameters $\boldsymbol{A}_I$ whose means $\boldsymbol{\mu}_I$ lie on the surface of a sphere, following the Voronoi scheme in section 3.5. When these anchor gaussians are normalized to create normalized counting functions, the counting regions take on the shape of angular sectors with boundaries that coincide with the spherical Voronoi diagram generated from the means of the anchor gaussians shown in figure 3.7. Second, we choose a set of N_R radii along which we will subdivide each of these angular sectors starting from the innermost shell and successively create spherical divisions between adjacent shells. To add a single radial shell of counting regions, for each anchor gaussian that lay on the current outermost shell, we place another at the same angular coordinates with parameters chosen such that the pair counting functions made from these two anchor gaussians have switching surfaces on a sphere at the prescribed radius, midway between the two anchor gaussians. This generates a total of $N_\Omega \cdot (N_R + 1)$ counting regions where the curvature of the radial partition is not limited by the number of angular divisions. As depicted in figure 3.8, the final angular sectors subdivided into radial shells can be visualized as radial extrusions of the initial spherical Voronoi partition.

3.7 Region Composition

Both of these schemes are useful and complete and each is most appropriate at describing correlations at different scales in the context of molecular electronic structure. The classical Voronoi partitioning scheme described in section 3.5 divides space into a set of atom-centered Voronoi cells, naturally correlating electronic populations between different atoms. The spherical partitioning scheme in section 3.6 instead divides space into a set of spherical sections and is most suited to capture correlations within a single atomic shell. As we look to apply these counting Jastrows in more complex systems, combining these two approaches within a single normalized counting function basis appears a promising way to describe correlations at both of these scales simultaneously. In addition, subdividing existing partitions while retaining the existing divisions allows us to approach the basis set limit

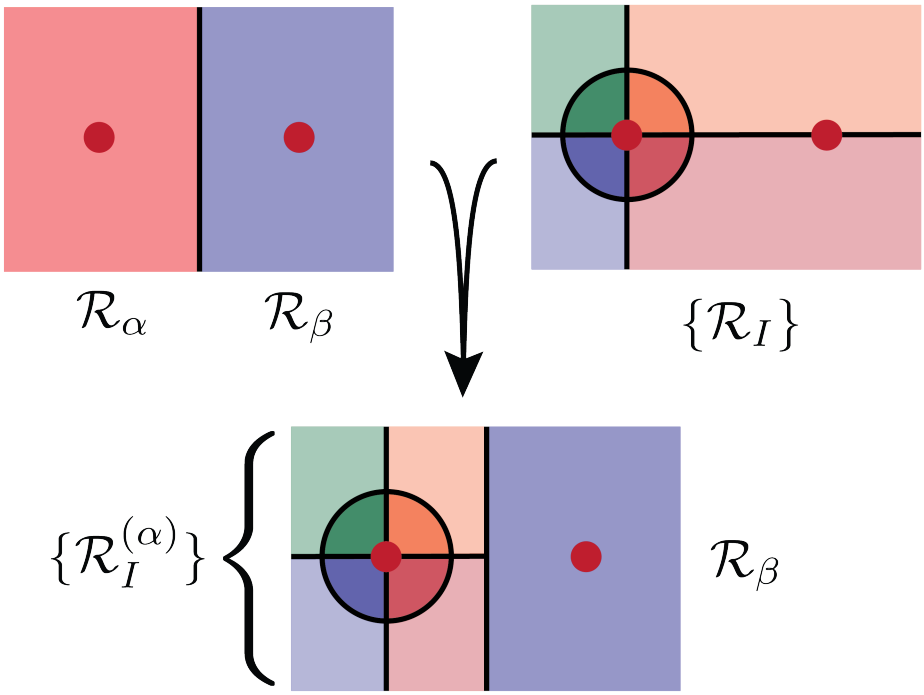

Figure 3.9: Schematic depicting a composition which subdivides a single atom-centered Voronoi region $\mathcal{R}_\alpha$ into a set of regions $\{\mathcal{R}_I^{(\alpha)}\}$ that best match the given partition $\{\mathcal{R}_I\}$. Counting regions are indicated by the shaded areas, atomic positions are indicated by the red circles, and switching surfaces are indicated by black lines. To divide space according to both partitions simultaneously, we must determine the $g_I^{(\alpha)}$ that match the conditions in equations 3.30 and 3.31.

for these counting functions in a systematic and chemically sensible way. In the following discussion, we will consider a rough partition consisting of two regions, $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$, and a fine-grained partition $\{\mathcal{R}_I\}$ which will be used to subdivide $\mathcal{R}_\alpha$, as depicted in figure 3.9. Our goal is to combine the two partitions into a single normalized counting function basis in a way that best preserves their boundaries simultaneously.

The property of clean additivity discussed in section 3.4 states that simple sums of adjacent counting functions act much like a single counting function over their combined region. This suggests a natural condition when splitting single counting function (C_α) into a set of pieces ($C_I^{(\alpha)}$):

$$C_\alpha(\boldsymbol{r}) = \sum_I C_I^{(\alpha)}(\boldsymbol{r}) \tag{3.29}$$

In order to retain the formal properties outlined in section 3.4, counting functions split in this way must be built from a single set of anchor gaussians. Our task is to find a set of

$g_I^{(\alpha)}(\boldsymbol{r})$ that both reproduces the divisions of the fine-grained partition internal to $\mathcal{R}_\alpha$:

$$C_I^{(\alpha)}(\boldsymbol{r}) = \frac{g_I^{(\alpha)}(\boldsymbol{r})}{\sum_I g_I^{(\alpha)}(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} = \frac{g_I(\boldsymbol{r})}{\sum_J g_J(\boldsymbol{r})} \tag{3.30}$$

and contracts to reproduce the counting function they replace:

$$C_\alpha(\boldsymbol{r}) = \frac{g_\alpha(\boldsymbol{r})}{g_\alpha(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} = \frac{\sum_J g_J^{(\alpha)}(\boldsymbol{r})}{\sum_J g_J^{(\alpha)}(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} \tag{3.31}$$

Since the degrees of freedom we use to fulfill these conditions are not immediately apparent, producing this subdivision may seem either trivial or impossible. A full explanation of how this can be done is given in appendix C, which we will summarize here by stating that we can exactly fulfill these conditions at a single point to second-order in a local Taylor expansion. We will use this scheme to subdivide two atom-centered Voronoi regions into spherical sections, and we choose to satisfy these conditions at the midpoint of the chemical bond. As the composite set of counting functions divides space into both atom-centered Voronoi regions and spherical subregions simultaneously, it strictly improves the granularity of the counting function basis. This composition scheme thus sets up a basis to capture population correlations between subregions while retaining the ability to enact number projections between distinct molecular fragments.

3.8 Results

Name	Symbol	Functional Form
One-Body Jastrow	J_1	$\sum_{i\alpha} u_1(r_{i\alpha})$
Two-Body Jastrow	J_2	$\sum_{i\neq j} u_2(r_{ij})$
Orbital rotation	$\mathbf{X}$	$\sum_{ab} X_{ab}\mathbf{a}_a^\dagger\mathbf{a}_b, \quad \boldsymbol{X}^\dagger = -\boldsymbol{X}$
Counting Jastrow	J_{NC}	$\sum_{ijIJ} F_{IJ}C_I(\boldsymbol{r}_i)C_J(\boldsymbol{r}_j)$

Table 3.2: Jastrow exponent functional forms.

Restricted Hartree-Fock (RHF), open-shell Hartree-Fock (ROHF), second-order Møller-Plesset perturbation theory (MP2), coupled-cluster singles and doubles (CCSD) with perturbative triples (CCSD(T)), complete active space self-consistent field (CASSCF), and multireference configuration interaction singles and doubles with Davidson correction (MRCI+Q) calculations were performed using GAMESS[267,268] and Molpro.[231–241] Due to the steep scaling of DMC calculations with atomic number,[269] the helium core of carbon and oxygen and the neon core of calcium were replaced by energy-consistent pseudopotentials found in the Stuttgart library[270,271] in all calculations. Variational Monte Carlo (VMC) and diffusion

Acronym	Functional Form
TJS	$e^{J_1}e^{J_2}\left\|\Psi_{HF}\right\rangle$
CJS	$e^{J_C}\left\|\Psi_{HF}\right\rangle$
CTJS	$e^{J_1}e^{J_2}e^{J_{NC}}\left\|\Psi_{HF}\right\rangle$
TJS-oo	$e^{J_1}e^{J_2}e^{\hat{X}}\left\|\Psi_{HF}\right\rangle$
CJS-oo	$e^{J_C}e^{\hat{X}}\left\|\Psi_{HF}\right\rangle$
CTJS-oo	$e^{J_C}e^{J_1}e^{J_2}e^{\hat{X}}\left\|\Psi_{HF}\right\rangle$

Table 3.3: Ansatz acronyms and their corresponding functional forms. The reference wavefunction $|\Psi_{HF}\rangle$ is a single-determinant spin-restricted Hartree-Fock wavefunction.

Monte Carlo (DMC) calculations[114] were performed in a modified version of QMCPACK[153] and the linear method[272,273] was employed to minimize VMC energies.

We will indicate the functional form of wavefunction ansatzes according to the Jastrow factor exponents in table 3.2 and acronyms in table 3.3. The CJS exponent was initially set to zero and each TJS exponent was initially set to a minimal 10-parameter cubic basis spline that fulfilled both electron-nuclear and electron-electron cusp conditions. Unless otherwise indicated, orbital optimizations were performed after pre-optimizing the Jastrow variables, and all variables were optimized together during orbital optimizations.

3.9 Ethene

We had previously shown[262] that a pair of bipartite planar counting functions recovered the correct energy ordering associated with nodal surfaces between orbitally-localized and orbitally-delocalized single-determinantal wavefunctions of ethene dissociating symmetrically along the carbon-carbon bond. The nearly-indistinguishable dissociation curves obtained from FD-CJS and NG2-CJS ansatzes in figure 3.10 demonstrates the functional equivalence of an explicit bipartite basis given in equation 3.10 and counting functions generated from appropriately parameterized carbon-centered normalized gaussians given in equation 3.12. The counting functions used in each of these wavefunctions are schematically depicted in figures 3.11, 3.12, and 3.13.

Using six atom-centered Voronoi counting regions built using the scheme in section 3.5, and depicted in figure 3.12, the NG6-CJS wavefunction achieves a lower variational energy than its two-region counterpart. In the NG6-CJS wavefunction, a contraction of the three counting functions that lay on the same methylene fragment behaves much like the counting functions in the NG2-CJS wavefunction, as the counting function gradient and curvature in the interior of the fragment cancels between neighboring regions due to the property of pointwise additivity described in section 3.4. Thus, the NG2-CJS counting function basis is (nearly exactly) contained in the space spanned by the NG6-CJS counting function basis, and the NG6-CJS wavefunction strictly improves upon the variational freedom present in the NG2-CJS wavefunction. As a result, we expect that the VMC energy of the NG6-CJS wavefunction to be bounded above by the VMC energy of the NG2-CJS wavefunction, which we do observe in figure 3.10. We can directly attribute this variational improvement to the participation of the hydrogen-centered counting functions in the NG6-CJS wavefunction,

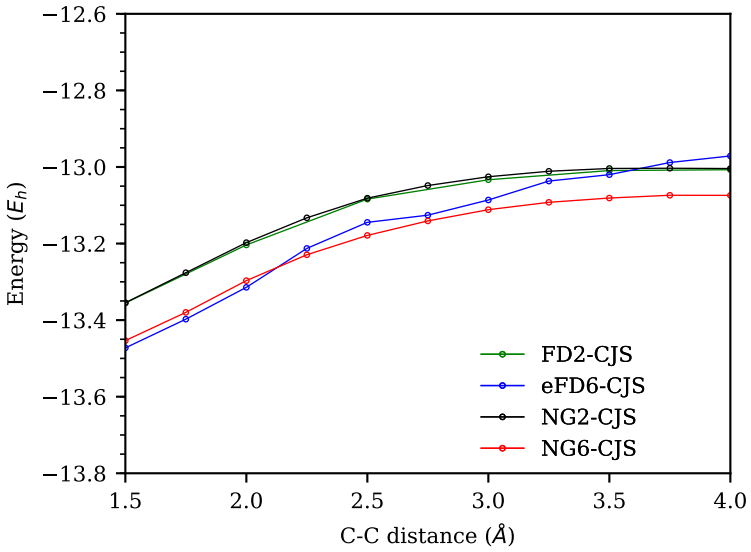

Figure 3.10: VMC energy as a function of ethene bondlength of CJS wavefunctions with a symmetry-adapted RHF reference using the counting function basis depicted in figures 3.11, 3.12, and 3.13. Fermi-Dirac style (FD) linear counting functions (equation 3.9) are explicitly sigmoidal counting functions which are equivalent to a pair of carbon-centered normalized gaussians (NG2). Subdividing these counting functions into a total of six atom-centered Voronoi cells (NG6) lowers the VMC energy. Increasing the size of the counting Jastrow basis by instead placing ellipsoidal counting functions (eFD6) around each atom did not consistently improve the variational energy for reasons discussed in section 3.4.

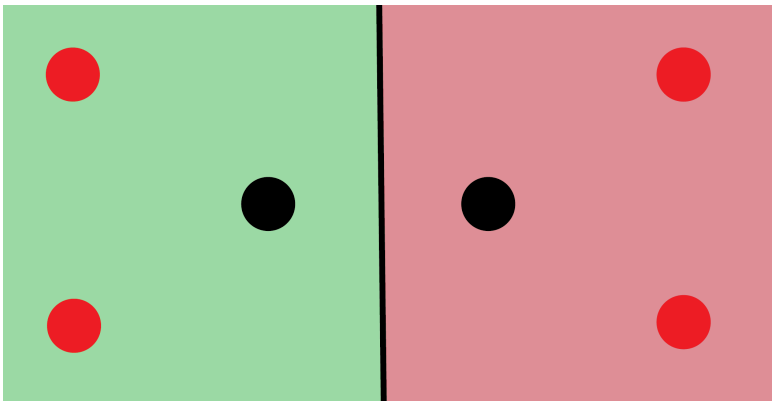

Figure 3.11: Illustration of the counting regions used in the NG2-CJS and FD2-CJS wavefunctions within the plane of the ethene molecule. Carbon and hydrogen atoms are shown as solid black and red circles respectively. Each colored region corresponds to the interior of a single counting region and approximate region boundaries are indicated by the solid black lines. Regions are generated either by explicitly constructing a planar counting function (equation 3.9) with a switching surface that bisects the central double bond (FD2) or by a two-region Voronoi tessellation using the carbon centers as Voronoi points (NG2).

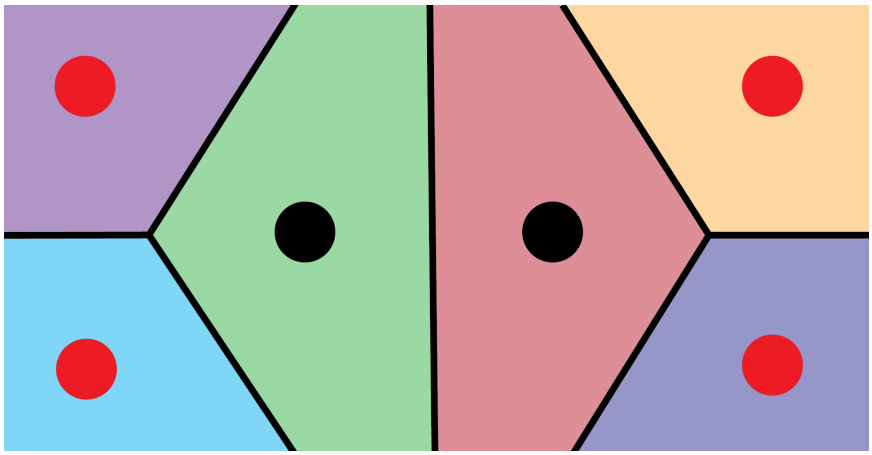

Figure 3.12: Illustration of the counting regions used in the NG6-CJS wavefunctions within the plane of the ethene molecule. Regions are built using the scheme outlined in section 3.5 which generates a Voronoi tessellation with atomic coordinates as the Voronoi points.

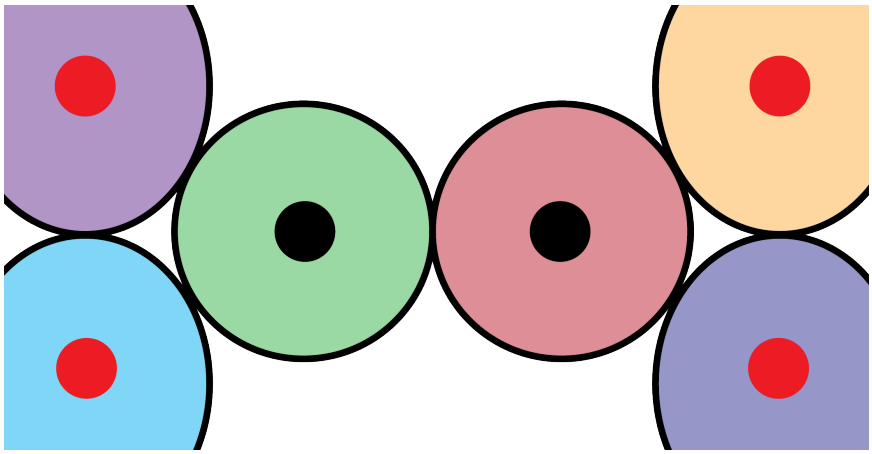

Figure 3.13: Illustration of the counting regions used in the eFD6-CJS wavefunctions within the plane of the ethene molecule. Regions are chosen to be packed ellipsoids using the Fermi-Dirac style ellipsoidal counting functions in equation 3.10. As discussed in section 3.4, the sigmoidal switches that occur between counting region interiors and the void regions disrupt number projections by introducing a kinetic energy cost at the boundary.

showing that these normalized gaussian counting functions can easily accommodate correlations between atoms in a single molecular fragment, a feat that was not achievable using the previous basis scheme due to the pathologies described in section 3.3. The same logic applies as we produce increasingly granular subdivisions of counting regions that strictly expand the span of the linear counting function space, and doing so systematically approaches the basis set limit for these Jastrow factors in a variational way.

3.10 Random Planar H_4

The bipartite counting function basis in earlier calculations[262] had been set up in an *ad hoc* way, and consisted of two counting functions whose collective planar switching surface was carefully placed to bisect a dissociating chemical bond. As shown in section 3.5, an atom-centered Voronoi tessellation can be easily generated using a normalized counting function basis described in section 3.4 by placing the anchor gaussian means at atomic coordinates. To investigate the effectiveness of this automatic generation scheme under less controlled conditions, we look at the fractional correlation energy recovered in various wavefunctions for 94 random planar geometries of H_4, with each atomic coordinate chosen randomly and uniformly within a 5Å × 5Å square. Geometries in which any two atoms were closer than 0.1Å were discarded from a total random sample of 100.

Wavefunction	% E_{corr} Average	%E_{corr} Std. Dev.
MP2	64.3%	6.1%
TJS-oo	78.6%	5.7%
CASSCF(4e,4o)	84.0%	12.5%
CCSD	87.1%	10.4%
CTJS-oo	95.0%	2.2%
CASPT2	97.0%	2.3%

Table 3.4: Mean and variance of the fractional correlation energies of 94 random H_4 geometries using the data set in figure 3.17.

Fractional correlation energies recovered by various methods are compared to those obtained from CCSD in figure 3.17, and indicate that these counting Jastrow factors recover a significant fraction of the total correlation energy in geometries where the other single-reference methods (TJS-oo and CCSD) struggle to do so. Addition of the counting Jastrow notably improves the correlation energy distribution relative to the standard TJS-oo wavefunctions across random geometries, shown in figure 3.14, and consistently recovers a high fraction of correlation energy on par with standard deterministic multireference methods in this simple system, as shown in table 3.4.

We can directly measure the projective effect of these Jastrows by plotting the density of region populations within the many-body wavefunction:

$$\rho_I(N) = \int |\Psi(\{\boldsymbol{r}_i\})|^2 \; \delta\left(N - \sum_i C_I(\boldsymbol{r}_i)\right) d\{\boldsymbol{r}_i\} \tag{3.32}$$

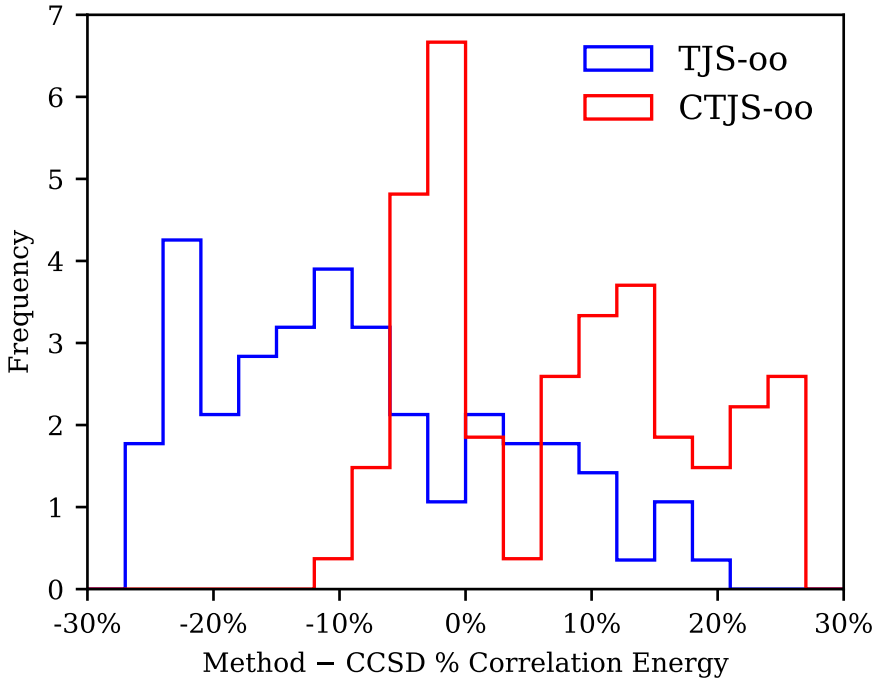

Figure 3.14: Frequency of fractional correlation energies recovered by the TJS-oo and CTJS-oo wavefunctions relative to a CCSD reference, using the data set plotted in figure 3.17. Fractional correlation energies for both wavefunctions were calculated using VMC, and the average fractional correlation energy error is 0.2%.

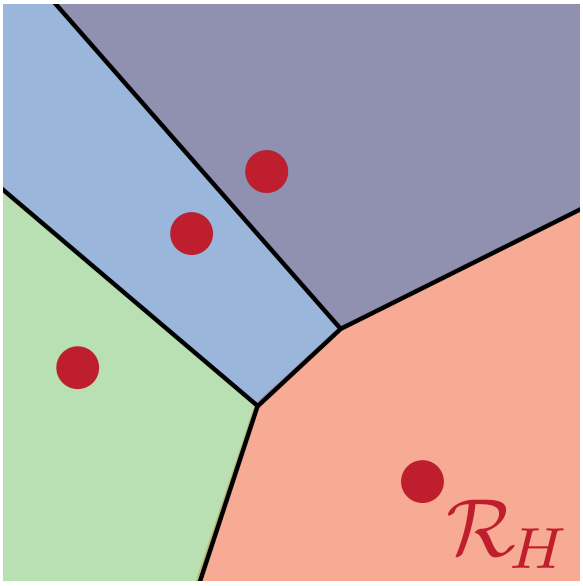

Figure 3.15: A representative random arrangement of H_4 atoms (indicated by solid red circles) and counting regions (indicated by colored areas) in a 5Å × 5Å plane. The population densities of the orange counting function $\mathcal{R}_H$ in the lower-right of the figure are calculated using equation 3.32 for several variationally optimized wavefunctions in figure 3.16.

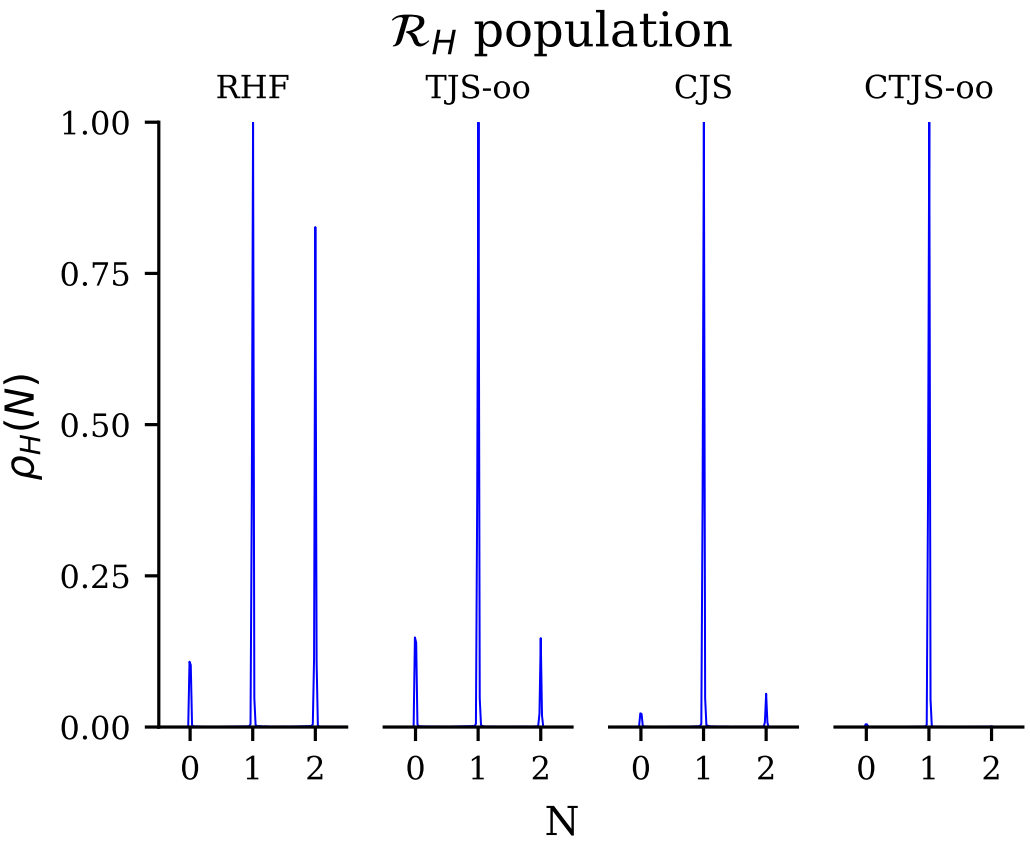

Figure 3.16: Sampled population density of the counting region $\mathcal{R}_H$ from figure 3.15 for various real-space wavefunctions using equation 3.33 with $\sigma = 0.03$. Narrow peak widths indicate both that there is little overlap between the reference wavefunction and the counting function switching surface and that this counting function acts much like a Fock-space number operator. The single-determinant reference contains high-energy ionic terms that are only partially removed by the one- and two-body Jastrow factors present in a TJS-oo wavefunction but are almost completely removed by the counting Jastrow factor in the CJS and CTJS-oo wavefunctions.

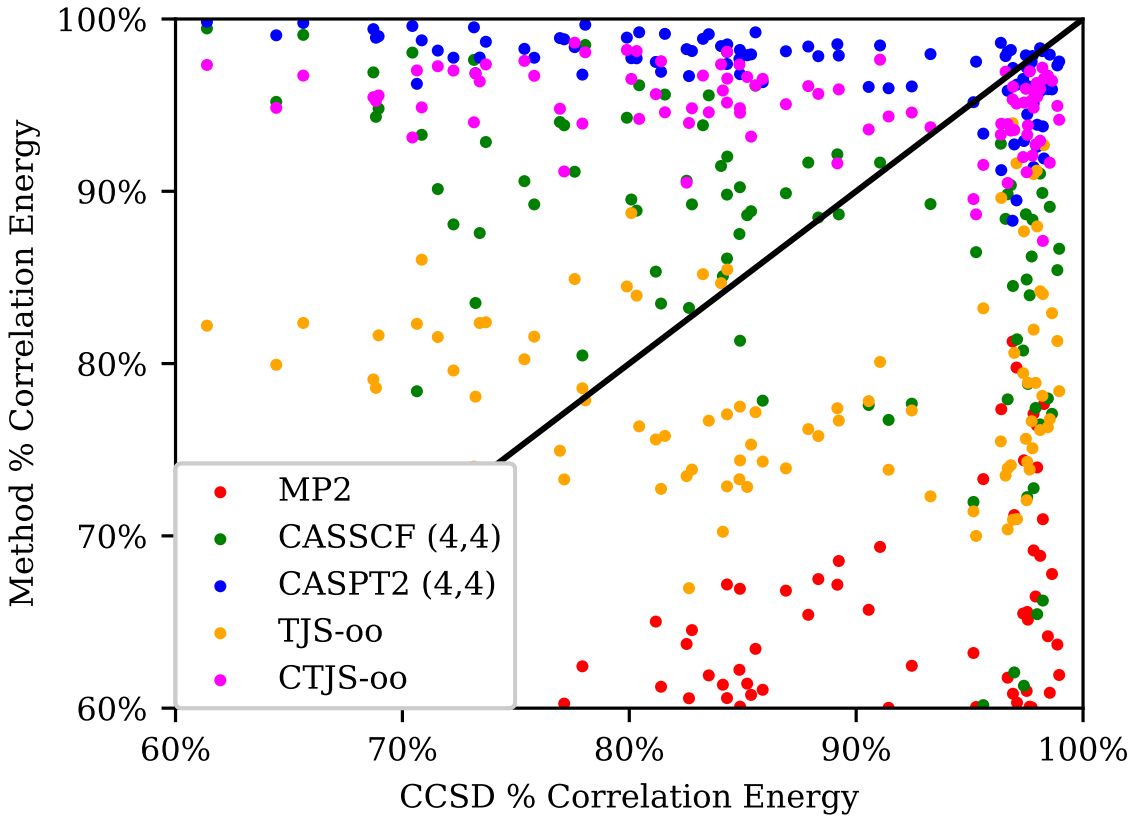

Figure 3.17: Fractional correlation energies recovered by various wavefunctions plotted against the fractional correlation energy recovered by a CCSD reference for 94 random arrangements of H_4 in in a 5Å × 5Å plane. Correlation energy values for the TJS-oo and CTJS-oo wavefunctions were calculated using Variational Monte Carlo (VMC), where the absolute stochastic error is less than 1 mE_h and the average error of fractional correlation energy is 0.2%, less than the plotted symbol size. Calculations were performed in the cc-pVQZ basis and all wavefunction parameters were optimized in variational calculations. CASSCF calculated were performed using a minimal 4e-4o active space and subsequent CASPT2 and MRCI+Q calculations considered all single and double excitations from this space. The counting regions used in CTJS-oo wavefunctions were generated according to section 3.5 in an atom-centered Voronoi arrangement, and an example counting region setup is given in figure 3.15. Benchmark correlation energies were calculated by performing a three-point (cc-pVDZ, cc-pVTZ, cc-pVQZ) basis set extrapolation on MRCI+Q energies.[1,2]

The function $\rho_I(N)$ is the density of electronic configurations in $|\Psi|^2$ for which the counting region $\mathcal{R}_I$ contains N electrons as determined by the counting function C_I.

The population density of the counting function $\mathcal{R}_H$ in the atomic geometry shown in figure 3.15 is plotted in figure 3.16. The value of this density function is approximated in a continuous way by stochastically sampling the wavefunction probability distribution $|\Psi|^2$, approximating the integral in equation 3.32 as a sum over these samples, and smoothing the resulting distribution using fixed-width gaussians:

$$\rho_I(N) \approx \frac{1}{\nu\sqrt{2\pi\sigma^2}} \sum_{\{\boldsymbol{r}_i\}\sim|\Psi|^2}^{\nu} \exp\left[-\left(N - \sum_i C_I(\boldsymbol{r}_i)\right)^2 / (2\sigma^2)\right] \tag{3.33}$$

The sampling procdure can be easily integrated into existing optimization techniques which rely on wavefunction derivatives because the counting function values directly correspond to the derivatives of linear counting Jastrow coefficients:

$$\frac{\partial J}{\partial G_I} = \sum_i C_I(\boldsymbol{r}_i) \tag{3.34}$$

Comparing the population density peak-heights between different wavefunctions reveals that the counting Jastrow factor suppresses ionic configurations in ways that one- and two-body Jastrow factors cannot. Ionic configurations corresponding to the peaks at $N = 0$ and $N = 2$ are only partially removed in the TJS wavefunction but more fully projected out in CJS wavefunctions using a simple atom-centered Voronoi basis. Based on the counting Jastrows' effectiveness when applied to these randomly generated systems, we conclude that these normalized counting function basis has potential as a black-box number projection factor in more general chemical systems, and are primed to perform number projections when variationally favorable.

3.11 Calcium Oxide

Our calculations have thus far focused on simple examples where the number projecting action of the counting Jastrow is applied to remove ionic terms from a symmetry-restricted single-reference wavefunction. A minimal, atom-centered Voronoi basis (section 3.5) of counting functions is sufficient to correctly dissociate the hydrogen molecule, recover the correct state nodal surface for ethene at dissociation,[262] and account for strong correlation in randomized planar H_4 geometries. However, in non-symmetric dissociation processes, fragment-based number projection accomplishes little, as molecular orbitals tend to already localize in SCF algorithms to avoid producing high-energy interfragment ionic terms. Table 3.5 demonstrates this effect prominently in the dissociation of molecular calcium oxide, where the 2-CJS wavefunction containing a counting Jastrow built from a pair of atom-centered counting functions fails to meaningfully improve the wavefunction's energy relative to the Hartree-Fock reference.

Section 3.4 shows how normalized counting functions can tile space in systematic, automatic, and flexible ways while avoiding kinetic energy problems that result from careless

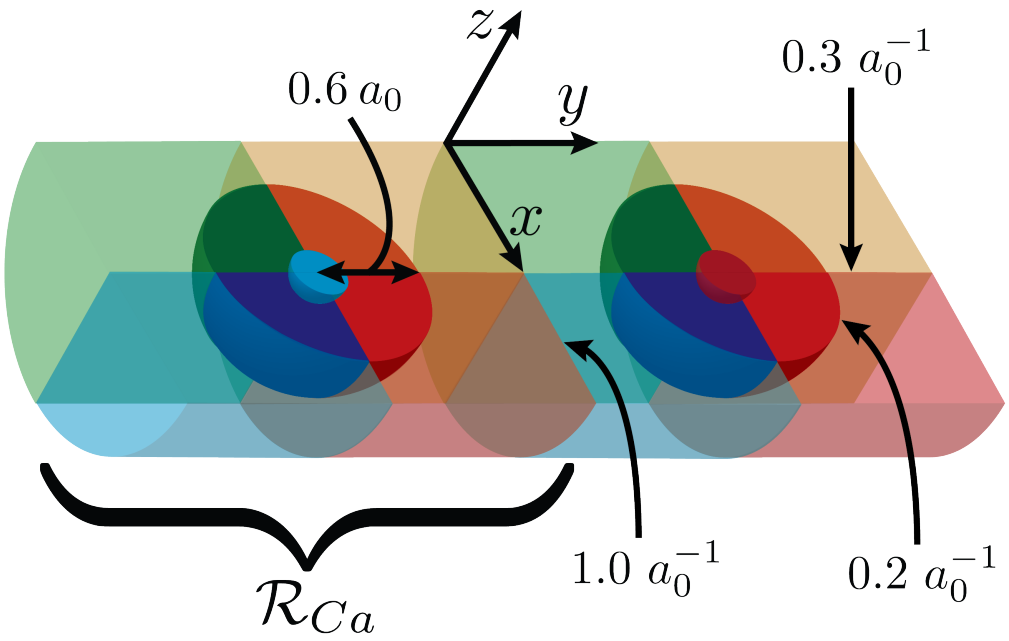

Figure 3.18: Cutaway of the counting regions used in the CaO counting Jastrow factor. The calcium atom is indicated by the solid blue half-sphere and the oxygen atom is indicated by the solid red half-sphere, and the remaining colored spherical sections indicate individual counting regions. Two atom-centered regions first divide space along a plane that bisects the bonding axis and attains a maximum slope of 1.0 Bohr^{-1} and are further subdivided into eight octants and two spherical shells to give a total of 32 counting regions. Only those below the x-y plane are shown here. Radial boundaries of both spherical partitions occur at a radius of 0.6 Bohr on which pair counting functions have a maximum slope of 0.2 Bohr^{-1}. Angular switches divide each atomic region into eight octants, and attain a maximum slope of 0.3 Bohr^{-1} across their entire switching surface. $\mathcal{R}_{Ca}$ corresponds to the union of sixteen regions that subdivide the Calcium-centered Voronoi region, and is the focus of the population density analyses in figures 3.19, 3.20, and 3.21.

placement of switching surfaces. Normalized counting functions are built from a set of anchor gaussians whose parameters can be chosen to divide space into classical (section 3.5) or spherical (section 3.6) Voronoi grids, and are not limited to the simple bipartite partitions previously used. As described in section 3.7, smaller grids may be embedded within a single larger grid so that the resulting patterns of counting regions preserves the structure of both grids simultaneously. Subdividing atomic Voronoi regions into spherical sections using this composition scheme gives counting Jastrows the ability to describe population correlations within and between atomic centers simultaneously.

Molecular calcium oxide undergoes a complex dissociation, and exhibits both a singlet-triplet crossing very near equilibrium geometries and a valence calcium d-orbital participating in the chemical double bond. In order to address the complex electronic correlations during the dissociation process, we embed a 16-region spherical partition within each of two atom-centered regions. These spherical partitions divide the space around each of the calcium and oxygen atoms into Cartesian octants and further into two radial shells, as depicted in figure 3.18. The composite counting regions very nearly meet the conditions discussed in appendix C as the counting functions each 16-region subdivision can be recombined to approximately form each of the two original atom-centered counting functions. Much like atomic orbital exponents,[274] the anchor gaussian exponent parameters are highly nonlinear – a small change in these parameters potentially changes multiple switching surfaces simultaneously – and we saw little variational benefit when optimizing them alongside other wavefunction parameters in these more complex partitions. As a result, these anchor gaussian parameters are held fixed in order to reduce the burden of parameter optimization and to simplify the population density analysis performed afterward. The geometric parameters of this spherical partition and the gradient values across the counting functions' switching surfaces were chosen based on those that minimized CJS variational energies for atomic beryllium, calcium, oxygen, and nitrogen whose counting functions had the same radial and angular divisions.

In this system, we focus our attention on the accuracy of the crossing point and the energy gap between the $^1\Sigma$ singlet and $^3\Pi$ triplet states of calcium oxide at bondlengths between 2Å and 3Å using a single-determinant Jastrow-Slater wavefunction. In contrast to the minimal, bipartite atom-centered Voronoi counting region partition, a counting Jastrow using the more intricate 32-region basis recovers a significant amount of correlation energy at the VMC level, and meaningfully improves both the singlet-triplet crossing point and the RMSE of the singlet-triplet gap function, as shown in table 3.5. The combined effect of both cusp-correcting and counting Jastrows in the 32-CTJS wavefunction moves the crossing point approximately 0.1Å further toward the benchmark MRCI+Q crossing point near 2.25 Å. Further optimizing the orbitals in the 32-CTJS-oo wavefunctions also improves the singlet-triplet gap RMSE by a modest 15.1 $\mathrm{mE_h}$ relative to the more conventional TJS-oo wavefunction.

We again look at population density functions defined in equation 3.32 to more precisely judge the number projecting effect of the counting Jastrow factor. Population densities of the aggregate calcium-centered counting regions, indicated by $\mathcal{R}_{Ca}$ in figure 3.18, are given for singlet and triplet wavefunctions in figures 3.19 and 3.21 respectively. Integrated peak areas of these population density plots – which signify the total fraction of the wavefunction that attains the indicated electronic population in the calcium-centered region – are given in table 3.6. These density functions clearly show that the counting Jastrow factors effectively

Method	Crossing (Å)	E_{corr}	Gap RMSE
2-CJS	1.87(5)	0(3)	204.4
2-CJS-oo	1.94(3)	6(3)	188.6
32-CJS	2.01(4)	360(2)	60.4
32-CJS-oo	1.994(2)	394(3)	43.4
TJS	2.02(2)	524(1)	54.5
TJS-oo	2.012(6)	546(1)	43.8
32-CTJS	2.08(1)	543(1)	49.9
32-CTJS-oo	2.081(4)	581(1)	28.7
CCSD(T)	2.32(4)	697	16.2
MRCI+Q	2.250(3)	675	N/A

Table 3.5: Singlet-triplet crossing points, the singlet's correlation energy E_{corr} in mE_h at 3 Å, and the root-mean square error (RMSE) of the singlet-triplet gap in mE_h calculated using VMC for a set of five Ca-O bondlengths at 0.25Å intervals from 2Å to 3Å compared to an (8e,80) MRCI+Q reference. Number-prefixes for CJS wavefunctions correspond to the size of the counting function basis, and are either atom-centered Voronoi regions (2-CJS) or set up according to the description in figure 3.18 (32-CJS). Crossing points and their uncertainties were determined by finding the roots of the singlet-triplet gap function using a curve fit given by a three-parameter exponential function $a + b\exp(-cx)$ and were found to be largely insensitive to the choice of fitting function. Estimates for the crossing points of the 2-CJS wavefunctions are particularly poor since, much like the bare single-determinant wavefunction, they did not exhibit a singlet-triplet crossing in the bondlength interval studied.

redistribute electrons between atoms when applied to singlet reference configurations and act most prominently when orbitals are coupled to the counting Jastrow factor, independent of the presence of cusp-correcting Jastrow factors. Since the two-region partition in the 2-CJS wavefunctions showed negligible effect on the wavefunction, we conclude that these are taking full advantage of this fine 32-region division to introduce intraatomic correlations that allow the Jastrow to reweigh configurations based on atomic populations.

In contrast to VMC energies that are straightforwardly affected by the amplitudes of the trial wavefunction, Diffusion Monte Carlo (DMC) energies are related to the trial wavefunction in a more indirect way. In order to ensure that the sampled distribution remains appropriately antisymmetric while avoiding the exponential sign problem,[146] DMC commonly employs the fixed-node approximation, where the Fermionic nodes of a trial wavefunction restrict the Monte Carlo sampling procedure.[114] In addition, due to the smaller timestep required to resolve the higher energy scale of core electrons, nonlocal pseudopotentials replace their explicit simulation, much like the frozen core approximation that reduce the number of excitations considered in post-Hartree-Fock methods. In DMC, these nonlocal components are evaluated through the locality approximation[275] or its variational analogue[276,277] (used here), and their accuracy directly relies on the quality of trial wavefunction amplitudes relative to the projected ground state. However, the relative magnitude of these approximations remains opaque due to the complexity of both the many-body nodal surface and nonlocal

Method	State	R (Å)	N=8	N=9	N=10
TJS-oo	$^1\Sigma$	2.75	0.202	0.589	0.204
32-CJS-oo	$^1\Sigma$	2.75	0.199	0.594	0.203
32-CTJS-oo	$^1\Sigma$	2.75	0.138	0.624	0.222
RHF	$^1\Sigma$	3.00	0.095	0.410	0.490
32-CJS	$^1\Sigma$	3.00	0.075	0.526	0.396
TJS-oo	$^1\Sigma$	3.00	0.133	0.524	0.336
32-CJS-oo	$^1\Sigma$	3.00	0.147	0.618	0.229
32-CTJS-oo	$^1\Sigma$	3.00	0.155	0.589	0.251
RHF	$^3\Pi$	3.00	0.017	0.816	0.161
32-CJS	$^3\Pi$	3.00	0.023	0.816	0.155
32-TJS-oo	$^3\Pi$	3.00	0.007	0.861	0.129
32-CJS-oo	$^3\Pi$	3.00	0.008	0.869	0.115
32-CTJS-oo	$^3\Pi$	3.00	0.004	0.875	0.120

Table 3.6: Peak areas for the population density peaks of the indicated wavefunction, state symmetry, and bond distance R, some of which are shown in figures 3.19 and 3.20, integrated over the domain $[N-0.5, N+0.5]$. These values roughly correspond to the fraction of the wavefunction in which the indicated number of electrons populate the aggregate calcium-centered region $\mathcal{R}_{Ca}$ shown in figure 3.18.

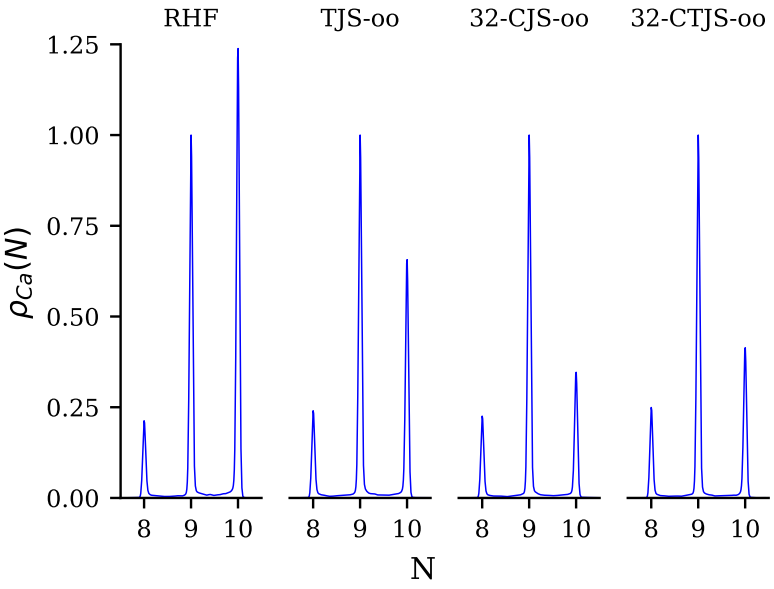

Figure 3.19: Population density distribution of the aggregate 16 regions surrounding the calcium atom indicated by $\mathcal{R}_{Ca}$ in figure 3.18 calculated using equation 3.33 with $\sigma = 0.03$ for the indicated singlet wavefunctions at 3 Å. Distributions are rescaled to unit peak height at $N = 9$ and integrated peak areas of the normalized distribution is given in table 3.6.

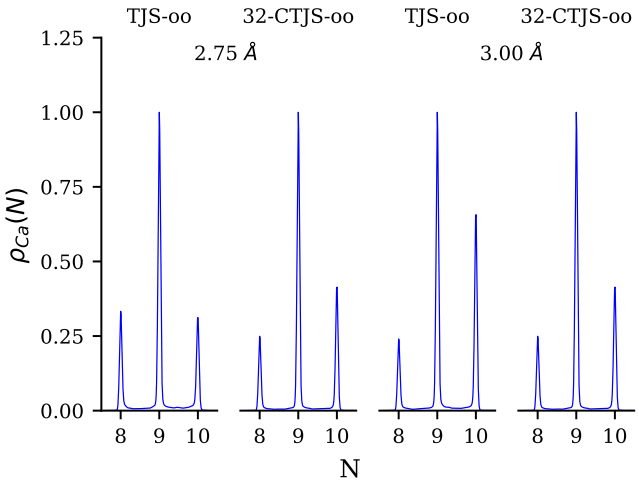

Figure 3.20: Population density distribution of the aggregate 16 regions surrounding the calcium atom indicated by $\mathcal{R}_{Ca}$ in figure 3.18 calculated using equation 3.33 with $\sigma = 0.03$ for the TJS-oo and 32-CTJS-oo singlet wavefunctions at 2.75 Å and 3 Å. Distributions are rescaled to unit peak height at $N = 9$ and integrated peak areas of the normalized distribution is given in table 3.6.

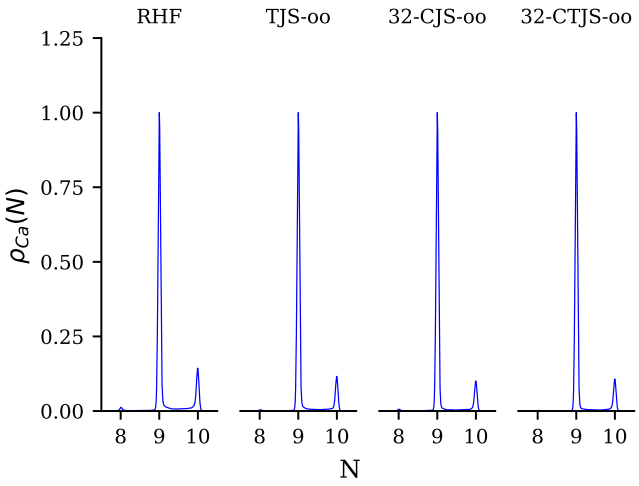

Figure 3.21: Population density distribution of the aggregate 16 regions surrounding the calcium atom indicated by $\mathcal{R}_{Ca}$ in figure 3.18 calculated using equation 3.33 with $\sigma = 0.03$ for the indicated $^3\Pi$ triplet wavefunctions at 3Å. Distributions are rescaled to unit peak height at $N = 9$ and integrated peak areas of the normalized distribution is given in table 3.6.

Method	Crossing (Å)	E_{corr}	Gap RMSE
2-CJS	2.20(3)	571(2)	22.9
2-CJS-oo	2.220(4)	569(2)	23.3
32-CJS	2.19(3)	621(1)	24.1
32-CJS-oo	2.22(5)	651(1)	4.9
TJS	2.21(1)	627(1)	23.2
TJS-oo	2.19(5)	647(1)	10.1
32-CTJS	2.19(2)	630(1)	23.6
32-CTJS-oo	2.220(1)	664(1)	4.2
CCSD(T)	2.32(3)	697	16.2
MRCI+Q	2.249(3)	675	N/A

Table 3.7: As table 3.5, but for DMC instead of VMC.

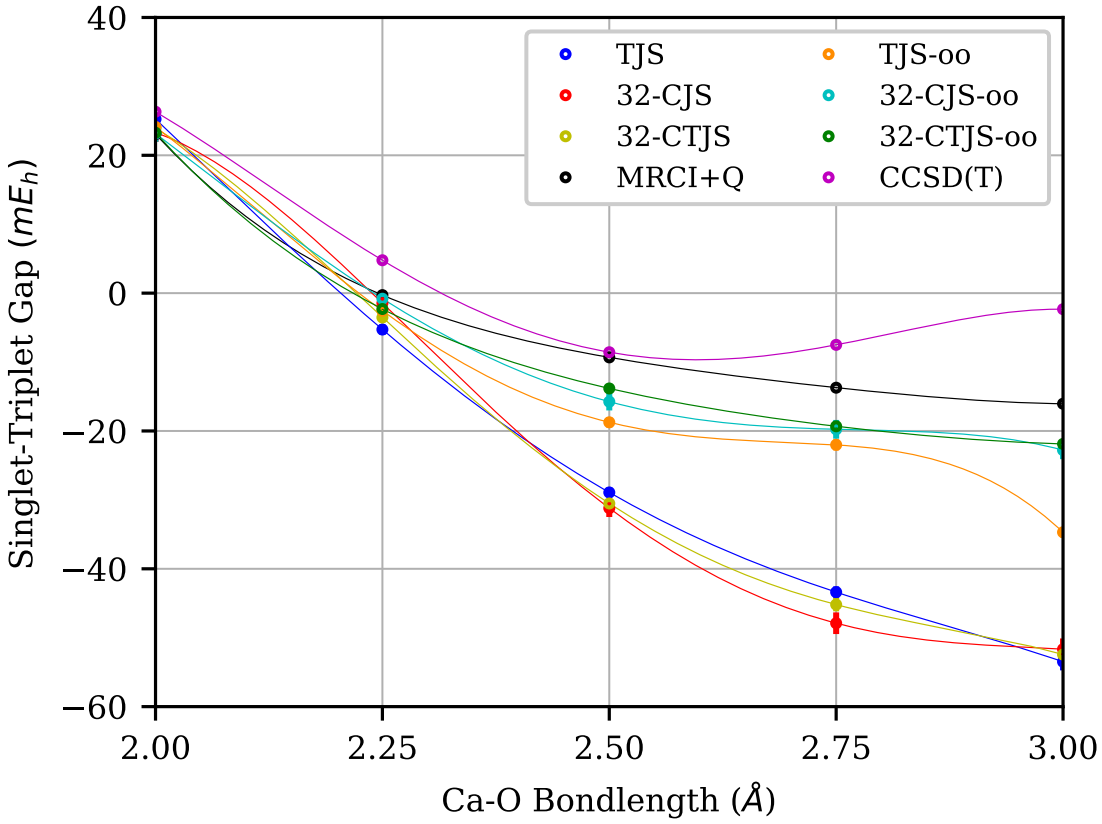

Figure 3.22: DMC singlet-triplet energy gap value for various wavefunctions as a function of Ca-O bondlength between 2Å and 3Å. Lines are a guide to the eye. The DMC timestep is 0.02 $\hbar$ $E_h{}^{-1}$ and absolute stochastic errors are less than 2 mE_h and are smaller than the plotted symbol size except for CJS and CJS-oo wavefunctions.

pseudopotential terms,[278] and it remains difficult to differentiate or address these sources of error directly.

The behavior of the singlet-triplet DMC energy gap, shown in figure 3.22 with RMSE given in table 3.7, roughly separates the real-space wavefunctions into two clusters: those without optimized orbitals (TJS, 32-CJS, 32-CTJS) and those with optimized orbitals (TJS-oo, 32-CJS-oo, 32-CTJS-oo) Despite their disparate VMC energies (table 3.5), those in the former cluster exhibit very similar DMC energy gaps between 2.0 Å and 3.0 Å. Since these wavefunctions all share the same Fermionic nodal surface, this strongly suggests that that cusp-correcting and counting Jastrow factors largely affect nonlocal pseudopotential evaluation uniformly in both singlet and triplet wavefunctions. Note that the difference in singlet correlation energy between the 2-CJS and 32-CTJS wavefunctions gives us an idea of the magnitude of the nonlocal pseudopotential error present, as the two-region counting Jastrow does very little to change the underlying Hartree-Fock reference. At around 60 mH, this error is surprisingly large, and yet the TJS and 32-CJS wavefunctions correct it to a similar degree (56 mH and 50 mH, respectively) and in similar ways (missing only 3 mH and 9 mH from the composite 32-CTJS result, respectively) despite their very different functional forms. By contrast, those in the latter cluster exhibit a marked improvement in the RMSE of the singlet-triplet gap, and as orbital rotations provide the only avenue for changing the wavefunction's nodal surface, it is safe to say these orbital rotations – through coupling to the Jastrow factors – are responsible for improvements to the nodal surface.

All of the DMC singlet-triplet gaps roughly match the MRCI+Q reference until the crossing point around 2.2Å, after which they start to deviate more strongly. For example, while the highly accurate single-reference Fock space wavefunction CCSD(T) is accurate near equilibrium, it overestimates the crossing and starts to diverge at longer bondlengths, producing a qualitatively incorrect singlet-triplet gap at stretched geometries. Interestingly, the inaccuracy of the DMC gap of the TJS-oo wavefunction at 3.0 Å is coincident with a stark change in the calcium population density, shown in figure 3.20 and table 3.6, a feature not shared by either the 32-CJS-oo or 32-CTJS-oo wavefunction. In addition, the singlet-triplet gap of the 32-CJS-oo wavefunction and the 32-CTJS-oo wavefunction are nearly on top of each other despite the much higher VMC energy of the former (table 3.5). These observations suggest that the counting Jastrow is particularly effective at coupling to the orbital parameters in this case, which is seemingly linked to their ability to describe long-range population correlations missing from cusp-correcting Jastrows.

3.12 Conclusions

In this chapter, we have shown how to construct a one-particle basis which partitions space in a complete and natural way. These basis functions are referred to as counting functions, and are designed to act as the position-space equivalent of Fock-space number operators. When used as the basis in a four-body Jastrow correlation factor these counting functions are able to correlate electronic populations between well-characterized counting regions. These counting functions are smooth, sigmoidal functions that are made by normalizing a set of three-dimensional gaussian functions and are complete, localizable, and cleanly additive. We show how this normalization condition is responsible for these attractive formal properties, and how it alleviates many of the problems posed by a real-space formulation of a population-

based projection factor. Simple sigmoidal functions closely approximate local facets of these counting regions, and the boundaries of these regions may be clearly visualized as a patchwork of linear and quadratic surfaces. We provide two parameterization schemes designed to describe correlations in orbital populations between and within atoms that arrange counting regions in patterns of atom-centered Voronoi cells or extruded spherical Voronoi partitions respectively. Finally, taking advantage of these basis functions' clean additivity, we show how a composition scheme can be used to subdivide these counting functions in a useful way.

In simple molecular systems, these number projections, alongside cusp-correcting Jastrows and orbital rotations, capture nodal surface details and correlation energies beyond what is currently achievable using several sophisticated single-reference methods. For example, in random planar arrangements of H_4, counting-Jastrow-augmented wavefunctions consistently outperform CCSD and TJF-oo, and on average recovers nearly as much fractional correlation energy as multi-determinant CASPT2, and can be optimized at a computational cost that scales no higher than existing real-space Jastrows. For nonsymmetric dissociations, like molecular calcium oxide, a simple pair of atom-based Voronoi counting regions are ineffective at improving the wavefunction through the counting Jastrow. A counting function basis that subdivides each of these atom-based cells into a coarse spherical grid meaningfully improves the 32-CJS-oo variational energies, and reduces the mean square error of the DMC singlet-triplet gap around the crossing point by a factor of two relative to the TJF-oo wavefunction and three relative to CCSD(T). The counting Jastrows' number projecting action is explicitly verified by inspecting measures of the counting region population density, and appears to be more effective at suppressing molecular terms with high charge-separation than standard real-space jastrow factors in H_4 and CaO.

As a multiplicative real-space factor, counting Jastrow factors are trivially compatible with existing real-space wavefunctions, and have the potential to augment them in a unique and compact way. Our present calculations have used only a single-determinant reference in an effort to demonstrate the power of these number projecting factors in a limited setting, and future work will explore complex references that span a larger portion of the configuration space, taking full advantage of recent advances in multi-Slater expansion optimization techniques[199,279] and compact functional forms like the antisymmeterized geminal power ansatz. The counting Jastrow does not increase the compuational cost-scaling of wavefunction evaluations compared to those already within the reach of DMC in larger chemical systems (such as TJS with $\sim$1000 electrons)[178–180] and can be straightforwardly applied at scale, subject to the development of optimization techniques[194] that can efficiently handle the growing number of variational parameters that these Jastrows and orbital rotations introduce. In particular, the counting Jastrow's efficient representation of population-based correlations suggests application to charge-transfer excited states and complements recent work in excited state variational theory.[33] As we look toward applying these counting Jastrows in more complex systems, the flexibility and adaptability built into this normalized gaussian basis will allow us to systematically construct compact and powerful number projections and optimize them at low-order polynomial cost.

4 Stable and Fast Excited State Orbital Optimization

4.1 Introduction

Electronically excited states are at the center of many chemical processes of practical interest, including photodynamics,[5,280] light harvesting,[281,282] and x-ray spectroscopy.[283] Electronic excitations that involve significant changes to electronic density, such as charge-transfer or core excitations,[283] can introduce complications[284–286] or inaccuracies[34,287] in affordable excited state methods. Accurate treatment of these long-range excitations involves also treating the surrounding electronic response, and the algorithms that compactly describe these these relaxation effects often involve finessing tricky non-linear optimization procedures to consistently converge.[288–291] Despite these difficulties, recent advancements in excited-state targeting algorithms and in the Variational Principles that enable them[31,33–35,37,203] have shown that these orbital relaxation effects can be integrated into even more flexible wavefunctions at mean-field cost. However, as these methods are under active development, straightforward numerical accelerations remain unimplemented and subtler algorithmic instabilities remain unaddressed. Improving these nascent methods on both counts is vital for them to see the widespread use that their theories rightly deserve.

Electronic ground states have the computational benefit of being global minima, and many methods rely on Variational Principles[39,43,53] to find them using well-established numerical minimization procedures.[57,292,293] By contrast, while excited states are also stationary points of the Hamiltonian, they often manifest as saddle points in the energy surface, and optimization algorithms that pursue them can require special modifications to consistently converge. There is convenient exception to this in linear spaces, however, where stationary points can be straightforwardly obtained by standard diagonalization algorithms, where orthogonalization against lower-energy states stabilizes higher-energy states against collapse. While this can be computationally inconvenient for even valence excitations, it is particularly so for core excitations, as obtaining a highly-excited core state can first require calculation of an unreasonably long tower of lower-energy states.

In practice, excited state methods are often built as direct extensions to ground state methods, and one popular framework for this is known as Linear Response theory.[23] Typically Linear Response (LR) theory involves calculating the ground state response to external monochromatic electromagnetic perturbations, and excitation energies are identified as those which provoke singular response. A close relative of LR theory instead constructs excited states directly in the wavefunction tangent space[294,295] (also called the linear response space[21,199]), a linear function space using wavefunction parameter derivatives as its basis:

$$|\Psi_{exc}\rangle = \sum_i c_i \left| \frac{\partial \Psi_{gs}}{\partial p_i} \right\rangle \tag{4.1}$$

Wavefunction parameters p_i characterize the ground state wavefunction, and can vary from CI coefficients, to orbital parameters, and cluster amplitudes,[53] and this space is often a natural starting point in which to search for excited states.

Perhaps the simplest of these LR methods is known as Configuration Interaction Singles (CIS),[296] in which excited states are formed in the tangent space of the Hartree-Fock

wavefunction, consisting of the space of singly-excited determinants. While it can be understood as a truncated CI method, CIS can also be characterized as an approximation[50] to time-dependent Hartree Fock theory.[297] Though it fails to capture finer details of correlation and can fail in more extreme situations, the CIS wavefunction qualitatively captures what is meant by a single-particle excitation, in which a single electron is promoted from a low-energy to high-energy orbital. This single-excitation is so ubiquitous in excited state analysis and has been extended to a wide-variety of excited state methods as Natural Transition Orbitals,[298–301] which can be used to obtain a picture of electronic density movement upon excitation.

In order to accurately describe many real electronic excitations, however, excited state methods must go beyond this single-excitation picture. For instance, CIS is most accurate when excited electrons don't travel far from their ground-state source, and infamously overestimates the energy of charge-transfer (CT) excitations by 1-2 eV[287] and core excitations by more than 10 eV.[302] As their names imply, both CT and core excitations involve drastic shifts in electronic density, and these inaccuracies are rooted in the fact that CIS does not effectively allow the many-body electronic environment to shift its density in response. This is an effect known as orbital relaxation, as this response involves a qualitative shift in the zeroth-order position of the electronic environment that is commonly described in terms of molecular orbitals. Perturbative approaches that aim to account for orbital relaxation are not always successful,[34,303] and variational methods have instead been developed[35–37,304] to perform the orbital optimization necessary to describe this shift. However, orbital optimization, especially of excited states, can be a fraught procedure, and its mathematical and numerical complexity helps explain its absence from standard methods.

Time-dependent density functional theory[22–24] (TD-DFT) is the most popular excited state method due to its mixture of speed and accuracy for many systems and is broadly effective at capturing details of weak correlation in both ground and excited states. As TD-DFT is based on an approximate treatment of electronic exchange, self-interaction error[305] can cause underestimation of charge transfer excitated state energies.[306–308] Though functionals that address the particular difficulties present in charge-transfer[306] and core excitations have been developed, results can nonetheless be highly sensitive to internal parameters[284–286,309] and functional choice.[283,310,311] While CIS by itself has little to offer over TD-DFT in terms of raw accuracy, the relative simplicity of CIS theory makes improvements straightforward (if not costly), while the theoretical underpinnings of TD-DFT can make improvements difficult to implement. The adiabatic approximation[312] that is used to adapt ground-state functionals to excited-state calculations complicates extending TD-DFT to account for doubly-excited states,[25–28] orbital relaxations,[29,313,314] or multi-determinant frameworks.[315,316]

Linear Response theory is not the only game in town, and Δ-SCF[288–290,317] methods have been developed as a variational alternative that efficiently folds in orbital relaxation effects. These Δ-SCF methods conveniently use the same fundamental mathematical approach as ground-state SCF methods, which variationally optimize the orbitals in a single-determinant wavefunction, and only require slight modification of existing code to use. They differ only by aiming at higher-energy solutions – now identified as excited electronic states – and the flexibility of these orbital variations means that doubly-excited states[289] can be accurately represented.

These approaches, however, are not without their drawbacks. Excited state wavefunc-

tions in Δ-SCF are written as a single Slater determinant, and as such, naturally miss the details of weak correlation that aren't captured by orbital variations. Strong forms of correlation common to single excitations, such as is present in open-shell singlets, can be resolved by spin-purification schemes,[318] but these do not address more general correlation effects. Excited states often manifest as saddle points on the excited state surface, and the orbital optimization at the heart of SCF algorithms are highly nonlinear, and may converge to the wrong state or fail to converge altogether. Because the excited-state orbitals are allowed to vary independently of the ground-state orbitals, there is no longer a formal guarantee of orthogonality between the two states, and re-orthogonalizing these states is not always trivial.

Many of the alternatives to these approaches, such as EOM-CCSD,[21] ADC,[319] and CASSCF,[30,100–104] can be prohibitively expensive and may still be unable to effectively account for orbital relaxation effects. As a result, there is ample reason to search for methods that can combine elements of both the CIS single excitation framework and orbital relaxations at low cost. Recent advancements in excited-state methods, including square-gradient minimization[289] and excited-state targeting functions[31,33,35] have shown success in stabilizing otherwise difficult orbital optimization procedures. Recently, these developments have been combined in the Excited State Mean-Field (ESMF) wavefunction,[31,33–35,37,203] in which orbitals of a CIS wavefunction are variationally-relaxed for a select excited state. Early implementations of ESMF have accurately predicted excitations energies of CT[34,37,320] and core[304] states, and show promise as a general excited state reference, upon which perturbative corrections or other extensions could be applied. Strengthing the analogy to ground state mean-field theory, SCF-like approaches for calculating the ESMF wavefunction have been developed where a self-consistent commutator condition is used to enforce stationarity,[321] mirroring a commutator condition formulation of ground-state RHF theory.[36] The ESMF wavefunction itself can be fully optimized at $O(N^4)$ mean-field cost, at parity with standard CIS theory, and its perturbative extension ESMP2[320] scales as $O(N^5)$ and rivals the accuracy of more expensive methods such as EOM-CCSD that scales as $O(N^6)$.

However, these methods are presently at an early stage of development, and while pilot implementations have proven their favorable accuracy and cost-scaling, some vital numerical accelerations have yet to be implemented. For instance, in the ESMF-GVP method,[35] some derivatives necessary for orbital optimization are calculated through an automatic differentiation[37,322] framework. While this avoids intractable analytic expressions, intermediate orbital transformations like those commonly used in other orbital optimizations,[323] can greatly reduce the complexity of derivative expressions. In addition, owing to the large cost-scaling of Newton-Raphson optimization, approximate Quasi-Newton (QN) methods such as BFGS[57] are instead used to generate orbital steps. However, the step-quality and convergence rate of these optimization methods relies heavily[56,323] on obtaining an accurate approximation to the second derivatives of the orbital parameters, which is unfortunately missing from the present ESMF-GVP implementation. And, as we will show, the composure of linear and nonlinear optimization presents subtle but catastrophic convergence issues that afflict wavefunction optimizations like ESMF unless special care is taken to avoid them.

This chapter will be organized as follows. First, we discuss in more detail some known pitfalls present in orbital optimizations as they manifest in single-determinant Δ-SCF methods and in multi-determinant CASSCF methods. Second, we show how and why certain

optimization procedures can fail in well-established multideterminant algorithms such as CASSCF even when these careful precautions are taken, and identify some theoretical ways to recover algorithmic stability. Finally, we show that upon adding a tuned Hessian preconditioner to an existing ESMF implementation, it is accelerated to match the speed of the current ESMF-SCF implementation while being far more resilient against variational collapse.

4.2 Variational Collapse

Excited state orbital-optmization can require special considerations not necessary for ground states. For instance, state-specific approaches are prone to "variational collapse",[31,289] where they erroneously converge to the ground state instead of an excited state. However, in linear wavefunctions, like CIS or CASCI, excited states are higher roots of the Hamiltonian matrix, and are not prone to collapse in the same way. In these algorithms, excited states are naturally ordered from lowest to highest energy, and varying nonlinear orbital parameters can cause this ordering to change. As a result, identifying excited states purely by their energy-ordering in this Hamiltonian can result in a nonconvergent behavior known as root-flipping.[31]

While variational collapse and root flipping are well-known issues and a variety of solutions have been proposed for each,[33,35,203,288–291] there is an additional collapse behavior that remains unaddressed. Energy degeneracies between multiple states combined with strong coupling between linear CI parameters and non-linear orbital parameters can cause variational collapse and root-flipping behaviors to occur in tandem and avoid measures designed to prevent them. Around these points, the energy surface undergoes an avoided crossing, and uncoupled optimizers are liable to take grossly inaccurate steps, and can fatally destabilize state-selective algorithms. Though these avoided crossings are similar in spirit to those encountered in non-adiabatic dynamic – as they are described by energy degeneracies and derivatives of inter-state coupling elements – these crossings occur only in wavefunction parameter space and bear no physical meaning. We will discuss this in more mathematical detail in section 4.4 and provide a worked example in section 4.5, but we start by reviewing variational collapse and root-flipping, as many of the same concepts apply.

Due to particle antisymmetry, single-determinant wavefunctions are perhaps the simplest, but still physically-accurate, representation of a many-body electronic state. Electronic structure calculations routinely begin by calculating single-determinant ground states upon which corrections are applied to account for the correlative effects of electronic repulsion. It is thus natural to extend these minimal wavefunctions directly for excited states, as they conveniently obey the same formal stationarity conditions as ground-state wavefunctions. Excited-state Δ-SCF methods take advantage of this by identifying alternative solutions to the SCF equations as excited states, and are particularly convenient to implement and study since the working equations are closely related.

As the only variables present in a single-determinant wavefunction are orbital coefficients, excited states in Δ-SCF methods can be viewed as an orbital optimization. However, excited-state stationary points can often manifest as saddle-points on energy surfaces, and converging to them can be more complicated as saddle-points are maxima along some directions and minima along others. This ambiguous curvature can confuse optimization algorithms which

can propose unreasonable steps, ultimately causing failed convergence or convergence to the more-stable ground state minimum.

However, this variational collapse of single-determinant excited-state methods is well-understood and can be sidestepped with proper adjustment. In the Δ-SCF algorithm, initial guesses for the excited state are generated by applying excitation operators to a reference ground state. The same SCF procedures used for ground-states are performed on this initial excited state wavefunction are used to vary the orbital basis in search of a nearby energy stationary point. However, these excited states are often described by a non-Aufbau occupation pattern, in which some high-energy orbitals are occupied, and some low-energy orbitals are left unoccupied. After orbitals are varied, the next iteration's excited state is constructed by populating a subset of these orbitals. In the ground state, the population pattern is obvious, and the orbitals with the lowest orbital energies are chosen to be fully populated. As the nonlinear diagonalization step can switch or mix orbitals, populating excited states is not so straightforward, and small variations in the orbital basis can correspond to drastic or discontinuous changes in the excited state character.

Unpredictable changes in the orbital basis can easily destabilize the optimization and ultimately produce oscillatory behavior or, given the relative stability of the Aufbau-populated ground state minimum, variational collapse. Methods that endeavor to patch these instabilities often scrutinize the orbital selection process, seeking to limit population to orbitals most like those in the previous iteration. For instance, in the Maximum Overlap Method,[288,317] orbitals that maximally overlap those from a previous iteration are populated in the next, preventing discrete switches in orbitals regardless of their energy ordering. In the STEP method,[290] a level-shift is applied to the Fock matrix to heighten the energy of unoccupied orbitals, allowing the more straightforward Aufbau filling to be used at each iteration.

Taking a step back, the approach these SCF algorithms are pursuing boils down to limiting orbital step-size in each iteration. An alternative approach would be to instead produce orbital steps using an algorithm unable to produce the global steps that can come about by re-populating molecular orbital each iteration. In Geometric Direct Minimization[323] (GDM), the orbitals within a single-determinant wavefunction are varied instead by locally approximating the energy surface around the current wavefunction. As opposed to the SCF eigenvector equations which generate a full set of orbitals each iteration, GDM instead produces an orbital step δX that minimizes the second-order Taylor expansion of the energy by approximately solving the following linear equation:

$$H_X \cdot \delta X = \nabla_X E \tag{4.2}$$

Whose matrix H_X and vector $\nabla_X E$ coefficients are the energy Hessian and energy gradient, respectively. The GDM algorithm has been found to be more stable than conventional SCF algorithms for ground states, as its local steps cannot produce the discrete orbital swaps that can confound SCF convergence.

Algorithms like GDM tend to approximate Newton-Raphson steps due to its expense, since it involves building and inverting the Hessian second-derivative matrix. Quasi-Newton methods, such as the popular BFGS algorithm,[324,325] instead start with an approximation to the Hessian inverse and iteratively updates it using finite-difference gradient information obtained from successive iterations. Worthy of note is that NR optimization is able to

converge to nearby stationary points, whether they be minima, saddle points, or maxima, as it contains the necessary second-derivative information to account for negative curvature. Due to the approximate Hessians used in Quasi-Newton methods, they are not always able to comfortably converge to saddle points, and can struggle to find them consistently. In order to partially offset their approximations, Quasi-Newton methods are often implemented with line-searches to find optimal step-sizes that will minimize function values along the direction generated by equation 4.2. Unfortunately, these line-search minimizations are inappropriate for saddle-point finding, and is but one example of why adapting QN algorithms for saddle-points can be a headache.

As saddle-point searches are fraught with complexities, it is often simpler to edit the energy function so that its desired stationary points become minima. These modified functions are known as target functions, and their design often must sacrifice some computational efficiency to perform this transformation. For instance, both the wavefunction variance, used in σ-SCF:[291]

$$\sigma^2 [\Psi] = \langle \Psi | (\mathbf{H} - E)^2 | \Psi \rangle \tag{4.3}$$

and the Square Gradient Minimization (SGM) used in Δ-SCF[289] and ESMF-GVP:[33,37]

$$f [\Psi] = |\nabla E|^2 \tag{4.4}$$

are minimized by every eigenstate of the Hamiltonian. However, the optimization of these target functions requires evaluation of the square Hamiltonian or higher-order energy derivatives. Thankfully, the cost of evaluating these target functions and their derivatives can be respectively reduced by resolving the identity between the Hamiltonian product:[31,203]

$$\sigma^2 [\Psi] \approx \sum_I \langle \Psi | (H - E) | I \rangle \langle I | (H - E) | \Psi \rangle \tag{4.5}$$

or by evaluating derivatives by finite-difference[289] approaches:

$$\nabla |\nabla E(p)|^2 \approx \frac{\nabla E(\boldsymbol{p} + \delta \cdot \nabla E) - \nabla E(\boldsymbol{p} - \delta \cdot \nabla E)}{\delta} \tag{4.6}$$

with only a small loss in accuracy. And while these target functions can still suffer from poor numerical conditioning[289] or fail to fulfill formal properties like size-consistency or state-selectivity[33] if special care is not taken, they dramatically improve the stability of excited-state orbital optimization.

4.3 Root Flipping

The orbitally-relaxed single-determinant wavefunctions used in Δ-SCF can be hugely effective for capturing the important details of excited states at mean-field cost, but their utility is limited by the simplicity of single-determinant ansatz. DFT implementations of Δ-SCF benefit from density functionals that can account for wide swaths of weak correlation and have proven successful at recovering some charge-transfer states, core excitations, and double excitations.[326] While simple forms of spin-correlation common to many single excitations can be recovered using relatively simple spin-projection methods, extending this further for

higher-spin or correlated excited states can prove challenging. For instance, multiconfigurational DFT is complicated by possible double-counting of weak-correlation effects, since both the multideterminant expansion and the density functional can account for the same components of electronic correlation.[315] Multideterminant methods therefore remain a useful and solid tool for accurately capturing complex excited states.

Multideterminant wavefunctions are most often expressed as a linear combination of orthonormal determinants, and the energy stationary points of wavefunctions in this space can be straightforwardly obtained as eigenvectors of a projected Hamiltonian. Though finding a single excited state in isolation is fraught with the same saddle-point issues discussed previously, the linearity of the solutions to the eigenvector problem provides a stable route to preventing variational collapse. By the Courant-Fischer principle,[327] higher eigenvectors of a matrix can be defined as those that minimize the Rayleigh Quotient in the space orthogonal to all of the lower eigenvectors. Many classic diagonalization procedures, such as the QR algorithm, reflect this, and effectively perform extremal eigenvalue-finding algorithms and this orthogonalization in tandem. In essence, to find an excited state, one generates a stack of eigenvectors, each minimized in the space orthogonal to eigenvectors lower than it, that reach up to the desired state. This can be straightforwardly extended to iterative procedures designed for efficiently calculating only extremal roots of a large, sparse Hamiltonian such as the Davidson method[56] commonly used in CI theory.

Once these roots have been determined, they are naturally ordered from lowest to highest in energy, but this is only the first step in an orbitally-optimized method. From these roots, an excited state is identified – perhaps naïvely by its place in the energy ordering – and an orbital-dependent energy expression is constructed using its eigenvector.

$$E_k(X) = \sum U_{ki} U_{kj} \langle D_i(X) | \mathbf{H} | D_j(X) \rangle \tag{4.7}$$

This energy expression can then be minimized by using a NR or QN descent method until a stationary point with respect to the orbitals is reached. After each orbital step, the orbital basis is folded into the wavefunction according to Thouless' theorem,[328] which amounts to rotating the orbital coefficient matrix using a unitary orbital rotation operator:

$$\tilde{\boldsymbol{\chi}} = \boldsymbol{\chi} e^{\mathbf{X}} \tag{4.8}$$

These two steps are alternated until an orbital stationary point is reached, at which point the optimization is complete.

While this approach appears straightforward, like with orbital orderings in Δ-SCF methods, the state ordering can abruptly change upon orbital variation. As orbitals drift away from those suited for a lower-energy state, that state's energy will increase and may cross with the targeted root. As this crossing point is traversed, selecting our state purely by energy ordering means that we will start to optimize a discretely different state, as illustrated in figure 4.1. Often, the orbital gradients of each of these two states oppose each other, and orbital steps zig-zag across the energy degeneracy seam. This is a phenomenon known as root-flipping, so named because orbital steps cause the indexed root to rapidly flip between two states, never to converge.

To resolve root-flipping, we need only change our state-selection criteria from energy ordering to some other measure that captures the state character. In some cases, such as

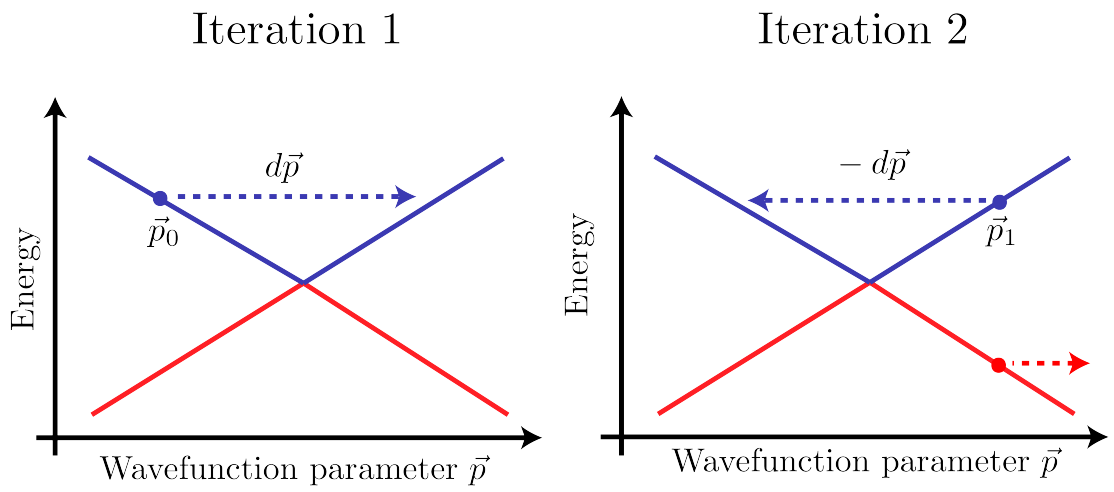

Figure 4.1: An Illustration of root flipping. In both plots, the upper curve (blue) and the lower curve (red) represent wavefunction energies as determined within the linear determinant space, ordered by their energies, as a function of a wavefunction parameter $\boldsymbol{p}$. At the crossing point, the character of the states associated with each state switches. To optimize an excited state, derivative information from the upper root is used to generate a parameter step. The wavefunction at each set of parameter values $(\boldsymbol{p}_0, \boldsymbol{p}_1)$ qualitatively changes character if the upper root is selected and the successive iteration will propose a parameter step contrary to the first, leading to oscillations about the crossing point (indicated by the blue arrows). Selecting the lower root instead continues optimization of a state with the consistent character (indicated by the red arrow), and is expected to converge.

when the two states that cross can be distinguished by their spin or symmetry, each state can be discretely identified on either side of the crossing. In the more general case, comparison might be done with observable state properties, but as many of these depend on the one-body density, density-based measures – similar in spirit to MOM's overlap metric – can be used instead, to similar effect. In a recent implementation,[31] this measure takes the form of a density difference norm, which measures each state density relative to a fixed reference (usually chosen as initial selected state):

$$M(\Gamma) = \|\Gamma_{ref} - \Gamma\|_F \tag{4.9}$$

Though these work well in many cases,[31,32] even these density-based measures fail when states mix their character. When this happens, selecting the right state can be an undoable task, as its components may already be spread amongst multiple roots.

4.4 Coupled Collapse

This strong state mixing in state-specific algorithms gives us direction for further investigation. While we could imagine simply taking some optimal mixture of states to minimize this metric, state-selection is not the only issue at hand. In regions where this coupling is significant and where mixing occurs, local expansions of the energy can misjudge the energy curvature and produce poor orbital steps. We can trace these errors back to equation 4.7 where linear coefficients and orbital parameters are assumed to be uncoupled, independent variables. The linear coefficients in U have an implicit dependence on the orbital basis that, owing to the complex dependence of eigenvectors on matrix elements, can be difficult to incorporate directly into the optimization algorithm. While there is a notable exception in modern CASSCF implementations[104] in which orbitals and linear coefficients are recoupled in order to accelerate convergence, this is so far limited to ground-state and state-averaged calculations.

The orbital dependence of eigenvectors is implicit and difficult to analyze directly, and so we will look at how the off-diagonal Hamiltonian coupling elements affect our local estimation of the energy surface. While off-diagonal coupling elements in the linear Hamiltonian are zero by definition within each orbital basis, rapid changes in these coupling elements can profoundly affect the curvature of the energy surface. To demonstrate how coupling can affect local estimates of the energy surface, we have projected the Hamiltonian onto two states for which we assume this coupling is significant.

$$H(X) = \begin{bmatrix} H_{11}(X) & H_{12}(X) \\ H_{21}(X) & H_{22}(X) \end{bmatrix}, \qquad H_{ij} = \sum_{kl} U_{ik} U_{jl} \langle D_i(X) | \mathbf{H} | D_j(X) \rangle \tag{4.10}$$

Because we are always working with eigenstates of the Hamiltonian in our current orbital basis, the coupling element H_{12} at our expansion point is zero, but its derivatives may not be. The projected 2×2 Hamiltonian can be diagonalized analytically, giving energy expressions for the upper (E_+) and lower (E_-) roots:

$$E_{av} = \frac{H_{11} + H_{22}}{2}, \qquad \Delta E = H_{22} - H_{11} \tag{4.11}$$

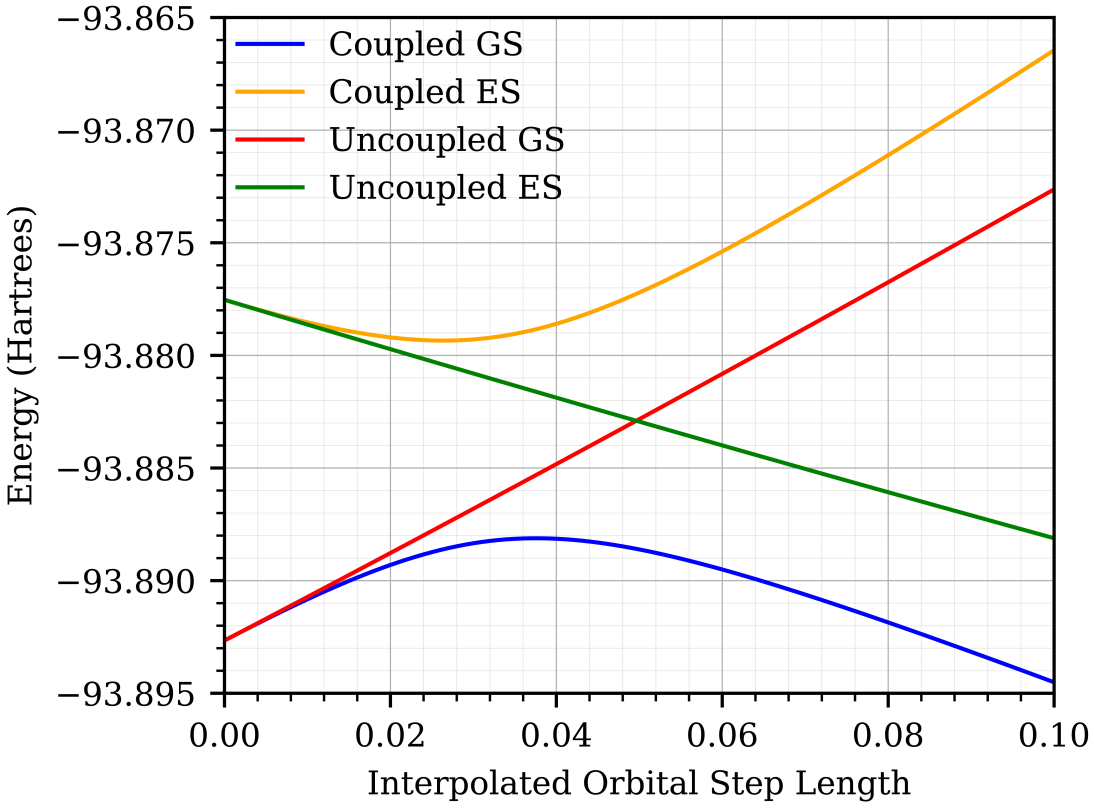

Figure 4.2: Coupled CASCI energies and uncoupled CASCI energies of ground and excited states of the methanimine molecule at interpolated orbital rotations. Coupled energies are obtained by diagonalizing the linear CASCI Hamiltonian within this basis, and correspond to the real energy surfaces that are being explored. Uncoupled energies are obtained by calculating expectation values using the CASCI vectors that diagonalize the Hamiltonian at $X = 0$, and corresponds to the energy surfaces as they are approximated by local expansion of equation 4.7.

$$E_{\pm} = E_{av} \pm \frac{\sqrt{\Delta E^2 + 4H_{12}^2}}{2} \tag{4.12}$$

The dependence of $E_{\pm}$ on the coupling element introduces an additional term into the Hessian, and differs from those derivatives of equation 4.7 by a coupling square-gradient term:

$$\frac{\partial^2 E_{\pm}}{\partial X^2} = \left[\frac{\partial^2 E_{av}}{\partial X^2} \pm \frac{1}{2}\frac{\partial^2 \Delta E}{\partial X^2}\right] + \frac{2}{\Delta E}\left(\frac{\partial H_{12}}{\partial X}\right)^2 \tag{4.13}$$

While one might rightly believe that this derivative coupling is negligible in many circumstances, we have found multiple simple equilibrium systems where omitting it can cause variational collapse. Figure 4.2 shows an example of this phenomenon in methanimine (discussed in the next section), where a poor estimate of energy curvature leads the optimizer to take poor orbital steps. In this plot, orbitals are rotated at each point along a linear-interpolation of the orbital step vector, and two sets of energy calculations are performed. As is plainly evident, the proposed orbital step at $X = 0.1$ drastically overshoots the coupled minimum at $X \approx 0.02$, and does so because it follows the vastly shallower curvature of the uncoupled energy curves. As will be discussed in the next section in more detail, the fate of these optimizations was often variational collapse to the ground state, as states would mix their character. This ultimately confuses the state-selection metric, which – after several poor orbital steps – selects the lower CASCI root, and promptly collapses to the ground state.

4.5 Methanimine

Methanimine, CH_2NH, is the smallest example of the imine functional group (fig. 4.3) and contains a total of sixteen electrons. The lone pair on the nitrogen is found in an sp^2 orbital perpendicular to the double bond, the first singlet excited state can be summarized as an $n \rightarrow \pi^*$ transition, and is well-represented as an open-shell singlet. While the ground state is a closed shell singlet, the orbital basis necessary to represent it as such is radically different from the optimal excited state orbital basis. As a result, when optimizing orbitals for the excited state, unless a sufficiently large active space is used, the ground state will increase in energy until the two states experience an avoided crossing in parameter space. As shown in the previous section, around this avoided crossing, the optimizer incorrectly underestimates the curvature of the energy, and makes erratic orbital steps before collapsing to the ground state.

Though this chapter will later focus on similar collapse behavior in the context of the ESMF wavefunction, state-selective CASSCF is used here because the span of complete active spaces are conveniently invariant to orbital rotations. With a large enough active space, any mismatch in ground state and excited state orbitals can be reversed within the linear configuration space. And while expanding the active space is instructive because it allows us to isolate the ground state orbitals disrupted by the state-specific optimization, it also points to a larger concern in state-specific orbital optimization in CASCI spaces.

In order to avoid both the complexities of state-specific optimization while still partially enjoying the benefits of orbital relaxation, active-space orbitals are often optimized with respect to a weighted average of multiple state energies.[30] Because multiple states feature

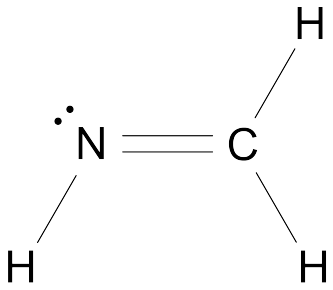

Figure 4.3: Methanimine

in this average, the orbitals are not optimal for any one state, and are instead thought to represent a reasonable compromise between states. Part of the appeal of state-specific optimization in active space methods is that one expects, contrary to state-averaged calculations, that active space orbitals can be tuned for a single excited state, and correlation present in other states need not be represented at all. In addition, state-specific orbitals generally reduce the size of active space necessary to represent a multireference state at a given level of accuracy. However, despite our state-specific approach, a larger active space appears necessary to converge a state-specific calculation which should only require a minimal one. This expanded active space is necessary to accurately represent the ground state and excited state at the same time and so prevent the two states from crossing as the optimizer approaches optimal excited state orbitals.

Here we show that in the case of methanimine, where the ground state is a closed-shell singlet (that can be represented with a single determinant) and the first excited is a simple open-shell singlet (that can be represented in a 2e-2o active space), that an uncoupled excited-state-specific algorithm will variationally collapse if anything less than a 6e-4o active space is used. Because the ground state is poorly represented in orbitals tuned for the excited state, unless an overly large active space is present to allow the linear space to reverse the orbital rotation, it steeply increases in energy before finally mixing with the excited state. This collapse is caused by a mis-estimation of energy surface curvature due to missing coupling terms between orbital parameters and CI coefficients as described in section 4.4 and depicted in figure 4.2. The state-selective metrics in place are powerless in the face of avoided crossings, as states can freely mix in the coupling region, and do nothing to prevent this collapse.

The natural orbital occupations of a reference 6e-5o CASCI wavefunction, given in table 4.1, confirms that the ground and excited states of methanimine can be – in their natural orbital bases – respectively represented as a single configuration or within a 2e-2o active space. However, active space orbitals in figure 4.8 indicate that these natural orbital bases optimal for each of these states are vastly different from each other. When – after optimizing orbitals for the excited state in the 6e-5o large active space – reducing the active space beyond 6e-4o, the energy of the CASCI ground state (table 4.2) increases by a massive 146 mE_h . At

Active space and State	Orbital 6	Orbital 7	Orbital 8	Orbital 9
6e-5o ground state	1.994	1.991	1.904	0.098
6e-5o excited state	1.990	1.978	1.010	1.009

Table 4.1: Natural orbital occupation numbers of the reference 6e-5o SA-CASSCF wavefunctions. The near-unit values of the ground-state natural orbitals occupations indicate that it is close to a single reference wavefunction (requiring no active space to qualitatively capture). The excited state natural orbital occupations are those of an open-shell singlet, suggesting it can be captured using a 2e-2o active space.

Active Space	Ground State Energy / E_h	Excited State Energy / E_h
6e-5o	-94.0715	-93.8916
6e-4o	-94.0582	-93.8825
4e-3o	-93.9119	-93.8776
2e-2o	-93.9115	-93.8764

Table 4.2: Energies of CASCI wavefunctions in a variety of active spaces using excited-state natural orbitals optimized specifically for the excited state in the 6e-5o active space. While ground-state energies sharply increase by 146 mE_h upon shrinking the active space from 6e-4o to 4e-3o, excited state energies are largely unaffected, indicating that the excited-state natural orbitals are unable to capture the ground state in less than a 6e-4o active space.

Active space	6e-5o	6e-4o	4e-3o	2e-2o
Ground state	0.976	0.978	0.866	0.867
Excited state	1.000	0.996	0.990	0.991

Table 4.3: Overlap of CASCI wavefunctions with the corresponding reference 6e-5o CASCI wavefunction. Each wavefunction uses state-specific orbitals optimized for the excited state in the 6e-5o active space. Following table 4.2, there is a sharp decrease in ground-state character of the lower root upon decreasing the size of the active space from 6e-4o to 4e-3o, while the excited-state character of the upper root is largely unaffected.

Active space change	C δq	N δq
6e-4o $\rightarrow$ 4e-3o ground state	0.725	-0.715
6e-4o $\rightarrow$ 4e-3o excited state	0.009	0.013
6e-4o $\rightarrow$ 2e-2o ground state	0.720	-0.705
6e-4o $\rightarrow$ 2e-2o excited state	0.013	-0.008

Table 4.4: Charge differences (in a.u.) on heavy atoms upon reducing the active space size as indicated. Charge differences are nearly unit for the ground state and negligible for the ground state, following the results from tables 4.2 and 4.3.

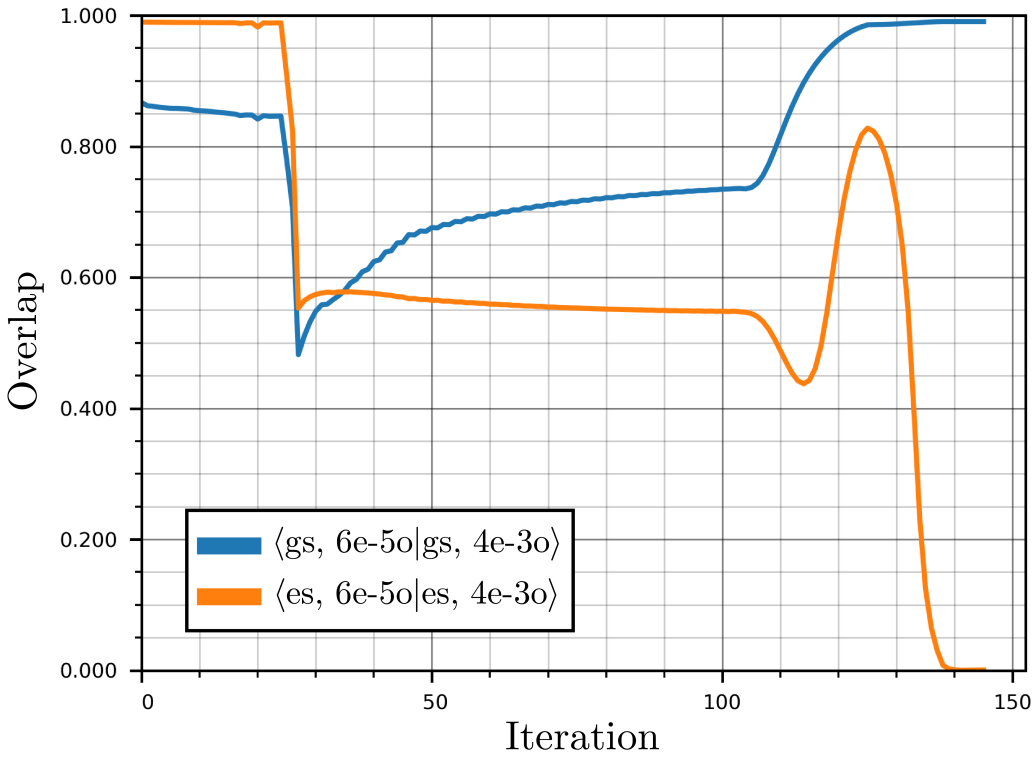

Figure 4.4: Overlap of each 4e-3o CASCI root with the corresponding 6e-5o reference root (i.e. overlap of lower root (4e-3o gs) with the reference lower root (6e-5o gs) and upper root (4e-3o es) with upper root (6e-5o es), as indicated in the legend) tracked across state-specific orbital optimizer iterations. The state-specific tracker initially targets the upper root (orange). Around iteration 25, the overlap of each of these roots with the corresponding reference root sharply decreases, coincident with a sharp increase in the overlap of each root with the opposite reference root shown in figure 4.5. After this point, the state-specific tracker tracks the lower root (blue), which slowly and consistently increases in reference lower-root (i.e. ground-state) character until convergence, completing variational collapse.

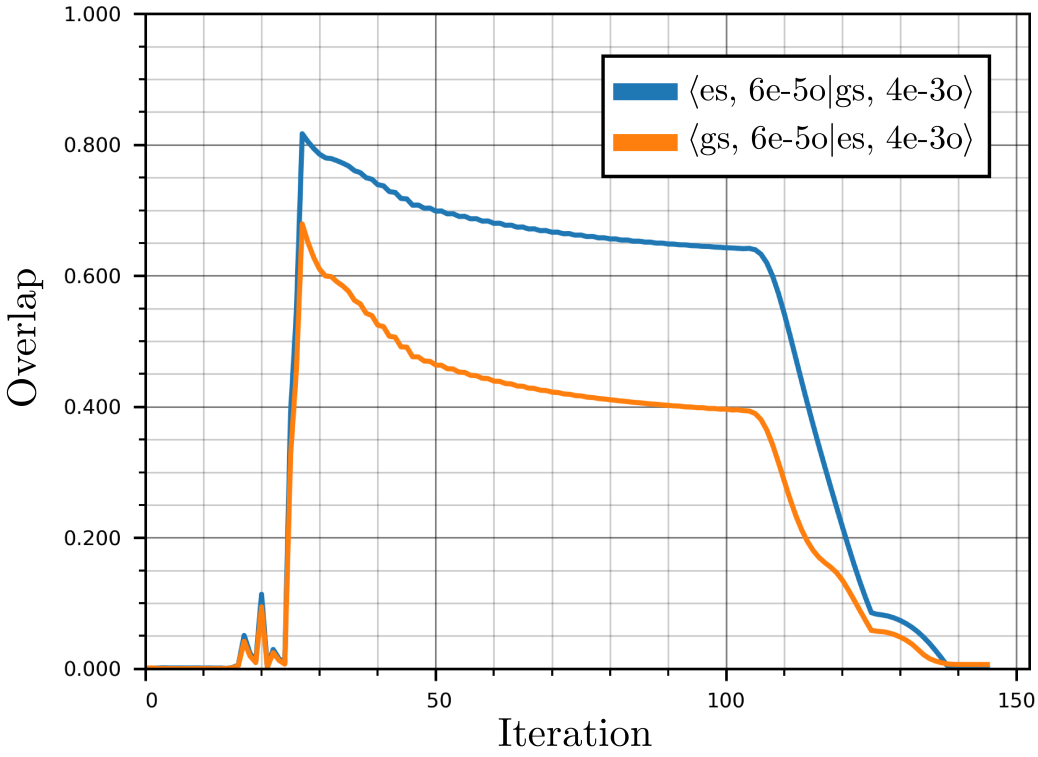

Figure 4.5: Overlap of each 4e-3o CASCI solution with the opposite 6e-5o reference root (i.e. overlap of the lower root (4e-3o gs) with the reference upper root (6e-5o es) and overlap of the upper root (4e-3o es) with the reference lower root (6e-5o gs), as indicated in the legend) tracked across state-specific orbital optimizer iterations. In the first 20 iterations, there is little overlap between roots, indicating that there is little mixing in the character of the lower two CASCI roots. Around iteration 25, there is a sharp spike in opposite-root overlap, coincident with the sharp decrease of same-root overlap in figure 4.4, indicating that these two states are strongly mixing. This overlap steadily decreases after the optimizer begins to optimize the lower root, which eventually collapses to the ground state.

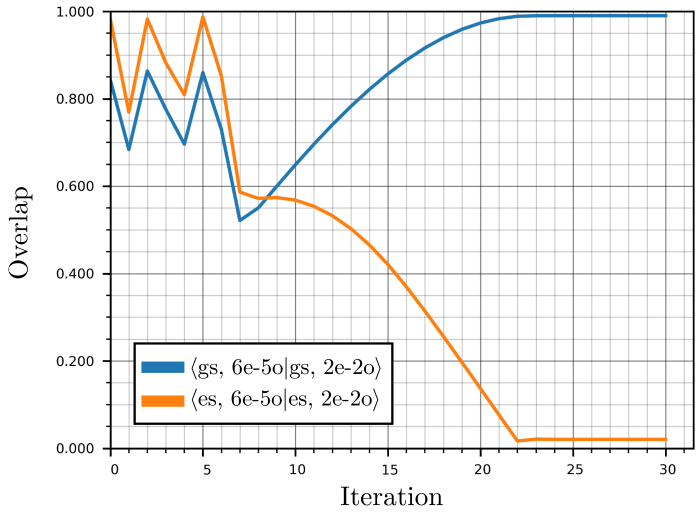

Figure 4.6: As figure 4.4, but in the 2e-2o active space. Behavior is qualitatively similar.

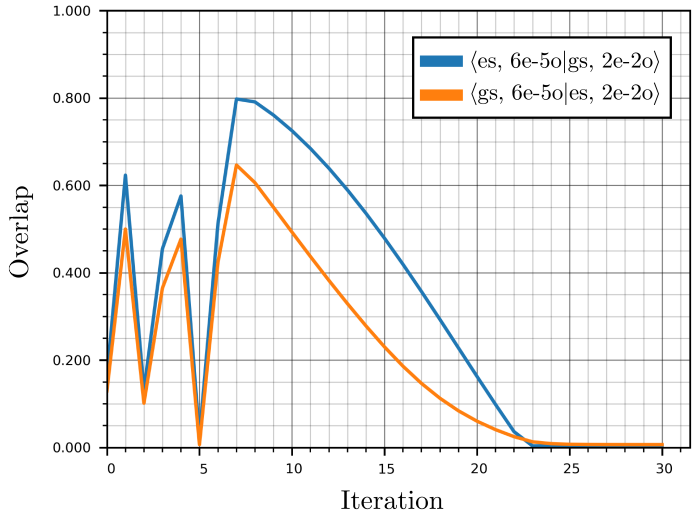

Figure 4.7: As figure 4.5, but in the 2e-2o active space. Behavior is qualitatively similar.

the same time, nearly an entire electron (table 4.4) shifts from carbon to nitrogen, and the overlap with the reference ground state (table 4.3) is cut by nearly a fifth. All of this is to show that the ground state wavefunction is poorly represented in the optimal excited-state basis.

Optimizer convergence in these smaller active spaces is chaotic, and very quickly start to mix ground and excited states before undergoing variational collapse. Over this optimization, we track the overlap (figures 4.4, 4.6) of CASCI roots with the 6e-5o CASCI reference states. A healthy optimization would show a steady, if small, increase in upper-root overlap with the reference excited state, and ultimately converge to a value near one. Lower-root overlap with the reference ground state might decline as the orbital basis drifted away from ground state orbitals, as the state-specific optimizer is free to ignore its affect on unselected roots. Instead, however, we see a sharp decrease (figure 4.4) or large oscillations (figure 4.6) in the overlap of *both* roots with their corresponding reference states. This decrease in overlap is matched by a coincident increase in the overlap with the opposite reference state (figures 4.5, 4.7), indicating large levels of mixing between ground and excited states.

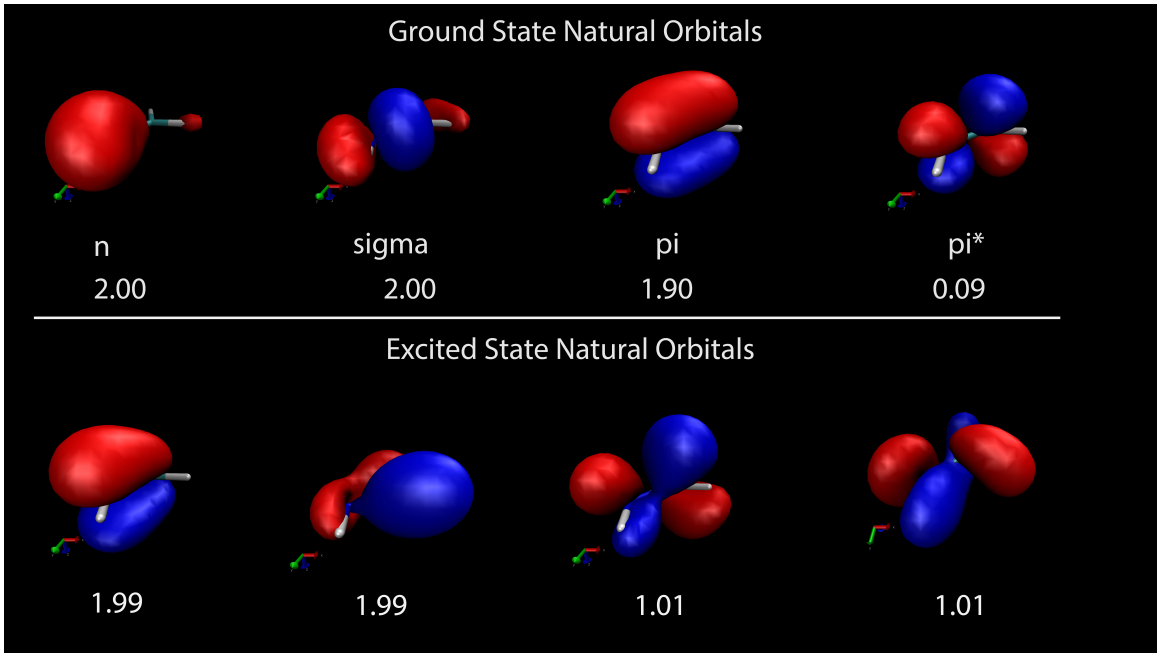

Figure 4.8: Methanimine state-specific natural orbitals in the 6e-4o active space. Natural orbital occupation is indicated by the number adjacent each orbital isosurface. While the span of these orbitals is very nearly the same, the difference in orbital shapes makes it difficult to represent one state using the natural orbital basis derived from the other state. When the active space is reduced to 4e-3o or smaller, the wavefunction has no ability to transform the basis between these two sets, and only one of the two states can be accurately represented.

Following the fall of both overlap measures, figures 4.4 and 4.6 show a steady increase in the overlap of the ground state character of the lower root, and all figures show a steady decrease of the excited state character of both roots. This change in behaviors corresponds to the state-selector refocusing on the lower root, dutifully minimizing it toward the ground

state and completing the variational collapse. These figures also indicate that, while the optimizer follows the lower root, the quality of the upper root degrades (with temporary exception in figure 4.4) until it has negligible overlap with both reference states. Just as the optimal excited state orbitals are incompatible with the ground state, the excited state is so poorly represented with the ground state orbitals that any component of the first excited state is banished to higher roots.

While it would be satisfying to close this section showing that a coupled, state-specific CASSCF algorithm converges correctly to the first excited state in a minimal 2e-2o active space, this is unfortunately outside the scope of this study. In the following sections we introduce the ESMF wavefunction alongside coupled and uncoupled optimizer implementations that exhibit all the same signs of this collapse behavior, but in a less obvious way. Unfortunately for our present purposes, owing to the different construction of the ESMF wavefunction, the first excited state of methanimine does not variationally collapse in either coupled or uncoupled ESMF optimizer implementations. Displaying a converged result from a coupled optimizer would not be meaningful; it simply does not fix this collapse problem in methanimine, because it wasn't present in the first place. Instead, we will investigate other molecules in which this type of collapse is present.

4.6 Excited State Mean Field

Following the discussion in the introduction, the Excited State Mean Field wavefunction is one that incorporates orbital relaxation effects into the CIS wavefunction. As orbital relaxation describes the many-body electronic response to the charge movement in electronic excitations, it is crucial to accurately depict the large-scale shifts in charge-transfer and core excitations. Multiple approaches to optimizing the ESMF wavefunction have been implemented, each with their own benefits and idiosyncracies. In the Generalized Variational Principle (GVP)[35,203] approach, a target function is used to guide the ESMF wavefunction to an excited state with selected energies and state properties. The GVP optimizer couples linear CIS parameters and nonlinear orbital parameters and treats them on the same footing, contrary to standard two-step approaches which alternatively perform independent linear optimizations and nonlinear orbital steps. By contrast, much like ground-state SCF theory, the ESMF-SCF implementation is based on a Lagrangian with orbital orthonormality constraints whose stationary point is found by iteratively and self-consistently fulfilling a commutator condition.[36]

The full ESMF wavefunction is written as:

$$|\Psi_{ESMF}\rangle = e^{\mathbf{X}} \left[c_0 |\psi\rangle + \sum_{ia} c_{ia} \mathbf{a}^\dagger \mathbf{i} |\psi\rangle \right] \tag{4.14}$$

where ψ is a closed-shell singlet reference state, $e^{\mathbf{X}}$ is an orbital rotation operator, and c_{ia} are linear CIS coefficients. For theoretical convenience, excitations are often constrained to singlet-pairs and the reference configuration is omitted, following the common observation that its contribution is negligible:

$$|\Psi_{ESMF}\rangle = e^{\mathbf{X}} \left[\sum_{ia} c_{ia} \left(\mathbf{a}_\uparrow^\dagger \mathbf{i}_\uparrow + \mathbf{a}_\downarrow^\dagger \mathbf{i}_\uparrow |\psi\rangle \right) \right] \tag{4.15}$$

Excited states are identified as stationary points of the energy with respect both to the linear CIS parameters c_{ia} and nonlinear orbital parameters x_{ia}. Following the logic in section 4.2, to prevent variational collapse, the stationary points of an energy-based target function are sought:

$$f(E) = \chi \left[(1-\mu)(\omega - E)^2 + \mu|\nabla E|^2\right] + (1-\chi)E \tag{4.16}$$

The different components of the GVP in equation 4.16 are each meant to serve different purposes when searching for a particular excited state. By varying the values of χ and μ in different regimes of optimization, each of the terms' unique strengths can be invoked when most needed. First we note that, regardless of the values of χ and μ, each component of the GVP – and therefore the entire GVP itself – shares stationary points with the bare energy function. Initially, the value of μ is set to a value of 0.5, χ to a value of 1 , and the value of ω is set to a value near the expected energy of the targeted excited state. The first term measures the energy-distance from this expected value, and its optimization biases the wavefunction toward that energy. As previous discussed, the second term converts energy saddle points to minima, and will push the optimization toward the closest excited state while avoiding variational collapse to lower-energy states. As optimization continues, the first term is assumed to have already done its job in locating the correct minimum basin, and μ is progressively decreased until it hits a final value of $\mu = 0$. Near the end of the optimization, χ may be switched off to a value of zero and the optimizer switched to a full NR optimizer in order to accelerate convergence.

In order to reduce the amount of moving parts in our target function, for our purposes we focus solely on the central square gradient magnitude component (i.e. $\mu = 1, \chi = 1$). Importantly, in the GVP implementation, optimization proceeds by a one-step procedure, in which a QN algorithm generates steps for linear and non-linear parameters simultaneously. While forgoing the independent linear optimization of CIS parameters means that the analysis in section 4.4 does not apply, parameter coupling can nonetheless be found in off-diagonal blocks of the Hessian. Empirically, table 4.5 shows that this coupled QN approach is effective at preventing collapse to ground-state-like wavefunctions.

The other implementation variant is known as ESMF-SCF, so named because it is based on a commutator condition that mirrors that of ground-state mean-field theory.[321] Inspired by single-determinant SCF methods, the ESMF-SCF[36] algorithm writes excited states as stationary points of a Lagrangian:

$$\mathcal{L} = E_{ESMF} + \mathrm{tr}\left[\left(\boldsymbol{I} - \boldsymbol{C}^T\boldsymbol{S}\boldsymbol{C}\ \right)\boldsymbol{\epsilon}\right] \tag{4.17}$$

whose second term enforces orthonormality of the molecular orbital basis using the Lagrange multipliers in $\boldsymbol{\epsilon}$. After differentiating, we set the gradient to zero to enforce the energy stationarity condition. This can be rearranged to give an equivalent commutator condition for pairs of density matrices $\boldsymbol{\Gamma}^{(i)}$ and Fock-like matrices $\boldsymbol{F}^{(i)}$:

$$\sum_i \left[\boldsymbol{C}^T\boldsymbol{F}^{(i)}\boldsymbol{C}, \boldsymbol{\Gamma}^{(i)}\right] = 0 \tag{4.18}$$

Which is a direct analogy to the commutator condition that may be used in RHF-SCF optimization:

$$\left[\boldsymbol{C}^T\boldsymbol{F}\boldsymbol{C}, \boldsymbol{\Gamma}\right] = 0 \tag{4.19}$$

The ESMF commutator condition in equation (4.18) can be written as a function of the orbital basis, linearized, recast as a linear equation, and numerically solved to generate an orbital step. The linear parameters c_{ia} are then optimized by independently solving the standard CIS equations, after which the process is repeated until the commutator condition is met.

As discussed in section 4.4, derivative couplings between linear and nonlinear parameters can be vitally necessary when the excited state stationary point is in the vicinity of a crossing region. In the ESMF-GVP implementation referenced above, linear CIS coefficients and nonlinear orbital parameters are optimized simultaneously in a coupled QN optimization, and so include this coupling information in off-diagonal blocks of the Hessian matrix. In the ESMF-SCF implementation, though both the $\mathbf{\Gamma}^{(i)}$ and $\boldsymbol{F}^{(i)}$ matrices depend on orbital parameters, this dependence is not present in the orbital step generation. As coupling between orbitals and CIS coefficients increases, ignoring this dependence risks variational collapse, which we will show several examples of below.

4.7 ESMF-SCF Collapse

The methanimine example in section 4.5 shows how unexpectedly small molecules can undergo variational collapse in state-targeted CASSCF algorithms even when careful precautions are taken. We will see that a form of variational collapse also affects the ESMF-SCF implementation in several small molecular examples. While the relative inscrutability of the ESMF-SCF commutator condition makes its orbital steps difficult to analyze, its linearization at least ensures that orbital steps are performed in a local way, and avoid the orbital population issues endemic to Δ-SCF algorithms. The coupled-collapse mechanism that methanimine follows can potentially arise in any multideterminant method, and the ESMF-SCF method is no exception, given its missing parameter coupling. However, stationary points in the ESMF-SCF energy expression are still saddle points, and the lack of a square-gradient target function or similar may also be partially responsible for the variational collapse that we observe.

Because closed-shell determinants are excluded in present ESMF wavefunctions, the endpoint of collapse is not the closed-shell ground state and collapse is thus more difficult to detect. However, there remain multiple signs that we can use to identify it, and with a particular orbital rotation, it is possible for such an open-shell wavefunction to recover half of the ground state. Consider a pair of orbitals, $\{\chi, \chi^*\}$ that correspond to bonding and anti-bonding configurations between orbitals $\{L, R\}$ on two molecular fragments.

$$\chi^\dagger = \frac{1}{\sqrt{2}}\left(\mathbf{L}^\dagger + \mathbf{R}^\dagger\right), \qquad \chi^{*\dagger} = \frac{1}{\sqrt{2}}\left(\mathbf{L}^\dagger - \mathbf{R}^\dagger\right) \tag{4.20}$$

and assume that the ground state wavefunction has occupied the two bonding configurations in a closed shell configuration:

$$|\psi_{gs}\rangle = \chi_\uparrow^\dagger \chi_\downarrow^\dagger |\psi\rangle \tag{4.21}$$

Equivalently, we can write out this ground state in the fragment basis, in which case:

$$\chi_\uparrow^\dagger \chi_\downarrow^\dagger |\psi\rangle = \frac{1}{2}\left[\mathbf{L}_\uparrow^\dagger \mathbf{L}_\downarrow^\dagger + \mathbf{R}_\uparrow^\dagger \mathbf{R}_\downarrow^\dagger + \mathbf{L}_\uparrow^\dagger \mathbf{R}_\downarrow^\dagger + \mathbf{R}_\uparrow^\dagger \mathbf{L}_\downarrow^\dagger\right] |\psi\rangle \tag{4.22}$$

By design, the ESMF wavefunction has the freedom to perform orbital rotations and single excitations, and it can use this freedom to rotate the $\{\chi, \chi^*\}$ basis to the $\{L, R\}$ basis:

$$e^{\mathbf{X}} \left[\chi_\uparrow \chi_\downarrow |\psi\rangle\right] = \left[\mathbf{L}_\uparrow^\dagger \mathbf{L}_\downarrow^\dagger\right] |\psi\rangle \tag{4.23}$$

Following this, a singlet excitation from the L to R orbital results in a wavefunction that significantly overlaps with the ground state:

$$|\phi\rangle = \frac{1}{\sqrt{2}} \left[\mathbf{R}_\uparrow^\dagger \mathbf{L}_\uparrow + \mathbf{R}_\downarrow^\dagger \mathbf{L}_\downarrow\right] \left[\mathbf{L}_\uparrow^\dagger \mathbf{L}_\downarrow^\dagger\right] |\psi\rangle = \frac{1}{\sqrt{2}} \left[\mathbf{R}_\uparrow^\dagger \mathbf{L}_\downarrow^\dagger + \mathbf{L}_\uparrow^\dagger \mathbf{R}_\downarrow^\dagger\right] |\psi\rangle \tag{4.24}$$

This wavefunction shares two configurations from the ground state wavefunction in equation 4.22, and so significantly overlaps:

$$\langle \phi | \psi_{gs} \rangle = \frac{1}{\sqrt{2}} \tag{4.25}$$

The significant ground state character of this open-shell state can often place this state lower in energy than any excited state.

We should briefly note here that the exclusion of the closed-shell determinant from this space makes analysis along the lines of section 4.4 less straightforward. Following the Brillouin condition, one can consider the RHF wavefunction a stationary point of the linear CIS equations. As our ESMF wavefunction begins as a standard CIS state, the Aufbau state plays the role of ground state, and the coupled-collapse analysis would anticipate the these two states crossing in energy and mixing components before variational collapse. While this can no longer occur, we have shown that a partial representation of the ground state may lurk in the linear space in particular orbital bases. This still admits the possibility of this partial ground state plummeting in energy and crossing with the targeted root as the optimizer approaches this particular basis, but we cannot be sure that this is occurring without more careful analysis.

It is still significant to say that target functions and coupling greatly improves the stability of ESMF optimization. Regardless of how collapse occurs, the orbital rotation necessary to converge to this partial ground state wavefunction is stark, and very likely to disrupt the energy of the Aufbau determinant. As mentioned, the Aufbau configuration begins as the RHF configuration, and in all cases shown, is lower in energy than all of its single excitations. In cases where collapse has occured, we often observe an inversion in the relative energies between the final Aufbau configuration and the collapsed, partial ground-state wavefunction.

To summarize, the following are signs of variational collapse in ESMF-SCF:

i) Unrealistically small excitation energies.

ii) Significant overlap between the ESMF wavefunction and the RHF ground state.

iii) Aufbau configuration energy higher than the ESMF wavefunction energy.

ESMF-SCF wavefunctions of molecules shown in table 4.5 meet all of these conditions, while ESMF-GVP wavefunctions meet none. While the GVP optimizer consistently avoids variational collapse, there are multiple cases where the final wavefunction nonetheless attains nontrivial overlap with the RHF ground state. In these cases, convergence was often considerably slower than in molecules of similar size, perhaps reflecting particularly strong coupling of wavefunction parameters or of partial double excitation character.

	Excitation Energy / mE_h				RHF Overlap	
molecule	CIS[a]	L-BFGS [b]	ESMF SCF[c]	Aufbau to SCF[d]	L-BFGS	ESMF-SCF
cyclopropene	253.3	247.8	125.5	-235.8	0.024	0.697
furan	247.2	240.1	129.9	-241.5	0.000	0.671
pyridine	229.6	224.6	153.2	-217.7	0.000	0.649
aniline	211.4	201.6	146.3	-189.6	0.000	0.603
furfural	208.2	189.4	111.5	-213.2	0.144	0.662
benzofuran	209.2	203.3	130.6	-219.4	0.103	0.606
indene	198.1	195.2	113.0	-205.3	0.006	0.684
indole	205.0	194.0	136.9	-190.8	0.103	0.650
chromone	208.2	201.4	134.6	-206.7	0.073	0.596
coumarin	182.3	173.2	105.0	-129.3	0.023	0.684
quinone	195.6	167.9	140.0	-138.3	0.134	0.623
biphenyl	132.7	130.5	66.4	-164.8	0.002	0.673

Table 4.5: [a]Excitation energy from the RHF to CIS wavefunction.
[b]From RHF to the L-BFGS-optimized ESMF wavefunction (see next section).
[c]From RHF to the ESMF-SCF wavefunction.
[d]From the Aufbau determinant to the ESMF-SCF wavefunction.

4.8 Quasi-Newton Preconditioning

While stable, the current ESMF-GVP implementation has several numerical drawbacks. BFGS, the QN method used in ESMF-GVP, iteratively constructs an inverse Hessian by numerically updating a initial inverse Hessian guess using finite difference information calculated between gradients in successive iterations. While step quality in BFGS relies heavily on a good approximation to the initial Hessian inverse – which here we will refer to as a "preconditioner" – the quality of this preconditioner must be balanced against its cost. In the following section, we explore several approximate Hessian preconditioners and measure their quality by cost-scaling, convergence speed, and wall-time in order to find such an optimal balance. However, before we do this, we have several formal approximations that we will later numerically justify.

First and second derivatives of the square-gradient magnitude target function used to construct the preconditioner can be obtained at the same cost-scaling as first and second derivatives of the energy function, with the exception of a contraction between the third-derivative tensor and the gradient. Following finite-difference approaches in the literature,[289] and to ease computational cost, this term can be approximated as the finite difference of Hessian matrices in the direction of the gradient:

$$\sum_k \frac{\partial^3 E}{\partial p_i \partial p_j \partial p_k} \cdot \nabla E_k \approx \lim_{\delta \to 0} \frac{\boldsymbol{H}(\boldsymbol{p} + \delta \nabla E) - \boldsymbol{H}(\boldsymbol{p} - \delta \nabla E)}{2\delta} \tag{4.26}$$

Theoretically, contraction with the energy gradient means that this term becomes vanishingly small as one approaches a stationary point. Empirically, dropping this term had little apparent adverse effect in exact Newton-Raphson optimizers, and to avoid both additional

compute-time and finite-difference error, this term was entirely omitted from future calculations. Though existing SGM implementations use finite difference directional derivatives[289] to avoid matrix-vector operations by the energy Hessian, our implementation performs these Hessian operations analytically at a lower total cost (8 TEI tensor contractions) than the finite-difference equivalent (9 TEI contractions).

Existing ESMF-GVP code relies on differentiating through the exponential orbital rotation operator $e^{\mathbf{X}}$ which formally involves evaluating an infinite series of product-rule terms:[329]

$$\frac{\partial}{\partial x_{ia}} e^{\mathbf{X}} = \sum_{n=0}^{\infty} \sum_{k=0}^{n} \frac{1}{(n+1)!} \mathbf{X}^k \frac{\partial \mathbf{X}}{\partial x_{ia}} \mathbf{X}^{k-n} \tag{4.27}$$

These derivatives are evaluated numerically by truncating the above expression to twelfth order. By instead applying the orbital rotation operator each iteration, the value of orbital rotation parameters are reset to zero so that all but the first term of this expansion vanish, greatly simplifying evaluation. However, in order to retain the same wavefunction before and after the transformation, the orbital rotation should be counter-balanced by an opposing rotation in the linear CI space to satisfy the identity:

$$\sum_i c_i \left|D_i\right\rangle = \sum_i \tilde{c}_i \left|\tilde{D}_i\right\rangle = \sum_i \tilde{c}_i e^{\mathbf{X}} \left|D_i\right\rangle \tag{4.28}$$

As this is a change of variables, derivatives are subject to Jacobian transformations, for instance:

$$\frac{\partial^2 E}{\partial \tilde{c}_i \partial \tilde{c}_j} = \sum_{kl} \frac{\partial c_k}{\partial \tilde{c}_i} \frac{\partial^2 E}{\partial c_k c_l} \frac{\partial c_l}{\partial \tilde{c}_j} = \sum_{kl} J_{ik} \frac{\partial^2 E}{\partial c_k \partial c_l} J_{lj}, \qquad J_{ik} = \frac{\partial c_k}{\partial \tilde{c}_i} \tag{4.29}$$

Whose elements can be obtained by projecting and differentiating equation 4.28:

$$\frac{\partial c_k}{\partial \tilde{c}_i} = \langle D_k | e^{\mathbf{X}} | D_i \rangle \tag{4.30}$$

Due to the form of the ESMF wavefunction, in order to evaluate equations 4.28 and 4.30 for the ESMF wavefunction, they must be projected onto the singles excitation space.

In practice, however, pilot implementations of both variable and Jacobian transformations did little to affect convergence behavior, and were accordingly omitted. Much like the omission of the third-derivative term, we can rationalize dropping these Jacobians by arguing that, in regions where large orbital steps are taken — where these Jacobian transformations would be significant — the accuracy of the approximate Quasi-Newton Hessian inverse is limited primarily by the accuracy of the BFGS update. Conversely, in regions where small orbital steps are taken, these Jacobian transformations are small.

4.9 Hessian Preconditioning

The analytical energy Hessian was implemented to serve both as a reference and a launching point for further approximation. In the following optimizations, derivatives of the square-gradient ESMF target function was were calculated as:

$$\nabla |\nabla E|^2 = 2 \boldsymbol{H}_E \cdot \nabla E, \qquad H_{|\nabla E|^2} = 2 \boldsymbol{H}^2 \tag{4.31}$$

Term name	formula	symbol	cost
Molecular Orbital Basis TEIs	$2(ia\|jb) - (ib\|ja)$	$\langle ij\|\|ab\rangle$	$O(N^5)$
Aufbau Fock Matrix	$h_{ia} + \sum_p \langle pi\|\|pa\rangle$	f_{ia}	$O(N^4)$
Contracted TEIs	$\sum_{ab} v_{bj} \langle bi\|\|ja\rangle$	$F[v]_{ia}$	$O(N^4)$
Half-Contracted TEIs	$\sum_{ic} c_{ic}(ja\|bc)$	—	$O(N^5)$
Double Half-Contrated TEIs	$\sum_{cd} c_{ic}c_{ad}(jc\|bd)$	—	$O(N^5)$

Table 4.6: Symbols and formulae for terms that appear in the ESMF energy Hessian in appendix D.

following the above discussion. Approximations to the energy Hessian were made to satisfy particular cost-scaling thresholds by dropping terms that exceeded them. In our experience, approximating the energy using physical arguments before differentiating would often provide ambiguous results, as the final expressions depended strongly on how one identified and grouped 'physically significant' terms.

Analytical formulae for the energy Hessian are given in appendix D, and can be broken down into classes of terms, named in table 4.6. Calculation of individual elements of the Hessian matrix is an $O(N^5)$ proposition, owing to molecular transformations of the two-electron integral (TEI) tensor necessary for both uncontracted and half-contracted TEI terms. Performing exact matrix-vector multiplies with the Hessian matrix instead scales as $O(N^4)$, as these terms can be recast as fully-contracted TEIs. As an example, we consider what we call a double half-contracted TEI term:

$$\sum_{cd} c_{ic}c_{ad}(jc|bd) \tag{4.32}$$

Which is calculated as a pair of one-index transformations on the TEIs, directly analogous to a pair of $O(N^5)$ orbital basis transformations. When we perform a matrix-vector multiplication, this can instead be written as a set of $O(N^3)$ matrix multiplies applied to a fully-contracted TEI term, $F[v]$.

$$\sum_{jb} H_{ia,jb}v_{jb} \quad \rightarrow \quad \sum_{cdjb} c_{ic}c_{ad}v_{jb}(jc|bd) \quad \rightarrow \quad \sum_{cd} c_{ic}c_{ad}F[v]_{cd} \tag{4.33}$$

As the vector v can be transformed to the atomic orbital basis before contraction, these terms can be calculated without requiring any TEI basis transformations, and the remaining full TEI contraction can be performed at $O(N^4)$.

For numerically difficult optimizations, Quasi-Newton solvers must be preconditioned.[323–325] The initial QN preconditioner is important since it serves as the foundation of a Hessian inverse that is iteratively updated over many steps. At the same time, we want to avoid building or inverting the full Hessian matrix, as its dimensions are $O(N^2)$, following the amount of orbital rotation and CIS linear parameters. Direct numerical inversion therefore costs $O(N^6)$, which serves as the cost-scaling upper-bound and is realized both by exact

Newton-Raphson optimization and BFGS optimization that uses the full Hessian inverse as its preconditioner. Alone, the BFGS algorithm attains a cost-scaling of $O(N^4)$, limited by the cost-scaling of gradient evaluation of the target function, which involves a single Hessian-gradient matrix-vector multiplication. While we will present the cost-scaling of Hessian preconditioners assuming exact matrix mathematics, sparsity arguments can be used to reduce the cost-scaling for the expensive TEI contractions to $O(N^3)$.[37] This appears to be closer to the minimum price of each BFGS iteration, given the multitude of matrix-multiplications present in the SGM derivative expressions.

We will begin by looking at the convergence behavior of Newton-Raphson optimization which uses the approximate SGM Hessian, given in the first column of table 4.7. As Hessians are calculated and inverted in full each step, these results represent a benchmark for step quality, as is evident by their low iteration count. Speedy convergence is achieved in spite having dropped the expensive and imprecise finite-difference third-derivative gradient contractions, and empirically justifies this approximation. In addition, since the Hessian is recalculated in full every step, the Jacobian transformations in equation 4.29 do not apply.

In the second column of table 4.7, we perform a BFGS optimization preconditioned with an fully inverted Hessian matrix. As this preconditioner is calculated a single time at the start of the optimization, the Jacobian transformations in equation 4.29 formally apply. Slowed convergence relative to the full NR optimization thus reflects inaccuracies both introduced by the approximate BFGS Hessian and omission of these variable transformations. While iteration count is roughly doubled in most cases, the cost per iteration has been reduced to the cost of $O(N^4)$ target function gradient evaluation from $O(N^6)$ matrix inversion, and the total optimization time has thus been cut by a significant factor. Though we have not separated the effects of the BFGS approximation and disregard of Jacobian transformations, the latter does not appear to hamper optimization speed enough to justify its implementation.

The third and fourth columns of table 4.7 test preconditioners at two significant tiers of cost-scaling. In the third column, the full Hessian inverse is replaced by its diagonal equivalent, and corresponds to perhaps the simplest way to sidestep the expensive dense matrix inversion otherwise required. As the calculation of individual Hessian matrix elements still requires the evaluation of MO-basis TEIs and half-contracted TEI terms, the cost of this diagonal inverse preconditioner is still $O(N^5)$, To lower this cost even further to match the ESMF gradient at $O(N^4)$ requires us to drop these terms, which we do in the "one-body" diagonal Hessian in column 4. This approximation is so-named because the only remaining expensive elements are full TEI contractions, and are so cost-limited by the same operation as the mean-field Fock matrix. While iteration count steadily increases as cheaper preconditioners are used, the total runtime of BFGS optimization steadily drops and finally becomes competitive with the ESMF-SCF implementation with the one-body Hessian preconditioner.

Identity-matrix preconditioned optimizers in the GVP and BFGS implementations are presented in columns 5 and 6 (respectively) of table 4.7 to serve as a concrete reminder of just how profoundly preconditioners impact BFGS convergence speed. In addition, these results help verify that the relatively slow convergence speed of the GVP implementation is primarily limited by its choice of preconditioner. The differences between these two methods can be explained by small implementation differences in GVP and BFGS, both in terms of derivative numerics and optimizer line-searches. In the GVP algorithm, orbital derivatives are calculated in a fixed orbital basis according to equation 4.27, and basis transformations

Method/ Preconditioner	Newton Raphson		L-BFGS Inverse		L-BFGS Diagonal		L-BFGS 1-Body Diagonal		L-BFGS Identity		ESMF-GVP Identity		ESMF-SCF
Cost Scaling	$O(N^6)$		$O(N^6)$		$O(N^5)$		$O(N^4)$		$O(N^4)$		$O(N^4)$		$O(N^4)$
Molecule	iter	time	iter	time	iter	time	iter	time	iter	time	iter	time	time
dihydrogen	7	0.0980	9	0.0802	16	0.1266	21	0.1783	33	0.5635	15	0.1037	0.0330
tetrahydrogen	10	1.2066	13	0.3580	30	0.4160	43	0.4346	129	3.6410	195	1.8533	1.0459
lithium hydride	12	1.2877	28	0.4439	35	0.4679	42	0.4589	116	3.0883	136	1.2606	0.6130
dilithium	8	3.5661	12	1.0057	26	0.8622	37	0.6495	158	5.9024	83	1.0679	0.4538
ammonia	10	8.9060	17	1.8946	38	1.1565	45	0.6570	362	18.8124	896	12.4677	1.2971
water	10	6.8645	292	12.8896	35	0.7211	58	0.6786	330	15.9113	636	7.9307	0.8682
formaldehyde	13	48.1102	23	7.5713	45	3.4301	58	1.2258	633	49.4131	785	16.3812	1.2490
hydrogen sulfide	8	12.7346	12	3.2764	89	8.2626	171	3.7532	521	33.3353	1000[a]	13.4606	1.4347
cyclopropene	12	356.2605	17	59.1773	69	33.5493	101	4.1464	1209	200.6113	1000[a]	84.3206	9.4412
diazomethane	15	227.1900	37	31.9987	78	15.5644	88	2.5219	1036	122.3400	1001[a]	54.0999	2.5539
ketene	13	205.7483	24	31.1079	57	14.7128	63	2.2744	943	113.1768	1001[a]	64.9124	2.5340
acetaldehyde	12	381.9266	24	62.8761	54	34.8132	59	2.6190	1079	182.0033	993	82.4376	3.7529
ethanol	13	838.2633	21	127.4991	—	—	65	4.2850	1241	293.6065	1001[a]	98.8830	4.5652
formamide	12	392.5720	36	67.7175	102	32.0470	95	3.3470	1163	160.6049	1001[a]	76.7921	4.4940
acetone	12	1866.0479	22	328.7256	56	210.3979	61	4.2673	1538	582.5048	844	123.1596	5.1629
pyrrole	11	2923.2304	24	525.3435	67	337.2110	73	5.5375	1703	942.2698	1001[a]	183.3615	8.9777

Table 4.7: ESMF optimization timings (in seconds) and iteration count by molecule, method, and Hessian preconditioner.
[a]Failed to converge before reaching the maximum iteration count.

are performed only when orbital rotation parameters become too large. In the present BFGS implementation, orbital transformations are performed at each single step to simplify analytic expressions, and so gradients may numerically differ between these two implementations. In addition, the GVP implementation uses more general and aggressive line search to control step-size, rigorously ensuring that each BFGS step follows the Wolfe conditions.[330,331] When a poor preconditioner is used, this step-size control is primarily responsible for its quicker convergence and higher stability relative to the present BFGS implementation, which uses no step-size control.

4.10 Conclusions

Many excited state methods have difficulty accurately predicting charge-transfer and core excitations due to the large shift in ground-state electronic density – and the following many-body response – that characterize these states. The ESMF wavefunction, which combines a CIS linear wavefunction with orbital relaxation effects, can more efficiently and accurately represent these states in a variational, state-selective way at mean-field cost. Though it does not include either double excitations or dynamic correlation effects, it appears to be an effective platform for further improvement.[320] Along these lines, the state-specific orbital optimization present in the ESMF wavefunction suggests that it may serve as excited-orbital reference, filling the same role as ground-state Hartree-Fock for excited state wavefunctions.

Through the worked example of methanimine, we identify a subtle collapse behavior that affects state-specific orbital optimization of multideterminant wavefunctions. When coupling between linear CI and nonlinear orbital parameters is ignored, state-mixing can cause chaos in the optimization process, ultimately causing variational collapse despite the best efforts of targeting functions and state-selection metrics. We trace this behavior to the omission of significant derivative coupling terms in the local expansion of eigenvalues in the linear space, and orbital steps are made on an energy surface whose curvature is severely underestimated. This behavior is directly observed in an uncoupled state-specific CASSCF algorithm and expected to contribute to the variational collapse found in the ESMF-SCF optimizer, but is notably absent from the coupled ESMF-GVP algorithm. While this coupling analysis mathematically echoes derivative couplings and avoided crossings found in non-adiabatic dynamics,[332] we emphasize that these features are artifacts of a limited wavefunction space, and are not physical in nature.

In order to accelerate the stable ESMF-GVP implementation, we explore several Hessian preconditioning schemes for coupled BFGS optimization. Implementation of analytical derivatives allows us to test the effect of intermediate orbital transformations, truncation of expensive third-derivative terms, and we find that all of these approximations can significantly improve convergence speed at marginal cost. Following cost-scaling arguments, we drop off-diagonal elements and expensive terms in the Hessian to arrive at a one-body diagonal Hessian that can be calculated at mean-field cost and accelerates QN methods to match the speed of gradient-only ESMF-SCF implementation. In addition, the stability introduced both by the the square-gradient target function and coupling present in the off-diagonal ellements of the iteratively-updated BFGS Hessian is enough to prevent variational collapse that can be found in ESMF-SCF.

Orbital relaxation is a significant effect in many electronic excitations that can be ac-

counted for at mean-field cost in state-specific ESMF optimizations. Though the non-linear optimization necessary to tune orbitals state-specifically can open up subtle and opaque avenues for variational collapse, we show that stability can be recovered by employing target functions and accounting for couplings in the wavefunction parameter space. While these optimizations may appear prohibitively expensive at first glance, appropriate numerical preconditioning and judicious approximation can greatly accelerate their convergence while ensuring variational stability. It is our hope that state-specific orbital optimization can be extended more broadly within quantum chemistry so that excited state wavefunctions can be efficiently and accurately represented in their native orbital basis.

5 Concluding Remarks

Chemical processes such as bond-breaking, charge-transfer excitations, and core excitations can require special theoretical treatment, and can often be difficult to describe with simpler methods. These states can be complex enough that an exhaustive description becomes too costly, and information-dense non-linear functional forms are an important tool to efficiently capture these states. Position-space QMC is a good setting for compact functional forms, but many position-space factors focus on general many-body correlation treatments, and are not specifically designed to describe the strong correlation present in bond-breaking processes. Excited states formalisms are limited by their adherence to a fixed reference point in linear response theory, effectively limiting the reach of the wavefunction, and methods to extend this directly can be unstable or difficult to implement. Addressing these limitations could go a long way toward making accurate wavefunctions much more compact and accelerate the methods that optimize them.

Following both of these threads, we implement and propose improvements for each. Using special basis functions that tessellate space, number-counting Jastrow factors can naturally account for bond-breaking in position space and variationally improves the wavefunction as the granularity of this partition increases. While state-specific excited-state optimization can compress and tune orbital parameters for an excited state, this can come with more costly target functions and variational instabilities, and we carefully document an instability unique to these procedures while showing that they can be accelerated to match more conventional algorithms. Both of these threads aim to compactify the wavefunction and its expression of complex electronic correlations in different contexts by qualitatively improving the zeroth-order reference state by variational optimization.

Number-counting Jastrows most clearly work to correct for the strong correlation of bond-breaking in small hydrogen clusters, and have been shown to correctly and strongly remove inappropriate configurations present in mean-field descriptions of stretched chemical bonds. In more complex dissociation processes, more careful subdivisions are necessary, and are surprisingly effective at improving the accuracy of the wavefunction. For instance, in the CaO dissociation, a composite 32-region partition made of Voronoi cells divided into spherical shells can effectively couple with orbital parameters and starkly improve the nodal surface of a single-determinant wavefunction. This is in contrast to both simpler partitions and shorter-range cusp-correcting Jastrow factors, and shows that the flexible and tunable subdivisions available from the normalized-gaussian basis functions are uniquely able to make these corrections.

Further research could investigate finer subdivisions more closely. Stretching ethene across the double bond produces two coupled, open-shell diradical fragments rather than a pair of closed-shell singlets, since electrons correlate their positions to spread out between the two degenerate valence p-orbitals present on each fragment. That the spherical subdivisions used for Calcium Oxide were vital in attaining an accurate result indicates that core-valence correlations are both significant and can be meaningfully captured by these number-counting Jastrow factors. While the one-region-per-atom Voronoi partitions can be generated automatically, these spherical subdivisions were instead prepared by hand due to their increased complexity. One next step would be to find an algorithmic way to generate these partitions, perhaps by adaptively adding subdivisions where they are needed most, as prescribed by

a mathematical metric. Though these Number-counting Jastrows were able to variationally order single-determinant wavefunctions for ethene, more exploration could be done here to more precisely capture the intra-molecular open-shell correlation present. Finally, given recent developments in multi-determinant wavefunctions and excited-state target-function optimization in VMC, these Jastrows could be used in more complex systems, where their ability to perform nodal surface relaxation could meaningfully improve DMC energies.

Our analysis of the coupled-collapse mechanism is general to multi-determinant state-specific optimizations and can readily occur in what otherwise appear to be very simple systems as we showed with methanimine. The coupled-collapse mechanism occurs through a poorly-modeled avoided crossing region and can be nigh-invisible to uncoupled optimization algorithms. This analysis is particularly important for state-specific CASSCF optimizations, because while avoiding this collapse can be done by simply expanding the active-space to include orbitals in each of the two crossing states, this nullifies one stated advantage of state-specific algorithms – that only the targetted state needs to be represented accurately. This is also important for methods like ESMF-SCF, where the wavefunction collapse can be obscured by a restricted functional form and an unconventional orbital step condition. The observation that iterative updates calculated by optimizing a square-gradient target function is enough to prevent this collapse is significant, and provides a straightforward way to generally include this coupling at an affortable cost. More generally, our analysis shows that adding this coupling to an uncoupled algorithm can be done without increasing the cost scaling, at approximately three times the cost. Future work could generalize state-selection metrics to detect and account for possible state-mixing and appropriately model these avoided crossings in parameter space when necessary.

The stability attained from the coupled algorithm is counterbalanced by the fact that it requires higher-order derivatives, both from QN optimization and from its square-gradient target function. Quasi-Newton optimizations are highly sensitive to the initial Hessian matrix, but through a number of approximations, an accurate-enough initial guess can be obtained by trimming expensive terms from analytical formulae, and evaluated at the same cost as the target function gradient. While they are exceedingly convenient for simplifying derivative expressions, intermediate orbital transformations formally transform linear wavefunction variables, and should (theoretically) be counterbalanced by transformations on the gradient history in QN algorithms. In a similar vein, third-derivative terms that appear in the square-gradient Hessian threatens to increase the cost of evaluation. However, both of these expensive and complex pieces can be dropped completely without meaningful loss in accuracy, theoretically justified by the already-approximate nature of Quasi-Newton optimization, and empirically justified by progressively increasing the level of approximation to study their effects. The preconditioned QN algorithm developed here enjoys the same cost-scaling and runtime as the ESMF-SCF method which calculates only the energy gradient and an enhanced variational stability, and is a worthy algorithmic platform on which to build further perturtabive improvements. The acceleration and stability shown here will help push it to the next stage in attaining widespread use, and lessons learned from this optimization procedure may be useful in implementing or analyzing these extensions.

The electronic structure of molecules is a difficult problem to solve and complex systems demand accuracy that pushes the boundaries of wavefunction size and intricacy. We have pursued multiple novel avenues to improve wavefunctions and their optimization that extends

the reach of variational methods in particular, in an effort to provide a concise baseline for states that can otherwise evade compact descriptions. While our exploration of these approaches has been limited to small molecular systems, there is room for each to be extended to larger systems and more accurate methodologies, and we hope this work serves as a solid foundation for future work.

References

[1] Thom H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen, *The Journal of Chemical Physics*, **90**(2), 1007–1023, (1989).

[2] A. J. C. Varandas, Extrapolating to the one-electron basis-set limit in electronic structure calculations, *The Journal of Chemical Physics*, **126**(24), 244105, (2007).

[3] Julian Tirado-Rives and William L. Jorgensen, Performance of B3LYP density functional methods for a large set of organic molecules, *Journal of Chemical Theory and Computation*, **4**(2), 297–306, (2008).

[4] Wolfgang Domcke and David R. Yarkony, Role of conical intersections in molecular spectroscopy and photoinduced chemical dynamics, *Annual Review of Physical Chemistry*, **63**, 325–352, (2012).

[5] Samer Gozem, Hoi Ling Luk, Igor Schapiro, and Massimo Olivucci, Theory and simulation of the ultrafast double-bond isomerization of biological chromophores, *Chemical Reviews*, **117**(22), 13502–13565, (2017).

[6] M. Garavelli, P. Celani, F. Bernardi, M.A. Robb, and M. Olivucci, The C5H6NH2+ protonated Shiff base: an ab initio minimal model for retinal photoisomerization, *Journal of the American Chemical Society*, **119**(29), 6891–6901, (1997).

[7] Marco Garavelli, Fernando Bernardi, Massimo Olivucci, Thom Vreven, Stèphane Klein, Paolo Celani, and Michael A. Robb, Potential-energy surfaces for ultrafast photochemistry Static and dynamic aspects, *Faraday Discuss.*, **110**, 51–70, (1998).

[8] A. Warshel and Z. T. Chu, Nature of the Surface Crossing Process in Bacteriorhodopsin: Computer Simulations of the Quantum Dynamics of the Primary Photochemical Event, *The Journal of Physical Chemistry B*, **105**(40), 9857–9871, (2001).

[9] Lihong Liu, Jian Liu, and Todd J. Martinez, Dynamical correlation effects on photoisomerization: Ab initio multiple spawning dynamics with MS-CASPT2 for a model trans-protonated Schiff base, *The Journal of Physical Chemistry B*, **120**(8), 1940–1949, (2016).

[10] Igor Schapiro, Mikhail Nikolaevich Ryazantsev, Luis Manuel Frutos, Nicolas Ferré, Roland Lindh, and Massimo Olivucci, The ultrafast photoisomerizations of rhodopsin and bathorhodopsin are modulated by bond length alternation and HOOP driven electronic effects, *Journal of the American Chemical Society*, **133**(10), 3354–3364, (2011).

[11] Jae Woo Park and Toru Shiozaki, On-the-fly CASPT2 surface-hopping dynamics, *Journal of Chemical Theory and Computation*, **13**(8), 3676–3683, (2017).

[12] Mark S. Gordon and Donald G. Truhlar, The Hartree-Fock dissociation of F2, *Theoretica Chimica Acta*, **71**(1), 1–5, (1987).

[13] Charles Alfred Coulson and Inga Fischer, XXXIV. Notes on the molecular orbital treatment of the hydrogen molecule, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **40**(303), 386–393, (1949).

[14] Noriyuki Minezawa and Mark S. Gordon, Optimizing conical intersections by spin-flip density functional theory: application to ethylene, *The Journal of Physical Chemistry A*, **113**(46), 12749–12753, (2009).

[15] Thomas M. Henderson and Gustavo E. Scuseria, Spin-projected generalized Hartree-Fock method as a polynomial of particle-hole excitations, *Physical Review A*, **96**(2), 022506, (2017).

[16] Hideo Fukutome, Unrestricted Hartree–Fock theory and its applications to molecules and chemical reactions, *International Journal of Quantum Chemistry*, **20**(5), 955–1065, (1981).

[17] Leslie L. Foldy, Antisymmetric functions and Slater determinants, *Journal of Mathematical Physics*, **3**(3), 531–539, (1962).

[18] S. Zarrabian, C.R. Sarma, and J. Paldus, Vectorizable approach to molecular CI problems using determinantal basis, *Chemical Physics Letters*, **155**(2), 183–188, (1989).

[19] C. David Sherrill and Henry F. Schaefer III, The configuration interaction method: Advances in highly correlated approaches, In *Advances in Quantum Chemistry* volume 34 , 143–269. Elsevier (1999).

[20] Jeppe Olsen, Poul Jørgensen, and Jack Simons, Passing the one-billion limit in full configuration-interaction (FCI) calculations, *Chemical Physics Letters*, **169**(6), 463–472, (1990).

[21] Anna I. Krylov, Equation-of-Motion coupled-cluster Methods for open-shell and electronically excited species: The hitchhiker's guide to Fock space, *Annual Review of Physical Chemistry*, **59**(1), 433–462, (2008).

[22] Erich Runge and Eberhard KU Gross, Density-functional theory for time-dependent systems, *Physical Review Letters*, **52**(12), 997, (1984).

[23] Andreas Dreuw and Martin Head-Gordon, Single-reference ab initio methods for the calculation of excited states of large molecules, *Chemical Reviews*, **105**(11), 4009–4037, (2005).

[24] Carsten A. Ullrich and Zeng-hui Yang, A brief compendium of time-dependent density functional theory, *Brazilian Journal of Physics*, **44**(1), 154–188, (2014).

[25] Neepa T. Maitra, Fan Zhang, Robert J. Cave, and Kieron Burke, Double excitations within time-dependent density functional theory linear response, *The Journal of Chemical Physics*, **120**(13), 5932–5937, (2004).

[26] Robert J. Cave, Fan Zhang, Neepa T. Maitra, and Kieron Burke, A dressed TDDFT treatment of the 21Ag states of butadiene and hexatriene, *Chemical Physics Letters*, **389**(1), 39–42, (2004).

[27] Peter Elliott, Sharma Goldson, Chris Canahui, and Neepa T Maitra, Perspectives on double-excitations in TDDFT, *Chemical Physics*, **391**(1), 110–119, (2011).

[28] Neepa T. Maitra, Perspective: Fundamental aspects of time-dependent density functional theory, *The Journal of Chemical Physics*, **144**(22), 220901, (2016).

[29] Young Choon Park, Mykhaylo Krykunov, and Tom Ziegler, On the relation between adiabatic time dependent density functional theory (TDDFT) and the ΔSCF-DFT method. Introducing a numerically stable ΔSCF-DFT scheme for local functionals based on constricted variational DFT, *Molecular Physics*, **113**(13-14), 1636–1647, (2015).

[30] Hans-Joachim Werner and Wilfried Meyer, A quadratically convergent MCSCF method for the simultaneous optimization of several states, *The Journal of Chemical Physics*, **74**(10), 5794–5801, (1981).

[31] Lan Nguyen Tran, Jacqueline A.R. Shea, and Eric Neuscamman, Tracking excited states in wave function optimization using density matrices and variational principles, *Journal of Chemical Theory and Computation*, **15**(9), 4790–4803, (2019).

[32] Lan Nguyen Tran and Eric Neuscamman, Improving excited-state potential energy surfaces via optimal orbital shapes, *The Journal of Physical Chemistry A*, **124**(40), 8273–8279, (2020).

[33] Jacqueline A. R. Shea and Eric Neuscamman, Size consistent excited states via algorithmic transformations between variational principles, *Journal of Chemical Theory and Computation*, **13**(12), 6078–6088, (2017).

[34] Jacqueline A.R. Shea and Eric Neuscamman, Communication: A mean field platform for excited state quantum chemistry, *The Journal of Chemical Physics*, **149**(8), 081101, (2018).

[35] Jacqueline A.R. Shea, Elise Gwin, and Eric Neuscamman, A generalized variational principle with applications to excited state mean field theory, *Journal of Chemical Theory and Computation*, **16**(3), 1526–1540, (2020).

[36] Tarini S. Hardikar and Eric Neuscamman, A self-consistent field formulation of excited state mean field theory, *The Journal of Chemical Physics*, **153**(16), 164108, (2020).

[37] Luning Zhao and Eric Neuscamman, Excited state mean-field theory without automatic differentiation, *The Journal of Chemical Physics*, **152**(20), 204112, (2020).

[38] Niels Bohr, LXXIII. On the constitution of atoms and molecules, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, **26**(155), 857–875, (1913).

[39] Donald Allan McQuarrie and John Douglas Simon, *Physical chemistry: a molecular approach* volume 1, University Science Books, Sausalito, CA (1997).

[40] Ramamurti Shankar, *Principles of Quantum Mechanics*, Springer Science & Business Media (2012).

[41] Claude Cohen-Tannoudji, Bernard Diu, Frank Laloe, and Bernard Dui, Quantum Mechanics (2006).

[42] Rafael De la Madrid, The role of the rigged Hilbert space in quantum mechanics, *European Journal of Physics*, **26**(2), 287, (2005).

[43] A. Szabo and N.S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*, Dover Books on Chemistry. Dover Publications (2012).

[44] Björn O. Roos, The complete active space self-consistent field method and its applications in electronic structure calculations, *Advances in Chemical Physics*, **69**, 399–445, (1987).

[45] Garnet Kin-Lic Chan and Sandeep Sharma, The density matrix renormalization group in quantum chemistry, *Annual Review of Physical Chemistry*, **62**, 465–481, (2011).

[46] Max Born and Robert Oppenheimer, Zur quantentheorie der molekeln, *Annalen der Physik*, **389**(20), 457–484, (1927).

[47] Wolfgang Pauli, The connection between spin and statistics, *Physical Review*, **58**(8), 716, (1940).

[48] S. Bartalucci, S. Bertolucci, M. Bragadireanu, M. Cargnelli, M. Catitti, C. Curceanu, S. Di Matteo, J.-P. Egger, C. Guaraldo, M. Iliescu, et al., New experimental limit on the Pauli exclusion principle violation by electrons, *Physics Letters B*, **641**(1), 18–22, (2006).

[49] Freeman J. Dyson, Ground-state energy of a finite system of charged particles, *Journal of Mathematical Physics*, **8**(8), 1538–1545, (1967).

[50] A.L. Fetter and J.D. Walecka, Quantum theory of many particle systems New York (1971).

[51] Warren J. Hehre, Robert Ditchfield, and John A. Pople, Self consistent molecular orbital methods. XII. Further extensions of Gaussian type basis sets for use in molecular orbital studies of organic molecules, *The Journal of Chemical Physics*, **56**(5), 2257–2261, (1972).

[52] T. Dunning, Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen, *Journal of Chemical Physics*, **90**, 1007, (1989).

[53] Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen, *Molecular electronic-structure theory*, John Wiley & Sons (2014).

[54] Olaf Melsheimer, Rigged Hilbert space formalism as an extended mathematical formalism for quantum systems. I. General theory, *Journal of Mathematical Physics*, **15**(7), 902–916, (1974).

[55] James W. Demmel, *Applied numerical linear algebra*, SIAM (1997).

[56] Michel Crouzeix, Bernard Philippe, and Miloud Sadkane, The davidson method, *SIAM Journal on Scientific Computing*, **15**(1), 62–76, (1994).

[57] Dong C. Liu and Jorge Nocedal, On the limited memory BFGS method for large scale optimization, *Mathematical Programming*, **45**(1-3), 503–528, (1989).

[58] Daniel S Levine, Diptarka Hait, Norm M Tubman, Susi Lehtola, K Birgitta Whaley, and Martin Head-Gordon, CASSCF with extremely large active spaces using the adaptive sampling configuration interaction method, *Journal of Chemical Theory and Computation*, **16**(4), 2340–2354, (2020).

[59] Norm M. Tubman, C. Daniel Freeman, Daniel S. Levine, Diptarka Hait, Martin Head-Gordon, and K. Birgitta Whaley, Modern approaches to exact diagonalization and selected configuration interaction with the adaptive sampling CI method, *Journal of Chemical Theory and Computation*, **16**(4), 2139–2159, (2020).

[60] Steven Weinberg, *Lectures on quantum mechanics*, Cambridge University Press (2015).

[61] T.S. Evans and D.A. Steer, Wick's theorem at finite temperature, *Nuclear Physics B*, **474**(2), 481–496, (1996).

[62] Ernest Davidson, *Reduced density matrices in quantum chemistry* volume 6, Elsevier (2012).

[63] Albert John Coleman and Vyacheslav I Yukalov, *Reduced density matrices: Coulsonś challenge* volume 72, Springer Science & Business Media (2000).

[64] Douglas Rayner Hartree and W Hartree, Self-consistent field, with exchange, for beryllium, *Proceedings of the Royal Society of London. Series A-Mathematical and Physical Sciences*, **150**(869), 9–33, (1935).

[65] Vladimir Fock, Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems, *Zeitschrift für Physik*, **61**(1-2), 126–148, (1930).

[66] Pierre Hohenberg and Walter Kohn, Inhomogeneous electron gas, *Physical Review*, **136**(3B), B864, (1964).

[67] Walter Kohn and Lu Jeu Sham, Self-consistent equations including exchange and correlation effects, *Physical Review*, **140**(4A), A1133, (1965).

[68] Robert G. Parr, Density functional theory of atoms and molecules, In *Horizons of quantum chemistry* , 5–15. Springer (1980).

[69] David C. Langreth and M.J. Mehl, Beyond the local-density approximation in calculations of ground-state electronic properties, *Physical Review B*, **28**(4), 1809, (1983).

[70] Axel D. Becke, Density-functional thermochemistry. I. The effect of the exchange-only gradient correction, *The Journal of Chemical Physics*, **96**(3), 2155–2160, (1992).

[71] E.K.U. Gross, J.F. Dobson, and M. Petersilka, Density functional theory of time-dependent phenomena, In *Density functional theory II* , 81–172. Springer (1996).

[72] Ylva Andersson, Erika Hult, Henrik Rydberg, Peter Apell, Bengt I. Lundqvist, and David C. Langreth, Van der Waals interactions in density functional theory, In *Electronic Density Functional Theory* , 243–260. Springer (1998).

[73] Muneaki Kamiya, Takao Tsuneda, and Kimihiko Hirao, A density functional study of van der Waals interactions, *The Journal of Chemical Physics*, **117**(13), 6010–6015, (2002).

[74] Erin R. Johnson and Gino A. DiLabio, Structure and binding energies in van der Waals dimers: Comparison between density functional theory and correlated ab initio methods, *Chemical Physics Letters*, **419**(4-6), 333–339, (2006).

[75] Aron J. Cohen, Paula Mori-Sánchez, and Weitao Yang, Fractional spins and static correlation error in density functional theory (2008).

[76] Fan Wang and Tom Ziegler, Time-dependent density functional theory based on a non-collinear formulation of the exchange-correlation potential, *The Journal of Chemical Physics*, **121**(24), 12191–12196, (2004).

[77] Aron J. Cohen and Nicholas C. Handy, Dynamic correlation, *Molecular Physics*, **99**(7), 607–615, (2001).

[78] Nicholas C. Handy and Aron J. Cohen, Left-right correlation energy, *Molecular Physics*, **99**(5), 403–412, (2001).

[79] Ernest R. Davidson and Leon L. Jones, Correlation splitting in the hydrogen molecule, *The Journal of Chemical Physics*, **37**(9), 1918–1922, (1962).

[80] Joanne M. Wittbrodt and H. Bernhard Schlegel, Some reasons not to use spin projected density functional theory, *The Journal of Chemical Physics*, **105**(15), 6574–6577, (1996).

[81] István Mayer, The spin-projected extended Hartree-Fock method, In *Advances in quantum chemistry* volume 12 , 189–262. Elsevier (1980).

[82] Carlos A. Jiménez-Hoyos, Thomas M. Henderson, Takashi Tsuchimochi, and Gustavo E. Scuseria, Projected hartree–fock theory, *The Journal of Chemical Physics*, **136**(16), 164109, (2012).

[83] Ernest R. Davidson and Weston Thatcher Borden, Symmetry breaking in polyatomic molecules: real and artifactual, *The Journal of Physical Chemistry*, **87**(24), 4783–4790, (1983).

[84] Obis Castaño and Peter Karadakov, The spin-unrestricted and the spin-projected HF methods: a comparison based on the stability properties of the closed-shell RHF solutions for large cyclic polyenes, *Chemical Physics Letters*, **130**(1-2), 123–126, (1986).

[85] Lorenzo Stella, Claudio Attaccalite, Sandro Sorella, and Angel Rubio, Strong electronic correlation in the hydrogen chain: A variational Monte Carlo study, *Physical Review B*, **84**(24), 245117, (2011).

[86] Johannes Hachmann, Wim Cardoen, and Garnet Kin-Lic Chan, Multireference correlation in long molecules with the quadratic scaling density matrix renormalization group, *The Journal of Chemical Physics*, **125**(14), 144101, (2006).

[87] Joshua Jortner, Mordechai Bixon, Thomas Langenbacher, and Maria E. Michel-Beyerle, Charge transfer and transport in DNA, *Proceedings of the National Academy of Sciences*, **95**(22), 12759–12765, (1998).

[88] Eunju Lee Tae, Seung Hwan Lee, Jae Kwan Lee, Su San Yoo, Eun Ju Kang, and Kyung Byung Yoon, A strategy to increase the efficiency of the dye-sensitized TiO2 solar cells operated by photoexcitation of dye-to-TiO2 charge-transfer bands, *The Journal of Physical Chemistry B*, **109**(47), 22513–22522, (2005).

[89] Haining Tian, Xichuan Yang, Jingxi Pan, Ruikui Chen, Ming Liu, Qingyu Zhang, Anders Hagfeldt, and Licheng Sun, A Triphenylamine Dye Model for the Study of Intramolecular Energy Transfer and Charge Transfer in Dye-Sensitized Solar Cells, *Advanced Functional Materials*, **18**(21), 3461–3468, (2008).

[90] Rodney J. Bartlett and Monika Musiał, Coupled-cluster theory in quantum chemistry, *Reviews of Modern Physics*, **79**(1), 291, (2007).

[91] Krishnan Raghavachari, Gary W. Trucks, John A. Pople, and Martin Head-Gordon, A fifth-order perturbation comparison of electron correlation theories, *Chemical Physics Letters*, **157**(6), 479–483, (1989).

[92] Chr. Møller and Milton S. Plesset, Note on an approximation treatment for many-electron systems, *Physical Review*, **46**(7), 618, (1934).

[93] Peter G. Szalay, Thomas Muller, Gergely Gidofalvi, Hans Lischka, and Ron Shepard, Multiconfiguration self-consistent field and multireference configuration interaction methods and applications, *Chemical Reviews*, **112**(1), 108–181, (2012).

[94] Stephen R. Langhoff and Ernest R. Davidson, Configuration interaction calculations on the nitrogen molecule, *International Journal of Quantum Chemistry*, **8**(1), 61–72, (1974).

[95] Stefano Evangelisti, Jean-Pierre Daudey, and Jean-Paul Malrieu, Convergence of an improved CIPSI algorithm, *Chemical Physics*, **75**(1), 91 – 102, (1983).

[96] Björn O. Roos, Peter R. Taylor, and Per E.M. Sigbahn, A complete active space SCF method (CASSCF) using a density matrix formulated super-CI approach, *Chemical Physics*, **48**(2), 157–173, (1980).

[97] Klaus Ruedenberg, Michael W. Schmidt, Mary M. Gilbert, and S.T. Elbert, Are atoms intrinsic to molecular electronic wavefunctions? I. The FORS model, *Chemical Physics*, **71**(1), 41–49, (1982).

[98] Björn O. Roos and K.P. Lawley, Ab initio methods in quantum chemistry II, *Advances in Chemical Physics*, **69**, 399–446, (1987).

[99] Per-Åke Malquist, Hans Ågren, and Björn O Roos, The lowest states of the O3+ ion, *Chemical Physics Letters*, **98**(5), 444–449, (1983).

[100] Hans-Joachim Werner and Peter J. Knowles, A second order multiconfiguration SCF procedure with optimum convergence, *The Journal of Chemical Physics*, **82**(11), 5053–5063, (1985).

[101] Peter J. Knowles and Hans-Joachim Werner, An efficient second-order MC SCF method for long configuration expansions, *Chemical Physics Letters*, **115**(3), 259–267, (1985).

[102] Hans-Joachim Werner and Wilfried Meyer, A quadratically convergent multiconfiguration–self-consistent field method with simultaneous optimization of orbitals and CI coefficients, *The Journal of Chemical Physics*, **73**(5), 2342–2356, (1980).

[103] Peter J. Knowles and Hans-Joachim Werner, Internally contracted multiconfiguration-reference configuration interaction calculations for excited states, *Theoretica Chimica Acta*, **84**(1-2), 95–103, (1992).

[104] David A. Kreplin, Peter J. Knowles, and Hans-Joachim Werner, Second-order MCSCF optimization revisited. I. Improved algorithms for fast and robust second-order CASSCF convergence, *The Journal of Chemical Physics*, **150**(19), 194106, (2019).

[105] Jeppe Olsen, Björn O. Roos, Poul Jo/rgensen, and Hans Jo/rgen Aa. Jensen, Determinant based configuration interaction algorithms for complete and restricted configuration interaction spaces, *The Journal of Chemical Physics*, **89**(4), 2185–2192, (1988).

[106] Per Åke Malmqvist, Alistair Rendell, and Björn O. Roos, The restricted active space self-consistent-field method, implemented with a split graph unitary group approach, *Journal of Physical Chemistry*, **94**(14), 5477–5482, (1990).

[107] Michael Klene, Michael A. Robb, Lluis Blancafort, and Michael J. Frisch, A new efficient approach to the direct restricted active space self-consistent field method, *The Journal of Chemical Physics*, **119**(2), 713–728, (2003).

[108] Kerstin Andersson, Per Aake Malmqvist, Björn O. Roos, Andrzej J. Sadlej, and Krzysztof Wolinski, Second-order perturbation theory with a CASSCF reference function, *Journal of Physical Chemistry*, **94**(14), 5483–5488, (1990).

[109] James Finley, Per-Åke Malmqvist, Björn O. Roos, and Luis Serrano-Andrés, The multi-state CASPT2 method, *Chemical Physics Letters*, **288**(2-4), 299–306, (1998).

[110] Vicenta Sauri, Luis Serrano-Andrés, Abdul Rehaman Moughal Shahi, Laura Gagliardi, Steven Vancoillie, and Kristine Pierloot, Multiconfigurational second-order perturbation theory restricted active space (RASPT2) method for electronic excited states: A benchmark study, *Journal of Chemical Theory and Computation*, **7**(1), 153–168, (2011).

[111] Celestino Angeli, Renzo Cimiraglia, S. Evangelisti, T. Leininger, and J.-P. Malrieu, Introduction of n-electron valence states for multireference perturbation theory, *The Journal of Chemical Physics*, **114**(23), 10252–10264, (2001).

[112] Celestino Angeli, Renzo Cimiraglia, and Jean-Paul Malrieu, n-electron valence state perturbation theory: A spinless formulation and an efficient implementation of the strongly contracted and of the partially contracted variants, *The Journal of Chemical Physics*, **117**(20), 9138–9153, (2002).

[113] Celestino Angeli, Mariachiara Pastore, and Renzo Cimiraglia, New perspectives in multireference perturbation theory: the n-electron valence state approach, *Theoretical Chemistry Accounts*, **117**(5), 743–754, (2007).

[114] W.M.C Foulkes, L. Mitas, R.J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Reviews of Modern Physics*, **73**(1), 33, (2001).

[115] George H. Booth, Alex J. W. Thom, and Ali Alavi, Fermion Monte Carlo without fixed nodes: A game of life, death, and annihilation in Slater determinant space, *The Journal of Chemical Physics*, **131**(5), 054106, (2009).

[116] George H. Booth, Andreas Grüneis, and Ali Alavi, Towards an exact description of electronic wavefunctions in real solids, *Nature*, **493**, 365–370, (2013).

[117] J. S. Spencer, N. S. Blunt, and W. M.C. Foulkes, The sign problem and population dynamics in the full configuration interaction quantum Monte Carlo method, *The Journal of Chemical Physics*, **136**(5), 054110, (2012).

[118] Uttam Sinha Mahapatra, Barnali Datta, and Debashis Mukherjee, A size-consistent state-specific multireference coupled cluster theory: Formal developments and molecular applications, *The Journal of Chemical Physics*, **110**(13), 6171–6188, (1999).

[119] Martin C. Gutzwiller, Effect of Correlation on the Ferromagnetism of Transition Metals, *Physical Review Letters*, **10**, 159, (1963).

[120] Martin C. Gutzwiller, Correlation of Electrons in a Narrow s Band, *Physical Review*, **137**, A1726–A1735, (1965).

[121] Mariapia Marchi, Sam Azadi, Michele Casula, and Sandro Sorella, Resonating valence bond wave function with molecular orbitals: Application to first-row molecules, *The Journal of Chemical Physics*, **131**(15), 154116, (2009).

[122] Elvira R. Sayfutyarova, Qiming Sun, Garnet Kin-Lic Chan, and Gerald Knizia, Automated construction of molecular active spaces from atomic valence orbitals, *Journal of Chemical Theory and Computation*, **13**(9), 4063–4078, (2017).

[123] Elvira R. Sayfutyarova and Sharon Hammes-Schiffer, Constructing molecular π-orbital active spaces for multireference calculations of conjugated systems, *Journal of Chemical Theory and Computation*, **15**(3), 1679–1689, (2019).

[124] B. Huron, J. P. Malrieu, and P. Rancurel, Iterative perturbation calculations of ground and excited state energies from multiconfigurational zeroth order wavefunctions, *The Journal of Chemical Physics*, **58**(12), 5745–5759, (1973).

[125] T. Kawakami, K. Miyagawa, Hiroshi Isobe, M. Shoji, S. Yamanaka, M. Katouda, T. Nakajima, K. Nakatani, M. Okumura, and K. Yamaguchi, Relative stability between the manganese hydroxide-and oxo-models for water oxidation by CCSD, DMRG CASCI, CASSCF, CASPT2 and CASDFT methods; Importance of static and dynamical electron correlation effects for OEC of PSII, *Chemical Physics Letters*, **705**, 85–91, (2018).

[126] Garnet Kin-Lic Chan, Anna Keselman, Naoki Nakatani, Zhendong Li, and Steven R. White, Matrix product operators, matrix product states, and ab initio density matrix renormalization group algorithms, *The Journal of Chemical Physics*, **145**(1), 014102, (2016).

[127] Sheng Guo, Zhendong Li, and Garnet Kin-Lic Chan, Communication: An efficient stochastic algorithm for the perturbative density matrix renormalization group in large active spaces, *The Journal of Chemical Physics*, **148**(22), 221104, (2018).

[128] Steven R. White, Density matrix formulation for quantum renormalization groups, *Physical Review Letters*, **69**, 2863–2866, (1992).

[129] Danny L. Yeager, Diane Lynch, Jeffrey Nichols, Poul Joergensen, and Jeppe Olsen, Newton-Raphson approaches and generalizations in multiconfigurational self-consistent field calculations, *The Journal of Physical Chemistry*, **86**(12), 2140–2153, (1982).

[130] Tosio Kato, On the eigenfunctions of many-particle systems in quantum mechanics, *Communications on Pure and Applied Mathematics*, **10**(2), 151–177, (1957).

[131] Wim Klopper, Robert Röhse, and Werner Kutzelnigg, CID and CEPA calculations with linear r12 terms, *Chemical Physics Letters*, **178**(5-6), 455–461, (1991).

[132] Wim Klopper and Werner Kutzelnigg, Møller-Plesset calculations taking care of the correlation cusp, *Chemical Physics Letters*, **134**(1), 17–22, (1987).

[133] Werner Kutzelnigg and Wim Klopper, Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory, *The Journal of Chemical Physics*, **94**(3), 1985–2001, (1991).

[134] Seiichiro Ten-no, Initiation of explicitly correlated Slater-type geminal theory, *Chemical Physics Letters*, **398**(1-3), 56–61, (2004).

[135] Seiichiro Ten-no, New implementation of second-order Møller-Plesset perturbation theory with an analytic Slater-type geminal, *The Journal of Chemical Physics*, **126**(1), 014108, (2007).

[136] Liguo Kong, Florian A. Bischoff, and Edward F. Valeev, Explicitly Correlated R12/F12 Methods for Electronic Structure, *Chemical Reviews*, **112**(1), 75–107, (2012).

[137] Eugene Wigner, On the interaction of electrons in metals, *Physical Review*, **46**(11), 1002, (1934).

[138] John Hubbard, Electron correlations in narrow energy bands, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **276**(1365), 238–257, (1963).

[139] Fabian H.L. Essler, Holger Frahm, Frank Göhmann, Andreas Klümper, and Vladimir E. Korepin, *The one-dimensional Hubbard model*, Cambridge University Press (2005).

[140] Dieter Cremer, Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **1**(4), 509–530, (2011).

[141] George D. Purvis III and Rodney J. Bartlett, A full couple-cluster singles and doubles model: The inclusion of disconnected triples, *Journal of Chemical Physics*, **76**, 1910, (1982).

[142] M.J. Gillan, F.R. Manby, M.D. Towler, and D. Alfè, Assessing the accuracy of quantum Monte Carlo and density functional theory for energetics of small water clusters, *The Journal of Chemical Physics*, **136**(24), 244105, (2012).

[143] Michael J. Bearpark, Francois Ogliaro, Thom Vreven, Martial Boggio-Pasqua, Michael J. Frisch, Susan M. Larkin, Marc Morrison, and Michael A. Robb, CASSCF calculations for photoinduced processes in large molecules: Choosing when to use the RASSCF, ONIOM and MMVB approximations, *Journal of Photochemistry and Photobiology A: Chemistry*, **190**(2-3), 207–227, (2007).

[144] Peter J. Reynolds, David M. Ceperley, Berni J. Alder, and William A. Lester Jr, Fixed-node quantum Monte Carlo for molecules a) b), *The Journal of Chemical Physics*, **77**(11), 5593–5603, (1982).

[145] William A. Lester Jr, Lubos Mitas, and Brian Hammond, Quantum Monte Carlo for atoms, molecules and solids, *Chemical Physics Letters*, **478**(1-3), 1–10, (2009).

[146] Shailesh Chandrasekharan and Uwe-Jens Wiese, Meron-cluster solution of fermion sign problems, *Physical Review Letters*, **83**(16), 3116, (1999).

[147] Jan Linderberg, The antisymmetrized geminal power, a simple correlated wave function for chemical bonding, *Israel Journal of Chemistry*, **19**(1-4), 93–98, (1980).

[148] Troy Van Voorhis and Martin Head-Gordon, A nonorthogonal approach to perfect pairing, *The Journal of Chemical Physics*, **112**(13), 5633–5638, (2000).

[149] W. J. Hehre, R. Ditchfield, and J. A. Pople, Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules, *Journal of Chemical Physics*, **56**, 2257, (1972).

[150] Wim Klopper, Frederick R. Manby, Seiichiro Ten-No, and Edward F. Valeev, R12 methods in explicitly correlated molecular electronic structure theory, *International Reviews in Physical Chemistry*, **25**(3), 427–468, (2006).

[151] N. D. Drummond, M. D. Towler, and R. J. Needs, Jastrow correlation factor for atoms, molecules, and solids, *Physical Review B*, **70**, 235119, (2004).

[152] Arne Lüchow, Alexander Sturm, Christoph Schulte, and Kaveh Haghighi Mood, Generic expansion of the Jastrow correlation factor in polynomials satisfying symmetry and cusp conditions, *Journal of Chemical Physics*, **142**(8), 084111, (2015).

[153] Jeongnim Kim, Andrew D. Baczewski, Todd D. Beaudet, Anouar Benali, M. Chandler Bennett, Mark A. Berrill, Nick S Blunt, Edgar Josué Landinez Borda, Michele Casula, David M. Ceperley, Simone Chiesa, Bryan K Clark, Raymond C. Clay III, Kris T. Delaney, Mark Dewing, Kenneth P. Esler, Hongxia Hao, Olle Heinonen, Paul R. C. Kent, Jaron T. Krogel, Ilkka Kylänpää, Ying Wai Li, M. Graham Lopez, Ye Luo, Fionn D. Malone, Richard M. Martin, Amrita Mathuriya, Jeremy McMinis, Cody A. Melton, Lubos Mitas, Miguel A. Morales, Eric Neuscamman, William D. Parker, Sergio D. Pineda Flores, Nichols A. Romero, Brenda M. Rubenstein, Jacqueline A. R. Shea, Hyeondeok Shin, Luke Shulenburger, Andreas F. Tillack, Joshua P. Townsend, Norm M. Tubman, Beatrice Van Der Goetz, Jordan E. Vincent, D. ChangMo Yang, Yubo Yang, Shuai Zhang, and Luning Zhao, QMCPACK : an open source ab initio quantum Monte Carlo package for the electronic structure of atoms, molecules and solids, *Journal of Physics: Condensed Matter*, **30**(19), 195901, (2018).

[154] Chien-Jung Huang, C. J. Umrigar, and M. P. Nightingale, Accuracy of electronic wave functions in quantum Monte Carlo: The effect of high-order correlations, *The Journal of Chemical Physics*, **107**(8), 3007–3013, (1997).

[155] Todd D. Beaudet, Michele Casula, Jeongnim Kim, Sandro Sorella, and Richard M. Martin, Molecular hydrogen adsorbed on benzene: Insights from a quantum Monte Carlo study, *Journal of Chemical Physics*, **129**(16), 164711, (2008).

[156] Sandro Sorella, Michele Casula, and Dario Rocca, Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods, *The Journal of Chemical Physics*, **127**(1), 014105, (2007).

[157] Fabio Sterpone, Leonardo Spanu, Luca Ferraro, Sandro Sorella, and Leonardo Guidoni, Dissecting the hydrogen bond: A quantum Monte Carlo approach, *Journal of Chemical Theory and Computation*, **4**(9), 1428–1434, (2008).

[158] Rüdiger Achilles and Andrea Bonfiglioli, The early proofs of the theorem of Campbell, Baker, Hausdorff, and Dynkin, *Archive for history of exact sciences*, **66**(3), 295–358, (2012).

[159] Peter A. Limacher, Paul W. Ayers, Paul A. Johnson, Stijn De Baerdemacker, Dimitri Van Neck, and Patrick Bultinck, A new mean-field method suitable for strongly correlated electrons: Computationally facile antisymmetric products of nonorthogonal geminals, *Journal of Chemical Theory and Computation*, **9**(3), 1394–1401, (2013).

[160] David W. Small, Keith V. Lawler, and Martin Head-Gordon, Coupled cluster valence bond method: Efficient computer implementation and application to multiple bond dissociations and strong correlations in the acenes, *Journal of Chemical Theory and Computation*, **10**(5), 2027–2040, (2014).

[161] Nicolas Dupuy, Samira Bouaouli, Francesco Mauri, Sandro Sorella, and Michele Casula, Vertical and adiabatic excitations in anthracene from quantum Monte Carlo: Constrained energy minimization for structural and electronic excited-state properties in the JAGP ansatz, *Journal of Chemical Physics*, **142**(21), (2015).

[162] Eric Neuscamman, Size consistency error in the antisymmetric geminal power wave function can be completely removed, *Physical Review Letters*, **109**(20), 203001, (2012).

[163] Eric Neuscamman, Improved optimization for the cluster jastrow antisymmetric geminal power and tests on triple-bond dissociations, *Journal of Chemical Theory and Computation*, **12**(7), 3149–3159, (2016).

[164] S.A. Baeurle, Computation within the auxiliary field approach, *Journal of Computational Physics*, **184**(2), 540–558, (2003).

[165] Kurt Binder, David M. Ceperley, J.-P. Hansen, M.H. Kalos, D.P. Landau, D. Levesque, H. Mueller-Krumbhaar, D. Stauffer, and J-J Weis, *Monte Carlo methods in statistical physics* volume 7, Springer Science & Business Media (2012).

[166] Paul Richard Charles Kent, *Techniques and applications of quantum Monte Carlo*, PhD thesis University of Cambridge (1999).

[167] Brian L. Hammond, William A. Lester, and Peter James Reynolds, *Monte Carlo methods in ab initio quantum chemistry* volume 1, World Scientific (1994).

[168] Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng, *Handbook of Markov chain Monte Carlo*, CRC press (2011).

[169] W.A. Majewski, The detailed balance condition in quantum statistical mechanics, *Journal of mathematical physics*, **25**(3), 614–616, (1984).

[170] Vasilios I. Manousiouthakis and Michael W. Deem, Strict detailed balance is unnecessary in Monte Carlo simulation, *The Journal of Chemical Physics*, **110**(6), 2753–2756, (1999).

[171] Siddhartha Chib and Edward Greenberg, Understanding the Metropolis-Hastings algorithm, *The american statistician*, **49**(4), 327–335, (1995).

[172] George S. Fishman, *Discrete-event simulation: modeling, programming, and analysis*, Springer Science & Business Media (2013).

[173] Martin Hohenadler and Thomas C. Lang, Autocorrelations in quantum monte carlo simulations of electron-phonon models, In *Computational Many-Particle Physics* , 357–366. Springer (2008).

[174] Jeongnim Kim, Andrew D. Baczewski, Todd D. Beaudet, Anouar Benali, M. Chandler Bennett, Mark A. Berrill, Nick S. Blunt, Edgar Josué Landinez Borda, Michele Casula, David M. Ceperley, et al., QMCPACK: an open source ab initio quantum Monte Carlo package for the electronic structure of atoms, molecules and solids, *Journal of Physics: Condensed Matter*, **30**(19), 195901, (2018).

[175] Paul R.C. Kent, Abdulgani Annaberdiyev, Anouar Benali, M. Chandler Bennett, Edgar Josué Landinez Borda, Peter Doak, Hongxia Hao, Kenneth D. Jordan, Jaron T. Krogel, Ilkka Kylänpää, et al., QMCPACK: Advances in the development, efficiency, and application of auxiliary field and real-space variational and diffusion quantum Monte Carlo, *The Journal of Chemical Physics*, **152**(17), 174105, (2020).

[176] Galin L. Jones et al., On the Markov chain central limit theorem, *Probability surveys*, **1**, 299–320, (2004).

[177] Roland Assaraf and Michel Caffarel, Zero-variance principle for Monte Carlo algorithms, *Physical Review Letters*, **83**(23), 4682, (1999).

[178] Enrique R. Batista, Jochen Heyd, Richard G. Hennig, Blas P. Uberuaga, Richard L. Martin, Gustavo E. Scuseria, C. J. Umrigar, and John W. Wilkins, Comparison of screened hybrid density functional theory to diffusion Monte Carlo in calculations of total energies of silicon phases and defects, *Physical Review B*, **74**, 121102, (2006).

[179] Jaehyung Yu, Lucas K. Wagner, and Elif Ertekin, Towards a systematic assessment of errors in diffusion Monte Carlo calculations of semiconductors: Case study of zinc selenide and zinc oxide, *The Journal of Chemical Physics*, **143**(22), 224707, (2015).

[180] Huihuo Zheng and Lucas K. Wagner, Computation of the correlated metal-insulator transition in vanadium dioxide from first principles, *Physical Review Letters*, **114**, 176401, (2015).

[181] J. Toulouse and C. J. Umrigar, Optimization of quantum Monte Carlo wave functions by energy minimization, *Journal of Chemical Physics*, **126**, 084102, (2007).

[182] J. Toulouse and C. J. Umrigar, Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules, *Journal of Chemical Physics*, **128**, 174101, (2008).

[183] Sandro Sorella, Green function Monte Carlo with stochastic reconfiguration, *Physical Review Letters*, **80**(20), 4558, (1998).

[184] Sandro Sorella and Luca Capriotti, Green function Monte Carlo with stochastic reconfiguration: An effective remedy for the sign problem, *Physical Review B*, **61**(4), 2599, (2000).

[185] Sandro Sorella, Generalized Lanczos algorithm for variational quantum Monte Carlo, *Physical Review B*, **64**(2), 024512, (2001).

[186] Sandro Sorella, Wave function optimization in the variational Monte Carlo method, *Physical Review B*, **71**(24), 241103, (2005).

[187] Lauretta R. Schwarz, Ali Alavi, and George H. Booth, Projector quantum Monte Carlo method for nonlinear wave functions, *Physical Review Letters*, **118**(17), 176403, (2017).

[188] Ankit Mahajan and Sandeep Sharma, Symmetry-projected Jastrow mean-field wave function in variational Monte Carlo, *The Journal of Physical Chemistry A*, **123**(17), 3911–3921, (2019).

[189] Iliya Sabzevari, Ankit Mahajan, and Sandeep Sharma, An accelerated linear method for optimizing non-linear wavefunctions in variational Monte Carlo, *The Journal of Chemical Physics*, **152**(2), 024111, (2020).

[190] Iliya Sabzevari and Sandeep Sharma, Improved speed and scaling in orbital space variational Monte Carlo, *Journal of Chemical Theory and Computation*, **14**(12), 6276–6286, (2018).

[191] Leon Otis and Eric Neuscamman, Complementary first and second derivative methods for ansatz optimization in variational Monte Carlo, *Physical Chemistry Chemical Physics*, **21**(27), 14491–14510, (2019).

[192] Leon Otis, Isabel Craig, and Eric Neuscamman, A hybrid approach to excited-state-specific variational Monte Carlo and doubly excited states, *The Journal of Chemical Physics*, **153**(23), 234105, (2020).

[193] Julien Toulouse and Cyrus J. Umrigar, Optimization of quantum Monte Carlo wave functions by energy minimization, *The Journal of Chemical Physics*, **126**(8), 084102, (2007).

[194] Luning Zhao and Eric Neuscamman, A blocked linear method for optimizing large parameter sets in variational Monte Carlo, *Journal of Chemical Theory and Computation*, **13**(6), 2604–2611, (2017).

[195] Claudia Filippi and C.J. Umrigar, Correlated sampling in quantum Monte Carlo: A route to forces, *Physical Review B*, **61**(24), R16291, (2000).

[196] Claudia Filippi and C.J. Umrigar, Interatomic forces and correlated sampling in quantum Monte Carlo, In *Recent Advances In Quantum Monte Carlo Methods Part II* , 12–29. World Scientific (2002).

[197] Sandro Sorella and Luca Capriotti, Algorithmic differentiation and the calculation of forces by quantum Monte Carlo, *The Journal of Chemical Physics*, **133**(23), 234111, (2010).

[198] Hao Shi and Shiwei Zhang, Infinite variance in fermion quantum Monte Carlo calculations, *Physical Review E*, **93**(3), 033303, (2016).

[199] Sergio D. Pineda Flores and Eric Neuscamman, Excited state specific multi-Slater Jastrow wave functions, *The Journal of Physical Chemistry A*, **123**(8), 1487–1497, (2019).

[200] Paul J. Robinson, Sergio D. Pineda Flores, and Eric Neuscamman, Excitation variance matching with limited configuration interaction expansions in variational Monte Carlo, *The Journal of Chemical Physics*, **147**(16), 164114, (2017).

[201] N.D. Drummond and R.J. Needs, Variance-minimization scheme for optimizing Jastrow factors, *Physical Review B*, **72**(8), 085124, (2005).

[202] C. J. Umrigar and Claudia Filippi, Energy and Variance Optimization of Many-Body Wave Functions, *Physical Review Letters*, **94**, 150201, (2005).

[203] Luning Zhao and Eric Neuscamman, An efficient variational principle for the direct optimization of excited states, *Journal of Chemical Theory and Computation*, **12**(8), 3436–3440, (2016).

[204] R.M. Fye, New results on Trotter-like approximations, *Physical Review B*, **33**(9), 6271, (1986).

[205] Lauri Lehtovaara, Jari Toivanen, and Jussi Eloranta, Solution of time-independent Schrödinger equation by the imaginary time propagation method, *Journal of Computational Physics*, **221**(1), 148–157, (2007).

[206] Michal Bajdich, Murilo L. Tiago, Randolph Q. Hood, Paul R.C. Kent, and Fernando A. Reboredo, Systematic reduction of sign errors in many-body calculations of atoms and molecules, *Physical Review Letters*, **104**(19), 193001, (2010).

[207] Lubos Mitas, Electronic structure by quantum Monte Carlo: atoms, molecules and solids, *Computer Physics Communications*, **96**(2-3), 107–117, (1996).

[208] Cody A. Melton and Lubos Mitas, Quantum Monte Carlo with variable spins: Fixed-phase and fixed-node approximations, *Physical Review E*, **96**(4), 043305, (2017).

[209] James B. Anderson, Quantum chemistry by random walk. H ^{2}P, H^+_3 D_{3h} ^{1} A' $_1$, H_2 ^{3} Σ^+_u, H_4 ^{1} Σ^+_g, Be ^{1}S, *The Journal of Chemical Physics*, **65**(10), 4121–4127, (1976).

[210] David M. Ceperley, Fermion nodes, *Journal of statistical physics*, **63**(5-6), 1237–1267, (1991).

[211] Lubos Mitas, Structure of fermion nodes and nodal cells, *Physical Review Letters*, **96**(24), 240402, (2006).

[212] K.M. Rasch and L. Mitas, Impact of electron density on the fixed-node errors in Quantum Monte Carlo of atomic systems, *Chemical Physics Letters*, **528**, 59–62, (2012).

[213] A.C. Calder, M.R. Curry, R.M. Panoff, and Y.J. Wong, Visualization of the local contribution to the nodal surface of a many-fermion wave function, *Physical Review E*, **53**(5), 5450, (1996).

[214] Michal Bajdich, Lubos Mitas, Gabriel Drobnỳ, and Lucas K. Wagner, Approximate and exact nodes of fermionic wavefunctions: coordinate transformations and topologies, *Physical Review B*, **72**(7), 075131, (2005).

[215] Pierre-François Loos and Dario Bressanini, Nodal surfaces and interdimensional degeneracies, *The Journal of Chemical Physics*, **142**(21), 214112, (2015).

[216] Kevin M. Rasch, Shuming Hu, and Lubos Mitas, Communication: Fixed-node errors in quantum Monte Carlo: Interplay of electron density and node nonlinearities (2014).

[217] Fernando A. Reboredo, Randolph Q. Hood, and Paul R.C. Kent, Self-healing diffusion quantum Monte Carlo algorithms: Direct reduction of the fermion sign error in electronic structure calculations, *Physical Review B*, **79**(19), 195117, (2009).

[218] Anouar Benali, Kevin Gasperich, Kenneth D. Jordan, Thomas Applencourt, Ye Luo, M. Chandler Bennett, Jaron T. Krogel, Luke Shulenburger, Paul R.C. Kent, Pierre-François Loos, et al., Toward a systematic improvement of the fixed-node approximation in diffusion Monte Carlo for solids - A case study in diamond, *The Journal of Chemical Physics*, **153**(18), 184111, (2020).

[219] Andrew C. Hurley, John Edward Lennard-Jones, and John Anthony Pople, The molecular orbital theory of chemical valency XVI. A theory of paired-electrons in polyatomic molecules, *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, **220**(1143), 446–455, (1953).

[220] Eric Neuscamman, The Jastrow antisymmetric geminal power in Hilbert space: Theory, benchmarking, and application to a novel transition state, *The Journal of Chemical Physics*, **139**(19), 194105, (2013).

[221] Eric Neuscamman, Subtractive manufacturing with geminal powers: making good use of a bad wave function, *Molecular Physics*, **114**(5), 577–583, (2016).

[222] Eric Neuscamman, Communication: A Jastrow factor coupled cluster theory for weak and strong electron correlation, *Journal of Chemical Physics*, **139**, 181101, (2013).

[223] Eric Neuscamman, Improved optimization for the cluster Jastrow antisymmetric geminal power and tests on triple-bond dissociations, *Journal of Chemical Theory and Computation*, **12**(7), 3149–3159, (2016).

[224] Emanuele Coccia, Daniele Varsano, and Leonardo Guidoni, Ab initio geometry and bright excitation of carotenoids: Quantum Monte Carlo and many body Green's function theory calculations on peridinin, *Journal of Chemical Theory and Computation*, **10**(2), 501–506, (2014).

[225] Sandro Sorella, Michele Casula, and Dario Rocca, Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods, *Journal of Chemical Physics*, **127**(1), 014105, (2007).

[226] Jeongnim Kim, Kenneth P. Esler, Jeremy McMinis, Miguel A. Morales, Bryan K. Clark, Luke Shulenburger, and David M. Ceperley, Hybrid algorithms in quantum Monte Carlo, In *Journal of Physics: Conference Series* volume 402 , 012008. IOP Publishing (2012).

[227] M. P. Nightingale and V. Melik-Alaverdian, Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters, *Physical Review Letters*, **87**, 043401, (2001).

[228] C. J. Umrigar, J. Toulouse, C. Filippi, S. Sorella, and R. G. Hennig, Alleviation of the Fermion-Sign Problem by Optimization of Many-Body Wave Functions, *Physical Review Letters*, **98**, 110201, (2007).

[229] Andreas Bergner, Michael Dolg, Wolfgang Küchle, Hermann Stoll, and Heinzwerner Preub, Ab initio energy-adjusted pseudopotentials for elements of groups 13-17, *Molecular Physics*, **80**(6), 1431–1441, (1993).

[230] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, Theresa L. Windus, Michel Dupuis, and John A. Montgomery, General atomic and molecular electronic structure system, *J. Comput. Chem.*, **14**(11), 1347–1363, (1993).

[231] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, and M. Schütz, Molpro: a general-purpose quantum chemistry program package, *WIREs Comput Mol Sci*, **2**, 242–253, (2012).

[232] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby, M. Schütz, P. Celani, W. Györffy, D. Kats, T. Korona, R. Lindh, A. Mitrushenkov, G. Rauhut, K. R. Shamasundar, T. B. Adler, R. D. Amos, S. J. Bennie, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, E. Goll, C. Hampel, A. Hesselmann, G. Hetzer, T. Hrenar, G. Jansen, C. Köppl, S. J. R. Lee, Y. Liu, A. W. Lloyd, Q. Ma, R. A. Mata, A. J. May, S. J. McNicholas, W. Meyer, T. F. Miller III, M. E. Mura, A. Nicklass, D. P. O'Neill, P. Palmieri, D. Peng, K. Pflüger, R. Pitzer, M. Reiher, T. Shiozaki, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, M. Wang, and M. Welborn, MOLPRO, version 2018.1, a package of ab initio programs (2018), see http://www.molpro.net.

[233] Roland Lindh, U. Ryu, and B. Liu, The reduced multiplication scheme of the Rys quadrature and new recurrence relations for auxiliary function based two-electron integral evaluation, *The Journal of Chemical Physics*, **95**(8), 5889–5897, (1991).

[234] Claudia Hampel, Kirk A. Peterson, and Hans-Joachim Werner, A comparison of the efficiency and accuracy of the quadratic configuration interaction (QCISD), coupled cluster (CCSD), and Brueckner coupled cluster (BCCD) methods, *Chemical Physics Letters*, **190**(1-2), 1–12, (1992).

[235] Claudia Knowles, Peter J .and Hampel and Hans-Joachim Werner, Coupled cluster theory for high spin, open shell reference wave functions, *The Journal of Chemical Physics*, **99**(7), 5219–5227, (1993).

[236] Miles J.O. Deegan and Peter J. Knowles, Perturbative corrections to account for triple excitations in closed and open shell coupled cluster theories, *Chemical Physics Letters*, **227**(3), 321–326, (1994).

[237] Hans-Joachim Werner and Peter J. Knowles, A second order multiconfiguration SCF procedure with optimum convergence, *The Journal of Chemical Physics*, **82**(11), 5053–5063, (1985).

[238] Peter J. Knowles and Hans-Joachim Werner, An efficient second-order MC SCF method for long configuration expansions, *Chemical Physics Letters*, **115**(3), 259–267, (1985).

[239] Hans-Joachim Werner and Peter J. Knowles, An efficient internally contracted multi-configurationâĂŞreference configuration interaction method, *The Journal of Chemical Physics*, **89**(9), 5803–5814, (1988).

[240] Peter J. Knowles and Hans-Joachim Werner, An efficient method for the evaluation of coupling coefficients in configuration interaction calculations, *Chemical Physics Letters*, **145**(6), 514–522, (1988).

[241] K.R. Shamasundar, Gerald Knizia, and Hans-Joachim Werner, A new internally contracted multi-reference configuration interaction method, *The Journal of Chemical Physics*, **135**(5), 054101, (2011).

[242] Axel D. Becke, Density functionals for static, dynamical, and strong correlation, *The Journal of Chemical Physics*, **138**(7), 074109, (2013).

[243] V. Adrian Parsegian, *Van der Waals forces: a handbook for biologists, chemists, engineers, and physicists*, Cambridge University Press (2005).

[244] Sebastian Keller, Katharina Boguslawski, Tomasz Janowski, Markus Reiher, and Peter Pulay, Selection of active spaces for multiconfigurational wavefunctions, *The Journal of Chemical Physics*, **142**(24), 244104, (2015).

[245] Steven Vancoillie, Per Åke Malmqvist, and Valera Veryazov, Potential Energy Surface of the Chromium Dimer Re-re-revisited with Multiconfigurational Perturbation Theory, *Journal of Chemical Theory and Computation*, **12**(4), 1647–1655, (2016).

[246] Yuki Kurashige, Garnet Kin-Lic Chan, and Yanai Takeshi, Entangled quantum electronic wavefunctions of the Mn4CaO5 cluster in photosystem II, *Nature Chemistry*, **5**, 662–666, (2013).

[247] Joonho Lee, David W. Small, and Martin Head-Gordon, Open-shell coupled-cluster valence-bond theory augmented with an independent amplitude approximation for three-pair correlations: Application to a model oxygen-evolving complex and single molecular magnet, *arXiv preprint arXiv:1808.06743*, (2018).

[248] D. R. Hartree, F. R. S., W. Hartree, and Bertha Swirles, Self-consistent field, including exchange and superposition of configurations, with some results for oxygen, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, **238**(790), 229–247, (1939).

[249] Steven R. White, Density-matrix algorithms for quantum renormalization groups, *Physical Review B*, **48**, 10345–10356, (1993).

[250] James E. T. Smith, Bastien Mussard, Adam A. Holmes, and Sandeep Sharma, Cheap and Near Exact CASSCF with Large Active Spaces, *Journal of Chemical Theory and Computation*, **13**(11), 5468–5478, (2017).

[251] Sandeep Sharma, Adam A. Holmes, Guillaume Jeanmairet, Ali Alavi, and C. J. Umrigar, Semistochastic Heat-Bath Configuration Interaction Method: Selected Configuration Interaction with Semistochastic Perturbation Theory, *Journal of Chemical Theory and Computation*, **13**(4), 1595–1604, (2017).

[252] Emmanuel Giner, Anthony Scemama, and Michel Caffarel, Using perturbatively selected configuration interaction in quantum Monte Carlo calculations, *Canadian Journal of Chemistry*, **91**(9), 879–885, (2013).

[253] Emmanuel Giner, Anthony Scemama, and Michel Caffarel, Fixed-node diffusion Monte Carlo potential energy curve of the fluorine molecule F2 using selected configuration interaction trial wavefunctions, *The Journal of Chemical Physics*, **142**(4), 044115, (2015).

[254] Brian M. Austin, Dmitry Yu. Zubarev, and William A. Lester, Quantum Monte Carlo and Related Approaches, *Chemical Reviews*, **112**(1), 263–288, (2012).

[255] Arne Lüchow, Alexander Sturm, Christoph Schulte, and Kaveh Haghighi Mood, Generic expansion of the Jastrow correlation factor in polynomials satisfying symmetry and cusp conditions, *The Journal of Chemical Physics*, **142**(8), 084111, (2015).

[256] P. López Ríos, P. Seth, N. D. Drummond, and R. J. Needs, Framework for constructing generic Jastrow correlation factors, *Physical Review E*, **86**, 036703, (2012).

[257] Michele Casula and Sandro Sorella, Geminal wave functions with Jastrow correlation: A first application to atoms, *The Journal of Chemical Physics*, **119**(13), 6500–6511, (2003).

[258] Michele Casula, Claudio Attaccalite, and Sandro Sorella, Correlated geminal wave function for molecules: An efficient resonating valence bond approach, *The Journal of Chemical Physics*, **121**(15), 7110–7126, (2004).

[259] Todd D. Beaudet, Michele Casula, Jeongnim Kim, Sandro Sorella, and Richard M. Martin, Molecular hydrogen adsorbed on benzene: Insights from a quantum Monte Carlo study, *The Journal of Chemical Physics*, **129**(16), 164711, (2008).

[260] Matteo Barborini, Sandro Sorella, and Leonardo Guidoni, Structural Optimization by Quantum Monte Carlo: Investigating the Low-Lying Excited States of Ethylene, *Journal of Chemical Theory and Computation*, **8**(4), 1260–1269, (2012).

[261] Andrea Zen, Ye Luo, Guglielmo Mazzola, Leonardo Guidoni, and Sandro Sorella, Ab initio molecular dynamics simulation of liquid water by quantum Monte Carlo, *The Journal of Chemical Physics*, **142**(14), 144111, (2015).

[262] Beatrice Van Der Goetz and Eric Neuscamman, Suppressing Ionic Terms with Number-Counting Jastrow Factors in Real Space, *Journal of Chemical Theory and Computation*, **13**(5), 2035–2042, (2017).

[263] James Albert Sethian, *Level set methods and fast marching methods: evolving interfaces in computational geometry, fluid mechanics, computer vision, and materials science* volume 3, Cambridge university press (1999).

[264] Geoffrey McLachlan, *Discriminant analysis and statistical pattern recognition* volume 544, John Wiley & Sons (2004).

[265] Geoffrey J. McLachlan and Kaye E. Basford, *Mixture models: Inference and applications to clustering* volume 84, Marcel Dekker (1988).

[266] Franz Aurenhammer, Rolf Klein, and Der-Tsai Lee, *Voronoi diagrams and Delaunay triangulations*, World Scientific Publishing Company (2013).

[267] Michael W. Schmidt, Kim K. Baldridge, Jerry A. Boatz, Steven T. Elbert, Mark S. Gordon, Jan H. Jensen, Shiro Koseki, Nikita Matsunaga, Kiet A. Nguyen, Shujun Su, et al., General atomic and molecular electronic structure system, *Journal of computational chemistry*, **14**(11), 1347–1363, (1993).

[268] Mark S. Gordon and Michael W. Schmidt, Chapter 41 - Advances in electronic structure theory: GAMESS a decade later, In Clifford E. Dykstra, Gernot Frenking, Kwang S. Kim, and Gustavo E. Scuseria, editors, *Theory and Applications of Computational Chemistry* , 1167 – 1189. Elsevier Amsterdam (2005).

[269] Brian L. Hammond, Peter J. Reynolds, and William A. Lester, Valence quantum Monte Carlo with ab initio effective core potentials, *The Journal of Chemical Physics*, **87**(2), 1130–1136, (1987).

[270] Andreas Bergner, Michael Dolg, Wolfgang Küchle, Hermann Stoll, and Heinzwerner Preuß, Ab initio energy-adjusted pseudopotentials for elements of groups 13–17, *Molecular Physics*, **80**(6), 1431–1441, (1993).

[271] Martin Kaupp, P. v. R. Schleyer, H. Stoll, and H. Preuss, Pseudopotential approaches to Ca, Sr, and Ba hydrides. Why are some alkaline earth MX2 compounds bent?, *The Journal of Chemical Physics*, **94**(2), 1360–1366, (1991).

[272] Julien Toulouse and C. J. Umrigar, Optimization of quantum Monte Carlo wave functions by energy minimization, *The Journal of Chemical Physics*, **126**(8), 084102, (2007).

[273] Julien Toulouse and C. J. Umrigar, Full optimization of Jastrow-Slater wave functions with application to the first-row atoms and homonuclear diatomic molecules, *The Journal of Chemical Physics*, **128**(17), 174101, (2008).

[274] Roy E. Kari, Paul G. Mezey, and Imre G. Csizmadia, Quality of Gaussian basis sets: Direct optimization of orbital exponents by the method of conjugate gradients, *The Journal of Chemical Physics*, **63**(1), 581–585, (1975).

[275] Luboš Mitáš, Eric L. Shirley, and David M. Ceperley, Nonlocal pseudopotentials and diffusion Monte Carlo, *The Journal of Chemical Physics*, **95**(5), 3467–3475, (1991).

[276] Michele Casula, Beyond the locality approximation in the standard diffusion Monte Carlo method, *Physical Review B*, **74**, 161102, (2006).

[277] Michele Casula, Saverio Moroni, Sandro Sorella, and Claudia Filippi, Size-consistent variational approaches to nonlocal pseudopotentials: Standard and lattice regularized diffusion Monte Carlo methods revisited, *The Journal of Chemical Physics*, **132**(15), 154113, (2010).

[278] Roman Nazarov, Luke Shulenburger, M. Morales, and Randolph Q. Hood, Benchmarking the pseudopotential and fixed-node approximations in diffusion Monte Carlo calculations of molecules and solids, *Physical Review B*, **93**(9), 094111, (2016).

[279] Roland Assaraf, S. Moroni, and Claudia Filippi, Optimizing the energy with quantum Monte Carlo: A lower numerical scaling for Jastrow-Slater expansions, *Journal of Chemical Theory and Computation*, **13**(11), 5273–5281, (2017).

[280] Arieh Warshel, Bicycle-pedal model for the first step in the vision process, *Nature*, **260**(5553), 679–683, (1976).

[281] I.N. Obotowo, I.B. Obot, and U.J. Ekpe, Organic sensitizers for dye-sensitized solar cell (DSSC): Properties from computation, progress and future perspectives, *Journal of Molecular Structure*, **1122**, 80–87, (2016).

[282] Wenjie Fan and Wei-qiao Deng, Incorporation of thiadiazole derivatives as π-spacer to construct efficient metal-free organic dye sensitizers for dye-sensitized solar cells: a theoretical study, *Computational Chemistry*, **1**, 152–170, (2013).

[283] Patrick Norman and Andreas Dreuw, Simulating X-ray Spectroscopies and Calculating Core-Excited States of Molecules, *Chemical Reviews*, **118**(15), 7208–7248, (2018).

[284] Ryan M. Richard and John M. Herbert, Time-Dependent Density-Functional Description of the 1La State in Polycyclic Aromatic Hydrocarbons: Charge-Transfer Character in Disguise?, *Journal of Chemical Theory and Computation*, **7**(5), 1296–1306, (2011).

[285] Marco Campetella, Federica Maschietto, Mike J. Frisch, Giovanni Scalmani, Ilaria Ciofini, and Carlo Adamo, Charge transfer excitations in TDDFT: A ghost-hunter index, *Journal of Computational Chemistry*, **38**(25), 2151–2156, (2017).

[286] Bushra Alam, Adrian F. Morrison, and John M. Herbert, Charge Separation and Charge Transfer in the Low-Lying Excited States of Pentacene, *The Journal of Physical Chemistry C*, **124**(45), 24653–24666, (2020).

[287] Joseph E. Subotnik, Communication: Configuration interaction singles has a large systematic bias against charge-transfer states, *The Journal of Chemical Physics*, **135**(7), 071104, (2011).

[288] Andrew T. B. Gilbert, Nicholas A. Besley, and Peter M. W. Gill, Self-consistent field calculations of excited states using the maximum overlap method (MOM), *The Journal of Physical Chemistry A*, **112**(50), 13164–13171, (2008).

[289] Diptarka Hait and Martin Head-Gordon, Excited state orbital optimization via minimizing the square of the gradient: General approach and application to singly and doubly excited states via density functional theory, *Journal of Chemical Theory and Computation*, **16**(3), 1699–1710, (2020).

[290] Kevin Carter-Fenk and John M. Herbert, State-targeted energy projection: A simple and robust approach to orbital relaxation of non-Aufbau self-consistent field solutions, *Journal of Chemical Theory and Computation*, **16**(8), 5067–5082, (2020).

[291] Hong-Zhou Ye and Troy Van Voorhis, Half-projected σ self-consistent field for electronic excited states, *Journal of Chemical Theory and Computation*, **15**(5), 2954–2965, (2019).

[292] Tjalling J. Ypma, Historical development of the Newton–Raphson method, *SIAM review*, **37**(4), 531–551, (1995).

[293] Janusz S. Kowalik and Michael Robert Osborne, *Methods for unconstrained optimization problems*, North-Holland (1968).

[294] Jutho Haegeman, Tobias J. Osborne, and Frank Verstraete, Post-matrix product state methods: To tangent space and beyond, *Physical Review B*, **88**, 075133, (2013).

[295] Laurens Vanderstraeten, Michaël Mariën, Frank Verstraete, and Jutho Haegeman, Excitations and the tangent space of projected entangled-pair states, *Physical Review B*, **92**, 201111, (2015).

[296] James B. Foresman, Martin Head-Gordon, John A. Pople, and Michael J. Frisch, Toward a systematic molecular orbital theory for excited states, *The Journal of Physical Chemistry*, **96**(1), 135–149, (1992).

[297] A. D. McLACHLAN and M. A. BALL, Time-Dependent Hartree–Fock Theory for Molecules, *Rev. Mod. Phys.*, **36**, 844–855, (1964).

[298] Felix Plasser, Michael Wormit, and Andreas Dreuw, New tools for the systematic analysis and visualization of electronic excitations. I. Formalism, *The Journal of Chemical Physics*, **141**(2), 024106, (2014).

[299] I. Mayer, Using singular value decomposition for a compact presentation and improved interpretation of the CIS wave functions, *Chemical Physics Letters*, **437**(4), 284–286, (2007).

[300] Richard L. Martin, Natural transition orbitals, *The Journal of Chemical Physics*, **118**(11), 4775–4777, (2003).

[301] A.V. Luzanov, A.A. Sukhorukov, and V.E. Umanskii, Application of transition density matrix for analysis of excited states, *Theoretical and Experimental Chemistry*, **10**(4), 354–361, (1976).

[302] Katherine J. Oosterbaan, Alec F. White, and Martin Head-Gordon, Non-orthogonal configuration interaction with single substitutions for the calculation of core-excited states, *The Journal of Chemical Physics*, **149**(4), 044116, (2018).

[303] Xinle Liu, Shervin Fatehi, Yihan Shao, Brad S. Veldkamp, and Joseph E. Subotnik, Communication: Adjusting charge transfer state energies for configuration interaction singles: Without any parameterization and with minimal cost, *The Journal of Chemical Physics*, **136**(16), 161101, (2012).

[304] Scott M. Garner and Eric Neuscamman, Core excitations with excited state mean field and perturbation theory, *The Journal of Chemical Physics*, **153**(15), 154102, (2020).

[305] Robert G. Parr and Weitao Yang, *Density-functional theory of atoms and molecules*, Oxford University Press New York (1989).

[306] Andreas Dreuw, Jennifer L. Weisman, and Martin Head-Gordon, Long-range charge-transfer excited states in time-dependent density functional theory require non-local exchange, *The Journal of Chemical Physics*, **119**(6), 2943–2946, (2003).

[307] Hiroya Nitta and Isao Kawata, A close inspection of the charge-transfer excitation by TDDFT with various functionals: An application of orbital-and density-based analyses, *Chemical Physics*, **405**, 93–99, (2012).

[308] Janus J. Eriksen, Stephan P.A. Sauer, Kurt V. Mikkelsen, Ove Christiansen, Hans Jørgen Aa Jensen, and Jacob Kongsted, Failures of TDDFT in describing the lowest intramolecular charge-transfer excitation in para-nitroaniline, *Molecular Physics*, **111**(9-11), 1235–1248, (2013).

[309] Yoshihiro Tawada, Takao Tsuneda, Susumu Yanagisawa, Takeshi Yanai, and Kimihiko Hirao, A long-range-corrected time-dependent density functional theory, *The Journal of Chemical Physics*, **120**(18), 8425–8433, (2004).

[310] Osamu Takahashi and Lars G.M. Pettersson, Functional dependence of core-excitation energies, *The Journal of Chemical Physics*, **121**(21), 10339–10345, (2004).

[311] Yutaka Imamura, Takao Otsuka, and Hiromi Nakai, Description of core excitations by time-dependent density functional theory with local density approximation, generalized gradient approximation, meta-generalized gradient approximation, and hybrid functionals, *Journal of computational chemistry*, **28**(12), 2067–2074, (2007).

[312] M.G.U.J. Petersilka, U.J. Gossmann, and E.K.U. Gross, Excitation energies from time-dependent density-functional theory, *Physical Review Letters*, **76**(8), 1212, (1996).

[313] Tom Ziegler, Michael Seth, Mykhaylo Krykunov, and Jochen Autschbach, A revised electronic Hessian for approximate time-dependent density functional theory, *The Journal of Chemical Physics*, **129**(18), 184114, (2008).

[314] Tom Ziegler, Michael Seth, Mykhaylo Krykunov, Jochen Autschbach, and Fan Wang, On the relation between time-dependent and variational density functional theory approaches for the determination of excitation energies and transition moments, *The Journal of Chemical Physics*, **130**(15), 154102, (2009).

[315] Luning Zhao and Eric Neuscamman, Density functional extension to excited-state mean-field theory, *Journal of Chemical Theory and Computation*, **16**(1), 164–178, (2019).

[316] Hong-Zhou Ye and Troy Van Voorhis, Half-projected σ self-consistent field for electronic excited states, *Journal of Chemical Theory and Computation*, **15**(5), 2954–2965, (2019).

[317] Giuseppe M. J. Barca, Andrew T. B. Gilbert, and Peter M. W. Gill, Simple models for difficult electronic excitations, *Journal of Chemical Theory and Computation*, **14**(3), 1501–1509, (2018).

[318] Diptarka Hait, Eric A Haugen, Zheyue Yang, Katherine J Oosterbaan, Stephen R Leone, and Martin Head-Gordon, Accurate prediction of core-level spectra of radicals at density functional theory cost via square gradient minimization and recoupling of mixed configurations, *The Journal of Chemical Physics*, **153**(13), 134108, (2020).

[319] Andreas Dreuw and Michael Wormit, The algebraic diagrammatic construction scheme for the polarization propagator for the calculation of excited states, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, **5**(1), 82–95, (2015).

[320] Rachel Clune, Jacqueline A. R. Shea, and Eric Neuscamman, N5-Scaling excited-state-specific perturbation theory, *Journal of Chemical Theory and Computation*, **16**(10), 6132–6141, (2020).

[321] Peter Pulay, Improved SCF convergence acceleration, *Journal of Computational Chemistry*, **3**(4), 556–560, (1982).

[322] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al., Tensorflow: Large-scale machine learning on heterogeneous distributed systems, *arXiv preprint arXiv:1603.04467*, (2016).

[323] Troy Van Voorhis and Martin Head-Gordon, A geometric approach to direct minimization, *Molecular Physics*, **100**(11), 1713–1721, (2002).

[324] Lianjun Jiang, Richard H Byrd, Elizabeth Eskow, and Robert B. Schnabel, A preconditioned L-BFGS algorithm with application to molecular energy minimization, Technical report COLORADO UNIV AT BOULDER DEPT OF COMPUTER SCIENCE (2004).

[325] Wenyong Pan, Kristopher A. Innanen, and Wenyuan Liao, Accelerating Hessian-free Gauss-Newton full-waveform inversion via L-BFGS preconditioned conjugate-gradient algorithm, *Geophysics*, **82**(2), R49–R64, (2017).

[326] Diptarka Hait and Martin Head-Gordon, Orbital optimized density functional theory for electronic excited states, *The Journal of Physical Chemistry Letters*, **12**, 4517–4529, (2021).

[327] Peter Lancaster and Miron Tismenetsky, *The Theory of Matrices: with Applications*, Elsevier (1985).

[328] David J. Thouless, *The quantum mechanics of many-body systems*, Courier Corporation (2014).

[329] Ravinder Rupchand Puri, Algebra of the exponential operator, In *Mathematical Methods of Quantum Optics* , 37–53. Springer (2001).

[330] Philip Wolfe, Convergence conditions for ascent methods, *SIAM review*, **11**(2), 226–235, (1969).

[331] Philip Wolfe, Convergence conditions for ascent methods. II: Some corrections, *SIAM review*, **13**(2), 185–188, (1971).

[332] John C. Tully, Perspective: Nonadiabatic dynamics theory, *The Journal of Chemical Physics*, **137**(22), 22A301, (2012).

A Linear Dependencies

The Linear Method[272,273] is a wavefunction minimization algorithm that begins by expanding the wavefunction into its first-order derivative, or tangent, space:

$$|\Psi_{lin}\rangle = d_0 \, |\Psi(\boldsymbol{p})\rangle + \sum_i d_i \, |\frac{\partial \Psi(\boldsymbol{p})}{\partial p_i}\rangle \tag{A.1}$$

As wavefunctions at an energy minimum will necessarily have a vanishing gradient, we solve for the linear coefficients d_i by setting:

$$\nabla_{\boldsymbol{d}} \left(\frac{\langle \Psi_{lin}| \, \boldsymbol{H} \, |\Psi_{lin}\rangle}{\langle \Psi_{lin}| \, \Psi_{lin}\rangle} \right) = 0 \tag{A.2}$$

which becomes a generalized eigenvalue equation whose lowest-energy eigenvector is used to update the wavefunction:

$$\boldsymbol{H}^{lin}\boldsymbol{d} = E^{lin}\boldsymbol{S}^{lin}\boldsymbol{d}, \quad |\Psi(\boldsymbol{p})\rangle \rightarrow |\Psi(\boldsymbol{p}+\boldsymbol{d})\rangle \tag{A.3}$$

Matrix elements are given by the expectation values of tangent space components:

$$\begin{aligned} \boldsymbol{H}^{lin}_{ij} &= \left\langle \frac{\partial \Psi(\boldsymbol{p})}{\partial p_i} \middle| \, \boldsymbol{H} \, \middle| \frac{\partial \Psi(\boldsymbol{p})}{\partial p_j} \right\rangle \\ \boldsymbol{S}^{lin}_{ij} &= \left\langle \frac{\partial \Psi(\boldsymbol{p})}{\partial p_i} \middle| \, \frac{\partial \Psi(\boldsymbol{p})}{\partial p_j} \right\rangle \end{aligned} \tag{A.4}$$

As often happens in numerical linear algebra, the solutions to this generalized eigenvalue equation become ill-conditioned as the eigenvalues of the overlap matrix $\boldsymbol{S}^{lin}$ approach zero. The numerical properties of the overlap matrix are strongly intertwined with the basis used to construct it, and becomes singular whenever this basis is not linearly independent. We will show exactly how redundant representations of the wavefunction in parameter space lead to linear dependencies in its tangent space basis, creating a singular overlap matrix. Though the normalization condition is the source of many of the formal benefits enjoyed by the basis introduced in section 3.4, it also produces exactly this type of parameter redundancy to the counting Jastrow factor. As a result, the identification and removal of excess degrees of freedom is crucial to obtain numerically well-behaved linear method equations when optimizing these counting Jastrow factors.

In this section, Φ represents a mapping from parameter space to the wavefunction Hilbert space:

$$\Phi : \mathbb{R}^n \rightarrow \mathbf{H} \tag{A.5}$$

First, we will assume that this mapping is continuous, and that there exists a continuous, nontrivial 'equivalence path' $\boldsymbol{p}$ along which Φ maps to the same wavefunction to within a normalization factor. For example, for:

$$\langle x|\Phi(a,b)\rangle = \exp\left(-abx^2\right) \tag{A.6}$$

this path would be:

$$\boldsymbol{p}(\alpha) = ab \begin{bmatrix} \alpha \\ 1/\alpha \end{bmatrix}, \quad \alpha \in (0,\infty) \tag{A.7}$$

so that:

$$\Phi(\boldsymbol{p}(1)) \equiv \Phi(\boldsymbol{p}(\alpha)), \text{ for all } \alpha \in (0, \infty) \tag{A.8}$$

In general, the derivative of this wavefunction mapping with respect to the path coordinate α of any such path is the zero vector in $\mathcal{H}$:

$$\frac{d\Phi(\boldsymbol{p}(\alpha))}{d\alpha} = \lim_{\Delta\alpha \to 0} \frac{\Phi(\boldsymbol{p}(\alpha + \Delta\alpha) - \Phi(\boldsymbol{p}(\alpha))}{\Delta\alpha} = \mathbf{0} \tag{A.9}$$

Expanding this derivative using the multivariate chain rule produces the linear dependency:

$$\mathbf{0} = \frac{d\Phi(\boldsymbol{p}(\alpha))}{d\alpha} = \sum_i \frac{\partial\Phi(\boldsymbol{p}(\alpha))}{\partial p_i} \frac{dp_i(\alpha)}{d\alpha} \tag{A.10}$$

As this linear combination is equal to the zero function, projecting any vector in the tangent space onto it produces a value of zero:

$$\left\langle \frac{\partial\Phi(\boldsymbol{p})}{\partial p_k} \middle| \sum_i \frac{\partial\Phi(\boldsymbol{p}(\alpha))}{\partial p_i} p_i'(\alpha) \right\rangle = 0 \tag{A.11}$$

This is equivalent to the action of the overlap matrix on the vector $\boldsymbol{p}'(\alpha)$:

$$\boldsymbol{S}^{lin}\boldsymbol{p}'(\alpha) = \mathbf{0} \tag{A.12}$$

showing that the overlap matrix is singular. Removing these linear dependencies can be done by identifying a formula for the coefficients of $\boldsymbol{p}'(\alpha)$, integrating to find the path $\boldsymbol{p}(\alpha)$, and restricting parameters to stop variations along the path-coordinate α. Once done, this removes the offending vectors from the linear method tangent space.

Our counting Jastrow factor is given by equation 3.1 with a normalized counting function basis from 3.12 and is written:

$$\Phi\left(\boldsymbol{F}, \boldsymbol{G}, \{g_I\}\right) = e^{J_C} \tag{A.13}$$

$$J_C = \sum_{ijIJ} F_{IJ} C_I(\boldsymbol{r}_i) C_J(\boldsymbol{r}_j) + \sum_{iI} G_I C_I(\boldsymbol{r}_i) \tag{A.14}$$

$$C_I(\boldsymbol{r}_i) = \frac{g_I(\boldsymbol{r}_i)}{\sum_J g_J(\boldsymbol{r}_i)} \tag{A.15}$$

$$g_I(\boldsymbol{r}_i) = \exp\left(\boldsymbol{r}^T \boldsymbol{A}_I \boldsymbol{r} - 2\boldsymbol{B}_I^T \boldsymbol{r} + K_I\right) \tag{A.16}$$

The normalization condition in the definition of C_I can be used to directly construct these equivalence paths for the $\boldsymbol{F}$ and $\boldsymbol{G}$ parameters. We first note that these parameter derivatives are given by:

$$\frac{\partial\Phi}{\partial G_I} = \sum_i C_I(\boldsymbol{r}_i)\Phi, \quad \frac{\partial\Psi}{\partial F_{IJ}} = \sum_{ij} C_I(\boldsymbol{r}_i) C_J(\boldsymbol{r}_j)\Phi \tag{A.17}$$

The sum of normalized counting function values for a single coordinate is always one, so the sum:

$$\sum_i \left(\sum_I C_I(\boldsymbol{r}_i) \right) = n_e \tag{A.18}$$

is the total number of electrons n_e. Consequently:

$$\sum_{IJ} \frac{\partial \Phi}{\partial F_{IJ}} - n_e^2 \Phi = \mathbf{0} \tag{A.19}$$

$$\sum_{J} \left(\frac{\partial \Phi}{\partial F_{IJ}} + \frac{\partial \Phi}{\partial F_{JI}} \right) - 2n_e \frac{\partial \Phi}{\partial G_I} = \mathbf{0} \tag{A.20}$$

As the zeroth-order wavefunction Φ is included in the linear method space, these are linear dependencies in the form of equation A.10. After integrating these constant linear coefficients, these supply us with the equivalence paths:

$$\boldsymbol{p}_1(\alpha) = (\boldsymbol{F} + \alpha \underline{\underline{\mathbf{1}}}, \boldsymbol{G}, \{g_I\}) \tag{A.21}$$

$$\boldsymbol{p}_{2,J}(\alpha) = (\boldsymbol{F} + \alpha(\underline{\mathbf{1}}^T \boldsymbol{e}_J + \boldsymbol{e}_J^T \underline{\mathbf{1}}), \boldsymbol{G} - \alpha(2n_e \boldsymbol{e}_J), \{g_I\}) \tag{A.22}$$

with:

$$\underline{\underline{\mathbf{1}}}_{ij} = 1, \quad \underline{\mathbf{1}}_i = 1, \quad (\boldsymbol{e}_j)_i = \delta_{ij} \tag{A.23}$$

We restrict variation along each of these paths by freezing the following parameter values:

$$\boldsymbol{G} = \mathbf{0}, \quad F_{n_C, n_C} = 0, \quad n_C = \mathrm{rank}(\boldsymbol{F}) \tag{A.24}$$

and omitting their parameter derivatives from the linear method tangent space.

Multiplying each of the anchor gaussians in $\{g_I\}$ by a single gaussian function g produces the same normalized counting function basis $\{C_I\}$, as this operation is equivalent to multiplying the numerator and denominator of equation 3.12 by g. Since multiplying two gaussian functions together produces another gaussian function, this directly gives the equivalence path:

$$\boldsymbol{p}_3(g) = (\boldsymbol{F}, \boldsymbol{G}, \{g_I / g\}) \tag{A.25}$$

The coefficients $\boldsymbol{p}_3'(g)$ that explicitly appear in the associated linear dependencies are more complex, as they depend on the value of gaussian parameters internal to g. To remove these dependencies, we find an equivalent set of anchor gaussians with ten fewer parameters by dividing each by a single anchor gaussian g_r:

$$\{\tilde{g}_I\} = \{1, g_2/g_r, \ldots, g_n/g_r\} \tag{A.26}$$

and omit the parameters of g_r from the linear method space.

B Spherical Voronoi Partitioning

Basis functions written in terms of spherical coordinates are a natural choice in chemical systems, as discussed in section 3.6. Here we provide an algorithm that generates counting functions that simultaneously divide space into angular sectors and radial shells. First, we define an partition on the surface of a sphere with a set of unit vectors $\hat{\mu}_i$. We will refer to anchor gaussians by the pair index (j, i):

$$g_{(j,i)}(\boldsymbol{r}) = \exp\left(-(\boldsymbol{r}^T \boldsymbol{A}_{(j,i)} \boldsymbol{r} - 2\boldsymbol{B}_{(j,i)}^T \boldsymbol{r} + K_{(j,i)})\right) \tag{B.1}$$

where j indexes the radial shell (in order from smallest to largest radius) and i indexes the angular position. We then specify the maximum slope S across switching surfaces set up by the angular partition, which happens to be consistent across radial shells, and set

$$\alpha = S/d_{min} \tag{B.2}$$

where d_{min} is the minimum Cartesian distance between the $\hat{\mu}_i$:

$$d_{min} = \min_{i \neq j} |\hat{\mu}_i - \hat{\mu}_j| \tag{B.3}$$

This gives the following parameters for the innermost shell of anchor gaussians:

$$\boldsymbol{A}_{(0,i)} = \alpha \boldsymbol{I}, \quad \boldsymbol{B}_{(0,i)} = \alpha \hat{\mu}_i, \quad K_{(0,i)} = \alpha \tag{B.4}$$

The pair counting region boundaries between any pair of these anchor gaussians is a plane that intersects the center of the sphere, and partitions space into a set of angular sectors. To further divide this partition along radial shells, we iteratively generate anchor gaussian parameters using those from the current outermost shell:

$$\boldsymbol{A}_{(j,i)} = \boldsymbol{A}_{(j-1,i)} \left(\frac{R_j - 2S_j}{R_j} \right) \tag{B.5}$$

$$\boldsymbol{B}_{(j,i)} = \boldsymbol{B}_{(j-1,i)} \tag{B.6}$$

$$K_{(j,i)} = K_{(j-1,i)} - 2S_j R_j \tag{B.7}$$

The pair counting function made from anchor gaussians on neighboring shells has a spherical switching boundary centered at the origin with radius R_j and with maximum slope S_j, which may be verified by comparing to equation 3.10:

$$C_{(j,i),(j-1,i)}(\boldsymbol{r}) = \frac{1}{1 + \exp\left(-2R_j S_j \left(\frac{r^2}{R_j^2} - 1\right)\right)} \tag{B.8}$$

C Region Composition

In sections 3.5 and 3.6 we describe how to partition space into atom-centered Voronoi cells and spherical Voronoi cells, and in section 3.7, we argue that subdividing regions within a single Jastrow basis is necessary to capture both interatomic and intraatomic population correlations simultaneously. We do this by 'composing' two existing, independent sets of counting functions into a single set while retaining the boundaries present in both partitions simultaneously as best we can. Conditions in equations 3.30 and 3.31 should be satisfied in this composition process and we provide a scheme that best meets these conditions here.

We will start by considering the case shown in figure 3.9, which starts with a rough partition of two counting functions (C_α and C_β) and a set of counting functions ($\{C_I\}$) that describe how C_α is to be subdivided. Our desired result is a set of anchor gaussians $\{g_I^{(\alpha)}\}$ that, when used to replace the anchor gaussian g_α in the initial two-region partition,

generates counting functions that reproduce the boundaries between the counting regions $\mathcal{R}_I$ within the region $\mathcal{R}_\alpha$:

$$\begin{aligned} C_I^{(\alpha)}(\boldsymbol{r}) &= \frac{g_I^{(\alpha)}(\boldsymbol{r})}{\sum_J g_J^{(\alpha)}(\boldsymbol{r}) + C_\beta(\boldsymbol{r})} \\ &= \frac{g_I(\boldsymbol{r})}{\sum_J g_J(\boldsymbol{r})}, \quad \text{whenever } \boldsymbol{r} \in \mathcal{R}_\alpha \end{aligned} \tag{C.1}$$

and whose counting functions can be added together to reproduce the counting function $C_\alpha(\boldsymbol{r})$ everywhere:

$$C_\alpha(\boldsymbol{r}) = \frac{g_\alpha(\boldsymbol{r})}{g_\alpha(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} = \frac{\sum_J g_J^{(\alpha)}(\boldsymbol{r})}{\sum_J g_J^{(\alpha)}(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} \tag{C.2}$$

The first condition will be satisfied if we set:

$$g_I^{(\alpha)} = g_I(\boldsymbol{r}) \cdot g_r(\boldsymbol{r}) \tag{C.3}$$

for any gaussian function $g_r(\boldsymbol{r})$, since the boundaries between the counting regions of $C_I^{(\alpha)}$ are characterized by the pair counting functions:

$$C_{I/J}^{(\alpha)}(\boldsymbol{r}) = \frac{g_I^{(\alpha)}(\boldsymbol{r})}{\sum_J g_J^{(\alpha)}(\boldsymbol{r})} = \frac{g_I(\boldsymbol{r})}{\sum_J g_J(\boldsymbol{r})} = C_{I/J}(\boldsymbol{r}) \tag{C.4}$$

which are the same as those of the original set $\{C_I\}$.

To satisfy the second condition, we substitute equation C.3 into the expression for $C_\alpha(\boldsymbol{r})$:

$$C_\alpha(\boldsymbol{r}) = \frac{g_\alpha(\boldsymbol{r})}{g_\alpha(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} = \frac{\sum_J g_J(\boldsymbol{r}) g_r(\boldsymbol{r})}{\left(\sum_J g_J(\boldsymbol{r})\right) g_r(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} \tag{C.5}$$

After solving for $g_r(\boldsymbol{r})$, this becomes:

$$g_r(\boldsymbol{r}) = \frac{g_\alpha(\boldsymbol{r})}{\sum_J g_J(\boldsymbol{r})} \tag{C.6}$$

The mismatch of the functional forms on the each side of this equation means that a solution for the gaussian parameters of g_r that holds for all $\boldsymbol{r}$ will not be possible except in trivial or limiting cases. We instead match a second-order Taylor expansion of the natural logarithm of each side at a single point $\boldsymbol{r}_0$:

$$\begin{aligned} \ln(g_r(\boldsymbol{r})) &= K_r - 2\boldsymbol{r}^T \boldsymbol{B}_r + \boldsymbol{r}^T \boldsymbol{A}_r \boldsymbol{r} \\ &= f(\boldsymbol{r}_0) + (\boldsymbol{r} - \boldsymbol{r}_0) \cdot \nabla f\big|_{\boldsymbol{r}_0} \\ &\quad + \frac{1}{2} (\boldsymbol{r} - \boldsymbol{r}_0)^T \boldsymbol{H}\big|_{\boldsymbol{r}_0} (\boldsymbol{r} - \boldsymbol{r}_0) \end{aligned} \tag{C.7}$$

Where $f(\boldsymbol{r}), \nabla f\big|_{\boldsymbol{r}}, \boldsymbol{H}\big|_{\boldsymbol{r}}$ are the right-hand side of equation C.6, its gradient, and its Hessian, evaluated at $\boldsymbol{r}$. Matching terms of the same order on each side gives us an explicit expression for each gaussian parameter in g_r:

$$K_r = \frac{1}{2}\boldsymbol{r}_0^T \boldsymbol{H}\big|_{\boldsymbol{r}_0}\boldsymbol{r}_0 - \nabla f\big|_{\boldsymbol{r}_0} \cdot \boldsymbol{r}_0 + f(\boldsymbol{r}_0) \tag{C.8}$$

$$\boldsymbol{B}_r = \frac{\boldsymbol{H}\big|_{\boldsymbol{r}_0}\boldsymbol{r}_0 - \nabla f\big|_{\boldsymbol{r}_0}}{2} \tag{C.9}$$

$$\boldsymbol{A}_r = \boldsymbol{H}\big|_{\boldsymbol{r}_0} \tag{C.10}$$

In this case, where we are subdividing an atom-centered Voronoi region into spherical shells, we choose $\boldsymbol{r}_0$ as a point midway between neighboring atomic centers, directly on the region boundary. In our simple two-region example, this is:

$$\boldsymbol{r}_0 = \frac{\boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\beta}{2} \tag{C.11}$$

This selection is made to best preserve the counting region boundary along the bond axis.

We can generalize this to replace both atom-centered regions $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ with multiple subpartitions $\{\mathcal{R}_{I,\alpha}\}$ and $\{\mathcal{R}_{I,\beta}\}$ simultaneously. Working through exactly the same logic with the first condition provides a similar prescription for $g_I^{(\alpha)}$ and $g_I^{(\beta)}$:

$$g_I^{(\alpha)}(\boldsymbol{r}) = g_{I,\alpha}(\boldsymbol{r}) g_r^{(\alpha)}(\boldsymbol{r}), \quad g_I^{(\beta)}(\boldsymbol{r}) = g_{I,\beta}(\boldsymbol{r}) g_r^{(\beta)}(\boldsymbol{r}) \tag{C.12}$$

We substitute this into the second condition, with:

$$S_\alpha(\boldsymbol{r}) = \sum_I g_{I,\alpha}(\boldsymbol{r}), \quad S_\beta(\boldsymbol{r}) = \sum_I g_{I,\beta}(\boldsymbol{r}) \tag{C.13}$$

which gives:

$$C_\alpha(\boldsymbol{r}) = \frac{g_\alpha(\boldsymbol{r})}{g_\alpha(\boldsymbol{r}) + g_\beta(\boldsymbol{r})} = \frac{S_\alpha(\boldsymbol{r}) g_r^{(\alpha)}(\boldsymbol{r})}{S_\alpha(\boldsymbol{r}) g_r^{(\alpha)}(\boldsymbol{r}) + S_\beta(\boldsymbol{r}) g_r^{(\beta)}(\boldsymbol{r})} \tag{C.14}$$

As we can divide both numerator and denominator on the right-hand side by $g_r^{(\alpha)}$, the behavior of the composed set of anchor gaussians is uniquely determined by the quotient of gaussians,

$$g_q(\boldsymbol{r}) = \frac{g_r^{(\beta)}(\boldsymbol{r})}{g_r^{(\alpha)}(\boldsymbol{r})} \tag{C.15}$$

which we solve for as before:

$$g_q(\boldsymbol{r}) = \frac{g_r^{(\beta)}(\boldsymbol{r})}{g_r^{(\alpha)}(\boldsymbol{r})} = \frac{S_\alpha(\boldsymbol{r}) g_\beta(\boldsymbol{r})}{S_\beta(\boldsymbol{r}) g_\alpha(\boldsymbol{r})} \tag{C.16}$$

We conclude exactly as before by matching the second-order Taylor series of the logarithms of both sides of equation C.16 to solve for the gaussian parameters of g_q. In cases where performing the composition serially may distort the counting region boundary, this latter scheme more faithfully reproduces the original boundaries of $\mathcal{R}_\alpha$ and $\mathcal{R}_\beta$ in the final composed basis.

D ESMF Derivatives

D.1 Wavefunction and Notation

Notation used in the ESMF derivatives follow. Spin components have been integrated out, and all formulae are in terms of spatial orbitals.

- ESMF wavefunction:
$$|\Psi\rangle = c_0 e^{\mathbf{X}} |\psi\rangle + e^{\mathbf{X}} \mathbf{C} |\psi\rangle \tag{D.1}$$
- Wavefunction Normalization
$$N = \langle\Psi|\Psi\rangle = |c_0|^2 + 2\sum_{ia} |c_{ia}|^2 \tag{D.2}$$
- Orbital rotation operator $\mathbf{X}$ and Single-excitation operator $\mathbf{C}$:
$$\mathbf{X} \equiv \sum_{ia} x_{ia} \left(\mathbf{a}^\dagger \mathbf{i} - \mathbf{i}^\dagger \mathbf{a}\right) \qquad \mathbf{C} \equiv \sum_{ia} c_{ia} \left(\mathbf{a}^\dagger \mathbf{i}\right) \tag{D.3}$$
- Occupied molecular orbitals in the reference ψ will be specified by $ijkl$ indices.
- Virtual molecular orbitals in the reference ψ will be specified by $abcd$ indicies.
- General orbitals will be specified by $pqrs$ indices.
- Quantities in the atomic orbital basis will be indicated by an overhead tilde (i.e. $\tilde{c}_{pq}$)

Numerical Quantities

- Integrals are as defined in equation 1.33.
- One-body Density matrix:
$$\gamma_{ij} = \begin{cases} 1 & i = j \text{ and } i \in \text{occ} \\ 0 & i \neq j \text{ or } i \notin \text{occ} \end{cases} \tag{D.4}$$
- Fock matrix elements:
$$f_{ia} = 2h_{ia} + \sum_p \gamma_{pp} \big(2(pp|ia) - (pi|pa)\big) \tag{D.5}$$
- Two-electron integral and Fock-like contractions:
$$J[d]_{pq} = \sum_{rs} d_{rs}(pq|rs) \tag{D.6}$$
$$K[d]_{pq} = \sum_{rs} d_{rs}(pr|qs) \tag{D.7}$$
$$F[d]_{pq} = 2J[d]_{pq} - K[d]_{pq} \tag{D.8}$$

- Basis transformations:

$$(ij|ab) = \sum_{pq} C_{pi}(\tilde{p}\tilde{q}|ab)C^T_{aq} \tag{D.9}$$

$$f_{ia} = \sum_{pq} C_{pi}\tilde{f}_{pq}C^T_{aq} \tag{D.10}$$

$$J[d]_{pq} = \sum_{ia} C_{pi}\tilde{J}[d]_{ia}C^T_{aq} \tag{D.11}$$

$$K[d]_{pq} = \sum_{ia} C_{pi}\tilde{K}[d]_{ia}C^T_{aq} \tag{D.12}$$

$$F[d]_{ia} = \sum_{pq} C_{pi}\tilde{F}[d]_{pq}C^T_{aq} \tag{D.13}$$

$$\tilde{c}_{pq} = \sum_{ia} C_{pi}c_{ia}C^T_{aq} \tag{D.14}$$

$$\left(\tilde{\boldsymbol{c}}^T\tilde{\boldsymbol{c}}\right)_{pq} = \sum_{ajb} C_{pa}c^T_{aj}c_{jb}C^T_{bq} \tag{D.15}$$

$$\left(\tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T\right)_{pq} = \sum_{iaj} C_{pi}c_{ia}c^T_{aj}C^T_{jq} \tag{D.16}$$

D.2 Energy

Reference Energy

$$E_0 = \sum_{pq} \tilde{\gamma}_{pq}\left(2\tilde{h}_{pq} + 2\tilde{J}[\tilde{\boldsymbol{\gamma}}]_{pq} - \tilde{K}[\tilde{\boldsymbol{\gamma}}]_{pq}\right) \tag{D.17}$$

ESMF Energy

$$E = \frac{1}{N}\left[NE_0 + 4c_0\sum_{pq}\tilde{c}_{pq}\tilde{f}_{pq} + 2\sum_{pq}\left(\tilde{\boldsymbol{c}}^T\tilde{\boldsymbol{c}} - \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T\right)_{pq}\tilde{f}_{pq} + 2\sum_{pq}\tilde{c}_{pq}\tilde{F}[\tilde{\boldsymbol{c}}]_{pq}\right] \tag{D.18}$$

D.3 First Derivatives

Reference coefficient

$$\left.\frac{\partial E}{\partial c_0}\right|_{X=0} = \frac{1}{N}\left[2c_0(E_0 - E) + 4\sum_{ia} c_{ia}f_{ia}\right] \tag{D.19}$$

CIS coefficient

$$\left.\frac{\partial E}{\partial c_{ia}}\right|_{X=0} = \frac{1}{N}\left[4c_{ia}(E_0 - E) + 4c_0 f_{ia} + 4\sum_b c_{ib}f_{ba} - 4\sum_j f_{ij}c_{ja} + 4F[\tilde{\boldsymbol{c}}]_{ia}\right] \tag{D.20}$$

Orbital Rotation coefficient

$$\begin{aligned}\left.\frac{\partial E}{\partial x_{ia}}\right|_{X=0} = \frac{1}{N}\Bigg[&4Nf_{ia} + 4c_0\sum_b c_{ib}f_{ba} - 4c_0\sum_j f_{ij}c_{ja} \\ &-4\sum_{jb} c^T_{aj}c_{jb}f_{bi} - 4\sum_{jb} c_{ib}c^T_{bj}f_{ja} \\ &+4c_0F[\tilde{\boldsymbol{c}}]_{ia} + 4c_0F[\tilde{\boldsymbol{c}}]_{ai} + 4F[\tilde{\boldsymbol{c}}^T\tilde{\boldsymbol{c}} - \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T]_{ia} \\ &+4\sum_c c_{ic}F[\tilde{\boldsymbol{c}}]_{ca} - 4\sum_k F[\tilde{\boldsymbol{c}}]_{ik}c_{ka}\Bigg]\end{aligned} \tag{D.21}$$

D.4 Second Derivatives

Reference coefficient, reference coefficient

$$\left.\frac{\partial^2 E}{\partial c_0^2}\right|_{X=0} = \frac{1}{N}\left[2\left(E_0 - E\right) - 4c_0\frac{\partial E}{\partial c_0}\right] \tag{D.22}$$

Reference coefficient, CIS coefficient

$$\frac{\partial^2 E}{\partial c_0 \partial c_{ia}} = \frac{1}{N}\left[4f_{ia} - 4c_{ia}\frac{\partial E}{\partial c_0} - 2c_0\frac{\partial E}{\partial c_{ia}}\right] \tag{D.23}$$

Reference coefficient, Rotation coefficient

$$\frac{\partial^2 E}{\partial c_0 \partial x_{ia}} = \frac{1}{N}\left[8c_0f_{ia} + 4\sum_b c_{ib}f_{ba} - 4\sum_j f_{ij}c_{ja} + 4F[\tilde{\boldsymbol{c}}]_{ia} + 4F[\tilde{\boldsymbol{c}}]_{ai}\right] \tag{D.24}$$

CIS coefficient, CIS coefficient

$$\begin{aligned}\left.\frac{\partial^2 E}{\partial c_{ia}\partial c_{jb}}\right|_{X=0} = \frac{1}{N}\Bigg[&4\delta_{ij}\delta_{ab}\left(E_0 - E\right) + 4\delta_{ij}f_{ab} - 4\delta_{ab}f_{ij} + 8(ia|jb) - 4(ij|ab) \\ &-4c_{ia}\frac{\partial E}{\partial c_{jb}} - 4c_{jb}\frac{\partial E}{\partial c_{ia}}\Bigg]\end{aligned} \tag{D.25}$$

CIS coefficient, Rotation coefficient

$$\begin{aligned}\left.\frac{\partial^2 E}{\partial x_{ia}\partial c_{jb}}\right|_{X=0} = \frac{1}{N}\Bigg[&-4\sum_k c_{kb}\left[2(ia|kj)-(ja|ki)\right]-4\sum_k c_{kb}\left[2(ia|kj)-(ij|ka)\right]\\ &-4\sum_k c_{ka}\left[2(jb|ki)-(ib|kj)\right]+4\sum_c c_{jc}\left[2(ia|cb)-(ib|ca)\right]\\ &+4\sum_c c_{jc}\left[2(ia|cb)-(ab|ci)\right]+4\sum_c c_{ic}\left[2(jb|ca)-(ja|cb)\right]\\ &+4c_0\left(2(ai|jb)-(aj|bi)\right)+4c_0\left(2(ai|jb)-(ab|ij)\right)\\ &+16c_{jb}f_{ia}-4c_{ib}f_{ja}-4c_{ja}f_{ib}-4c_{jb}\frac{\partial E}{\partial x_{ia}}\\ &+4c_0\delta_{ij}f_{ab}+4\delta_{ij}F[\tilde{\boldsymbol{c}}]_{ab}-4\delta_{ij}\sum_k c_{kb}f_{ka}\\ &-4c_0\delta_{ab}f_{ij}-4\delta_{ab}F[\tilde{\boldsymbol{c}}]_{ji}-4\delta_{ab}\sum_c c_{jc}f_{ic}\Bigg]\end{aligned} \tag{D.26}$$

Rotation coefficient, Rotation coefficient

$$
\begin{aligned}
\left.\frac{\partial^2 E}{\partial x_{ia}\partial x_{jb}}\right|_{X=0} = \frac{1}{N}\Bigg[& 4N\delta_{ij}f_{ab} - 4N\delta_{ab}f_{ij} + 4N\Big(2(ia|jb) - (ij|ab)\Big) + 4N\Big(2(ia|jb) - (ja|ib)\Big) \\
& -4c_0c_{ib}f_{ja} - 4c_0c_{ja}f_{ib} - 4c_{ib}F[\tilde{\boldsymbol{c}}]_{aj} - 4c_{ja}F[\tilde{\boldsymbol{c}}]_{bi} + 4f_{ij}\sum_k c_{kb}c_{ka} - 4f_{ab}\sum_c c_{jc}c_{ic} \\
& -2c_0\delta_{ij}\sum_k c_{ka}f_{kb} - 2c_0\delta_{ij}\sum_k f_{ak}c_{kb} - 2c_0\delta_{ab}\sum_c c_{ic}f_{cj} - 2c_0\delta_{ab}\sum_c f_{ic}c_{jc} \\
& -2\delta_{ij}\sum_{kc} c_{ka}c_{kc}f_{cb} - 2\delta_{ij}\sum_{kc} c_{kb}c_{kc}f_{ac} + 2\delta_{ab}\sum_{kc} c_{ic}c_{kc}f_{kj} + 2\delta_{ab}\sum_{kc} c_{jc}c_{kc}f_{ik} \\
& +4\delta_{ij}c_0F[\tilde{\boldsymbol{c}}]_{ab} + 4\delta_{ij}c_0F[\tilde{\boldsymbol{c}}]_{ba} - 4\delta_{ab}c_0F[\tilde{\boldsymbol{c}}]_{ij} - 4\delta_{ab}c_0F[\tilde{\boldsymbol{c}}]_{ji} \\
& -2\delta_{ij}\sum_k F[\tilde{\boldsymbol{c}}]_{ak}c_{kb} - 2\delta_{ij}\sum_k c_{ka}F[\tilde{\boldsymbol{c}}]_{kb} + 4\delta_{ij}F[\tilde{\boldsymbol{c}}^T\tilde{\boldsymbol{c}} - \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T]_{ab} \\
& -2\delta_{ab}\sum_c F[\tilde{\boldsymbol{c}}]_{ic}c_{jc} - 2\delta_{ab}\sum_c c_{ic}F[\tilde{\boldsymbol{c}}]_{cj} - 4\delta_{ab}4F[\tilde{\boldsymbol{c}}^T\tilde{\boldsymbol{c}} - \tilde{\boldsymbol{c}}\tilde{\boldsymbol{c}}^T]_{ij} \\
& -4c_0\sum_k c_{kb}\Big(2(ia|kj) - (ij|ka)\Big) - 4c_0\sum_k c_{ka}\Big(2(jb|ki) - (ji|kb)\Big) \\
& -4c_0\sum_k c_{kb}\Big(2(ia|kj) - (ja|ki)\Big) - 4c_0\sum_k c_{ka}\Big(2(jb|ki) - (ib|kj)\Big) \\
& +4c_0\sum_c c_{jc}\Big(2(ia|cb) - (ib|ca)\Big) + 4c_0\sum_c c_{ic}\Big(2(jb|ca) - (ja|cb)\Big) \\
& +4c_0\sum_c c_{jc}\Big(2(ia|cb) - (ab|ci)\Big) + 4c_0\sum_c c_{ic}\Big(2(jb|ca) - (ba|cj)\Big) \\
& -4\sum_{kc} c_{ka}c_{kc}\Big(2(jb|ci) - (ib|cj)\Big) - 4\sum_{kc} c_{kb}c_{kc}\Big(2(ia|cj) - (ja|ci)\Big) \\
& -4\sum_{kc} c_{ic}c_{kc}\Big(2(jb|ka) - (aj|kb)\Big) - 4\sum_{kc} c_{jc}c_{kc}\Big(2(ia|kb) - (ib|ka)\Big) \\
& -4\sum_{kc} c_{ic}c_{kc}\Big(2(jb|ka) - (ab|kj)\Big) - 4\sum_{kc} c_{jc}c_{kc}\Big(2(ia|kb) - (ab|ki)\Big) \\
& -4\sum_{kc} c_{ka}c_{kc}\Big(2(jb|ci) - (ij|cb)\Big) - 4\sum_{kc} c_{kb}c_{kc}\Big(2(ia|cj) - (ij|ca)\Big) \\
& -4\sum_{kc} c_{kb}c_{ic}\Big(2(kj|ca) - (ka|cj)\Big) - 4\sum_{kc} c_{jc}c_{ka}\Big(2(ki|cb) - (kb|ci)\Big) \\
& +4\sum_{cd} c_{jc}c_{id}\Big(2(cb|ad) - (cd|ab)\Big) + 4\sum_{kl} c_{kb}c_{la}\Big(2(kj|li) - (kl|ij)\Big)\Bigg]
\end{aligned}
\tag{D.27}
$$