Eigenvector continuation in chemistry on a quantum computer

Few-parameter representations of the wave function

Simon Elias Schrader

Thesis submitted for the degree of
Master in Computational Science: Chemistry
60 credits

Department of Chemistry
The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

Spring 2022

Eigenvector continuation in chemistry on a quantum computer

Few-parameter representations of the wave function

Simon Elias Schrader

Eigenvector continuation in chemistry on a quantum computer

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

An important part of quantum chemistry is to solve the time-independent Schrödinger equation to obtain the ground state and the corresponding energy. Approximate single reference-methods such as coupled cluster (CC) are qualitatively correct close to the equilibrium structure, but can fail qualitatively upon dissociation. We used subspace projected CC eigenvector continuation to analytically continue the ground state and obtained qualitatively correct dissociation curves for the nitrogen molecule and improved potential energy surfaces for beryllium hydride. We developed a variant of eigenvector continuation that acts on the cluster operator and found that few sample states are sufficient to obtain chemical accuracy compared to the CCSD energy along wide ranges of the potential energy surface, while cutting off more than 50% of computing time using an approximate calculation scheme for the cluster operator. We found that eigenvector continuation cannot be used on quantum computers when sample points are close to one another due to a nearly singular overlap matrix, but it is an efficient interpolation method that can be used to solve some types of multi-reference problems.

Contents

List of Figures

List of Tables

Preface

Part I

Introduction

The purpose of this thesis is to study how eigenvector continuation can be used in the context of quantum chemistry. eigenvector continuation [1] is a novel method where the ground state eigenvector of a Hermitian matrix $\mathbf{H}(\vec{\alpha}_\odot)$, analytically dependent on some parameter $\vec{\alpha}$, is approximated as a linear combination of ground state eigenvectors at different parameters $\vec{\alpha}_1, \ldots, \vec{\alpha}_L$. With eigenvector continuation, it has been shown [1] to be possible to extrapolate the solutions from "easy" regimes to "difficult" regimes at reduced computational cost. Furthermore, ground state vectors in many-dimensional vector spaces can be well-represented with few parameters [2]. A variant of eigenvector continuation, applicable to CCSD wave functions [3], has decreased the assumed run time for a given calculation from 20 years down to one hour with very small accuracy loss, and eigenvector continuation has been shown to have better convergence properties than perturbation theory [2]. All of these considerations make eigenvector continuation a promising method to obtain approximate solutions of the ground state of Hermitian matrices.
Quantum mechanics is a natural application of eigenvector continuation. To our knowledge, however, it has not been applied to quantum chemistry. One of the main concerns in quantum chemistry is to find the ground state of the electronic Schrödinger equation. Often, it is of interest to find a potential energy surface, e.g. finding the ground state energy as a function of the molecular geometry, expressed by the nuclear positions $\vec{R}$. As finding the exact solution in a given basis set scales exponentially with the number of electrons, it is necessary to use approximations that have polynomial scaling. The CCSD approximation is a high-accuracy approximation method and scales only as $O(N^6)$ with the number of electrons. This is still expensive. Furthermore, the CCSD method fails for systems with strong multi-reference character and can give results that are both quantitatively and qualitatively wrong. It is hence of interest to both improve the CCSD wave function as well as approximate it by cheaper methods. In this thesis, we study if and how eigenvector continuation can be used to extrapolate the solutions from "easy" geometries to "hard" geometries, and if and how it can be used to reduce numerical cost. An extra difficulty in quantum chemistry is that each nuclear configuration gets assigned a different basis and hence spans a different vector space, which is something we focus on particularly throughout this thesis.
Quantum computing is a promising technology for quantum chemistry, with the potential of gaining "quantum advantage" and thus outperform classical computing. Calculations on state of the art quantum computers and near-future devices all rely on the VQE algorithm [4], where the wave function is parametrized by an ansatz, and expectation values, including the energy, are sampled. As sampling introduces insecurity, quantum computing is a natural source to evaluate how eigenvector continuation works when errors are introduced. At the same time, specific operations which are considered very expensive on classical devices have a low polynomial or even linear scaling on quantum computers, such as the trotterized unitary coupled cluster method and methods derived from it. We have thus studied how eigenvector continuation can be performed with data sampled from quantum computers and which difficulties and advantages are introduced.

This thesis is organized in the following way: In chapter 1, we give a brief and non-exhaustive introduction of quantum chemical methods and theory necessary to understand the context and the aim of this thesis. In chapter 2, we explain eigenvector continuation, what it does, how it works and the way it is implemented, and how it has been adapted to coupled cluster wave functions. In chapter 3, we describe the basics of quantum computing and how it can be applied to chemistry, focusing on relevant formulas and circuits to implement eigenvector continuation. In chapter 4, we will discuss necessary theory, methods and considerations to implement eigenvector continuation for quantum chemistry, with the main focus to obtaining potential energy surfaces. In this chapter, we will both introduce new methods as well as adaptions of previously described methods in such a way that they can be applied to quantum chemistry. It hence reflects our thinking and reasoning as well as our own work. This is followed by a chapter on how we implemented algorithms numerically. In the next chapters 6, 7 and 8, we show and discuss our results for single-reference, multi-reference and quantum computing methods, respectively, and finish

with a conclusion and an outlook. Finally, the appendix contains different figures and proofs which did not belong in the main part of this thesis.

A small note on notation

Quantum mechanics has a lot of notation. In order to not repeatedly state what parameters stand for, the following set of parameters will, unless stated otherwise, have the following meaning:

- N stands for the number of electrons in a system.
- M stands for the number of molecular spin orbitals.
- L stands for the number of sampling vectors included in eigenvector continuation algorithm.
- α or $\vec{\alpha}$ is a general single- or multidimensional parameter not further specified, while $\vec{R}$ stands for the position of the nuclei within the BO approximation.
- The indices $a, b, c, \ldots$ represent unoccupied molecular orbitals, $i, j, k, \ldots$ represent occupied molecular orbitals, and $p, q, r, \ldots$ represent either.
- $|\Psi\rangle$ (possibly with an index, a tilde, ...) will be used for general wave functions, unless otherwise stated. A Slater determinant (or a CSF) will be written $|\Phi^{\mathrm{SD}}\rangle$, and $|\Phi^{\mathrm{HF}}\rangle$ will be a Hartree-Fock Slater determinant. One-particle functions will be written with small Greek letters.
- We will write operators with a hat $\hat{O}$, with the exception of creation/annihilation operators and Pauli spin operators, matrices with large, bold face letters $\mathbf{A}$, and vectors either as small, boldface letters $\mathbf{v}$ or with a vector symbol $\vec{v}$. On quantum computers, the operator hat or the boldfacing will sometimes be dropped for notational convenience, as the difference between gates, operators in a basis and matrices is sometimes hard to distinguish.
- For elements of a matrix $\mathbf{M}$, we write $(\mathbf{M})_{ij} = M_{ij}$. For elements of inverse matrices, we write $(\mathbf{M})^{-1}_{ij} = M^{-1}_{ij}$.

Part II

Theory

Chapter 1

Quantum chemical methods

The aim of this chapter is to give an introduction to quantum chemical methods that are relevant for this thesis. We will describe the standard methods, with special focus on Hartree-Fock theory and coupled cluster theory as well as the analytic properties of specific choices of Molecular orbitals upon geometric perturbations of the atoms, which is important for the application of eigenvector continuation. This text is by no means meant to be a complete summary, but should be seen as a text that builds up to and emphasises methods and concepts which are important for methods suggested and developed in the methods part of the thesis. Almost everything in this chapter is taken from the literature, except for the analytical properties of the Hamiltonian in a basis and the natural orbitals as functions of geometric perturbations.

1.1 Basics of quantum chemistry

Quantum chemistry is the application of quantum mechanics to chemical systems. To be able to predict the dynamics of a system, which is governed by the time dependent Schrödinger equation, it is necessary to find the eigenfunctions of the time independent Schrödinger equation

$$\hat{H}\left|\Psi\right\rangle = E\left|\Psi\right\rangle \tag{1.1}$$

where $\hat{H}$ for molecules and atoms is the electronic Hamiltonian

$$\begin{aligned}\hat{H} &= \sum_{i=1}^{N}\left[-\frac{1}{2}\vec{\nabla}_i^2 + \sum_{A=1}^{N_A}\frac{Z_A}{\left|\vec{r}_i - \vec{R}_A\right|}\right] + \sum_{j<i}^{N}\frac{1}{\left|\vec{r}_i - \vec{r}_j\right|} + h_{\text{nuc}} \\ &= \sum_{i=1}^{N}\hat{h}_i + \sum_{j<i}^{N}\hat{v}_{ij} + h_{\text{nuc}} \\ &= \hat{H}_1 + \hat{H}_2 + h_{\text{nuc}}\end{aligned} \tag{1.2}$$

where N_A is the number of atoms, $\vec{r}_i$ the spatial position of electron i, $\vec{R}_A$ the spatial position of nucleus A with corresponding charge Z_A, h_{nuc} the repulsion between the (stationary) nuclei, and all units in Hartree. This Hamiltonian arises from considering the full Hamiltonian under the BO approximation [5]. From statistical mechanics, we know that the ground state is most prevalent, and it is used in many methods, such as ab-initio molecular dynamics [6]. However, for two electrons or more, eq. (1.1) has not been solved analytically, making it necessary to do approximations. The first step is the insertion of a basis, that is, expressing the wave function as

$$\left|\Psi\right\rangle = \sum_{i=1}^{\infty} c_i\left|\Phi_i\right\rangle \tag{1.3}$$

where $\left|\Phi_i\right\rangle$ are orthonormal. This parameterization is exact, but it is computationally infeasible to sum up to ∞, and we will consider finite bases in section 1.2.1.

In the following we use the standard notation $\mathbf{x} = \{\mathbf{r}, \omega\}$, where $\omega \in \{\alpha, \beta\}$. That is, we write $|\psi(\mathbf{x})\rangle \hat{=} |\psi(\mathbf{r})\rangle \otimes |\omega\rangle$, and we will use the notation

$$\int d\mathbf{x} \hat{=} \sum_{\omega \in \{\alpha,\beta\}} \int d\mathbf{r}. \tag{1.4}$$

Unless explicitly necessary, we will simply write $|\psi\rangle$ instead of $|\psi(\mathbf{x})\rangle$ or $\psi(\mathbf{x})$. Furthermore, we will use the notation $H^{\wedge N}$ for the antisymmetric Hilbert space of N particles, and simply write H for $H^{\wedge 1}$.

1.1.1 Slater determinants

A possible parameterization of the exact wave function is to write it as a linear combination of Slater-determinants. For one-electron functions $|\phi_1(\mathbf{x})\rangle, \ldots, |\phi_N(\mathbf{x})\rangle \in H$, we define the *Slater determinant* as

$$|\Phi^{\mathrm{SD}}\rangle = |\phi_1\phi_2\cdots\phi_N\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & \phi_2(\mathbf{x}_1) & \cdots & \phi_N(\mathbf{x}_1) \\ \phi_1(\mathbf{x}_2) & \phi_2(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{x}_N) & \phi_2(\mathbf{x}_N) & \cdots & \phi_N(\mathbf{x}_N) \end{vmatrix}. \tag{1.5}$$

Being a determinant, the Slater determinant $|\Phi^{\mathrm{SD}}\rangle$ is antisymmetric with respect to particle exchange, hence $|\Phi^{\mathrm{SD}}\rangle \in H^{\wedge N}$. If all one-electron functions entering the Slater determinant are mutually orthonormal, the prefactor $1/\sqrt{N!}$ guarantees normalization of the Slater determinant. Unless otherwise stated, Slater determinants $|\Phi^{\mathrm{SD}}\rangle$ will be normalized and the one-electron functions that enter it, will be mutually orthonormal in the remainder of this thesis.
The same state can also be represented as an *occupation number (ON) vector* or *bit string*

$$\begin{aligned} |\mathbf{k}\rangle = |k_1, k_2, \ldots, k_M, \ldots\rangle, k_p = \begin{cases} 1 & \phi_p \text{ occupied} \\ 0 & \phi_p \text{ virtual} \end{cases} \\ \sum_{p=1}^{\infty} k_p = N \end{aligned} \tag{1.6}$$

where the term *occupied* means that $\phi_p(\mathbf{x})$ enters the Slater determinant, and *virtual* means not occupied - this notation, of course, only makes sense if the one-electron functions are mutually orthogonal. It can be shown [7] that if a set B is a complete basis for the one-electron Hilbert space H, then the set of all possible Slater determinants with N electrons that can be constructed from B, will be a complete basis for the antisymmetric N-electron Hilbert space $H^{\wedge N}$. In the truncated case where B is a set of finite size $|B| = M$, this still means that one needs $\binom{M}{N}$ coefficients in order to express any N-electron wave function exactly in the truncated Hilbert space. This is infeasible for even moderately sized M and explains why further approximations are necessary, even though it is much less than the M^N possible Hartree products.

1.2 Second quantization

Second quantization is a formalism where both wave functions and operators are expressed in terms of so-called creation and annihilation operators. This section will only repeat some of the most important concepts of the formalism, derivations and details can be found in, for example, any of Refs. [7–9]. The *Fock space* of antisymmetric wave functions is defined as

$$\mathcal{F} = \bigoplus_{N=0}^{\infty} H^{\wedge N}. \tag{1.7}$$

An element $|\Psi\rangle \in \mathcal{F}$ is hence a linear combination of states containing 0 to ∞ particles. The state $|\rangle$ is a basis for the Fock space $H^{\wedge 0}$ and referred to as *vacuum state*. The set of all Slater determinants that can be constructed from an orthonormal 1-electron basis, together with the vacuum state, form a basis for the Fock space $\mathcal{F}$.
Assuming an orthonormal basis B, we define the *creation operator*[1] $a_p^\dagger$, which is defined by the way it acts on a Slater determinant

$$a_p^\dagger |\phi_q \cdots \phi_s\rangle = |\phi_p \phi_q \cdots \phi_s\rangle . \tag{1.8}$$

It hence "creates" a particle in the sense that a row and a column are added to the Slater determinant. Because (Slater) determinants change sign when interchanging two columns and because they equal zero when two columns are identical, we have the following anticommutation rule

$$\left[a_p^\dagger, a_q^\dagger\right]_+ = 0. \tag{1.9}$$

Similarly, we have the annihilation operator a_p (the adjoint of $a_p^\dagger$), which removes a row and a column from the Slater determinant

$$a_p |\phi_p \phi_q \cdots \phi_s\rangle = |\phi_q \cdots \phi_s\rangle . \tag{1.10}$$

Similarly to eq. (1.9), we have

$$\left[a_p, a_q\right]_+ = 0 \tag{1.11}$$

and by considering how $a_p a_q^\dagger$ and $a_q^\dagger a_p$ act on a Slater determinant, it can be shown that

$$\left[a_p^\dagger, a_q\right]_+ = a_p^\dagger a_q + a_q a_p^\dagger = \delta_{pq}. \tag{1.12}$$

In ON-vector representation, Slater determinants of N particles can elegantly be expressed as

$$|\mathbf{k}\rangle = \left[\prod_{p=1}^{\infty} \left(a_p^\dagger\right)^{k_p}\right] |\rangle \tag{1.13}$$

where the product goes over the whole basis, and $\sum_{p=1}^{\infty} k_p = N$. Importantly, it is possible to express n-body operators using the second quantization formalism by defining them in such a way that they act on the basis of the Fock space $\mathcal{F}$ in the same way as they do in first quantization. Doing this, the Hamiltonian takes the form

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger a_s a_r + h_{\text{nuc}} \tag{1.14}$$

where

$$h_{pq} = \langle \phi_p | \hat{h} | \phi_q \rangle = \int \phi_p^*(\mathbf{x}) \hat{h} \phi_q(\mathbf{x}) d\mathbf{x} \tag{1.15}$$

$$\langle pq \| rs \rangle = \langle \phi_p \phi_q \| \phi_r \phi_s \rangle = \langle \phi_p \phi_q | \phi_r \phi_s \rangle - \langle \phi_p \phi_q | \phi_s \phi_r \rangle \tag{1.16}$$

$$\langle \phi_p \phi_q | \phi_r \phi_s \rangle = \int \phi_p^*(\mathbf{x}_1) \phi_q^*(\mathbf{x}_2) \hat{v}_{1,2} \phi_r(\mathbf{x}_1) \phi_s(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \tag{1.17}$$

and h_{nuc} is the potential energy between the nuclei. This representation is completely equivalent to eq. (1.2), as long as the basis is complete. Observe that the Hamiltonian has no strings of creation/annihilation operators of length 5 or higher, as the action of the one-electron part of the Hamiltonian is characterized by the way it acts on the one-electron basis, and the action of the two-electron part of the Hamiltonian is characterized by the way it acts on the two-electron basis: The tensor product of the one-electron basis functions $|\phi_p\rangle \otimes |\phi_q\rangle$.

[1] although being operators, due to their prevalence, we will always write creation and annihilation operators without the operator-typical hat.

1.2.1 Incomplete and non-orthogonal basis sets

From a computational point of view, it is impractical to use infinite bases, hence we introduce a finite basis to express the Hamiltonian in. To do this, we consider a set consisting of M one-electron functions $A = \{\chi_p\}_{p=1}^{M}$ which we assume to be linearly independent, but not generally orthogonal.

We define the overlap matrix $\mathbf{S}$ with elements

$$S_{pq} = \langle \chi_p | \chi_q \rangle = \int \phi_p^*(\mathbf{x}) \phi_q(\mathbf{x}) d\mathbf{x}. \tag{1.18}$$

One way to form an orthonormal basis B is by symmetric orthonormalization [10]

$$|\phi_p\rangle = \sum_{q=1}^{M} S_{qp}^{-\frac{1}{2}} |\chi_q\rangle \tag{1.19}$$

where $B = \{\phi_p\}_{p=1}^{M}$. This choice of orthonormalization is not unique. The introduction of a basis gives raise to a projection $\hat{P}_B$, projecting the (infinite-dimensional) one-electron basis on the finite-dimensional basis B

$$\hat{P}_B = \sum_{p}^{M} |\phi_p\rangle\langle\phi_p| = \sum_{p,q}^{M} |\chi_p\rangle \, S_{pq}^{-1} \, \langle\chi_q| \, . \tag{1.20}$$

A consequence of this is that Slater determinants expressed using the basis B now cannot have more than M electrons, as that would make the underlying matrix linearly dependent and the Slater determinant equal to zero. Furthermore, ON-vectors are now bitstrings of length M. The use of M orthonormal basis functions induces a new Fock space

$$\mathcal{F}(M)_B = \bigoplus_{N=0}^{M} H^{\wedge N} \tag{1.21}$$

with dimension $\dim(\mathcal{F}(M)) = 2^M$, as can be seen by considering the ON-vector representation. As the basis for the Fock space is precisely the 2^M Slater determinants, we define the projector

$$\hat{P}_{\mathcal{F}(M)_B} = \sum_{i=0}^{2^M-1} |i_{(2)}\rangle\langle i_{(2)}| \tag{1.22}$$

where the subscript $_{(2)}$ stands for the binary, e.g. ON-vector, representation of a Slater determinant.[2] In the new basis, the effect of the Hamiltonian can be expressed by a effective Hamiltonian, which corresponds to the Hamiltonian projected on the new basis.

$$\hat{H}_{\text{eff}} = \hat{P}_{\mathcal{F}(M)_B} \hat{H} \hat{P}_{\mathcal{F}(M)_B} = \sum_{pq}^{M} h_{pq} a_p^\dagger a_q + \frac{1}{4} \sum_{pqrs}^{M} \langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger a_s a_r + h_{\text{nuc}} \tag{1.23}$$

where the excitation operators now are with respect to the new basis B.

Basis sets

A standard way to create a set A, which we used to construct the basis B, is first to introduce a set of spatial one-electron functions $BS = \{\eta_\mu(\mathbf{r})\}_{\mu=1}^{M/2}$, wich are individually referred to as *atomic orbitals (AOs)*, while the set BS is referred to as a *basis set*. From those $M/2$ atomic orbitals, which need to be linearly independent, but not mutually orthogonal or normalized, we can construct M linearly independent *atomic spin orbitals*

$$\chi_\mu(\mathbf{x}) = \begin{cases} \eta_\mu(\mathbf{r}) \, |\alpha\rangle & \text{if } \mu > M/2 \\ \eta_{\mu - M/2}(\mathbf{r}) \, |\beta\rangle & \text{if } \mu \leq M/2 \end{cases} \tag{1.24}$$

[2] For example, for $M = 4$, $|5_{(2)}\rangle$ corresponds to $|0101\rangle$.

which the set A will now consist of. When constructing a basis set BS, one associates a fixed number of atom-specific AOs with each atom. Those AOs will be centered at the nuclear position of that atom. Each individual AO is usually expressed as a linear combination of Cartesian Gaussian functions (GTOs), though other choices exist. The set of AOs used in a polyatomic system will then be the union of the AOs of each individual atom that is part of the system. There are many different basis sets with different levels of accuracy, size and computational cost [8, ch. 7, 8].

Analyticity under geometric perturbations of standard basis sets

Using Gaussian basis sets, we will assume that the overlap matrix $\mathbf{S}$ will be invertible unless we place two nuclei at the same position (in which case the Hamiltonian is not defined).[3] The overlap matrix as function of the nuclear positions $\mathbf{S}(\vec{R})$ will be analytic, which follows from the fact that each integral is a polynomial in the geometric perturbation, which again follows from the fact that every Gaussian basis function can be written as a linear combination of primitive GTOs, the Obara–Saika scheme being a finite linear combination of integrals, and the Gaussian product rule [8, ch. 9]. In addition, due to the linear independence of the overlap matrix $\mathbf{S}(\vec{R})$ (except for the singularities), it will be invertible as well as positive-definite [11], and then $\mathbf{S}^{-\frac{1}{2}}(\vec{R})$ will be analytic. But then the projection $\hat{P}_B(\vec{R})$ will also be analytic, as translation of the basis functions is analytic as well, from which we also see that the projector $\hat{P}_{\mathcal{F}(M)_B}$ is analytic, as Slater determinants simply are sums of antisymmetrized tensor products over the analytic one-electron basis. Finally, we see that, as both the projection $\hat{P}_{\mathcal{F}(M)_B}$ and the Hamiltonian $H(\vec{R})$ are analytic, so is the effective Hamiltonian $\hat{H}_{\text{eff}}(\vec{R})$ as product of analytic functions. We will from now on drop the subscript $_{\text{eff}}$ and simply keep in mind that we are solving the Schrödinger equation in a finite basis.

1.2.2 Wick's theorem

Wick's theorem [7] is a very powerful method to calculate the expectation values of strings of creation and annihilation operators $\hat{A} = abc...xyz$. It states that

$$abc\cdots xyz = \{abc\cdots xyz\}_v + \sum_{\text{singles}} \{\overbracket{abc}\cdots xyz\}_v + \sum_{\text{doubles}} \{\overbracket{abc\cdots xyz}\}_v + \cdots \tag{1.25}$$

where $\{\hat{A}\}_v$ is the *normal ordering* of an operator $\hat{A}$ and $\overbracket{ab}$ is a *contraction*. The normal ordering is a "reordering" of the individual operators such that all annihilation operators stand to the right of all creation operators. It is not unique. The contraction (which turns out to be a number) is defined as

$$\overbracket{ab} = ab - \{ab\}_v. \tag{1.26}$$

Wick's theorem is extremely useful for the calculation of vacuum expectation values. Assume that the operator $\hat{A}$ has an equal number of creation and annihilation operators. It is easy to see that the expectation value of any normal-ordered operator $\langle| \{\hat{A}\}_v |\rangle$ will be zero, as there will always be an annihilation operator on the right, and $a_p |\rangle = 0$. For that reason, only fully contracted terms will contribute.
It should be mentioned that there also is more general form of Wick's theorem, which states that contractions do not need to be considered within products of normal-ordered strings, only between them [9].

1.2.3 The Fermi vacuum

Using Wick's theorem, the matrix element between any two Slater-determinants in the same MO-basis can be effectively evaluated, as we can always write it as a vacuum expectation

[3] From a computational point of view, basis set overcompleteness is a general problem, e.g. some eigenvalues of $\mathbf{S}$ are very small. This is however not a problem in the strictly mathematical sense and can also be avoided by removing some basis functions alltogether.

value

$$\langle \mathbf{l}| \hat{A} |\mathbf{k}\rangle = \langle | \left[\prod_{p=1}^{M} \left(a_p^\dagger\right)^{l_p}\right]^\dagger \hat{A} \left[\prod_{p=1}^{M} \left(a_p^\dagger\right)^{k_p}\right] |\rangle \,. \tag{1.27}$$

However, for a large number of particles, these expectation values become very long. Also, in quantum chemical calculations, one is often interested in matrix elements between the Hartree-Fock states and excited determinants. Wick's theorem keeps its validity upon redefinition of what serves as vacuum simply by redefining what serves as annihilation and creation operators. With respect to a Slater determinant $|\Phi^{\text{SD}}\rangle$, where $i, j, \ldots$ are occupied and $a, b, \ldots$ are virtual orbitals, we have that

$$a_i^\dagger |\Phi^{\text{SD}}\rangle = a_a |\Phi^{\text{SD}}\rangle = 0 \tag{1.28}$$

while $a_i, a_a^\dagger$ acting on $|\Phi^{\text{SD}}\rangle$ will produce a $N-1$ particle Slater determinants. Hence, using $|\Phi^{\text{SD}}\rangle$ as *Fermi vacuum*. The MOs corresponding to virtual orbitals are called *particle states*, and the ones for the occupied orbitals are called *hole states*. $a_i^\dagger$ and a_a serve as *q-annihilation* operators as they annihilate hole/particle states, similarly, a_i and $a_a^\dagger$ are *q-creation* operators. Normal ordering now moves all q-annihilation operators to the right. The only non-zero contractions are

$$\begin{aligned} \overbrace{a_i^\dagger a_j} &= \delta_{ij} \\ \overbrace{a_a a_b^\dagger} &= \delta_{ab}. \end{aligned} \tag{1.29}$$

We will write $\{\}_v$ for normal ordering with respect to the true vacuum, while $\{\}$ will be used with respect to the Fermi vacuum.

1.2.4 Second quantization of the Hamiltonian, normal ordering of the Hamiltonian

As described above, in second quantization, the Hamiltonian reads

$$\hat{H} = \sum_{pq} h_{pq} a_p^\dagger a_q + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger a_s a_r. \tag{1.30}$$

Using Wick's theorem, the strings $a_p^\dagger a_q$ and $a_p^\dagger a_q^\dagger a_s a_r$ can be written as sum of normal-ordered strings. Using Wick's theorem and some algebra, it can be shown that the Hamiltonian can be rewritten as

$$\hat{H} = \sum_{pq} f_{pq} \left\{ a_p^\dagger a_q \right\} + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \left\{ a_p^\dagger a_q^\dagger a_s a_r \right\} + \langle \Phi^{\text{SD}}| \hat{H} |\Phi^{\text{SD}}\rangle \tag{1.31}$$

where f_{pq} is the Fock matrix described in the next section and $|\Phi^{\text{SD}}\rangle$ is a reference Slater determinant. This gives rise to the *normal-ordered Hamiltonian*

$$\hat{H}_N = \hat{H} - \langle \Phi^{\text{SD}}| \hat{H} |\Phi^{\text{SD}}\rangle \,. \tag{1.32}$$

1.3 Hartree-Fock method

The Hartree-Fock method [5, ch. 3] is a variational method that at finding the Slater determinant $|\Phi^{\text{SD}}\rangle$ that minimizes the expectation energy:

$$|\Phi^{\text{HF}}\rangle = \arg\min_{|\Phi\rangle} \langle \Phi| \hat{H} |\Phi\rangle \quad \text{subject to} \quad |\Phi\rangle \text{ being a normalized Slater determinant.} \tag{1.33}$$

As a Slater determinant consists of N orthonormal one-electron functions $\phi_i(\mathbf{x})$, $i = 1, \ldots, N$ that determine it uniquely, solving eq. (1.33) corresponds to finding those N functions.

In practice, one does not aim to find the global minimum of eq. (1.33), but finds the stationary points of the Lagrangian

$$\mathcal{L} = E - \sum_{i,j}^{N} \epsilon_{i,j} \left[\langle \phi_i | \phi_j \rangle - 1 \right] \tag{1.34}$$

where E is the expectation energy of a Slater determinant with orthonormal elements

$$E = \langle \Phi | \hat{H} | \Phi \rangle = \sum_{i}^{N} \langle \phi_i | \hat{h} | \phi_i \rangle + \frac{1}{2} \sum_{i,j}^{N} \langle \phi_i \phi_j || \phi_i \phi_j \rangle \tag{1.35}$$

and the Lagrangian multipliers $\epsilon_{i,j}$ are included to ascertain orthonormality of the MOs and hence normalization of the Slater determinant. In a basis-independent representation, the stationary points are characterized by the following eigenvalue problem [12]:

$$\hat{F}[\phi] \, |\phi_i\rangle = \epsilon_i \, |\phi_i\rangle \tag{1.36}$$

where $i = 1, \ldots, N$ and the action of the Hermitian *Fock operator* $\hat{F}$ is defined in the way it acts on a one-electron function $|\psi\rangle$ as

$$\hat{F}[\phi] \, |\psi\rangle = \left[\hat{h} + \sum_{j=1}^{N} (\hat{J}_j[\phi] - \hat{K}_j[\phi]) \right] |\psi\rangle \tag{1.37}$$

$$\hat{J}_j[\phi] \, |\psi\rangle = \left(\int \frac{\phi_j^* \left(x'\right) \phi_j \left(x'\right)}{|\boldsymbol{r} - \boldsymbol{r}'|} d\mathbf{x}' \right) |\psi\rangle \, , \tag{1.38}$$

$$\hat{K}_j[\phi] \, |\psi\rangle = \left(\int \frac{\phi_j^* \left(x'\right) \psi \left(x'\right)}{|\boldsymbol{r} - \boldsymbol{r}'|} d\mathbf{x}' \right) |\phi_j\rangle \tag{1.39}$$

where we used the notation $[\phi]$ to indicate that the Fock operator is a functional of a set of N one-electron functions. Solving (1.36) analytically is hardly possible, hence we introduce an orthogonal basis:

$$\phi_i(\mathbf{x}) = \sum_{\mu=1}^{M} C_{i,\mu} \chi_\mu(\mathbf{x}) \tag{1.40}$$

where the coefficients $C_{i,\mu}$ are chosen in such a way the basis is orthonormal, and $\chi_\mu(\mathbf{x})$ is a set of atomic spin orbitals. The $M \times M$ matrix $\mathbf{C}$ with elements $C_{i,\mu}$ is referred to as *coefficient matrix*. The energy of a Slater determinant is then given as

$$\begin{aligned} \langle \Phi | \hat{H} | \Phi \rangle &= \sum_{i}^{N} \langle \phi_i | \hat{h} | \phi_i \rangle + \frac{1}{2} \sum_{i,j}^{N} \langle \phi_i \phi_j || \phi_i \phi_j \rangle \\ &= \sum_{a,b}^{M} \sum_{i}^{N} C_{i,\mu}^* C_{i,\nu} \left\langle \chi_a \right| h \left| \chi_b \right\rangle + \frac{1}{2} \sum_{\mu,\nu,\sigma,\tau}^{M} \sum_{i,j}^{N} C_{i,\mu}^* C_{j,\nu}^* C_{i,\sigma} C_{j,\tau} \langle \chi_\mu \chi_\nu || \chi_\sigma \chi_\tau \rangle. \end{aligned} \tag{1.41}$$

With the problem reduced to finding the coefficients $C_{i,\mu}$, the solution of eq. (1.33) gives raise to the self-consistent field method

$$\mathbf{F}[\mathbf{C}]\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\varepsilon} \tag{1.42}$$

where the *overlap matrix* $\mathbf{S}$ is given by eq. (1.18) and the Hermitian *Fock matrix* $\mathbf{F}$

$$S_{pq} = \langle \chi_p | \chi_q \rangle \tag{1.43}$$

$$F_{pq}[\mathbf{C}] = \langle \chi_p | \, \hat{h} \, | \chi_q \rangle + \sum_{r,s}^{M} D_{rs} \langle \chi_p \chi_s || \chi_q \chi_r \rangle \tag{1.44}$$

where we defined the *density matrix* $\mathbf{D} = \mathbf{C}\mathbf{C}^\dagger$. Eq. (1.42) is a generalized Hermitian eigenvalue problem which can be solved efficiently for $\mathbf{C}$, cf. sec. 5.1.2, p. 60. However,

it is an iterative procedure as $\mathbf{F}$ depends on $\mathbf{C}$ and requires a start guess $\mathbf{C}_0$. Having solved the HF-equations corresponds to $\mathbf{F}$ being diagonal in the MO-basis.
To obtain the Hartree-Fock wave function, those N molecular orbitals are chosen that minimize eq. (1.33), which correspond to those with the eigenvalues ϵ_i, $i = 1, \ldots, N$ where $\epsilon_1 \leq \epsilon_2 \leq \cdots \leq \epsilon_M$. A set of orthonormal one-electron functions that correspond (by. eq. (1.40)) to the columns of $\mathbf{C}$, is called *molecular orbitals (MOs)*. *Canonical MOs* are the set of MOs that are obtained when solving eq. (1.42). In addition to the N *occupied* MOs that make up the Hartree-Fock state $|\Phi^{\text{HF}}\rangle$, there are also $M - N$ *virtual* MOs. Based on this discussion, we will from now on use the notation

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}^o & \mathbf{C}^v \end{bmatrix} \tag{1.45}$$

where $\mathbf{C}^{o/v}$ is the part of the coefficient matrix containing the occupied/virtual orbitals.

Important attributes of the HF-state and MOs

There is a number of attributes that are relevant later on, hence they are briefly discussed here.

1. Slater determinants built from orthonormal MOs are invariant under unitary operations within those MOs. To see this, let $\tilde{\phi}_p = \sum_p^N \phi_q U_{qp}$ with $\mathbf{U}$ unitary. Let $\mathbf{A}$ be a matrix with elements $A_{pq} = \phi_p(\mathbf{x}_q)$, and $\tilde{\mathbf{A}}$ correspondingly. Observing that $1/\sqrt{N!}|\mathbf{A}| = |\Phi^{\text{SD}}\rangle$, we see that

$$|\tilde{\Phi}^{\text{SD}}\rangle = |\tilde{\mathbf{A}}| = |\mathbf{AU}| = |\mathbf{A}||\mathbf{U}| = e^{i\pi\theta}\,|\Phi^{\text{SD}}\rangle \tag{1.46}$$

   where $e^{i\pi\theta}$ is simply a phase which does not change the physical state (for real $\mathbf{U}$, $\theta \in \{0, 1\}$).

2. For canonical MOs, a direct consequence of eq. (1.42) is that $\mathbf{F}$ is diagonal in a basis of canonical MOs. We will refer to those diagonal elements as $F_{pp} = \epsilon_p$. Unitary operations applied to the occupied MOs and the virtual MOs, respectively, put the Fock matrix in a block-diagonal form. In this MO-basis, simple calculations show that $\langle\Phi^{\text{SD}}|\, a_a^\dagger a_i\, |\Phi^{\text{SD}}\rangle = F_{ia} = 0$. This is *Brillouin's theorem* [5].

3. For Slater determinants, motion of electrons with opposite spins is uncorrelated and thus they are non-interacting [5]. The Hartree-Fock theory is thus a mean field theory, as each electron is subject to the mean field of the other electrons. The Hartree-Fock method is necessarily an approximation due to lack of electron correlation. For accurate results, it is necessary to use methods that are more advanced than the Hartree-Fock method. Those methods that build on an (approximate) Hartree-Fock wave function, are referred to as *post-Hartree-Fock* methods.

4. Solving the SCF equations without any restrictions on the MOs gives raise to the *general* Hartee-Fock method (GHF), and the resulting Slater determinant will not necessarily be an eigenfunction of the spin projection $\hat{S}_z$, and each MO will have an α and a β spin contribution. Insisting that each MO has definite α or β spin, gives raise to the *unrestricted* Hartree-Fock method (UHF), which turns it into a spin projection eigenfunction. Finally, the *restricted* Hartree-Fock method imposes that the resulting Slater determinant is a spin $\hat{S}^2$ eigenfunction by producing pairs of MOs that have the same spatial part. Using the UHF and RHF methods is justified because the exact wave function is an eigenfunction of both spin and spin projection and because these restrictions reduce computational cost. In addition, we restrict the MOs to be *real* functions. This is justified because the Hamilton operator $\hat{H}$ is Hermitian and we can thus choose its eigenfunctions to be real. Using real AOs, this means that the coefficient matrix $\mathbf{C}$ is real. Finally, even though the exact wave function can be chosen to be real and an eigenfunction of both $\hat{S}^2$, $\hat{S}_z$, this does not mean that the RHF wave function

has the same energy as the GHF/UHF wave function: The GHF/UHF wave function can have a lower energy. This has been coined "Löwdin's symmetry dilemma" [13] - physically correct symmetries in the Hartee-Fock approximation may lead to higher energies of the Slater determinants.

5. In addition to spin symmetry, there is also point group symmetry of molecules to consider, which can further reduce computational cost. For molecules with point group symmetry, one may define the occupancy of the MOs in each irreducible representation beforehand. This leads to block-diagonal Fock and overlap matrices and reduces computational cost by several orders of magnitude. How the occupancy of the MOs is chosen, impacts the energy of the Slater determinant. It is noteworthy that there are cases where, along a given trajectory, the Slater determinant with lowest energy will not necessarily be the same in terms of occupancy per irreducible representation. This changes the canonical orbitals as well as their eigenvalues discontinuously, and with BeH_2, we will discuss such a case in the results section.

1.3.1 Analyticity of the Hartree-Fock wave function

The Fock operator $\hat{F}$ is a integro-differential operator and much more complicated than the linear Hamilton operator. When we have a diagonal Fock operator, that only means that we have found a stationary point of the Lagrangian in eq. (1.34), but not the Slater determinant with the lowest energy. The resulting solution $\mathbf{C}$ depends heavily on the initial starting guess of the SCF procedure. More generally, we can ask what can be said about the analyticity of a solution to the HF equations as function of a geometric perturbation, independently of whether it is the ground state or not. As the basis is analytic as function of a geometric perturbation, the question boils down to wether the solutions of the stationarity condition

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \operatorname{Re}(C_{i,\mu})} = \frac{\partial \mathcal{L}}{\partial \operatorname{Im}(C_{i,\mu})} = 0 \quad i = 1, \dots, N; \mu = 1, \dots, M \\ \frac{\partial \mathcal{L}}{\partial \epsilon_{i,j}} = 0 \quad i, j = 1, \dots, N \end{aligned} \tag{1.47}$$

(where $\mathcal{L}$ is the Lagrangian in eq. (1.34) in a basis) are analytic as a function of the geometric perturbation. In general, the answer to this question is no. The Lagrangian is a function of not only the coefficients $C_{i,\mu}$, but also their complex conjugates, and as, in general, $\frac{\partial \mathcal{L}}{\partial C^*_{i,\mu}} \neq 0$, the Cauchy-Riemann conditions are not fulfilled and thus the Lagrangian itself is not analytic. [4]
Solutions can disappear, for both UHF and RHF, even in case of a single-parameter dependency. The analytic variant of the Hartree-Fock equations, where $C_{i,\mu}$ is allowed to be complex, but $C^*_{i,\mu}$ is replaced with $C_{i,\mu}$, has a constant number of solutions at all geometries [14, 15], and the authors observed smooth and continuous potential energy surfaces as well as coinciding solutions with the HF equations whenever solutions existed, albeit not proving analyticity. Hence, we can generally assume that the coefficients $C_{i,\mu}$ can be chosen to be continuous and analytic for non-disappearing solutions, except for an isolated set of geometries. More generally, the question boils down to when the solution to a set of polynomial equations in several variables are analytic in the parameters, which goes beyond the scope of this thesis.

It should be noted that this discussion is of theoretical nature. Although solutions to the HF equations, and especially the ground state are not generally analytic and can even disappear, they are often analytic in practice, as can, for example, be seen from the general smoothness of HF energies.

[4] Indeed, we "tricked" in the way we defined the stationarity condition, defining the derivatives with respect to real variables $\operatorname{Im}(C_{i,\mu}), \operatorname{Re}(C_{i,\mu})$ to obtain well-defined derivatives.

1.4 Change of basis operations

Let us have two sets of MOs $A = \{|^w\phi_p\rangle\}_{p=1}^M, B = \{|^x\phi_p\rangle\}_{p=1}^M$ spanning the same space with $|^x\phi_p\rangle = \sum_q |^w\phi_q\rangle U_{qp}$ and $\mathbf{U}$ unitary. To set A, we associate a set of creation and annihilation operators $a_p^\dagger, a_p$, and to set B we associate $b_p^\dagger, b_p$. Assume we want to express a Slater determinant expressed in terms of the MOs from set B as a linear combination of Slater determinants expressed in the terms of MOs from set A. As $\mathbf{U}$ is unitary, we can write it in terms of an anti-Hermitian matrix $e^{-\kappa}$. It can now be shown [8] that the creation and annihilation operators between those sets are related as

$$\begin{aligned} b_p^\dagger &= \exp(-\hat{\kappa}) a_p^\dagger \exp(\hat{\kappa}) \\ b_p &= \exp(-\hat{\kappa}) a_p \exp(\hat{\kappa}) \\ |^x\Phi^{\mathrm{SD}}\rangle &= \exp(-\hat{\kappa}) \, |^w\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{1.48}$$

where $|^x\Phi^{\mathrm{SD}}\rangle$ means that we "replaced" a_p with b_p, and

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} a_p^\dagger a_q \tag{1.49}$$

$$\kappa_{pq} = [\ln(\mathbf{U})]_{pq}. \tag{1.50}$$

On a side note, from the last line of eq. (1.48) and the linearity of $\exp(-\hat{\kappa})$, we see that this applies to linear combinations of Slater determinants and hence all wave functions.

1.5 Generalized Slater-Condon rules and Thouless' theorem: Non-orthogonal Slater determinants

In this thesis, we will calculate matrix elements of Slater determinants in different basis sets, with MOs that are not mutually orthogonal, and the normal Slater-Condon rules [5] do not apply. The *generalized Slater-Condon rules* are used for calculating the matrix element between two non-orthogonal Slater determinants [11]. Let us consider two sets of orthonormal molecular orbitals $\{^w\phi_i\}_{i=1}^N$, $\{^x\phi_i\}_{i=1}^N$ that are not mutually orthonormal,

$$\begin{aligned} \langle^w\phi_i|^w\phi_j\rangle &= \delta_{ij} \\ \langle^x\phi_i|^x\phi_j\rangle &= \delta_{ij} \\ \langle^w\phi_i|^x\phi_j\rangle &\neq \delta_{ij}. \end{aligned} \tag{1.51}$$

We define the overlap matrix ${}^{wx}\mathbf{S}$ with elements ${}^{wx}S_{ij} = \langle^w\phi_i|^x\phi_j\rangle$. Observe that ${}^{wx}\mathbf{S}$ is generally not symmetric. Given two Slater determinants built from these MOs,

$$\begin{aligned} |^w\Phi^{\mathrm{SD}}\rangle &= |^w\phi_1 \ldots {}^w\phi_N\rangle \\ |^x\Phi^{\mathrm{SD}}\rangle &= |^x\phi_1 \ldots {}^x\phi_N\rangle \end{aligned} \tag{1.52}$$

the overlap between those two Slater determinants is now given by

$$\langle^w\Phi^{\mathrm{SD}}|^x\Phi^{\mathrm{SD}}\rangle = |{}^{wx}\mathbf{S}|. \tag{1.53}$$

Consider now the one-electron operator $\hat{H}_1 = \sum_{i=1}^N \hat{h}(i)$. Its matrix element is given by

$$\langle^w\Phi^{\mathrm{SD}}|\hat{H}_1|^x\Phi^{\mathrm{SD}}\rangle = \sum_{i,j=1}^N {}^{wx}h_{ij} \, {}^{wx}S_{ij}^{(1)} \tag{1.54}$$

where ${}^{wx}\mathbf{S}^{(1)}$ is the is *cofactor* matrix of ${}^{wx}\mathbf{S}$, e.g. ${}^{wx}S_{ij}^{(1)}$ is the determinant of ${}^{wx}\mathbf{S}$ when removing row i and column j. Similarly, for the two-electron operator $\hat{H}_2 = \sum_{i<j}^N \hat{v}(i,j)$, we find

$$\langle^w\Phi^{\mathrm{SD}}|\hat{H}_2|^x\Phi^{\mathrm{SD}}\rangle = \sum_{i<j}^N \sum_{k<l}^N {}^{wx}\langle ij\|kl\rangle \, {}^{wx}S_{ij,kl}^{(2)} \tag{1.55}$$

where ${}^{wx}\mathbf{S}^{(2)}$ is the second-order cofactor matrix, e.g. ${}^{wx}S^{(2)}_{ij,kl}$ is the determinant of ${}^{wx}\mathbf{S}$ when removing rows i, j and columns k, l. It is straightforward to show that the standard Slater-Condon rules [5, ch. 2] can be derived from the Generalized Slater-Condon rules, as ${}^{wx}\mathbf{S}$ will be diagonal with elements $S_{ii} \in \{0, 1\}$. A proof of the generalized Slater-Condon rules is found in the appendix, sec. 10.3, p. 92. The GSC rules allow for the use of correlated calculations with non-orthogonal Slater determinants, which we briefly discuss in section 1.10, p. 22.
Slightly related, we have Thouless' theorem [16]. Let $|{}^{w}\Phi^{\mathrm{SD}}\rangle$, $|{}^{x}\Phi^{\mathrm{SD}}\rangle$ be two arbitrary Slater determinants (expressed in different bases). Then it can be shown that the following formula holds, unless $\langle {}^{w}\Phi^{\mathrm{SD}}|{}^{x}\Phi^{\mathrm{SD}}\rangle = 0$:

$$|{}^{x}\Phi^{\mathrm{SD}}\rangle \propto e^{\hat{T}_1} |{}^{w}\Phi^{\mathrm{SD}}\rangle \tag{1.56}$$

where

$$\hat{T}_1 = \sum_{ia} t_i^a a_a^\dagger a_i \tag{1.57}$$

and the parameters t_i^a need to be determined.

1.6 Configuration interaction

The Hartree-Fock method is not exact. Because we know that the set of all possible Slater determinants that can be created from a set of molecular orbitals is a basis for the truncated Hilbert space, finding the coefficients c_n that minimize the energy expectation value, gives raise to the exact wave function within a basis set

$$|\Psi^{\mathrm{FCI}}\rangle = \sum_n c_n^{\mathrm{FCI}} |\Phi_n^{\mathrm{SD}}\rangle \tag{1.58}$$

where the sum goes over all Slater determinants and $\mathbf{c}^{\mathrm{FCI}}$ is chosen such that

$$\mathbf{c}^{\mathrm{FCI}} = \arg\min_{\mathbf{c}} \left(\sum_{nm} c_m c_n^* \langle \Phi_n^{\mathrm{SD}}| \hat{H} |\Phi_m^{\mathrm{SD}}\rangle \right) \quad \text{subject to} \quad \mathbf{c}^\dagger \mathbf{c} = 1. \tag{1.59}$$

This approach is called configuration interaction (CI). In solving this equation, a matrix of dimension $\binom{M}{N} \times \binom{M}{N}$ needs to be diagonalized. While there exist very efficient ways to do avoid diagonalizing the whole matrix [8], this approach is nevertheless not feasible even for small molecules. One alternative is to only include singly and doubly excited determinants, as, by the Slater-Condon rules $\langle \Phi^{\mathrm{SD}}| \hat{H}\hat{A} |\Phi^{\mathrm{SD}}\rangle = 0$ when $\hat{A}$ is an operator that brings three or more electrons from occupied into virtual orbitals - this seems reasonable, as doubly excited determinants (doubles) thus should the most important contribution to the energy. This approach is called CISD. However, CISD is not size consistent and performs worse for larger systems. This also applies to all other truncated CI methods. While the FCI method is independent of the choice of MOs, all truncated CI methods are not and require a reasonable choice - usually, the HF wave function is chosen. There are size consistent methods that are about as expensive as CISD, hence truncated CI methods are rarely used nowadays. The FCI wave function is exact in a given basis and thus remains important in order to check how well approximate methods work.

1.7 Coupled cluster theory

Coupled cluster (CC) theory is a post Hartree-Fock method that tries to improve the wave function beyond the single determinant picture, while maintaining a polynomial scaling of the operation count in the number of electrons/number of basis functions. We will here describe the most important aspects of CC theory, including unitary coupled cluster theory. Here, $|\Phi^{\mathrm{SD}}\rangle$ will be a single Slater determinant. It is customary that $|\Phi^{\mathrm{SD}}\rangle = |\Phi^{\mathrm{HF}}\rangle$ is the ground HF-state as these serve as a good reference and give computational advantages compared to more general Slater determinants, but the theory remains the same independent of the choice of the reference determinant $|\Phi^{\mathrm{SD}}\rangle$.

1.7.1 The exponential ansatz

In coupled cluster theory, the wave functions takes the form

$$|\Psi\rangle \approx e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \hat{T}^n |\Phi^{\mathrm{SD}}\rangle \tag{1.60}$$

where the *cluster operator* $\hat{T}$ is an excitation operator of the form

$$\hat{T} = \sum_{i=1}^{\infty} T_i = \sum_{\mu} t_\mu \hat{\tau}_\mu = \sum_{ia} t_i^a a_a^\dagger a_i + \sum_{iajb} \frac{1}{4} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i + \dots \tag{1.61}$$

and $\hat{\tau}_\mu |\Phi^{\mathrm{SD}}\rangle = |\mu\rangle$ represents an excited determinant of any order. We will sometimes denote the collection of all amplitudes t_i^a, t_{ij}^{ab} by a vector $\mathbf{t}$. It should be noted that the excitation operators commute, that is $[\hat{\tau}_\mu, \hat{\tau}_\nu] = 0$ for all μ, ν. Furthermore, the Taylor expansion of $e^{\hat{T}}$ is finite, as contributions for $n > N$ where N is the number of electrons are zero. Assuming the ground state wave function can indeed be represented in this form, insertion in the Schrödinger equation yields

$$\hat{H} |\Psi\rangle = \hat{H} e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = E e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = E |\Psi\rangle . \tag{1.62}$$

Assuming the reference Slater determinant $|\Phi^{\mathrm{SD}}\rangle$ to be normalized, we have intermediate normalization, and thus

$$\langle \Phi^{\mathrm{SD}}| \hat{H} |\Psi\rangle = E. \tag{1.63}$$

By considering the operator $\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$, the so-called *similarity-transformed Hamiltonian*, it is straightforward to show that a true eigenfunction of the Hamiltonian satisfies

$$\langle \Phi^{\mathrm{SD}}| \bar{H} |\Psi\rangle = E. \tag{1.64}$$

This is the energy equation. In addition, we have the *amplitude equations*

$$\langle \mu| \bar{H} |\Psi\rangle = 0 \tag{1.65}$$

where $\langle \mu| = \langle \Phi^{\mathrm{SD}}| \hat{\tau}_\mu^\dagger$. This set of equations is called the *linked coupled-cluster equations* and needs to be solved in order to obtain the amplitudes t_μ.

1.7.2 Truncation of the cluster operator

The number of coefficients t_μ is identical to the number of coefficients in a full CI calculation when

$$\hat{T} = \sum_{i=1}^{N} \hat{T}_i \tag{1.66}$$

where N is the number of electrons. In that case, we can recover the full CI solution, but the number of parameters will then scale exponentially with the system size. A remedy is to truncate the cluster operator at some order k, such that

$$\hat{T} = \sum_{i=1}^{k} \hat{T}_i = \sum_{\mu \in S,D,T,\dots,K} t_\mu \hat{\tau}_\mu \tag{1.67}$$

where $k < N$. The case $k = 2$ yields the *coupled-cluster singles-and-doubles* (CCSD) model. From the Taylor expansion of $e^{\hat{T}}$, one can see that the resulting wave function $|\Psi\rangle$ can contain contributions from all determinants, unlike truncated CI expansions. The exponential ansatz makes the CCSD wave function size consistent [8], which is one of the biggest advantages of the method. As the cluster operator is truncated, the variational solution, with $|\tilde{\Psi}\rangle = \exp\left(\sum_i^k \hat{T}_i\right) |\Phi^{\mathrm{SD}}\rangle$ that minimizes the expectation value of the energy

$$E_{VCC} = \min_{\hat{T}} \frac{\langle \tilde{\Psi}| H |\tilde{\Psi}\rangle}{\langle \tilde{\Psi}|\tilde{\Psi}\rangle} \tag{1.68}$$

will not be the true ground state, and we cannot expect equations 1.64 and 1.65 to be correct for that state. However, as eq. (1.68) becomes intractable to solve, the default approach is to solve 1.64 and 1.65, where only singles and doubles are projected in order for the number of amplitude equations to match the number of parameters. Thus, the energy is no longer variational, which is a serious drawback of truncated coupled-cluster methods. While there are some theoretical studies, such as Ref. [17, 18] that solve the variational equation (1.68), giving raise to the *variational coupled cluster* (VCC) method, it is not considered a standard method due to its high numerical cost.

1.7.3 The similarity-transformed Hamiltonian

We remember that the similarity-transformed Hamiltonian, which takes a central role in coupled cluster theory, is given by

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}. \tag{1.69}$$

To evaluate this, we could use the Taylor expansions of $e^{\pm T}$, or one can use the *Baker-Campbell-Hausdorff* formula (BCH formula), which, when applied to $\bar{H}$, gives

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{4!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}] + \cdots. \tag{1.70}$$

Direct calculation of the commutators using commutator rules for creation and annihilation operators [8, ch. 1] shows that each commutator of the Hamiltonian with the excitation operators $\hat{T}$ removes one general index annihilation/creation operator from the Hamiltonian $\hat{H}$. Because the Hamiltonian $\hat{H}$ has four general index operators, that means that all commutators nested more than four times disappear independent of the truncation k of the cluster operator, which gives

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{4!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}]. \tag{1.71}$$

1.7.4 Solving the coupled cluster equations

In order to determine the excitation coefficients t_μ, the amplitude equations (eq. (1.65)) need to be solved. This requires explicit expressions in terms of the excitation coefficients t_μ. In practice, one does not consider the similarity transformed Hamiltonian, but the similarity transformed normal-ordered Hamiltonian $\bar{H}_N = \hat{H} - \langle \Phi^{\mathrm{SD}} | \hat{H} | \Phi^{\mathrm{SD}} \rangle$, which simplifies the evaluation of the equations. This can be done by repeatedly applying (the generalized) Wick's theorem. This is a tedious task and the resulting equations will not be given here, as they are side-long. The derivation of these equations can also done using computer-assisted tools or diagrammatic methods. In particular, the `Drudge` and `Gristmill` packages [19] have functionalities for evaluation and code generation of commutators and Fermi-vacuum expectation values of second-quantized operators. The obtained expressions are nonlinear and are solved iteratively in practice. For CCSD, they scale as $O(M^8)$, but a rewrite in terms of intermediates reduces this scaling to $O(M^6)$, with the most expensive term scaling as $O(M_v^4 M_o^2)$ with $M_{o/v}$ being the number of occupied/virtual orbitals. An explicit form of the resulting equations is given in, for example, Ref. [20]. In the iterative process, Newtons method can be used with an initial guess of the amplitudes based on the MP1 wave function (described in the next section), where it is usual that the Jacobian

$$J_{\mu\nu}(\mathbf{t}) = \frac{\partial}{\partial \nu} \langle \mu | \bar{H}(\mathbf{t}) | \Phi^{\mathrm{SD}} \rangle = \langle \mu | e^{-\hat{T}(\mathbf{t})} \left[\hat{H}, \hat{\tau}_\nu \right] e^{\hat{T}(\mathbf{t})} | \Phi^{\mathrm{SD}} \rangle \tag{1.72}$$

is approximated by a constant diagonal matrix forming the dominating part of the Jacobian when using canonical orbitals [8, ch. 13]:

$$J_{\mu\mu}(\mathbf{t}) = \epsilon_\mu \tag{1.73}$$

where $\epsilon_{aibj} = \epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j$, $\epsilon_{ai} = \epsilon_a - \epsilon_i$.

1.7.5 Equation of Motion coupled cluster

For a truncated cluster operator $\hat{T}$, the state $|\Psi\rangle = e^{\hat{T}} |\Phi^{\text{SD}}\rangle$ is the approximate ground state of the Hamiltonian $\hat{H}$. However, we can just as well consider it to be the true solution to the similarity transformed Hamiltonian $\bar{H}$, which is an approximation to the true Hamiltonian. In the *equation of motion coupled cluster* (EOM-CC) approach, one projects the similarity transformed Hamiltonian on all excited determinants up to order k, where k is the truncation rank of the cluster operator. Thus, we get a projection operator

$$\hat{P} = \sum_{\mu \in S,D,\ldots} |\mu\rangle\langle\mu| + |\Phi^{\text{SD}}\rangle\langle\Phi^{\text{SD}}| \tag{1.74}$$

$$\bar{H}' = \hat{P}\bar{H}\hat{P} \tag{1.75}$$

giving raise to a matrix $\bar{\mathbf{H}}'$ with matrix elements $\bar{H}'_{\mu\nu} = \langle\mu|\, \bar{H}\, |\nu\rangle$. Because eq. (1.65) holds, we have that $\langle\mu|\, \bar{H}'\, |\Phi^{\text{SD}}\rangle = 0$, thus $|\Phi^{\text{SD}}\rangle$ is an eigenstate of $\bar{H}'$. Because $\bar{H}'$ is not Hermitian, the left and right eigenstates are not the same, but have the same eigenvalues, and there is a left eigenstate such that

$$\langle\tilde{\Psi}|\, E = \langle\tilde{\Psi}|\, \bar{H}' \tag{1.76}$$

where

$$\langle\tilde{\Psi}| = \langle\Phi^{\text{SD}}|\, (\mathbb{1} + \hat{\Lambda}_1 + \hat{\Lambda}_2 + \ldots) = \langle\Phi^{\text{SD}}|\, \left(\mathbb{1} + \sum_{ia} \lambda_i^a a_i^\dagger a_a + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab} a_i^\dagger a_j^\dagger a_a a_b + \ldots\right) \tag{1.77}$$

and we see that $\langle\tilde{\Psi}|\Psi\rangle = 1$. Using CCSD, $k = 2$ and the similarity transformed Hamiltonian is expressed in the basis of singles and doubles, hence $\hat{\Lambda}_l = 0$ for $l > k$. EOM-CC can be used to find approximate excited states. The EOM-CC Hamiltonian $\bar{H}'$ has left and right eigenvalues

$$\begin{aligned} \langle\tilde{\Psi}^{(j)}|\, E^{(j)} &= \langle\tilde{\Psi}^{(j)}|\, \bar{H}' \\ E^{(j)}\, |\Psi^{(j)}\rangle &= \bar{H}'\, |\Psi^{(j)}\rangle \end{aligned} \tag{1.78}$$

where

$$\begin{aligned} \langle\tilde{\Psi}^{(j)}| &= \langle\Phi^{\text{SD}}|\, \hat{\mathcal{L}}^{(j)} \\ |\Psi^{(j)}\rangle &= \hat{\mathcal{R}}^{(j)}\, |\Phi^{\text{SD}}\rangle \end{aligned} \tag{1.79}$$

are biorthogonal:

$$\langle\tilde{\Psi}^{(j)}|\Psi^{(k)}\rangle = \delta_{jk} \tag{1.80}$$

and $\hat{\mathcal{L}}^{(j)}, \hat{\mathcal{R}}^{(j)}$ are deexcitation and excitation operators up to truncation order k, respectively. The states

$$e^{\hat{T}}\hat{\mathcal{R}}^{(j)}\, |\Phi^{\text{SD}}\rangle = \hat{\mathcal{R}}^{(j)} e^{\hat{T}}\, |\Phi^{\text{SD}}\rangle \tag{1.81}$$

are approximate excited state solutions of the real Hamiltonian. While the left and right eigenstates are biorthogonal, the right eigenstates are not mutually orthogonal among each other. Furthermore, eigenvalues can be complex, which is usually not a problem in practice [8, ch. 13]. The error in the energies does not stem from the truncation of the cluster operator, but from the projection of the similarity-transformed Hamiltonian on the excited states up to order $k < N$. If one were to project on the all possible excited states, this would correspond to a similarity transform of the FCI Hamiltonian matrix, an operation that conserves eigenvalues.

1.7.6 Non-canonical reference determinants

In this thesis, we consider the use of reference states that are not canonical HF-solutions with a non-diagonal, but block-diagonal Fock matrix $f_{ij} \neq 0$, $f_{ab} \neq 0$, $f_{ai} = 0$, as well as determinants that are not HF-solutions at all, with non-block diagonal Fock matrices $f_{ai} \neq 0$. One might expect that the latter is not a suitable reference state for coupled cluster calculations. However, by Thouless' theorem (cf. sec. 1.5), any Slater determinant $|\Phi^{\text{SD}}\rangle$ can be turned into another Slater determinant $|\tilde{\Phi}^{\text{SD}}\rangle$ by application of an $e^{\hat{T}_1}$ excitation operator: $|\tilde{\Phi}^{\text{SD}}\rangle \propto e^{\hat{T}_1} |\Phi^{\text{SD}}\rangle$ given that $\langle\tilde{\Phi}^{\text{SD}}|\Phi^{\text{SD}}\rangle \neq 0$. This makes the CCSD wave functions relatively independent of the choice of the reference determinant, as long as the reference used is not too different from a "good" reference determinant.

1.7.7 Unitary coupled cluster methods

In the *unitary coupled cluster* (UCC) approach, the wave function is parametrized as

$$|\Psi\rangle = e^{\hat{\sigma}} |\Phi^{\text{SD}}\rangle \tag{1.82}$$

and the Hermitian similarity transformed Hamiltonian reads

$$\bar{H} = e^{-\hat{\sigma}} \hat{H} e^{\hat{\sigma}} \tag{1.83}$$

where $\hat{\sigma} = \hat{T} - \hat{T}^\dagger$ is anti-Hermitian: $\hat{\sigma}^\dagger = \left(\hat{T} - \hat{T}^\dagger\right)^\dagger = -\left(\hat{T} - \hat{T}^\dagger\right) = -\hat{\sigma}$, hence $e^{\hat{\sigma}}$ is unitary. Thus, the unitary coupled cluster approach yields variational energies

$$E_{\text{UCC}} = \min_{\hat{T}} \left(\langle \Phi^{\text{SD}}| e^{-\hat{\sigma}} \hat{H} e^{\hat{\sigma}} |\Phi^{\text{SD}}\rangle \right) \tag{1.84}$$

as the wave function is normalized independently of $\hat{T}$: $\langle\Psi|\Psi\rangle = \langle\Phi^{\text{SD}}|\Phi^{\text{SD}}\rangle = 1$. It has been theorized that UCC and VCC (cf. eq. (1.68)) have the same (global) energy minimum, but both theoretical [17] and practical [17, 18] studies have shown that this is not the case. Furthermore, the unitary coupled cluster method has been shown to be likely superior to the standard coupled cluster method [21].
On classical computers, the unitary coupled cluster approach is not used as a standard method as the similarity transformed Hamiltonian (eq. (1.83)) has a BCH-expansion that does not terminate and no projection approach as in conventional coupled cluster theory (cf. eq. (1.65)) is possible, as the operator is $e^{-\left(\hat{T}-\hat{T}^\dagger\right)}$ is neither a pure excitation or de-excitation operator. However, unitary coupled cluster methods and methods derived from it work well on quantum computers, which is discussed in more detail in section 3.5, p. 33.

1.8 Perturbation theory

An alternative to CISD and CCSD is to use perturbation theory. In *Møller-Plesset* perturbation theory, we write the Hamiltonian as

$$\hat{H} = \hat{F} + \hat{\Phi} \tag{1.85}$$

where $\hat{F}$ is the Fock operator and $\hat{\Phi}$ is the *fluctuation potential*. Assuming the perturbation $\hat{\Phi}$ to be "small", we can write the eigenvalues $E(\lambda)$ and eigenvectors $|\Psi(\lambda)\rangle$ of $\hat{H} = \hat{F} + \lambda\hat{\Phi}$ as Taylor series

$$E(\lambda) = \sum_{n=0}^{\infty} \frac{1}{n!} \lambda^n \frac{\partial^n E(\lambda)}{\partial \lambda^n}|_{\lambda=0} = \sum_{n=0}^{\infty} \lambda^n E^{(n)} \tag{1.86}$$

$$|\Psi(\lambda)\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \lambda^n \frac{\partial^n |\Psi(\lambda)\rangle}{\partial \lambda^n}|_{\lambda=0} = \sum_{n=0}^{\infty} \lambda^n |\Psi^{(n)}\rangle \tag{1.87}$$

where λ is set to 1 to obtain the desired solution. We will not go through the derivation, but only present the solution to the first-order correction of the wave function. Remembering that $|\Psi^{(0)}\rangle = |\Phi^{\text{HF}}\rangle$ is a Slater determinant, the first order correction to the wave function is [8, ch. 14]

$$|\text{MP1}\rangle = |\Psi^{(1)}\rangle = \hat{T}_2^{(1)} |\Psi^{(0)}\rangle \tag{1.88}$$

$$\hat{T}_2^{(1)} = \sum_{\mu \in D} t_\mu^{(1)} \hat{\tau}_\mu = \sum_{\substack{a>b \\ i>j}} t_{ij}^{ab(1)} a_a^\dagger a_i a_b^\dagger a_j \tag{1.89}$$

where

$$t_{ij}^{ab(1)} = -\frac{\langle \Psi^{(0)}| \left[a_j^\dagger a_b a_i^\dagger a_a, \hat{H}\right] |\Psi^{(0)}\rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{1.90}$$

and the second order correction to the energy is (the first order correction to the energy gives simply the Hartree-Fock energy)

$$E^{(2)} = \langle \Psi^{(0)} | \hat{H} | \Psi^{(1)} \rangle \tag{1.91}$$

Obtaining $E^{(2)}$ and $|\Psi^{(1)}\rangle$ generally scales as $O(M^5)$ and is thus chaper than CCSD, and is hence the method of choice to obtain improved energies and wave functions. MP perturbation theory does not give variational energies, and the radius of convergence of $\hat{H}(\lambda)$ is often less than 1 (see also sec. 2.1), hence perturbation theory may not converge and give wrong results.

1.9 1-RDM and natural orbitals

Assume that we have found the exact ground state wave function $|\Psi\rangle$. The corresponding wave function in position/spin space is given by

$$\langle \vec{\mathbf{x}} | \Psi \rangle = \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{1.92}$$

We define the *density matrix*

$$\Gamma(\mathbf{x}_1', \ldots, \mathbf{x}_N'; \mathbf{x}_1, \ldots, \mathbf{x}_N) = \Psi^*(\mathbf{x}_1', \ldots, \mathbf{x}_N') \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{1.93}$$

and the *p-reduced density matrix* (p-RDM)

$$\begin{aligned} &\Gamma^{(p)}(\mathbf{x}_1', \ldots, \mathbf{x}_p'; \mathbf{x}_1, \mathbf{x}_p) = \\ &\binom{N}{p} \int \Gamma(\mathbf{x}_1', \mathbf{x}_2', \ldots, \mathbf{x}_p', \mathbf{x}_{p+1}, \ldots, \mathbf{x}_N; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_N) d\mathbf{x}_{p+1} \ldots d\mathbf{x}_N \end{aligned} \tag{1.94}$$

For the 1-RDM, we will use the notation $\gamma(\mathbf{x}', \mathbf{x})$. The p-RDMs can be expanded in a basis. For the 1-RDM, we have that

$$\gamma(\mathbf{x}', \mathbf{x}) = \sum_{p,q}^{M} {}^1D_{pq} \phi_p^*(\mathbf{x}') \phi_q(\mathbf{x}) \tag{1.95}$$

where the matrix ${}^1\mathbf{D}$ also is called the 1-RDM. In second quantization, the 1-RDM can be expressed as

$${}^1D_{pq} = \langle \Psi | \, a_q^\dagger a_p \, | \Psi \rangle \tag{1.96}$$

which can be shown by considering the action of an arbitrary one-electron operator. From the second-quantized representation, it is easy to see that the 1-RDM is Hermitian. Furthermore, it has diagonal elements between 0 and 1, and its trace is given by the number of electrons, $\text{tr}\left({}^1\mathbf{D}\right) = N$. Being a Hermitian matrix, it can be diagonalized,

$${}^1\mathbf{D} = \mathbf{U}\mathbf{d}\mathbf{U}^\dagger. \tag{1.97}$$

The set of MOs $\{|\xi_p\rangle\}_{p=1}^{M}$ that diagonalize the 1-RDM is called *natural orbitals* (NOs), first introduced by Löwdin in 1955 [22]:00

$$|\xi_p\rangle = \sum_{q=1}^{M} |\phi\rangle_q U_{qp}. \tag{1.98}$$

The corresponding eigenvalues d_{pp} all lie between zero and one and sum to N by the previous analysis and the invariance of the trace to unitary operations, or equally, the fact that the wave function is an eigenfunction of the number operator

$$\hat{N} = \sum_{p=1}^{\infty} a_p^\dagger a_p. \tag{1.99}$$

The numbers d_{pp} are referred to as the *natural occupation numbers*. By expressing $|\Psi\rangle$ in terms of its FCI-expansion, we see that those NOs with large occupation numbers correspond to "important" orbitals, in the sense that the expansion coefficients of the Slater determinants that contain those NOs, are comparatively large in absolute value, while those with small occupation numbers contribute little. This leads to a truncation scheme: NOs with occupation numbers below some threshold can be removed from a calculation without a large loss of accuracy. This has been shown for several post-Hartree-Fock methods including CISD [23] and CCSD [24]. Given the scaling $O(M_v^4 M_o^2)$ for CCSD, reducing the number of virtual orbitals gives very large savings. In both cases, a variant of the NOs was considered, where the occupied orbitals of the HF wave function are not mixed with the unoccupied orbitals, such that the reference wave function is unchanged. For larger basis sets, they have shown that a lager percentage of NOs can be discarded. Calculations where natural orbitals with low occupation numbers were removed by some threshold, are referred to as "frozen natural orbital" methods.Even though the NOs are formally defined from the FCI wave function, using approximate NOs obtained from a first order perturbation theory wave function are very close to the exact NOs in most situations and can hence be used just as well [25], and we will simply call them NOs.

1.9.1 Analyticity of the NOs

We have previously discussed the analyticity of the Hamilton operator $\hat{H}(\vec{R})$ in a basis under geometric perturbations of the nuclei. As the Hamiltonian is Hermitian, its ground state can be expressed as an analytic function as long as it is non-degenerate and does not cross with an excited state (cf sec. 2.1). If the 1-electron basis is analytic, the 1-RDM ${}^1\mathbf{D}$ is analytic as well. The natural orbitals can hence be chosen as analytic functions $|\xi_p(\vec{R})\rangle$ with associated eigenvalues $d_{pp}(\vec{R})$. This does not necessarily hold true when the occupied and the unoccupied blocks of the 1-RDM are diagonalized individually when expressed in terms of canonical orbitals, or when approximate NOs are used. Even the exact NOs may cross. When insisting on the order $d_{11}(\vec{R}) \geq \cdots \geq d_{MM}(\vec{R})$ for all $\vec{R}$, the functions $|\xi_p(\vec{R})\rangle$ will not necessarily be analytic, and neither will the eigenvalues. Thus, in order to obtain analytic functions, this ordering needs to be given up. This is illustrated in figure 1.1. However, as standard linear algebra programs automatically sort eigenvectors by their lowest eigenvalue, some algorithm needs to be employed to ascertain the analytical ordering. This is not trivial and discussed in the numerical methods part of the thesis, sec. 5.3, p. 62.

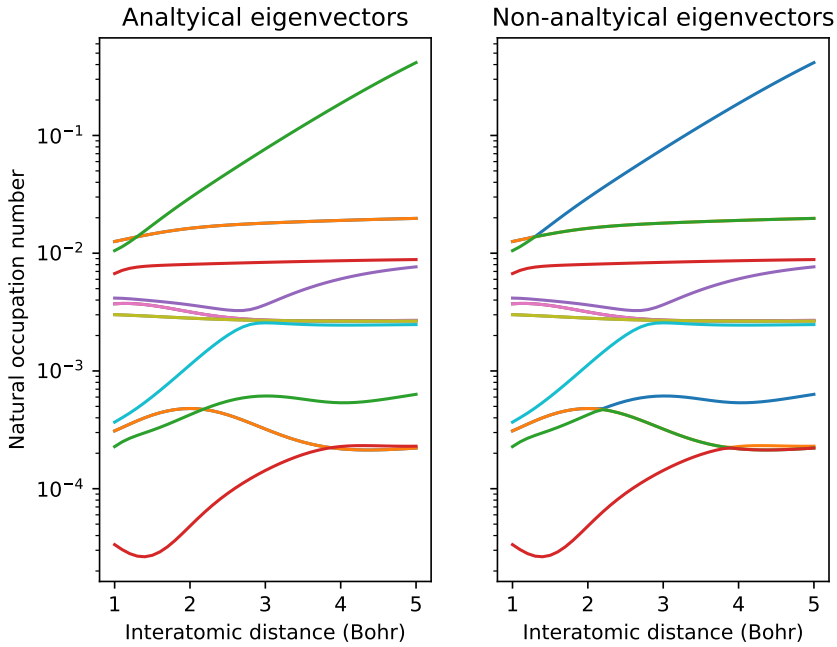


Figure 1.1: Trajectory of natural occupation numbers of the virtual block of the HF molecule in the cc-pVDZ basis set under dissociation of the molecule. Left: Analytic functions for the eigenvalues and hence the eigenvectors. Right: Ordering of the natural occupation numbers by value. We see that the ordered eigenvalues are non-analytic at crossing points.

1.10 Non-orthogonal multi-reference approaches

All methods considered so far are single-reference methods that aim at improving a the wave function represented by a reference Slater determinant (usually a HF-state) through successive addition of excited determinants, either in a linear (CI) or in an exponential fashion (CC). They are *single reference* methods. Multireference methods put several determinants on an equal footing. In the *multi-configurational self-consistent field* (MCCSF) method, the wave function is expressed as sum of Slater determinants $|\Psi_{MCSCF}\rangle = \sum_i c_i |\Phi_i^{\mathrm{SD}}\rangle$, optimizing not just the expansion coefficients c_i, but also the set of MOs that enter into the Slater determinants $|\Psi_i\rangle$ [5]. In the *CASPT*-approach, perturbation theory is applied to a multi-reference wave function [8].
Further generalizations are possible when allowing two Slater determinants to be non-orthogonal, which effectively means that they are built from different sets of MOs. In the *NOCI* approach [26], the wave function is expressed as a linear combination of non-orthogonal Slater determinants. In the NOCISD-approach [27], the wave function is built from all singles and doubles that can be reached from each non-orthogonal Slater determinant included. In the *NO-PT2* approach [26], perturbation theory is applied to a NOCI wave function. Those methods are an active field of research and, as of today, limited by the computational cost of the generalized Slater-Condon rules and Wick's theorem [28].

Chapter 2

Eigenvector continuation

This chapter introduces the eigenvector continuation algorithm, its implementation and its mathematical justification, as well as a comparison to perturbation theory. Furthermore, it introduces how eigenvector continuation has been used to solve quantum mechanical problems in previous studies and how a variant of it has been applied to the coupled cluster formalism. We will use bra-ket notation for vectors $\mathbf{v} \to |\psi\rangle$ and their conjugate transpose $\mathbf{v}^\dagger \to \langle\psi|$. The mathematical behaviour and justification of eigenvector continuation requires mathematical theory which would require several pages of introduction, which we will not do, but we will present a summary of the most important results from finite dimensional perturbation theory.

2.1 Main results from perturbation theory

We have already introduced the term *perturbation theory* in the previous chapter (cf. sec. 1.8). However, perturbation theory extends to operators that depend analytically on some parameter α, asking the question how the eigenvectors $|\psi(\alpha)\rangle$ and eigenvalues $\lambda(\alpha)$. We will here consider finite-dimensional matrices $\mathbf{M}(\alpha)$, following tightly chapter two of the work by Kato [29]. We will not prove any results, but simply state the most important results for the following discussions.

1. Eigenvalues $\lambda(\alpha)$ are continuous everywhere, and analytic everywhere except at an isolated set of *exceptional points*. These exceptional points lie where the trajectory of at least two eigenvalues coincide [29, Ch. 2, §1].

2. The eigenprojections $\mathbf{P}(\alpha)$, e.g. the projection operators that project an eigenvector of $\mathbf{M}(\alpha)$ on itself, are analytical everywhere except at the exceptional points for that eigenvalue (even though they can be analytical there). The eigenvectors can thus be chosen to be analytical everywhere except at those exceptional points [29, Ch. 2, §1].

3. The convergence radius of perturbation series for the eigenvalues and eigenprojections (and hence eigenvalues) are precisely given by the distance from the center of expansion to the first exceptional point which is exceptional for a given eigenvalue/eigenprojection pair (that means that from a given center of expansion, different eigenvalues have different radii of convergence). This can be interpreted as that the trajectory becomes "unpredictable" afterwards [29, Ch. 2, §3].

4. The eigenvalues and eigenvectors of Hermitian operators $\mathbf{H}(\alpha) = \mathbf{H}^\dagger(\alpha)$ can chosen to be analytic on the real axis (even when eigenvalues collide at special points) [29, Ch.2, Theorem 6.1]. Their perturbation series have however not an infinite radius of convergence as there might be exceptional points in the complex plane nearby. When two eigenvalues $\lambda_1(\alpha), \lambda_2(\alpha)$ come very close on the real line (avoided crossings), we can generally suspect small radii of convergence.

5. In general, when we consider matrices of more than one parameter $\mathbf{M}(\vec{\alpha})$, those results still hold [29, Ch. 2, Theorem 5.16]. There are however cases where the eigenvalues of symmetric matrices are not analytic, but only continuous.

2.2 Eigenvector continuation: formulation and theory

Eigenvector continuation (EVC) [1] is a novel, variational method to estimate the eigenvector-eigenvalue pair $E(\vec{\alpha}), |\psi(\vec{\alpha})\rangle$ of a Hermitian matrix $\mathbf{H}(\vec{\alpha})$ analytic in $\vec{\alpha}$. While Ref. [1] only considered a linear dependency of the form $\mathbf{H}(\alpha) = \mathbf{H}_0 + \alpha\mathbf{H}_1$, the method can easily be extended to multivariate parameters $\vec{\alpha}$ and non-linear perturbations.
When interested in the ground state, instead of explicitly diagonalizing $\mathbf{H}(\vec{\alpha})$ at each value $\vec{\alpha}_\odot$ of interest, the eigenvector continuation method approximates the eigenvector $|\psi(\vec{\alpha}_\odot)\rangle$ as a linear combination of the form

$$|\psi(\vec{\alpha}_\odot)\rangle = \sum_{i=1}^{L} c_i \, |\psi(\vec{\alpha_i})\rangle \tag{2.1}$$

where $|\psi(\vec{\alpha_i})\rangle$ are (approximate) ground state for a pre-defined set of sample points $\vec{\alpha}_i \in \{\vec{\alpha}_0, \vec{\alpha}_1, \ldots, \vec{\alpha}_L\}$. The coefficients c_i are chosen such that the eigenvalue is minimized. This is accomplished by projecting the true matrix $\mathbf{H}(\vec{\alpha}_\odot)$ on the space spanned by the eigenvectors $|\psi(\vec{\alpha}_i)\rangle$, such that

$$\begin{aligned} H^{\text{EVC}}_{i,j} &= \langle\psi(\vec{\alpha}_i)| \, \mathbf{H}(\vec{\alpha}_\odot) \, |\psi(\vec{\alpha}_j)\rangle \\ S^{\text{EVC}}_{i,j} &= \langle\psi(\vec{\alpha}_i)|\psi(\vec{\alpha}_j)\rangle \end{aligned} \tag{2.2}$$

and solving the generalized eigenvector problem $\mathbf{H}^{\text{EVC}}\mathbf{C} = \mathbf{S}^{\text{EVC}}\mathbf{C}\boldsymbol{\epsilon}$. An intended use of eigenvector continuation is to use as sample points a set of $\vec{\alpha}_i \in I$, where I is a region where $|\psi(\vec{\alpha}_i)\rangle$ is "easy" to obtain and a good approximation to the exact ground state, while the points $\vec{\alpha}_\odot$ lie in a region where these approximation methods fail. However, when EVC works well in a given region, it can also be considered as an effective interpolation method, and it might possibly be an effective dimensionality reduction tool.

2.2.1 Mathematical justification

We will here consider the one-dimensional case where $\mathbf{H}(\alpha)$ only depends on a single parameter α, as was done in Ref. [1]. As $\mathbf{H}(\alpha)$ is Hermitian and analytic, its eigenvalues are real, and unless there is degeneracy in the ground state energy for some α, the ground state and ground state energy can be chosen to be analytic in α. Around the point $\alpha = 0$, we can expand the ground state eigenvector and eigenvalue as a Taylor series

$$\begin{aligned} |\psi(\alpha)\rangle &= \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} \, |\psi^{(n)}(0)\rangle \\ E(\alpha) &= \sum_{n=0}^{\infty} \frac{\alpha^n}{n!} E^{(n)}(0) \end{aligned} \tag{2.3}$$

with non-zero radius of convergence. Let z be the closest singularity to $\alpha = 0$, which will be an exceptional point. Then the Taylor series 2.3 will converge for $|\alpha| < |z|$. Let now w be a real number such that $|w| < |z|$. We can choose w such that $|w - |z|| < |w - z|$. We can now perform an analytic continuation and expand the eigenvector-eigenvalue pair around $\alpha = w$:

$$\begin{aligned} |\psi(\alpha)\rangle &= \sum_{n=0}^{\infty} \frac{(\alpha - w)^n}{n!} \, |\psi^{(n)}(w)\rangle \\ E(\alpha) &= \sum_{n=0}^{\infty} \frac{(\alpha - w)^n}{n!} E^{(n)}(w) \end{aligned} \tag{2.4}$$

This series will converge for $|\alpha - w| < |z - w|$, given that z is still the closest singularity from the point w. We can express the vectors $|\psi^{(n)}(w)\rangle$ in terms of $|\psi^{(n)}(0)\rangle$. Thus, one can approximate $|\psi(\alpha)\rangle$ as linear combination of the eigenvectors $|\psi^{(n)}(0)\rangle$ also in the region $|\alpha - w| < |z - w|$. As this procedure can be repeated as often as desired, we can express any $|\psi(\alpha)\rangle$ as linear combination of the vectors $|\psi^{(n)}(0)\rangle$. For any given accuracy ϵ the number of necessary vectors n will be finite, and depend on the number of expansions needed, as well as the order-by-order convergence of each individual series expansion. In figure 2.1, we show an example of how well EVC performs on a system with a relatively complicated shape of the ground-state eigenvalues and several avoided crossings when sampling far away from the first avoided crossing.

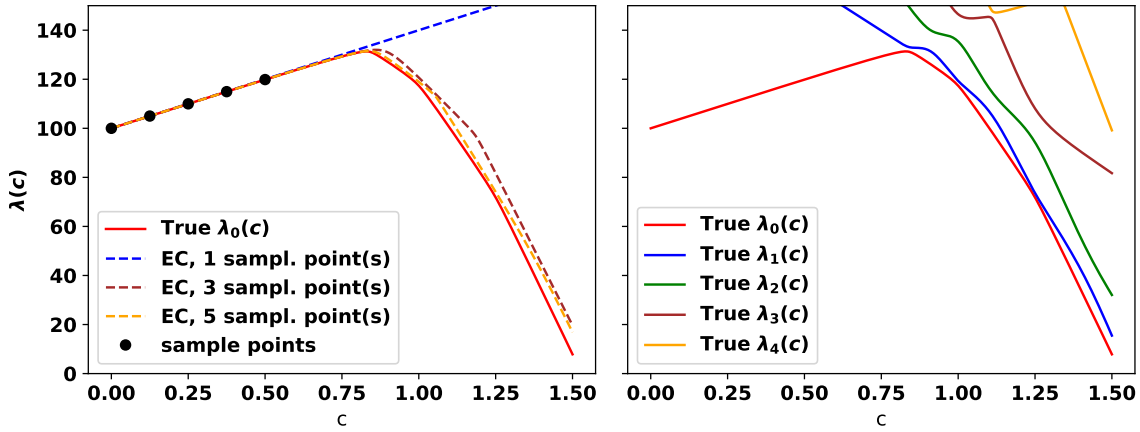


Figure 2.1: Left: eigenvector continuation of the lowest lying eigenvalue using 1, 3 and 5 sample points (the leftmost points were used). Right: The five lowest-lying eigenvalues. The matrix function giving raise to these eigenvalues is a 500×500 *matrix of the type* $\mathbf{M}(c) = \mathbf{M}_0 + c\mathbf{M}_1$ *and described in sec. 9.1 in the appendix. We see how using 5 sample points along the "straight line" gives a qualitatively correct eigenvalue even after passing several avoided crossings.*

Behaviour close to exceptional points

This explanation largely follows Ref. [1], but uses the behaviour of the eigenprojections lined out by Kato [29, ch. 2] rather than the eigenvectors themselves. Let z be an exceptional point. Let $E(\alpha), |\psi(\alpha)\rangle$ be a general eigenvector-eigenvalue pair, not necessarily the ground state. Let D be a portion of the Riemann surface associated with $|\psi(\alpha)\rangle$. When performing a counterclock rotation in the complex plane around $\alpha = z$ (a monodromy transform $M(z)$) and $E(z)$ is a unique eigenvalue of $\mathbf{H}(z)$, it is the only element in its λ-group, hence $E(z)$ and the eigenprojection $\mathbf{P}(z)$ are analytic at $\alpha = z$ and there occurs no splitting. $\mathbf{P}(z)$ being analytic guarantees that the monodromy-transformed eigenvector $|\psi'(\alpha)\rangle$ is proportional to $|\psi(\alpha)\rangle$, hence no issues in the convergence of eigenvector continuation arise. Let now $E(\alpha)$ coincide with $k-1$ other eigenvalues $\tilde{E}_1(\alpha), \ldots, \tilde{E}_{k-1}(\alpha)$ and assume that the λ-group only consists of one cycle. The monodromy transform will lead to a cyclic permutation of the eigenvalues and eigenvectors. The total projection for the λ-group is analytic at $\alpha = z$ and has dimension k, and is hence invariant under the monodromy transform. Thus, the subspace corresponding to the total projection is left invariant. But then one should include, at each sample point α_i, k vectors such that the subspace corresponding to the total projection is spanned by them, making the inclusion of excited state vectors necessary to assure the best convergence beyond z.

2.2.2 Convergence compared to Perturbation theory

Based on the previous discussion, choosing sample points α_i infinitesimally close to $\alpha = 0$, the sample vectors $|\psi(\alpha_i)\rangle$ will span the same space as the derivative vectors $|\psi^{(i)}(0)\rangle$. When these sample points are chosen, it can be proven that eigenvector continuation

converges faster than perturbation theory [2]. This is because the derivative vectors in perturbation theory are not mutually orthogonal, leading to constructive and destructive interference implying a slower convergence. This is called *Differential Folding* and does not occur for eigenvector continuation - although eigenvector continuation uses the same basis vectors, the expansion coefficients are optimized with respect to energy, avoiding this problem. This also explains why eigenvector continuation will not necessarily diverge outside of the radius of convergence, and adding more sample points will not increase the eigenvalue. Demol et al. [30] have considered a partitioning of a many-body Hamiltonian as $\mathbf{H}(\alpha) = \mathbf{H}^0 + \alpha\mathbf{H}^1$ where $\mathbf{H} = \mathbf{H}(1)$ is the exact Hamiltonian and exact eigenvectors of $\mathbf{H}^0$ are readily available, and constructed the corrections to the wave function up to p'th order $|\psi^{(p)}\rangle$, constructing the EVC-Hamiltonian as $H^{\text{EVC}}_{i,j} = \langle\Psi^{(i)}|\,\mathbf{H}\,|\Psi^{(j)}\rangle$ for $i,j = 0,\ldots,p$ and the overlap matrix correspondingly. While perturbation theory results diverged, they got rapid and monotonically converging series using the EVC-approach, corresponding to a resummation of the individual terms.

2.2.3 Useful properties of eigenvector continuation

When used to find the ground state of a system, the eigenvector continuation approach has the useful property that inclusion of sample eigenvectors that are bad approximations to the real eigenvectors, does not worsen the resulting EVC solution. This follows directly from the Hylleraas-Undheim theorem [8] and is summarized in the following lemma.

Lemma 2.2.1. *Let x_0 be one of the sampling points of the EVC algorithm and $\mathbf{H}(x)$ a Hermitian matrix. If $|\psi_0(x_0)\rangle$ is the exact ground state with the corresponding eigenvalue $E(x_0)$, then the EVC algorithm is exact at $x_\odot = x_0$. If $|\psi_0(x_0)\rangle$ is an approximate normalized eigenvector with expectation value $E_0 = \langle\psi_0(x_0)|\,\mathbf{H}(x_0)\,|\psi_0(x_0)\rangle$, then the EVC algorithm will return an approximate eigenvector with eigenvalue equal to or lower than E_0 at $x_\odot = x_0$.*

Proof. By construction, the matrix with elements $H^{\text{EVC}}_{i,j} = \langle\psi_0(x_i)|\,\mathbf{H}(x_\odot)\,|\psi_0(x_j)\rangle$ is Hermitian if $\mathbf{H}(x_\odot)$ is Hermitian. Expressed in the basis consisting of only $|\psi_0(x_0)\rangle$, we have that $\mathbf{H}(x_0) \to \langle\psi_0(x_0)|\,\mathbf{H}(x_\odot)\,|\psi_0(x_0)\rangle = E_0$. When expanding the basis by adding vectors $|\psi_0(x_1)\rangle, \ldots, |\psi_0(x_n)\rangle$, constructing the matrix $\mathbf{H}^{\text{EVC}}(x_\odot)$, E_0 will be a (trivially Hermitian) 1×1 submatrix with eigenvalue E_0. By the Hylleraas-Undheim theorem, E_0 will then be an upper bound for the lowest eigenvalue of $\mathbf{H}^{\text{EVC}}(x_\odot)$, proving the second statement. If E_0 is the exact lowest eigenvalue of $\mathbf{H}(x_\odot)$, then the lowest eigenvalue of $\mathbf{H}^{\text{EVC}}(x_\odot)$ cannot be lower than E_0 by the variational principle and not be higher than E_0 by the second statement, thus proving the first statement. □

2.3 Coupled cluster eigenvector continuation

When applying eigenvector continuation to coupled cluster wave functions, we will consider two different approaches. The first approach is to write the correct approximate wave function as a linear combination of sample coupled cluster wave functions, called *wave function approach* (WF-CCEVC). In the methods section (sec. 4.3), we will introduce the second approach, which is to write the cluster operator as a linear combination of the sample cluster operators, called *amplitude approach* (AMP-CCEVC).

2.3.1 Wave Function approach

A direct application of the EVC approach (eq. 2.2) to a coupled cluster wave function $|\Psi(\vec{\alpha})\rangle = e^{\hat{T}(\vec{\alpha})}\,|\Phi^{\text{SD}}\rangle$ would require the evaluation of the following matrix elements and overlaps

$$\langle\Psi(\vec{\alpha}_i)|\Psi(\vec{\alpha}_j)\rangle = \langle\Phi^{\text{SD}}|\,e^{\hat{T}^\dagger(\vec{\alpha}_i)}e^{\hat{T}(\vec{\alpha}_j)}\,|\Phi^{\text{SD}}\rangle \tag{2.5}$$

$$\langle\Psi(\vec{\alpha}_i)|\hat{H}(\vec{\alpha}_\odot)|\Psi(\vec{\alpha}_j)\rangle = \langle\Phi^{\text{SD}}|\,e^{\hat{T}^\dagger(\vec{\alpha}_i)}\hat{H}(\vec{\alpha}_\odot)e^{\hat{T}(\vec{\alpha}_j)}\,|\Phi^{\text{SD}}\rangle\,. \tag{2.6}$$

As evaluating these functions is as expensive as evaluating the corresponding equations for Full CI states, Ekström and Hagen [3] have considered how eigenvector continuation can be applied to coupled cluster wave functions more effectively. They decided to apply a bi-orthogonal approach, where the left and the right eigenvectors of the Hamiltonian differ, with

$$\langle \tilde{\Psi}(\vec{\alpha})| = \langle \Phi^{\mathrm{SD}}| \left[\mathbb{1} + \hat{\Lambda}(\vec{\alpha})\right] e^{-\hat{T}(\alpha)} \tag{2.7}$$

$$|\Psi(\vec{\alpha})\rangle = e^{\hat{T}(\vec{\alpha})} |\Phi^{\mathrm{SD}}\rangle . \tag{2.8}$$

They called this approach for subspace-projected CC, we will call it *WF-CCEVC*. $\hat{\Lambda}(\vec{\alpha})$ can, for example, be found by solving the left eigenvalue problem of the full EOM-CC Hamiltonian (cf. sec. 1.7.5, p. 18) or the coupled cluster Lagrangian [8, ch. 13]. Unlike EOM-CC, however, they do not consider a similarity-transformed Hamiltonian directly, but express the EVC-Hamiltonian at the target reference as

$$\begin{aligned} \langle \tilde{\Psi}'|\Psi\rangle &= \langle \Phi^{\mathrm{SD}}| \left[\mathbb{1} + \hat{\Lambda}'\right] e^{-\hat{T}'} e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle \\ \langle \tilde{\Psi}'|\hat{H}(\vec{\alpha}_\odot)|\Psi\rangle &= \langle \Phi^{\mathrm{SD}}| e^{-\hat{T}'} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{2.9}$$

where we made the dependence of the parameters $\vec{\alpha}_i$ implicit. Letting $\hat{X} = \hat{T} - \hat{T}'$, this can be rewritten as

$$\begin{aligned} \langle \tilde{\Psi}'|\Psi\rangle &= \langle \Phi^{\mathrm{SD}}| \left[\mathbb{1} + \hat{\Lambda}'\right] e^{\hat{X}} |\Phi^{\mathrm{SD}}\rangle \\ \langle \tilde{\Psi}'|\hat{H}(\vec{\alpha}_\odot)|\Psi\rangle &= \langle \Phi^{\mathrm{SD}}| e^{-\hat{T}'+\hat{T}} e^{-\hat{T}} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = \langle \Phi^{\mathrm{SD}}| e^{\hat{X}} \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{2.10}$$

where we introduced the similarity-transformed Hamiltonian

$$\bar{H}(\vec{\alpha}_\odot) = e^{-\hat{T}} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}} \tag{2.11}$$

Observe that $\hat{X}$ is also an excitation operator. In the CCSD formulation, we have $\hat{T} = \hat{T}_1 + \hat{T}_2$ and $\hat{\Lambda} = \hat{\Lambda}_1 + \hat{\Lambda}_2$, where

$$\begin{aligned} \hat{T}_1 &= \sum_{ia} t_i^a a_a^\dagger a_i \\ \hat{T}_2 &= \frac{1}{4} \sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i \\ \hat{\Lambda}_1 &= \sum_{ia} \lambda_i^a a_i^\dagger a_a \\ \hat{\Lambda}_2 &= \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab} a_i^\dagger a_j^\dagger a_b a_a \end{aligned} \tag{2.12}$$

Using Wick's theorem, we can get an explicit expression for eq. (2.10), yielding

$$\begin{aligned} \langle \tilde{\Psi}'|\Psi\rangle =& 1 + \sum_{ia} \lambda_i^{a\prime} x_i^a + \frac{1}{2} \sum_{ijab} \lambda_{ij}^{ab\prime} x_i^a x_j^b + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab\prime} x_{ij}^{ab} \\ \langle \tilde{\Psi}'|\hat{H}(\vec{\alpha}_\odot)|\Psi\rangle =& \langle \tilde{\Psi}'|\Psi\rangle \langle \Phi^{\mathrm{SD}}| \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle + \sum_{ia} \lambda_i^{a\prime} \langle \Phi_i^a| \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle \\ &+ \sum_{ijab} \lambda_{ij}^{ab\prime} x_i^a \langle \Phi_j^b| \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab\prime} \langle \Phi_{ij}^{ab}| \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{2.13}$$

where we recognize that $\langle \Phi_\tau| \bar{H}(\vec{\alpha}_\odot) |\Phi^{\mathrm{SD}}\rangle$ correspond to the CCSD amplitude equations. The total scaling of WF-CCEVC is hence $O(M_o^2 M_v^4 L)$, where L again is the number of sampling points, given that the similarity transformed Hamiltonian needs to be calculated for each $\vec{\alpha}_\odot$. However, computational time is saved because the method is noniterative. An explicit derivation for the overlap is found in the appendix, sec. 10.4.1, p. 94. It should be mentioned that the Hamiltonian matrix is non-Hermitian and that the eigenvalues will be

generally complex, as in EOM-CC. However, the imaginary part is usually very small and will be discarded. The resulting wave function will be of the form

$$|\Psi^{\text{WF-EVC}}(\vec{\alpha})\rangle = \sum_{m=1}^{L} c_m e^{\hat{T}(\vec{\alpha}_m)} |\Phi^{\text{SD}}\rangle = \sum_{m=1}^{L} c_m \left(\mathbb{1} + \hat{T}_1(\vec{\alpha}_m) + \frac{1}{2}\hat{T}_1(\vec{\alpha}_m)^2 + \hat{T}_2(\vec{\alpha}_m) + \ldots\right)|\Phi^{\text{SD}}\rangle \tag{2.14}$$

and can usually not be expressed as a coupled cluster wave function. This is, in principle, a more flexible representation of the wave function, and one can hope that this may serve to overcome shortcomings in CC theory. From that point of view, this ansatz might not only be usable in terms of eigenvector continuation, but for linear combinations of CC wave functions in general.

It should be noted that a big drawback of WF-CCEVC is that the results of lemma 2.2.1 do not hold because WF-CCEVC is not a variational method, as the EVC-Hamiltonian is non-Hermitian. Thus, inclusion of "bad" sample wave functions and energies can yield wrong results.

Furthermore, the generalized eigenvalue problem $\mathbf{HC} = \mathbf{SC}\boldsymbol{\epsilon}$ that arises is not symmetric. To solve the generalized eigenvalue problem with symmetric matrices, one can use canonical orthogonalization [5, 10, 31], where the vanishing eigenvalues and corresponding eigenvalues of the overlap matrix are discarded. In the non-symmetric case, this problem gets harder [32]. We will delay suitable algorithms to section 5.1.3.

Chapter 3

Quantum computing

In this chapter, we will introduce the basics of error-free quantum computing and introduce relevant algorithms to calculate the ground state energy of a molecule, as well as algorithms relevant for eigenvector continuation on a quantum computer. We will also describe state of the art methods and algorithms and their scaling in order to justify the huge potential of quantum computing in quantum chemistry.

3.1 Why quantum computing?

The idea to use quantum mechanical systems to simulate other quantum mechanical systems was first put forward by Feynman, who argued that quantum computers can evolve a given input state into a desired state, demanding only polynomial resources and evolution time [33]. While the computational requirements for representing an arbitrary wave function increases exponentially with the number of particles on regular computers, the resource requirements on a quantum computer are linear [34]. State of the art quantum computers cannot yet outcompete regular computers for chemical applications, as they have too few qubits and are too error-prone. Given the low-polynomial and possibly linear time scaling of recent, very accurate quantum chemistry methods on quantum computers, which are discussed in section 3.9, the field has a vast potential. As we are entering the area of NISQ (*noisy intermediate scale quantum*) devices, there is hope that advances in algorithm and hardware design will lead to quantum supremacy for chemical applications in the near future.

3.2 Mathematical preliminaries

3.2.1 Qubits

A qubit is a two-level system, with a state described by a superposition of two basis states $|0\rangle, |1\rangle$:

$$|\psi\rangle = \alpha |0\rangle + \beta |1\rangle \hat{=} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \quad \alpha, \beta \in \mathbb{C} \tag{3.1}$$

where α, β are chosen such that $|\psi\rangle$ is normalized. In general, a set of Q qubits spans a 2^Q dimensional Hilbert space of distinguishable qubits, denoted by $H^{\otimes Q}$. As basis vectors, the *computational basis* $|s\rangle \in \{0,1\}^Q$ is usually chosen, which can be written in bitstring form

$$|k_1\rangle \otimes |k_2\rangle \otimes \cdots \otimes |k_{Q-1}\rangle \otimes |k_Q\rangle = |k_1 k_2 \ldots k_{Q-1} k_Q\rangle, \quad k_i \in \{0,1\} \tag{3.2}$$

where the subscripts are usually not written out. We will sometimes use a decimal instead of a binary representation when convenient, e.g. $|5_{(2)}\rangle \hat{=} |101\rangle$, and we will write $|0^m\rangle = |0_1 \ldots 0_m\rangle$. It should be noted that elements in this Hilbert space in general cannot be written as tensor products of states: $|01\rangle = |0\rangle \otimes |1\rangle$, but such a decomposition is not possible for

the state $|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ independently of the choice of basis - qubits 1 and 2 are *entangled*.

3.2.2 Unitary operations and measurements

A postulate of quantum mechanics is that time evolution of quantum systems is described by unitary operations. In order to transform one state into another, unitary operations $\hat{U}$ can be applied to transform a state $|\Psi_1\rangle \in H^{\otimes Q}$ into a different state $|\Psi_2\rangle \in H^{\otimes Q}$:

$$|\Psi_2\rangle = \hat{U}\,|\Psi_1\rangle \tag{3.3}$$

where $\hat{U}$ being unitary follows from the fact that unitary matrices are norm-preserving [35] and quantum mechanics being reversible[1]. Hence, quantum computing is concerned about applying a set of unitary operations to qubits in order to obtain a specific state, and then measure (some of) those qubits in order to obtain information of interest. Measurements lead to a (partial) collapse of the wave function and are mathematically represented by projection operators, hence they are not unitary.

3.2.3 Quantum circuits

In order to portray the actions performed on a quantum computer in a conceptually easy way, quantum circuits are used, which can express any arbitrary quantum computation. Consider the following example in figure 3.1, transforming the state $|000\rangle$ into a state $|\Phi\rangle = (U_1^\dagger \otimes \mathbb{1} \otimes U_3)(U_2)(U_1 \otimes \mathbb{1} \otimes \mathbb{1})\,|000\rangle$:

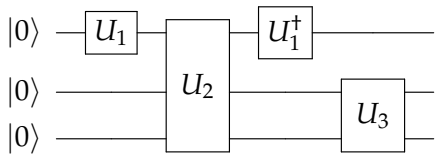
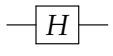
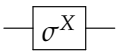
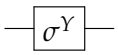
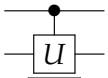
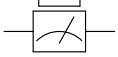


Figure 3.1: Quantum circuit transforming the state $|000\rangle$ into the state $|\Phi\rangle = (\mathbb{1} \otimes \mathbb{1} \otimes U_3)(U_1^\dagger \otimes \mathbb{1} \otimes \mathbb{1})(U_2)(U_1 \otimes \mathbb{1} \otimes \mathbb{1})\,|000\rangle$

Quantum circuits are written from the left to the right, with gates to the left applied before gates to the right. Sometimes, a single wire may represent more than one qubit, which is clear from the context. Table 3.1 contains the most important gates and their actions in computational basis.

[1] alternatively, one might argue that it is physically impossible to apply non-unitary operations to a system of qubits, as every application is necessarily an evolution in time and hence unitary.

Table 3.1: Non-exhaustive list of the most important quantum gates and their matrix representation / action. For one-qubit gates, the matrix representation is with respect to the basis $\{|0\rangle, |1\rangle\}$, for two-qubit gates, it is with respect to the basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$

Gate name	Symbol	Action
Hadamard		$\frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$
Pauli-X		$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}$
Pauli-Y		$\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}$
Pauli-Z		$\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$
CNOT		$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$
SWAP		$\begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$
Controlled-U		$\|0\rangle \otimes \|\Psi\rangle \rightarrow \|0\rangle \otimes \|\Psi\rangle$, $\|1\rangle \otimes \|\Psi\rangle \rightarrow \|1\rangle \otimes \hat{U}\|\Psi\rangle$
Measurement		Projection on $\|0\rangle$, $\|1\rangle$

3.2.4 Pauli matrices

The three Pauli gates and their matrix representation were introduced in table 3.1. They play an important role in quantum chemistry on quantum computers, as the Hamiltonian can be decomposed in terms of them (cf. section 3.6). It is hence necessary to describe some of their properties as well as introduce notation. Let σ^i be any Pauli matrix, $i \in \{X, Y, Z\}$. Furthermore, let $\tilde{\sigma}^i$ be any Pauli matrix or the identity operator $i \in \{0, X, Y, Z\}$, where we use the notation $\sigma^0 = \mathbb{1}$.

- The Pauli matrices have the following commutation rules:

$$[\sigma^X, \sigma^Y] = 2i\sigma^Z$$
$$[\sigma^Y, \sigma^Z] = 2i\sigma^X$$
$$[\sigma^Z, \sigma^X] = 2i\sigma^Y$$

- The Pauli matrices are unitary and involutory:

$$\left(\tilde{\sigma}^i\right)^2 = \left(\tilde{\sigma}^i\right)^\dagger \left(\tilde{\sigma}^i\right) = \mathbb{1}$$

  And this also applies to any string $\hat{P}_m$ of Pauli matrices

$$\hat{P}_m^2 = \left(\tilde{\sigma}_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q}\right)^2 = \mathbb{1}^{\otimes Q}$$

  where the tensor product symbol was omitted, and $\mathbb{1}^{\otimes Q} = \mathbb{1}_1 \otimes \cdots \otimes \mathbb{1}_Q$.

- All Pauli matrices and strings of Pauli matrices have an eigenvalue of ± 1, which is a direct consequence of them being an involution.

- The set $\{\sigma^0, \sigma^X, \sigma^Y, \sigma^Z\}^Q$ is a basis for Hermitian matrices of dimension $2^Q \times 2^Q$ with only real expansion coefficients, and more generally, a basis for all matrices of dimension $2^Q \times 2^Q$ when the expansion coefficients may be complex.

- Measurements of any Pauli string $\sigma_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q}$ on an arbitrary state $|\Phi\rangle$, e.g. measuring $\langle\Phi|\,\tilde{\sigma}_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q}\,|\Phi\rangle$, can be transformed into measuring σ_1^Z by application of 1-qubit gates and CNOT gates. Measuring σ_1^Z is equivalent to measuring whether qubit 1 is in the state $|0\rangle$ or $|1\rangle$.

3.2.5 Trotter formula

Let $\mathbf{A}_1$,$\mathbf{A}_2$ be complex $m \times m$ matrices that may or may not commute. Then it can be shown [36] that the *Trotter formula* holds

$$e^{(\mathbf{A}_1+\mathbf{A}_2)} = \lim_{n\to\infty}\left(e^{\mathbf{A}_1/n}e^{\mathbf{A}_2/n}\right)^n \tag{3.4}$$

where n is the number of *Trotter steps*.

3.3 Physical devices

3.3.1 Decomposition of unitaries into gates

On physical quantum computers, it is not possible to directly implement any arbitrary unitary operator $\hat{U}$ directly. On most quantum computers, unitary operations cannot act on more than two qubits at once. Even then, it is usually restricted what type of two-qubit operations can be implemented (only CNOT-gates, for example), and it is not even possible to have those gates act on an arbitrary pair of qubits (i, j). However, one can show that any unitary operation can be decomposed to any desired accuracy using only a set of few one-qubit gates and two-qubit CNOT gates [36]. Furthermore, using SWAP gates, the problem of not being able to perform 2-qubit on arbitrary pairs, can be resolved.

3.3.2 Restrictions due to errors

Quantum algorithms do not work perfectly due to imperfections in gate implementations and interaction with the environment. For example, a single qubit might suddenly have a phase added to it: $\alpha\,|0\rangle + \beta\,|1\rangle \;\rightarrow\; \alpha\,|0\rangle + e^{i\theta}\beta\,|1\rangle$, a qubit might flip $|0\rangle \;\rightarrow\; |1\rangle$, or a state might be reset to $|0\rangle$. As noise is unpredictable, we do not know when these errors occur. This reduces our knowledge about the system. A standard method to reduce decoherence is by representing a *logical qubit* by several *physical qubits*, where noise changes the state of the physical qubits, but not the logical qubit. *Quantum error correction* [37] is an active field of research on its own and we will here not go into detail how it works, but we want to underline that it is not possible to implement arbitrary N-qubit algorithms on an N-qubit devices on today's NISQ devices, and as physically implementable error-correction is not fully fault-tolerant and also expensive, circuit length is not arbitrary either. In addition to quantum error correction, there also is *quantum error mitigation*, which does not aim at reducing the error, but use knowledge about the error and the system to modify the circuits in such a way that the correct result is achieved, or to extrapolate the correct measurement value from a set of noisy measurements [38].

3.4 Variational quantum eigensolver

While there exist algorithms that can accurately calculate the (exact) ground-state eigenvalue when a sufficiently close approximate state is prepared on the quantum computer, such as quantum phase estimation (QPE) [39, 40], NISQ devices will not be capable of performing

those calculations due to a reduced number of qubits and short coherence times. Thus, hybrid quantum computing algorithms were developed, using the full power of near-future NISQ devices, while performing part of the calculation on a classical computer. The most prominent one for quantum chemistry is the *variational quantum eigensolver* (VQE) method [4]. The algorithm works as follows:

1. Prepare a state $|\Psi(\vec{\theta}_0)\rangle$ on a quantum computer, where $|\Psi(\vec{\theta})\rangle$ is, in general, a state that can be prepared at reasonable cost for all $\vec{\theta}$, and $\vec{\theta}_0$ being some starting guess.
2. Set $\vec{\theta} = \vec{\theta}_0$.
3. Estimate the state's expectation energy $\langle\Psi(\vec{\theta})| \hat{H} |\Psi(\vec{\theta})\rangle$. This is done using *Hamiltonian averaging* on the quantum computer [40], described in more detail in sec. 3.8.
4. Update the parameter $\vec{\theta}$ *classically* to get a new state $|\Psi(\vec{\theta})\rangle$ based on a classical optimization procedure.
5. Repeat 2, 3 until some convergence criteria is fulfilled.

The effectiveness of the VQE ansatz depends on whether the ansatz $|\Psi(\vec{\theta})\rangle$ can represent the ground state well enough; and whether the classical optimization manages to find a good set of parameters. This optimization is generally NP-hard [41], which is not related to the "intrinsic hardness" of finding the ground state, but the optimization procedure itself, with local minima slowing down or halting optimization. This is however not always a problem in practice.
Gradients for a variety of ansätze can be measured or approximated. On noisy devices, numerical differentiation based on finite-difference schemes tends to give insufficient results and be costly. However, analytical partial derivatives can often be explicitly constructed and measured. This is the case for the qubit representation of the UCC-based ansätze described in the next section, where the parameter-shift rules apply [42], which allow for relatively cheap measurement of gradients in a circuit which is very similar to the one measuring the expectation value.

3.5 Ansätze

3.5.1 Unitary coupled cluster on a quantum computer

One possible parameterization of the wave function $|\Psi(\vec{\theta})\rangle$ is to use a unitary coupled cluster wave function (eq. (1.82)), e.g.

$$|\Psi(\vec{\theta})\rangle = e^{\hat{\sigma}(\vec{\theta})} |\Phi_0\rangle = e^{\hat{T}(\vec{\theta}) - \hat{T}^\dagger(\vec{\theta})} |\Phi_0\rangle \tag{3.5}$$

where $|\Phi_0\rangle$ is usually, but not necessarily, a Slater determinant. Slater determinants are very simple to prepare on quantum computers when a second quantized Hamiltonian is used. In the singles- and doubles- approximation, the UCCSD exponent thus reads

$$\begin{aligned}
e^{\hat{\sigma}(\vec{\theta})}_{\text{UCCSD}} &= \exp\left(\hat{T}_1(\vec{\theta}) - \hat{T}_1^\dagger(\vec{\theta}) + \hat{T}_2(\vec{\theta}) - \hat{T}_2^\dagger(\vec{\theta})\right) \\
&= \exp\left(\sum_{ia} \theta_i^a \left(a_a^\dagger a_i - a_i^\dagger a_a\right) + \sum_{ijab} \theta_{ij}^{ab} \left(a_a^\dagger a_b^\dagger a_i a_j - a_j^\dagger a_i^\dagger a_b a_a\right)\right) \\
&\approx \exp\left(\sum_{ia} \theta_i^a \left(a_a^\dagger a_i - a_i^\dagger a_a\right)\right) \exp\left(\sum_{ijab} \theta_{ij}^{ab} \left(a_a^\dagger a_b^\dagger a_i a_j - a_j^\dagger a_i^\dagger a_b a_a\right)\right) \\
&\approx \prod_{ia} \exp\left(\theta_i^a \left(a_a^\dagger a_i - a_i^\dagger a_a\right)\right) \prod_{ijab} \exp\left(\theta_{ij}^{ab} \left(a_a^\dagger a_b^\dagger a_i a_j - a_j^\dagger a_i^\dagger a_b a_a\right)\right) \\
&= \prod_{\mu \in S,D} \exp\left(\theta_\mu \hat{\sigma}_\mu\right)
\end{aligned} \tag{3.6}$$

where the Trotter formula (3.4) with $n = 1$ was used and the last equality is just a shorthand notation. This might seem to be a crude approximation - however, as has been analyzed in Ref. [43], increasing the number of Trotter steps n leads to small energy improvements, although the optimal parameter $\vec{\theta}$ changes as n is varied. As the individual exponentials do not commute, the ordering has a big impact on the resulting wave function [44], and sometimes going to $n \geq 2$ with independent parameters is necessary to achieve chemical accuracy. The ansatz for $n = 1$ has been called *disentangled UCC (dUCC)*, and it has been shown that it is exact when the cluster operator is not truncated and a correct ordering of the elements $\exp\left(\theta_\mu \hat{\sigma}_\mu\right)$ is chosen [45]. Furthermore, it is always possible to parameterize the exact wave function using an infinite number of singles and doubles excitations [45], which means that increasing the number of parameters systematically improves the wave function:

$$|\Psi^{\mathrm{FCI}}\rangle = \prod_{k}^{\infty} \prod_{\mu \in S,D} \exp\left(\theta_\mu^{(k)} \hat{\sigma}_\mu\right) |\Psi_0\rangle \tag{3.7}$$

Further modifications of the unitary coupled cluster method: k-UpCCGSD

A generalization of the UCCSD method is the *UCCGSD* method, where generalized excitations and de-exictations are employed:

$$\begin{aligned} e^{\hat{\sigma}(\vec{\theta})}_{\mathrm{UCCGSD}} &= \exp\left(\sum_{pq} \theta_p^q \left(a_p^\dagger a_q - a_q^\dagger a_p\right) + \sum_{pqrs} \theta_{pq}^{rs} \left(a_r^\dagger a_s^\dagger a_p a_q - a_p^\dagger a_q^\dagger a_s a_r\right)\right) \\ &\approx \prod_{pq} \exp\left(\theta_p^q \left(a_p^\dagger a_q - a_q^\dagger a_p\right)\right) \prod_{pqrs} \exp\left(\theta_{pq}^{rs} \left(a_r^\dagger a_s^\dagger a_p a_q - a_p^\dagger a_q^\dagger a_s a_r\right)\right) \end{aligned} \tag{3.8}$$

While the motivation of the ansatz is based on the Nooijen conjecture [46] of the exactness of the non-unitary CCGSD, which was later disproved [47], it serves as a starting point for further methods.
The *UpCCGSD* ansatz is a restriction on the UCCGSD ansatz, where the T_2 amplitudes are restricted to only exciting pairs of electrons, that is

$$\begin{aligned} e^{\hat{\sigma}(\vec{\theta})}_{\mathrm{UpCCGSD}} &= \exp\left(\sum_{pq} \theta_p^q \left(a_p^\dagger a_q - a_q^\dagger a_p\right) + \sum_{pq} \theta_{p_\alpha p_\beta}^{q_\alpha q_\beta} \left(a_{q_\alpha}^\dagger a_{q_\beta}^\dagger a_{p_\beta} a_{p_\alpha} - a_{p_\alpha}^\dagger a_{p_\beta}^\dagger a_{q_\beta} a_{q_\alpha}\right)\right) \\ &\approx \prod_{pq} \exp\left(\theta_p^q \left(a_p^\dagger a_q - a_q^\dagger a_p\right)\right) \prod_{pq} \exp\left(\theta_{p_\alpha p_\beta}^{q_\alpha q_\beta} \left(a_{q_\alpha}^\dagger a_{q_\beta}^\dagger a_{p_\beta} a_{p_\alpha} - a_{p_\alpha}^\dagger a_{p_\beta}^\dagger a_{q_\beta} a_{q_\alpha}\right)\right) \end{aligned} \tag{3.9}$$

The clear advantage is the scaling of the number of variables as $O(N^2)$, one shortcoming is that it is not invariant to unitary rotations applied to the occupied/unoccupied MOs. To partially overcome this issue and allow for a better parameterization, we finally define the *k-UpCCGSD* ansatz [48], where the excitations are described by k repetitions of the UpCCGSD-ansatz:

$$|\Psi(\vec{\theta})\rangle = \prod_{l=1}^{k} e^{\hat{\sigma}(\vec{\theta}^{(l)})}_{\mathrm{UpCCGSD}} |\Phi_0\rangle \tag{3.10}$$

where $\vec{\theta}^{(l)}$ is different for each l, thus there are $O(kN^2)$ variational parameters. This ansatz has been shown to work well for small, single-digit k, and is considered to be one of the most promising ansätze in quantum computing.

Starting guess

There are several ways to choose the initial guess vector $\vec{\theta}$. In this thesis, we will choose it randomly, though it should be noted that smart guesses for unitary coupled cluster methods exist, such as using MP2 amplitudes [49], which accelerate convergence.

3.5.2 Alternative ansätze

The UCC approach and the approaches derived from it are *physically motivated ansätze (PMA)* where the wave function is parametrized in such a way that the FCI limit is approached systematically, based on chemical intuition. It is however not necessarily the most effective way in terms of circuit depth of creating an effective state. The alternative are the so-called *hardware heuristic ansätze (HHA)*, which 1. try to reduce the resource use of the quantum-hardware (also called *hardware efficient ansätze (HEA)*) 2. are tailored to the naturally available unitary operations on the quantum computer [40]. The trotterized UCCSD method and its derived trotterized methods lie somewhere between HEA and PMA. One prominent example of annother algorithm in between HEA and PMA is the adaptive algorithm ADAPT-VQE described in Ref. [50], where the true ground state is parameterized as a finite product reminiscent of eq. (3.8), which adaptively adds terms $\exp\left(\theta_\mu \hat{\sigma}_\mu\right)$ until some convergence criteria is hit - in comparison to UCCSD, not all excitations are included, while others are repeated several times. Pure HEA algorithms, such as the ones described in Refs. [43, 51], are usually based on combinations of one-qubit rotations and entanglement unitaries

$$|\Psi\rangle = U(\vec{\theta})\,|\Phi_0\rangle = \left(\prod_{i=1}^{D} U_{rot}(\vec{\theta}^{(D)})U_{\text{ENT}}\right) U_{rot}(\vec{\theta}^{(0)})\,|\Phi_0\rangle \tag{3.11}$$

where the entanglers U_{ENT} can for example be a set of CNOT-gates or a time development unitary $e^{-i\tau\hat{H}_0}$ for some qubit-Hamiltonian $\hat{H}_0$ which is independent of the system (the molecule's) Hamiltonian, while the one-qubit rotations are, for example, products of spin rotation matrices $R_j(\theta) = \exp\left(\frac{i\theta}{2}\sigma^j\right)$, $j \in \{X, Y, Z\}$.

3.6 Mapping creation and annihilation operators on qubits

In a basis consisting of M molecular orbitals, each Slater determinant can be expressed as a bit-string $|\Phi\rangle = |k_1, k_2, \ldots, k_M\rangle$ where k_p are occupation numbers. As $k_p \in \{0, 1\}$, the number of possible Slater determinants is 2^M, spanning a 2^M dimensional fermionic Fock space. The Hilbert space spanned by M qubits is 2^M dimensional too. The aim of this section is to describe how a 2^M dimensional Fock space can be mapped isomorphically on the Hilbert space spanned by M qubits. Because the action of any operator in second quantization can be expressed using complex numbers and second quantized operators, one only needs to find a mapping that conserves the algebraic properties of creation and annihilation operators. The result will be a Hamiltonian of the form

$$\hat{H} = \sum_{i_1, i_2, \ldots i_Q} c_{i_1, i_2, \ldots i_Q} \tilde{\sigma}^{i_1} \tilde{\sigma}^{i_2} \ldots \tilde{\sigma}^{i_Q} \tag{3.12}$$

which will consist of $O(M^4)$ terms.

3.6.1 The Jordan-Wigner mapping

The Jordan-Wigner mapping [52] is a particularly simple way to map from the creation operators for electronic orbitals to qubits. We have to find a way to associate a creation operator $a_j^\dagger$, describing an electronic orbital, to a qubit j. Intuitively, one might try to set

$$\begin{aligned} a_j^\dagger &\leftrightarrow \sigma_j^+ = \left(\sigma_j^X + i\sigma_j^Y\right)/2 \\ a_j &\leftrightarrow \sigma_j^- = \left(\sigma_j^X - i\sigma_j^Y\right)/2 \end{aligned} \tag{3.13}$$

However, this is not a valid representation. The anticommutator is not zero, as needed; the commutator is: $\{\sigma_j^+, \sigma_i^-\} \neq [\sigma_j^+, \sigma_i^-] = 0$. The following mapping resolves this problem:

$$\begin{aligned} a_j^\dagger &\leftrightarrow \left(\sigma^Z\right)_1 \otimes \ldots \otimes \left(\sigma^Z\right)_{j-1} \otimes \left(\sigma^+\right)_j \otimes \mathbb{I} \\ a_j &\leftrightarrow \left(\sigma^Z\right)_1 \otimes \ldots \otimes \left(\sigma^Z\right)_{j-1} \otimes \left(\sigma^-\right)_j \otimes \mathbb{I} \end{aligned} \tag{3.14}$$

that means, the Pauli σ_z matrix is applied to all preceding qubits, which applies a phase. This way, a correspondence between the second quantized basis states

$$\left(a_M^\dagger\right)^{k_M} \dots \left(a_1^\dagger\right)^{k_1} |0\rangle \tag{3.15}$$

and the quantum computer basis states

$$|k_1\rangle \otimes |k_2\rangle \otimes \dots \otimes |k_M\rangle \tag{3.16}$$

is induced. Using this mapping, the anticommutation relations are fulfilled. A drawback of this mapping is that, in general, one single electronic operation requires $O(M)$ qubit operations due to the non-locality of the mapped creation and annihilation operators. Accessing the occupation of a state is hence a local operation, while the parity is delocalized. The action of the σ^Z strings is to induce a phase change of -1 for odd parities, or do nothing for even parities, where the parity of a set of qubits is just the sum mod 2 of the numbers representing the state they are in [53].

3.6.2 The parity mapping

The parity mapping stores the parity of each state. Going from the Jordan-Wigner mapping to the parity mapping, we apply the following transfomration

$$|k_1, k_2, \dots, k_i, \dots, k_M\rangle \to |k_1, k_1 + k_2, \dots, \sum_{p=1}^{i} k_p, \dots, \sum_{p=1}^{M} k_p\rangle \tag{3.17}$$

where all sums are taken mod 2. While reading off the parity is now straightforward, when adding a particle at index j (that is, the action of $a_j^\dagger$), the parity of all indices $> j$ need to be inverted (using σ^X). In addition, whether to act with σ_j^+ or σ_j^- depends on the parity at position $j-1$, and the correct sign needs to be chosen to fulfill the anticommutation relations. The computational cost of applying a one-electron operator thus scales as $O(M)$, too.

3.6.3 The Bravyi-Kitaev mapping

In the Jordan-Wigner basis, the parity is stored non-locally, while the occupation number is stored locally; the parity basis does the opposite. The Bravyi-Kitaev basis hits of a compromise between these two, storing the parity and the occupation up to index j delocalized over $O(log(M))$ qubits. The actual form of the Bravyi-Kitaev mapping is rather complicated and depends on wether j is even or odd, and is given in Ref. [53]. It should be noted that the original Bravyi-Kitaev mapping only works when $M = 2^n$, $n \in \mathbb{N}$ [40], however, there are tree-based, very similar methods with the same scaling $O(log(M))$. Although the scaling of the Bravyi-Kitaev mapping is superior to the Jordan-Wigner mapping, it has a considerable overhead, and it has been estimated that for $M < 32$, the Jordan-Wigner mapping can be faster [54]. Furthermore, the Bravyi-Kitaev mapping loses some of the appealing interpretability of the parity mapping and the Jordan-Wigner mapping.

3.6.4 Reducing the number of qubits: Qubit tapering

In the mappings described, the number of qubits is identical to the number of spin orbitals M. However, different symmetries can be applied to reduce the number of qubits. The number of particles is conserved, and a proof by induction shows that the number of Slater determinants is $\binom{M}{N} \leq \binom{M}{M/2} \leq 2^{M-1}$. We can hence find a mapping to $M-1$ qubits. Furthermore, as the number of electrons with spin up and spin down $N_{\uparrow/\downarrow}$ is conserved as well $\left[\hat{N}_{\uparrow/\downarrow}, \hat{H}\right] = 0$, only a fraction of the 2^M states in the Fock space need to be considered, reducing the dimensionality further. As molecules have different types of point group symmetries and further spin symmetries, further dimensionality reduction is possible. This

has been discussed in, among others, Refs. [55, 56] and has been implemented in experiment [57] and works for all three types of mappings discussed.
Qubit tapering has several advantages: First of all, it may guarantee that the solution is an eigenfunction of some symmetry operation (most importantly, particle number conservation), it may reduce the computational cost (on simulated quantum computers) and the number of operations (on real quantum computers), and it allows to simulate larger systems with M molecular orbitals on quantum computers with a given number of qubits Q when $M - Q$ qubits can be tapered off.

2-qubit tapering in the parity mapping

In the parity mapping, qubit tapering can be understood intuitively. Writing the Slater determinant in bitstring from by first writing all alpha and then all beta electrons

$$|\Psi\rangle = |k_{\alpha_1}, \ldots, k_{\alpha_{M_\alpha}}, k_{\beta_1}, \ldots, k_{\beta_{M_\beta}}\rangle \tag{3.18}$$

and using that $\sum_i^{M_\alpha} k_{\alpha_i} = N_\alpha$ and identically for β, we see that all allowed states in the parity mapping will have

$$k_{\alpha_{M_\alpha}} = \begin{cases} 1 & \text{if } N_\alpha \text{ is even} \\ -1 & \text{if } N_\alpha \text{ is odd} \end{cases} \tag{3.19}$$

and similarly for $k_{\beta M_\beta}$. Thus, when spin is conserved, two qubits will always have the same value for all configuration and consequently do not need to be simulated, as we can replace the sum of all Pauli matrices in the Pauli strings acting on qubit $k_{\alpha_{M_\alpha}}$ by a number as the Hamiltonian is spin-number preserving.

General qubit tapering

With the details described in Ref. [56], we will here give an overview over how further symmetries can lead to further qubit reduction. For a Hamiltonian in the form of eq. (3.12), we are after a transformation such that the Pauli matrix in each Pauli string of eq. (3.12) acting on qubit number q is either the identity matrix or at most one Pauli matrix σ^i. The Pauli matrix can then be replaced by its eigenvalues ± 1, reducing the qubit requirement by one. It is possible to find a set of Pauli strings $\{\tau_1, \ldots, \tau_L\}$ that each commute with each string of the Hamiltonian (and thus the Hamiltonian). Furthermore, it is possible to find a set of unitary operators $\{U_1, \ldots, U_L\}$ that transforms those into $\sigma_q^X(1), \ldots, \sigma_q^X(L)$ acting trivially on all qubits but qubit $q(L)$. Each term of the unitary transformed Hamiltonian $U_i H U_i^\dagger$ will then commute with $\sigma_q^X(i)$, which is only possible if each Pauli string acts with σ^X or trivially on qubit $q(i)$, allowing for the aforementioned reduction. When k qubits are removed, 2^k possible eigenvalue combinations of the tapered qubits are possible, and the right symmetry sector has to be found by trial and error . How the sets $\{\tau_1, \ldots, \tau_L\}$, $\{U_1, \ldots, U_L\}$ are found is described in the articles by Bravyi et al. [56] and Setia et al. [58]. The paper by Setia et al. also expresses molecular point group symmetries in second quantization and consequently in terms of Pauli matrices and thus manages to taper off qubits by block-diagonalizing the Hamiltonian.

3.7 Implementation of Pauli matrix exponentials

Exponentials of creation/annihilation operators in second quantization can effectively be decomposed as an exponential of a sum of Pauli matrix Kronecker products. It remains to discuss how those can be implemented on quantum computers. Let

$$H = \bigotimes_{i=1}^{n} \sigma_{c(i)}^i \quad ; \quad U = e^{-i\tau H} \quad ; \quad \tau \in \mathbb{R} \tag{3.20}$$

σ^X and σ^Y terms can be transformed into σ^Z terms (stated differently, σ^X and σ^Y correspond to σ^Z terms in a different basis), $\sigma^X = R\sigma^Z R^\dagger$ and $\sigma^Y = V\sigma^Z V^\dagger$ for unitary R, V, and using repeatedly that $AC \otimes BD = (A \otimes B)(C \otimes D)$. To exemplify, we have, using $H = \sigma_1^X \otimes \sigma_2^Z \otimes \sigma_3^Y$

$$\begin{aligned} e^{-i\tau H} &= e^{-i\tau U_1 \sigma_1^Z R_1^\dagger \otimes \sigma_2^Z \otimes V_3 \sigma_3^Z V_3^\dagger} = e^{(R_1 \otimes \mathbb{1}_2 \otimes V_3)(-i\tau \sigma_1^Z \otimes \sigma_2^Z \otimes \sigma_3^Z)(R_1^\dagger \otimes \mathbb{1}_2 \otimes V_3^\dagger)} \\ &= (R_1 \otimes \mathbb{1}_2 \otimes V_3) e^{(-i\tau \sigma_T^Z 1 \otimes \sigma_2^Z \otimes \sigma_3^Z)} (R_1^\dagger \otimes \mathbb{1}_2 \otimes V_3^\dagger) \end{aligned} \tag{3.21}$$

where it was used that $e^{\Omega M \Omega^\dagger} = \Omega e^M \Omega^\dagger$ for unitary Ω and that the tensor product of unitary matrices is unitary. The element $e^{(-i\tau\sigma_T^Z 1 \otimes \sigma_2^Z \otimes \sigma_3^Z)}$ is diagonal in the computational basis and has elements $e^{\pm i\tau}$ depending on the parity of the n qubits. Albeit being highly non-local (acting on all n qubits), one can see by direct calculation that the inclusion of a single ancilla qubit can turn this into a local operation as described in Ref. [36]. This gives the quantum circuit shown in figure 3.2.

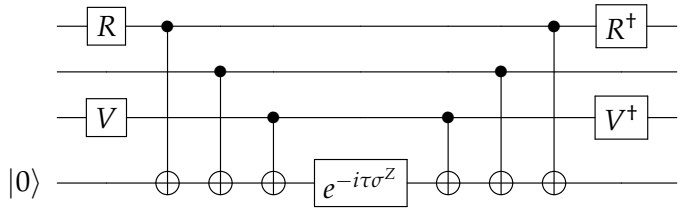


Figure 3.2: Quantum circuit for simulating $e^{-i\tau\sigma_1^X \otimes \sigma_2^Z \otimes \sigma_3^Y}$

A general decomposition of each term of the Jordan-Wigner transformed electronic Hamiltonian following this approach can be found in Ref. [59]. While we are not after a decomposition of the exponated electronic Hamiltonian, we need exponentials of creation and annihilation operators mapped to qubits. These are strings of Pauli operators of the form of eq. 3.20, which, for the Jordan-Wigner transform, is shown in Ref. [43].

3.8 Hamiltonian averaging

With the Hamiltonian in the form of eq. (3.12) and some ansatz $|\Psi(\vec{\theta})\rangle$, we now want to minimize the expectation value of the energy

$$\begin{aligned} \langle\Psi(\vec{\theta})| \hat{H} |\Psi(\vec{\theta})\rangle &= \langle\Psi(\vec{\theta})| \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q} \tilde{\sigma}^{i_1}\tilde{\sigma}^{i_2}\ldots\tilde{\sigma}^{i_Q} |\Psi(\vec{\theta})\rangle \\ &= \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q} \langle\Psi(\vec{\theta})| \tilde{\sigma}^{i_1}\tilde{\sigma}^{i_2}\ldots\tilde{\sigma}^{i_Q} |\Psi(\vec{\theta})\rangle \\ &= \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q} \langle\tilde{\sigma}^{i_1}\tilde{\sigma}^{i_2}\ldots\tilde{\sigma}^{i_Q}\rangle \\ &= \langle\hat{H}\rangle \end{aligned} \tag{3.22}$$

where we used the linearity of expectation values and the notation $\langle\Psi(\vec{\theta})| \hat{O} |\Psi(\vec{\theta})\rangle = \langle\hat{O}\rangle$. This decomposition into spin matrix expectation values has been labeled *Hamiltonian averaging* [60]. On simulated quantum computers, these expectation values can be calculated directly, on physical quantum computers, this is not the case as it would require measuring all qubits simultaneously, and each expectation value needs to be estimated by sampling. In Ref. [60], a simple prescription on the numbers of measurements necessary to obtain the expectation value to within a set tolerance ε is given. Because every Pauli string has eigenvalues ± 1, the variance of each Pauli string $\hat{P}_m$ is $\mathrm{Var}(\hat{P}_m) \leq 1$. In general, a new state preparation is necessary for each measurement as the wave function is collapsed, hence the covariance between two different Pauli strings $\mathrm{Cov}(\hat{P}_m \hat{P}_n) = 0$, thus

$$\mathrm{Var}(\hat{H}) = \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q}^2 \mathrm{Var}(\tilde{\sigma}^{i_1}\tilde{\sigma}^{i_2}\ldots\tilde{\sigma}^{i_Q}) \leq \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q}^2 \tag{3.23}$$

assuming real coefficients $c_{i_1,i_2,\ldots i_Q}$, and by the central limit theorem

$$\mathrm{Var}(\langle \hat{H} \rangle) = \frac{\mathrm{Var}(\hat{H})}{n_m} \tag{3.24}$$

where n_m is the number of repeated measurements of all Pauli strings, where each measurement of all Pauli strings consists of n_{noz} measurements, where be n_{noz} is number of non-zero coefficients $c_{i_1,i_2,\ldots i_Q}$. Thus, in order to obtain a variance $\mathrm{Var}(\langle \hat{H} \rangle) \leq \varepsilon^2$, the number of necessary measurements N_m becomes [61]

$$N_m = n_{noz} \frac{\sum_{i_1,i_2,\ldots i_Q} c^2_{i_1,i_2,\ldots i_Q}}{\varepsilon^2} \tag{3.25}$$

This number can be reduced by allowing a different n_m for each Pauli string, based on the value of $\mathrm{Var}(\tilde{\sigma}^{i_1}\tilde{\sigma}^{i_2}\ldots\tilde{\sigma}^{i_Q})$. Further reductions are possible by measuring those coefficients $c_{i_1,i_2,\ldots i_Q}$ with small absolute value less often, at the cost of introducing a bias, or through several measurements of commuting Pauli matrices which does not require a new state preparation, but triggers makes the mutual covariance non-zero [61]. In addition, the number of necessary measurements can be reduced by rewriting the Hamiltonian, or by using statistical methods, which we discuss briefly in the next section.

3.9 Scaling of methods

Having discussed how to map ansätze on qubits, how the circuits look like, and how to calculate expectation values, a final point to consider their scaling of these methods as function of the number of electrons N and basis size M. Generally speaking, the exact number of gates necessary to implement a circuit are strongly hardware-dependent and depend on the connectivity between qubits, natively available gates, number of qubits etc.. Further complications arise when considering what gates can be implemented in parallel. It is nevertheless possible to consider a general computational scaling in terms of the number of qubits, gate counts and the *circuit depth*, which is defined as the the number of "time steps" a circuit would take from the beginning to the end if each implemented gate took the same time to be executed and perfect parallel execution were possible. Circuit depth hence refers to time complexity, while the number of qubits and the gate count refer to hardware requirements. Often, only two-qubit gates are considered in the gate count, being the most expensive gates.

For trotterized unitary coupled cluster with k Trotter steps, the number of parameters is given by $k(M-N)^2N^2$ and thus scales as $O(kM^4)$ when parameters are allowed to vary independently for each Trotter step and $O(M^4)$ otherwise. A serial implementation on a quantum computer hence requires $O(fkM^4)$ gates, where f depends on way the (de)excitation operators are mapped on qubits. for the Jordan-Wigner mapping, $f = O(M)$, for the Bravyi-Kitaev mapping, $f = \log(M)$. It should however be noted that the prefactor f can be eliminated in the Jordan-Wigner and parity mappings [62] by a clever decomposition of the strings of creation/annihilation operators, which is based by a double factorization of the UCC-amplitudes and the Hamiltonian, writing the rank 4 tensors as products of lower-rank tensors (*tensor hypercontraction*) [63]. We will thus drop the f dependency. Methods such as k-UpGCCSD hence scale as $O(kM^2)$. k-UPGCCSD has a circuit depth of $O(kM)$, while UCCSD scales as $O(kM^3)$ [48]. k-UPGCCSD is the algorithm with the lowest scaling we are aware of [64]. For HEA-ansätze, the gate count varies depending on the ansatz, but the one given in eq. (3.11) scales as $O(DM)$ when U_{ENT} concists of applying a linear number of CNOT gates per qubit, and has a small circuit depth of $O(D)$ when those gates are applied in parallel.

Hamiltonian averaging requires a vast number of measurements, and using a formula similar to eq. (3.25), where ideal distribution of measurements is assumed to reduce the error, the number of measurements required still scales as $O(M^4/\epsilon^2)$, with the $O(M^4)$ term stemming from the number of terms in the Hamiltonian. When requiring chemical accuracy,

the required number of measurements can be prohibitively large. Based on the same tensor hypercontraction as the one presented in Ref. [62], the number of terms to measure in the Hamiltonian can be reduced to $O(M)$ [65], while adding an overhead of $O(M)$ to the state preparation. This is the strongest reduction we are aware of [64]. The scaling in $1/\epsilon^2$ can also be reduced. The likelihood function to measure 0 as function of the expectation value of a Pauli string $f(0|P)$ is linear, but following the approach described in Ref. [66], it is possible to apply gates to turn the likelihood function into a Chebyshev polynomial. Using Bayesian interference, it is possible to obtain a distribution for the expectation value which has a much smaller standard deviation, at the cost of more expensive circuits. This way, it is possible to find a scaling lying between $1/\epsilon$ (which is the lowest possible limit) and $1/\epsilon^2$, depending on the longest acceptable circuit length, which depends on decoherence times and gate errors.

3.10 Calculation of matrix elements on a quantum computer

Assume we have two wave functions $|\Psi_i\rangle$ and $|\Psi_j\rangle$, and we want to calculate $\langle\Psi_i|\hat{O}|\Psi_j\rangle$ for some observable expressed by the Hermitian operator $\hat{O}$. For eigenvector continuation, $\hat{O} \in \{\mathbb{1}, \hat{H}\}$. Diagonalizing the Hamiltonian in a basis of several non-orthogonal wave functions on a quantum computer has been done in previous work [67–69] which all considered similar approaches, and we will here describe the approach by Parrish et al. [69]. Let $|\Phi\rangle = |0\ldots0\rangle$ be the initial state, and $|\Psi_i\rangle = \hat{U}_i|\Phi\rangle$, $|\Psi_j\rangle = \hat{U}_j|\Phi\rangle$ for some unitary operators $\hat{U}_i$, $\hat{U}_j$. The first step is to create the state

$$|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \otimes \hat{U}_i|\Phi\rangle + |1\rangle \otimes \hat{U}_j|\Phi\rangle\right) \tag{3.26}$$

which can be done with the circuit in figure 3.3.

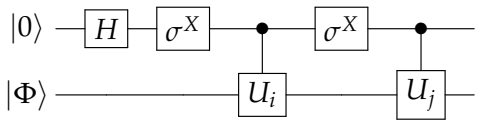


Figure 3.3: Quantum circuit for creating the state $|\Omega\rangle$. The first σ^X gate can be ommitted, but was added to signify a basis change operation.

These controlled unitary operations, which are costly, are not strictly necessary, the same state can be obtained using SWAP gates which interchange the values of qubits at locations i and j, at the expense of requiring an extra system with the same number of qubits as the state $|\Phi\rangle$ [67]. It is simple to verify that expectation values can be calculated as

$$\langle\Psi_i|\hat{O}|\Psi_j\rangle = \langle\Omega|\hat{\sigma}_+ \otimes \hat{O}|\Omega\rangle = \langle\Omega|\hat{\sigma}^X \otimes \hat{O}|\Omega\rangle + i\langle\Omega|\hat{\sigma}^Y \otimes \hat{O}|\Omega\rangle \tag{3.27}$$

where $\langle\Omega|\hat{\sigma}^X \otimes \hat{O}|\Omega\rangle$ and $\langle\Omega|\hat{\sigma}^Y \otimes \hat{O}|\Omega\rangle$ are the real and the imaginary part of the matrix element $\langle\Psi_i|\hat{O}|\Psi_j\rangle$, respectively.
In the case that we only are interested in the overlap $\langle\Psi_i|\Psi_j\rangle$, a method to calculate this without the need of controlled unitaries or extra qubits is to simply measure

$$\langle\Psi_i|\Psi_j\rangle = \langle\Phi|\hat{U}_i^\dagger\hat{U}_j|\Phi\rangle\,. \tag{3.28}$$

3.11 Finding excited states on quantum computers

There are many ways to find excited states on quantum computers, but we will here follow the approach by Higgot et al. [70]. They suggested a method to find excited states on

quantum computers by introducing a modified Hamilton operator

$$\hat{H}_k = \hat{H} + \sum_{i=0}^{k-1} \beta_i \left|\Psi_i\right\rangle \left\langle\Psi_i\right| \tag{3.29}$$

where $|\Psi_i\rangle$ is the i'th eigenfunction of $\hat{H}$ and β_i is a real, nonnegative parameter. Clearly, as $\beta_i \to \infty$ for $i < k$, the $k'th$ excited state of $\hat{H}$ will be the ground state of $\hat{H}_k$, though using ∞ is not necessary, as the energy gap between two eigenfunctions of the Hamiltonian cannot be larger than $2\left\|\hat{H}\right\|$ for the operator norm. Hence, the variational principle can be used with the effective Hamiltonian $\hat{H}_k$ to find excited states of $\hat{H}$. For a sample wave function $|\Phi(\theta)\rangle$, we will then find the energy of a sample wave function to be

$$\langle\Phi(\theta)|\hat{H}_k|\Phi(\theta)\rangle = \langle\Phi(\theta)|\hat{H}|\Phi(\theta)\rangle + \sum_{i=0}^{k-1} \beta_i |\left\langle\Phi(\theta)|\Psi_i\right\rangle|^2 \tag{3.30}$$

which is to be minimized for $|\Phi\rangle = \hat{U}(\theta)\left|\Phi^{\text{ref}}\right\rangle$. It is important that the ansatz $\hat{U}(\theta)$ is sufficiently flexible to represent excited states, which is the case at least for the first excited state for GCCSD and k-UpCCSD as well as repeated applications of UCCSD [48].

3.12 Linear combination of unitaries

Sometimes, one wants to implement on a quantum computer a normalized state proportional to

$$|\Psi\rangle = \hat{M}|\Phi\rangle = \sum_{i=0}^{N-1} c_i \hat{U}_i |\Phi\rangle \tag{3.31}$$

where $\hat{M}$ is a matrix which is not necessarily unitary and $|\Phi\rangle$ is an initial state. In order to do so, one can use the *linear combination of unitaries (LCU)* algorithm [71, 72]. Observe that decomposing any matrix as linear combination of unitaries is always possible, for example, in terms of Pauli matrices (cf. sec. 3.2.4).

In order to implement the LCU algorithm, we first make the coefficients c_i real and positive. This is always possible by putting the phase on the unitaries instead. Furthermore, renormalize such that $\sum_{i=0}^{N-1} c_i^2 = 1$. Then we define

$$\begin{aligned} n &= \lceil \log_2(N) \rceil \\ c &= \sum_{i=0}^{N-1} c_i \\ \hat{V}\left|0_1..0_n\right\rangle &= \sum_{i=0}^{N-1} \sqrt{\frac{c_i}{c}} \left|i_{(2)}\right\rangle \quad \text{where } \hat{V}^\dagger \hat{V} = I \\ \hat{U} &= \sum_{i=0}^{N-1} \left|i_{(2)}\right\rangle \left\langle i_{(2)}\right| \hat{U}_i. \end{aligned} \tag{3.32}$$

It should be noted that the unitary $\hat{V}$ is not uniquely defined, only its action on the state $|0_1..0_n\rangle$. Then, a calculation shows that

$$\left(\hat{V}^\dagger \otimes \mathbb{1}\right) \hat{U} \left(\hat{V} \otimes \mathbb{1}\right) |0^n\rangle \otimes |\Phi\rangle = \frac{1}{c} |0^n\rangle \otimes \hat{M}|\Phi\rangle + |\Phi^\perp\rangle \tag{3.33}$$

where $|\Phi^\perp\rangle$ is orthogonal to $|0^n\rangle \otimes |\Psi\rangle$. $\hat{U}$ can be implemented as a product of controlled unitaries, controlled by the state $|i\rangle$. In order to remove the undesired part $|\Phi^\perp\rangle$, one needs to measure the n ancilla qubits in the computational basis and obtain the state $|0^n\rangle$, the success probability of which is given by $p_s = \frac{1}{c^2} \langle\Phi| \hat{M}^\dagger \hat{M} |\Phi\rangle$.

3.13 Errors on quantum computers

There are many possible sources that can lead to wrong energy calculations on a quantum computer. Within a given basis set, the following errors lead to deviations from exact diagonalization of the Hamiltonian:

1. The ansatz $\hat{U}(\vec{\theta})\,|\Phi^{\text{SD}}\rangle$ not being able to represent the exact FCI wave function.
2. Optimization errors leading to $\langle\Psi(\vec{\theta}^{\text{opt.}})|\,\hat{H}\,|\Psi(\vec{\theta}^{\text{opt.}})\rangle < \langle\Psi(\vec{\theta}^{\text{obt.}})|\,\hat{H}\,|\Psi(\vec{\theta}^{\text{obt.}})\rangle$, where $\vec{\theta}^{\text{opt.}}$ is the optimal set of parameters $\vec{\theta}$, while $\vec{\theta}^{\text{obt.}}$ is the obtained $\vec{\theta}$ after some optimization procedure. This may happen when $\langle\Psi(\vec{\theta}^{\text{obt.}})|\,\hat{H}\,|\Psi(\vec{\theta}^{\text{obt.}})\rangle$ is a local minimum.
3. Sampling errors when sampling $\langle\hat{H}\rangle$. This is relevant for optimization as well as the final energy calculation.
4. Noise due to faulty gates, decoherence, etc.

The first two points are not unique to quantum computers, as exact methods often are prohibitively expensive on standard computers, and optimization methods can give wrong results. However, the second point is especially true for HEA ansätze on quantum computers, as rotation operators are periodic functions. Hence, a global minimum does not exist, and there are no physically motivated starting guesses $\vec{\theta}_0$. The third problem is also relevant in quantum Monte Carlo methods [73], but unusual for standard wave-function based approaches. The last point is unique for quantum computers in the sense that we do not need to consider "faulty" hardware on modern computers anymore. Furthermore, it should be understood that points 3 and 4 have a strong impact on point 2, as expectation values and gradients, which are necessary for parameter optimization, are compromised by inexact expectation values.

Part III

Methods

Chapter 4

Methods

The aim of this chapter is to present how eigenvector continuation can be used in chemistry. As eigenvector continuation requires analyticity of the Hamiltonian as a function of a parameter $\vec{\alpha}$, we focus on ways to make the second-quantized Hamiltonian analytic. As the second-quantized Hamiltonian is directly related to the expansion coefficients of the molecular orbitals, we consider in some detail how unitary operations can be used to make changes in the molecular orbitals both analytic and small. We will first consider the general requirements that the MOs as a function of geometry have to fulfill in order for eigenvector continuation to work. We will then consider several choices of MOs that can be used by using symmetric orthonormalization, Cholesky decomposition and solving the orthogonal Procrustes problem, as well as natural orbitals. After that, we will introduce the AMP-CCEVC approach where the amplitudes, not the wave functions, are continued. Based on the unitary operations, we will introduce different ways to apply eigenvector continuation to obtain a potential energy surface based on multi-configurational and on single-configurational methods. Finally, we will consider how eigenvector continuation can be implemented on a quantum computer and the impact of sampling errors there, as well as advantages of quantum computing.

4.1 Eigenvector continuation for quantum chemistry

In chapter 2, we have considered eigenvector continuation for eigenvectors of matrices $\mathbf{M}(\vec{\alpha})$. There, the underlying vector space did not change. We have also considered coupled cluster wave functions of the form $e^{\hat{T}(\vec{\alpha})}|\Phi^{\mathrm{SD}}\rangle$ - there, the reference determinant $|\Phi^{\mathrm{SD}}\rangle$ was the same for all values of $\vec{\alpha}$. The latter approach might work in quantum chemistry when the reference determinant $|\Phi^{\mathrm{SD}}\rangle$ remains a good guess, as for perturbations to the Hamiltonian of the form $\hat{H}(\alpha) = \hat{H}_0 + \alpha\hat{V}$, where $\hat{V}$ has an impact on the electronic state of the wave function, but does not impact the relative positions of the nuclei. However, throughout this thesis, we want to perform eigenvector continuation to obtain a potential energy surface. We consider hence a Hamiltonian $\hat{H}(\vec{R})$, where $\vec{R}$ describes nuclear positions. A reference state, expressed in a basis localized at $\vec{R} = \vec{R}_{\mathrm{ref}}$, will be a bad reference state for all other values than $\vec{R} = \vec{R}_{\mathrm{ref}}$ because we cannot expect that a wave function centered around $\vec{R}! = \vec{R}_{\mathrm{ref}}$ can be expressed in a basis centered around $\vec{R}_{\mathrm{ref}}$. When the basis set moves with the nuclei, not only the onebody and the twobody elements $h_{pq}(\vec{R})$ and $\langle pq\|rs\rangle(\vec{R})$ are changed, the creation and annihilation operators $a_p(\vec{R})$,$a_q^\dagger(\vec{R})$ now refer to a molecular orbital at a different geometry. Thus, performing "naive" eigenvector continuation using the states $|\Psi(\vec{R})\rangle$ directly is bound to fail.
However, we can consider how we can still use a basis set centered at a geometry $\vec{R}$, while keeping information from geometry $\vec{R}_{\mathrm{ref}}$. This is best expressed in second quantization, where all basis dependence can be removed from the creation and annihilation operators. As the eigenvalues of the Hamiltonian are invariant to unitary operations, including orbital rotations, there are infinitely many choices of MOs.

4.1.1 Analyticity of the Hamiltonian

In second quantization, the Hamiltonian with nuclei placed at positions $\vec{R}$ reads (cf. eq. (1.14)):

$$\hat{H}(\vec{R}) = \sum_{pq} h_{pq}(\vec{R}) a_p^\dagger(\vec{R}) a_q(\vec{R}) + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle(\vec{R}) a_p^\dagger(\vec{R}) a_q^\dagger(\vec{R}) a_s(\vec{R}) a_r(\vec{R}) + h_{\text{nuc}}(\vec{R}) \tag{4.1}$$

both the coefficients and the creation and annihilation operators are position dependent. For the creation and annihilation operators, as they refer to an MO that is a linear combination of AOs at geometry $\vec{R}$. However, one might also consider the creation and annihilation operators to be position independent, as this basis dependence can be neglected: ON-strings are basis independent, and as long as the MOs at different geometries remain orthogonal, we consider the Fock spaces at different geometries as one Fock space, with creation and annihilation operators as constant entities, as was done in Ref. [74]. From that point of view, the geometric dependence of the creation and annihilation operators is irrelevant in Fock space. All geometry information is contained in the coefficients $C_{i\mu}$ and the one-body and two-body integrals. We will hence write the Hamiltonian

$$\begin{aligned} \hat{H}(\vec{R}) &= \sum_{pq} h_{pq}(\vec{R}) a_p^\dagger a_q + \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle(\vec{R}) a_p^\dagger a_q^\dagger a_s a_r + h_{\text{nuc}}(\vec{R}) \\ &= \sum_{\mu\nu} \sum_{pq} C_{\mu p}^*(\vec{R}) C_{\nu q}(\vec{R}) h_{\mu\nu}(\vec{R}) a_p^\dagger a_q \\ &\quad + \frac{1}{4} \sum_{\mu\nu\sigma\tau} \sum_{pqrs} C_{\mu p}^*(\vec{R}) C_{\nu q}^*(\vec{R}) C_{\sigma r}(\vec{R}) C_{\tau s}(\vec{R}) \langle \mu\nu \| \sigma\tau \rangle(\vec{R}) a_p^\dagger a_q^\dagger a_s a_r + h_{\text{nuc}}(\vec{R}) \end{aligned} \tag{4.2}$$

where the elements $h_{\mu\nu}(\vec{R})$, $\langle \mu\nu \| \sigma\tau \rangle(\vec{R})$ are a functions of the geometry and the choice of basis, while the expansion coefficients $C_{\mu p}(\vec{R})$ can be chosen freely as long as the resulting MOs are orthonormal for all $\vec{R}$.
In order for eigenvector continuation and perturbation theory to work, eq. (4.2) needs to be analytic as function of the nuclear positions $\vec{R}$. We have already considered a special choice of MOs that assure this in section 1.2.1, but this choice is not unique: The analyticity and orthogonality of the natural orbitals makes those a viable choice too. It is sufficient to make the expansion coefficients $C_{\mu p}(\vec{R})$ analytic, making the Hamiltonian analytic, making thus a given choice of the eigenvectors analytic too. The complicated geometry dependence makes it nevertheless clear that it is unlikely to get convergence as good as the one shown in figure 2.1.

4.2 Unitary operations applied to the coefficient matrix: orbital rotations

4.2.1 Symmetric orthonormalization of MOs

Assume we have solved the Hartree-Fock equations at a molecular geometry $\vec{R}$. This gives rise to a coefficient matrix $\mathbf{C}_{\vec{R}}(\vec{R})$ and a Hartree-Fock ground state wave function $|\Phi_{\vec{R}}^{\text{HF}}(\vec{R})\rangle$ (the need of the subscripts will become clear after reading this section). The question that arises is how to describe the same configuration at a different geometry $\vec{R}'$. At the new geometry, the atom-centered basis functions change, as the basis functions are now centered at different positions, with atom number i at position R_i', and $R_i' \neq R_i$ for at least one i. Using the same coefficient matrix $\mathbf{C}_{\vec{R}}(\vec{R})$ at geometry $\vec{R}'$ leads to some complications, as the change in the basis functions will make the MOs non-orthogonal and the wave-function non-normalized, complicating the mathematics greatly, hence the need to re-orthonormalize the MOs by defining an orbital connection [75]. This procedure is not unique, and we will here simply use symmetric orthonormalization as described in Ref. [74], which has the advantage that the new MOs are as similar to the old ones as possible in a least-square sense [76]. Let

$\mathbf{S}(\vec{R}')$ be the overlap matrix between the MOs of configuration $\vec{R}$, with the atomic orbitals centered at configuration $\vec{R}'$. At geometry $\vec{R}$, this is the identity matrix, but this will not be the case at other geometries. An orthonormal, new coefficient matrix can now be expressed as

$$\mathbf{C}_{\vec{R}}(\vec{R}') = \mathbf{S}^{-\frac{1}{2}}(\vec{R}')\mathbf{C}_{\vec{R}}(\vec{R}) \tag{4.3}$$

which explains the notation: The coefficient matrix corresponds now to a set of orthonormal MOs created from basis functions centered at $\vec{R}'$, but it is not the Hartree-Fock solution, but represents the same configuration as the one solved at $\vec{R}$ in a least-square sense.
A further complication is the consideration of the coefficient matrix as consisting of a virtual and an occupied part. Either we convert the whole coefficient matrix according to eq. (4.3), an approach which will mix the occupied and the virtual coefficient matrices, which might lead to larger changes of the occupied block. We will hence follow also consider a second approach, converting the coefficient matrices independently

$$\mathbf{C}_{\vec{R}}^{(o/v)}(\vec{R}') = \left(\mathbf{S}^{(o/v)}\right)^{-\frac{1}{2}}(\vec{R}')\mathbf{C}_{\vec{R}}^{(o/v)}(\vec{R}) \tag{4.4}$$

where we previously subtracted the projections of all occupied MOs from each virtual orbital, as to ascertain orthonormality between occupied and virtual molecular orbitals. We will refer to the approach where the whole coefficient matrix is converted as "general symmetric orthonormalization", while the approach with independent transformations will be called "symmetric orthonormalization". Because the expectation energy of the new Slater determinant built from $\mathbf{C}_{\vec{R}}(\vec{R}')$ will likely not have a stationary energy, Brillouin's theorem (cf. sec. 1.3, p. 12) does not hold anymore.

4.2.2 Cholesky MOs

It is a fact [77] that any Hermitian positive-definite matrix $\mathbf{A}$ can be Cholesky decomposed as

$$\mathbf{A} = \mathbf{L}\mathbf{L}^\dagger = \mathbf{R}^\dagger\mathbf{R} \tag{4.5}$$

where $\mathbf{L}$ ($\mathbf{R}$) is lower (upper) triangular. For real matrices $\mathbf{A}$ we have that $\mathbf{L}^\dagger = \mathbf{L}^T$.
When allowing for pivoting, the Cholesky decomposition extends to positive-semidefinite matrices [77] and reads, for real matrices

$$\mathbf{P}^T\mathbf{A}\mathbf{P} = \mathbf{L}\mathbf{L}^T \tag{4.6}$$

$$\mathbf{A} = (\mathbf{P}\mathbf{L})(\mathbf{P}\mathbf{L})^T \tag{4.7}$$

where $\mathbf{P}$ is a permutation matrix. An explicit algorithm that also considers stopping criteria is found in Ref. [78]. It should be noted that pivoting and requiring $\mathbf{L}$ to have positive diagonal entries makes the decomposition unique, unless some of the diagonal elements of $\mathbf{A}$ are equal at some point in the algorithm, in which uniqueness can still be enforced through some additional criteria.
The pivoted Cholesky decomposition can be applied to get a unique representation of molecular orbitals [79]. While they were proposed as a way to localize molecular orbitals, we are mainly interested in the uniqueness and guaranteed convergence of the algorithm. All "physics" of the HF-state are determined by the density matrix $\mathbf{D}^o$ (for real coefficients and assuming restricted Hartree Fock)

$$\mathbf{D}^{(o/v)} = \mathbf{C}^{(o/v)}\left(\mathbf{C}^{(o/v)}\right)^T = (\mathbf{C}\mathbf{U})^{(o/v)}\left((\mathbf{C}\mathbf{U})^{(o/v)}\right)^T \tag{4.8}$$

There is an infinite number of coefficient matrices $\mathbf{C}^{(o/v)}$ satisfying this equation , but a very restricted number of pivoted Cholesky decompositions - uniqueness occurs when there are no molecular symmetries, and even then, uniqueness can be enforced. Applying the pivoted Cholesky decomposition to the density matrix, we get

$$\mathbf{D}^{(o/v)} = \mathbf{C}^{(o/v)}\left(\mathbf{C}^{(o/v)}\right)^T = ((\mathbf{P}\mathbf{L})^{(o/v)})\left((\mathbf{P}\mathbf{L})^{(o/v)}\right)^T \tag{4.9}$$

and the first N vectors in $(\mathbf{PL})^{(o)}$ will describe a new set of occupied MOs and the first M vectors in $(\mathbf{PL})^{(v)}$ will describe a new set of virutal MOs, giving raise to the *Cholesky molecular orbitals*. We have thus constructed a unique representation of the HF coefficient matrix.

4.2.3 Minimizing distance between Molecular orbitals at different geometries: Procrustes orbitals

Assume that we have a two coefficient matrices, representing a Hartree-Fock wave function, that were obtained at two different geometries, $\mathbf{C}(\vec{R})$ and $\mathbf{C}(\vec{R}')$. We are interested in performing a unitary operation $\mathbf{\Omega}^{(o/v)}$ on the occupied and virtual MOs in $\mathbf{C}(\vec{R})^{(o/v)}$ in such a way that $\mathbf{C}(\vec{R}')^{(o/v)}$ is as close to $\mathbf{C}(\vec{R})^{(o/v)}$ as possible. We then obtain a new coefficient matrix $\boldsymbol{C}_{\vec{R}}(\vec{R}')^{(o/v)} = \mathbf{C}(\vec{R}')\mathbf{\Omega}^{(o/v)}$ for unitary $\mathbf{\Omega}^{(o/v)}$. This does not change the Slater determinant $|\Psi^{\mathrm{SD}}(\vec{R})\rangle$ except for a phase, as stated in section 1.3. Mathematically, we try to find

$$\mathbf{Q}^{(o/v)} = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega} - \mathbf{C}(\vec{R})^{(o/v)} \right\|_F \quad \text{subject to} \quad \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{1} \tag{4.10}$$

where $\|\cdot\|_F$ is the Frobenius norm. The Frobenius norm of an $m \times n$ matrix $\mathbf{A}$ is defined as

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|^2} = \sqrt{\mathrm{tr}\left(\mathbf{A}^\dagger\mathbf{A}\right)} = \sqrt{\mathrm{vec}(\mathbf{A})^\dagger \mathrm{vec}(\mathbf{A})} \tag{4.11}$$

In principle, any other norm can be chosen, but using the Frobenius norm, this problem, known as the *orthogonal Procrustes problem*, has an analytical solution [80]: Letting $\mathbf{M} = \mathbf{C}(\vec{R}')^T\mathbf{C}(\vec{R}) = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ using the Singular Value Decomposition [81], the solution is given by

$$\mathbf{Q} = \mathbf{U}\mathbf{V}^T. \tag{4.12}$$

The proof of this can be found in the appendix, sec. 10.1, p. 90. We will refer to the orbitals expressed by $\boldsymbol{C}_{\vec{R}}(\vec{R}') = \begin{bmatrix} \boldsymbol{C}_{\vec{R}}(\vec{R}')^{(o)} & \boldsymbol{C}_{\vec{R}}(\vec{R}')^{(v)} \end{bmatrix}$ as *Procrustes orbitals* with respect to a geometry $\vec{R}$. Procrustes orbitals are thus the set of orbitals that minimize the change in expansion coefficients wrt. to a reference geometry $\vec{R}$. It should be noted that the application of the orthogonal Procrustes problem to orbital localization is not new and has been done in, for example, Ref. [82]. A small lemma is that applying unitary rotations to the reference $\mathbf{C}(\vec{R})^{(o/v)}$ will not give a better fit.

Lemma 4.2.1. *The Procrustes-problem is basis-independent in the sense that a unitary operation applied to the reference MO does not change the distance metric.*

Proof. Let $\mathbf{U}, \mathbf{\Omega}$ be unitary matrices. Then

$$\begin{aligned} &\min_{\mathbf{\Omega},\mathbf{U}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega} - \mathbf{C}(\vec{R})^{(o/v)}\mathbf{U} \right\|_F \\ &= \min_{\mathbf{\Omega},\mathbf{U}} \left\| \left(\mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega}\mathbf{U}^{-1} - \mathbf{C}(\vec{R})^{(o/v)} \right) \mathbf{U} \right\|_F \\ &= \min_{\mathbf{\Omega},\mathbf{U}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega}\mathbf{U}^{-1} - \mathbf{C}(\vec{R})^{(o/v)} \right\|_F \end{aligned} \tag{4.13}$$

where we used that unitary operations preserve the Frobenius norm. Observing that $\mathbf{\Omega}\mathbf{U}^{-1}$ is unitary, finishes the proof. □

Generalized Procrustes orbitals

The Procrustes orbitals method can be trivially extended by solving the orthogonal Procrustes problem for the whole coefficient matrix, allowing for mixing of the occupied and unoccupied orbitals. By lemma 4.2.1, the minimal distance obtained is then independent of

the choice of $\mathbf{C}(\vec{R}')$. We prove that the use of generalized Procrustes orbitals guarantees the analyticity of the coefficient matrix in the appendix, theorem 10.2.1, p. 91. This generalizes to non-generalized Procrustes orbitals too, but requires the analiticity of the span of the occupied canonical orbitals and virtual canonical orbitals individually, e.g. the analyticity of the HF state, which is not always the case.

4.2.4 Natural orbitals

We have discussed natural orbitals and their analiticity in section 1.9, p. 20. Even though we stated that analiticity does not necessarily hold for approximate NOs and for NOs obtained by diagonalizing the occupied-occupied and virtual-virtual block of the 1-RDM individually, they are analytic when the canonical MOs are analytic. Using natural orbitals, we can discard of those NOs with the smallest natural occupation numbers. For this to work, however, we require that natural occupation numbers remain reasonably constant along the whole reaction coordinate $\vec{R}$. We have observed that this is usually the case. This is exemplified in figure 4.1, showing a random selection of natural occupation numbers as function of the dissociation of hydrogen fluoride and the symmetric stretch of beryllium dihydride. We see that albeit the magnitudes change along the reaction path, most small natural occupation numbers remain small. This leads to the following idea: Only keep those natural orbitals that are important at any of the sampling geometries and discard the rest.

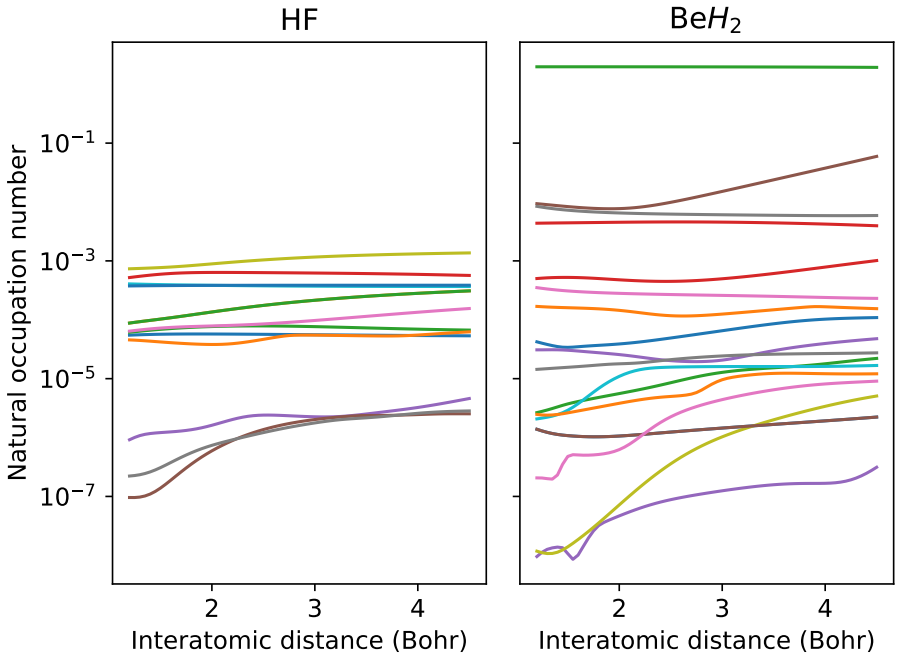


Figure 4.1: Natural occupation numbers of 20% *randomly chosen natural orbitals obtained from the first order perturbation theory wave function with the occupied and the virtual blocks of the 1-RDM diagonalized individually, as function of the intermolecular distance for the symmetric dissociation of hydrogen fluoride and beryllium dihydride in the aug-cc-pVTZ basis set. While the natural occupation numbers vary greatly, a general distinction in "important" and "unimportant" can easily be made.*

4.3 AMP-CCEVC

In the WF-CCEVC approach (sec. 2.3, p. 26), the resulting wave function is a linear combination of coupled cluster wave functions. A different approach is to perform eigenvector continuation on the coupled cluster amplitudes t_μ. Instead of writing the resulting wave function as a linear combination of sample wave functions, one can express

the cluster operator as linear combination of sample cluster operators. Thus, the AMP-CCEVC wave function will take the shape of a coupled cluster wave function,

$$|\Psi^{\text{AMP-CCEVC}}(\vec{\alpha})\rangle = e^{\hat{T}^*}|\Phi^{\text{SD}}\rangle = \exp\left\{\sum_{m=1}^{L} c_m \hat{T}(\vec{\alpha}_m)\right\}|\Phi^{\text{SD}}\rangle\,. \tag{4.14}$$

Technically speaking, no eigenvector continuation is performed, but a dimensionality reduction, as there are only L variables to be determined. The parameters c_m are determined using the standard way, which is by solving a set of amplitude equations 1.65

$$\langle\Phi^{\text{SD}}|\,\hat{T}(\vec{\alpha}_m)^\dagger \bar{H}\,|\Phi^{\text{SD}}\rangle = 0 \quad \text{for all } m \tag{4.15}$$

$$\langle\Phi^{\text{SD}}|\,\bar{H}\,|\Phi^{\text{SD}}\rangle = E. \tag{4.16}$$

This is valid because we made no assumption about the states $\langle\mu|$ in the derivation of the amplitude equations 1.65, other than that they are orthogonal to the reference determinant $|\Phi^{\text{SD}}\rangle$. It is also customary that the states are mutually orthogonal and normalized $\langle\mu|\nu\rangle = \delta_{\mu\nu}$, which is not the case here

$$\langle\Phi^{\text{SD}}|\,\hat{T}(\vec{\alpha}_i)^\dagger \hat{T}(\vec{\alpha}_j)\,|\Phi^{\text{SD}}\rangle \neq \delta_{ij} \tag{4.17}$$

however, given that eq. (4.15) holds, then it holds for any linear combination of the states $\langle\Phi^{\text{SD}}|\sum_{m=1}^{L} c_m \hat{T}(\vec{\alpha}_m)^\dagger$, including an orthonormal one. Hence, the approaches are equivalent. We will discuss how to implement AMP-CCEVC in greater detail in the numerical methods section, sec. 5.4, p. 62).

4.3.1 Parameter reduction

In the amplitude approach, there are only L parameters to optimize, compared to CCSD, where there are $O(N^2M^2)$. However, there is no direct numerical advantage, as one still needs to calculate $\langle\mu|\,\hat{H}\,|\Phi^{\text{SD}}\rangle$ for $\mu \in S, D, \ldots$:

$$\mathrm{e}(c_m) = \langle\Phi^{\text{SD}}|\,\hat{T}(\vec{\alpha}_m)^\dagger \bar{H}\,|\Phi^{\text{SD}}\rangle = \sum_{\mu} t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\text{SD}}\rangle\,. \tag{4.18}$$

$e(c_m)$, the projection error, is a linear combination of the projection errors of each excitation and thus not cheaper to calculate. This is also the case for the Jacobian - even when using a Quasi-Newton approach, one still needs to consider the whole diagonal part of the CCSD-Jacobian when calculating

$$\frac{\partial}{\partial c_n} e(c_m) \tag{4.19}$$

However, it is not unreasonable to assume that both the Jacobian and the error vector are well approximated by considering only a small subset of the excitations, that is, assuming that

$$e(c_m) = \langle\Phi^{\text{SD}}|\,\hat{T}(\vec{\alpha}_m)^\dagger \bar{H}\,|\Phi^{\text{SD}}\rangle = \sum_{\text{all }\mu} t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\text{SD}}\rangle \approx \tilde{e}(c_m) = \sum_{\text{some }\mu} t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\text{SD}}\rangle \tag{4.20}$$

Because of the shape of the coupled cluster amplitude equations, this choice of "some μ" is important. Selecting a given percentage randomly, or by the largest absolute value of the amplitudes will not necessarily give any computational savings. In the CCSD approximation, the projection error reads

$$e(c_m) = \sum_{ia} t_i^a(\vec{\alpha}_m)\,\langle\Phi^{\text{SD}}|\,a_i^\dagger a_a \bar{H}\,|\Phi^{\text{SD}}\rangle + \frac{1}{4}\sum_{ijab} t_{ij}^{ab}(\vec{\alpha}_m)\,\langle\Phi^{\text{SD}}|\,a_i^\dagger a_j^\dagger a_b a_a \bar{H}\,|\Phi^{\text{SD}}\rangle\,. \tag{4.21}$$

In order to cut down the number of floating point operations, one has to not sum over all virtual indices, but only a subset thereof, and possibly do the same for occupied indices. We

will now justify how this reduces computational cost. The most computationally expensive equations in CCSD-theory in a direct product decomposition is the calculation of the $\mathcal{W}_{abef}$ intermediate and its contribution to the t_{ij}^{ab} amplitudes, which both scale as $O(M_v^4 M_o^2)$ [20]:

$$\mathcal{W}_{abef} \leftarrow \frac{1}{4}\sum_{mn} \tau_{mn}^{ab} \langle mn \| ef \rangle \tag{4.22}$$

$$t_{ij}^{ab} D_{ij}^{ab} \leftarrow \frac{1}{2}\sum_{ef} \tau_{ij}^{ef} \mathcal{W}_{abef} \tag{4.23}$$

where

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_j^b t_i^a \tag{4.24}$$

$$D_{ij}^{ab} = F_{ii} + F_{jj} - F_{aa} - F_{bb} \tag{4.25}$$

and we see that we do not need the whole $\mathcal{W}_{abef}$ tensor, as we are only interested in some a, b, not all of them. Thus when only using a fraction $p < 1$ of virtual orbitals, the number of floating point operations is reduced by a constant factor of $\frac{1}{p^2}$. Similar considerations apply to other intermediates when not including all virtual or occupied orbitals. If this is a feasible approach, the approximate solution vector $(c_1, \ldots, c_L)^T_{\text{approx}}$ is extremely close to the exact one, $(c_1, \ldots, c_L)^T$, and we obtain a good approximation to $\hat{T}^*$ at reduced cost. We found that the the values $\Theta_i = \sum_{abj} \left| t_{ij}^{ab} \right|^2$, $\Theta_a = \sum_{bij} \left| t_{ij}^{ab} \right|^2$ work as a reasonable selection criterion: We include only those virtual orbitals with large Θ_a and those occupied orbitals with large Θ_i.

4.3.2 Justification of the parameter reduction approach

Assume that the AMP-CCEVC approach is exact, e.g. there is a linear combination such that it solves the CC amplitude equations

$$\langle \mu | \, e^{-\hat{T}^*} \hat{H} e^{\hat{T}^*} \, | \Phi^{\text{SD}} \rangle = 0 \quad \text{for all } \mu \tag{4.26}$$

$$\hat{T}^* = \sum_{m=1}^{L} c_m \hat{T}(\vec{\alpha}_m). \tag{4.27}$$

Then the amplitude equations are fulfilled for any arbitrary linear combinations of states, e.g.

$$\sum_{\text{some } \mu} c_\mu \, \langle \mu | \, e^{-\hat{T}^*} \hat{H} e^{\hat{T}^*} \, | \Phi^{\text{SD}} \rangle = 0 \tag{4.28}$$

for arbitrary c_μ. Thus, we see especially that both eqs. (4.18) and (4.20) will have zero projection error and thus share the same solution. This is not the case if the AMP-CCCEVC approach does not give exact CC wave function, in which case the roots differ. However, if the AMP-CCEVC approach is approximately correct, the solutions of eq. (4.20) will be very similar to those of eq. (4.18). When AMP-CCEVC solution is no good approximation to the CCSD amplitudes, we expect the parameter reduction solution to be qualitatively different.

4.4 Multi-reference EVC for chemistry

We will here describe in detail how we considered eigenvector continuation in chemistry.

4.4.1 "Naive" EVC along a trajectory

Let $\vec{R}(x)$ describe an analytic trajectory. The "naive" approach of doing eigenvector continuation along a trajectory is to solve the Schrödinger equation (approximately) at a

set of molecular geometries $\vec{r}_i$ for $i = 1, \ldots, N$, giving raise to a set of wave functions $|\Psi(\vec{r}_i)\rangle$. Then, we approximate the true wave function at a target geometry $\vec{r}_T$ as

$$|\Psi_{\mathrm{EVC}}(\vec{R}_\odot)\rangle = \sum_{i=1}^{L} c_i |\Psi(\vec{R}_i)\rangle \tag{4.29}$$

We have discussed previously why this fails. For each $|\Psi(\vec{R}_i)\rangle$, the atomic orbital basis is centered at different positions. This basis is very badly suited to describing a molecule in a different geometry $\vec{R}_\odot \neq \vec{r}_i$. This is why eq. (4.29) will be a very poor fit for the true wave function at that geometry.

4.4.2 NOCI along a trajectory

Let $\vec{R}_\odot$ be the geometry of interest. Instead of using the sample wave functions $|\Phi(\vec{R}_i)\rangle$ directly, we convert the coefficient matrix using symmetric orthonormalization (sec. 4.2.1) or use generalized Procrustes orbitals to represent the state at the new geometry. Doing this leads to a state $|\Phi_{\vec{R}_\odot}(\vec{R}_i)\rangle$, where the creation and annihilation operators of the basis set are expressed with respect to the geometry $\vec{R}_\odot$:

$$|\Psi_{\mathrm{EVC}}(\vec{R}_\odot)\rangle = \sum_{i=1}^{L} c_i |\Psi_{\vec{R}_\odot}(\vec{R}_i)\rangle \tag{4.30}$$

This can be considered as "moving" the wave function $|\Phi(\vec{R}_i)\rangle$ to a position $\vec{R}_\odot$, without performing any re-optimization of the MOs, the expansion coefficients etc. This is a NOCI approach (cf. sec. 1.10, p. 22), as two states $|\Phi_{\vec{R}_\odot}(\vec{R}_i)\rangle$, $|\Phi_{\vec{R}_\odot}(\vec{R}_j)\rangle$ will have two different sets of molecular orbitals (but be built from the same basis). A computational bottleneck arises because the molecular orbitals corresponding to $|\Phi_{\vec{R}_\odot}(\vec{R}_i)\rangle$ are non-orthogonal to the the ones belonging to $|\Phi_{\vec{R}_\odot}(\vec{R}_j)\rangle$, making this approach generally intractable in, for example, coupled cluster theory.

4.4.3 Multi-reference EVC at one geometry by tweaking the electron-electron repulsion

We can split the electronic Hamiltonian in a constant, a one-body part and a two-body part. We can create a new Hamiltonian by strengthening or weakening the electron-electron repulsion,

$$\hat{H}_{\mathrm{e}}(\lambda) = -\sum_i \frac{1}{2}\nabla_i^2 - \sum_{i,A} \frac{Z_A}{r_{iA}} + \lambda \sum_{i>j} \frac{1}{r_{ij}} + \sum_{B>A} \frac{Z_A Z_B}{R_{AB}} \tag{4.31}$$

Now, for each value of λ, we can obtain a HF-state. For $\lambda = 0$, a single Slater determinant is the exact solution (as the particles do not interact), but the single particle picture is worse for increasing λ. The approach is to select a set of $\lambda_i \in [0.5, 1.1]$, $i = 1, \ldots, L$ and obtain Slater determinants $|\Phi^{\mathrm{SD}}(\lambda_i)\rangle$ and use those to build the EVC space. It should be noted that this method is not really an EVC method, in the sense that we are not "continuing" a solution, but rather try to build a small space of non-orthogonal determinants that we hope might approximate the accurate solution.

4.5 Single-reference EVC

While the multi reference approach is more general, calculating matrix elements for non-orthogonal wave functions is expensive, and complicates the mathematics of multi-reference problems further. Here, we consider how we can transform states from one geometry to another such that a single-reference picture is kept. We exemplify this for coupled cluster wave functions, but the approach is exactly the same for configuration interaction. Assume

we have solved the CCSD equations for a system at geometry $\vec{R}_i$. Then, for example, the $\hat{T}_1$ operator is given by

$$\hat{T}_1(\vec{R}_i) = \sum_{ia} t_i^a(\vec{R}_i) a_a^\dagger(\vec{R}_i) a_i(\vec{R}_i) = \mathbf{t}(\vec{R}_i) \cdot \mathbf{a}(\vec{R}_i) \tag{4.32}$$

We then express the corresponding operator at a different geometry $\vec{R}_\odot$ as

$$\hat{T}_{1,\vec{R}_\odot}(\vec{R}_i) = \sum_{ia} t_i^a(\vec{R}_i) a_a^\dagger(\vec{R}_\odot) a_i(\vec{R}_\odot) = \mathbf{t}(\vec{R}_i) \cdot \mathbf{a}(\vec{R}_\odot) \tag{4.33}$$

where some rule was used to convert the creation/annihilation operators from some geometry to another. The same transformation can be applied to the operators $\hat{T}_2, \hat{T}_3, \ldots$ or the CI operators $\hat{C}_1, \hat{C}_2, \ldots$.
Put differently, we keep the amplitudes from an excitation $\hat{T}$, but change molecular orbitals analytically (cf. sec. 4.1.1). Then, it is likely that the CC-amplitudes/CC-expansion coefficients are analytic as well. It remains how to map the excitation and annihilation operators. This is the same as considering how to map the MOs. Here, we can use any of the methods described in section 4.2.

1. Define a reference geometry $\vec{R}_{\text{ref}}$, solve the HF-equations and obtain $\mathbf{C}(\vec{R}_{\text{ref}})$. Then, for each $\vec{R}$, solve the HF equations to obtain $\mathbf{C}(\vec{R})$. Then, follow the approach described in section 4.2.3 for the occupied and the virtual MOs individually to find the Procrustes orbitals with corresponding coefficient matrix $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R})$. As this is done on the occupied and the virtual orbitals individually, this guarantees that a Hartree-Fock state is chosen as reference determinant. The cluster operator $\hat{T}(\vec{R})$ (or the CI operator $\hat{C}(\vec{R})$) is found with respect to the new orbitals $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R})$. This can also be done with general Procrustes orbitals, which guarantees analyticity and a smaller change in the Hamiltonian, but yields non-Hartree-Fock reference states.

2. Define a reference geometry $\vec{R}_{\text{ref}}$, solve the HF-equations and obtain $\mathbf{C}(\vec{R}_{\text{ref}})$. We use this state (converted to the new geometry using symmetric orthonormalization) as reference state for finding the cluster operator. All excitations can then be found with respect to the new MOs, which will not correspond to HF states.

3. For each $\vec{R}$, use an orbital localization algorithm on the canonical orbitals. The cluster operator $\hat{T}(\vec{R})$ (or the CI operator $\hat{C}(\vec{R})$) is found with respect to the localized orbitals. The suggested localization algorithm is the use of Cholesky MOs (cf. sec. 4.2.2, p. 46) as they are found using a noniterative algorithm. This guarantees that a Hartree-Fock state is chosen as reference determinant and is independent of a reference.

4. Calculate the cluster operator with respect to natural orbitals, taking special care of the correct ordering and degenerate NOs such that the natural orbitals are analytic. This approach allows for freezing of natural orbitals that have a small occupation number along the whole geometry. It should be noted that this is in principle also possible with canonical orbitals.

Each of these approaches will give rise to a uniquely defined, transformed reference Slater determinant $|\Phi^{\text{tr.}}(\vec{R})\rangle$ for each $\vec{R}$, including the sample points $\vec{R}_i$. It is then straightforward to apply either AMP-CCEVC, WF-CCEVC or any other EVC-based method.

Frozen and inactive orbitals

In order to reduce the number of parameters, it is usual to define an active space for methods such as Full-CI (leading to CASCI [8]), but also for methods such as CC-theory, Unitary CC-theory, and similar. This procedure is applicable to eigenvector continuation too, and give exactly the same energies to be comparable. For Procrustes orbitals, this can be done by first obtaining the coefficient matrix $\mathbf{C}(\vec{R})$, and then applying the Procrustes procedure only to the active occupied orbitals and the active virtual orbitals. The frozen and inactive orbitals

remain untouched. When freezing two core electrons, for example, the first vector of the HF coefficient matrix will remain the same. For natural orbitals or canonical orbitals, the same natural orbitals need to be frozen.

4.6 Comparison of choices of MOs

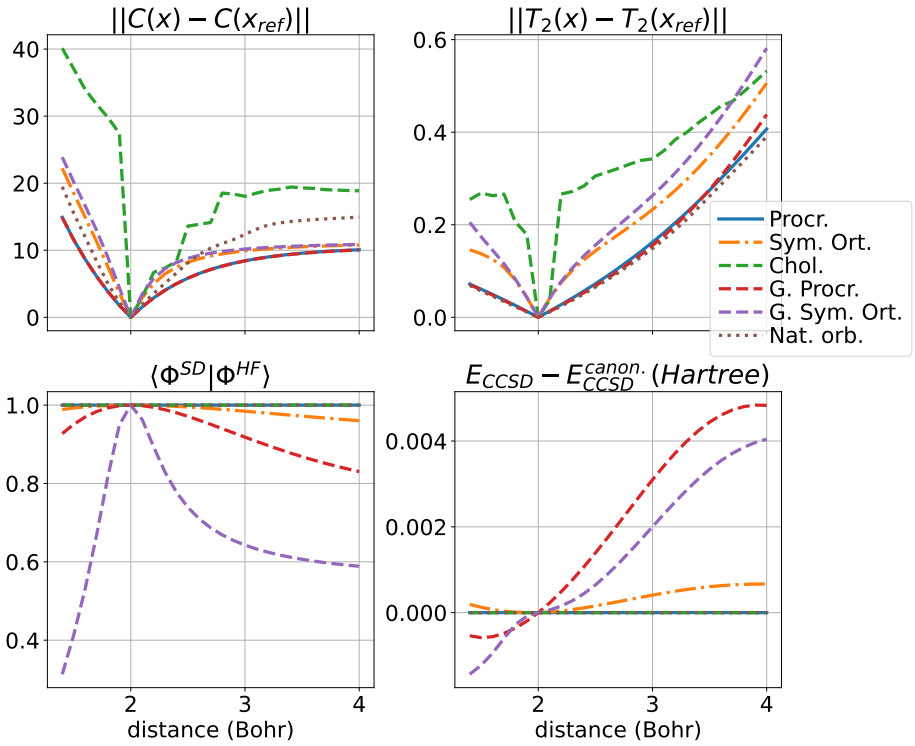


Figure 4.2: Using $x_{ref} = 2$ Bohr, the norm of the difference in the coefficient matrices (upper left), $\hat{T}_2$-amplitudes (upper right), the overlap to the HF state (lower left) and the CCSD energy compared to canonical orbitals (lower right) are depicted for the hydrogen fluoride molecule in the cc-pVTZ basis set.

Each of the methods described in sec. 4.2 have advantages and disadvantages and have different applications. Procrustes orbitals require to solve the HF-equations at each geometry of interest. To obtain natural orbitals, it is also required to acquire and diagonalize an approximate 1-RDM using, for example, Møller-Plesset perturbation theory. As the analyticity of the canonical orbitals is not guaranteed, using Procrustes or natural orbitals might fail. This can happen when orbital energies cross at a geometry $\vec{R}_c$, or even more gravely for systems such as the BeH_2 insertion reaction, where the principal configuration changes at a critical geometry $\vec{R}_c$ [83], both cases leading to a discontinuity in the spans of the occupied and virtual MOs, respectively. Both Procrustes orbitals and natural orbitals work well for Post Hartree-Fock methods, as they both have the ground state HF wave function as a reference. An additional advantage of that is that numerical methods developed which are optimized for the use of canonical orbitals can still be applied, because we can always perform unitary operations among occupied/virtual orbitals, respectively, and transform the sample wave functions (the cluster operator and the reference) , expressed in Procrustes/natural orbitals, to canonical orbitals at the geometry of interest.
Both generalized Procrustes orbitals and Procrustes orbitals require an arbitrarily chosen reference state which has no physical meaning, but they do reduce the change in the expansion coefficients, which aims at minimizing the change in the Hamiltonian. A huge advantage of Procrustes orbitals over natural orbitals and canonical orbitals, and the reason why they are our first choice for single-reference problems, is that they are extremely simple to obtain. For natural orbitals (or canonical orbitals), there is an assignment problem to be solved. For this to give the correct results, we require that two geometries $\vec{R}_1$,$\vec{R}_2$ are close.

But this means that we have to solve the HF-equations with a relatively high resolution (e.g. dense spacing) along the whole PES *before* obtaining the sample geometries in order to assign them correctly, as discussed in sec. 5.3. With Procrustes orbitals, this is not necessary.
Symmetric orthonormalization, among others, is used to obtain gradients of potential energy surfaces and also minimizes the change of MOs in a least-square sense, but does not lead to a Hartree-Fock wave function.
Finally, it should be noted that there are many other possible approaches which we did not consider. For example, the use of Procrustes orbitals weighs all orbitals equally, but implementing a different weighing is straightforward. It is also clear that an approach similar than the one we applied to natural orbitals can be used for canonical orbitals, too, which can be chosen to be analytic whenever the Hartree-Fock state is. Furthermore, we did not explicitly minimize $\left\| \hat{H}(\vec{R}_{\text{ref}}) - \hat{H}(\vec{R}) \right\|$ for some norm, which would require an iterative minimization and analyticity would not be sure to be achieved.
In figure 4.2, we compare different orbital rotation schemes for the dissociation of HF, comparing the change in the coefficient matrix wrt. to the reference, the change in the CCSD $\hat{T}_2$ amplitudes, the overlap to the Hartree-Fock wave function at that geometry and the canonical CCSD energy using this wave function as reference, compared to canonical orbitals. We see that Cholesky orbitals, albeit unique, lead to discontinuous changes in the coefficients and the $\hat{T}_2$ amplitudes and are hence insufficient choices. Procrustes orbitals, natural orbitals and general Procrustes orbitals, unlike symmetric orthonormalization, cause small changes in the coefficient matrix and the $\hat{T}_2$ amplitudes. However, symmetric orthonormalization leads to a Slater determinant that has much more overlap with the exact HF solution than general Procrustes orbitals do. Finally, we see that Procrustes orbitals, natural orbitals and Cholesky MOs lead to the same CCSD energies as using canonical orbitals, while the other choices do not. This change is larger the further the reference Slater determinant lies from the HF Slater determinant. Even though Thouless' theorem (cf. sec. 1.5, p. 14) implies that CCSD is relatively independent of the reference state, we see how too large deviations from the "good" reference nevertheless lead to quantitatively wrong results.

4.7 EVC on a quantum computer

4.7.1 Single-reference approach

On a quantum computer, the formalism for EVC becomes simpler than both WF-CCEVC and AMP-CCEVC. As the EVC-matrix is Hermitian, we can again use lemma 2.2.1. Thus, including extra sampling points will only lead to improvements in the expectation value of the energy. The principal idea follows the same steps as before:

1. For a set of sample geometries $\vec{R}_i$, $i = 1, \ldots, L$, obtain approximate ground state wave functions $|\Psi_i\rangle = \hat{U}_i |\Phi^{\text{SD}}(\vec{R}_i)\rangle$, where $\hat{U}_i$ is a unitary operator obtained, for example, using one of the ansätze in sec. 3.5 and $|\Phi^{\text{SD}}(\vec{R}_i)\rangle$ is created using any of the methods described in sec. 4.2.

2. For each $\vec{R}_\odot$, calculate the $L \times L$ matrices $\mathbf{H}^{\text{EVC}}$, $\mathbf{S}^{\text{EVC}}$ with elements

$$H_{ij}^{\text{EVC}}(\vec{R}_\odot) = \langle \Psi_{\vec{R}_\odot}(\vec{R}_i) | \hat{H}(\vec{R}_\odot) | \Psi_{\vec{R}_\odot}(\vec{R}_j) \rangle \tag{4.34}$$

$$S_{ij}^{\text{EVC}}(\vec{R}_\odot) = \langle \Psi_{\vec{R}_\odot}(\vec{R}_i) | \Psi_{\vec{R}_\odot}(\vec{R}_j) \rangle \tag{4.35}$$

   where $\Psi_{\vec{R}_\odot}(\vec{R}_i) = U_i |\Phi^{\text{SD}}(\vec{R}_\odot)\rangle$, and find the approximate ground state. It should be noted that using Procrustes and generalized Procrustes orbitals, the states $|\Psi_{\vec{R}_\odot}(\vec{R}_j)\rangle$ are independent of $\vec{R}_\odot$.

In a second quantized picture, the only explicit reference to $\vec{R}_\odot$ is hence in the Hamiltonian $H_{ij}^{\text{EVC}}(\vec{R}_\odot) = \langle \Psi(\vec{R}_i)| \hat{H}(\vec{R}_\odot) |\Psi(\vec{R}_j)\rangle$. Each Hamiltonian is a sum of Pauli strings

$$\hat{H}(\vec{R}) = \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q}(\vec{R}) \tilde{\sigma}^{i_1} \tilde{\sigma}^{i_2} \ldots \tilde{\sigma}^{i_Q} \tag{4.36}$$

where the dependence on the geometry is in the coefficients only, corresponding to eq. (4.2). Independently of the geometry, the Hamiltonians will have the same set of coefficients $c_{i_1,i_2,\ldots i_Q}$ that are nonzero.[1] For that reason, when calculating $\mathbf{H}^{\text{EVC}}(\vec{R})$, we can re-use the same expectation values $\langle \tilde{\sigma}^{i_1} \tilde{\sigma}^{i_2} \ldots \tilde{\sigma}^{i_Q} \rangle_{ij}$ as they are the same for each $\vec{R}$. Thus, when calculating a potential energy surface with $k > L$ points of interest, we only need to sample $O(L^2M^4)$ expectation values. In addition, we only need to optimize L sets of parameters, which are costly optimizations.

4.7.2 Sampling Error

An extra complication on quantum computers is that there is a sampling error to take into consideration. As matrix elements need to be sampled, the matrices $\mathbf{H}^{\text{EVC}}$, $\mathbf{S}^{\text{EVC}}$ will be random matrices, and thus, the eigenvectors and eigenvalues will be random variables and it is necessary to do a statistical analysis of the problem. We follow an approach similar to the one implemented by Huggins et al. [67] to quantify the uncertainty in the ground state energy. Let $\tilde{h}_{ij}, \tilde{s}_{ij}$ be the sample means. Let $\tilde{\sigma}^2_{S_{ij}}$ be the sample variance, while we used eq. (3.23) for $\tilde{\sigma}^2_{H_{ij}}$. Observe that this is an overestimation. We then treat H_{ij}, S_{ij} as random variables, which we approximate as normal distributions

$$\begin{aligned} H_{i,j} &\sim \mathcal{N}\left(\tilde{h}_{ij}, \tilde{\sigma}_{H_{ij}} / \sqrt{N_{H_{ij}}}\right) \\ S_{i,j} &\sim \mathcal{N}\left(\tilde{s}_{ij}, \tilde{\sigma}_{S_{ij}} \sqrt{N_{S_{ij}}}\right) \end{aligned} \tag{4.37}$$

where $N_{H_{ij}}, N_{S_{ij}}$ are the number of measurements of the respective elements. giving rise to matrices $\tilde{\mathbf{H}}^{\text{EVC}}$, $\tilde{\mathbf{S}}^{\text{EVC}}$. Then, we will use the the Bootstrapping method [84] to diagonalize N_B matrices $\tilde{\mathbf{H}}_i^{\text{EVC}}$, $\tilde{\mathbf{S}}_i^{\text{EVC}}$, $i = 1, \ldots, N_B$ drawn from the distribution (4.37) to obtain a guess of the eigenvalue $\tilde{E}_0$ and the uncertainty $\tilde{\sigma}_{E_0}$. It should be noted that this approach will lead to biased estimators for the eigenvalues, as $\tilde{h}_{ij}$, $\tilde{s}_{ij}$ are only estimates of the true mean. Finally, we also follow an approach similar to the adaptive protocol by Huggins et. al.. As long as $\tilde{\sigma}_{E_0} > x$ for some threshold x, we need to increase the number of measurements to reduce that uncertainty. We then choose to measure the element $S_{i,j}$, $H_{i,j}$ where increasing the number of measurements decreases the uncertainty $\tilde{\sigma}_{E_0}$ the most. For each k, l, all elements follow the distribution (4.37), except for

$$H_{k,l} \sim \mathcal{N}\left(\tilde{h}_{k,l}, \tilde{\sigma}_{H_{k,l}} / \sqrt{N_{H_{ij}} + N^+}\right) \tag{4.38}$$

or similarly for $S_{k,l}$, where N^+ is the number of extra measurements to be performed. This gives rise to to $\tilde{\sigma}^{H/S}_{E_0}(k,l)$. We then pick the values k, l that minimize $\tilde{\sigma}^{H/S}_{E_0}(k,l)$. This procedure does not require quantum measurements and can be considered as a way to decrease uncertainty without increasing accuracy (which is achieved by doing quantum measurements) - we hence choose the element where increasing accuracy decreases uncertainty in the eigenvalue $\tilde{\sigma}_{E_0}$ the most. This can be done iteratively until $\tilde{\sigma}_{E_0} < x$.

4.7.3 Solving multi-reference problems on quantum computers

From the discussion in the previous sections, we know that using generalized Procrustes orbitals with respect to a reference geometry $\vec{R}_{\text{ref}}$ leads to a coefficient matrix that changes

[1] Some coefficients $c_{i_1,i_2,\ldots i_Q}(\vec{R})$ may of course be coincidentally zero at $\vec{R} = \vec{R}_i$, but not at $\vec{R}_j$. But as the $O(M^4)$ creation and annihilation operators in the Hamiltonian are independent of geometry and mapped on the same set of Pauli strings, the same $O(M^4)$ coefficients will be nonzero at all geometries.

as little as possible in Frobenius norm, which leads to a Hamiltonian that changes little. Furthermore, general Procrustes orbitals do not require the HF equations to be solved and are analytic as function of the geometric perturbation. Let $\hat{K}(\vec{R}_i)\,|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle$ be a wave function at a geometry $\vec{R}_i$. $|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle$ is a Slater determinant with coefficient matrix $\mathbf{C}(\vec{R}_i)$, and $\hat{K}(\vec{R}_i) \in \{\hat{C}(\vec{R}_i), e^{\hat{T}(\vec{R}_i)}, \hat{U}(\vec{R}_i)\}$ is a post-HF ansatz, either unitary or not (though unitary on a quantum computer). Let $\mathbf{C}(\vec{R}_{\mathrm{ref}})$ be any coefficient matrix a geometry $\vec{R}_{\mathrm{ref}}$. We define the *best orbital fit* of that wave function at geometry $\vec{R}_{\mathrm{ref}}$ as

$$|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)\rangle = \hat{U}_i\hat{K}(\vec{R}_i)\,|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle \tag{4.39}$$

where $\hat{U}_i = e^{-\hat{\kappa}_i}$ is a change of basis operator defined from the matrix $\mathbf{U}_i$ as described in section 1.4, where $\mathbf{U}_i$ is obtained by solving the generalized Procrustes problem

$$\mathbf{U}_i = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}_{\mathrm{ref}}) - \mathbf{C}(\vec{R}_i)\mathbf{\Omega} \right\| \tag{4.40}$$

We are hence finding the unitary transformation applied at geometry $\vec{R}_i$ that makes the resulting molecular orbitals closest to the ones that were used at geometry $\vec{R}_{\mathrm{ref}}$. Unlike the non-general Procrustes approach, the resulting wave function $|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)\rangle$ will have a more complicated structure than $|\Psi(\vec{R}_i)\rangle$: If, for example, $|\Psi(\vec{R}_i)\rangle$ is a Slater determinant in its reference basis, then $|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)\rangle$ will generally not be a Slater determinant in the new basis, as we are mixing occupied and unoccupied orbitals. On a regular computer, this is an exclusion criterion in the sense that there are no efficient ways to calculate the overlap and matrix elements, which we will show two paragraphs below. On quantum computers, however, this basis change operation is cheap and readily implemented [85]. It should however be noted that standard qubit tapering is not possible with this method, requiring the Jordan-Wigner mapping without qubit tapering. Freezing core electrons and selecting an active space is still possible however. This leads to the following algorithm:

1. For $i = 1, \ldots, L$, find approximate ground states $|\Psi(\vec{R}_i)\rangle = K(\vec{R}_i)\,|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle$. The use of Procrustes orbitals for the reference determinant $|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle$ is possible whenever desired, but using e.g. canonical orbitals works too.
2. Define an (arbitrary) reference geometry $\vec{R}_{\mathrm{ref}}$.
3. For $i = 1, \ldots, L$, find the states $|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)\rangle$.
4. At the desired geometry $\vec{R}_\odot$, express the Hamiltonian $\hat{H}_{\vec{R}_{\mathrm{ref}}}(\vec{R}_\odot)$ in the basis defined by the coefficient matrix $\mathbf{C}(\vec{R}_\odot)\mathbf{U}_\odot$.
5. Find the matrix elements
$$H_{ij} = \langle \Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)|\hat{H}_{\vec{R}_{\mathrm{ref}}}(\vec{R}_\odot)|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_j)\rangle$$
$$S_{ij} = \langle \Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_i)|\Psi_{\vec{R}_{\mathrm{ref}}}(\vec{R}_j)\rangle\,.$$
6. Solve the EVC equations.

The arbitrary choice of the MOs to be used in step 1 is a clear advantage. The method can be applied to systems where the reference changes discontinuously, unlike normal Procrustes orbitals. We perform a unitary transform of the wave function *after* the application of a post-HF ansatz into a basis where the Hamiltonian changes continuously.
One might argue why the use of a reference geometry even is necessary. Indeed, not using a reference has a conceptual advantage. For every geometry of interest, we could instead directly transform to the geometry of interest, that is, build a state

$$|\Psi_{\vec{R}_\odot}(\vec{R}_i)\rangle = \hat{U}_{i\to\odot}\hat{K}(\vec{R}_i)\,|\Phi^{\mathrm{SD}}(\vec{R}_i)\rangle \tag{4.41}$$

with

$$\mathbf{U}_{i\to\odot} = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}_\odot) - \mathbf{C}(\vec{R}_i)\mathbf{\Omega} \right\|. \tag{4.42}$$

This approach destroys the advantages of eigenvector continuation on quantum computers. Using $\hat{U}_i$ with a fixed reference, the overlap matrix $\mathbf{S}^{\text{EVC}}$ is independent of $\vec{R}_\odot$, and as previously, sampling only needs to be performed once. Using $|\Psi_{\vec{R}_\odot}(\vec{R}_i)\rangle$, $\mathbf{S}^{\text{EVC}}$ will differ at each each geometry, hence the overlap matrix and Pauli matrix expectation values become state-dependent and need to be resampled each time.
We will now show why this method is not applicable on a regular computer. Let $A = \{|\phi_p^a\rangle\}_{p=1}^M, B = \{|\phi_p^b\rangle\}_{p=1}^M$ and $C = \{|\phi_p^c\rangle\}_{p=1}^M$ be different sets of MOs build from the same set of AOs, with the creation operators defined correspondingly, with the corresponding creation and annihilaton operators denoted as a_p, b_p, c_p and $a_p^\dagger, b_p^\dagger, c_p^\dagger$. Let $|\Psi_i^a\rangle$, $|\Psi_j^b\rangle$ be linear combination of ON-vectors expressed in terms of $a_p, a_p^\dagger$ and $b_p, b_p^\dagger$ respectively, with a Hamiltonian $\hat{H}_c$ in second quantization expressed in terms of $c_p, c_p^\dagger$. When we want to calculate

$$H_{ij} = \langle \Psi_i^a | \, \hat{H}_c \, | \Psi_j^b \rangle \,, \tag{4.43}$$

we need to express $|\Psi_i^a\rangle, |\Psi_j^b\rangle$ in terms of MOs C by performing a unitary transformation on $|\Psi_i^a\rangle$, $|\Psi_i^b\rangle$ as described in section 1.4. This leaves the state unchanged, we write $|\Psi_i^a\rangle = \exp(-\hat{\kappa}_a)\,|\Psi_i^c\rangle$ and similarly for $|\Psi_i^b\rangle = \exp(-\hat{\kappa}_b)\,|\Psi_i^c\rangle$. Now

$$H_{ij} = \langle \Psi_i^c | \exp(\hat{\kappa}_b) \hat{H}_c \exp(-\hat{\kappa}_a) \, |\Psi_j^c\rangle \tag{4.44}$$

applying the operations $\exp(-\hat{\kappa})$ to the wave functions is out of question, as this, in general, leads to a wave function with no simple form. It is this application that is cheap on the quantum computer. Application to the Hamiltonian is too complicated, too: if $\hat{\kappa}_b$ were equal to $\hat{\kappa}_a$, this would correspond to a change of basis operation of the Hamiltonian. However, as this is not the case, there is no simple expression.

4.7.4 Failed ideas to make overlap matrix approximately diagonal

In the result section, we will see how very small eigenvalues in the overlap matrix $\mathbf{S}$ will give rise to an increased sampling need. We consider here some ideas we had to make the overlap matrix approximately diagonal. As the diagonal elements will necessarily be equal to one, $S_{ii}^{\text{EVC}} = 1$, we can write $\mathbf{S}^{\text{EVC}} = \boldsymbol{I} - \mathbf{\Theta}$ and apply the Neumann series where $\mathbf{\Theta}$ fulfills $\|\mathbf{\Theta}\| < 1$ in the operator norm:

$$\left(\boldsymbol{S}^{\text{EVC}}\right)^{-1} = (\boldsymbol{I} - \mathbf{\Theta})^{-1} = \sum_{i=0}^{\infty} \mathbf{\Theta}^i \tag{4.45}$$

from this, we see that by making the off-diagonal elements Θ_{ij} "small", we obtain a matrix inverse with elements that do not "blow up" and stay close to identity, decreasing the sampling need.

These methods ultimately turned out not to work, the first one because it is too expensive on a quantum computer, the second one because it is not exact.

Overlap reduction with LCU

As the span of a set of linear independent vectors is the same for almost any linear combinations between these vectors, we might instead consider not using the sample vectors directly in the EVC procedure, but linear combinations between them. For example, when the EVC-space only consists of two vectors, we might use $|\Psi_2'\rangle = \frac{|\Psi_2\rangle - c|\Psi_1\rangle}{\||\Psi_2\rangle - c|\Psi_1\rangle\|}$ where ideally, c is chosen such that $\langle \Psi_2' | \Psi_1 \rangle$ is small. For example, if we know that the overlap is large $|\langle \Psi_1 | \Psi_2 \rangle| > 0.98$, then $c = 1$ is a reasonable choice, and a (re-normalized) $|\Psi_2'\rangle$ will have less than 0.1 overlap with $|\Psi_1\rangle$. It is important that $|\Psi_2'\rangle$ is renormalized, otherwise, the eigenvalues of the new overlap matrix $\mathbf{S}'$ would correspond to those of $\mathbf{S}$. Based on LCU

(cf. sec. 3.12, p. 41), we developed a method to calculate matrix elements of arbitrary linear combinations of wave functions. Let $|\Psi_i'\rangle$, $|\Psi_j'\rangle$ be linear combinations of unitaries. Then, one can use the LCU algorithm to implement the state

$$|\Omega\rangle = \frac{1}{N}\left(\frac{1}{p_i}|0^n\rangle|0\rangle|\Psi_i'\rangle + \frac{1}{p_j}|0^n\rangle|1\rangle|\Psi_j'\rangle\right) \tag{4.46}$$

where n is the number of ancillas necessary for the LCU algorithm, N is a normalization constant and $1/p_i$, $1/p_j$ are success probabilities for the implementation of those states. From eq. (4.46), we can then measure $I^{\otimes n} \otimes \sigma^+ \otimes \hat{O}$ for some Hermitian $\hat{O}$ and obtain the desired matrix elements. However, this approach is prohibitively expensive. In order to implement the state $|\Omega\rangle$ by means of LCU, one needs to implement unitary operations controlled by at least two qubits. In the case of a σ^X-gate, this corresponds to a Toffoli gate, which requires at least 6 CNOT-gates [86], which are considered expensive. Furthermore, the success probability of the individual measurements will be very small, as the (unnormalized) state will have a very small norm when c is large. Even though there exist methods such as oblivious amplitude amplification to increase the success probability [87], they require further expensive operations. Finally, as the overlap is not exactly known, the success probabilities need to be sampled as well, which makes the success of this approach questionable. To conclude, although it is possible to create an algorithm that can make the overlap matrix diagonally dominant based on LCU, the resource requirements are not viable on NISQ devices.

Overlap reduction with excited states

Instead of finding a linear combination of the sample vectors that approximately diagonalize the overlap matrix, we might also consider finding different sampling vectors with the same span. In order to do this, we adapt the approach described in section 3.11. For a set of sample geometries $\vec{R}_1, \ldots, \vec{R}_L$, we find the kth sample state $|\Psi(\vec{R}_k)\rangle$ using the effective Hamiltonian $\hat{H}_{k-1}$ such that the resulting ground state has "little" overlap with the states $|\Psi(\vec{R}_i)\rangle$ where $i \in \{1, \ldots, k-1\}$. As the states $|\Psi(\vec{R}_i)\rangle$ will likely not be ground states at geometry $\vec{R}_k$, we will not get excited states, but low-lying energy states that are near-orthogonal to $|\Psi(\vec{R}_i)\rangle$. However, this method fails, and we will now justify with an example why it does so.

Assume a symmetric $3x3$ matrix $\mathbf{M}(\alpha)$. Let $\mathbf{v}_0(0)$ be the ground state at $\alpha = 0$. At $\alpha = x$, we can write $\mathbf{v}_0(0) = c_0\mathbf{v}_0(x) + c_1\mathbf{v}_1(x) + c_2\mathbf{v}_2(x)$, where $\mathbf{v}_i(x)$ are eigenvectors of $\mathbf{M}(x)$, with eigenvalues $E_0 < E_1 < E_2$. Using the proposed algorithm to create a state $\tilde{\mathbf{v}}(x)$ which is orthogonal to $\mathbf{v}_0(0)$ will in general give rise to a state $\tilde{\mathbf{v}}(x) = c_0'\mathbf{v}_0(x) + c_1'\mathbf{v}_1(x) + c_2'\mathbf{v}_2(x)$. The problem is that it is in general impossible to create a linear combination $a\tilde{\mathbf{v}}(x) + b\mathbf{v}_0(0) = \mathbf{v}_0(x)$, e.g. diagonalizing the EVC matrix spanned by the vectors $\tilde{\mathbf{v}}(x)$ and $\mathbf{v}_0(0)$ will yield an energy $E > E_0$, which makes the method fail. This is rooted in the fact that the EVC space when used for extrapolation has the same span as the n first derivatives of $\mathbf{v}_0(0)$ wrt. α, while this algorithm approximately produces $\text{span}\{\mathbf{v}_0(0), \ldots, \mathbf{v}_n(0)\}$, not $\text{span}\{\mathbf{v}_0(0), \ldots, \mathbf{v}_0^{(n)}(0)\}$. Using perturbation theory, it is possible to show that the first excited state usually has a large contribution to the derivative $\mathbf{v}_0^{(n)}(0)$ [8, ch. 14], but this is not sufficient to do proper eigenvector continuation.

Chapter 5

Numerical implementation and algorithms

In this chapter, we will explain how many of the methods described in the previous chapters were implemented. We will describe numerical details of some algorithms. We also describe how some problems can be solved practically, especially when no simple implementation exists in existing programming libraries. Furthermore, we will discuss our code implementation, which libraries and which functionalities were used, and where to find our code.

5.1 Numerical methods

5.1.1 The Moore-Penrose Pseudoinverse

Let $\mathbf{S}$ be a matrix with small singular values, that is, some elements of the SVD of $\mathbf{S}$ are small

$$\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger \tag{5.1}$$

When inverting $\mathbf{S}$, we see that its inverse

$$\mathbf{S}^{-1} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^\dagger \tag{5.2}$$

will have very large singular values and hence eigenvalues and be unstable with respect to small deviations in the smallest eigenvalues. From a computational point of view [5], one can consider some threshold ϵ, and remove the smallest eigenvalues

$$\tilde{\Sigma}_{ii} = \begin{cases} \Sigma_{ii} & \text{if } \Sigma_{ii} > \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

and use instead

$$\tilde{\Sigma}^+_{ii} = \begin{cases} 1/\tilde{\Sigma}_{ii} & \text{if } \tilde{\Sigma}_{ii} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

which leads to a more numerically stable inverse

$$\mathbf{S}^{-1} \approx \mathbf{S}^+ = \mathbf{V}\tilde{\boldsymbol{\Sigma}}^+\mathbf{U}^\dagger \tag{5.5}$$

where the parameter ϵ is a tunable parameter. $\mathbf{S}^+$ is called *the Moore-Penrose Pseudoinverse* of $\mathbf{S}$ [88]. Removing the smallest eigenvalues of $\mathbf{S}$ leads generally to a very small change in the Frobenius norm $\left\|\mathbf{S} - \tilde{\mathbf{S}}\right\|_F^2 = \sum_{i=1}^{R} \Sigma_{ii}^2$ where R is the number of removed eigenvalues and we sum over the smallest singular values, which justifies this procedure.

5.1.2 The Hermitian generalized eigenvalue problem: Canonical orthonormalization

The Hermitian eigenvalue problem is given by

$$\mathbf{HC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{5.6}$$

where $\mathbf{H}, \mathbf{S}$ are Hermitian matrices. The Hermitian generalized eigenvalue problem can be solved by canonical orthogonalization [5, 10, 31]. As $\mathbf{S}$ is Hermitian, it can be diagonalized as

$$\mathbf{S} = \mathbf{U}\mathbf{s}\mathbf{U}^\dagger \tag{5.7}$$

$\mathbf{s}$ might have some very small diagonal elements, corresponding to small eigenvalues of $\mathbf{S}$ which might lead to round-off errors. We can define a threshold ϵ and use the Moore-Penrose pseudoinverse $\mathbf{s}^+$ instead of $\mathbf{s}^{-1}$. Letting $\mathbf{X} = \mathbf{U}\sqrt{\mathbf{s}^+}$ and observing that $\mathbf{X}^\dagger\mathbf{S}\mathbf{X} = \mathbf{I}$ (with zeros, instead of ones, at those indices corresponding to the removed eigenvalues), we can instead solve the standard eigenvalue problem

$$\begin{aligned} \mathbf{HC} &= \mathbf{SC}\boldsymbol{\epsilon} \\ \mathbf{HXC}' &= \mathbf{SXC}'\boldsymbol{\epsilon} \\ \left(\mathbf{X}^\dagger\mathbf{HX}\right)\mathbf{C}' &= \mathbf{X}^\dagger\mathbf{SXC}'\boldsymbol{\epsilon} \\ \mathbf{H}'\mathbf{C}' &= \mathbf{C}'\boldsymbol{\epsilon} \end{aligned} \tag{5.8}$$

where we used $\mathbf{XC}' = \mathbf{C}$ and $\mathbf{X}^\dagger\mathbf{HX} = \mathbf{H}'$. It should be noted that there exists an algorithm due to Fix and Heiberger [89] specifically developed to solve a symmetric generalized eigenvalue problem, but we will still resort to canonical orthogonalization, which is the standard algorithm for those types of problems [90].

5.1.3 The non-Hermitian generalized eigenvalue problem

The non-Hermitian generalized eigenvalue problem is given by

$$\mathbf{HC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{5.9}$$

where $\mathbf{H}, \mathbf{S}$ are not Hermitian matrices. This is known to be a difficult problem [32], we will here present two ways to solve it. If $\mathbf{S}$ is invertible, the problem can in principle be solved by diagonalizing $\mathbf{H}' = \mathbf{S}^{-1}\mathbf{H}$, or when $\mathbf{S}$ has very small singular values, we can use the Moore-Penrose pseudoinverse $\mathbf{S}^{-1} \approx \mathbf{S}^+$. It should be noted that, there exist more advanced methods for those type of problems, such as the generalized Schur decomposition [91, chap. 7.7], which is less prone to numerical errors, but may fail for ill-conditioned problems. With the `GUPTRI` algorithm [92], there exists an algorithm which generalizes the generalized Schur decomposition and finds the regular eigenvalues and is hence applicable to ill-conditioned problems. However, it requires several parameters that require substantial knowledge of the underlying matrix structure. Furthermore, we found that it overestimated the eigenvalues. A possible alternative, which we will use here, is to instead invert $\tilde{\mathbf{S}} = \mathbf{S} + \epsilon\exp\{\mathbf{S}/\epsilon\}$ for a small constant $\epsilon \in [10^{-5}, 10^{-14}]$. e.g. diagonalizing $\tilde{\mathbf{S}}^{-1}\mathbf{H}$ instead. If $\mathbf{S}$ has large complex eigenvalues, we might also set $\tilde{\mathbf{S}} = \mathbf{S} + \epsilon\mathbb{1}$ [93]. We have found that WF-CCEVC is extremely sensitive to the choice of method. Generally, we found that using the Moore-Penrose pseudoinverse of $\mathbf{S}$ or the Guptri algorithm, the eigenvalues were overestimated. We hence opted for the method to use $\tilde{\mathbf{S}}$. We found that the correct choice of the parameter ϵ the number at which the eigenvalues become unstable or nonsencial, depend heavily on the problem, and larger numbers give higher eigenvalues. The same is also true for the cutoff in the Pseudoinverse: Usually, 10^{-8} is the smallest valid cutoff, but for some problems, we had to go down to as much as 10^{-6}.

5.2 Calculation of matrix elements between non-orthogonal Slater determinants

We have previously mentioned the Generalized Slater-Condon rules, which can be used to calculate matrix elements between Slater determinants in different bases. These formulas are correct and fully general, but have an inconvenient scaling. This can be avoided using *biorthogonalization*. Let $\{v_i\}_{i=1}^{M}$ and $\{u_i\}_{i=1}^{M}$ be sets of vectors. They are said to be biorthogonal if

$$\langle v_i, u_j \rangle = \delta_{ij} \tag{5.10}$$

for some inner product $\langle \cdot, \cdot \rangle$. This concept can be used to diagonalize the overlap matrix S, turning the first- and second- order cofactor matrices $\mathbf{S}^{(i,j)}$ and $\mathbf{S}^{(ij,kl)}$ into diagonal matrices. Computing of matrix elements without biorthogonalization has been considered previously in Refs.[94] and [95], but scales less conveniently. The biorthogonal approach is due to Thom et al. [96]. We will use first quantization to describe the necessary calculations. Consider two non-orthgonal Slater determinants for $|{}^{x}\Psi^{\mathrm{SD}}\rangle$ and $|{}^{w}\Phi^{\mathrm{SD}}\rangle$, respectively, consisting of a set of molecular orbitals $|{}^{x}\phi_i\rangle$ and $|{}^{w}\phi_i\rangle$, respectively.

We have already defined the overlap matrix $\langle {}^{w}\Phi^{\mathrm{SD}} | {}^{x}\Phi^{\mathrm{SD}} \rangle = |{}^{wx}\mathbf{S}|$, with ${}^{wx}\mathbf{S}_{ij} = \langle {}^{w}\phi_i | {}^{x}\phi_j \rangle = \left({}^{w}\mathbf{C}^{T} \times \mathbf{S}_{bas} \times {}^{x}\mathbf{C} \right)_{ij}$ where $\mathbf{C}$ is the coefficient matrix and $\mathbf{S}_{bas}$ is the overlap between the atomic orbitals. Using a variant of the singular value decomposition (SVD) [81], we can diagonalize the overlap matrix as

$${}^{wx}\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger} \tag{5.11}$$

where $\mathbf{U}$, $\mathbf{V}$ are *special unitary matrices* with determinant $+1$ and $\mathbf{D}$ is diagonal. This gives

$$\left(\mathbf{U}^{\dagger} {}^{w}\mathbf{C}^{T} \right) \times \mathbf{S}_{bas} \times \left({}^{x}\mathbf{C}\mathbf{V} \right) = \mathbf{D}. \tag{5.12}$$

We have now found a way to transform the coefficient matrices in such a way that the overlap matrix is diagonal. This simplifies the calculations tremendously. As we used a singular value decomposition, $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices, hence the Slater-determinants $|{}^{x}\Phi^{\mathrm{SD}}\rangle$ and $|{}^{w}\Phi^{\mathrm{SD}}\rangle$ will still represent the same state, and the molecular orbitals remain orthogonal. To use these formulas, it is useful to define the overlap and the reduced overlap

$$S = \prod_{i} D_{ii} \tag{5.13}$$

$$\tilde{S} = \prod_{\substack{i \\ D_{ii} \neq 0}} D_{ii} \tag{5.14}$$

$$\tag{5.15}$$

Depending on the number of zeros along the diagonal of $\mathbf{D}$, the expression for the Hamiltonian matrix changes. We find that

1. No zeros:

$$\begin{aligned} \langle {}^{w}\Phi^{\mathrm{SD}} | {}^{x}\Phi^{\mathrm{SD}} \rangle &= S = \tilde{S} \\ \langle {}^{w}\Phi^{\mathrm{SD}} | \hat{H} | {}^{x}\Phi^{\mathrm{SD}} \rangle &= \tilde{S} \left(\sum_{i<j} \frac{\left(\langle {}^{w}\phi_i {}^{w}\phi_j \| {}^{x}\phi_i {}^{x}\phi_j \rangle \right)}{\langle {}^{w}\phi_i | {}^{x}\phi_i \rangle \langle {}^{w}\phi_j | {}^{x}\phi_j \rangle} + \sum_{i} \frac{\langle {}^{w}\phi_i | \hat{h} | {}^{x}\phi_i \rangle}{\langle {}^{w}\phi_i | {}^{x}\phi_i \rangle} + h_{nuc} \right) \end{aligned}$$

2. One zero, $D_{ii} = 0$:

$$\begin{aligned} \langle {}^{w}\Phi^{\mathrm{SD}} | {}^{x}\Phi^{\mathrm{SD}} \rangle &= S = 0 \\ \langle {}^{w}\Phi^{\mathrm{SD}} | \hat{H} | {}^{x}\Phi^{\mathrm{SD}} \rangle &= \tilde{S} \left(\sum_{j} \frac{\left(\langle {}^{w}\phi_i {}^{w}\phi_j \| {}^{x}\phi_i {}^{x}\phi_j \rangle \right)}{\langle {}^{w}\phi_j | {}^{x}\phi_j \rangle} + \langle {}^{w}\phi_i | \hat{h} | {}^{x}\phi_i \rangle \right) \end{aligned}$$

3. Two zeros, $D_{ii} = D_{jj} = 0$:

$$\langle {}^w\Phi^{\mathrm{SD}} | {}^x\Phi^{\mathrm{SD}} \rangle = S = 0$$
$$\langle {}^w\Phi^{\mathrm{SD}} | \hat{H} | {}^x\Phi^{\mathrm{SD}} \rangle = \tilde{S} \langle {}^w\phi_i {}^w\phi_j \| {}^x\phi_i {}^x\phi_j \rangle$$

4. Three or more zeros:

$$\langle {}^w\Phi^{\mathrm{SD}} | {}^x\Phi^{\mathrm{SD}} \rangle = S = 0$$
$$\langle {}^w\Phi^{\mathrm{SD}} | \hat{H} | {}^x\Phi^{\mathrm{SD}} \rangle = 0$$

These formulas follow directly from the Generalized Slater-Condon rules derived previously and deriving them is straightforward. In Ref. [96], an integral-direct matrix formulation of these equations which also takes into account that RHF/UHF MOs have definite spin, can be found. This is what was implemented by us in practice. The most expensive calculation is the SVD, scaling as $O(N^3)$.

5.3 Ascertaining the correct eigenvalue trajectories

Let $\mathbf{C}_1, \mathbf{C}_2$ be coefficient matrices corresponding to natural orbitals at geometries $\vec{R}_1, \vec{R}_2$, with corresponding overlap matrices $\mathbf{S}_1$ and $\mathbf{S}_2$. Because we know that $\mathbf{C}_i^\dagger \mathbf{S}_i \mathbf{C}_i = \mathbb{1}$ $(i = 1, 2)$ [5], we can consider the matrix

$$\mathbf{O}(\vec{R}_1, \vec{R}_2) = \mathbf{C}_1^\dagger \mathbf{S}_1^{1/2} \mathbf{S}_2^{1/2} \mathbf{C}_2 \tag{5.16}$$

which can be seen as an overlap matrix between the NOs at different geometries. If $\vec{R}_1$ is "close" to $\vec{R}_2$ and no crossings occur and no NOs are degenerate, then $\mathbf{O}(\vec{R}_1, \vec{R}_2)$ will have values close to ± 1 along the diagonal, which follows the orthogonality and the analyticity of the NOs: For an analytical set of orthonormal vectors $\vec{v}_i(\vec{R})$, we have that, assuming real NOs

$$\vec{v}_i(\vec{R})^T \vec{v}_j(\vec{R} + \delta\vec{R}) \approx \delta_{ij} \tag{5.17}$$

If crossings occur, then we will have that some off-diagonal elements will be close to ± 1, with some diagonal elements close to zero. But this tells us exactly that a crossing has occurred, and which NOs have crossed, making it possible to keep track of the analytical NOs. This comes at a price, because if $\vec{R}_1$ is "far" from $\vec{R}_2$, the assignment might fail. Thus, we also need to construct the NOs at some intermediate geometries

$$\vec{R}_i' = \vec{R}_1 + \frac{i}{K}(\vec{R}_2 - \vec{R}_1) \tag{5.18}$$

for some $K > 1$ and $i = 1, \ldots, K-1$. Finally, we need to consider what happens when some NOs are degenerate for all geometries $\vec{R}$, which might happen due to symmetry. One option is to symmetry adapt them. Alternatively, we can use the fact that if they have the same natural occupation number, any (normalized) linear combination between them will also be degenerate. We can hence use the orthogonal Procrustes algorithm to make them as similar as possible to some reference, which is chosen arbitrarily, to guarantee analyticity.

5.4 Implementation of AMP-CCEVC

When discussing the AMP-CCEVC and its reduced form, we did not talk about its practical implementation. As we do not use canonical orbitals, the quasi-Newton approach with diagonal, constant Jacobian (eq. (1.73), p. 17) is less justified. However, we did still used this approximation, using F_{pp} instead of ϵ_p (which are only identical in the case that the Fock matrix is diagonal). Because the amplitudes at close geometries tend to be very similar, we used SVD to obtain a set of orthogonal amplitude vectors $\mathbf{t}_m$ that span the same space, which ascertains that the coefficients c_m, $m = 1, \ldots, L$ do not grow arbitrarily large, even

though any other orthogonalization scheme could be used, too. We also implemented DIIS [97, 98], which has negligible cost, but accelerated convergence by orders of magnitude. For the parameter reduction approach, almost no adaptions needed to be made, except for removing all unused amplitudes from the calculation of the approximate Jacobian as well as some changes to the CCSD amplitude equations in order to ascertain correct summation.

5.5 Code implementation

All code is available on Github.[1] All code was written in Python, which offers a variety of libraries for numerical analysis, scientific computing in general and computational chemistry in particular. Numerical algorithms were imported from the standard libraries `NumPy` [99] and `SciPy` [100]. Methods not existing in those libraries were written from scratch, often relying on `NumPy` datastructures.

5.5.1 Quantum chemistry

For quantum chemical calculations, we used both in-house code for coupled cluster calculations and `PySCF` [101, 102] for most other algorithms, such as Hartree-Fock, Full CI or calculation of natural orbitals. We also used `PySCF` to calculate overlap matrices as well as one-electron and two-electron integrals. `Drudge/Gristmill` [19] was used to obtain optimized CC amplitude equations.

5.5.2 Tensor operations

In several calculations, such as the CCSD amplitude equations, WF-CCEVC, and calculation of nonorthogonal Slater determinant matrix elements, we needed to perform sums over multidimensional arrays/tensors. This is a computational bottleneck, and we used the `opt_einsum` package [103] to speed up those calculations.

5.5.3 Quantum computation

For simulations on a quantum computer, `qiskit` [104] was used. No calculations were run on actual quantum computers and all simulations were run on our home computer. As simulating quantum computers has exponential overhead, the largest number of qubits we simulated was $N_q = 12$. `qiskit` has the ability to both simulate error-free and noisy quantum computers, and can be used to calculate numerically exact expectation values without sampling, but it can also perform a quasi-stochastic sampling. `qiskit` also has a variety of pre-implemented features and ansätze, including repeated UCC, pUCCD, k-UpGCCSD and HEA-ansätze. For the change of basis operation $\exp(-\boldsymbol{\kappa})$ explained in section 4.7.3, we obtained the relevant circuit using `OpenFermion` [105] where it is implemented. For optimizing the ansatz $\hat{U}(\vec{\theta})\,|\Phi^{\text{ref}}\rangle$, we used the *BFGS* optimization algorithm, which is the most used Quasi-Newton approach and works well for optimization [106]. We avoided the problem of sampled expectation values and gradients by using simulated exact ones, unless we wanted to study the effect of sampling. After obtaining working circuits, heavy use of `qiskits transpile(optimize=2)` function was made, which collapses several gates into one and uses gate identities to minimize computational cost.

[1] https://github.com/schraderSimon/masterthesis

Part IV

Results and Discussion

Problems

We will here discuss results obtained using the different methods. As sample systems, we picked a variety of small, but possibly complicated systems, where standard methods fail qualitatively or give unsatisfying results quantitatively. We only considered closed shell singlet systems and used restricted Hartree-Fock Slater determinants as well as symmetry-adapted CC to decrease computational cost. These problems include

1. The linear dissociation reaction of BeH_2, where the distances between the beryllium atom and the hydrogen atoms are varied, while the angle is kept constant at $\gamma = 180°$. Occasionally, we will only consider the symmetric linear dissociation. This is an interesting problem because of strong static correlation in off-equilibrium [68], and CCSD gives energies that differ from the exact energy by more than chemical accuracy.

2. The BeH_2 insertion reaction [18, 83, 107] is difficult problem, describing the reaction $Be + H_2 \rightarrow BeH_2$, where the beryllium atom is placed in the origin and the position of the hydrogen atoms is given by $R_x = x$, $R_y = \pm(2.54 - 0.46x)$ (in Bohr). The system has C_{2v} symmetry along the reaction pathway. At $x = 0$ (corresponding to BeH_2), the system has a dominant Slater determinant $|\Phi_1\rangle = |(1a1)^2(2a1)^2(1b1)^2\rangle$, at $x = 4$ (Be+H_2), the dominating Slater determinant is $|\Phi_2\rangle = |(1a1)^2(2a1)^2(3a1)^2\rangle$. Thus, the system is relatively well-described by a single Slater determinant at those geometries, and CCSD is correct to chemical accuracy when choosing the correct reference determinant. However, for $x \in [2.5, 3.2]$, the states $|\phi_1\rangle$,$|\phi_2\rangle$ both have a large contribution, and the dominant Slater determinant switches $|\phi_1\rangle \rightarrow |\phi_2\rangle$ at $x \sim 2.85$ (the exact value being basis dependent). The state $|\phi_1\rangle$ remains important for $x > 3$. Figure 5.1 shows the FCI energies, the energies of $|\Phi_1\rangle$ and $|\Phi_2\rangle$, as well as the CCSD-energies when using either $|\Phi_1\rangle$ or $|\Phi_2\rangle$ as reference determinant.

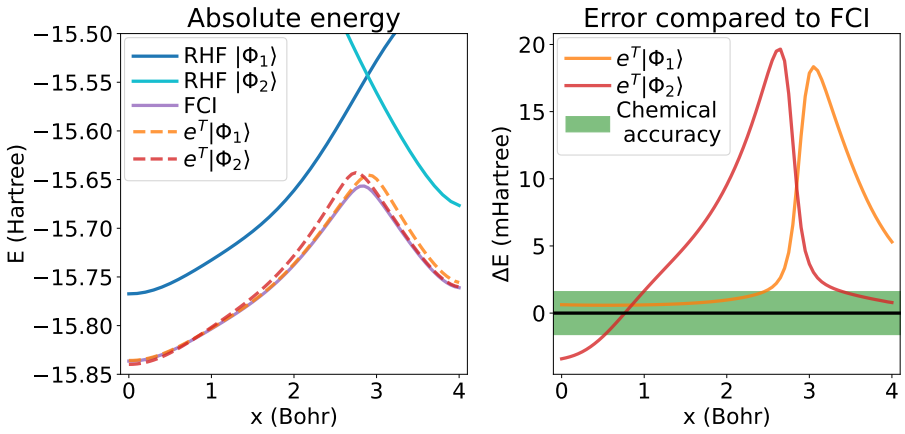


Figure 5.1: BeH_2 insertion reaction using the cc-pVDZ basis set. Left: Energy of the two dominant Slater determinants and the CCSD energy using either as reference, as well as FCI energy. Right: Deviation from FCI energy for the two CCSD wave functions.

3. The linear dissociation of N_2 [108] is an interesting problem because strong correlation effects arise when breaking the triple bond, to the point that CCSD and even CCSDT first predict a qualitatively wrong shape of the potential energy surface and eventually fail to find a solution (when an MP1 guess is used).

4. The linear dissociation of HF is an interesting test case, going from a single reference into a multireference state upon dissociation, with the CCSD energy well overshooting the FCI energy [18].

Chapter 6

Eigenvector continuation with Hartree Fock

6.1 Sampling along the potential energy surface

6.1.1 Hydrogen Fluoride

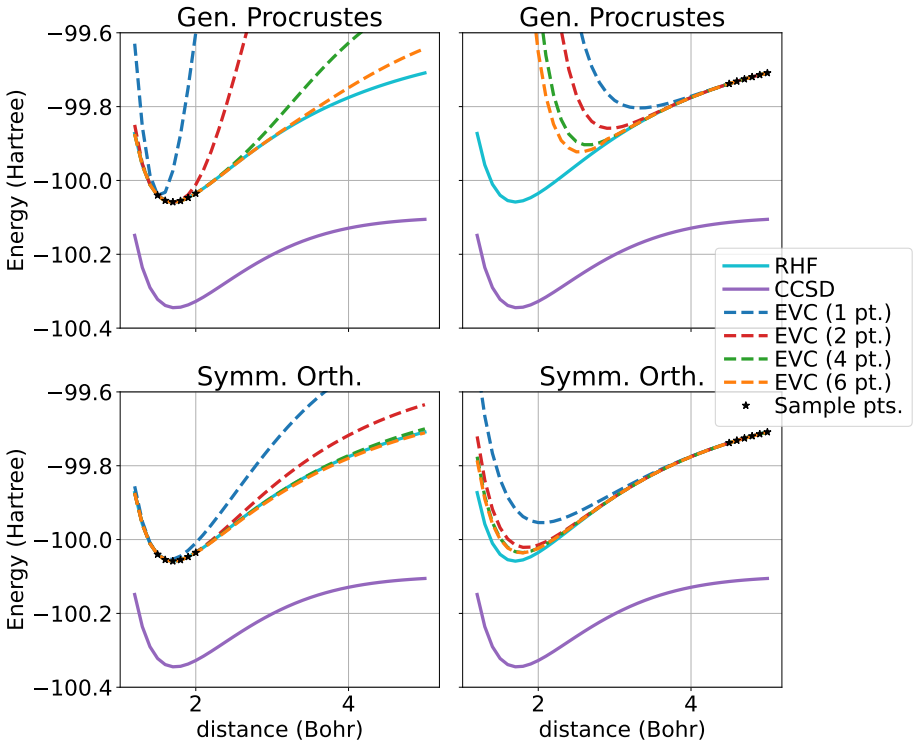


Figure 6.1: Eigenvector continuation PES for Hydrogen Fluoride in the cc-pVTZ basis with sampling points close to equilibrium (left) or far from equilibrium (right) using generalized Procrustes orbitals (top) or symmetric orthonormalization (bottom). When not all sampling points are included, only the leftmost sampling points are included.

We followed the procedure described in section 4.4.2 with restricted Hartree-Fock wave functions for the Hydrogen Fluoride molecule. This is shown in figure 6.1. As suggested in the methods section, we see how symmetric orthonormalization of the sample geometries leads to a faster convergence to the RHF energy compared to generalized Procrustes orbitals because the overlap to the HF state is larger. Sampling near the reference geometry gives us a number of Slater determinants that can reproduce the PES of the RHF solution quite well, but at the same time, it does not improve the energy further to the CCSD energy or beyond. To understand what is happening, for the same system, we also calculated the overlap between

the RHF-solution at a point $\vec{r}_T$ and the EC-solution at the same point $|\langle\psi_{\mathrm{HF}}(\vec{r}_T)|\psi_{\mathrm{EC}}(\vec{r}_T)\rangle|^2$. We present values for the lower right graph of figure 6.1 in table 6.1, which shows that we are indeed getting states that have a very large overlap with the HF wave function - the HF state is being continued.

Table 6.1: Overlap between EVC and Hartree-Fock wave functions $\left|\langle\psi_{HF}(\vec{r}_T)|\psi_{EC}^{N_{EC}}(\vec{r}_T)\rangle\right|^2$*, where* N_{EC} *stands for the number of sampling points. The same sampling points as in the left lower graph of figure 6.1 were chosen. We do not write the dependence of the wave functions on* $\vec{r}_T$ *for brevity.*

Interatomic distance (Bohr)	$\left\|\langle\psi_{\mathrm{HF}}\|\psi_{\mathrm{EC}}^1\rangle\right\|^2$	$\left\|\langle\psi_{\mathrm{HF}}\|\psi_{\mathrm{EC}}^2\rangle\right\|^2$	$\left\|\langle\psi_{\mathrm{HF}}\|\psi_{\mathrm{EC}}^4\rangle\right\|^2$	$\left\|\langle\psi_{\mathrm{HF}}\|\psi_{\mathrm{EC}}^6\rangle\right\|^2$
1.5	1.000	1.000	1.000	1.000
3	0.907	0.964	0.998	0.998
4.5	0.807	0.891	0.986	0.991

Inclusion of singlets

We studied again Hydrogen Fluoride, this time including all possible spin-adapted singly excited Slater determinants. We observe that Brillouin's theorem does not hold anymore, hence singly excited determinants at a different geometry will contribute. This is shown in figure 6.2.

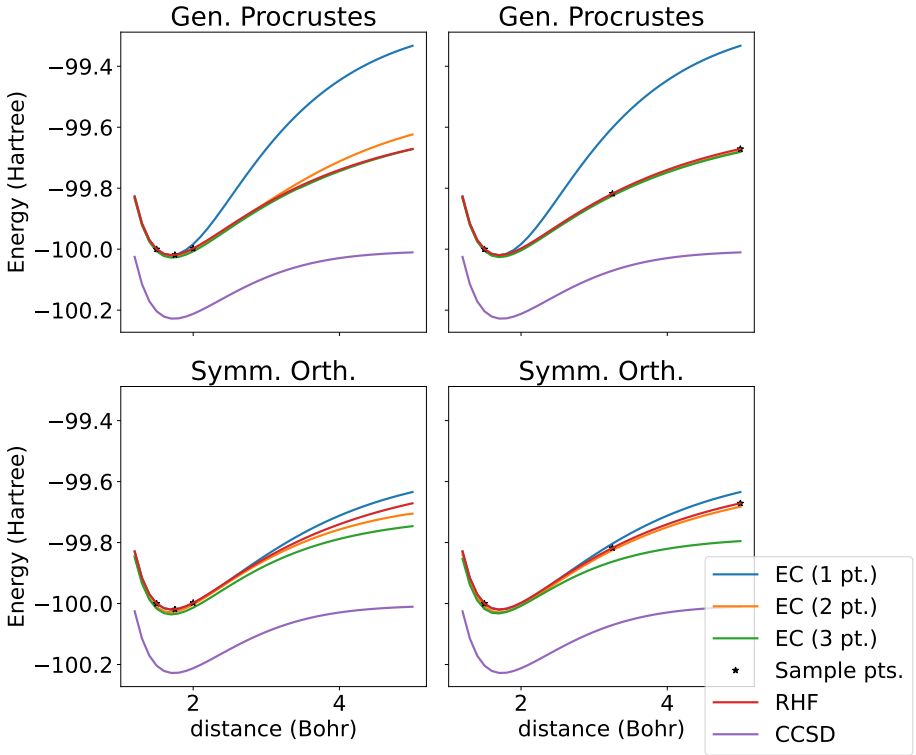


Figure 6.2: Eigenvector continuation PES for Hydrogen Fluoride in the cc-pVDZ basis with sampling points close to equilibrium (left) or far from equilibrium (right) using generalized Procrustes orbitals (top) or symmetric orthonormalization (bottom), including all possible singly excited determinants that preserve $M_s = 0$*. When not all sampling points are included, only the leftmost sampling points are included.*

When sampling the ground state and the singly excited determinants at just one sample point, it is possible to almost reproduce the RHF-solution, while the inclusion of several points leads to improved results. However, naive addition of several sampling points, as we did here, does not lead to substantial improvements in general.

6.1.2 Beryllium Dihydride

Having described the BeH_2 dissociation reaction in the beginning of the result section, figure 6.3 shows the potential energy surface of the BeH2 molecule with eigenvector continuation. When only including sampling points close to $x = 0$, we see how the energy stays close to that of $|\Phi_1\rangle$ for a longer part of the trajectory, but the energy is not able to "dip" down to state $|\Phi_2\rangle$. This changes when including sampling points from "both sides", which gives a qualitatively right energy curve without "cusp", even though more sampling points would have been necessary to get the energy below the RHF energy. To improve upon the HF energy, we see that it is necessary to include both states $|\Phi_1\rangle$ and $|\Phi_2\rangle$ into the sample space, leading to a large dicrease in the energy, as is visible in the lower right graph.

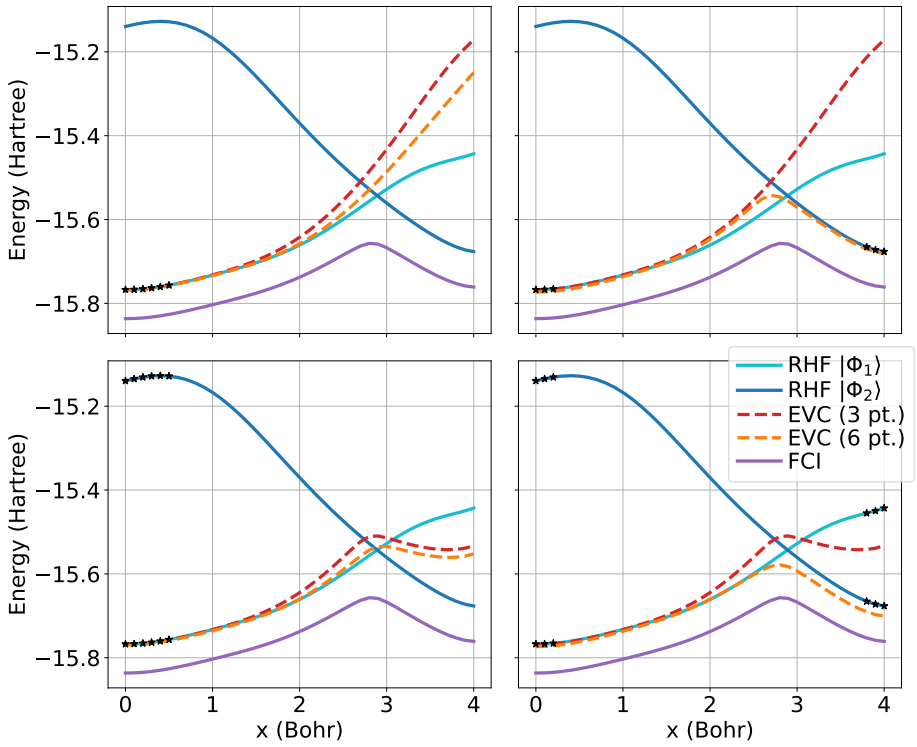


Figure 6.3: Eigenvector continuation PES for BeH_2 in the cc-pVDZ basis (the largest basis we could do Full CI in) with sampling points on the left (left), left and right (right), and without (upper) or with (lower) inclusion of both states $|\Phi_1\rangle$ and $|\Phi_2\rangle$. In all cases, symmetric orthonormalization was used to transform the Slater determinants. When not all sampling points are included, only the leftmost sampling points are included.

6.2 Sampling at one geometry

We considered again the HF-molecule and the BeH_2 molecule, this time creating several Slater determinants at the same geometry by tweaking the strength of the two-body interaction with a parameter λ, as we described in section 4.4.3, p. 51 and we see the results in figure 6.4. The costly part here is to create the sample Slater determinants, not creating the EVC matrices. First of all, we see how we, for HF essentially recover a qualitatively correct multi-reference solution which follows the qualitatively correct curve for both basis sets considered. There is a similar improvement for BeH_2, where we included both states $|\Phi_1\rangle$ and $|\Phi_1\rangle$ for each λ, even though it closes less in on the exact FCI solution. For HF, there is basically no difference between 10 and 13 sample determinants. Furthermore, going from 4 to 7 improves the energy at the dissociation limit, but not along the dissociation line, while going from 7 to 10 does not improve the PES at dissociation limit, but along the dissociation line. For BeH_2, even the improvement going from 4 to 13 sample determinants is very slight. We see how the energy gap between the HF solution and the FCI/CCSD solution grows for

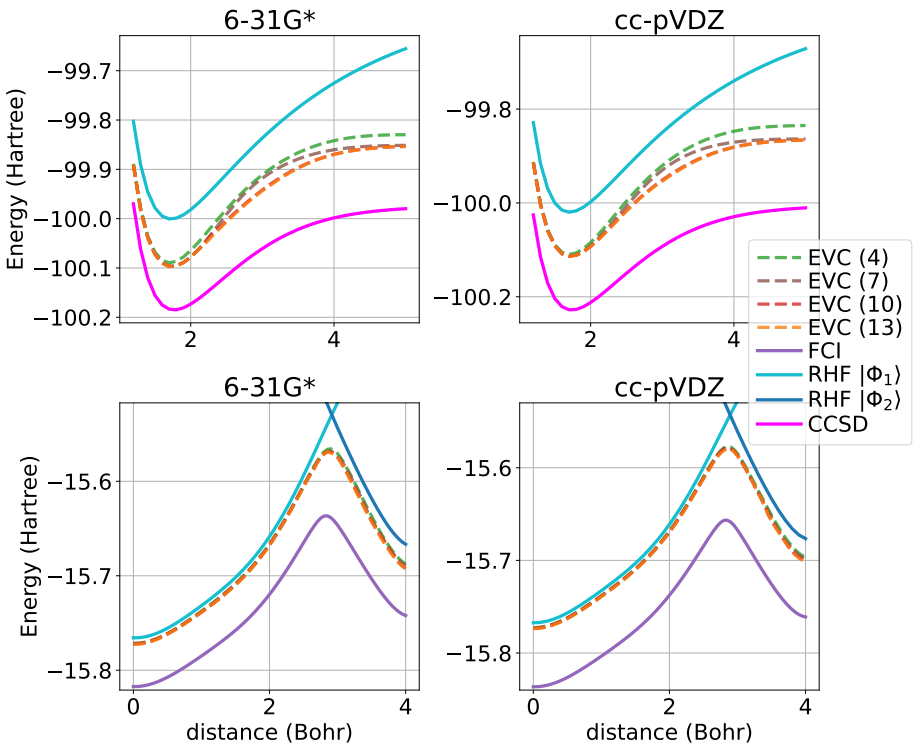


Figure 6.4: Multireference EVC for the BeH_2 insertion and HF with a tweaked 2-electron repulsion. The number in parentheses EVC(N) refers to the number of λ chosen for $\lambda \in [1.1, 1.05, \ldots, 0.55, 0.5]$, where we picked the first N ones. For BeH_2, we included two determinants with different point group symmetry per λ, corresponding to $|\Phi_1\rangle$ and $|\Phi_2\rangle$ at $\lambda = 1$.

larger basis sets, which is expected, as the number of sample points stays the same at an increased size of the Hilbert space.

6.3 Discussion

Generally, we see that few sampling points along a potential energy surface and applying the EVC procedure are sufficient to reproduce the RHF energy. This might be useful when there are troubles to converge the SCF equations and instead solve the problem of finding the Slater determinant that maximizes the overlap with the EVC solution. In systems where the principal configuration changes, such as BeH_2, smooth energy surfaces are obtained compared to the lowest lying HF-state. For simpler systems, the RHF energy is sufficiently extrapolated, while the change of principal configuration in BeH_2 makes it necessary to sample at geometries from both principal configurations to obtain qualitatively correct potential energy surface. Inclusion of determinants from several principal configurations at each sampling point leads to improved potential energy surfaces, but this is then no longer a black box method. The inclusion of singles, which is an expensive calculation scaling as $O(N^8)$ [28], leads to minor improvements of the PES. We believe that the inclusion of doubles would lead to strongly improved energies, but we did not consider it due to the prohibitively expensive scaling.

When using HF states from different electron interaction strengths as basis at each geometry, we see that the results vary a lot, in the sense that the qualitative energy shape improves greatly for some systems (e.g. HF, but less for other systems (BeH_2). The improvement when adding more points is seemingly unpredictable, with bigger improvements at some geometries than at others. Furthermore, there is no "natural" way to improve the final solution any further without adding singles/doubles, which we discussed is prohibitively expensive. While those results are interesting, we do not consider this method to be generally applicable.

Chapter 7

Coupled Cluster eigenvector continuation

7.1 Results

7.1.1 Hydrogen fluoride

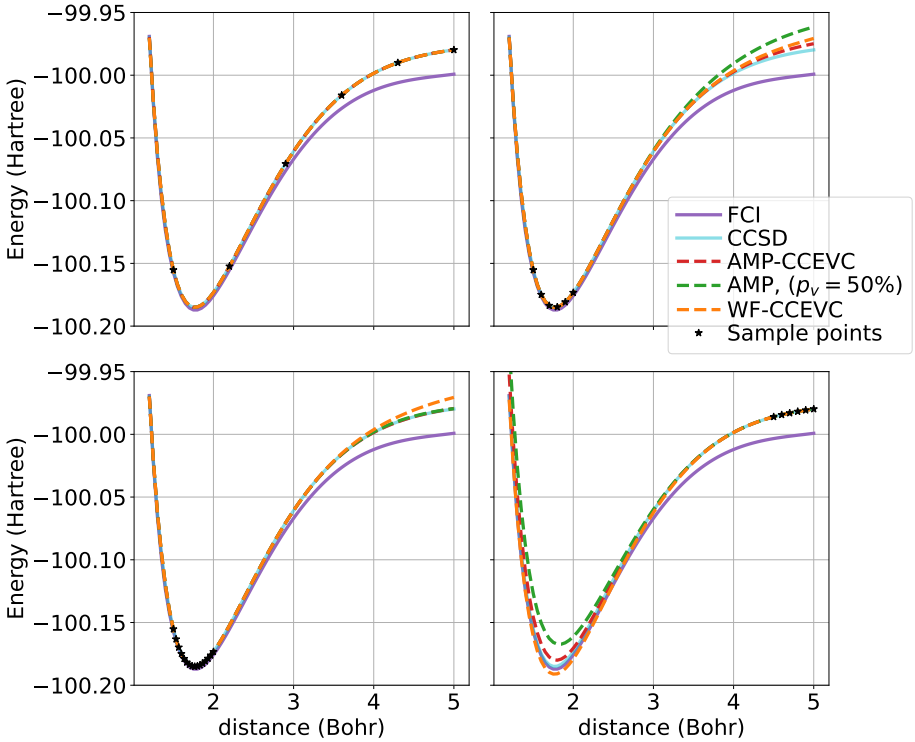


Figure 7.1: Energies of the HF molecule when dissociating using FCI, CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs with the reference geometry $x = 1.75$ Bohr where used. The 6-31G basis set was used, being the largest basis set for which we could calculate FCI energies. Top left: 6 evenly spread sampling points. Top right: 6 sampling points close to equilibrium geometry. Bottom left: 16 sampling points close to equilibrium geometry. Bottom right: 6 sampling points far from equilibrium geometry.*

Figure 7.1 shows the energy of the HF molecule at different bond lengths using different methods and Procrustes orbitals in a small basis set (6-31G*). We see that both WF-CCEVC and AMP-CCEVC interpolate extremely well, and both extrapolate well, even though the AMP-CCEVC energy, unlike WF-CCEVC, converges to the CCSD energy as more sample points are added. Increasing the number of sample points close to the equilibrium geometry has essentially no advantage for WF-CCEVC, but it does for AMP-CCEVC. For intrapolation,

that is, sampling across several geometries of the PES, as few as 6 sample points manage to reproduce the full CCSD potential energy surface for AMP-CCEVC, WF-CCEVC and the approximation to AMP-CCEVC. In each case, the FCI energy remains out of reach. We also see that the parameter reduction approach in AMP-CCEVC leads to extremely good interpolation results and also works well when many sample points are used, but it is inferior to full AMP-CCEVC for extrapolation when few sampling points where used. This behaviour is expected and follows our discussion in the methods section.

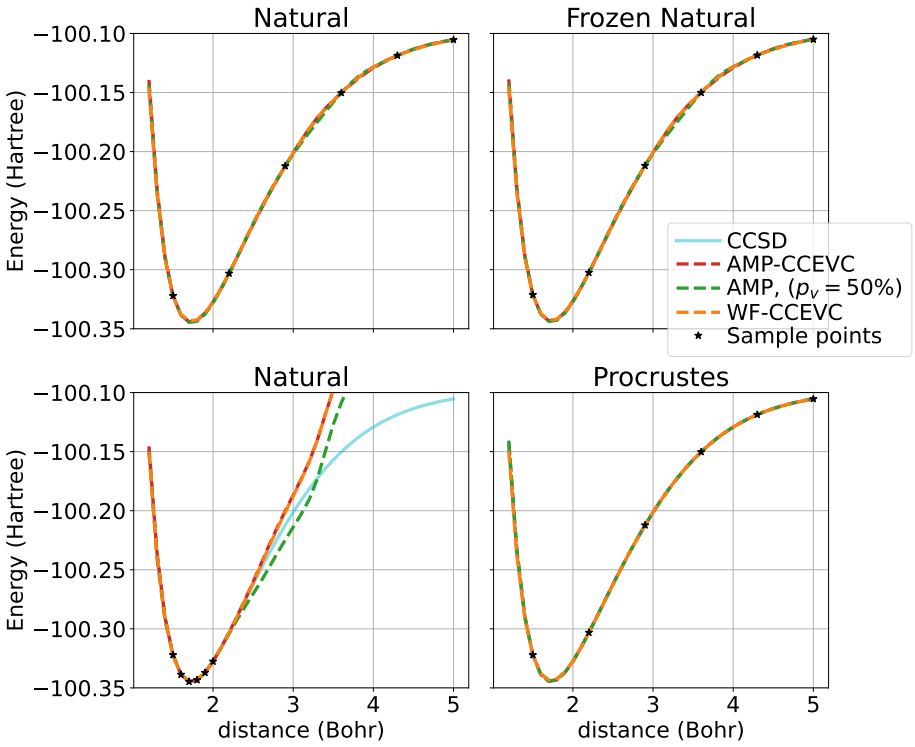


Figure 7.2: Energies of the HF molecule when dissociating using CCSD, WF-CCEVC and AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. The cc-pVTZ basis set was used. Procrustes MO's with the reference geometry $x = 1.75$ Bohr where used in the bottom right corner, while the other plots used natural orbitals. The plot in the top right corner used frozen orbitals, freezing 5 of 44 with an occupation number steadily below 10^{-4}, which lead to a maximum deviation in energy of 0.8 mHartree and a minimum deviation of 0.2 mHartree.

Figure 7.2 shows again the dissociation of HF, where we used a much larger basis set (cc-pVTZ). In the bottom right corner, we used Procrustes orbitals, which shows that the methods' interpolation ability is not impaired by larger basis sets. With the conservative cutoff of 10^{-4} for the natural occupation number, only 5 of 44 NOs were frozen, but the energy deviations were well below chemical accuracy. Following the discussion in Ref. [24], it is however clear that a larger percentage of natural orbitals can be frozen in larger basis sets with a lower loss of correlation energy, and that using 39 NOs obtained from a larger basis set would give higher parts of the correlation energy, which we did not consider here in order to compare results to the non-cutoff case. For example, using the cc-pVQZ basis set with the same cutoff, we would only use 40 out of 85 NOs (aug-cc-pVTZ: 39 out of 69). We see that natural orbitals both interpolate and extrapolate worse than Procrustes orbitals. This is especially visible in the parameter reduced AMP-CCEVC in the upper two graphs for interpolation. This result not surprising, given that Procrustes orbitals try to reduce the change in the expansion coefficients of the MOs and thus the Hamilton operator, which is not the case for natural orbitals.

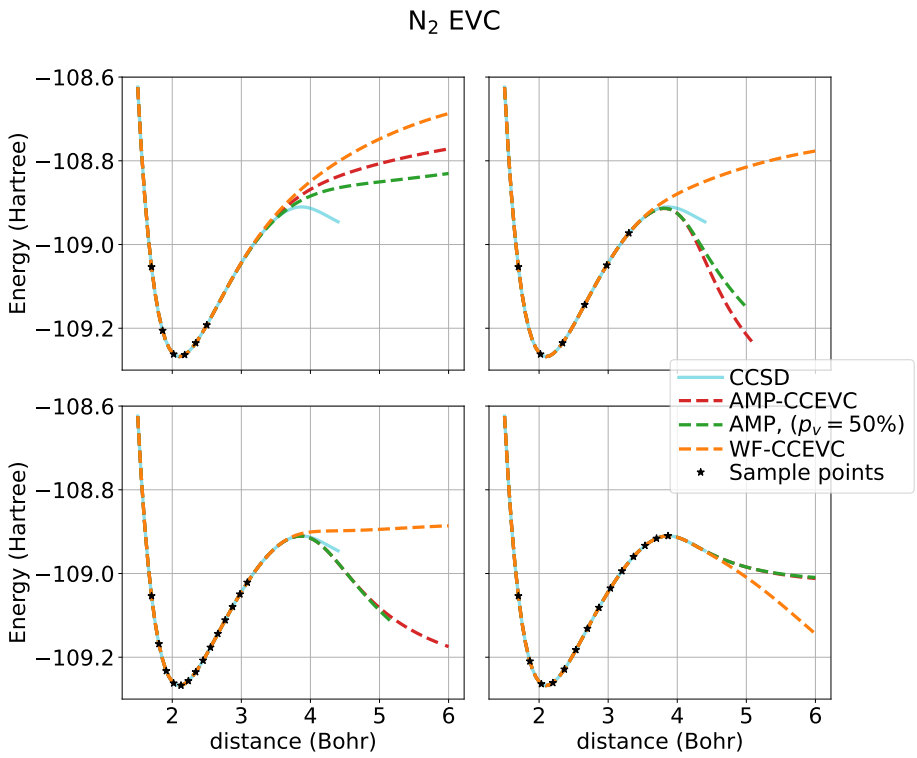


Figure 7.3: Energies of the N_2 molecule when dissociating using CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs with the reference geometry $x = 2$ Bohr where used. Procrustes MO's with the reference geometry $x = 2$ Bohr where used. The cc-pVDZ basis set was used. Energies where no solution to the amplitude equations where found, are not included.

7.1.2 Nitrogen

Figure 7.3 shows the energy of the N_2 molecule at different bond lengths using different sample points. The N_2 molecule is an interesting molecule to study, as the CCSD energy *decreases* as the molecule dissociates, which is physically wrong. Bulik et al. [108] have studied N_2 and shown that inclusion of triple excitations in the cluster operator (CCSDT) leads to the same catastrophic breakdown. They however proposed a generalization to pCCSD (we discussed the meaning of "p" when we introduced k-UpGCCSD) where this brea down does not occur, at the cost of getting higher energies even where CCSD is approximately correct. Eigenvector continuation attains CCSD energies without this breakdown. One sees that the AMP-CCEVC energy follows the CCSD well beyond the sample points, but does the same, unphysical "knick" as the CCSD energy, unless we include few sample points that lie far away from the unphysical region, which we consider to be a coincidence. Using only a subset of amplitudes in the AMP-CCEVC approach gives similar interpolation, but different extrapolation results, as was the case for HF. The results get much more interesting for the WF-CCEVC approach. We see that the upper left and lower right subplots are qualitatively wrong, while the other two are qualitatively correct, with the lower left one almost becoming straight, such as the correct FCI energy [108]. That shows the importance of the sampling points: Sampling too far away from where the "difficult" region starts and sampling too few points, gives a bad fit that overestimates the energy. Similarly, sampling wrong points makes the WF-CCEVC algorithm, which is not variational, take over the wrong behaviour. However, a careful choice of the sample points may give proper results. Inclusion of several good sample points clearly improves the WF-CCEVC energy.

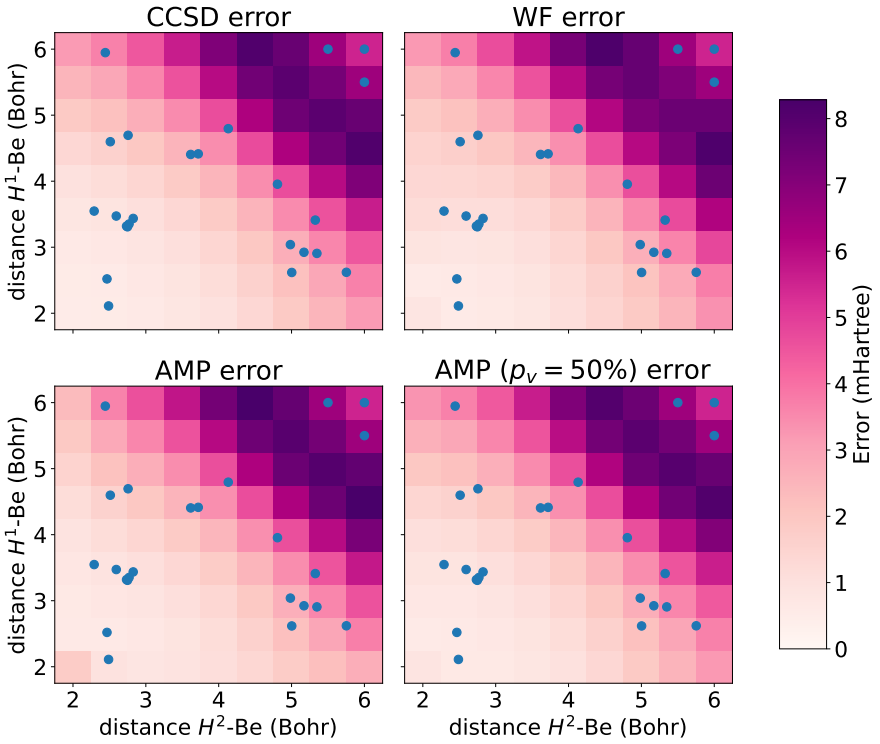


Figure 7.4: Energy difference compared to Full CI solution of the asymmetric linear stretch of the BeH_2 molecule using CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs with $R_1 = R_2 = 1$ Bohr were used. The cc-pVDZ basis set was used. The 26 sample points are marked in blue and were chosen randomly except for three that were purposely put in the top right corner, while the "difficult" area was purposefully not included in the sampling space.

7.1.3 Beryllium hydride stretch

Figure 7.4 shows the derivation from the FCI energy at different geometries of the molecule. We varied the distance of each hydrogen atom to the Beryllium atom (R_1, R_2) individually, while keeping the angle constant at 180°. This figure shows that eigenvector continuation can also be applied to multidimensional problems and still has a substantial extrapolation ability. By sampling outside of the difficult region (e.g. in the white and pink, but not the violet region) on both sides, the WF-CCEVC approach gives decreased errors in the difficult region, such as at $R_1 = R_2 = 5$ Bohr. Those observations do not hold for the AMP-CCEVC method and the parameter reduced AMP-CCEVC, which essentially recover the potential energy surface of CCSD, except at $R_1 = R_2 = 2$ Bohr.

7.1.4 Beryllium hydride insertion

When considering the Berillyum hydride insertion using EVC, special care needs to be taken that the same reference ($|\Phi_1\rangle$ or $|\Phi_2\rangle$) is used along the whole PES, as both AMP-CCEVC and WF-CCEVC require the reference to remain the same along the whole region of interest. We only present results for WF-CCEVC, as we have already established that AMP-CCEVC can be used as an approximation to CC, while WF-CCEVC has the ability to improve the guesses. We have to follow a different approach compared to the previous chapter, as we have to use either $|\Phi_1\rangle$ or $|\Phi_2\rangle$ as references at all geometries, but not both at different geometries.[1] In figure 7.5, we show the WF-CCEVC approach using different types of sampling points and either $|\Phi_1\rangle$ or $|\Phi_2\rangle$ as reference. In the upper figures, we see that choosing a set number of points along the PES essentially recovers the PES. When using $|\Phi_2\rangle$ as reference, however, the WF-CCEVC algorithm finds a wrong, lower lying state. This is not a numerical error and happens even when adding a very large $\epsilon \sim 10^{-8}$ to the overlap matrix. The correct

[1] We will show in the next chapter that this is not necessarily the case on a quantum computer.

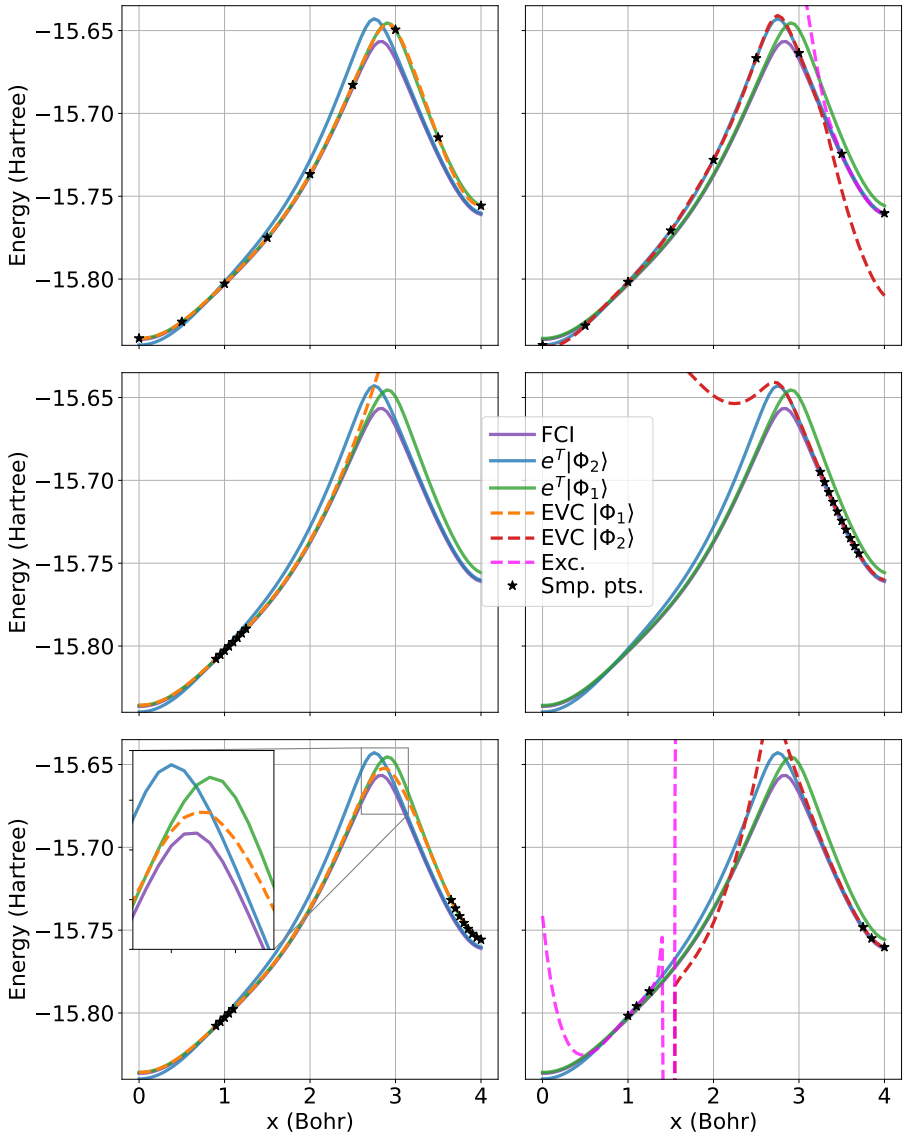


Figure 7.5: Energies of the BeH_2 insertion reaction energies in the cc-pVDZ basis set. Left: Using $|\Phi_1\rangle$ as a reference. Right: Using $|\Phi_2\rangle$ as a reference. Different sampling point geometries were used. In the upper and lower plot to the right, we also plotted the excited EVC solution arising in the matrix diagonalization, in magenta.

energy is now found in the first excited state. In the middle, we see how sampling only in the region where $|\Phi_1\rangle$, $|\Phi_2\rangle$ are good reference determinants, respectively, does not suffice to obtain the qualitatively correct PES along the whole geometry. In the bottom left, we see that when sampling on both sides, even though $|\Phi_1\rangle$ is not a good reference for $x > 3$, improves the CCSD energy drastically and also gives a correct shape of the PES. While we did not obtain chemical accuracy, which is because we used sample points where there is no chemical accuracy, the WF-CCEVC procedure lead to improved energy near the transition point. On the right, we see that we cannot do the same with $|\Phi_2\rangle$, as it is again the excited state that has (qualitatively) the correct energy, while the ground state energy is nonsensical for $x \sim 1.5$. This is not a numerical error either. A possible explanation is that $|\Phi_2\rangle$ is a very bad reference on the left site and that the amplitudes hence are very large, which generally impacts the applicability of CCSD and thus WF-CCEVC too.

7.1.5 Difference between HF, BeH_2 and N_2 for WF-CCEVC

WF-CCEVC performs better than CCSD for N_2 and gives qualitatively correct results when using the correct sample points, while the HF solution is not improved at all compared

to CCSD. The BeH2 insertion solution improves when sampling from both "sides". To understand why, we can consider the singular values of the geometry-independent overlap matrix to see how many different dominating configurations there are. By the discussion in section 5.1.1, the eigenvectors with the largest singular values contribute the most to the overlap matrix. Th We found the following (all x in Bohr):

1. For HF, sampling at 6 evenly distributed points $x \in [1.5, 4]$ gives one singular value above 1 ($\sigma_1 = 5.91, \sigma_2 = 0.093, \sigma_3 = 0.007$). Adding the points $4.5, 5.0$ does not change that ($\sigma_1 = 7.78, \sigma_2 = 0.24, \sigma_3 = 0.007$).

2. For N_2, sampling at 5 evenly distributed points $x \in [1.5, 3.5]$ gives one singular value above 1 ($\sigma_1 = 4.88, \sigma_2 = 0.13$). Adding the points $4.0, 4.5$ increases that number to two ($\sigma_1 = 6.39, \sigma_2 = 1.08, \sigma_3 = 0.048$).

3. For BeH_2 (using configuration $|\Phi_1\rangle$), sampling along the whole PES (first graph) gives two out of 9 singular values above 1 ($\sigma_1 = 8.4, \sigma_2 = 1.3, \sigma_3 = 0.92$). Sampling at one geometry on the left gives one out of eight singular values above 1 ($\sigma_1 = 7.9995, \sigma_2 = 0.00048$). Sampling on both sides has 2 out of 13 singular values above one ($\sigma_1 = 11.9, \sigma_2 = 1.58, \sigma_3 = 0.021$).

When two singular values are large (e.g. above 1), this corresponds to the situation that there are two qualitatively different important states. For N_2, we see how there arises a second, faulty configuration when including the two extra sample points. This configuration is not included when only sampling in the correct region. Eigenvector continuation hence finds a solution within the configurations from the "good" region. For HF, there is only one qualitatively important state, and the second most important state does not suddenly "pop up" when including more points. For BeH_2, we also see how there is a second qualitatively important state showing up when including points from the right side - unlike N_2, it is desirable to include that state.

7.1.6 Water around equilibrium structure

Having established that EVC works well for 1-dimensional and 2-dimensional problems when a sufficient number of sample points is given, we considered the stochastic convergence of the CC-EVC energy to the CCSD energy for a three-dimensional problem, namely, water around its equilibrium structure. Free parameters are the distance between the oxygen and the hydrogen atoms r_1, r_2 as well as the angle α. We took sampling points randomly from uniform distributions for for $r_1, r_2 \in [1.3, 4.0]$ $(Bohr)$ and $\alpha \in [99.5°, 109.5°]$. Figure 7.6 shows the mean deviation from the CCSD energy as function of the number of sampling points included. We see a monotone convergence of the mean and all quantiles for AMP-CCEVC and for allmost all values with WF-CCEVC. Given that AMP-CCEVC tries to approximate the cluster operator, while WF-CCEVC does not, this is not surprising. For all four cases, we see that the mean is higher than the median, which converges to chemical accuracy faster, showing that outliers with high energy deviations pull up the mean energy for both WF-CCEVC and AMP-CCEVC. Furthermore, we see that as little as 12-13 sampling points are sufficient to have the means converge to chemical accuracy for full AMP-CCEVC and WF-CCEVC. When using parameter-reduced AMP-CCEVC, convergence is slower and slightly noisy, both for the mean and the median. This is not surprising either, as these methods are approximations. We nevertheless see the same trend and convergence of the mean to just above chemical accuracy, justifying the use of those methods further.

7.2 Time usage of AMP-CCEVC

The time-consuming step is to calculate the error vector $e(c_i)$ $i = 1, \ldots, L$ at each iteration. This is the value which we want to be equal to zero. In the parameter-reduced version, we approximate $e(c_i)$, and the cost is reduced. This might impact the accuracy of the

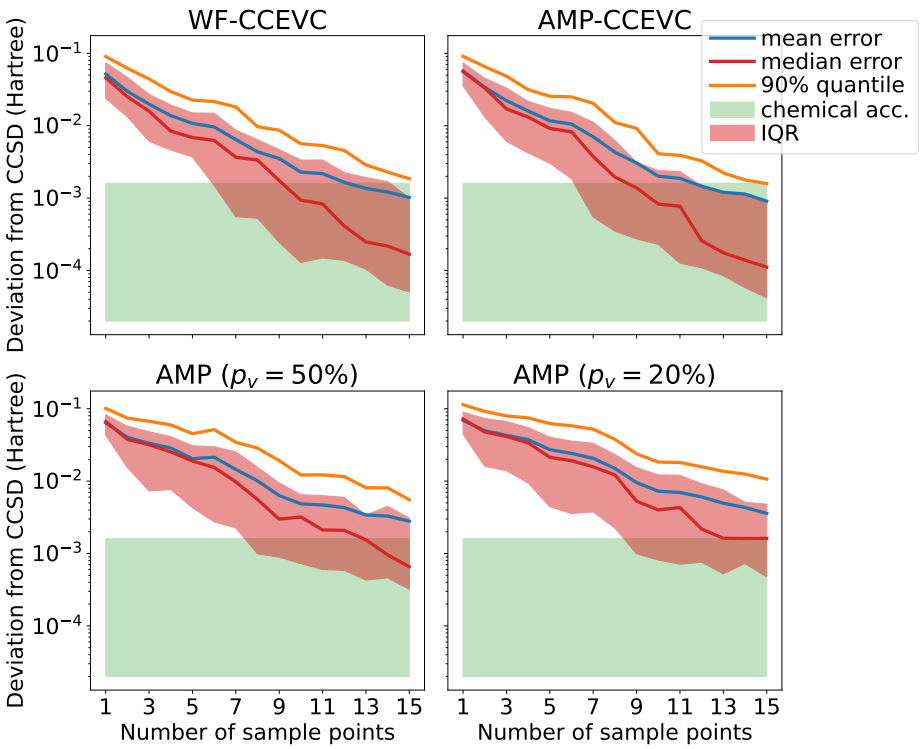


Figure 7.6: Absolute deviation from CCSD energy (mean and median) as well as IQR (range of 2nd and 3rd quartile) for WF-EVC and AMP-EVC ($p_v \in \{100\%, 50\%, 20\%\}$) as function of the number of sampling points for water in the geometry described in the main text using the aug-cc-pVDZ basis set. We used 16 different test geometries from the same range and repeated the experiment 3 times to avoid "lucky" geometries.

amplitude calculations. In table 7.1, we compared AMP-CCEVC with the parameter reduced version with different percentages of included virtual orbitals using Procrustes orbitals for the symmetric stretch of BeH_2. The same is shown for HF in table 7.2. We considered a pure interpolation problem, having established that these methods work worse for extrapolation.

Table 7.1: Number of iterations in the quasi-Newton approach to reach a projection error $e(c_i)$ below 10^{-5} for each i, total time use per geometry and average projection error as function of the percentage of included virtual orbitals (rounded) for the symmetric stretch of BeH_2 in the cc-pVTZ basis set. As initial guess, we used the amplitudes of the sampling points closest to the geometry in question. Procrustes orbitals were used. As sample geometries, we used 11 evenly spaced samples for the interatomic distance $R \in [1.5, 5.5]$. Values were obtained for 33 different geometries evenly spaced for the interatomic distance $R \in [1.6, 5.]$. `opt_einsum` [103] was used for tensor operations.

Percentage of included $M_{\text{virt.}}$	100	80	40	20	10
Number of iterations per geometry	7 ± 4	7 ± 4	7 ± 4	7 ± 4	7 ± 4
Total time per geometry (s)	3.6 ± 2.1	3.2 ± 1.8	2.1 ± 1.1	1.4 ± 0.7	0.7 ± 0.4
Absolute max. projection error	0.0013	0.0013	0.0013	0.0012	0.0033

Table 7.2: Number of iterations in the quasi-Newton approach to reach a projection error $e(c_i)$ below 10^{-5} for each i, total time use per geometry and average projection error as function of the percentage of included virtual orbitals (rounded) for the stretch of HF in the cc-pVTZ basis set. As initial guess, we used the amplitudes of the sampling points closest to the geometry in question. Procrustes orbitals were used. As sample geometries, we used 9 evenly spaced samples for the interatomic distance $R \in [1.0, 5.0]$. Values were obtained for 33 different geometries evenly spaced for the interatomic distance $R \in [1.3, 4.8]$. `opt_einsum` [103] was used for tensor operations.

Percentage of included $M_{\text{virt.}}/M_{\text{occ.}}$	100/100	80/80	40/60	20/60	10/60
Number of iterations per geometry	6 ± 3	6 ± 3	7 ± 3	8 ± 4	9 ± 4
Total time per geometry (s)	2.7 ± 1.0	2.6 ± 0.9	2.2 ± 0.9	2.0 ± 0.9	1.3 ± 0.6
Absolute max. projection error	0.0047	0.0043	0.0060	0.0074	0.0079

We see that using only a fraction of amplitudes has almost no impact on the number of iterations as well as the projection errors for BeH_2, which justifies using the method for interpolation. Only when including less than 20% of virtual orbitals did the error increase. For HF, there is a slight increase in the number of iterations as more orbitals are neglected, but the projection error did not increase much. In sufficiently large basis sets, inclusion of as little as 10% of the virtual orbitals thus still yield good results and approximate the exact CCSD amplitudes closely. We see that by reducing the number of parameters, we can cut computation time of `opt_einsum` down by more than 50% without losing relevant accuracy for interpolation. The reduction in time usage even more drastic when using `einsum` from the `numpy` library, which is less efficient at reusing results, instead.

7.3 Discussion

The AMP-CCEVC method works very well at "learning" the true CCSD parameters in a few-variable representation, in one, two and three dimensions and also manages to extrapolate well for some range, as long as the sample points lie "close" to the geometry of interest or a sufficient number of sampling amplitudes were included. Furthermore, the dimensionality reduction method works well and gives almost the same energies as the non-reduced case for interpolation, even though the extrapolation results differ. This is in agreement with theoretical considerations of the parameter reduction approach. It is a very efficient way to achieve speedup of CCSD calculations, and we believe that even larger speedups are possible by implementing the parameter-reduced AMP-CCEVC when considering different types of numerical optimizations. The AMP-CCEVC method is a very good approximation to a full CCSD calculation and can be used on its own, but also to give a good initial guess to CCSD amplitudes when a higher accuracy of the amplitudes is desired. As AMP-CCEVC is used to obtain an approximation to the exact CCSD cluster operator, every quantitative or qualitative improvement that AMP-CCEVC gives over CCSD, as for nitrogen, is not an inherent strength of the method, in the sense that inclusion of extra sample points close to the original sampling points will likely recover the erroneous behaviour, as we have seen. We found that the energy obtained using AMP-CCEVC converges monotonously to CCSD, even when a parameter reduction approach is used. We conclude that AMP-CCEVC has theoretical value as it shows that the CCSD amplitudes can be expressed with few parameters, and practical value as it can reduce the cost of computing the CCSD amplitudes using the parameter reduction approach.
Unlike AMP-CCEVC, the WF-CCEVC method is not solely an approximation to CCSD, as the wave function itself is written as a linear combination of several CCSD wave functions and cannot be written as a CCSD wave function. It interpolates the CCSD wave function as good as AMP-CCEVC, and it can be used to improve the CCSD results by careful selection of the sampling points, as shown for N_2 & BeH_2. It was especially possible to obtain a qualitatively correct dissociation curve for N_2 with the same quality of energy and computational scaling as CCSD. The drawback is that it is by no means a black box method when used for this type of extrapolation, and knowledge of the chemical problem at hand is

required whenever an improvement of the CCSD energy is desired, even though analyzing the singular values of the overlap matrix tells us when a "new" state is found. Furthermore, as the method is not variational, naive inclusion of further sampling points can increase the error and even give rise to nonphysical potential energy surfaces, as we have observed for nitrogen. For the beryllium hydride insertion reaction, we saw an additional error, were there arised an nonphysical lower-lying state, while the physically correct result was actually found in an excited state. An increase of the number of sampling points gives an increased extrapolation range, but also leads to an increased computational extra cost. The speedup of WF-CCEVC for calculating potential energy surfaces is orders of magnitudes lower than the one attained in Ref. [3], who finished their calculations in one hour, compared to a full CCSD calculation which they estimated would have taken 20 years. Because their Hamiltonian was of the shape

$$\hat{H}(\vec{\alpha}) = \sum_{i=0}^{N_\alpha} \alpha_i \hat{H}_i, \tag{7.1}$$

they only had to similarity-transform each Hamiltonian $\hat{H}_i$ once per sample cluster operator. Unlike them, we had to recalculate the similarity-transformed Hamiltonian at every geometry due to the form of the Hamiltonian, and thus cannot avoid the $O(M^6)$ scaling. In addition, our Hamiltonian is not linear in the perturbation, and due to this complication, we did not obtain the same quality of extrapolation that they achieved.

Just like AMP-CCEVC, WF-CCEVC shows that the CCSD wave function along a whole geometry can be expressed with very few parameters (compared to $O(M^2N^2)$), in the sense that chemical accuracy is obtained when including few wave functions. Both WF-CCEVC and the parameter reduced AMP-CCEVC make use of this and may reduce the computational cost with little accuracy loss.

Both methods are single-reference methods and can generally not be applied to multi-reference problems. However, given the fact that WF-CCEVC is relatively simple to implement when the amplitudes and the λ-amplitudes are available, it is an easy to use method with a cost no higher than CCSD, which can give substantial improvements when applied correctly.

We have also shown that there are several choices to choose orbitals that make the Hamiltonian change analytically. Using natural orbitals, it is still possible to freeze those with low occupation numbers, as long as they remain frozen at every geometry. We found that using natural orbitals is inferior to using Procrustes orbitals in terms of interpolation and extrapolation ability. This is however not a problem, as one can remove the least important natural orbitals and then use Procrustes orbitals on top of that. We do not claim that Procrustes orbitals are the best choice, but we found that they worked well for eigenvector continuation.

Chapter 8

Quantum computer eigenvector continuation

8.1 Results

8.1.1 EVC with Procrustes orbitals: BeH_2 stretch

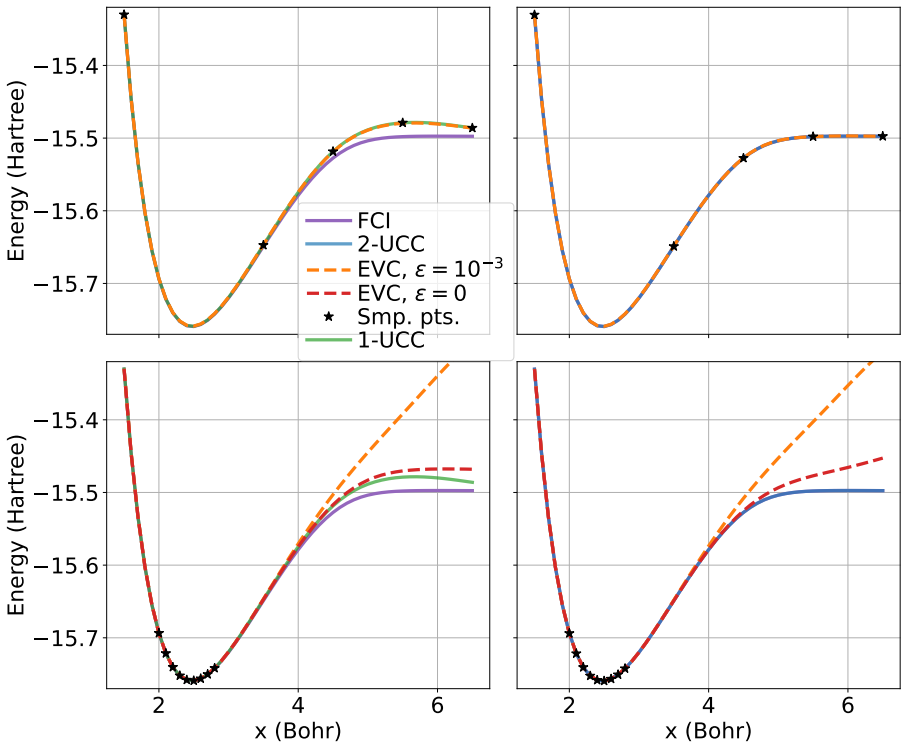


Figure 8.1: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes orbitals. The STO-6G basis set was used and the two core electrons were frozen. 5 additional qubits were tapered off. We used either a 1-UCCSD (left) or a 2-UCCSD (right) ansatz (e.g. either one or two Trotter steps with independent parameters) to obtain the sample wave functions. Note that the 2-UCCSD ansatz is essentially exact. We used two sets of cutoffs ϵ as cutoffs for canonical orthonormalization.

We considered the symmetric linear dissociation of BeH_2 molecule. We used either a 1-UCCSD (left) or a 2-UCCSD (right) ansatz (e.g. either one or two Trotter steps with independent parameters) to obtain the sample wave functions. While UCCSD methods are not the most efficient methods, we got energies well below the chemical accuracy threshold for 2-UCCSD, and we can hence consider it to essentially give the exact wave function, while 1-UCCSD is an approximation. The EVC potential energy curve using Procrustes orbitals for this system is shown in figure 8.1. The STO-6G basis set was used and the two core electrons

were frozen using `Qiskit's ActiveSpaceTransformer` [109]. Qubit tapering was used to reduce the number of qubits from 12 to 7. Expectation values of the matrices $\mathbf{H}^{\text{EVC}}$ and $\mathbf{S}^{\text{EVC}}$ were calculated, not sampled, and are hence numerically exact. In order to solve the generalized eigenvalue problem $\mathbf{H}^{\text{EVC}}\mathbf{C} = \mathbf{S}^{\text{EVC}}\mathbf{C}\boldsymbol{\epsilon}$, we used canonical orthogonalization as described in Refs. [10, 31], discarding eigenvalues of $\mathbf{S}^{\text{EVC}}$ smaller than $s < \epsilon$. This is indeed a standard approach which is used in quantum subspace diagonalization methods, where a similar generalized eigenvalue problem is solved [90].
Using only 5 sample points along the whole PES, the PES for both 1-UCCSD and 2-UCCSD is reproduced to chemical accuracy (even slightly improving he 1-UCCSD wave function at dissociation). While we cannot test this for larger basis sets due to the exponential scaling in the number of simulated qubits, we cannot confirm that this is the case for larger basis sets, too, but since basis set effects were minor for WF-CCEVC, we believe this to be the case for EVC on a quantum computer too. We see how a threshold of $\epsilon = 10^{-3}$ for canonical orthonormalization suffices. Indeed, both for 1-UCCSD and 2-UCCSD, only a single eigenvector of the overlap matrix is discarded. When using many close sample points, we see that choosing a large value for $\epsilon = 10^{-3}$ reduces the method's extrapolation ability. This is because all but two eigenvectors are discarded. Using all eigenvectors ($\epsilon = 0$) increases the extrapolation range, with the 1-UCC extrapolation being qualitatively correct. Finally, although 2-UCC and 1-UCC both are chemically accurate in the interpolation regime, and the wave functions thus very similar, the 2-UCC extrapolation follows the 2-UCC trajectory, with the 1-UCC extrapolation following the 1-UCC trajectory. The 1-UCC wave function is qualitatively correct at dissociation, while the 2-UCC one is not, which we believe to be coincidental.

8.1.2 Sampling need

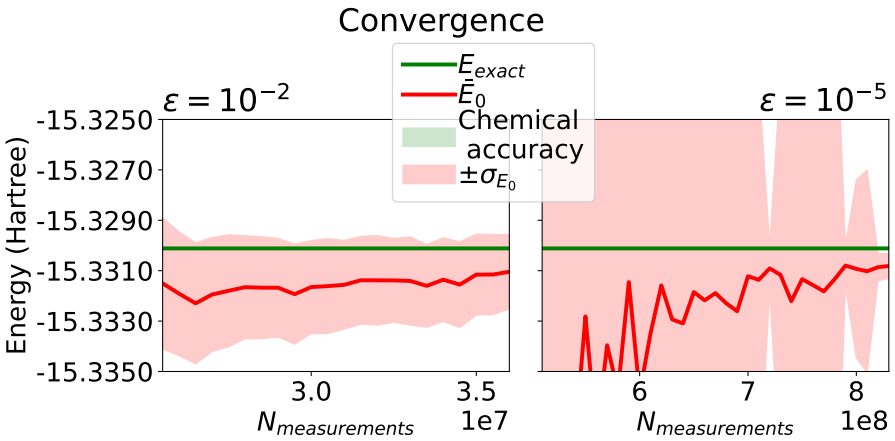


Figure 8.2: Convergence of the expectation value $\bar{E}_0(1.5)$ as numbers of measurements is increased for two values of ϵ. We used $N_B = 100$ bootstrapping samples, and at each iteration. Left: We increased the number of measurements of "the best" matrix element by $5 \cdot 10^5$, having initially measured each matrix element 10^6 times. We used a threshold $\epsilon = 10^{-2}$ in the canonical orthogonalization procedure. Right: We increased the number of measurements of "the best" matrix element by $1 \cdot 10^7$, having initially measured each matrix element $2 \cdot 10^7$ times. We used a threshold $\epsilon = 10^{-5}$ in the canonical orthogonalization procedure.

As discussed in the theory and method section, the matrices $\mathbf{H}^{\text{EVC}}$ and $\mathbf{S}^{\text{EVC}}$ need to be sampled on real quantum computers. We will now address the sampling requirements to obtain chemical accuracy for different cutoff parameters ϵ for both the interpolation and the extrapolation of the 2-UCC case described previously. As a convergence criteria, we demanded that σ_{E_0} (which is overestimated) is less than chemical accuracy upon convergence and that energy fluctuations between consecutive iterations should be below acc_{chem} for 20 consecutive increases. This is shown in figure 8.2.

We observe that the choice of the threshold is paramount. We see how the number of required measurements for $\epsilon = 10^{-2}$ is more than an order of magnitude smaller than the

one for $\epsilon = 10^{-5}$. Small cutoff values lead to a need to measure more often by several magnitudes, as even small fluctuations in the smallest non-cutoff eigenvalues s lead to large changes in the square root inverse of the diagonal $\mathbf{s}^{-\frac{1}{2}}$ which is calculated in canonical orthogonalization, and hence one gets unacceptable errors in the eigenvalues unless $\mathbf{S}^{\text{EVC}}$ is sampled very accurately. In figure 8.3, we see that the expectation values obtained at one geometry (here, $x = 1.5$ Bohr) can be used to obtain the full potential energy surface and only need to be sampled once, as we claimed in the methods section. Furthermore, we do not get extra accuracy along the PES when using a larger value for ϵ.

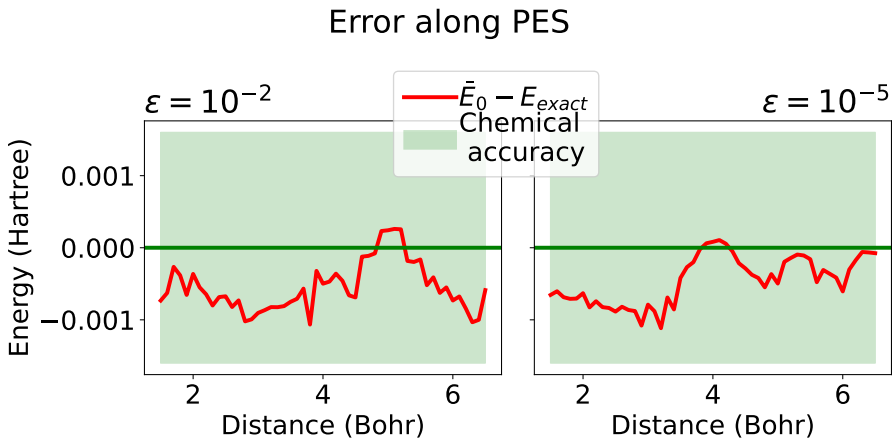


Figure 8.3: Expectation value $\bar{E}_0$ along the PES, using the sampled expectation values corresponding to the samples in figure 8.2. The Pauli matrix expectation values were obtained at $x = 1.5$ Bohr.

8.1.3 Smoothening of noisy potential energy surfaces

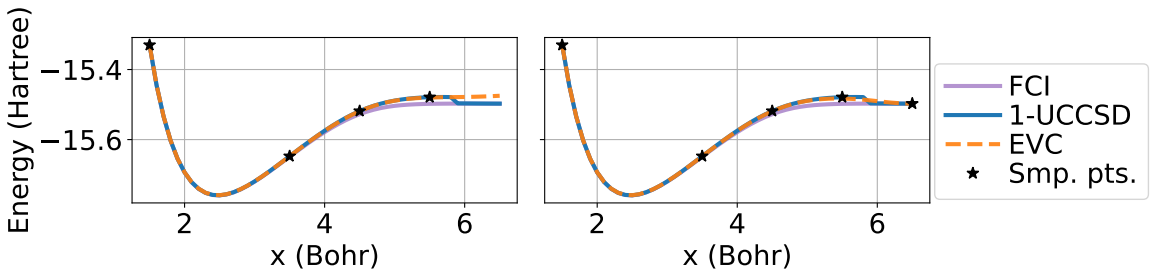


Figure 8.4: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes orbitals. The STO-6G basis set was used and the two core electrons were frozen. The Jordan-Wigner mapping without tapering was used. We used a 1-UCCSD reference. Observe the nonphysical "jump" at $x = 5.8$ Bohr. Left: Including only sample points from one continuous region. Right: Including sampling points from both regions.

Even on a noiseless quantum computer, it is possible that the PES obtained using the same ansatz changes discontinuously. This can happen when an optimization procedure finds a set of parameters $\vec{\theta}(\vec{R}_1)$ that differ a lot from those at a nearby point $\vec{\theta}(\vec{R}_2)$ (assuming an analytic change in the Hamiltonian) by, for example, jumping out of a local minimum and thus improving the energy, or when using ansätze that are hard to optimize. This is precisely what happened to us when simulating the linear symmetric stretch of BeH_2 using the Jordan-Wigner mapping without qubit tapering on 12 qubits. Eigenvector continuation can be used to overcome those problems by coupling the solutions. This is shown in figure 8.4.

The choice to make is which sample points to include. In this case, we obtain a qualitatively correct PES by ignoring the region after $x = 5.8$ Bohr, even though adding several sample points $x_i \in [5.8, 6.5]$ might remove the nonphysical shape in the right plot.

8.1.4 Comparison between general Procrustes and regular Procrustes orbitals

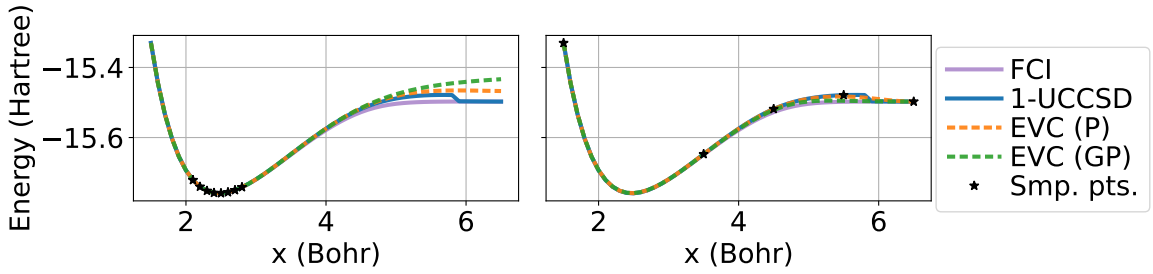


Figure 8.5: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes (P) or generalized Procrustes (GP) orbitals. The STO-6G basis set was used and the two core electrons were frozen. The Jordan-Wigner mapping without tapering was used. We used a 1-UCCSD reference. Observe the unphysical "jump" at $x = 5.8$ Bohr. Left: Tight sampling in equilibrium region. Right: Sampling along the whole PES, including beyond the jump.

We tested whether the use of generalized Procrustes orbitals as described in section 4.7.3 can give an advantage over the use of regular Procrustes orbitals. This requires that the Jordan-Wigner mapping is used an hence comes at a price, as also qubit tapering becomes harder. In figure 8.5, we compare the extrapolation and interpolation of generalized Procrustes orbitals with regular Procrustes orbitals.

Which method works better is generally inconclusive. Sampling only near the equilibrium region, Procrustes orbitals work better in this case, while along the whole PES, generalized Procrustes orbitals work better. There is an explanation why using Procrustes orbitals works better.

Using generalized Procrustes orbitals, unoccupied orbitals are mixed with occupied orbitals. Using Procrustes orbitals, the HF state at geometry $\vec{R}_1$ corresponds to the HF state at geometry $\vec{R}_2$. With generalized Procrustes orbitals, this is no longer the case when using the approach of section 4.7.3. In the limit where only the HF state is used for EVC at every sample geometry, using Procrustes orbitals would lead to an all-one overlap matrix, which is not the case for generalized Procrustes orbitals. When a geometry is dominated by the HF state, hence, using general Procrustes orbitals will lead to a state that changes faster along the geometry. Generalized Procrustes orbitals reduce the change in the coefficient matrix as a whole, which also leads to a lesser change in the Hamiltonian, but this does not mean that the change in the eigenstates is minimized. Figure 8.6 contains the overlap matrix for the right plot of figure 8.5, which confirms this, as the general Procrustes orbital overlap matrix has smaller off-diagonal elements compared to the Procrustes one. It should also be noted that the overlap matrix for general Procrustes orbitals has larger eigenvalues than the one for Procrustes orbitals, which follows from the smaller off-diagonal elements - from that point of view, general Procrustes orbitals reduce sampling need of the overlap matrix.

$$\mathbf{S}_P^{\mathrm{EVC}} = \begin{bmatrix} 1 & 0.979 & 0.883 & 0.711 & 0.467 \\ 0.979 & 1 & 0.949 & 0.795 & 0.512 \\ 0.883 & 0.949 & 1 & 0.939 & 0.614 \\ 0.711 & 0.795 & 0.939 & 1 & 0.678 \\ 0.467 & 0.512 & 0.614 & 0.678 & 1 \end{bmatrix} \quad \mathbf{S}_{GP}^{\mathrm{EVC}} = \begin{bmatrix} 1 & 0.959 & 0.806 & 0.511 & 0.197 \\ 0.959 & 1 & 0.928 & 0.676 & 0.302 \\ 0.806 & 0.928 & 1 & 0.895 & 0.497 \\ 0.511 & 0.676 & 0.895 & 1 & 0.657 \\ 0.197 & 0.302 & 0.497 & 0.657 & 1 \end{bmatrix}$$

Figure 8.6: The overlap matrices from the right plot in figure 8.5, where the left one (P) stands for Procrustes and the right one for generalized Procrustes (GP)

8.1.5 EVC with generalized Procrustes orbitals: BeH_2 insertion reaction

To see how well EVC works with general Procrustes orbitals method for systems where the HF state changes discontinuously, we considered again the BeH_2 insertion reaction.

As sample ansatz, we used the FCI wave function, as all ansätze implemented in `qiskit` converge to the triplet state at those geometries where its energy is lower. This is however not a problem, as existing methods, such as (U)CCSD, give results well below chemical accuracy at the sample points considered, too, and it is hence likely that using any of those wave functions as references would yield similar results. The resulting PES is shown in figure 8.7.

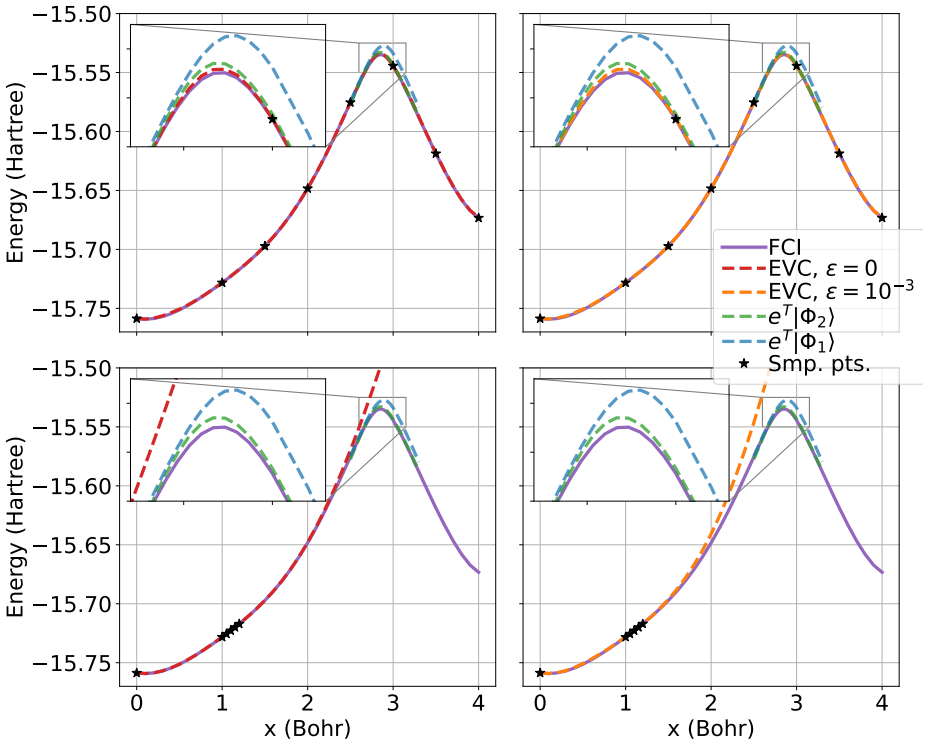


Figure 8.7: Potential energy curve of the BeH_2 insertion reaction. General Procrustes orbitals were used. We used FCI at sample points and the STO-6G basis set. Two core electrons were frozen using `Qiskit`'s `ActiveSpaceTransformer` [109]. For comparison, we also depict the CCSD solutions in the "difficult" region. Left: using no threshold $\epsilon = 0$. Right: Using a large threshold $\epsilon = 10^{-3}$.

The results essentially match with the results from the previous sections. CCSD fails qualitatively also in minimal basis. We see how EVC can be used for interpolation and "overcome" the difficult region $x \in [2.5, 3.0]$, where CCSD fails. We see that when the sample points are spread "far" apart, then discarding the smallest eigenvalues of the overlap matrix yields essentially the same results, as the two upper plots look identical. This is however not the case when sampling "tightly". The extrapolation strength of the method requires a large number of eigenvectors, and is strongly compromised when several eigenvectors are discarded. However, discarding those is necessary when trying to minimize the necessary number of sampling points on a quantum computer, as increasing the threshold s increases the number of required samples exponentially.

8.2 Discussion

By the discussion of section 4.7 and the results presented here, it is clear that running eigenvector continuation on a quantum computer is very attractive, as Pauli matrix measurements - and thus energy and overlap measurements - only need to be performed once to calculate the energy for the whole PES - however, $O(L^2)$ matrix elements need to be obtained. This is an advantage on quantum computers compared to regular computers - in WF-CCEVC one has one has to similarity transform L Hamiltonians at each geometry of interest. Furthermore, the generalized Procrustes approach makes it possible to use eigenvector continuation for problems that are not tractable on regular computers due to

a qualitative change in the HF reference, as basis change operations are easily implemented. It is not clear whether Procrustes orbitals or generalized Procrustes orbitals are the better method to use. Procrustes orbitals cannot always be used, but when the same HF state is an alright reference, the sample states have a large overlap and change less. On the other hand, on a quantum computer, the non-orthogonal basis functions lead to an increased sampling need by several magnitudes in order to obtain accurate energies when using EVC for extrapolation. The problem relating to the increased sampling need has been observed previously in Ref. [67]. Careful selection of the cutoff of the eigenvalues of the overlap matrix $\mathbf{S}^{\text{EVC}}$, $s \leq \epsilon \to 0$, needs to be performed. This is reflected in the fact that the number of measurements increase with L and decrease for larger ϵ. Both is not surprising as a larger sampling space can lead to near-singularities.
The sampling problem extends to the application of EVC outside of chemistry, too. Consider the problem shown in figure 2.1. The overlap matrix for 5 sample points is shown in figure 8.8 below. It is almost singular, with three eigenvalues below 10^{-5}. When using $\epsilon = 10^{-5}$ as threshold (and we observed previously that this increases the measurement requirements a lot already), the quality of the EVC eigenvalues deteriorates, which is shown in figure 9.1 in the appendix. This shows that EVC for extrapolation is not applicable on quantum computers, simply because sampling the overlap matrix becomes infeasible. We thus conclude that eigenvector continuation on a quantum computer is an extremely powerful interpolation tool, as only few wave functions need to be found and optimized and the accuracy of the overlap matrix $\mathbf{S}$ does not need to be so large, and because expectation values do not need to be sampled at every geometry when using Procrustes orbitals. Furthermore, eigenvector continuation can be used with general Procrustes orbitals to solve complicated multi-reference problems. As an extrapolation method, however, eigenvector continuation fails, as it is precisely the almost-vanishing eigenvectors that give the method its extrapolation strength, but those are prohibitively expensive to sample.

$$\mathbf{S}^{\text{EVC}} = \begin{bmatrix} 1 & 0.99999 & 0.99992 & 0.99974 & 0.99923 \\ 0.99999 & 1 & 0.99997 & 0.99985 & 0.99942 \\ 0.99992 & 0.99997 & 1 & 0.99995 & 0.99964 \\ 0.99974 & 0.99985 & 0.99995 & 1 & 0.99986 \\ 0.99923 & 0.99942 & 0.99964 & 0.99986 & 1 \end{bmatrix}$$

Figure 8.8: The overlap matrix from the example in figure 2.1 for five sample points, which has eigenvalues $\lambda_0 = 1.7 \cdot 10^{-13}, \lambda_1 = 1.9 \cdot 10^{-10}, \lambda_2 = 3.7 \cdot 10^{-7}, \lambda_3 = 9.8 \cdot 10^{-4}, \lambda_4 = 4.999$ *and is hence very expensive to sample on a quantum computer.*

Part V

Conclusion

In this thesis, we have analyzed how eigenvector continuation (EVC) can be applied to efficiently reproduce and improve the potential energy surface of molecules. Performing EVC with Hartree-Fock states, we found that very few Slater determinants can essentially reproduce the whole PES for relatively simple systems. Furthermore, it can "smoothen" the PES for systems where the HF energy has cusps. Including a number of determinants at each geometry that were generated as HF-solutions to a modified potential with a strengthened/weakened electron-repulsion, we found that qualitative and quantitative improvements are possible. This method is however unreliable and hard to improve further. Our main focus in this thesis was application of eigenvector Continuation to (U)CCSD wave functions. We found that the CC-reformulation of EVC, which we termed WF-CCEVC [3] (sec. 2.10), can be applied to quantum chemistry and yield promising results. We found that it can be used both as an interpolation tool, where it works essentially as a blackbox method, or in order to improve CCSD energies, in which case it requires knowledge of the system and why CCSD fails. Doing this, we obtained improved potential energy curves for complicated systems as BeH_2 and N_2. Furthermore, we developed AMP-CCEVC (sec. 4.3) and found that it can be used as a few-parameter representation of the wave function. In addition, we developed a way to use AMP-CCEVC in such a way that it can reduce the cost and time of solving the CCSD amplitude equations approximately.
We found that quantum computers have some particular advantages to the application of EVC. As change of basis operations are not prohibitively expensive on quantum computers, EVC on quantum computers can be applied to more systems than WF-CCEVC, including systems where the reference determinant changes discontinuously. In particular, the original variational formulation of EVC is restored. Furthermore, we found that expectation values only need to be sampled at one single geometry and are then available everywhere using EVC on a quantum computer. At the same time, we found that eigenvector continuation cannot be used when sampling points are "close", as this would increase sampling need on a quantum computer by several orders of magnitude.
As basis functions and thus the vector space spanned by them change at each nuclear position $\vec{R}$, we considered different ways to make the Hamiltonian change analytically. In order to do so, we compared several schemes to make the molecular orbitals and thus the second quantized Hamiltonian in a basis change smoothly. We found that using symmetric orthogonalization works well when one wants to change the MOs as little as possible in a least square sense, leading to wave functions that share a substantial amount of overlap with the correct HF wave function. Orbitals obtained by solving the orthogonal Procrustes problem are a powerful tool when the reference wave function should be a Hartree-Fock wave function. Natural orbitals can be used for that, too, even though both require that the HF wave function is analytic.

Outlook

The research performed in this thesis has by no means depleted the possibilities that eigenvector continuation might have in computational quantum chemistry. First of all, with the minor interlude of the "strength parameter λ", we have not considered other parameters than the molecular geometry. There are hence possibilities to use eigenvector continuation for, for example, the application of external electric and magnetic fields. With magnetic fields, the Hamiltonian is linear in some strength parameter, but using e.g. London atomic orbitals, extra complications arise. Many of the applications here can be considered as a "proof of concept" without fully exploring computational and numerical optimizations. As such, we have not considered how concepts such as tensor hypercontractions [110] can be used to effectively reduce the size of the amplitude matrices in AMP-EVC and CC-EVC, and the time improvement using AMP-CCEVC was nothing more but a p. Another point should be made about the use of Procrustes orbitals. Procrustes orbitals are just a special choice of a more general problem - orbitals could be weighted differently. It is possible that eigenvector continuation can also be used not only with Natural orbitals, but also with pair natural orbitals

[111] or local natural orbitals [112, 113]. Furthermore, we have not seen the big computational savings as Ekstrøm and Hagen [3] in the WF-CCEVC approach, because the shape of the geometry-dependent Hamiltonian in quantum chemistry still required to calculate the projection errors, which scales as $O(M^6)$. Unlike standard CC theory, however, this is not an iterative procedure, as no new cluster operator needs to be found. Because the same set of operators $\hat{T}(\vec{R}_i)$ is similarity transformed at each geometry $\vec{R}$, it is possible that machine learning might be used to get approximate projection errors for each $\hat{T}(\vec{R}_i)$. Furthermore, we might use WF-CCEVC to write the wave function as a linear combination of two approximations to CCSD that do not lead to qualitatively wrong energies for systems with strong static correlation, such as CCSD0 and CCSD1 [108], in the hope of achieving a qualitatively correct lower energy. It might even be of interest to develop a variant of WF-CCEVC which can be applied to multireference CCSD wave functions. For the Mk-MRCC ansatz, Lambda equations are available [114].
Another question to be answered is how the correct sampling points for EVC are chosen, independently of the method. This number should be as small as possible to keep computational cost down, while still yielding correct results. For some chemical problems, the correct range might be deduced from the problem at hand, such as nitrogen.
From a mathematical point of view, studying the mathematical properties of EVC when sample points are not close and their span deviates from that of the first n derivatives of $|\Psi(0)\rangle$, as has been pointed out in Ref. [2], is an important task. On quantum computers, it is also relevant to consider the canonical orthogonalization cutoff ϵ in more detail.

The AMP-CCEVC approach and the corresponding parameter reduction has a natural extension to learning the best parameters $\vec{\theta}$ for some ansatz on a quantum computer by writing the parameter $\vec{\theta}$ as a linear combination of sampling parameters at different geometries. Even though there are no amplitude equations to be solved, the parameter reduction approach can just as well be applied the stationarity condition $\frac{\partial E}{\partial c_n} = 0$ for $n = 1, \ldots, L$. Thus, we might be able to obtain approximate parameters at reduced cost. The parameter reduction might furthermore be applicable for general CC calculations on top of DIIS, approximating the "best" linear combination of different iterations of cluster operators and thus reducing the total number of iterations.

Part VI

Appendix

Chapter 9

Figures, Tables and miscellaneous

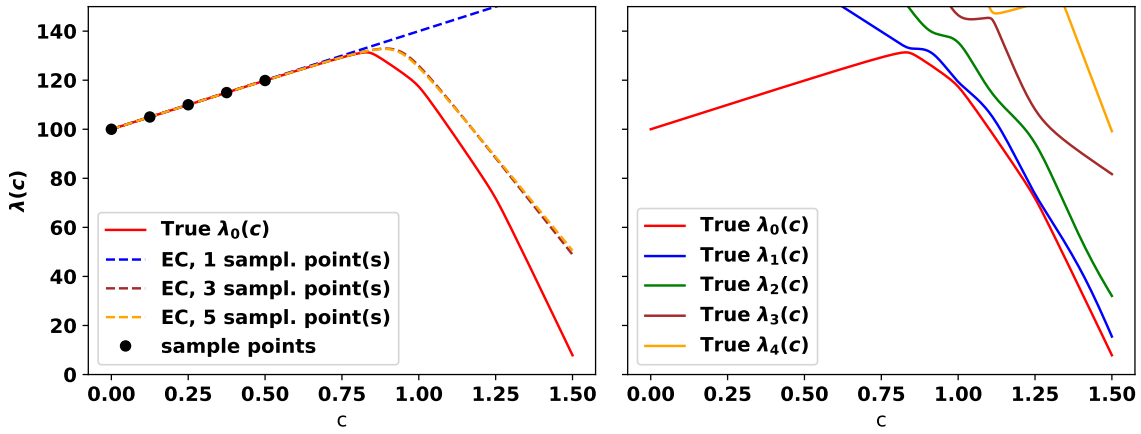


Figure 9.1: EVC applied to the problem described in figure 2.1 when putting an eigenvalue cutoff of the overlap matrix with value 10^{-5}, discarding the lowest three eigenvalues of the overlap matrix. We see how adding additional sample points makes no qualitative difference and the eigenvalues of the matrix $\mathbf{M}(c)$ differ qualitatively and strongly from the EVC eigenvalues.

9.1 Toy model for Eigenvector Continutation

This model corresponds to model 3 explained in the Supplemental Material of Ref. [2]. The matrix function is given by $\mathbf{M}(c) = \mathbf{M}_0 + c\mathbf{M}_1$, where $\mathbf{M}_0$, $\mathbf{M}_1$ are 500×500 matrices. The non-zero elements are given by

$$
\begin{aligned}
(\mathbf{M}_0)_{n,n} &= 100n \quad \text{for } n = 1, \ldots, 500 \\
(\mathbf{M}_1)_{1,1} &= 40 \\
(\mathbf{M}_1)_{2,2} &= -80 \\
(\mathbf{M}_1)_{3,3} &= -180 \\
(\mathbf{M}_1)_{4,4} &= -260 \\
(\mathbf{M}_1)_{5,5} &= -320 \\
(\mathbf{M}_1)_{6,6} &= -335 \\
(\mathbf{M}_1)_{n,n} &= 50n \quad \text{for } n = 7, \ldots, 500 \\
(\mathbf{M}_1)_{n,n+1} &= (\mathbf{M}_1)_{n+1,n} = 2 \quad \text{for } n = 1, \ldots, 499 \\
(\mathbf{M}_1)_{n,n+2} &= (\mathbf{M}_1)_{n+2,n} = 5 \quad \text{for } n = 1, \ldots, 498 \\
(\mathbf{M}_1)_{n,n+3} &= (\mathbf{M}_1)_{n+3,n} = 5 \quad \text{for } n = 1, \ldots, 497
\end{aligned}
$$

Chapter 10

Proofs

10.1 Orthogonal Procrustes problem

Consider the optimization problem

$$\mathbf{Q} = \arg\min_{\mathbf{\Omega}} \|\mathbf{A\Omega} - \mathbf{B}\|_F \quad \text{subject to} \quad \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{1} \tag{10.1}$$

where $\mathbf{A}, \mathbf{B}$ are real $n \times m$ matrices and $\mathbf{\Omega}$ an orthogonal $m \times m$ matrix. This problem is, by the non-negativity of a norm, identical to

$$\mathbf{Q} = \arg\min_{\mathbf{\Omega}} \|\mathbf{A\Omega} - \mathbf{B}\|_F^2 = \arg\min_{\mathbf{\Omega}} \sum_{i=1}^{m} \|(\mathbf{A\Omega} - \mathbf{B})_i\|^2 \quad \text{subject to} \quad \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{1} \tag{10.2}$$

where the sum goes over all columns, and we used the index $_i$ to indicate the ith column. We will find the solution to a more general problem where we allow a different weighting of the column differences

$$\mathbf{Q} = \arg\min_{\mathbf{\Omega}} \sum_{i=1}^{m} w_i \|(\mathbf{A\Omega} - \mathbf{B})_i\|^2 \quad \text{subject to} \quad \mathbf{\Omega}^T\mathbf{\Omega} = \mathbf{1} \tag{10.3}$$

where $w_i \geq 0$ for all i. This problem is, in a sense, a generalization of both Wahba's problem, where $\mathbf{Q} \in SO(m)$ is a rotation matrix, and the orthogonal Procrustes problem, where $w_i = 1$ for all i.

Theorem 10.1.1. *The solution to eq. 10.3 is given by* $\mathbf{Q} = \mathbf{U}\mathbf{V}^T$, *where* $\mathbf{U}$ *and* $\mathbf{V}$ *are obtained from the SVD of* $\mathbf{A}^T\mathbf{B}\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, *where* $\mathbf{W} = diag[w_1, \ldots, w_m]$.

Proof.

$$
\begin{aligned}
\mathbf{Q} &= \arg\min_{\mathbf{\Omega}} \sum_{i=1}^{m} w_i \|(\mathbf{A\Omega} - \mathbf{B})_i\|^2 \\
&= \arg\min_{\mathbf{\Omega}} \sum_{i}^{m} w_i \left[\langle (\mathbf{A\Omega})_i, (\mathbf{A\Omega})_i \rangle + \langle \mathbf{B}_i, \mathbf{B}_i \rangle - 2 \langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle\right] \\
&= \text{const.} \; - 2 \arg\min_{\mathbf{\Omega}} \sum_{i}^{m} w_i \langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle \\
&= \arg\max_{\mathbf{\Omega}} \sum_{i}^{m} w_i \langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle \\
&= \arg\max_{\mathbf{\Omega}} \sum_{i}^{m} \langle (\mathbf{A\Omega})_i, w_i \mathbf{B}_i \rangle \\
&= \arg\max_{\mathbf{\Omega}} \langle \mathbf{A\Omega}, \mathbf{BW} \rangle \\
&= \arg\max_{\mathbf{\Omega}} \operatorname{Tr}\left((\mathbf{A\Omega})^T \mathbf{BW} \right) \\
&= \arg\max_{\mathbf{\Omega}} \operatorname{Tr}\left(\mathbf{\Omega}^T \mathbf{A}^T \mathbf{BW} \right) \\
&= \arg\max_{\mathbf{\Omega}} \langle \mathbf{\Omega}, \mathbf{A}^T \mathbf{BW} \rangle \\
&= \arg\max_{\mathbf{\Omega}} \langle \mathbf{\Omega}, \mathbf{U\Sigma V}^T \rangle \\
&= \arg\max_{\mathbf{\Omega}} \langle \mathbf{U}^T \mathbf{\Omega V}, \mathbf{\Sigma} \rangle \\
&\Rightarrow \mathbf{Q} = \mathbf{UV}^T
\end{aligned}
\tag{10.4}
$$

where we made use of the fact that that the identity matrix is the matrix is closest to a diagonal matrix with nonnegative entries [115]. This proof was inspired by a similar proof by deRuiter and Forbes [116]. □

10.2 Analyticity of general Procrustes orbitals

Theorem 10.2.1. *General Procrustes orbitals are analytic whenever the overlap matrix* $\mathbf{S}(\vec{R})$ *is analytic and non-singular.*

Proof. Let $\mathbf{B} \in \mathbb{R}^{n \times n}$ be a matrix corresponding to the reference coefficient matrix at $\vec{R}_{\text{ref}}$. In the case of molecular orbitals, where $\mathbf{S}(\vec{R}) \in \mathbb{R}^{n \times n}$ is the overlap matrix as a function of a parameter $\vec{R}$, by Lemma 4.2.1, we can always choose an arbitrary reference which corresponds to any state with orthonormal MOs. Let us choose $\mathbf{S}^{-\frac{1}{2}}(\vec{R})$, which is analytic everywhere except for the obvious singularities (placing two atoms at the same position). Consider now the Procrustes problem

$$
\mathbf{Q}(\vec{R}) = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{S}^{-\frac{1}{2}}(\vec{R}) \mathbf{\Omega} - \mathbf{B} \right\|_F \tag{10.5}
$$

with solution

$$
\mathbf{Q}(\vec{R}) = \mathbf{U}(\vec{R}) \mathbf{V}^{\dagger}(\vec{R}) \tag{10.6}
$$

where

$$
\mathbf{U}(\vec{R}) \mathbf{\Sigma}(\vec{R}) \mathbf{V}^{\dagger}(\vec{R}) = \left(\mathbf{S}^{-\frac{1}{2}}(\vec{R}) \right)^{\dagger} \mathbf{B} = \mathbf{S}^{-\frac{1}{2}}(\vec{R}) \mathbf{B} \tag{10.7}
$$

now, because $\mathbf{S}^{-\frac{1}{2}}(\vec{R})$ is analytic, so is $\mathbf{Y}(\vec{R}) = \mathbf{S}^{-\frac{1}{2}}(\vec{R})\mathbf{B}$. Furthermore, as $\mathbf{Y}(\vec{R})\mathbf{Y}^{\dagger}(\vec{R})$ and $\mathbf{Y}^{\dagger}(\vec{R})\mathbf{Y}(\vec{R})$ are Hermitian and analytic, their eigenvectors, the columns of $\mathbf{U}(\vec{R})$, $\mathbf{V}(\vec{R})$, respectively, can be chosen to be analytic. Thus, $\mathbf{Q}(\vec{R})$ is analytic, and so is $\mathbf{Q}(\vec{R})\mathbf{S}^{-\frac{1}{2}}(\vec{R})$. Hence, the change in the general Procrustes MOs is analytic. □

10.3 Generalized Slater-Condon rules

When we have two Slater determinants, $|{}^w\Psi\rangle$ and $|{}^x\Psi\rangle$ that are not orthogonal, the Slater-Condon rules cannot be used. In order to calculate the overlap $\langle {}^w\Psi|{}^x\Psi\rangle$ and the matrix element $\langle {}^w\Psi|\,\hat{H}\,|{}^x\Psi\rangle$, it is hence necessary to find a set of rules which can express these in terms of the basis. These rules are called the *Generalized Slater-Condon rules* and we will re-derive them here in a somewhat different way than in Ref. [11], where first quantization was used. To do so, we need the following lemma:

Lemma 10.3.1. *Let there be two sets of orthonormal molecular orbitals* $\{{}^w\phi_i\}_{i=1}^{M_1}$, $\{{}^x\phi_i\}_{i=1}^{M_2}$ *that are not mutually orthonormal, and* $|{}^w\phi_i\rangle = a_i^\dagger\,|\rangle$, $|{}^x\phi_i\rangle = b_i^\dagger\,|\rangle$. *Then the anticommutator is given by*

$$[a_i, b_j^\dagger]_+ = \langle|\, a_i b_j^\dagger\,|\rangle = \langle {}^w\phi_i|{}^x\phi_j\rangle = {}^{wx}S_{ij} \tag{10.8}$$

Proof. We start by expanding $|{}^x\phi_i\rangle$ in a complete basis. Let the first M_1 elements of that basis correspond to the elements $|{}^w\phi_i\rangle$. That is

$$b_j^\dagger\,|\rangle = |{}^x\phi_j\rangle = \sum_{i=1}^{\infty} \langle {}^w\phi_i|{}^x\phi_j\rangle\,|{}^w\phi_i\rangle = \sum_{i=1}^{\infty} {}^{wx}S_{ij}\,|{}^w\phi_i\rangle = \sum_{i=1}^{\infty} {}^{wx}S_{ij} a_i^\dagger\,|\rangle \tag{10.9}$$

Thus, $b_j^\dagger = \sum_{i=1}^{\infty} {}^{wx}S_{ij} a_i^\dagger$. Then

$$[a_i, b_j^\dagger]_+ = \sum_{k=1}^{\infty} {}^{wx}S_{kj}[a_i, a_k^\dagger]_+ = \sum_{k=1}^{\infty} {}^{wx}S_{kj}\left(a_i a_k^\dagger + a_k^\dagger a_i\right) = \sum_{k=1}^{\infty} {}^{wx}S_{kj}\delta_{ik} = {}^{wx}S_{ij} \tag{10.10}$$

□

10.3.1 The overlap between two Slater determinants in different basis

We will now prove eq. (1.53), p. 14.

Lemma 10.3.2. *The overlap between nonorthogonal Slater determinants is given by* $\langle {}^w\Psi|{}^x\Psi\rangle = |\mathbf{S}|$.

Proof. This proof is by induction on the number of particles. Clearly, in the 1×1 case, this holds true. Assume this holds true for n particles. Let $|{}^x\Psi\rangle, |{}^w\Psi\rangle$ be $n+1$ particle Slater determinants in different bases. Then

$$\begin{aligned}
\langle {}^w\Psi|{}^x\Psi\rangle &= \langle|\, a_{n+1}\ldots\left(a_1 b_1^\dagger\right)\ldots b_{n+1}^\dagger\,|\rangle \\
&= \langle|\, a_{n+1}\ldots a_2\left({}^{wx}S_{11} - b_1^\dagger a_1\right) b_2^\dagger\ldots b_{n+1}^\dagger\,|\rangle \\
&= {}^{wx}S_{11}\,\langle|\, a_{n+1}\ldots a_2 b_2\ldots b_{n+1}\,|\rangle - \langle|\, a_{n+1}\ldots a_2 b_1^\dagger\left({}^{wx}S_{12} - b_2^\dagger a_1\right) b_3^\dagger\ldots b_{n+1}^\dagger\,|\rangle \\
&= \sum_{i=1}^{n+1} {}^{wx}S_{1,i}(-1)^{1+i}\,\langle|\, a_{n+1}\ldots a_2 b_1^\dagger\ldots b_{i-1}^\dagger a_1 b_i^\dagger b_{i+1}^\dagger b_{n+1}\,|\rangle
\end{aligned} \tag{10.11}$$

where, for $i=1$, $b_1^\dagger\ldots b_{i-1}^\dagger = 1$, and for $i=2$, $b_1^\dagger\ldots b_{i-1}^\dagger = b_1^\dagger$. By the induction hypothesis, the inner products are given by the determinants of $\mathbf{S}$ when removing the first row and the ith column of $\mathbf{S}$

$$\langle|\, a_{n+1}\ldots a_2 b_1^\dagger\ldots b_{i-1}^\dagger a_1 b_i^\dagger b_{i+1}^\dagger b_{n+1}\,|\rangle = {}^{wx}S_{1,i}^{(1)} \tag{10.12}$$

Recognizing ${}^{wx}S_{1,i}(-1)^{1+i}$ as elements of the cofactor matrix and the last line of eq. (10.11) as the Laplace expansion of $|\mathbf{S}|$ finishes the proof. □

On a side note, this formula also applies to calculating the nuclear repulsion part of two Slater determinants, as

$$\langle {}^w\Psi|h_{\text{nuc}}|{}^x\Psi\rangle = h_{\text{nuc}}\,\langle {}^w\Psi|{}^x\Psi\rangle \tag{10.13}$$

10.3.2 The one-body matrix element

We will here use a second-quantized approach. First, observe that we can rewrite the Hamiltonian $\hat{H}_1$ in second quantization as

$$\hat{H}_1 = \sum_{p,q} h_{pq} b_p^\dagger b_q = \sum_{p,q} h_{pq} \sum_{r=1}^{\infty} S_{rp} a_r^\dagger b_q = \sum_{rq} \left(\sum_p S_{rp} h_{pq} \right) a_r^\dagger b_q = \sum_{rq} {}^{wx}h_{rq} a_r^\dagger b_q \tag{10.14}$$

where ${}^{wx}h_{rq} = \langle {}^w\phi_r | \hat{h} | {}^x\phi_q \rangle$, and we used the same infinite basis as in the previous section. The one-body matrix element then reads

$$\begin{aligned}
\langle {}^w\Psi | \hat{H}_1 | {}^x\Psi \rangle &= \sum_{rq} \langle | \, a_N \dots a_1 \, {}^{wx}h_{rq} a_r^\dagger b_q b_1^\dagger \dots b_N^\dagger \, | \rangle \\
&= \sum_{ij}^{N} \langle | \, a_N \dots a_1 \, {}^{wx}h_{ij} a_i^\dagger b_j b_1^\dagger \dots b_N^\dagger \, | \rangle \\
&= \sum_{ij}^{N} (-1)^{i+j} \, {}^{wx}h_{ij} \, \langle | \, a_N \dots a_{i+1} a_{i-1} \dots a_1 b_1^\dagger \dots b_{j-1}^\dagger b_{j+1}^\dagger \dots b_N^\dagger \, | \rangle \\
&= \sum_{ij}^{N} (-1)^{i+j} \, {}^{wx}h_{ij} \, {}^{wx}S_{i,j}^{(1)}
\end{aligned} \tag{10.15}$$

10.3.3 The two-body matrix element

Finding a formula for the two-body matrix element is essentially the same as calculating the one-body matrix element. We write

$$\hat{H}_2 = \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, b_p^\dagger b_q^\dagger b_r b_s = \frac{1}{4} \sum_{pqrs} {}^{wx}\langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger b_r b_s \tag{10.16}$$

where the sums for p, q go to infinity (but that does not matter, hence it is implicit).

$$\begin{aligned}
&\langle {}^w\Psi | \hat{H}_2 | {}^x\Psi \rangle = \frac{1}{4} \sum_{pqrs} \langle | \, a_N \dots a_1 \, {}^{wx}\langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger b_r b_s b_1^\dagger \dots b_N^\dagger \, | \rangle \\
&= \frac{1}{4} \sum_{ijkl}^{N} {}^{wx}\langle ij \| kl \rangle \, \langle | \, a_N \dots a_1 a_i^\dagger a_j^\dagger b_k b_l b_1^\dagger \dots b_N^\dagger \, | \rangle \\
&= \frac{1}{4} \sum_{ij}^{N} (-1)^{i+j+k+l} \, {}^{wx}\langle ij \| kl \rangle \, \langle | \, a_N \dots a_{i+1} a_{i-1} \dots a_{j+1} a_{j-1} \dots a_1 b_1^\dagger \dots b_{k-1}^\dagger b_{k+1}^\dagger \dots b_{l-1}^\dagger b_{l+1}^\dagger \dots b_N^\dagger \, | \rangle \\
&= \frac{1}{4} \sum_{ijkl}^{N} (-1)^{i+j+k+l} \, {}^{wx}\langle ij \| kl \rangle \, {}^{wx}S_{ij,kl}^{(2)}
\end{aligned} \tag{10.17}$$

where the matrix with elements $(-1)^{i+j+k+l} \, {}^{wx}\langle ij \| kl \rangle \, {}^{wx}S_{ij,kl}^{(2)}$ is the second-order cofactor matrix.

10.4 Proofs regarding eigenvector continuation

10.4.1 Overlap between EVC-CC-states

We will here prove the first part of eq. 2.13 for the overlap. Calculating the matrix elements for the Hamiltonian follows the same pattern.

$$\langle \tilde{\Psi} | \Psi \rangle = \langle \Phi^{\mathrm{SD}} | \left[\mathbb{1} + \hat{\Lambda}' \right] e^{\hat{X}} | \Phi^{\mathrm{SD}} \rangle \tag{10.18}$$

$$= \langle \Phi^{\mathrm{SD}} | \left[\mathbb{1} + \hat{\Lambda}'_1 + \hat{\Lambda}'_2 \right] \left[\mathbb{1} + \hat{X} + \frac{1}{2}\hat{X}^2 + \frac{1}{6}\hat{X}^3 + \dots \right] | \Phi^{\mathrm{SD}} \rangle \tag{10.19}$$

$$= \langle \Phi^{\mathrm{SD}} | \left[\mathbb{1} + \hat{\Lambda}'_1 + \hat{\Lambda}'_2 \right] \left[\mathbb{1} + \hat{X}_1 + \hat{X}_2 + \frac{1}{2}\hat{X}_1^2 \right] | \Phi^{\mathrm{SD}} \rangle \tag{10.20}$$

$$= \langle \Phi^{\mathrm{SD}} | \Phi^{\mathrm{SD}} \rangle + \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_1 \hat{X}_1 | \Phi^{\mathrm{SD}} \rangle + \frac{1}{2} \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_1^2 | \Phi^{\mathrm{SD}} \rangle + \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_2 | \Phi^{\mathrm{SD}} \rangle \tag{10.21}$$

$$= 1 + \sum_{iajb} \lambda_i^{a\prime} x_j^b \langle \Phi^{\mathrm{SD}} | \hat{a}_i^\dagger \hat{a}_a \hat{a}_b^\dagger \hat{a}_j | \Phi^{\mathrm{SD}} \rangle + \frac{1}{2} \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_1^2 | \Phi^{\mathrm{SD}} \rangle + \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_2 | \Phi^{\mathrm{SD}} \rangle \tag{10.22}$$

$$= 1 + \sum_{iajb} \lambda_i^{a\prime} x_j^b \overbrace{\hat{a}_i^\dagger \hat{a}_a \hat{a}_b^\dagger \hat{a}_j} + \frac{1}{2} \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_1^2 | \Phi^{\mathrm{SD}} \rangle + \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_2 | \Phi^{\mathrm{SD}} \rangle \tag{10.23}$$

$$= 1 + \sum_{ia} \lambda_i^{a\prime} x_i^a + \frac{1}{2} \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_1^2 | \Phi^{\mathrm{SD}} \rangle + \langle \Phi^{\mathrm{SD}} | \hat{\Lambda}'_2 \hat{X}_2 | \Phi^{\mathrm{SD}} \rangle \tag{10.24}$$

$$= 1 + \sum_{ia} \lambda_i^{a\prime} x_i^a + \sum_{ijklabcd} \lambda_{ij}^{ab\prime} \left(\frac{1}{8} x_k^c x_l^d + \frac{1}{16} x_{kl}^{cd} \right) \langle \Psi^{\mathrm{SD}} | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k | \Psi^{\mathrm{SD}} \rangle \tag{10.25}$$

the last expression can also be resolved using Wick's theorem, giving

$$\langle \Psi^{\mathrm{SD}} | \hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k | \Psi^{\mathrm{SD}} \rangle \tag{10.26}$$

$$= \overbrace{\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k} + \overbrace{\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k} + \overbrace{\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k} + \overbrace{\hat{a}_i^\dagger \hat{a}_j^\dagger \hat{a}_b \hat{a}_a \hat{a}_c^\dagger \hat{a}_d^\dagger \hat{a}_l \hat{a}_k} \tag{10.27}$$

$$= \delta_{il}\delta_{jk}\delta_{bc}\delta_{ad} - \delta_{ik}\delta_{jl}\delta_{bc}\delta_{ad} - \delta_{il}\delta_{jk}\delta_{bd}\delta_{ac} + \delta_{ik}\delta_{jl}\delta_{bd}\delta_{ac} \tag{10.28}$$

Using the antisymmetry of the λ_{ij}^{ab} and relabeling of dummy variables, we proved eq. 2.13. Going from eq. 10.19 to eq. 10.20, we used that $\langle \Phi^{\mathrm{SD}} | \left[1 + \hat{\Lambda}'_1 + \hat{\Lambda}'_2 \right]$ is at most a doubly excited state that has zero overlap with the triple excitation operators $\hat{X}_1\hat{X}_2$, $\hat{X}_1^3$, $\hat{X}_3$ and those of higher order. Similar excitation level arguments were used going from eq. 10.20 to eq. 10.21.

Bibliography

1. Frame, D. *et al.* Eigenvector Continuation with Subspace Learning. *Phys. Rev. Lett.* **121,** 032501. https://link.aps.org/doi/10.1103/PhysRevLett.121.032501 (3 July 2018).
2. Sarkar, A. & Lee, D. Convergence of Eigenvector Continuation. *Phys. Rev. Lett.* **126,** 032501. https://link.aps.org/doi/10.1103/PhysRevLett.126.032501 (3 Jan. 2021).
3. Ekström, A. & Hagen, G. Global Sensitivity Analysis of Bulk Properties of an Atomic Nucleus. *Phys. Rev. Lett.* **123,** 252501. https://link.aps.org/doi/10.1103/PhysRevLett.123.252501 (25 Dec. 2019).
4. Peruzzo, A. *et al.* A variational eigenvalue solver on a photonic quantum processor. *Nature Communications* **5,** 4213. ISSN: 2041-1723. https://doi.org/10.1038/ncomms5213 (July 2014).
5. Szabo, A. & Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* First (Dover Publications, Inc., 1982).
6. Allen, M. P. & Tildesley, D. J. *Computer simulation of liquids* (Oxford university press, 2017).
7. Gross, E., Runge, E. & Heinonen, O. *Many-Particle Theory* (Adam Higler, Bristol, Philadelphia and New York, 1991).
8. Helgaker, T., Jørgensen, P. & Olsen, J. *Molecular Electronic Structure Theory* (John Wiley & Sons, LTD, Chichester, 2000).
9. Crawford, T. D. & Schaefer III, H. F. in *Reviews in Computational Chemistry* 33–136 (John Wiley & Sons, Ltd, 2000). ISBN: 9780470125915. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470125915.ch2. https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470125915.ch2.
10. Löwdin, P.-O. in (ed Löwdin, P.-O.) 185–199 (Academic Press, 1970). https://www.sciencedirect.com/science/article/pii/S0065327608603391.
11. Mayer, I. *Simple Theorems, Proofs, and Derivations in Quantum Chemistry* en (ed Mezey, P. G.) ISBN: 9781441933898. http://link.springer.com/10.1007/978-1-4757-6519-9 (2021) (Springer US, Boston, MA, 2003).
12. Echenique, P. & Alonso, J. L. A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry. *Molecular Physics* **105,** 3057–3098. ISSN: 1362-3028. http://dx.doi.org/10.1080/00268970701757875 (Dec. 2007).
13. Lykos, P. & Pratt, G. W. Discussion on The Hartree-Fock Approximation. *Rev. Mod. Phys.* **35,** 496–501. https://link.aps.org/doi/10.1103/RevModPhys.35.496 (3 July 1963).
14. Hiscock, H. G. & Thom, A. J. W. Holomorphic Hartree–Fock Theory and Configuration Interaction. *Journal of Chemical Theory and Computation* **10,** 4795–4800. ISSN: 1549-9618. https://doi.org/10.1021/ct5007696 (Nov. 2014).
15. Burton, H. G. A., Gross, M. & Thom, A. J. W. Holomorphic Hartree–Fock Theory: The Nature of Two-Electron Problems. *Journal of Chemical Theory and Computation* **14,** 607–618. ISSN: 1549-9626. http://dx.doi.org/10.1021/acs.jctc.7b00980 (Jan. 2018).

16. Thouless, D. Stability conditions and nuclear rotations in the Hartree-Fock theory. en. *Nuclear Physics* **21,** 225–232. ISSN: 00295582. https://linkinghub.elsevier.com/retrieve/pii/0029558260900481 (2021) (Nov. 1960).

17. Harsha, G., Shiozaki, T. & Scuseria, G. E. On the difference between variational and unitary coupled cluster theories. *The Journal of Chemical Physics* **148,** 044107. eprint: https://doi.org/10.1063/1.5011033. https://doi.org/10.1063/1.5011033 (2018).

18. Evangelista, F. A. Alternative single-reference coupled cluster approaches for multireference problems: The simpler, the better. *The Journal of Chemical Physics* **134,** 224102. eprint: https://doi.org/10.1063/1.3598471. https://doi.org/10.1063/1.3598471 (2011).

19. Zhao, J. & Scuseria, G. E. *Drudge/Gristmill* https://tschijnmo.github.io/drudge/ (2021).

20. Stanton, J. F., Gauss, J., Watts, J. D. & Bartlett, R. J. A direct product decomposition approach for symmetry exploitation in many-body methods. I. Energy calculations. *The Journal of Chemical Physics* **94,** 4334–4345 (1991).

21. Taube, A. G. & Bartlett, R. J. New perspectives on unitary coupled-cluster theory. *International Journal of Quantum Chemistry* **106,** 3393–3401. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.21198. https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.21198 (2006).

22. Löwdin, P.-O. Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction. *Phys. Rev.* **97,** 1474–1489. https://link.aps.org/doi/10.1103/PhysRev.97.1474 (6 Mar. 1955).

23. Shavitt, I., Rosenberg, B. J. & Palalikit, S. Comparison of configuration interaction expansions based on different orbital transformations. *International Journal of Quantum Chemistry* **10,** 33–46. eprint: https://onlinelibrary-wiley-com.ezproxy.uio.no/doi/pdf/10.1002/qua.560100804. https://onlinelibrary-wiley-com.ezproxy.uio.no/doi/abs/10.1002/qua.560100804 (1976).

24. Taube, A. G. & Bartlett, R. J. Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory. *Collection of Czechoslovak chemical communications* **70,** 837–850 (2005).

25. Hay, P. J. On the calculation of natural orbitals by perturbation theory. *The Journal of Chemical Physics* **59,** 2468–2476. eprint: https://doi.org/10.1063/1.1680359. https://doi.org/10.1063/1.1680359 (1973).

26. Burton, H. *Holomorphic Hartree-Fock Theory: Moving Beyond the Coulson-Fischer Point* PhD thesis (University of Cambridge, Feb. 2020). https://doi.org/10.17863/CAM.55079.

27. Nite, J. M. & Jiménez-Hoyos, C. A. *Efficient Multi-Configurational Wavefunction Method with Dynamical Correlation Using Non-Orthogonal Configuration Interaction Singles and Doubles (NOCISD)* Department of Chemistry, Wesleyan University. Dec. 2019. https://doi.org/10.26434/chemrxiv.11369646.v1.

28. Burton, H. G. A. Generalized nonorthogonal matrix elements: Unifying Wick's theorem and the Slater–Condon rules. *The Journal of Chemical Physics* **154,** 144109 (2021).

29. Kato, T. *Perturbation theory for linear operators; 2nd ed.* https://cds.cern.ch/record/101545 (Springer, Berlin, 1976).

30. Demol, P. *et al.* Improved many-body expansions from eigenvector continuation. *Phys. Rev. C* **101,** 041302. https://link.aps.org/doi/10.1103/PhysRevC.101.041302 (4 Apr. 2020).

31. Löwdin, P.-O. Quantum theory of cohesive properties of solids. *Advances in Physics* **5,** 1–171. eprint: https://doi.org/10.1080/00018735600101155. https://doi.org/10.1080/00018735600101155 (1956).

32. Hochstenbach, M. E., Mehl, C. & Plestenjak, B. Solving Singular Generalized Eigenvalue Problems by a Rank-Completing Perturbation. *SIAM Journal on Matrix Analysis and Applications* **40,** 1022–1046. ISSN: 1095-7162. http://dx.doi.org/10.1137/18M1188628 (Jan. 2019).

33. Feynman, R. P. Simulating physics with computers. *International Journal of Theoretical Physics* **21,** 467–488. ISSN: 1572-9575. https://doi.org/10.1007/BF02650179 (June 1982).

34. Bauer, B., Bravyi, S., Motta, M. & Chan, G. K.-L. Quantum Algorithms for Quantum Chemistry and Quantum Materials Science. *Chemical Reviews* **120,** 12685–12717. ISSN: 0009-2665. https://doi.org/10.1021/acs.chemrev.9b00829 (Nov. 2020).

35. Horn, R. A. & Johnson, C. R. *Matrix Analysis* (Cambridge University Press, 1985).

36. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition* ISBN: 9781107002173 (Cambridge University Press, 2011).

37. Devitt, S. J., Munro, W. J. & Nemoto, K. Quantum error correction for beginners. *Reports on Progress in Physics* **76,** 076001. https://doi.org/10.1088/0034-4885/76/7/076001 (June 2013).

38. Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation. *Journal of the Physical Society of Japan* **90,** 032001. eprint: https://doi.org/10.7566/JPSJ.90.032001. https://doi.org/10.7566/JPSJ.90.032001 (2021).

39. Cleve, R., Ekert, A., Macchiavello, C. & Mosca, M. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454,** 339–354. ISSN: 1471-2946. http://dx.doi.org/10.1098/rspa.1998.0164 (Jan. 1998).

40. Cao, Y. *et al.* Quantum Chemistry in the Age of Quantum Computing. eprint: arXiv:1812.09976 (2018).

41. Bittel, L. & Kliesch, M. Training Variational Quantum Algorithms Is NP-Hard. *Physical Review Letters* **127.** ISSN: 1079-7114. http://dx.doi.org/10.1103/PhysRevLett.127.120502 (Sept. 2021).

42. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Physical Review A* **99.** ISSN: 2469-9934. http://dx.doi.org/10.1103/PhysRevA.99.032331 (Mar. 2019).

43. Barkoutsos, P. K. *et al.* Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions. *Phys. Rev. A* **98,** 022322. https://link.aps.org/doi/10.1103/PhysRevA.98.022322 (2 Aug. 2018).

44. Grimsley, H. R., Claudino, D., Economou, S. E., Barnes, E. & Mayhall, N. J. Is the Trotterized UCCSD Ansatz Chemically Well-Defined? *Journal of Chemical Theory and Computation* **16,** 1–6. ISSN: 1549-9626. http://dx.doi.org/10.1021/acs.jctc.9b01083 (Dec. 2019).

45. Evangelista, F. A., Chan, G. K.-L. & Scuseria, G. E. Exact parameterization of fermionic wave functions via unitary coupled cluster theory. *The Journal of Chemical Physics* **151,** 244112. eprint: https://doi.org/10.1063/1.5133059. https://doi.org/10.1063/1.5133059 (2019).

46. Nooijen, M. Can the Eigenstates of a Many-Body Hamiltonian Be Represented Exactly Using a General Two-Body Cluster Expansion? *Phys. Rev. Lett.* **84,** 2108–2111. https://link.aps.org/doi/10.1103/PhysRevLett.84.2108 (10 Mar. 2000).

47. Mukherjee, D. & Kutzelnigg, W. Some comments on the coupled cluster with generalized singles and doubles (CCGSD) ansatz. *Chemical Physics Letters* **397,** 174–179. ISSN: 0009-2614. https://www.sciencedirect.com/science/article/pii/S0009261404013181 (2004).

48. Lee, J., Huggins, W. J., Head-Gordon, M. & Whaley, K. B. Generalized Unitary Coupled Cluster Wave functions for Quantum Computation. *Journal of Chemical Theory and Computation* **15,** 311–324. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.8b01004 (Jan. 2019).

49. Romero, J. *et al.* Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz. *Quantum Science and Technology* **4,** 014008. https://doi.org/10.1088/2058-9565/aad3e4 (Oct. 2018).

50. Grimsley, H. R., Economou, S. E., Barnes, E. & Mayhall, N. J. An adaptive variational algorithm for exact molecular simulations on a quantum computer. *Nature Communications* **10,** 3007. ISSN: 2041-1723. https://doi.org/10.1038/s41467-019-10988-2 (July 2019).

51. Kandala, A. *et al.* Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets. *Nature* **549,** 242–246. ISSN: 1476-4687. https://doi.org/10.1038/nature23879 (Sept. 2017).

52. Jordan, P. & Wigner, E. Über das Paulische Äquivalenzverbot. *Zeitschrift für Physik* **47,** 631–651. ISSN: 0044-3328. https://doi.org/10.1007/BF01331938 (July 1928).

53. Seeley, J. T., Richard, M. J. & Love, P. J. The Bravyi-Kitaev transformation for quantum computation of electronic structure. *The Journal of Chemical Physics* **137,** 224109. eprint: https://doi.org/10.1063/1.4768229. https://doi.org/10.1063/1.4768229 (2012).

54. Tranter, A., Love, P. J., Mintert, F. & Coveney, P. V. A Comparison of the Bravyi–Kitaev and Jordan–Wigner Transformations for the Quantum Simulation of Quantum Chemistry. *Journal of Chemical Theory and Computation* **14.** PMID: 30189144, 5617–5630. eprint: https://doi.org/10.1021/acs.jctc.8b00450. https://doi.org/10.1021/acs.jctc.8b00450 (2018).

55. Moll, N., Fuhrer, A., Staar, P. & Tavernelli, I. Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer. *Journal of Physics A: Mathematical and Theoretical* **49,** 295301. https://doi.org/10.1088/1751-8113/49/29/295301 (June 2016).

56. Bravyi, S., Gambetta, J. M., Mezzacapo, A. & Temme, K. *Tapering off qubits to simulate fermionic Hamiltonians* 2017. arXiv: 1701.08213 [quant-ph].

57. O'Malley, P. J. J. *et al.* Scalable Quantum Simulation of Molecular Energies. *Physical Review X* **6.** ISSN: 2160-3308. http://dx.doi.org/10.1103/PhysRevX.6.031007 (July 2016).

58. Setia, K. *et al.* Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries. *Journal of Chemical Theory and Computation* **16,** 6091–6097. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.0c00113 (Oct. 2020).

59. Whitfield, J. D., Biamonte, J. & Aspuru-Guzik, A. Simulation of electronic structure Hamiltonians using quantum computers. *Molecular Physics* **109,** 735–750. eprint: https://doi.org/10.1080/00268976.2011.552441. https://doi.org/10.1080/00268976.2011.552441 (2011).

60. McClean, J. R., Babbush, R., Love, P. J. & Aspuru-Guzik, A. Exploiting Locality in Quantum Computation for Quantum Chemistry. *The Journal of Physical Chemistry Letters* **5,** 4368–4380. https://doi.org/10.1021/jz501649m (Dec. 2014).

61. McClean, J. R., Romero, J., Babbush, R. & Aspuru-Guzik, A. The theory of variational hybrid quantum-classical algorithms. *New Journal of Physics* **18,** 023023. ISSN: 1367-2630. http://dx.doi.org/10.1088/1367-2630/18/2/023023 (Feb. 2016).

62. Motta, M. *et al.* Low rank representations for quantum simulation of electronic structure. *npj Quantum Information* **7,** 83. ISSN: 2056-6387. https://doi.org/10.1038/s41534-021-00416-z (May 2021).

63. Peng, B. & Kowalski, K. Highly Efficient and Scalable Compound Decomposition of Two-Electron Integral Tensor and Its Application in Coupled Cluster Calculations. *Journal of Chemical Theory and Computation* **13,** 4179–4192. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.7b00605 (Sept. 2017).

64. Tilly, J. *et al.* The Variational Quantum Eigensolver: a review of methods and best practices. *arXiv preprint arXiv:2111.05176* (2021).

65. Huggins, W. J. *et al.* Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers. *npj Quantum Information* **7,** 23. ISSN: 2056-6387. https://doi.org/10.1038/s41534-020-00341-7 (Feb. 2021).

66. Wang, G., Koh, D. E., Johnson, P. D. & Cao, Y. Minimizing Estimation Runtime on Noisy Quantum Computers. *PRX Quantum* **2.** ISSN: 2691-3399. http://dx.doi.org/10.1103/PRXQuantum.2.010346 (Mar. 2021).

67. Huggins, W. J., Lee, J., Baek, U., O'Gorman, B. & Whaley, K. B. A non-orthogonal variational quantum eigensolver. *New Journal of Physics* **22,** 073009. ISSN: 1367-2630. http://dx.doi.org/10.1088/1367-2630/ab867b (July 2020).

68. Stair, N. H., Huang, R. & Evangelista, F. A. A Multireference Quantum Krylov Algorithm for Strongly Correlated Electrons. *Journal of Chemical Theory and Computation* **16.** PMID: 32091895, 2236–2245. eprint: https://doi.org/10.1021/acs.jctc.9b01125. https://doi.org/10.1021/acs.jctc.9b01125 (2020).

69. Parrish, R. M. & McMahon, P. L. *Quantum Filter Diagonalization: Quantum Eigendecomposition without Full Quantum Phase Estimation* 2019. arXiv: 1909.08925 [quant-ph].

70. Higgott, O., Wang, D. & Brierley, S. Variational Quantum Computation of Excited States. *Quantum* **3,** 156. ISSN: 2521-327X. http://dx.doi.org/10.22331/q-2019-07-01-156 (July 2019).

71. Childs, A. M., Kothari, R. & Somma, R. D. Quantum Algorithm for Systems of Linear Equations with Exponentially Improved Dependence on Precision. *SIAM Journal on Computing* **46,** 1920–1950. ISSN: 1095-7111. http://dx.doi.org/10.1137/16M1087072 (Jan. 2017).

72. Ralli, A., Love, P. J., Tranter, A. & Coveney, P. V. Implementation of measurement reduction for the variational quantum eigensolver. *Physical Review Research* **3.** ISSN: 2643-1564. http://dx.doi.org/10.1103/PhysRevResearch.3.033195 (Aug. 2021).

73. Dos Santos, R. R. *Introduction to Quantum Monte Carlo simulations for fermionic systems* 2003. arXiv: cond-mat/0303551 [cond-mat.str-el].

74. Helgaker, T. U. & Almlöf, J. A second-quantization approach to the analytical evaluation of response properties for perturbation-dependent basis sets. en. *International Journal of Quantum Chemistry* **26,** 275–291. ISSN: 1097-461X. http://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560260211 (2021) (1984).

75. Olsen, J., Bak, K. L., Ruud, K., Helgaker, T. & Jørgensen, P. Orbital connections for perturbation-dependent basis sets. *Theoretica chimica acta* **90,** 421–439 (1995).

76. Carlson, B. C. & Keller, J. M. Orthogonalization Procedures and the Localization of Wannier Functions. en. *Physical Review* **105,** 102–103. ISSN: 0031-899X. https://link.aps.org/doi/10.1103/PhysRev.105.102 (2021) (Jan. 1957).

77. Higham, N. J. Cholesky factorization. en. *Wiley Interdisciplinary Reviews: Computational Statistics* **1,** 251–254. ISSN: 1939-5108, 1939-0068. https://onlinelibrary.wiley.com/doi/10.1002/wics.18 (2021) (Sept. 2009).

78. Bastos, L. S. & O'Hagan, A. Pivoting Cholesky Decomposition applied to Emulation and Validation of computer models. en, 28.

79. Aquilante, F., Bondo Pedersen, T., Sánchez de Merás, A. & Koch, H. Fast noniterative orbital localization for large molecules. en. *The Journal of Chemical Physics* **125,** 174101. ISSN: 0021-9606, 1089-7690. http://aip.scitation.org/doi/10.1063/1.2360264 (2021) (Nov. 2006).

80. Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. en. *Psychometrika* **31,** 1–10. ISSN: 0033-3123, 1860-0980. http://link.springer.com/10.1007/BF02289451 (2021) (Mar. 1966).

81. Banerjee, S. & Roy, A. *Linear Algebra and Matrix Analysis for Statistics* https://doi.org/10.1201/b17040 (Chapman and Hall/CRC, June 2014).

82. Sax, A. F. Localization of molecular orbitals on fragments. *Journal of Computational Chemistry* **33,** 1495–1510. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.22980. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.22980 (2012).

83. Purvis III, G. D., Shepard, R., Brown, F. B. & Bartlett, R. J. C2V Insertion pathway for BeH2: A test problem for the coupled-cluster single and double excitation model. *International Journal of Quantum Chemistry* **23,** 835–845. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/qua.560230307. https://onlinelibrary.wiley.com/doi/abs/10.1002/qua.560230307 (1983).

84. Davison, A. C. & Hinkley, D. V. *Bootstrap methods and their application* **1** (Cambridge university press, 1997).

85. Kivlichan, I. D. *et al.* Quantum Simulation of Electronic Structure with Linear Depth and Connectivity. *Physical Review Letters* **120.** ISSN: 1079-7114. http://dx.doi.org/10.1103/PhysRevLett.120.110501 (Mar. 2018).

86. Shende, V. V. & Markov, I. L. *On the CNOT-cost of TOFFOLI gates* 2008. arXiv: 0803.2316 [quant-ph].

87. Guerreschi, G. G. Repeat-until-success circuits with fixed-point oblivious amplitude amplification. *Physical Review A* **99.** ISSN: 2469-9934. http://dx.doi.org/10.1103/PhysRevA.99.022306 (Feb. 2019).

88. Penrose, R. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society* **51,** 406–413 (1955).

89. Fix, G. & Heiberger, R. An Algorithm for the Ill-Conditioned Generalized Eigenvalue Problem. *SIAM Journal on Numerical Analysis* **9,** 78–88. eprint: https://doi.org/10.1137/0709009. https://doi.org/10.1137/0709009 (1972).

90. Epperly, E. N., Lin, L. & Nakatsukasa, Y. *A Theory of Quantum Subspace Diagonalization* 2021. https://arxiv.org/abs/2110.07492.

91. Golub, G. H. & Van Loan, C. F. Matrix computations (2013).

92. Demmel, J. & Kågström, B. The Generalized Schur Decomposition of an Arbitrary Pencil A–λB—Robust Software with Error Bounds and Applications. Part II: Software and Applications. *ACM Trans. Math. Softw.* **19,** 175–201. ISSN: 0098-3500. https://doi.org/10.1145/152613.152616 (June 1993).

93. Ghojogh, B., Karray, F. & Crowley, M. *Eigenvalue and Generalized Eigenvalue Problems: Tutorial* 2019. https://arxiv.org/abs/1903.11240.

94. Prosser, F. & Hagstrom, S. On the rapid computation of matrix elements. eng. *International journal of quantum chemistry* **2,** 89–99. ISSN: 0020-7608 (1968).

95. Verbeek, J. & Van Lenthe, J. H. On the evaluation of non-orthogonal matrix elements. en. *Journal of Molecular Structure: THEOCHEM* **229,** 115–137. ISSN: 01661280. https://linkinghub.elsevier.com/retrieve/pii/0166128091901416 (2021) (May 1991).

96. Thom, A. J. W. & Head-Gordon, M. Hartree–Fock solutions as a quasidiabatic basis for nonorthogonal configuration interaction. *The Journal of Chemical Physics* **131,** 124113. eprint: https://doi.org/10.1063/1.3236841. https://doi.org/10.1063/1.3236841 (2009).

97. Pulay, P. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters* **73,** 393–398. ISSN: 0009-2614. https://www.sciencedirect.com/science/article/pii/0009261480803964 (1980).

98. Scuseria, G. E., Lee, T. J. & Schaefer, H. F. Accelerating the convergence of the coupled-cluster approach: The use of the DIIS method. *Chemical Physics Letters* **130,** 236–239. ISSN: 0009-2614. https://www.sciencedirect.com/science/article/pii/0009261486804614 (1986).

99. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585,** 357–362. https://doi.org/10.1038/s41586-020-2649-2 (Sept. 2020).

100. Virtanen, P. *et al.* SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **17,** 261–272 (2020).

101. Sun, Q. *et al.* PySCF: the Python-based simulations of chemistry framework. *WIREs Computational Molecular Science* **8,** e1340. eprint: https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wcms.1340. https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1340 (2018).

102. Sun, Q. *et al.* Recent developments in the PySCF program package. *The Journal of Chemical Physics* **153,** 024109. eprint: https://doi.org/10.1063/5.0006074. https://doi.org/10.1063/5.0006074 (2020).

103. A. Smith, D. G. & Gray, J. opt_einsum - A Python package for optimizing contraction order for einsum-like expressions. *Journal of Open Source Software* **3,** 753. https://doi.org/10.21105/joss.00753 (2018).

104. ANIS, M. S. *et al. Qiskit: An Open-source Framework for Quantum Computing* 2021.

105. McClean, J. R. *et al.* OpenFermion: the electronic structure package for quantum computers. *Quantum Science and Technology* **5,** 034014. https://doi.org/10.1088/2058-9565/ab8ebc (June 2020).

106. Nocedal, J. & Wright, S. J. *Numerical optimization* (Springer, 1999).

107. Bodenstein, T. & Kvaal, S. A state-specific multireference coupled-cluster method based on the bivariational principle. *The Journal of Chemical Physics* **153,** 024106. eprint: https://doi.org/10.1063/5.0009429. https://doi.org/10.1063/5.0009429 (2020).

108. Bulik, I. W., Henderson, T. M. & Scuseria, G. E. Can Single-Reference Coupled Cluster Theory Describe Static Correlation? *Journal of Chemical Theory and Computation* **11.** PMID: 26575754, 3171–3179. eprint: https://doi.org/10.1021/acs.jctc.5b00422. https://doi.org/10.1021/acs.jctc.5b00422 (2015).

109. Rossmannek, M., Barkoutsos, P. K., Ollitrault, P. J. & Tavernelli, I. Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems. *The Journal of Chemical Physics* **154,** 114105. ISSN: 1089-7690. http://dx.doi.org/10.1063/5.0029536 (Mar. 2021).

110. Koch, H., Sánchez de Merás, A. & Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *The Journal of Chemical Physics* **118,** 9481–9484. eprint: https://doi.org/10.1063/1.1578621. https://doi.org/10.1063/1.1578621 (2003).

111. Neese, F., Wennmohs, F. & Hansen, A. Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method. *The Journal of Chemical Physics* **130,** 114108. eprint: https://doi.org/10.1063/1.3086717. https://doi.org/10.1063/1.3086717 (2009).

112. Rolik, Z. & Kállay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *The Journal of Chemical Physics* **135,** 104111. eprint: https://doi.org/10.1063/1.3632085. https://doi.org/10.1063/1.3632085 (2011).

113. Nagy, P. R., Samu, G. & Kállay, M. Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Improved Algorithm and Benchmark Applications. *Journal of Chemical Theory and Computation* **14,** 4193–4215. ISSN: 1549-9618. https://doi.org/10.1021/acs.jctc.8b00442 (Aug. 2018).

114. Prochnow, E., Evangelista, F. A., Schaefer, H. F., Allen, W. D. & Gauss, J. Analytic gradients for the state-specific multireference coupled cluster singles and doubles model. *The Journal of Chemical Physics* **131,** 064109. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.3204017. https://aip.scitation.org/doi/abs/10.1063/1.3204017 (2009).

115. Fan, K. & Hoffman, A. J. Some metric inequalities in the space of matrices. *Proceedings of the American Mathematical Society* **6,** 111–116 (1955).

116. De Ruiter, A. H. J. & Forbes, J. R. On the Solution of Wahba's Problem on $SO(n)$. *The Journal of the Astronautical Sciences* **60,** 1–31. ISSN: 2195-0571. https://doi.org/10.1007/s40295-014-0019-8 (Mar. 2013).