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
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

Acknowledgements

I would like to thank my main supervisor Simen Kvaal for his support, for our fruitful (and sometimes very long) conversations, and for letting me go in the direction I wanted to with this project and explore the topic independently. Special thanks go to him for introducing me to eigenvector continuation and the mathematical aspects of perturbation theory - I did not know chemistry could be so mathematical. Furthermore, I would like to thank my co-supervisor Joakim Bergli for his support and our regular meetings, which really helped me restructure my thoughts and develop new ideas. I would also like to thank Håkon Kristiansen for helping me with `PySCF` and for sharing his coupled cluster code with me. Finally, I would like to give a shout-out to my fiancée Evan, who does not know quantum chemistry, for proofreading this thesis, and for listening to what probably felt like endless talk about the contents of this thesis.

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 the coupled cluster singles and doubles (CCSD) method give qualitatively correct energies close to the equilibrium geometry, but can fail qualitatively upon dissociation. We used subspace projected CCSD eigenvector continuation to continue the ground state and obtained a qualitatively correct dissociation curve for the nitrogen molecule and an improved potential energy surface for the asymmetric stretch of 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 for large parts of the potential energy surface, while cutting off more than 50% of computation 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, leading to extremely large sample needs. It is however an efficient interpolation method that can be used to solve some types of multi-reference problems on quantum computers.

Contents

List of Figures

List of Tables

Introduction

The purpose of this thesis is to study how eigenvector continuation can be used in quantum chemistry. Eigenvector continuation [1], [2] is a novel method where the ground state eigenvector of a Hermitian matrix $\mathbf{H}(\vec{\alpha}_\odot)$, where $\mathbf{H}(\vec{\alpha})$ is analytically dependent on some parameter $\vec{\alpha}$, is approximated as a linear combination of ground state eigenvectors obtained for different parameters $\vec{\alpha}_1, \ldots, \vec{\alpha}_L$. With eigenvector continuation, it has been shown that it is possible to extrapolate the ground state from "easy" regions to "difficult" regions at reduced computational cost [2]. Furthermore, ground state eigenvectors in many-dimensional vector spaces can in many cases be well-represented with few sample states [3]. A variant of eigenvector continuation, applicable to coupled cluster singles and doubles (CCSD) wave functions [4], has decreased the presumed run time for a given calculation from 20 years down to one hour with little accuracy loss, and eigenvector continuation has been shown to have better convergence properties than perturbation theory [3]. All of these aspects make eigenvector continuation a promising method to approximate 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 (PES), 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 method is a high-accuracy 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 quantitatively or 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, it is studied if and how eigenvector continuation can be used to extrapolate the solutions from "easy" geometries, where CCSD gives accurate results, to "hard" geometries, and if and how eigenvector continuation can be used to reduce computational cost. An extra difficulty in quantum chemistry is that each nuclear geometry gets assigned a different basis, so the vector space changes as nuclei are moved. This is something which is focused on in particular 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 use the VQE algorithm [5], where the wave function is parametrized by an ansatz, and expectation values, including the energy, are sampled. As sampling introduces uncertainty, quantum computing can be used 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 a linear scaling on quantum computers, such as the trotterized unitary coupled cluster method [6] and methods derived from it. In this thesis, it is thus studied how eigenvector continuation can be performed with data sampled

from quantum computers and which advantages and difficulties it poses.

This thesis is organized in the following way: In chapter 1, we give a brief and non-exhaustive introduction to 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 used to do quantum chemistry, focusing on relevant formulas and circuits to implement eigenvector continuation. In chapter 4, we discuss necessary theory, methods and considerations to implement eigenvector continuation for quantum chemistry, with the main focus on obtaining PESs. In this chapter, we 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 chapter 5, discussing the numerical implementation of our methods. In the next chapters 6, 7 and 8, we show and discuss our results for multi-reference, single-reference and quantum computing methods, respectively, and finish with a conclusion and an outlook. Finally, the appendix contains figures and proofs that do 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 and to make clear what a shorthand notation means, we use the following notation:

- N stands for the number of electrons in a system.
- M stands for the number of molecular orbitals.
- L stands for the number of sample vectors/states included in the 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 Born-Oppenheimer approximation.
- The indices $a, b, c, \ldots$ represent unoccupied molecular spin orbitals, $i, j, k, \ldots$ represent occupied molecular spin orbitals, and $p, q, r, \ldots$ represent either. Greek indices $\mu, \nu, \sigma, \ldots$ are used for atomic (spin) orbitals.
- $|\Psi\rangle$ (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 Slater determinant representing a Hartree-Fock state. One-particle functions will be written with small Greek letters.
- Operators are written with a hat $\hat{O}$, with the exception of creation/annihilation operators and Pauli spin operators. Matrices are written 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.
- For elements of a matrix $\mathbf{M}$, we write $(\mathbf{M})_{ij} = M_{ij}$. For elements of inverse matrices, we write $\left(\mathbf{M}^{-1}\right)_{ij} = M^{-1}_{ij}$. An equivalent notation is used for fractional matrix powers.
- When writing sums, we will sometimes use a shorthand notation. $\sum_{pq}^{M}$, for example, is to be interpreted as $\sum_{p=1}^{M}\sum_{q=1}^{M}$.

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 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 (e.g. moving them), which is important for the application of eigenvector continuation. This chapter is by no means meant to be a complete description of each method included and does not aim to give an overview of the field as a whole, 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 chapter 4 of this thesis. Almost everything in this chapter is taken from the literature, except for the analyticity 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. Often, it is of interest 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, which in atomic units reads

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

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 charge Z_A, and h_{nuc} the repulsion between the (stationary) nuclei. This Hamiltonian arises when considering the full Hamiltonian under the Born-Oppenheimer approximation [7, ch. 2]. From statistical mechanics, we know that the ground state is the most prevalent state, and ground state potential energy surfaces are used in many applications, such as ab-initio molecular dynamics [8], hence it is of interest to obtain the ground state. However, for two electrons or more, eq. (1.1) has not been solved analytically, making it necessary to make approximations. The first step is the insertion of a basis, that is, expressing the wave function as

$$\left|\Psi\right\rangle = \sum_{n=1}^{\infty} c_n \left|\Phi_n\right\rangle \tag{1.3}$$

where $\{|\Phi_n\rangle\}_{n=1}^{\infty}$ is a complete orthonormal basis. This parameterization is exact, but it is computationally infeasible to sum to infinity, and we consider finite bases in section 1.2.1.

Electrons are indistinguishable spin $\frac{1}{2}$ particles, and a wave function describing N electrons needs to be antisymmetric with respect to particle interchange. In the following, we use the standard notation $\mathbf{x} = \{\mathbf{r}, \omega\}$, where $\omega \in \{\alpha, \beta\}$, e.g. spin up $|\alpha\rangle = |\uparrow\rangle$ and spin down $|\beta\rangle = |\downarrow\rangle$. We write $|\psi(\mathbf{x})\rangle \equiv |\psi(\mathbf{r})\rangle \otimes |\omega\rangle$, and make use of the notation

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

Unless explicitly necessary, we simply write $|\psi\rangle$ instead of $|\psi(\mathbf{x})\rangle$ or $\psi(\mathbf{x})$. Furthermore, we 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* [9]. For one-electron functions $|\phi_1(\mathbf{x})\rangle, \ldots, |\phi_N(\mathbf{x})\rangle \in H$, a Slater determinant is defined 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 for the remainder of this thesis.

The same state can also be represented as an *occupation number (ON) vector* or *bit-string*

$$|\mathbf{k}\rangle = |k_1, k_2, \ldots, k_M, \ldots\rangle, k_p = \begin{cases} 1 & \phi_p \text{ occupied} \\ 0 & \phi_p \text{ not occupied} \end{cases}, \qquad \sum_{p=1}^{\infty} k_p = N \tag{1.6}$$

where the term *occupied* means that $\phi_p(\mathbf{x})$ enters the Slater determinant. This notation only makes sense if the one-electron functions are mutually orthogonal. It can be shown [9] 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 $|B| = M$, one still needs $\binom{M}{N}$ determinants 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 more approximations than the introduction of a finite basis are necessary to calculate wave functions in practice.

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 any of Refs. [9]–[11]. 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 a linear combination of states containing 0 to ∞ particles.

The state $|\rangle$ is a basis for the space $H^{\wedge 0}$ and referred to as *vacuum state*. The set of all Slater determinants that can be constructed from an orthonormal one-electron basis, together with the vacuum state, form a basis for the Fock space $\mathcal{F}$.

Assuming an orthonormal one-electron basis $B = \{|\phi_p\rangle\}_{p=1}^{\infty}$, we define the *creation operator*[1] $a_p^\dagger$, which is defined by the way it acts on a Slater determinant

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

It "creates" a particle in the sense that a row and a column are added to the Slater determinant while keeping it normalized. 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

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

where $[A, B]_+ \equiv AB + BA$. Similarly, there is 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

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

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

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

where

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

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

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

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

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, that is, the tensor product of the one-electron basis functions $|\phi_p\rangle \otimes |\phi_q\rangle$.

1.2.1 Incomplete and non-orthogonal basis sets

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

We define the Hermitian, positive semi-definite overlap matrix $\mathbf{S}$ [12] with elements

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

One way to form an orthonormal basis $B = \{|\phi_p\rangle\}_{p=1}^{M}$ is by symmetric orthogonalization [13]

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

This choice of orthogonalization is not unique. The introduction of a basis gives rise to a projection $\hat{P}_B$, projecting the (infinite-dimensional) one-electron basis on the finite-dimensional basis B

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

A consequence of this is that Slater determinants expressed using only one-electron functions from the basis B cannot have more than M electrons, as that would make the columns of the underlying matrix linearly dependent and the Slater determinant equal to zero. Furthermore, ON vectors are now bit-strings 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 of finite size, the effect of the Hamiltonian can be expressed by an effective Hamiltonian, e.g. the Hamiltonian projected on the new basis

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

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

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

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, a so-called *basis set* $BS = \{\eta_\mu(\mathbf{r})\}_{\mu=1}^{M/2}$, which contains spatial *atomic orbitals (AOs)*. 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*

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

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, called Cartesian GTOs, though other choices exist. The AOs used in a poly-atomic 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 [10, ch. 7, 8].

There are infinitely many ways to construct M orthonormal functions from M atomic spin orbitals. Any such set of orthonormal functions constructed from atomic spin orbitals is from now on referred to as set of *molecular orbitals (MOs)*. We use the term molecular orbitals for any orthonormal set of one-electron functions, even when they are not constructed from atomic spin orbitals. Observe that molecular orbitals do not need to have a definite spin.

Analyticity under geometric perturbations of standard basis sets

Using Gaussian basis sets, we assume that the overlap matrix $\mathbf{S}$ is invertible unless we place two nuclei at the same position (in which case the Hamiltonian is not defined).[3] The overlap matrix as a function of the nuclear positions $\mathbf{S}(\vec{R})$ is analytic under geometric perturbations. To see this, we first recognize that each integral S_{pq} consists of a finite sum of integrals over primitive Cartesian GTOs. Each integral over primitive Cartesian GTOs can be expressed as a finite sum of terms analytic in the geometric perturbation using the Obara-Saika scheme and the Gaussian product rule [10, ch. 9]. Thus, the overlap matrix is analytic. In addition, due to the non-singularity of the overlap matrix $\mathbf{S}(\vec{R})$ (except for the singularities), it will be invertible as well as positive-definite, 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 molecular orbitals. 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 $\hat{H}_{\text{fin}}(\vec{R})$ as a product of analytic functions. From now on, the subscript $_{\text{fin}}$ is dropped, keeping in mind that the Schrödinger equation is solved in a finite basis.

1.2.2 Wick's theorem

Wick's theorem [9] is a very powerful method to calculate the expectation values of strings of creation and annihilation operators $\hat{A} = abc\ldots xyz$. It states that

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

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

where $\{\hat{A}\}_v$ is the *normal ordering* of an operator $\hat{A}$, $\overset{\sqcap}{ab}$ is a *contraction*, and the "singles" sum goes over all possible combinations of single contractions and similarly for the other terms. The normal ordering is a reordering of the individual operators such that all annihilation operators stand to the right of all creation operators. The contraction (which turns out to be a number) is defined as

$$\overset{\sqcap}{ab} = ab - \{ab\}_v, \tag{1.26}$$

where it should be noted that in order to calculate contractions of the type $\overbrace{abc\cdots xyz}$, it is necessary to apply anticommutation rules to move the operators being contracted next to one another, which may induce sign changes. 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 a 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 [11].

1.2.3 The Fermi vacuum

Using Wick's theorem, the matrix element two Slater determinants expressed in the same MO basis can be effectively evaluated, as one can always write it as a vacuum expectation 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 expressions are very long. This is also a common calculation, as matrix elements between Slater determinants play an important role in quantum chemistry. Wick's theorem keeps its validity upon redefinition of what serves as vacuum by redefining what serves as annihilation and creation operators. With respect to a Slater determinant $|\Phi^{\mathrm{SD}}\rangle$, where $i, j, \ldots$ are occupied and $a, b, \ldots$ are unoccupied orbitals, we have that

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

while $a_a^\dagger, a_i$ acting on $|\Phi^{\mathrm{SD}}\rangle$ produce a $N \pm 1$ particle Slater determinant. $|\Phi^{\mathrm{SD}}\rangle$ serves as *Fermi vacuum*. The molecular orbitals corresponding to unoccupied 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

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

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

1.2.4 Normal ordering of the Hamiltonian

As described above, in second quantization, the Hamiltonian reads

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

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

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

where $|\Phi^{\text{SD}}\rangle$ is a reference Slater determinant and F_{pq} is the Fock matrix described in the next section. 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 The Hartree-Fock method

The *Hartree-Fock (HF)* method [7, ch. 3] is a variational method that aims at finding the Slater determinant $|\Phi^{\text{SD}}\rangle$ that minimizes the expectation energy

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

As a Slater determinant consists of N orthonormal one-electron functions $|\phi_i(\mathbf{x})\rangle$, $i = 1, \ldots, N$ that determine it, solving eq. (1.33) can be carried out by finding those N one-electron functions.

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

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

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

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

and the Lagrangian multipliers ϵ_{ij} are included to ascertain orthonormality of the one-electron functions. Because the expectation energy of Slater determinants is invariant under unitary operations among these one-electron functions (which we prove on the next page), it is possible to make further assumptions about them. With these assumptions, the stationary points are characterized by the following eigenvalue problem [14]:

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

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

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

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

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

where we used the notation $[\phi]$ to indicate that the Fock operator is a functional of N one-electron functions $|\phi_i\rangle$, $i = 1,\ldots,N$. Those N functions will be eigenfunctions of the Fock operator, but it should be observed that the Fock operator has additional eigenfunctions. The eigenvalues ϵ_i are referred to as *orbital energies*. Solving eq. (1.36) analytically is not possible for molecules [7, ch. 3], hence we introduce an orthogonal basis:

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

where the coefficients $C_{\mu i}$ need to be chosen in such a way the functions $\{|\phi_i(\mathbf{x})\rangle\}_{i=1}^{M}$ are orthonormal, and $\{|\chi_\mu(\mathbf{x})\rangle\}_{i=1}^{M}$ is a set of atomic spin orbitals. The $M \times M$ matrix $\mathbf{C}$ with elements $C_{\mu i}$ is referred to as *coefficient matrix*. Even though $\mathbf{C}$ is an $M \times M$ matrix, it can naturally be divided into blocks

$$\mathbf{C} = [\mathbf{C}^o \quad \mathbf{C}^v] \tag{1.41}$$

where $\mathbf{C}^o$ contains the first N columns describing those MOs that will enter in the Slater determinant, and $\mathbf{C}^v$ contains the last $M-N$ columns describing MOs that will not enter in the Slater determinant. In this basis, the energy of a Slater determinant is given as

$$\begin{aligned}\left\langle \Phi^{\mathrm{SD}}\middle|\hat{H}\middle|\Phi^{\mathrm{SD}}\right\rangle &= \sum_{i}^{N}\langle\phi_i|\hat{h}|\phi_i\rangle + \frac{1}{2}\sum_{i,j}^{N}\langle\phi_i\phi_j||\phi_i\phi_j\rangle \\ &= \sum_{\mu,\nu}^{M}\sum_{i}^{N} C_{\mu i}^* C_{\nu i}\,\langle\chi_\mu|\,\hat{h}\,|\chi_\nu\rangle + \frac{1}{2}\sum_{\mu,\nu,\sigma,\tau}^{M}\sum_{i,j}^{N} C_{\mu i}^* C_{\nu j}^* C_{\sigma i} C_{\tau j}\langle\chi_\mu\chi_\nu||\chi_\sigma\chi_\tau\rangle.\end{aligned} \tag{1.42}$$

Observe that the energy does not depend on $\mathbf{C}^v$, only $\mathbf{C}^o$. With the problem reduced to finding the coefficients $C_{\mu i}$, finding the stationary points of the Lagrangian gives rise to the *self-consistent field (SCF)* method

$$\mathbf{F}[\mathbf{C}^o]\mathbf{C} = \mathbf{S}\mathbf{C}\boldsymbol{\varepsilon} \tag{1.43}$$

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

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

where we defined the *density matrix* $\mathbf{D}^o = \mathbf{C}^o\,(\mathbf{C}^o)^\dagger$. The eigenvalue problem in eq. (1.43) is a generalized Hermitian eigenvalue problem which can be solved efficiently, cf. sec. 5.1.2, p. 62. However, an extra complication is that the correct $\mathbf{C}^o$ needs to be known beforehand in order to construct the Fock matrix $\mathbf{F}[\mathbf{C}^o]$. Using a wrong $\tilde{\mathbf{C}}^o$ will give raise to the following eigenvalue problem

$$\mathbf{F}[\tilde{\mathbf{C}}^o]\tilde{\mathbf{C}}' = \mathbf{S}\tilde{\mathbf{C}}'\boldsymbol{\varepsilon} \tag{1.45}$$

where $\tilde{\mathbf{C}}' \neq \tilde{\mathbf{C}} \neq \mathbf{C}$, where only the correct $\mathbf{C}$ makes the Lagrangian stationary. The hope is now that $\tilde{\mathbf{C}}'$ is closer to $\mathbf{C}$ than $\tilde{\mathbf{C}}$ is, given that $\tilde{\mathbf{C}}$ itself is a reasonable guess. This defines

an iterative procedure, where a new matrix diagonalization is carried out, this time using the Fock operator $\mathbf{F}[(\tilde{\mathbf{C}}')^o]$ and continuing until the input coefficient matrix is approximately identical to the output coefficient matrix, e.g. until convergence.

After convergence, to obtain the Hartree-Fock Slater determinant $|\Phi^{\mathrm{HF}}\rangle$, those N one-electron functions are chosen which correspond to the columns of $\mathbf{C}^o$. Those are the N functions with the smallest N eigenvalues of the Fock matrix $\epsilon_1 \leq \cdots \leq \epsilon_N \leq \cdots \leq \epsilon_M$. The MOs corresponding to the columns of $\mathbf{C}$ after having solved eq. (1.43) are referred to as *canonical molecular orbitals*. The unoccupied MOs corresponding to the columns of $\mathbf{C}^v$ are referred to as *virtual* orbitals when $\mathbf{C}$ is the coefficient matrix of a HF state. We use the term virtual orbitals even when the Slater determinant serving as a reference (for normal-ordering, for example) is not a HF state.

1.3.1 Important attributes of the HF state and MOs

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

1. Slater determinants built from orthonormal MOs are invariant except for a phase under unitary operations within those MOs. To see this, let $|{}^x\phi_i\rangle = \sum_j^N |{}^w\phi_j\rangle\, U_{ji}$ where $\mathbf{U}$ is unitary. Let ${}^w\mathbf{A}$ be a matrix with elements ${}^wA_{ij} = {}^w\phi_i(\mathbf{x}_j)$, and ${}^x\mathbf{A}$ correspondingly. Observing that $1/\sqrt{N!}|{}^w\mathbf{A}| = |{}^w\Phi^{\mathrm{SD}}\rangle$, we see that

$$|{}^x\Phi^{\mathrm{SD}}\rangle = \frac{1}{\sqrt{N!}}|{}^x\mathbf{A}| = \frac{1}{\sqrt{N!}}|{}^w\mathbf{A}\mathbf{U}| = \frac{1}{\sqrt{N!}}|{}^w\mathbf{A}||\mathbf{U}| = e^{i\pi\theta}\,|{}^w\Phi^{\mathrm{SD}}\rangle \tag{1.46}$$

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

2. Upon convergence, the SCF method results in a state that is a stationary point of the Lagrangian (eq. (1.34)) in a basis. There is no guarantee that a global minimum (or a minimum, at all) is found. We will refer to all stationary points of the Lagrangian as HF states, even when the energy is not minimized.

3. For canonical MOs, a direct consequence of eq. (1.43) is that $\mathbf{F}$ is diagonal in a basis of canonical MOs. We 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, and it can be shown that $\langle\Phi^{\mathrm{SD}}|\, a_a^\dagger a_i\, |\Phi^{\mathrm{SD}}\rangle = F_{ai} = 0$. This is *Brillouin's theorem* [12].

4. For Slater determinants, motion of electrons with opposite spins is uncorrelated, thus they are non-interacting [7, ch. 2]. Hartree-Fock theory is 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 do not restrict the wave function to be a single Slater determinant. Methods that build on an (approximate) Hartree-Fock wave function, are referred to as *post-Hartree-Fock* methods.

5. Solving the SCF equations without any restrictions on the MOs gives rise to the *general* Hartree-Fock method (GHF). The resulting Slater determinant will not necessarily be an eigenfunction of the spin projection $\hat{S}_z$, and each MO might have an α and a β spin contribution. Insisting that each MO has definite α or β spin, gives rise 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 with spin α and β respectively, that have the same spatial function. 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, for RHF and UHF, we restrict the MOs to be *real* functions. This is justified because the Hamilton operator $\hat{H}$ is Hermitian and real, and we can 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 $\hat{S}^2$ and $\hat{S}_z$, 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" [15], stating that physically correct symmetries in the Hartree-Fock approximation may lead to higher energies of the Slater determinant.

6. 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 [16]. 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 discuss such a case in the results chapters.

1.3.2 Analyticity of the Hartree-Fock wave function

In standard quantum chemical calculations, it is customary to use real atomic orbitals, as the exact ground state is real and computational cost is reduced. It is important to know whether real Hartree-Fock wave functions are analytic under geometric perturbations. The answer is generally no. Even for hydrogen in minimal basis, there are three UHF states at large internuclear separation, while there is only one at small internuclear separation [17]. This shows that states can appear or disappear, and similar behaviour can be observed for RHF states of other small molecules [18]. Furthermore, the potential energy curves of two HF states might cross, which implies that the HF ground state changes discontinuously. The Lagrangian (eq. (1.34)) is not complex analytic, even though there exist methods to make it analytic and keep the number of solutions constant, which however leads to complex energies [17], [18]. We are not aware of any explicit real analyticity or smoothness results regarding the HF states in the regions where they exist, but we assume for the remainder of this thesis that the RHF states, unless they disappear, are analytic, which seems reasonable to us, as the Hartree-Fock energy and the molecular energies ϵ_i are usually smooth.

1.4 Change of basis operations

Let there be 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 $|^x\phi_p\rangle = \sum_q^M |^w\phi_q\rangle U_{qp}$ 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 $\boldsymbol{\kappa}$ as $\mathbf{U} = e^{-\boldsymbol{\kappa}}$. It can now be shown [10, ch. 3] that the

creation and annihilation operators between those sets are related as

$$\begin{aligned} b_p^\dagger &= \exp(-\hat{\kappa}) a_p^\dagger \exp(\hat{\kappa}) \\ b_p &= \exp(-\hat{\kappa}) a_p \exp(\hat{\kappa}) \\ |{}^x\Phi^{\mathrm{SD}}\rangle &= \exp(-\hat{\kappa}) \, |{}^w\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{1.47}$$

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

$$\hat{\kappa} = \sum_{pq} \kappa_{pq} a_p^\dagger a_q \tag{1.48}$$

$$\kappa_{pq} = [-\ln(\mathbf{U})]_{pq}. \tag{1.49}$$

From the last line of eq. (1.47) 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

In this thesis, we calculate matrix elements between Slater determinants expressed in terms of different sets of MOs that are not mutually orthogonal, and the Slater-Condon rules [7, ch. 2] for the calculation of matrix elements between Slater determinants built from the same set of MOs no longer apply. The *generalized Slater-Condon (GSC) rules* are used for calculating matrix elements between two *non-orthogonal Slater determinants*, e.g. Slater determinants build from non-orthogonal sets of MOs [12]. Let us consider two sets of orthonormal molecular orbitals $\{|{}^w\phi_i\rangle\}_{i=1}^N$, $\{|{}^x\phi_i\rangle\}_{i=1}^N$ that are not mutually orthonormal. Defining the overlap matrix ${}^{wx}\mathbf{S}$ with elements ${}^{wx}S_{ij} = \langle {}^w\phi_i | {}^x\phi_j \rangle$, we have in general that ${}^{wx}S_{ij} \neq \delta_{ij}$, and ${}^{wx}\mathbf{S}$ will generally not be symmetric. Given two Slater determinants built from these MOs,

$$\begin{aligned} |{}^w\Phi^{\mathrm{SD}}\rangle &= |{}^w\phi_1 \dots {}^w\phi_N\rangle \\ |{}^x\Phi^{\mathrm{SD}}\rangle &= |{}^x\phi_1 \dots {}^x\phi_N\rangle \end{aligned} \tag{1.50}$$

the overlap between those two Slater determinants is now given by

$$\langle {}^w\Phi^{\mathrm{SD}} | {}^x\Phi^{\mathrm{SD}} \rangle = |{}^{wx}\mathbf{S}|. \tag{1.51}$$

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

$$\langle {}^w\Phi^{\mathrm{SD}} | \hat{H}_1 | {}^x\Phi^{\mathrm{SD}} \rangle = \sum_{i,j=1}^N {}^{wx}h_{ij} \, {}^{wx}S_{ij}^{(1)} \tag{1.52}$$

where ${}^{wx}h_{ij} = \langle {}^w\phi_i | \hat{h} | {}^x\phi_j \rangle$, and ${}^{wx}\mathbf{S}^{(1)}$ is the *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

$$\langle {}^w\Phi^{\mathrm{SD}} | \hat{H}_2 | {}^x\Phi^{\mathrm{SD}} \rangle = \sum_{i<j}^N \sum_{k<l}^N {}^{wx}\langle ij \| kl \rangle \, {}^{wx}S_{ij,kl}^{(2)} \tag{1.53}$$

where ${}^{wx}\langle ij \| kl \rangle = \langle {}^w\phi_i \, {}^w\phi_j \| {}^x\phi_k \, {}^x\phi_l \rangle$, and ${}^{wx}\mathbf{S}^{(2)}$ is the second-order cofactor matrix, e.g. ${}^{wx}S_{ij,kl}^{(2)}$ 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 can be derived from the GSC rules, as ${}^{wx}\mathbf{S}$ will be diagonal with elements $S_{ii} \in \{0, 1\}$. A proof of the GSC rules is found in the appendix, sec. B.3, p. 95. The GSC rules allow for the use of correlated calculations with non-orthogonal Slater determinants, which we briefly discuss in section 1.10, p. 21.

Slightly related, there is Thouless' theorem [19]. 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 $\langle {}^{w}\Phi^{\mathrm{SD}}|{}^{x}\Phi^{\mathrm{SD}}\rangle = 0$:

$$|{}^{x}\Phi^{\mathrm{SD}}\rangle \propto e^{\hat{T}_1} |{}^{w}\Phi^{\mathrm{SD}}\rangle \tag{1.54}$$

where

$$\hat{T}_1 = \sum_{ia} t_i^a a_a^\dagger a_i \tag{1.55}$$

and the parameters t_i^a need to be determined.

1.6 Configuration interaction

Because 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 rise to the exact wave function within a basis set

$$|\Psi^{\mathrm{FCI}}\rangle = \sum_n c_n^{\mathrm{FCI}} |\Phi_n^{\mathrm{SD}}\rangle \tag{1.56}$$

where the sum goes over all basis Slater determinants (e.g. $\binom{M}{N}$ elements) and $\mathbf{c}^{\mathrm{FCI}}$ is chosen such that the expectation energy is minimized

$$\mathbf{c}^{\mathrm{FCI}} = \arg\min_{\mathbf{c}} \left(\sum_{nm} c_m c_n^* \langle \Phi_n^{\mathrm{SD}}| \hat{H} |\Phi_m^{\mathrm{SD}}\rangle \right) \quad \text{subject to} \quad \mathbf{c}^\dagger \mathbf{c} = 1. \tag{1.57}$$

This method is called *full configuration interaction (FCI)*. In obtaining the FCI solution, a matrix of dimension $\binom{M}{N} \times \binom{M}{N}$ needs to be diagonalized. While there exist very efficient ways to avoid diagonalizing the whole matrix and symmetry can be taken into account [10, ch. 11], this approach is nevertheless not feasible even for small molecules. We can consider approximations where the wave function is written as a linear combination of some Slater determinants, which is called *configuration interaction*. In CISD, one writes the wave function as a linear combination of the HF state and all singly and doubly *excited determinants* (*singles* and *doubles*), e.g. determinants that arise when replacing one or two occupied MOs by virtual MOs. 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, which seems reasonable, as doubles should bring the most important contribution to the energy. However, CISD is not size-consistent, meaning that the combined CISD energy of two isolated systems A and B is not identical to the energy of the super-system A and B at infinite separation: $E_{\mathrm{CISD}}(A+B) \neq E_{\mathrm{CISD}}(A) + E_{\mathrm{CISD}}(B)$. 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, canonical orbitals are 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 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 describe the most important aspects of CC theory following Ref. [10, ch. 13], as well as 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 HF ground state as it serves as a good reference and gives computational advantages compared to arbitrary Slater determinants, but the theory remains the same independent of the choice of the reference determinant $|\Phi^{\mathrm{SD}}\rangle$. We refer to operators that act on a reference Slater determinant by replacing one or more occupied MOs by one or more virtual ones as *excitation operators* producing excited determinants. Linear combinations of excitation operators will also be referred to as excitation operators. Sometimes, a shorthand notation for singles $|\Phi_i^a\rangle \equiv a_a^\dagger a_i |\Phi^{\mathrm{SD}}\rangle$ and doubles $|\Phi_{ij}^{ab}\rangle \equiv a_a^\dagger a_b^\dagger a_j a_i |\Phi^{\mathrm{SD}}\rangle$ is used.

1.7.1 The exponential ansatz

In coupled cluster theory, the wave function takes the form

$$|\Psi\rangle = e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = \sum_{n=0}^{\infty} \frac{1}{n!} \hat{T}^n |\Phi^{\mathrm{SD}}\rangle \tag{1.58}$$

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

$$\hat{T} = \sum_{i=1}^{\infty} T_i = \sum_{\mu} t_\mu \hat{\tau}_\mu = \sum_{ia} t_i^a a_a^\dagger a_i + \sum_{iajb} \frac{1}{4} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i + \dots \tag{1.59}$$

and $\hat{\tau}_\mu |\Phi^{\mathrm{SD}}\rangle = |\mu\rangle$ represents an excited determinant of any order. Sometimes, we denote the collection of all amplitudes t_i^a, t_{ij}^{ab} by a vector $\mathbf{t}$. It should be noted that the individual 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

$$\hat{H} |\Psi\rangle = \hat{H} e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = E_{\mathrm{CC}} e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle = E_{\mathrm{CC}} |\Psi\rangle . \tag{1.60}$$

As the reference Slater determinant $|\Phi^{\mathrm{SD}}\rangle$ is normalized, we have intermediate normalization ($\langle\Phi^{\mathrm{SD}}|\Psi\rangle = 1$), and

$$\langle\Phi^{\mathrm{SD}}| \hat{H} |\Psi\rangle = E_{\mathrm{CC}}. \tag{1.61}$$

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 the *energy equation*

$$\langle\Phi^{\mathrm{SD}}| \bar{H} |\Phi^{\mathrm{SD}}\rangle = E_{CC} \tag{1.62}$$

and the *amplitude equations*

$$\langle\mu| \bar{H} |\Phi^{\mathrm{SD}}\rangle = 0. \tag{1.63}$$

This set of equations is called the *linked coupled-cluster equations*.

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

$$\hat{T} = \sum_{i=1}^{N} \hat{T}_i \tag{1.64}$$

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

$$\hat{T} = \sum_{i=1}^{k} \hat{T}_i = \sum_{\mu \in S,D,T,\dots,K} t_\mu \hat{\tau}_\mu \tag{1.65}$$

where $k < N$, and the notation $\mu \in \{S, D, T, ...K\}$ for all possible single, double, ..., k-tuple excitations is used. 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 $\binom{M}{N}$ basis determinants, unlike truncated CI expansions. The exponential ansatz makes the CCSD wave function size-consistent, which is one of the biggest advantages of the method. As the cluster operator is truncated, the variational solution, with $|\tilde{\Psi}\rangle = \exp\left(\sum_{i=1}^{k} \hat{T}_i\right) |\Phi^{\mathrm{SD}}\rangle$ that minimizes the expectation value of the energy

$$E_{VCC} = \min_{\hat{T}} \frac{\langle\tilde{\Psi}|\,\hat{H}\,|\tilde{\Psi}\rangle}{\langle\tilde{\Psi}|\tilde{\Psi}\rangle} \tag{1.66}$$

will not be the true ground state, and we cannot expect equations (1.62) and (1.63) to be correct for that state. However, as eq. (1.66) becomes intractable to solve, the default approach is to nevertheless solve eqs. (1.62) and (1.63), where in the CCSD model, 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 Refs. [20], [21] that solve the variational equation (1.66), giving rise 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

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}. \tag{1.67}$$

To evaluate this, one can 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

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{4!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}] + \cdots. \tag{1.68}$$

Direct calculation of the commutators using commutator rules for creation and annihilation operators [10, 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

$$\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}} = \hat{H} + [\hat{H}, \hat{T}] + \frac{1}{2!}[[\hat{H}, \hat{T}], \hat{T}] + \frac{1}{3!}[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}] + \frac{1}{4!}[[[[\hat{H}, \hat{T}], \hat{T}], \hat{T}], \hat{T}]. \tag{1.69}$$

1.7.4 Solving the coupled cluster equations

In order to determine the excitation coefficients t_μ, the amplitude equations (1.63) 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, which simplifies the evaluation of the equations. Explicit expressions are then obtained 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 page-long. The derivation of these equations can also done using computer-assisted tools or diagrammatic methods. In particular, the `Drudge` and `Gristmill` packages [22]

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 common 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 (with the number of occupied orbitals M_o being identical to the number of electrons N, of course). An explicit form of the resulting equations is given in, for example, Ref. [23], where symmetry aspects are considered too. In the iterative process, Newton's method can be used with an initial guess of the amplitudes based on the MP1 wave function (described in section 1.8), where it is usual that the Jacobian

$$J_{\mu\nu}(\mathbf{t}) = \frac{\partial}{\partial \nu} \langle \mu | \bar{H}(\mathbf{t}) | \Phi^{\mathrm{SD}} \rangle = \langle \mu | e^{-\hat{T}(\mathbf{t})} \left[\hat{H}, \hat{\tau}_\nu \right] e^{\hat{T}(\mathbf{t})} | \Phi^{\mathrm{SD}} \rangle \tag{1.70}$$

is approximated by a constant diagonal matrix forming the dominating part of the Jacobian when using canonical orbitals, leading to a Quasi-Newton method:

$$J_{\mu\mu}(\mathbf{t}) = \epsilon_\mu \tag{1.71}$$

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

1.7.5 Equation of motion coupled cluster

For a truncated cluster operator $\hat{T}$, the state $|\Psi\rangle = e^{\hat{T}} |\Phi^{\mathrm{SD}}\rangle$ is an approximation to the ground state of the Hamiltonian $\hat{H}$. However, we can just as well consider $|\Phi^{\mathrm{SD}}\rangle$ to be the exact ground state of the truncated similarity-transformed Hamiltonian $\bar{H}$. In the *equation of motion coupled cluster* (EOM-CC) method, 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

$$\hat{P} = \sum_{\mu \in S,D,\ldots,K} |\mu\rangle\langle\mu| + |\Phi^{\mathrm{SD}}\rangle\langle\Phi^{\mathrm{SD}}| \tag{1.72}$$

$$\bar{H}' = \hat{P}\bar{H}\hat{P} \tag{1.73}$$

giving rise to a matrix $\bar{\mathbf{H}}'$ with matrix elements $\bar{H}'_{\mu\nu} = \langle \mu | \bar{H} | \nu \rangle$. There is a close connection between these matrix elements and the CC Jacobian (eq. (1.70)), as we can write

$$\bar{H}'_{\mu\nu} = \langle \mu | \bar{H} | \nu \rangle = \langle \mu | e^{-\hat{T}} \left[\hat{H}, \hat{\tau}_\nu \right] e^{\hat{T}} | \Phi^{\mathrm{SD}} \rangle + \langle \mu | \hat{\tau}_\nu \bar{H} | \Phi^{\mathrm{SD}} \rangle = J_{\mu\nu} + \delta_{\mu\nu} E_{\mathrm{CC}}. \tag{1.74}$$

Because the amplitude equations (1.63) holds, we have that $\langle \mu | \bar{H}' | \Phi^{\mathrm{SD}} \rangle = 0$, and $|\Phi^{\mathrm{SD}}\rangle$ 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

$$\langle \tilde{\Psi} | E_{CC} = \langle \tilde{\Psi} | \bar{H}' \tag{1.75}$$

where

$$\langle \tilde{\Psi} | = \langle \Phi^{\mathrm{SD}} | \left(\mathbb{1} + \hat{\Lambda}_1 + \hat{\Lambda}_2 + \ldots \right) = \langle \Phi^{\mathrm{SD}} | \left(\mathbb{1} + \sum_{ia} \lambda_i^a a_i^\dagger a_a + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab} a_i^\dagger a_j^\dagger a_a a_b + \ldots \right) \tag{1.76}$$

and we see that $\langle \tilde{\Psi} | \Psi \rangle = 1$. When truncated, the similarity-transformed Hamiltonian is expressed in the basis of excited determinants up to order k, 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

$$\begin{aligned} \langle \tilde{\Psi}^{(j)} | \bar{H}' &= \langle \tilde{\Psi}^{(j)} | E^{(j)} \\ \bar{H}' | \Psi^{(j)} \rangle &= E^{(j)} | \Psi^{(j)} \rangle \end{aligned} \tag{1.77}$$

where

$$\begin{aligned} \langle \tilde{\Psi}^{(j)}| &= \langle \Phi^{\mathrm{SD}}| \, \hat{\mathcal{L}}^{(j)} \\ |\Psi^{(j)}\rangle &= \hat{\mathcal{R}}^{(j)} \, |\Phi^{\mathrm{SD}}\rangle \end{aligned} \tag{1.78}$$

are *biorthogonal*:

$$\langle \tilde{\Psi}^{(j)} | \Psi^{(k)} \rangle = \delta_{jk} \tag{1.79}$$

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

$$e^{\hat{T}} \hat{\mathcal{R}}^{(j)} \, |\Phi^{\mathrm{SD}}\rangle = \hat{\mathcal{R}}^{(j)} e^{\hat{T}} \, |\Phi^{\mathrm{SD}}\rangle \tag{1.80}$$

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. 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 $k < N$. The full similarity-transformed Hamiltonian matrix itself is a similarity transformation of the Hamiltonian matrix, an operation that conserves eigenvalues.

1.7.6 Non-canonical reference determinants

In this thesis, we consider the use of reference states $|\Phi^{\mathrm{HF}}\rangle$ constructed from non-canonical MOs, with a non-diagonal, but block-diagonal Fock matrix $F_{ij} \neq 0$, $F_{ab} \neq 0$, $F_{ai} = 0$. Those give the same CC energies as when using canonical orbitals. We also consider reference determinants $|\Phi^{\mathrm{SD}}\rangle$ that are not HF states, with non-block diagonal Fock matrices $F_{ai} \neq 0$. One might expect that those are not suitable reference states for coupled cluster calculations. However, by Thouless' theorem (cf. sec. 1.5), any Slater determinant $|\Phi^{\mathrm{SD}}\rangle$ can be turned into another Slater determinant $|\tilde{\Phi}^{\mathrm{SD}}\rangle$ by application of an $e^{\hat{T}_1}$ excitation operator: $|\tilde{\Phi}^{\mathrm{SD}}\rangle \propto e^{\hat{T}_1} \, |\Phi^{\mathrm{SD}}\rangle$ given that $\langle \tilde{\Phi}^{\mathrm{SD}} | \Phi^{\mathrm{SD}} \rangle \neq 0$. This makes 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

In the *unitary coupled cluster* (UCC) method [24], the wave function is parametrized as

$$|\Psi\rangle = e^{\hat{\sigma}} \, |\Phi^{\mathrm{SD}}\rangle \tag{1.81}$$

and the similarity-transformed Hamiltonian, which is Hermitian, reads

$$\bar{H} = e^{-\hat{\sigma}} \hat{H} e^{\hat{\sigma}} \tag{1.82}$$

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

$$E_{\mathrm{UCC}} = \min_{\hat{T}} \left(\langle \Phi^{\mathrm{SD}}| \, e^{-\hat{\sigma}} \hat{H} e^{\hat{\sigma}} \, |\Phi^{\mathrm{SD}}\rangle \right) \tag{1.83}$$

as the wave function is normalized independently of $\hat{T}$: $\langle \Psi | \Psi \rangle = \langle \Phi^{\mathrm{SD}} | \Phi^{\mathrm{SD}} \rangle = 1$. It has been theorized that UCC and VCC (cf. eq. (1.66)) have the same (global) energy minimum, but both theoretical [20] and practical [20], [21] 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 when the same truncation k for the cluster operator is used [24].

On classical computers, the unitary coupled cluster method is not used as a standard method because the similarity-transformed Hamiltonian (eq. (1.82)) has a BCH-expansion that does not terminate and no projection approach as in conventional coupled cluster theory (cf. eq. (1.63)) is possible, even when the operator $\hat{T}$ is restricted to single and double excitations, as the operator $e^{-\left(\hat{T}-\hat{T}^\dagger\right)}$ is neither a pure excitation or de-excitation operator. However, truncated 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. 34.

1.8 Perturbation theory

An alternative to CISD and CCSD is to use perturbation theory [10, ch. 14]. In *Møller-Plesset* (MP) perturbation theory, we write the Hamiltonian as

$$\hat{H} = \hat{F} + \hat{\Phi} \tag{1.84}$$

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

$$E(\lambda) = \sum_{n=0}^{\infty} \frac{1}{n!}\lambda^n \left.\frac{\partial^n E(\lambda)}{\partial\lambda^n}\right|_{\lambda=0} = \sum_{n=0}^{\infty} \lambda^n E^{(n)} \tag{1.85}$$

$$|\Psi(\lambda)\rangle = \sum_{n=0}^{\infty} \frac{1}{n!}\lambda^n \left.\frac{\partial^n |\Psi(\lambda)\rangle}{\partial\lambda^n}\right|_{\lambda=0} = \sum_{n=0}^{\infty} \lambda^n |\Psi^{(n)}\rangle \tag{1.86}$$

where λ is set to 1 to obtain the desired solution and we assumed intermediate normalization, e.g. $\langle\Phi^{\mathrm{HF}}|\Psi(\lambda)\rangle = 1$. We do not go through the derivation, but only present the solution to the first-order correction of the wave function. Remembering that $|\Psi^{(0)}\rangle = |\Phi^{\mathrm{HF}}\rangle$ is a Slater determinant and using canonical orbitals, the first order correction to the wave function is

$$|\mathrm{MP1}\rangle = |\Psi^{(1)}\rangle = \hat{T}_2^{(1)} |\Psi^{(0)}\rangle \tag{1.87}$$

$$\hat{T}_2^{(1)} = \sum_{\mu\in D} t_\mu^{(1)} \hat{\tau}_\mu = \sum_{\substack{a>b \\ i>j}} t_{ij}^{ab(1)} a_a^\dagger a_i a_b^\dagger a_j \tag{1.88}$$

where

$$t_{ij}^{ab(1)} = -\frac{\langle\Psi^{(0)}| \left[a_j^\dagger a_b a_i^\dagger a_a, \hat{H}\right] |\Psi^{(0)}\rangle}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j} \tag{1.89}$$

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

$$E^{(2)} = \langle\Psi^{(0)}|\hat{H}|\Psi^{(1)}\rangle\,. \tag{1.90}$$

Obtaining $E^{(2)}$ and $|\Psi^{(1)}\rangle$ generally scales as $O(M^5)$ and is cheaper than CCSD, and MP perturbation theory is the method of choice to obtain a first improvement of the energy and the wave function. Energies are not variational, and the radius of convergence of the Taylor series 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 (normalized) exact ground state wave function $|\Psi\rangle$. The corresponding wave function in position-spin space is given by

$$\langle \vec{\mathbf{x}} | \Psi \rangle = \Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{1.91}$$

We define the *density matrix* [25]

$$\Gamma(\mathbf{x}'_1, \ldots, \mathbf{x}'_N; \mathbf{x}_1, \ldots, \mathbf{x}_N) = \Psi^*(\mathbf{x}'_1, \ldots, \mathbf{x}'_N)\Psi(\mathbf{x}_1, \ldots, \mathbf{x}_N) \tag{1.92}$$

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

$$\begin{aligned} &\Gamma^{(p)}(\mathbf{x}'_1, \ldots, \mathbf{x}'_p; \mathbf{x}_1, \ldots, \mathbf{x}_p) = \\ &\binom{N}{p} \int \Gamma(\mathbf{x}'_1, \mathbf{x}'_2, \ldots, \mathbf{x}'_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_N; \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p, \mathbf{x}_{p+1}, \ldots, \mathbf{x}_N) d\mathbf{x}_{p+1} \ldots d\mathbf{x}_N \end{aligned} \tag{1.93}$$

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

$$\gamma(\mathbf{x}', \mathbf{x}) = \sum_{pq}^{M} {}^1D_{pq} \phi_p^*(\mathbf{x}') \phi_q(\mathbf{x}) \tag{1.94}$$

where the matrix ${}^1\mathbf{D}$ also is called the 1-RDM. In second quantization, the 1-RDM can be expressed as [10, ch. 1]

$${}^1D_{pq} = \langle \Psi | \, a_q^\dagger a_p \, | \Psi \rangle \tag{1.95}$$

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, $\mathrm{Tr}\left({}^1\mathbf{D}\right) = N$. Being a Hermitian matrix, it can be diagonalized,

$${}^1\mathbf{D} = \mathbf{U}\mathbf{d}\mathbf{U}^\dagger. \tag{1.96}$$

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 [25]

$$|\xi_p\rangle = \sum_{q=1}^{M} |\phi_q\rangle \, U_{qp}. \tag{1.97}$$

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

$$\hat{N} = \sum_{p=1}^{M} a_p^\dagger a_p. \tag{1.98}$$

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 are "important" orbitals, in the sense that the expansion coefficients belonging to Slater determinants that contain mostly NOs with large occupation numbers, are comparatively large in absolute value. 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 [26] and CCSD [27]. 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 Hartree-Fock reference remains unchanged. For larger basis sets, it was shown that a larger percentage of NOs can be discarded. Calculations where natural orbitals with occupation numbers below some threshold are removed, 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 [28], and we simply call them NOs.

1.9.1 Analyticity of the NOs

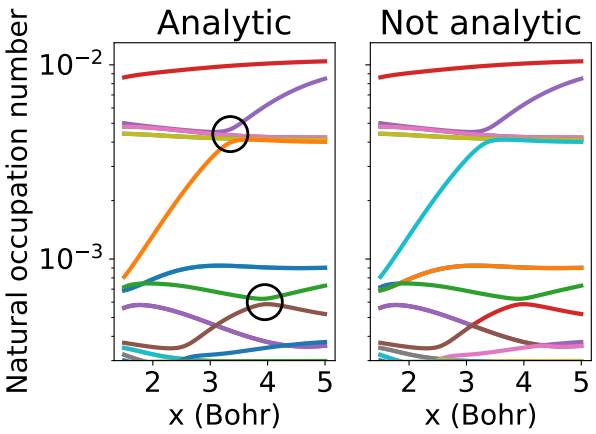


Figure 1.1: Trajectory of some natural occupation numbers as function of internuclear distance x for hydrogen fluoride in the cc-pVTZ basis set. Left: Crossing analytic functions for the eigenvalues. Right: Ordering of the natural occupation numbers by value, so they cross. The ordered eigenvalues are not analytic at crossing points. Two avoided crossings are marked with black circles.

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 one-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 natural occupation numbers can cross. When insisting on the ordering $d_{11}(\vec{R}) \geq \cdots \geq d_{MM}(\vec{R})$ for all $\vec{R}$, the functions $|\xi_p(\vec{R})\rangle$ are not necessarily analytic because of those crossings, and neither are the eigenvalues. 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 eigenvalues, some algorithm needs to be employed to ascertain the analytic ordering. This is not trivial and is discussed in the numerical methods part of the thesis, sec. 5.3, p. 64. We also marked two avoided crossings in the left graph. Those avoided crossings play an important role later, as they impact how natural orbitals perform in eigenvector continuation, which we discuss in chapter 4.

1.10 Non-orthogonal multi-reference methods

All methods considered so far are single-reference methods that aim at improving the quality of the wave function represented by a Slater determinant (usually the HF state)

through successive addition of excited determinants, either in a linear (CI) or in an exponential fashion (CC). They are *single-reference* methods. *Multi-reference* methods put several determinants on an equal footing. In the *multi-configurational self-consistent field* (MCSCF) method, the wave function is expressed as sum of Slater determinants $|\Psi_{MCSCF}\rangle = \sum_n c_n |\Phi_n^{\mathrm{SD}}\rangle$, optimizing not just the expansion coefficients c_n, but also the set of MOs that enter into the Slater determinants $|\Phi_n^{\mathrm{SD}}\rangle$ [7, ch. 4]. In the *CASPT* method, perturbation theory is applied to a multi-reference wave function [10, ch. 12].

Further generalizations are possible when allowing two Slater determinants to be non-orthogonal, meaning that they are constructed using different sets of MOs. In the *NOCI* method [29], the wave function is expressed as a linear combination of non-orthogonal Slater determinants. In the NOCISD-approach [30], 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* method [29], 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 [31].

Chapter 2

Eigenvector continuation

This chapter introduces eigenvector continuation, its implementation and its mathematical justification. It is discussed how it performs compared to perturbation theory, how it has been used to solve quantum mechanical problems in previous studies and how a variant of it has been adapted to work with coupled cluster wave functions. We use bra-ket notation for vectors $\mathbf{v} \equiv |\psi\rangle$ and their conjugate transpose $\mathbf{v}^\dagger \equiv \langle\psi|$ and call the smallest eigenvalue of a Hermitian matrix *ground state eigenvalue* and the corresponding eigenvector *ground state*. The mathematical behaviour and justification of eigenvector continuation requires knowledge about perturbation theory which would require several pages of introduction, which we do not do, instead presenting only 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 $\lambda(\alpha)$ behave as a function of α. We consider finite-dimensional matrices $\mathbf{M}(\alpha)$, following chapter two in Ref. [32]. No results are proved, we only state those results that are most important for the following discussions.

1. Eigenvalues $\lambda(\alpha)$ are continuous everywhere, and analytic everywhere except at a finite number of *exceptional points*. These exceptional points lie where the trajectory of at least two eigenvalues coincide [32, Ch. 2, §1].

2. The eigenprojections $\mathbf{P}(\alpha)$, e.g. the projection operators that project vectors onto eigenvectors of $\mathbf{M}(\alpha)$, are analytic everywhere except possibly at the exceptional points for that eigenvalue. The eigenvectors can thus be chosen to be analytic everywhere except at those exceptional points [32, Ch. 2, §1].

3. The convergence radius of perturbation series for the eigenvalues and eigenprojections (and hence eigenvectors) 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 [32, Ch. 2, §3].

4. The eigenvalues and eigenvectors of Hermitian operators $\mathbf{H}(\alpha) = \mathbf{H}^\dagger(\alpha)$ can be chosen to be analytic on the real axis (even when eigenvalues collide at special points) [32, 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, which is the case when two eigenvalues $\lambda_1(\alpha), \lambda_2(\alpha)$ come very close on the real line (avoided crossings).

5. In general, the aforementioned points still hold for matrices that depend on more than one parameter $\mathbf{M}(\vec{\alpha})$ [32, 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) [2] is a novel, variational method to estimate the ground state $|\psi(\vec{\alpha})\rangle$ and the corresponding eigenvalue $\lambda_0(\vec{\alpha})$ of a Hermitian matrix $\mathbf{H}(\vec{\alpha})$ analytic in $\vec{\alpha}$. While Ref. [2] only considered an operator 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 matrix functions $\mathbf{H}(\vec{\alpha})$ that are not linear in $\vec{\alpha}$ [1]. Instead of explicitly diagonalizing $\mathbf{H}(\vec{\alpha})$ at each value $\vec{\alpha}_\odot$ of interest, the eigenvector continuation method approximates the ground state $|\psi(\vec{\alpha}_\odot)\rangle$ as a linear combination of the form

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

where $|\psi(\vec{\alpha}_i)\rangle$ are (approximate) ground states for a predefined set of sample points $\vec{\alpha}_i$, $i = 1, \ldots, L$. The coefficients $\mathbf{c} = \begin{bmatrix} c_1 & \ldots & c_L \end{bmatrix}^T$ are chosen such that the expectation value

$$\lambda_0^{\mathrm{EVC}}(\vec{\alpha}_\odot) = \min_{\mathbf{c}} \frac{\langle\psi^{\mathrm{EVC}}(\vec{\alpha}_\odot)| \, \mathbf{H}(\vec{\alpha}_\odot) \, |\psi^{\mathrm{EVC}}(\vec{\alpha}_\odot)\rangle}{\langle\psi^{\mathrm{EVC}}(\vec{\alpha}_\odot)|\psi^{\mathrm{EVC}}(\vec{\alpha}_\odot)\rangle} \tag{2.2}$$

is minimized. This is accomplished by projecting the true matrix $\mathbf{H}(\vec{\alpha}_\odot)$ on the space spanned by the ground states $|\psi(\vec{\alpha}_i)\rangle$ giving rise to the matrices $\mathbf{H}^{\mathrm{EVC}}$ and $\mathbf{S}^{\mathrm{EVC}}$ with elements

$$\begin{aligned} H_{ij}^{\mathrm{EVC}} &= \langle\psi(\vec{\alpha}_i)| \, \mathbf{H}(\vec{\alpha}_\odot) \, |\psi(\vec{\alpha}_j)\rangle \\ S_{ij}^{\mathrm{EVC}} &= \langle\psi(\vec{\alpha}_i)|\psi(\vec{\alpha}_j)\rangle \end{aligned} \tag{2.3}$$

and solving the generalized eigenvalue problem $\mathbf{H}^{\mathrm{EVC}}\mathbf{C} = \mathbf{S}^{\mathrm{EVC}}\mathbf{C}\boldsymbol{E}$. We then obtain the eigenvectors $\mathbf{C}_i$ and the eigenvalues E_{ii} of that generalized eigenvalue problem. The column vector $\mathbf{C}_k$ corresponding to the smallest E_{kk} contains the coefficients $\mathbf{c}$, while $E_{kk} = \lambda_0^{\mathrm{EVC}}(\vec{\alpha}_\odot)$ minimizes the expectation value.

One possible use of eigenvector continuation is to use as sample points a set of $\vec{\alpha}_i \in I$, where I is a region where an approximate $|\psi(\vec{\alpha}_i)\rangle$ is "easy" to obtain with a specific approximation method, while the target point(s) $\vec{\alpha}_\odot$ lie in a region where the approximation method fails. EVC can also be used as an effective interpolation method [1], and thus an effective dimensionality reduction tool. Figure 2.1 illustrates how EVC performs for a system with a relatively complicated shape of the eigenvalues and several avoided crossings. Using only 5 sample points gives a good approximation to the true ground state for both interpolation and extrapolation, but while the interpolation result is essentially exact, we see that the extrapolation solution gets worse after passing through avoided crossings, e.g. when exceptional points are nearby. This is because the state which was an excited state before an avoided crossing becomes qualitatively similar to the ground state after said avoided crossing [33].

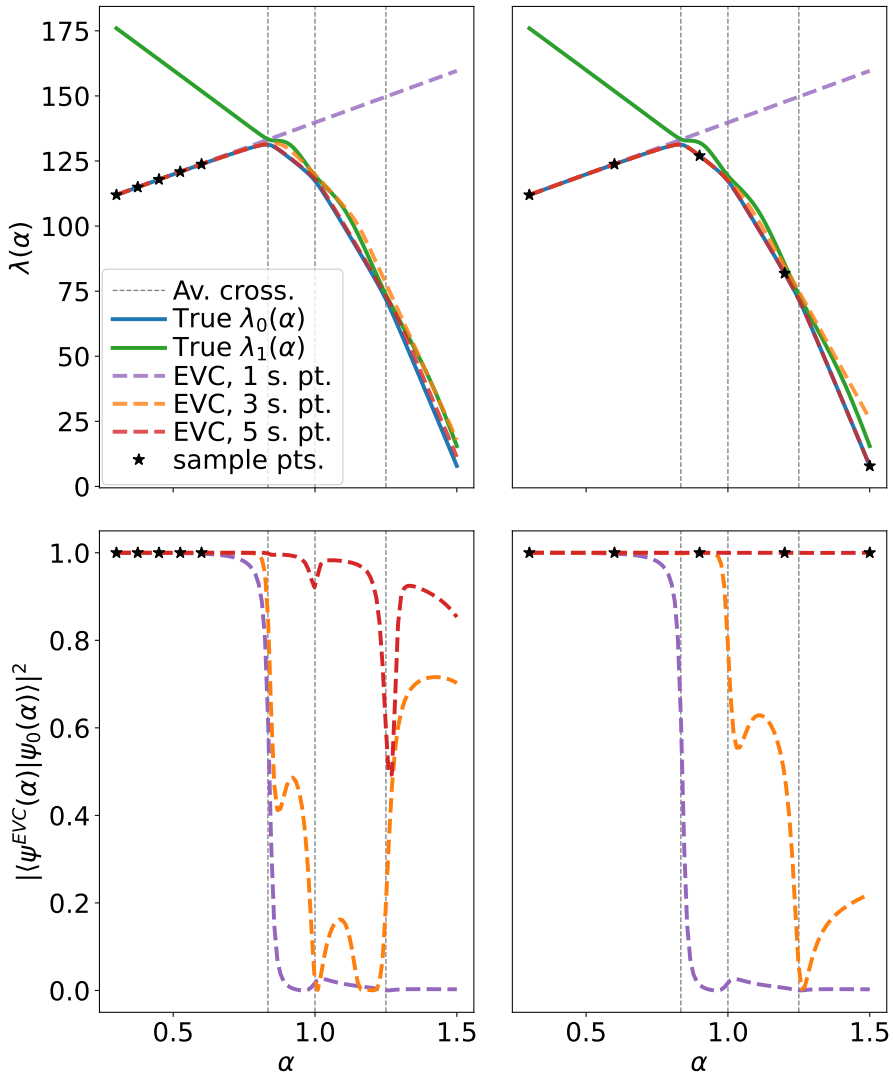


Figure 2.1: Eigenvector continuation of the ground state using 1, 3 and 5 sample points (using the leftmost ones). Avoided crossings are marked. Top: Eigenvalues $\lambda(c)$ and EVC approximations to the ground state eigenvalue. Bottom: Overlap between the true ground state and the EVC approximation $\left|\langle\psi^{EVC}(\alpha)|\psi(\alpha)\rangle\right|^2$. Left: Sampling tightly in the same region. Right: Sampling across the whole curve. The matrix function giving rise to these eigenvalues is a 500×500 *matrix of the type* $\mathbf{M}(\alpha) = \mathbf{M}_0 + \alpha\mathbf{M}_1$ *and can be found in sec. A.1 in the appendix.*

2.2.1 Mathematical justification for extrapolation

Let $\mathbf{H}(\alpha)$ depend on a single parameter α. This justification follows Ref. [2] closely. As $\mathbf{H}(\alpha)$ is Hermitian and analytic, 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 $\alpha = 0$, the ground state and the corresponding eigenvalue can be expanded as a Taylor series

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

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 (2.4) will converge for $|\alpha| < |z|$. Let now w be a

real number such that $|w| < |z|$. It is possible to choose w such that $|w - |z|| < |w - z|$ and perform an analytic continuation, expanding the eigenvector-eigenvalue pair around $\alpha = w$:

$$\begin{aligned} |\psi(\alpha)\rangle &= \sum_{n=0}^{\infty} \frac{(\alpha - w)^n}{n!} |\psi^{(n)}(w)\rangle \\ E(\alpha) &= \sum_{n=0}^{\infty} \frac{(\alpha - w)^n}{n!} E^{(n)}(w) \end{aligned} \tag{2.5}$$

The series will converge for $|\alpha - w| < |z - w|$, given that z is still the closest singularity from the point w. The derivative vectors at $|\psi^{(m)}(w)\rangle$ can be expressed as a linear combinations of the derivative vectors $|\psi^{(n)}(0)\rangle$. Thus, one can approximate $|\psi(\alpha)\rangle$ as a linear combination of the derivative vectors $|\psi^{(n)}(0)\rangle$ also in the region $|\alpha - w| < |z - w|$. As this procedure can be repeated as often as desired, one can express any $|\psi(\alpha)\rangle$ as a linear combination of the vectors $|\psi^{(n)}(0)\rangle$. For any given accuracy ϵ, the number of necessary derivative vectors $|\psi^{(n)}(0)\rangle$ to include will be finite and depend on the number of expansions needed as well as the order-by-order convergence of each individual series expansion. Eigenvector continuation works because it is not necessary to use the derivative vectors $|\psi^{(n)}(0)\rangle$ themselves, but vectors that span the same vector space. This is precisely the case for $|\psi(\alpha_1)\rangle, \ldots, |\psi(\alpha_L)\rangle$ when the sample points $\alpha_1, \ldots, \alpha_L$ lie infinitesimally close to $\alpha = 0$, justifying why eigenvector continuation works for extrapolation.

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 $|\psi^{(i)}(0)\rangle$. In Ref. [3], it has been proven that eigenvector continuation converges faster than perturbation theory when these sample points are chosen. This is because the derivative vectors in perturbation theory are not mutually orthogonal, so there is constructive and destructive interference that leads to a slower convergence. This is called *Differential Folding* and does not occur for eigenvector continuation - although the eigenvector continuation sample vectors span the same vector space and are not mutually orthogonal either, the expansion coefficients are optimized with respect to energy. There is no destructive interference. This also explains why eigenvector continuation can give good results outside of the radius of convergence of perturbation theory. Adding more sample vectors does not increase the approximate eigenvalue, while perturbation series might show a rather erratic convergence.

In Ref. [34] a partitioning of a many-body Hamiltonian as $\mathbf{H}(\alpha) = \mathbf{H}^0 + \alpha\mathbf{H}^1$ was considered, where $\mathbf{H} = \mathbf{H}(1)$ is the exact Hamiltonian and exact eigenvectors of $\mathbf{H}^0$ are readily available. Corrections to the wave function up to p^{th} order $\{|\psi^{(i)}\rangle\}_{i=0}^{p}$ were found, and the EVC-Hamiltonian was constructed as $H_{ij}^{\text{EVC}} = \langle\psi^{(i)}|\,\mathbf{H}\,|\psi^{(j)}\rangle$ for $i, j = 0, \ldots, p$ and the overlap matrix correspondingly. While the perturbation series diverged, rapidly and monotonic convergence using eigenvector continuation were observed, being a resummation of the individual terms.

2.2.3 Behaviour for approximate sample eigenvectors

In this thesis, we use eigenvector continuation with approximate ground states, e.g. the sample vectors $|\psi(\vec{\alpha}_i)\rangle$ will only be approximations to the true ground state at $\vec{\alpha}_i$. It is possible that these approximations are qualitatively wrong. The EVC algorithm remains unaltered when using approximate sample eigenvectors. Furthermore, because $\mathbf{H}(\vec{\alpha})$ is Hermitian, we can use the Hylleraas-Undheim theorem [35] and the variational principle to reason that the approximate eigenvalue $\lambda_0^{\text{EVC}}(\vec{\alpha}_\odot)$ obtained using EVC is always an upper

bound for the true eigenvalue, and that the inclusion of approximate or wrong eigenstates in the EVC sample space does not lead to an increase in the approximate EVC eigenvalue. This also means that EVC is exact whenever the true ground state $|\psi(\vec{\alpha}_\odot)\rangle$ can be written as a linear combination of (some of) the sample vectors.

2.3 Coupled cluster eigenvector continuation

When applying eigenvector continuation to coupled cluster wave functions, two different approaches are considered in this thesis. The first method is to approximate the wave function as a linear combination of sample coupled cluster wave functions, called *wave function method* (WF-CCEVC). In the methods section (sec. 4.1), we introduce a second method, where the cluster operator is written as a linear combination of sample cluster operators, called *amplitude method* (AMP-CCEVC).

2.3.1 WF-CCEVC

A direct application of eigenvector continuation (eq. (2.3)) to a coupled cluster wave function $|\Psi(\vec{\alpha})\rangle = e^{\hat{T}(\vec{\alpha})}|\Phi^{\mathrm{SD}}\rangle$ would require the evaluation of the following matrix elements and overlaps

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

$$\langle\Psi(\vec{\alpha}_i)|\hat{H}(\vec{\alpha}_\odot)|\Psi(\vec{\alpha}_j)\rangle = \langle\Phi^{\mathrm{SD}}|\, e^{\hat{T}^\dagger(\vec{\alpha}_i)} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}(\vec{\alpha}_j)}\,|\Phi^{\mathrm{SD}}\rangle\,. \tag{2.7}$$

As evaluating these functions is as expensive as evaluating matrix elements for full CI states, Ekström and Hagen [4] have considered how eigenvector continuation can be applied to coupled cluster wave functions more efficiently. They applied a biorthogonal approach reminiscent of EOM-CC, where the left and the right eigenvectors of the Hamiltonian differ, with

$$\langle\tilde{\Psi}(\vec{\alpha})| = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}(\vec{\alpha})\right] e^{-\hat{T}(\vec{\alpha})} \tag{2.8}$$

$$|\Psi(\vec{\alpha})\rangle = e^{\hat{T}(\vec{\alpha})}|\Phi^{\mathrm{SD}}\rangle\,. \tag{2.9}$$

$\hat{\Lambda}(\vec{\alpha})$ can be found by solving the left ground state eigenvalue problem of the full EOM-CC Hamiltonian (cf. sec. 1.7.5, p. 17) or the coupled cluster Lagrangian [10, ch. 13]. The EVC overlap and the Hamiltonian at the target reference are then expressed as

$$S_{ij}^{\mathrm{EVC}} = \langle\tilde{\Psi}'|\Psi\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}'\right] e^{-\hat{T}'} e^{\hat{T}}\,|\Phi^{\mathrm{SD}}\rangle \tag{2.10}$$

$$H_{ij}^{\mathrm{EVC}}(\vec{\alpha}_\odot) = \langle\tilde{\Psi}'|\hat{H}(\vec{\alpha}_\odot)|\Psi\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}'\right] e^{-\hat{T}'} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}}\,|\Phi^{\mathrm{SD}}\rangle \tag{2.11}$$

where the dependence of the parameters $\vec{\alpha}_i$ is made implicit, with $\hat{T}$ and $\hat{\Lambda}$ obtained for $\vec{\alpha} = \vec{\alpha}_j$ and $\hat{T}'$ and $\hat{\Lambda}'$ obtained for $\vec{\alpha} = \vec{\alpha}_i$. Letting $\hat{X} = \hat{T} - \hat{T}'$, this can be rewritten as

$$\langle\tilde{\Psi}'|\Psi\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}'\right] e^{\hat{X}}\,|\Phi^{\mathrm{SD}}\rangle \tag{2.12}$$

$$\langle\tilde{\Psi}'|\hat{H}(\vec{\alpha}_\odot)|\Psi\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}'\right] e^{-\hat{T}'+\hat{T}} e^{-\hat{T}} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}}\,|\Phi^{\mathrm{SD}}\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1} + \hat{\Lambda}'\right] e^{\hat{X}} \bar{H}(\vec{\alpha}_\odot)\,|\Phi^{\mathrm{SD}}\rangle \tag{2.13}$$

where the similarity-transformed Hamiltonian depends on $\hat{T}$ only, not $\hat{T}'$

$$\bar{H}(\vec{\alpha}_\odot) = e^{-\hat{T}} \hat{H}(\vec{\alpha}_\odot) e^{\hat{T}}. \tag{2.14}$$

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

$$\begin{aligned}
\hat{T}_1 &= \sum_{ia} t_i^a a_a^\dagger a_i \\
\hat{T}_2 &= \frac{1}{4} \sum_{ijab} t_{ij}^{ab} a_a^\dagger a_b^\dagger a_j a_i \\
\hat{\Lambda}_1 &= \sum_{ia} \lambda_i^a a_i^\dagger a_a \\
\hat{\Lambda}_2 &= \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab} a_i^\dagger a_j^\dagger a_b a_a.
\end{aligned} \tag{2.15}$$

Using Wick's theorem, explicit expressions for eqs. (2.12) and (2.13) can be obtained, yielding

$$\langle \tilde{\Psi}' | \Psi \rangle = 1 + \sum_{ia} \lambda_i^{a\prime} x_i^a + \frac{1}{2} \sum_{ijab} \lambda_{ij}^{ab\prime} x_i^a x_j^b + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab\prime} x_{ij}^{ab} \tag{2.16}$$

$$\begin{aligned}
\langle \tilde{\Psi}' | \hat{H}(\vec{\alpha}_\odot) | \Psi \rangle = \langle \tilde{\Psi}' | \Psi \rangle \langle \Phi^{\mathrm{SD}} | \bar{H}(\vec{\alpha}_\odot) | \Phi^{\mathrm{SD}} \rangle + \sum_{ia} \lambda_i^{a\prime} \langle \Phi_i^a | \bar{H}(\vec{\alpha}_\odot) | \Phi^{\mathrm{SD}} \rangle \\
+ \sum_{ijab} \lambda_{ij}^{ab\prime} x_i^a \langle \Phi_j^b | \bar{H}(\vec{\alpha}_\odot) | \Phi^{\mathrm{SD}} \rangle + \frac{1}{4} \sum_{ijab} \lambda_{ij}^{ab\prime} \langle \Phi_{ij}^{ab} | \bar{H}(\vec{\alpha}_\odot) | \Phi^{\mathrm{SD}} \rangle
\end{aligned} \tag{2.17}$$

where we recognize that $\langle \mu | \bar{H}(\vec{\alpha}_\odot) | \Phi^{\mathrm{SD}} \rangle$ is the left-hand side of the CCSD amplitude equations (1.63). The total scaling of WF-CCEVC is then $O(M_o^2 M_v^4 L)$, where L again is the number of sample points, because L similarity-transformed Hamiltonians need to be calculated for each $\vec{\alpha}_\odot$. However, computational time is saved for small L because the method is not iterative. An explicit derivation for the overlap equation (2.16) is found in the appendix, sec. B.4, p. 97.

It should be mentioned that both the resulting Hamiltonian matrix $\mathbf{H}^{\mathrm{EVC}}$ and the overlap matrix $\mathbf{S}^{\mathrm{EVC}}$ are not Hermitian and that the obtained WF-CCEVC eigenvalues can be complex, as is the case for EOM-CC. However, the imaginary part is usually very small and will be discarded. The resulting wave function will be of the form

$$\begin{aligned}
|\Psi^{\text{WF-CCEVC}}(\vec{\alpha}_\odot)\rangle &= \sum_{m=1}^{L} c_m e^{\hat{T}(\vec{\alpha}_m)} |\Phi^{\mathrm{SD}}\rangle \\
&= \sum_{m=1}^{L} c_m \left(\mathbb{1} + \hat{T}_1(\vec{\alpha}_m) + \frac{1}{2} \hat{T}_1(\vec{\alpha}_m)^2 + \hat{T}_2(\vec{\alpha}_m) + \dots \right) |\Phi^{\mathrm{SD}}\rangle
\end{aligned} \tag{2.18}$$

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 parameterization might not only be usable in terms of eigenvector continuation, but for linear combinations of CC wave functions in general.

A big drawback of WF-CCEVC is that the discussion in section 2.2.3 does not apply, because WF-CCEVC is not a variational method, as the EVC-Hamiltonian is not 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 Hermitian. To solve the generalized eigenvalue problem with Hermitian matrices, one can use canonical orthogonalization [13], [36], where the vanishing eigenvalues and corresponding eigenvalues of the overlap matrix are discarded. In the non-Hermitian case, this problem gets harder [37]. We delay the discussion of suitable algorithms to section 5.1.3.

Chapter 3

Quantum computing

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

3.1 Why quantum computing?

The idea of using 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 [38]. While the computational requirements for representing an arbitrary wave function increase exponentially with the number of particles on regular computers, the resource requirements on a quantum computer are linear [39]. 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, but given the low-polynomial and possibly linear time scaling of recent, very accurate quantum chemistry methods on quantum computers, discussed in sections 3.5 and 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

The content in this section is largely based on Ref. [40].

3.2.1 Qubits

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

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

where α, β are chosen such that $|\psi\rangle$ is normalized. A set of Q qubits spans a 2^Q dimensional Hilbert space of distinguishable particles, 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 bit-string form

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

where the subscripts are usually not written out. Sometimes, a decimal instead of a binary representation is used when convenient, e.g. when $Q = 3$, $|5_{(2)}\rangle = |101\rangle$, and we write $|0^Q\rangle = |0_1 \ldots 0_Q\rangle$. It should be noted that elements in this Hilbert space in general cannot be written as tensor products of states. While $|01\rangle = |0\rangle \otimes |1\rangle$, 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 [41] and quantum mechanics being reversible.[1] Quantum computing is concerned about applying a set of unitary operations to qubits in order to obtain a specific state, and then measuring (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, which 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 $|\Phi\rangle = (\hat{U}_1^\dagger \otimes \hat{U}_2)(\hat{U}_3)(\hat{U}_1 \otimes \mathbb{1} \otimes \mathbb{1})\,|000\rangle$, where $\hat{U}_1$, $\hat{U}_2$ and $\hat{U}_3$ act on 1, 2, and 3 qubits, respectively:

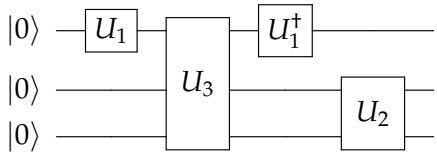
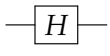
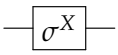
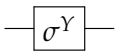
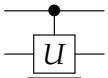
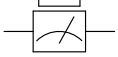


Figure 3.1: Quantum circuit transforming the state $|000\rangle$ into the state $|\Phi\rangle = (\mathbb{1} \otimes \hat{U}_2)(U_1^\dagger \otimes \mathbb{1} \otimes \mathbb{1})(\hat{U}_3)(\hat{U}_1 \otimes \mathbb{1} \otimes \mathbb{1})\,|000\rangle$.

Quantum circuits are written from the left to the right, with gates to the left applied before gates to the right, where *gates* are the physical realization of individual unitary operations and are the building block of quantum circuits. Sometimes, a single *wire* may represent more than one qubit, which is clear from the context. It is customary that a system is initially in the state $|0^Q\rangle$. 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 thereby 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 $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ (in that order).

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 $\tilde{\sigma}^0 = \mathbb{1}$.

- 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, Hermitian 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 of Pauli matrices

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

- Pauli matrices and strings of Pauli matrices have eigenvalues ± 1, which is a direct consequence of them being involutory matrices.
- 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 $|\Psi\rangle$, e.g. measuring $\langle\Psi|\,\tilde{\sigma}_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q}\,|\Psi\rangle$, 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 [40] 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*. Application of the Trotter formula is called *trotterization*.

3.3 Physical devices

3.3.1 Decomposition of unitaries into gates

On physical quantum computers, it is not possible to 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 [40]. Furthermore, using SWAP gates, the problem of not being able to perform 2-qubit operations on arbitrary pairs, can be resolved.

3.3.2 Restrictions due to errors

Quantum algorithms do not work perfectly on physical devices due to imperfect 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 $|0\rangle$. This noise is unpredictable and reduces our knowledge about the system. A standard method to reduce the effect of noise 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* [42] is an active field of research on its own and we do not attempt to describe how it is carried out, but we want to underline that full error correction cannot yet be carried out on today's NISQ devices, and implementing arbitrary Q-qubit algorithms with arbitrary circuit length on an Q-qubit device is still infeasible. 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 calculate the correct measurement value from a set of noisy measurements [43].

3.4 Variational quantum eigensolver

While there exist quantum algorithms that allow measuring the exact ground state eigenvalue of a Hamiltonian when a sufficiently close approximate ground state is prepared on the quantum computer, such as quantum phase estimation (QPE) [44], [45], implementing these algorithms on NISQ devices is infeasible due to a small number of qubits and short coherence times. Thus, hybrid quantum computing algorithms have been 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) [5]. The algorithm works as follows:

1. Define a unitary ansatz $\hat{U}(\vec{\theta})$ parametrized by parameters $\vec{\theta} = \begin{bmatrix}\theta_1 & \dots & \theta_n\end{bmatrix}^T$ acting on Q qubits, which can be implemented at reasonable cost on a quantum computer. Possible forms of $\hat{U}(\vec{\theta})$ are discussed in the next section.
2. Set $\vec{\theta} = \vec{\theta}_0$, where $\vec{\theta}_0$ is a starting guess.
3. Prepare the state $|\Psi(\vec{\theta})\rangle = \hat{U}(\vec{\theta})\,|0^Q\rangle$ on a quantum computer. Observe that it it always possible to write $\hat{U}'(\vec{\theta})\,|\Phi_0\rangle = \hat{U}'(\vec{\theta})\hat{V}\,|0^Q\rangle = \hat{U}(\vec{\theta})\,|0^Q\rangle$ for arbitrary $|\Phi_0\rangle = \hat{V}\,|0^Q\rangle$, justifying why the system can always be in the state $|0^Q\rangle$ initially.
4. Estimate the state's expectation energy $E(\vec{\theta}) = \langle\Psi(\vec{\theta})|\,\hat{H}\,|\Psi(\vec{\theta})\rangle$. This is done using *Hamiltonian averaging* on the quantum computer, described in more detail in sec. 3.8.
5. Update the parameter $\vec{\theta}$ *classically* to get a new state $|\Psi(\vec{\theta})\rangle$ based on a classical optimization procedure which aims at minimizing $E(\vec{\theta})$.
6. Repeat 3, 4, 5 until some convergence criteria is fulfilled.

The effectiveness of the VQE method depends on whether the ansatz $|\Psi(\vec{\theta})\rangle$ can approximate the ground state well enough, and whether the classical optimization manages to find a good set of parameters. This optimization is generally NP-hard [46], 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.

Many classical optimization procedures require partial derivatives $\frac{\partial}{\partial\theta_i}E(\vec{\theta})$ for each parameter θ_i, $i = 1, \dots, n$. On a quantum computer, these need to be measured. Numerical differentiation based on finite-difference schemes tends to give insufficient results on noisy devices and can be costly. However, it is often possible to explicitly construct circuits and operators to measure partial derivatives, avoiding the need to use finite-difference schemes. This is the case for the UCC-based ansätze described in the next section, where the *parameter shift* rules apply [47]. Those allow for cheap measurement of partial derivatives, as the expectation value of the partial derivative can be found by measuring the original expectation value with shifted parameters, e.g. $\frac{\partial}{\partial\theta_i}E(\vec{\theta}) \propto \left[E(\vec{\theta} + \vec{\theta}_i') - E(\vec{\theta} - \vec{\theta}_i')\right]$, where $\vec{\theta}_i'$ is a parameter shift. While this resembles numerical differentiation, the parameter shift $\vec{\theta}_i'$ is not necessarily small and the formula is not an approximation, but exact.

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 the unitary coupled cluster wave function (eq. (1.81)), 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 ansatz reads (incorporating the prefactor $1/4$ into the parameters θ_{ij}^{ab})

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

where the Trotter formula (3.4) with $n = 1$ was used and the last equality is just a shorthand notation with $\hat{\sigma}_\mu = \hat{\tau}_\mu - \hat{\tau}_\mu^\dagger$. Even though $\hat{\sigma}_\mu$ is a linear combination of excitation and de-excitation operators, we will still refer to it as excitation. Trotterization with $n = 1$ might seem to be a crude approximation - however, as has been analyzed in Ref. [6], increasing the number of Trotter steps n leads only to small energy improvements, although the optimal parameter $\vec{\theta}$ changes as n is varied. As the individual exponential terms do not commute ($[\hat{\sigma}_\mu, \hat{\sigma}_\nu] \neq 0$), the ordering has a big impact on the resulting wave function [48], and sometimes going to $n \geq 2$ with independent parameters is necessary to achieve chemical accuracy (an energy difference of less than 1.6 mHartree=1 kcal/mol from the FCI energy). 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 [49]. Furthermore, it is always possible to parameterize the exact wave function using an infinite number of exponentials of singles and doubles unitary excitations of the form $\exp\left(\theta_\mu \hat{\sigma}_\mu\right)$ [49], which means that increasing the number of parameters systematically improves the wave function:

$$|\Psi^{\text{FCI}}\rangle = \prod_{k}^{\infty} \prod_{\mu \in S,D} \exp\left(\theta_\mu^{(k)} \hat{\sigma}_\mu\right) |\Phi_0\rangle \tag{3.7}$$

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

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

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

While the motivation of the ansatz is based on the Nooijen conjecture [50] of the exactness of the non-unitary CCGSD, which was later disproved [51], it serves as a starting point for further methods.

The *UpCCGSD* ansatz, which acts on an RHF reference determinant, is a restriction on the UCCGSD ansatz, where the $\hat{T}_2$ amplitudes are restricted to only act on pairs of electrons, e.g. both the α and the β electron are excited from an occupied MO to a virtual MO. The ansatz then reads

$$\begin{aligned} e^{\hat{\sigma}(\vec{\theta})}_{\text{UpCCGSD}} &= \exp\left(\sum_{pq}\theta^q_p\left(a^\dagger_p a_q - a^\dagger_q a_p\right) + \sum_{pq}\theta^{q_\alpha q_\beta}_{p_\alpha p_\beta}\left(a^\dagger_{q_\alpha}a^\dagger_{q_\beta}a_{p_\beta}a_{p_\alpha} - a^\dagger_{p_\alpha}a^\dagger_{p_\beta}a_{q_\beta}a_{q_\alpha}\right)\right) \\ &\approx \prod_{pq}\exp\left(\theta^q_p\left(a^\dagger_p a_q - a^\dagger_q a_p\right)\right)\prod_{pq}\exp\left(\theta^{q_\alpha q_\beta}_{p_\alpha p_\beta}\left(a^\dagger_{q_\alpha}a^\dagger_{q_\beta}a_{p_\beta}a_{p_\alpha} - a^\dagger_{p_\alpha}a^\dagger_{p_\beta}a_{q_\beta}a_{q_\alpha}\right)\right) \end{aligned} \tag{3.9}$$

The clear advantage is the scaling of the number of variables as $O(M^2)$, but 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 [52], where the generalized excitations are repeated k times:

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

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 for chemistry [53].

Starting guess $\vec{\theta}_0$

There are several ways to choose the initial guess vector $\vec{\theta}_0$. In this thesis, it is chosen randomly, though it should be noted that smart guesses for unitary coupled cluster methods exist, such as using MP2 amplitudes [54], which accelerate convergence.

3.5.2 Alternative ansätze

The UCC method and the methods 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 to produce a good approximate ground state in terms of gate count. The alternative are the so-called *hardware efficient ansätze (HEA)*, which try to reduce the resource use of the quantum-hardware. They are called *hardware heuristic ansätze (HHA)* when they are additionally tailored to the naturally available unitary operations on the quantum computer [45]. The trotterized UCCSD method and the methods derived from it lie somewhere between HEA and PMA. A prominent example of another algorithm between HEA and PMA is the adaptive ADAPT-VQE ansatz [55], where the wave function 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. [6], [56], are usually based on combinations of one-qubit rotations $U_{\text{rot}}(\vec{\theta})$ and entanglement unitaries U_{ENT}

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

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 Hamiltonian and depends on the architecture of the quantum computer itself, while the one-qubit rotations $U_{\text{rot}}(\vec{\theta})$ 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. In this section, it is described 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

$$\hat{H} = \sum_{i_1, i_2, \ldots i_Q} c_{i_1, i_2, \ldots i_Q} \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q}, \tag{3.12}$$

where each $i_j \in \{0, X, Y, Z\}$. This corresponds in theory to 4^M terms, but only $O(M^4)$ terms are nonzero, as each single string of creation and annihilation operators corresponds to a single string of Pauli matrices, which we show below.

3.6.1 The Jordan-Wigner mapping

The Jordan-Wigner mapping [57] is a particularly intuitive way to map creation and annihilation operators 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

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

However, this is not a valid representation. The anticommutator is not zero when as needed, the commutator is: $\left[\sigma_j^+, \sigma_i^-\right]_+ \neq [\sigma_j^+, \sigma_i^-] = 0$, $i \neq j$,. The following mapping resolves this problem:

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

The Pauli σ^Z matrix is applied to all qubits preceding qubit j, which applies a phase. This way, a correspondence between the second quantized basis states

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

and the quantum computer basis states

$$|k_1 k_2 \ldots k_M\rangle \tag{3.16}$$

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 [58].

3.6.2 The parity mapping

The parity mapping [58] stores the parity of each state. Going from the Jordan-Wigner mapping to the parity mapping, the following transformation is applied

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

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 acting with 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 off 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 whether j is even or odd, and is given in Ref. [58]. It should be noted that the original Bravyi-Kitaev mapping only works when $M = 2^n$, $n \in \mathbb{N}$ [45], however, there are very similar tree-based 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 [59]. 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, which is called *qubit tapering*. 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 represent M molecular orbitals on $M-1$ 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. [60], [61]. It has been implemented in experiment [62] and works for all three types of mappings discussed.

Qubit tapering has several advantages. First of all, it guarantees 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 gate count (on real quantum computers), and it allows to simulate systems with M molecular orbitals on quantum computers with $Q < M$ qubits when at least $M - Q$ qubits can be tapered off.

Tapering off two qubits in the parity mapping

In the parity mapping, qubit tapering can be understood intuitively. When spin MOs are used, one can represent the Slater determinant as a bit-string by first writing all alpha and then all beta electrons

$$|\Phi^{\mathrm{SD}}\rangle = |k_{\alpha_1}, \ldots, k_{\alpha_{M_\alpha}}, k_{\beta_1}, \ldots, k_{\beta_{M_\beta}}\rangle \tag{3.18}$$

and using that the Hamiltonian conserves the number of electrons with α spin $\sum_i^{M_\alpha} k_{\alpha_i} = N_\alpha$, 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 configurations and consequently do not need to be considered, as the action of a Pauli matrix acting on qubit $k_{\alpha_{M_\alpha}}$ can be replaced by a number.

General qubit tapering

For a Hamiltonian in the form of eq. (3.12), we are after a unitary transformation $\hat{U}$ of the Hamiltonian $\hat{H}' = \hat{U}\hat{H}\hat{U}^\dagger$ such that each individual Pauli string of the transformed Hamiltonian acts with only σ^Z or σ^0 on qubits $i_1, i_2, \ldots, i_T$ ($T < Q$), where T is the number of qubits that can be tapered off. The Pauli matrices acting on qubits $i_1, i_2, \ldots, i_T$ can then be replaced by their eigenvalues ± 1, because we know that the exact eigenstate of the Hamiltonian will have each qubit $i_1, i_2, \ldots, i_T$ either in state $|0\rangle$ or $|1\rangle$. We can then consider each of the 2^T cases individually. This way, we do not simulate qubits at positions $i_1, i_2, \ldots, i_T$, and they can be tapered off. There exist explicit algorithms to find the unitary operator $\hat{U}$ that accomplishes this, which are described in Refs. [61], [63]. Ref. [63] connects qubit tapering of molecular Hamiltonians with molecular point group symmetries by expressing the symmetries in second quantization (and thus as strings of Pauli operators when mapped to a quantum computer), which removes the need to iterate over all 2^T cases.

3.7 Implementation of Pauli matrix exponentials

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

$$W = \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q} \quad ; \quad U = e^{-i\tau W} \quad ; \quad \tau \in \mathbb{R} \tag{3.20}$$

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. To

exemplify, we have, using $W = \sigma_1^X \otimes \sigma_2^Z \otimes \sigma_3^Y$

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

where it was used that $AC \otimes BD = (A \otimes B)(C \otimes D)$, and 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 $e^{(-i\tau \sigma_1^Z \otimes \sigma_2^Z \otimes \sigma_3^Z)}$ 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 Q qubits), one can see by direct calculation that the inclusion of a single *ancilla qubit*, e.g. an extra qubit to help with the calculation, can turn this into a local operation [40]. This gives the quantum circuit shown in figure 3.2.

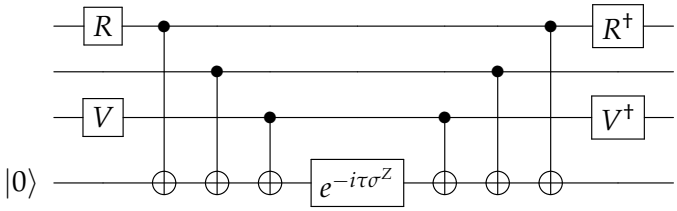


Figure 3.2: Efficient quantum circuit for simulating $e^{-i\tau \sigma_1^X \otimes \sigma_2^Z \otimes \sigma_3^Y}$ with one ancilla qubit.

For the trotterized UCCSD ansatz $\hat{U}(\vec{\theta})$ (eq. (3.6)) in the Jordan-Wigner mapping, a decomposition into

$$\hat{U}(\vec{\theta}) = \prod_i e^{-i\tau W_i} \tag{3.22}$$

where each W_i is a single Pauli string, can be found in Ref. [6]. This type of decomposition is applicable to other UCCSD-based ansätze and other mappings than the Jordan-Wigner mapping, too.

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

$$\begin{aligned} \langle \Psi(\vec{\theta}) | \hat{H} | \Psi(\vec{\theta}) \rangle &= \langle \Psi(\vec{\theta}) | \sum_{i_1, i_2, \ldots i_Q} c_{i_1, i_2, \ldots i_Q} \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q} | \Psi(\vec{\theta}) \rangle \\ &= \sum_{i_1, i_2, \ldots i_Q} c_{i_1, i_2, \ldots i_Q} \langle \Psi(\vec{\theta}) | \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q} | \Psi(\vec{\theta}) \rangle \\ &= \sum_{i_1, i_2, \ldots i_Q} c_{i_1, i_2, \ldots i_Q} \langle \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q} \rangle \\ &= \langle \hat{H} \rangle \end{aligned} \tag{3.23}$$

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 Pauli matrix expectation values has been labeled *Hamiltonian averaging* [64]. On simulated quantum computers, these expectation values can be calculated exactly, on physical quantum computers, they need to be sampled. We cannot sample $\langle \hat{H} \rangle$ directly as it would require measuring all qubits simultaneously, so each expectation value $\langle \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \ldots \tilde{\sigma}_Q^{i_Q} \rangle$ needs to be sampled individually. In Ref. [64], 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

$$\mathrm{Var}(\hat{H}) = \sum_{i_1,i_2,\ldots i_Q} c^2_{i_1,i_2,\ldots i_Q} \mathrm{Var}(\tilde{\sigma}_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q}) \leq \sum_{i_1,i_2,\ldots i_Q} c^2_{i_1,i_2,\ldots i_Q} \tag{3.24}$$

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

$$\mathrm{Var}(\langle \hat{H} \rangle) = \frac{\mathrm{Var}(\hat{H})}{n_m} \tag{3.25}$$

where n_m is the number of repeated measurements of all Pauli strings. The total number of measurements N_m in order to obtain a variance $\mathrm{Var}(\langle \hat{H} \rangle) \leq \varepsilon^2$ becomes [65]

$$N_m = n_{PS} \frac{\sum_{i_1,i_2,\ldots i_Q} c^2_{i_1,i_2,\ldots i_Q}}{\varepsilon^2} \tag{3.26}$$

where n_{PS} is the number of Pauli strings in the Hamiltonian and scales as $O(M^4)$. N_m 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. It does however make the mutual covariance non-zero [65]. In addition, the number of necessary measurements can be reduced by clever decompositions of the Hamiltonian, or by using statistical methods, which we discuss briefly in the next section.

3.9 Scaling of methods

Having discussed how different ansätze can be expressed in terms of operators acting on qubits, how they can be decomposed into basic gates, and how to measure expectation values, a final point to consider is the scaling of these methods as a function of the number of electrons N and the basis size M. Generally speaking, the exact number of gates necessary to implement a circuit is strongly hardware-dependent and depends on the connectivity between qubits, natively available gates, number of qubits, and similar. 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 count, 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 refers to time complexity, while the number of qubits and the gate count refer to hardware requirements. Only two-qubit gates are considered in the gate count, being the most expensive gates.

For a trotterized unitary coupled cluster ansatz with k Trotter steps, the number of parameters 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 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, $f = O(\log(M))$. It should however be noted that the prefactor f can be eliminated in the Jordan-Wigner and parity mappings [66] by a clever decomposition of the strings of creation/annihilation operators, which is based on a double factorization of the UCC-amplitudes and the Hamiltonian, writing the rank 4 tensors as products of lower-rank tensors (*tensor hypercontraction*) [67]. We thus neglect the f dependency from now on, as

it possible to make $f = O(1)$, then the number of gates scales as $O(kM^4)$. k-UpGCCSD has a gate count of $O(kM^2)$. k-UPGCCSD has a circuit depth of $O(kM)$, while UCCSD has a circuit depth of $O(kM^3)$ [52]. k-UPGCCSD is the most promising ansatz we are aware of and gives excellent results, with the circuit depth scaling linearly in the basis size [53]. 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} corresponds to a number of CNOT gates scaling linearly in the number of qubits, and has a small circuit depth of $O(D)$ when those CNOT gates are applied in parallel.

Hamiltonian averaging requires a vast number of measurements, and using a formula similar to eq. (3.26) to estimate the variance, the number of necessary measurements to achieve chemical accuracy scales as $O(M^4/\epsilon^2)$, with the $O(M^4)$ term stemming from the number of terms in the Hamiltonian, and the formula additionally taking into account an optimal distribution of the measurements [53]. Naively, this also requires $O(M^4/\epsilon^2)$ state preparations. This number of measurements and state preparations can be prohibitively large. Based on the same tensor hypercontraction as the one presented in Ref. [66], the number of state preparations to measure all $O(M^4)$ terms in the Hamiltonian can be reduced to $O(M)$ [68], adding an overhead of $O(M)$ to the state preparation. This is the strongest reduction we are aware of [53]. The scaling in $1/\epsilon^2$ can also be reduced. The likelihood function to measure $|0\rangle$ for a single qubit as a function of the expectation value P of a Pauli string $f(0|P)$ is linear, but following the approach described in Ref. [69], it is possible to apply gates to turn the likelihood function into a Chebyshev polynomial. Using Bayesian interference, it is then possible to obtain a distribution for the expectation value which has a much smaller standard deviation, at the cost of a higher gate count and circuit depth. This way, it is possible to find a scaling lying between $1/\epsilon$ (which is the best possible) and $1/\epsilon^2$, depending on the longest acceptable gate count and circuit depth, which depend 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 [70]–[72] which all considered similar approaches. We follow the approach in Ref. [72]. Let $|0^Q\rangle$ be the initial state, and $|\Psi_i\rangle = \hat{U}_i|0^Q\rangle$, $|\Psi_j\rangle = \hat{U}_j|0^Q\rangle$ for some unitary operators $\hat{U}_i$, $\hat{U}_j$. The first step is to prepare the state

$$|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \otimes \hat{U}_i|0^Q\rangle + |1\rangle \otimes \hat{U}_j|0^Q\rangle\right) \tag{3.27}$$

which can be done with the circuit in figure 3.3.

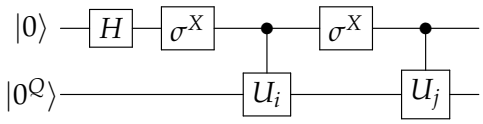


Figure 3.3: Quantum circuit to prepare the state $|\Omega\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle \otimes \hat{U}_i|0^Q\rangle + |1\rangle \otimes \hat{U}_j|0^Q\rangle\right)$. The first σ^X gate can be omitted, but was added to signify a basis change operation.

It should be noted that the controlled unitary operations, which are costly, can be avoided. The state $|\Omega\rangle$ can be obtained without controlled unitaries when a total of $2Q+1$ qubits is available, e.g. using $Q+1$ ancilla qubits [70]. Expectation values can then be

calculated as

$$\langle\Psi_i|\,\hat{O}\,|\Psi_j\rangle = 2\,\langle\Omega|\,\sigma^+\otimes\hat{O}\,|\Omega\rangle = \langle\Omega|\,\sigma^X\otimes\hat{O}\,|\Omega\rangle + i\,\langle\Omega|\,\sigma^Y\otimes\hat{O}\,|\Omega\rangle \tag{3.28}$$

where $\langle\Omega|\,\sigma^X\otimes\hat{O}\,|\Omega\rangle$ and $\langle\Omega|\,\sigma^Y\otimes\hat{O}\,|\Omega\rangle$ are the real and the imaginary part of the matrix element $\langle\Psi_i|\,\hat{O}\,|\Psi_j\rangle$, respectively.

In the case that only the overlap $\langle\Psi_i|\Psi_j\rangle$ is needed, a method to calculate it without the need of controlled unitaries or ancilla qubits is to simply measure

$$\langle\Psi_i|\Psi_j\rangle = \langle 0^Q|\,\hat{U}_i^\dagger\hat{U}_j\,|0^Q\rangle\,. \tag{3.29}$$

3.11 Sampling error in the context of EVC

An extra complication on quantum computers is that there is a sampling error to take into account when performing eigenvector continuation. Assume that we have a set of sample wave functions $|\Psi(\vec{\alpha}_i)\rangle$, $i = 1,\ldots,L$. The EVC matrices then read, for some target parameter $\vec{\alpha}_\odot$

$$\begin{aligned} S_{ij}^{\text{EVC}} &= \langle\Psi(\vec{\alpha}_i)|\Psi(\vec{\alpha}_j)\rangle \\ H_{ij}^{\text{EVC}} &= \langle\Psi(\vec{\alpha}_i)|\hat{H}(\vec{\alpha}_\odot)|\Psi(\vec{\alpha}_j)\rangle\,. \end{aligned} \tag{3.30}$$

As matrix elements need to be sampled on quantum computers, the matrices $\mathbf{H}^{\text{EVC}}$, $\mathbf{S}^{\text{EVC}}$ are random matrices, so the eigenvectors and eigenvalues are random variables and it is necessary to do a statistical analysis of the problem. For the remainder of this section, we drop the superscript $^{\text{EVC}}$. In Ref. [70], a method is described to quantify the uncertainty in the ground state energy. Let $\tilde{h}_{ij}, \tilde{s}_{ij}$ be the sample means. Let $\tilde{\sigma}^2_{S_{ij}}, \tilde{\sigma}^2_{H_{ij}}$ be the sample variances. We then treat H_{ij}, S_{ij} as random variables, which are approximated as normal distributions

$$\begin{aligned} \tilde{H}_{ij} &\sim \mathcal{N}\left(\tilde{h}_{ij}, \tilde{\sigma}_{H_{ij}}/\sqrt{N_{H_{ij}}}\right) \\ \tilde{S}_{ij} &\sim \mathcal{N}\left(\tilde{s}_{ij}, \tilde{\sigma}_{S_{ij}}/\sqrt{N_{S_{ij}}}\right) \end{aligned} \tag{3.31}$$

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, the Bootstrapping method [73] can be used 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 (3.31) to obtain a guess of the eigenvalue $E_{bootstrap}$ and the standard deviation in the energy σ_E. 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. As long as $\sigma_E > x$ for some threshold x, it is necessary to increase the number of measurements to reduce that uncertainty. The aim is to measure the element S_{ij}, H_{ij} where increasing the number of measurements decreases the uncertainty σ_E the most. Following the algorithm in Ref. [70], in order to determine which element that is, one iterates over all matrix elements $S_{k,l}$, $H_{k,l}$ individually and assumes that they follow the distribution

$$H_{k,l} \sim \mathcal{N}\left(\tilde{h}_{k,l}, \tilde{\sigma}_{H_{k,l}}/\sqrt{N_{H_{ij}} + N^+}\right) \tag{3.32}$$

and similarly for $S_{k,l}$, where N^+ is the number of extra measurements to be performed. All other elements S_{ij}, H_{ij} are assumed to follow the distribution (3.31). For each k,l and for H and S individually, this gives a new estimate for the standard deviation σ_E, which is denoted as $\tilde{\sigma}_E^{H/S}(k,l)$. After having obtained these estimates, the element to be actually measured on the quantum computer is the one with the smallest $\tilde{\sigma}_E^{H/S}(k,l)$. Calculating $\tilde{\sigma}_E^{H/S}(k,l)$ does

not require quantum measurements. This way, uncertainty is decreased without increasing accuracy, which requires quantum measurements. The hope is that measurements are performed where increasing accuracy decreases uncertainty in the eigenvalue σ_E the most. This can be done iteratively until σ_E is less than x.

3.12 Finding excited states on quantum computers

There are many ways to find excited states on quantum computers. We follow the approach in Ref. [74]. They suggested a method to find excited states on quantum computers by introducing a modified Hamilton operator

$$\hat{H}_k = \hat{H} + \sum_{i=0}^{k-1} \beta_i \left|\Psi_i\right\rangle\!\left\langle\Psi_i\right| \tag{3.33}$$

where $|\Psi_i\rangle$ is the i^{th} excited state (with $i = 0$ being the ground state) 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 eigenstate of the Hamiltonian cannot be larger than $2\left\|\hat{H}\right\|$ for the operator norm. 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 representing the k^{th} excited state $|\Psi_k(\vec{\theta})\rangle = \hat{U}_k(\vec{\theta})\,|0^Q\rangle$, we minimize the modified expectation energy

$$\langle\Psi_k(\vec{\theta})|\hat{H}_k|\Psi_k(\vec{\theta})\rangle = \langle\Psi_k(\vec{\theta})|\hat{H}|\Psi_k(\vec{\theta})\rangle + \sum_{i=0}^{k-1} \beta_i |\langle\Psi_k(\vec{\theta})|\Psi_i\rangle|^2 \tag{3.34}$$

with respect to $\vec{\theta}$ using the VQE method. It is important that the ansatz $\hat{U}_k(\vec{\theta})$ is sufficiently flexible to represent the k^{th} excited state, which is the case at least for the first excited state for GCCSD and k-UpGCCSD as well as repeated applications of UCCSD [52].

3.13 Linear combination of unitaries

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

$$|\Psi\rangle = \hat{M}\,|\Phi_0\rangle = \sum_{i=0}^{N-1} c_i \hat{U}_i\,|\Phi_0\rangle \tag{3.35}$$

where $\hat{M}$ is a matrix which is not necessarily unitary and $|\Phi_0\rangle$ is an initial state. In order to do so, one can use the *linear combination of unitaries (LCU)* algorithm [75], [76]. Observe that decomposing any matrix into a 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, rescale such that $\sum_{i=0}^{N-1} c_i^2 = 1$. Then we define

$$\begin{aligned} n &= \lceil \log_2(N) \rceil \\ c &= \sum_{i=0}^{N-1} c_i \\ \hat{V}\,|0^n\rangle &= \sum_{i=0}^{N-1} \sqrt{\frac{c_i}{c}}\,|i_{(2)}\rangle \quad \text{where } \hat{V}^\dagger\hat{V} = I \\ \hat{U} &= \sum_{i=0}^{N-1} |i_{(2)}\rangle\!\langle i_{(2)}|\,\hat{U}_i. \end{aligned} \tag{3.36}$$

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

$$\left(\hat{V}^\dagger \otimes \mathbb{1}\right)\hat{U}\left(\hat{V} \otimes \mathbb{1}\right)|0^n\rangle \otimes |\Phi_0\rangle = \frac{1}{c}|0^n\rangle \otimes \hat{M}|\Phi_0\rangle + |\Phi_0^\perp\rangle \tag{3.37}$$

where $|\Phi_0^\perp\rangle$ 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 $p_s = \frac{1}{c^2}\langle\Phi|\hat{M}^\dagger\hat{M}|\Phi_0\rangle$.

3.14 Errors on quantum computers

There are many possible error sources that can lead to wrong energies when using a quantum computer. For a given basis set, the following errors lead to deviations from exact diagonalization of the Hamiltonian:

1. The ansatz $\hat{U}(\vec{\theta})|\Phi_0\rangle$ not being able to represent the exact FCI wave function.
2. Optimization errors leading to $\langle\Psi(\vec{\theta}^{\text{opt.}})|\hat{H}|\Psi(\vec{\theta}^{\text{opt.}})\rangle < \langle\Psi(\vec{\theta}^*)|\hat{H}|\Psi(\vec{\theta}^*)\rangle$, where $\vec{\theta}^{\text{opt.}}$ is the optimal set of parameters $\vec{\theta}$, while $\vec{\theta}^*$ is the obtained $\vec{\theta}$ after some optimization procedure. This may happen when $\langle\Psi(\vec{\theta}^*)|\hat{H}|\Psi(\vec{\theta}^*)\rangle$ is a local minimum.
3. Sampling errors when sampling $\langle\hat{H}\rangle$ and its derivatives with respect to the parameters θ_i. This is relevant for optimization as well as the final energy calculation.
4. Physical noise due to faulty gates, decoherence, and similar.

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, such as rotation operators are periodic functions. Thus, there are infinitely many global minimum points, and there are no physically motivated starting guesses $\vec{\theta}_0$. The third problem is also relevant in quantum Monte Carlo methods [77], but unusual for standard wave-function based methods. 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.

Chapter 4

Methods

The aim of this chapter is to present how eigenvector continuation can be used in chemistry, as well as introduce the AMP-CCEVC method which we developed. Eigenvector continuation requires analyticity of the Hamiltonian matrix as a function of a parameter $\vec{\alpha}$, so we focus on ways to make the Hamiltonian matrix analytic as function of the nuclear positions. In a second-quantized representation of the Hamiltonian, the analytic properties of the Hamiltonian matrix are connected to the expansion coefficients of the molecular orbitals in terms of atomic orbitals, hence we consider in some detail how unitary operations can be used to make changes in the molecular orbitals both analytic and small.

First, we introduce the AMP-CCEVC method, where the CC amplitudes, not the wave function, are written as linear combinations of sample amplitudes. Then, we consider the general requirements that the MOs as a function of geometry have to fulfill in order for eigenvector continuation to work to obtain potential energy surfaces. After that, we consider several choices of MOs that fulfill those requirements by using symmetric orthonormalization and solving the orthogonal Procrustes problem, as well as simply using canonical or natural orbitals. We then introduce different ways to use eigenvector continuation to obtain a potential energy surface based on either multi-reference or single-reference methods, for which different choices of MOs should be employed. Finally, we consider how eigenvector continuation can be implemented on a quantum computer, describing possible advantages and disadvantages.

4.1 AMP-CCEVC

In the WF-CCEVC approach (sec. 2.3.1, p. 27), the resulting wave function is a linear combination of coupled cluster wave functions. A related approach is to write the cluster operator $\hat{T}(\vec{\alpha}_\odot)$ as a linear combination of sample cluster operators. Thus, the AMP-CCEVC wave function is a coupled cluster wave function,

$$|\Psi^{\text{AMP-CCEVC}}(\vec{\alpha}_\odot)\rangle = e^{\hat{T}^*(\vec{\alpha}_\odot)}\,|\Phi^{\text{SD}}\rangle = \exp\left\{\sum_{m=1}^{L} c_m \hat{T}(\vec{\alpha}_m)\right\}|\Phi^{\text{SD}}\rangle\,. \tag{4.1}$$

Technically speaking, no eigenvector continuation is performed, but a dimensionality reduction, as there are only L variables to be determined. The parameters $\mathbf{c} = \begin{bmatrix} c_1 & \dots & c_L \end{bmatrix}^T$ and the energy can be determined in the standard way, which is by solving a set of amplitude equations (1.63) and the energy equation (1.62)

$$\langle \Phi^{\text{SD}}|\, \hat{T}^\dagger(\vec{\alpha}_m)\bar{H}(\mathbf{c})\,|\Phi^{\text{SD}}\rangle = 0 \quad \text{for all } m \tag{4.2}$$

$$\langle \Phi^{\text{SD}}|\, \bar{H}(\mathbf{c})\,|\Phi^{\text{SD}}\rangle = E^*_{CC}. \tag{4.3}$$

This approach is justified because we made no assumption about the states $\langle\mu|$ in setting up the amplitude equations (1.63), other than that they are orthogonal to the reference determinant $|\Phi^{\mathrm{SD}}\rangle$. We can hence project with either excited determinants $\langle\mu|$ or linear combinations of excited determinants. Furthermore, the number of states to project with should be equal to the number of parameters, which is the case here. The implementation of AMP-CCEVC is discussed in greater detail in the numerical methods section, sec. 5.4, p. 64.

4.1.1 Parameter reduction

In AMP-CCEVC, there are only L parameters to obtain, compared to CCSD, where there are $O(N^2M^2)$. However, there is no direct numerical advantage, as one still needs to calculate the individual projection errors $\langle\mu|\,\bar{H}(\mathbf{c})\,|\Phi^{\mathrm{SD}}\rangle$ for $\mu \in S, D, \ldots$:

$$\mathrm{e}(c_m) = \langle\Phi^{\mathrm{SD}}|\,\hat{T}^\dagger(\vec{\alpha}_m)\bar{H}\,|\Phi^{\mathrm{SD}}\rangle = \sum_\mu t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle\,, \tag{4.4}$$

where we simply wrote $\bar{H}$ for $\bar{H}(\mathbf{c})$. The projection error $\mathrm{e}(c_m)$ is a linear combination of the projection errors of each excitation $\langle\mu|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle$ and thus not cheaper to calculate. This is also the case for the Jacobian. Even when using a Quasi-Newton approach (eq. (1.71)), one still needs to consider the whole diagonal part of the CCSD-Jacobian when calculating

$$J_{mn}(\mathbf{c}) = \frac{\partial}{\partial c_n}\mathrm{e}(c_m). \tag{4.5}$$

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

$$\mathrm{e}(c_m) = \langle\Phi^{\mathrm{SD}}|\,\hat{T}^\dagger(\vec{\alpha}_m)\bar{H}\,|\Phi^{\mathrm{SD}}\rangle = \sum_{\text{all }\mu} t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle \approx \tilde{\mathrm{e}}(c_m) = \sum_{\text{some }\mu} t_\mu(\vec{\alpha}_m)\,\langle\mu|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle\,. \tag{4.6}$$

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

$$\mathrm{e}(c_m) = \sum_{ia} t_i^a(\vec{\alpha}_m)\,\langle\Phi_i^a|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle + \frac{1}{4}\sum_{ijab} t_{ij}^{ab}(\vec{\alpha}_m)\,\langle\Phi_{ij}^{ab}|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle\,. \tag{4.7}$$

In order to cut down the number of floating point operations, one needs to find a way to avoid summing over all virtual indices, but only a subset thereof, and possibly do the same for occupied indices. The most computationally expensive equations in CCSD theory in a direct product decomposition [23] is the calculation of the $\mathcal{W}_{abef}$ intermediate and its contribution to the projection error $\langle\Phi_{ij}^{ab}|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle$, which both scale as $O(M_v^4M_o^2)$:

$$\mathcal{W}_{abef} = \frac{1}{4}\sum_{mn}\tau_{mn}^{ab}\,\langle mn\|ef\rangle + \ldots \tag{4.8}$$

$$\langle\Phi_{ij}^{ab}|\,\bar{H}\,|\Phi^{\mathrm{SD}}\rangle = \frac{1}{2}\sum_{ef}\tau_{ij}^{ef}\mathcal{W}_{abef} + \ldots \tag{4.9}$$

where we only wrote the terms scaling as $O(M_v^4M_o^2)$, and where

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_j^b t_i^a. \tag{4.10}$$

When omitting some virtual indices, we do not need to calculate every $\langle \Phi_{ij}^{ab} | \bar{H} | \Phi^{\mathrm{SD}} \rangle$ and the whole $\mathcal{W}_{abef}$ tensor is not required, as we are only interested in some virtual indices 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 $\frac{1}{p^2}$ for calculating the intermediate $\mathcal{W}_{abef}$ and the projection error $\langle \Phi_{ij}^{ab} | \bar{H} | \Phi^{\mathrm{SD}} \rangle$. Similar considerations apply to other intermediates when not including all virtual or occupied orbitals.The parameter-reduced approach will yield an approximate $\hat{T}^*(\vec{\alpha}_\odot)$ at reduced cost. In this thesis, we use $\Theta_i = \sum_{abj} \left| t_{ij}^{ab} \right|^2$, $\Theta_a = \sum_{bij} \left| t_{ij}^{ab} \right|^2$ as a selection criterion and only include the fraction p of virtual orbitals with the largest Θ_a and those occupied orbitals with large Θ_i.

4.1.2 Justification of the parameter reduction approach

Assume that the AMP-CCEVC approach is exact, e.g. there is a parameter $\mathbf{c}$ that solves the CC amplitude equations

$$\langle \mu | \, e^{-\hat{T}^*(\vec{\alpha}_\odot)} \hat{H} e^{\hat{T}^*(\vec{\alpha}_\odot)} \, | \Phi^{\mathrm{SD}} \rangle = 0 \quad \text{for all } \mu \in S, D, \ldots \tag{4.11}$$

$$\hat{T}^*(\vec{\alpha}_\odot) = \sum_{m=1}^{L} c_m \hat{T}(\vec{\alpha}_m). \tag{4.12}$$

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

$$\sum_{\text{some } \mu} t_\mu \, \langle \mu | \, e^{-\hat{T}^*(\vec{\alpha}_\odot)} \hat{H} e^{\hat{T}^*(\vec{\alpha}_\odot)} \, | \Phi^{\mathrm{SD}} \rangle = 0 \tag{4.13}$$

for arbitrary t_μ. Thus, both eqs. (4.4) and (4.6) will be equal to zero and thus share the same solution. If the AMP-CCEVC approach does not give an exact CC wave function (eq. (4.11) does not hold), the parameter-reduced approach will have different roots than "full" AMP-CCEVC. However, if the AMP-CCEVC approach is approximately correct, the solutions of eq. (4.6) will be similar to those of eq. (4.4). When the AMP-CCEVC amplitudes are not a good approximation to the CC amplitudes, we expect the parameter reduction solution to be qualitatively different from both the AMP-CCEVC and the CC solution.

4.2 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 $e^{T(\vec{\alpha})} | \Phi^{\mathrm{SD}} \rangle$ where 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 reference for all values of $\vec{\alpha}$, such as for perturbations to the Hamiltonian of the form $\hat{H}'(\alpha) = \hat{H} + \alpha \hat{V}$, where $\hat{H}$ is the electronic Hamiltonian and $\hat{V}$ impacts the electronic state of the wave function, but not the positions of the nuclei. However, throughout this thesis, we want to perform eigenvector continuation to obtain potential energy surfaces. We consider hence a Hamiltonian $\hat{H}(\vec{R})$, where $\vec{R}$ stands for the position of the nuclei. A state $| \Psi(\vec{R}_i) \rangle$ is constructed from a finite basis set localized at $\vec{R} = \vec{R}_i$. It cannot be expressed with a finite basis set centered at $\vec{R} = \vec{R}_j$ unless $\vec{R}_j \approx \vec{R}_i$. When the basis set moves with the nuclei, not only the one-body and the two-body elements $h_{\mu\nu}(\vec{R})$ and $\langle \mu\nu \| \sigma\tau \rangle(\vec{R})$ change: the creation and annihilation operators $a_\mu^\dagger(\vec{R})$, $a_\mu(\vec{R})$ now refer to atomic orbitals at a different geometry. Thus, performing "naive" eigenvector continuation using the sample states $| \Psi(\vec{R}_i) \rangle$, which all lie in different regions

of space, is bound to fail, as we cannot expect that

$$|\Psi^{\mathrm{EVC}}(\vec{R}_\odot)\rangle = \sum_{m=1}^{L} c_m \, |\Psi(\vec{R}_m)\rangle \tag{4.14}$$

resembles the true ground state $|\Psi(\vec{R}_\odot)\rangle$: Because the sample states lie in different region of space, the overlap matrix elements $S^{\mathrm{EVC}}_{mn} = \langle\Psi(\vec{R}_m)|\Psi(\vec{R}_n)\rangle$, $m \neq n$, decay rapidly with increasing $\left\|\vec{R}_m - \vec{R}_n\right\|$, and the matrix elements of the Hamiltonian $H^{\mathrm{EVC}}_{mn} = \langle\Psi(\vec{R}_m)|\hat{H}(\vec{R}_\odot)|\Psi(\vec{R}_n)\rangle$ get large. We thus have to consider how one can still "reuse information" from a state constructed from a basis set centered at $\vec{R}_i$ at a new geometry $\vec{R}_j \neq \vec{R}_i$ by transporting the state from the geometry $\vec{R}_i$ to $\vec{R}_j$. In the next subsection, we discuss how to adapt the second quantization formalism in such a way that all geometry dependence can be removed from the creation and annihilation operators.

4.2.1 Geometry-independent representation of the Hamiltonian

In second quantization, the Hamiltonian with nuclei placed at positions $\vec{R}$ reads (eq. (1.23))

$$\hat{H}(\vec{R}) = \sum_{pq}^{M} h_{pq}(\vec{R}) a_p^\dagger(\vec{R}) a_q(\vec{R}) + \frac{1}{4}\sum_{pqrs}^{M} \langle pq\|rs\rangle(\vec{R}) a_p^\dagger(\vec{R}) a_q^\dagger(\vec{R}) a_s(\vec{R}) a_r(\vec{R}) + h_{\mathrm{nuc}}(\vec{R}). \tag{4.15}$$

The one-body integrals, the two-body integrals and the creation and annihilation operators depend on the nuclear positions. Following the discussion in Ref. [78], one might however consider the creation and annihilation operators to be geometry-independent: ON vectors are geometry-independent. The HF state, for example, can be represented as $|1_1 \ldots 1_N 0_{N+1} \ldots 0_M\rangle$ in ON vector representation, independently of geometry. This geometry independent representation can be used for all 2^M basis states of the Fock space. As long as the MOs are orthogonal at every geometry, one can thus consider the Fock spaces at different geometries as the same Fock space. Doing this, creation and annihilation operators can be considered as constant entities that do not depend on geometry. $a_p^\dagger$ acts on an ON vector by simply turning 0_p into 1_p (or yield 0) and possibly change the sign. Thus, the geometry dependence can be removed from the creation and annihilation operators. Within a given basis set, all geometric information is then contained in the expansion coefficients $C_{i\mu}(\vec{R})$ of the MOs in terms of AOs, and the one-body integrals $h_{\mu\nu}(\vec{R})$ and two-body integrals $\langle\mu\nu\|\sigma\tau\rangle(\vec{R})$ of the AOs. We hence write the Hamiltonian with geometry-independent creation and annihilation operators and express the one-body integrals and two-body integrals in terms of the (geometry-dependent) AO basis (cf. eq. (1.40), p. 10)

$$\begin{aligned}
\hat{H}(\vec{R}) &= \sum_{pq}^{M} h_{pq}(\vec{R}) a_p^\dagger a_q + \frac{1}{4}\sum_{pqrs}^{M} \langle pq\|rs\rangle(\vec{R}) a_p^\dagger a_q^\dagger a_s a_r + h_{\mathrm{nuc}}(\vec{R}) \\
&= \sum_{\mu\nu}^{M}\sum_{pq}^{M} C^*_{\mu p}(\vec{R}) C_{\nu q}(\vec{R}) h_{\mu\nu}(\vec{R}) a_p^\dagger a_q \\
&\quad + \frac{1}{4}\sum_{\mu\nu\sigma\tau}^{M}\sum_{pqrs}^{M} C^*_{\mu p}(\vec{R}) C^*_{\nu q}(\vec{R}) C_{\sigma r}(\vec{R}) C_{\tau s}(\vec{R}) \, \langle\mu\nu\|\sigma\tau\rangle(\vec{R}) a_p^\dagger a_q^\dagger a_s a_r + h_{\mathrm{nuc}}(\vec{R})
\end{aligned} \tag{4.16}$$

where the elements $h_{\mu\nu}(\vec{R})$, $\langle\mu\nu\|\sigma\tau\rangle(\vec{R})$ are functions of the molecular geometry and the choice of basis set, 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 this representation, all geometry independence is now in the Hamiltonian, with the basis (e.g. Slater determinants) being geometry-independent.

In order for eigenvector continuation and perturbation theory to work, the Hamiltonian matrix needs to be analytic as a function of the nuclear positions $\vec{R}$. This is precisely the case when the expansion coefficients $C_{\mu p}(\vec{R})$ are analytic (we have already established analyticity for $h_{\mu\nu}(\vec{R})$ and $\langle\mu\nu\|\sigma\tau\rangle(\vec{R})$ in section 1.2.1, p. 6). Then the eigenstates of the Hamiltonian can be chosen to be analytic. However, there are infinitely many ways to choose those expansion coefficients $C_{\mu p}(\vec{R})$. Any set of orthonormal MOs can be used, and it is not necessary that $C_{\mu p}(\vec{R})$ are analytic or even continuous. We have already considered a special choice of MOs that assure analyticity in section 1.2.1, p. 6, but this choice is not unique: The analyticity and orthogonality of the natural orbitals make them a viable choice too. We will thus discuss different ways to ascertain the analyticity of the expansion coefficients $C_{\mu p}(\vec{R})$.

4.3 Possible choices of MOs

In this section, we describe four possible choices of orbitals that make the MO coefficient matrix $\mathbf{C}(\vec{R})$ analytic. These methods have been described previously in the literature, but they have not been considered in the context of eigenvector continuation.

4.3.1 Symmetric orthonormalization of MOs

Assume we have solved the Hartree-Fock equations at a molecular geometry $\vec{R}$. This gives rise to an MO coefficient matrix $\mathbf{C}(\vec{R})$ and a Hartree-Fock ground state wave function $|\Phi^{\mathrm{HF}}(\vec{R})\rangle$. The question that arises is how to describe the same configuration at a different geometry $\vec{R}'$. At the new geometry, the AOs change, as the AOs are now centered at different positions. Using the same coefficient matrix $\mathbf{C}(\vec{R})$ at geometry $\vec{R}'$ leads to some complications, as the change in the basis functions will make the MOs non-orthonormal and the Slater determinant non-normalized, complicating the mathematics greatly, hence the need to re-orthonormalize the MOs by defining an orbital connection [79]. This procedure is not unique, and we simply use symmetric orthonormalization as described in Ref. [78], which has the advantage that the new MOs are as similar to the old ones as possible in a least-square sense [80]. Let $\mathbf{S}_{\vec{R}}(\vec{R}')$ be the overlap matrix between the MOs obtained at geometry $\vec{R}$ and expressed in terms of the coefficient matrix $\mathbf{C}(\vec{R})$, with the atomic orbitals centered at configuration $\vec{R}'$, that is

$$\left(\mathbf{S}_{\vec{R}}(\vec{R}')\right)_{pq} = \sum_{\mu,\nu}^{M} C_{\mu p}(\vec{R})^* C_{\nu q}(\vec{R}) S^{\mathrm{AO}}_{\mu\nu}(\vec{R}') \tag{4.17}$$

where $\mathbf{S}^{\mathrm{AO}}(\vec{R}')$ is the overlap matrix between the non-orthogonal atomic spin orbitals at geometry $\vec{R}'$. If $\vec{R}' = \vec{R}$, $\mathbf{S}_{\vec{R}}(\vec{R}')$ is the identity matrix because MOs are orthonormal, but this will not be the case otherwise. A new coefficient matrix can now be expressed as

$$\mathbf{C}_{\vec{R}}(\vec{R}') = \mathbf{S}_{\vec{R}}^{-\frac{1}{2}}(\vec{R}')\mathbf{C}(\vec{R}). \tag{4.18}$$

The coefficient matrix $\mathbf{C}_{\vec{R}}(\vec{R}')$ corresponds now to a set of orthonormal MOs created from basis functions centered at $\vec{R}'$, but it is not a Hartree-Fock state, instead representing 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. (4.18), an approach which will mix the occupied and the virtual coefficient matrices, which might lead to larger changes of the occupied block. Hence, we follow a second approach,

converting the occupied and virtual parts of the coefficient matrix independently

$$\mathbf{C}_{\vec{R}}^{(o/v)}(\vec{R}') = \left(\mathbf{S}_{\vec{R}}^{(o/v)}(\vec{R}')\right)^{-\frac{1}{2}} \mathbf{C}_{\vec{R}}^{(o/v)}(\vec{R}) \tag{4.19}$$

where we previously subtracted the projections of all occupied MOs from each virtual orbital, as to ascertain orthogonality 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. 1.3.1, p. 11) does not hold anymore.

4.3.2 Procrustes orbitals

Assume that we have two MO coefficient matrices, possibly representing Hartree-Fock wave functions, that were obtained at two different geometries, $\mathbf{C}(\vec{R})$ and $\mathbf{C}(\vec{R}')$. We are interested in performing a unitary transformation $\mathbf{Q}^{(o/v)}$ 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 $\boldsymbol{C}_{\vec{R}}(\vec{R}')^{(o/v)} = \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{Q}^{(o/v)}$. This does not change the Slater determinant $|\Psi^{\mathrm{SD}}(\vec{R}')\rangle$ except for a phase (cf. sec. 1.3.1, p. 11). Mathematically, we try to find

$$\mathbf{Q}^{(o/v)} = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega} - \mathbf{C}(\vec{R})^{(o/v)} \right\|_F \quad \text{subject to} \quad \mathbf{\Omega}^\dagger\mathbf{\Omega} = \mathbf{1} \tag{4.20}$$

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

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} |A_{ij}|^2} = \sqrt{\mathrm{Tr}\left(\mathbf{A}^\dagger\mathbf{A}\right)} \tag{4.21}$$

In principle, any other norm can be chosen, but using the Frobenius norm, this problem, known as the *orthogonal Procrustes problem* for real matrices, has an analytical solution [81]: Letting

$$\mathbf{M} = \left(\mathbf{C}(\vec{R}')^{(o/v)}\right)^\dagger \mathbf{C}(\vec{R})^{(o/v)} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger \tag{4.22}$$

using the *Singular Value Decomposition (SVD)* [82], the solution is given by

$$\mathbf{Q} = \mathbf{U}\mathbf{V}^\dagger. \tag{4.23}$$

The proof of this can be found in the appendix, sec. B.1, p. 93. We will refer to the orbitals expressed by $\boldsymbol{C}_{\vec{R}}(\vec{R}') = \left[\boldsymbol{C}_{\vec{R}}(\vec{R}')^{(o)} \quad \boldsymbol{C}_{\vec{R}}(\vec{R}')^{(v)}\right]$ 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 with respect 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. [83]. A small lemma is that applying unitary rotations to the reference $\mathbf{C}(\vec{R})^{(o/v)}$ will not give a better fit.

Lemma 4.3.1. *The Procrustes problem is independent of the reference in the sense that a unitary operation applied to the reference coefficient matrix does not change the distance metric.*

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

$$\begin{aligned}
&\min_{\mathbf{\Omega},\mathbf{U}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega} - \mathbf{C}(\vec{R})^{(o/v)}\mathbf{U} \right\|_F \\
&= \min_{\mathbf{\Omega},\mathbf{U}} \left\| \left(\mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega}\mathbf{U}^{-1} - \mathbf{C}(\vec{R})^{(o/v)} \right) \mathbf{U} \right\|_F \\
&= \min_{\mathbf{\Omega},\mathbf{U}} \left\| \mathbf{C}(\vec{R}')^{(o/v)}\mathbf{\Omega}\mathbf{U}^{-1} - \mathbf{C}(\vec{R})^{(o/v)} \right\|_F
\end{aligned} \tag{4.24}$$

where we used that unitary operations preserve the Frobenius norm. Observing that $\mathbf{\Omega U}^{-1}$ 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 4.3.1, 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 B.2.1, p. 94. Normal Procrustes orbitals are analytic when the span of the occupied canonical orbitals and virtual canonical orbitals individually are analytic, which is the case when the HF ground state is analytic, but this is not always the case.

4.3.3 Natural orbitals

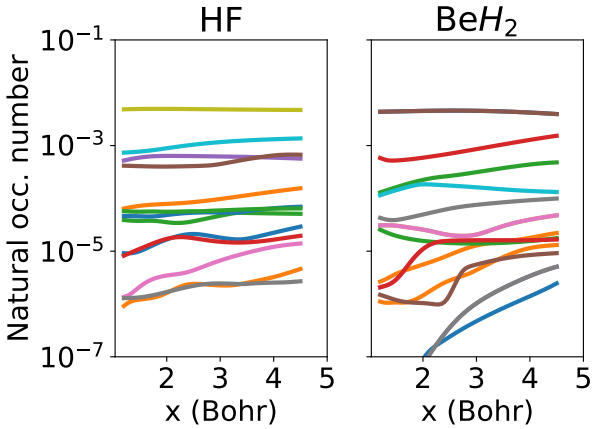


Figure 4.1: Natural occupation numbers of 20% *randomly chosen natural orbitals obtained from the first order perturbation theory wave function with the occupied and the virtual blocks of the 1-RDM diagonalized individually, as function of the internuclear distance x for the symmetric dissociation of hydrogen fluoride and beryllium hydride in the aug-cc-pVTZ basis set.*

We have discussed natural orbitals and their analyticity in section 1.9, p. 20. Even though we stated that analyticity 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 HF state is analytic. Using natural orbitals, we can discard of those NOs with the smallest natural occupation numbers. For this to work, however, the natural occupation numbers need to remain reasonably constant along the whole reaction coordinate $\vec{R}$, as we do not want to discard NOs which are unimportant at some geometries, but important at other geometries. We have observed that small NOs remain reasonably small. This is exemplified in figure 4.1, showing a random selection of natural occupation numbers as a function of the internuclear distance for the dissociation of hydrogen fluoride and the symmetric stretch of beryllium hydride. Albeit some natural occupation numbers change by orders of magnitude along the reaction path, most small natural occupation numbers remain small. Thus, a general distinction between "important" and "unimportant" can still be made. We can thus only keep those natural orbitals that are important at any of the sampling geometries and discard the rest. It should be noted that natural orbitals can be combined with Procrustes orbitals in the sense that one can apply the

orthogonal Procrustes algorithm on the coefficient matrix containing NOs after removing the least important NOs.

4.4 Multi-reference EVC for chemistry

4.4.1 NOCI along a trajectory

Let $\vec{R}_\odot$ be the target geometry. The MO coefficient matrix $\mathbf{C}(\vec{R}_m)$ is converted into $\mathbf{C}_{\vec{R}_m}(\vec{R}_\odot)$ for each sample geometry $\vec{R}_m$, $m = 1, \ldots, L$ using symmetric orthonormalization or generalized Procrustes orbitals to represent the state at the new geometry. Doing this leads to a state $|\Psi_{\vec{R}_m}(\vec{R}_\odot)\rangle$, where the MOs are now built as a linear combination of AOs centered at $\vec{R}_\odot$. This can be considered as "moving" the wave function $|\Psi(\vec{R}_m)\rangle$ to a position $\vec{R}_\odot$, without performing any re-optimization of the MOs. The EVC wave function at the target geometry is then written as

$$|\Psi^{\mathrm{EVC}}(\vec{R}_\odot)\rangle = \sum_{m=1}^{L} c_m \, |\Psi_{\vec{R}_m}(\vec{R}_\odot)\rangle \, . \tag{4.25}$$

This is a NOCI approach (cf. sec. 1.10, p. 21), as two states $|\Psi_{\vec{R}_m}(\vec{R}_\odot)\rangle$, $|\Psi_{\vec{R}_n}(\vec{R}_\odot)\rangle$ will be constructed from two different sets of molecular orbitals (which, however, are constructed from the same AOs). These two sets of MOs will be non-orthogonal, making it necessary to use the generalized Slater-Condon rules (cf. sec. 1.5, p. 13) for the calculation of matrix elements. Thus, this approach is too expensive for methods where the wave functions $|\Psi(\vec{R}_m)\rangle$ consist of more than a few Slater determinants, such as CISD or CCSD, and we will in this thesis only consider the case where $|\Psi(\vec{R}_m)\rangle$ is a Slater determinant.

4.4.2 Multi-reference EVC with tweaked 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,

$$\hat{H}(\lambda) = \sum_{i=1}^{N} \left[-\frac{1}{2}\vec{\nabla}_i^2 - \sum_{A=1}^{N_A} \frac{Z_A}{\left|\vec{r}_i - \vec{R}_A\right|} \right] + \lambda \sum_{j<i}^{N} \frac{1}{|\vec{r}_i - \vec{r}_j|} + h_{\mathrm{nuc}} = \hat{H}_1 + \lambda \hat{H}_2 + h_{\mathrm{nuc}} \tag{4.26}$$

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 gets less accurate as λ increases. The approach is to select a set of $\lambda_i \in [0.5, 1.1]$, $i = 1, \ldots, L$ and obtain Slater determinants $|\Phi^{\mathrm{SD}}(\lambda_i)\rangle$ which serve as EVC sample vectors. The range $[0.5, 1.1]$ is relatively arbitrary, but should contain values close to 1 in order not to deviate too much from the real physics of the system. It should be noted that this method is not an EVC method: we are neither interpolating nor continuing a solution, but try to build a small space of non-orthogonal Slater determinants that might approximate the accurate solution. This manifests in the fact that we aim to obtain a wave function at the physically correct value $\lambda = 1$ which improves the HF-state $|\Phi^{\mathrm{SD}}(\lambda = 1)\rangle = |\Phi^{\mathrm{HF}}\rangle$. The state $|\Phi^{\mathrm{HF}}\rangle$ itself can be one of the sample states.

4.5 Single-reference EVC for chemistry

While the multi-reference approach is more general, calculating matrix elements for non-orthogonal wave functions is prohibitively expensive for e.g. CC wave functions. Here, we

consider how we can transform states from one geometry to another such that a single-reference picture is kept for post-Hartree-Fock methods. 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 $\vec{R}_m$. Then, for example, the $\hat{T}_1$ operator is given by

$$\hat{T}_1(\vec{R}_m) = \sum_{ia} t_i^a(\vec{R}_m) a_a^\dagger(\vec{R}_m) a_i(\vec{R}_m) \tag{4.27}$$

where the creation and annihilation operators refer to MOs expressed in an AO basis which is centered at $\vec{R}_m$. Following the discussion in sec. 4.2.1, the basis states and thus the creation and annihilation operators can be made geometry independent, and all geometry dependence is in the Hamiltonian. We will thus simply write

$$\hat{T}_1(\vec{R}_m) = \sum_{ia} t_i^a(\vec{R}_m) a_a^\dagger a_i. \tag{4.28}$$

The same geometry-independent representation is also used for any operator expressed in terms of creation and annihilation operators, e.g. $\hat{T}_2, \hat{T}_3, \ldots$ or the CI operators $\hat{C}_1, \hat{C}_2, \ldots$.

It is important that the excitation coefficients $t_i^a(\vec{R})$ etc. are obtained with respect to MOs that change analytically for EVC to work. It remains to discuss how to choose the MO expansion coefficients, which impacts the final energy of the EVC wave function. Any of the methods described in section 4.3 can be used. There are some details and steps to consider for each method, which are described in the following list.

1. Procrustes orbitals: Define a reference geometry $\vec{R}_{\text{ref}}$, solve the Hartree-Fock equations and obtain an MO coefficient matrix $\mathbf{C}(\vec{R}_{\text{ref}})$, where the occupied MOs correspond to a HF state. For each sample geometry $\vec{R}_m$, solve the HF equations to obtain $\mathbf{C}(\vec{R}_m)$. Then, follow the approach described in section 4.3.2 for the occupied and the virtual MOs individually to find the Procrustes orbitals with corresponding coefficient matrix $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_m)$. As this is done on the occupied and the virtual orbitals individually, this guarantees that a Hartree-Fock state is chosen as the reference determinant at the sample geometries. The cluster operator $\hat{T}(\vec{R}_m)$ is found with respect to the Procrustes orbitals expressed by $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_m)$. At the target geometry $\vec{R}_\odot$, use again the Procrustes algorithm to obtain the Procrustes coefficient matrix $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_\odot)$. This can also be done with generalized Procrustes orbitals, which guarantees analyticity of the MO coefficient matrix and a smaller change in the Hamiltonian, but yields non-Hartree-Fock reference states.

2. Symmetric orthonormalization: Similarly to Procrustes orbitals, define a reference geometry $\vec{R}_{\text{ref}}$, solve the HF equations and obtain $\mathbf{C}(\vec{R}_{\text{ref}})$. For the sample geometries, the cluster operator is obtained with respect to the MOs in $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_m)$, where $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_m)$ is the symmetrically orthogonalized set of MOs $\mathbf{C}(\vec{R}_{\text{ref}})$ at the sample geometry $\vec{R}_m$. This will not correspond to a HF state. Similarly, at the target geometry, $\boldsymbol{C}_{\vec{R}_{\text{ref}}}(\vec{R}_\odot)$ is used.

3. Natural/canonical orbitals: Calculate the cluster operator $\hat{T}(\vec{R}_m)$ at the sample geometries $\vec{R}_m$ with respect to natural orbitals or canonical orbitals, taking special care of the correct ordering and degenerate canonical/natural orbitals such that they are analytic. This approach allows for freezing of natural orbitals that have a small occupation number along the whole geometry. At the target geometry $\vec{R}_\odot$, natural or canonical orbitals are used, too.

Each of these approaches defines how the MOs are mapped from one geometry to another. Having obtained sample cluster operators $\hat{T}(\vec{R}_m)$ and knowing which MOs to use at the

target geometry $\vec{R}_\odot$, it is then straightforward to apply either AMP-CCEVC, WF-CCEVC or another EVC method. Consider AMP-CCEVC using Procrustes orbitals as an example. The wave function is written as

$$|\Psi^{\text{AMP-CCEVC}}(\vec{R}_\odot)\rangle = \exp\left(\sum_{m=1}^{L} c_m \hat{T}(\vec{R}_m)\right)|\Phi^{\text{HF}}(\vec{R}_\odot)\rangle \tag{4.29}$$

where the amplitudes $t_i^a(\vec{R}_m), t_{ij}^{ab}(\vec{R}_m)$ were obtained using Procrustes orbitals with reference $\vec{R}_{\text{ref.}}$ at geometry $\vec{R}_m$, and both the occupied and the virtual orbitals at geometry $\vec{R}_\odot$ also are Procrustes orbitals with reference $\vec{R}_{\text{ref.}}$.

4.6 Comparison of choices of MOs

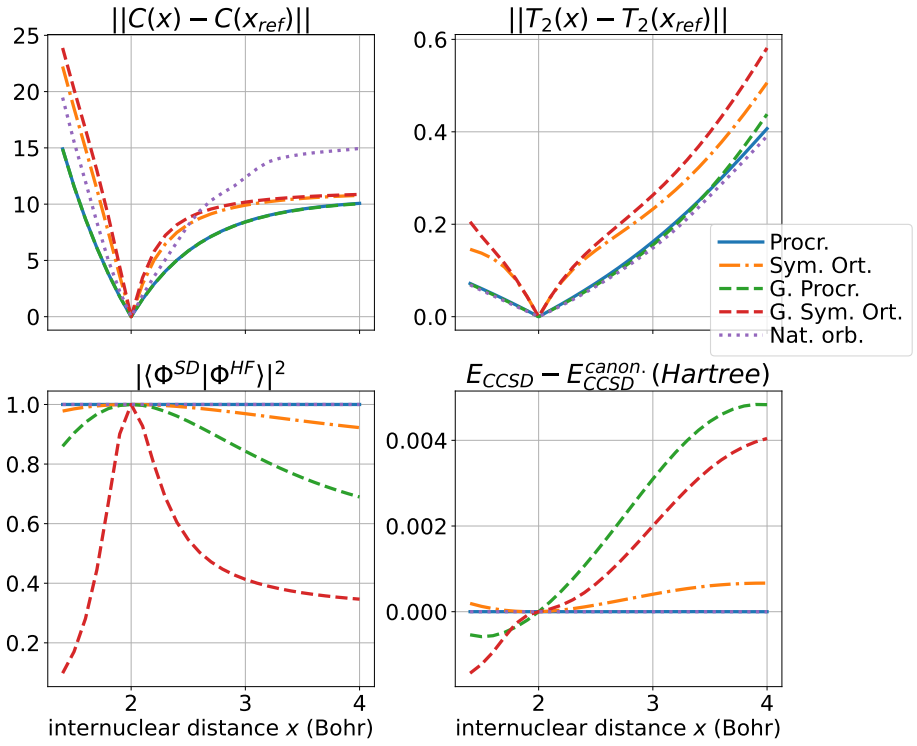


Figure 4.2: Using $x_{ref} = 2$ Bohr, the norm of the difference in the coefficient matrices (upper left), $\hat{T}_2$-amplitudes (upper right), the overlap to the HF state (lower left) and the CCSD energy compared to canonical orbitals (lower right) are depicted for the hydrogen fluoride molecule in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

Each of the methods described in sec. 4.3 has advantages and disadvantages and are useful for different applications.

Procrustes orbitals require solving the HF equations at each target geometry. Furthermore, Procrustes orbitals require the choice of an arbitrary reference geometry $\vec{R}_{\text{ref.}}$ which has no physical meaning. Because the use of Procrustes orbitals guarantees that the reference determinant is a HF state, they work as well as canonical orbitals for post HF methods. Furthermore, Procrustes orbitals guarantee that the coefficient matrices $\mathbf{C}^{(o/v)}$ change as little as possible. The use of Procrustes orbitals fails when the HF ground state changes discontinuously.

Generalized Procrustes orbitals are similar to Procrustes orbitals in that they require a reference geometry $\vec{R}_{\text{ref.}}$. While no HF state needs to be obtained, the reference will not be a HF state. However, the coefficient matrix $\mathbf{C}$ as a whole changes as little as possible.

Natural orbitals require that an approximate 1-RDM is diagonalized, where the 1-RDM can be obtained using, for example, Møller-Plesset perturbation theory. They work as good as canonical orbitals for post HF methods. The use of NOs fails when the HF ground state changes discontinuously. For natural orbitals (or canonical orbitals), there is an assignment problem to be solved in order to assure analyticity (jf. sec. 1.9.1, p. 21). For this assignment problem at two geometries $\vec{R}_1$,$\vec{R}_2$ to give the correct results, the geometries need close. This means that the HF equations need to be solved with relatively high resolution along the whole PES *before* obtaining the sample wave functions in order to assign them correctly, which is discussed in greater detail in sec. 5.3.

Symmetric orthonormalization minimizes the change of MOs in a least-square sense, but does not lead to a Hartree-Fock wave function and, when being used for single-reference problems, requires an arbitrary reference geometry $\vec{R}_{\text{ref.}}$.

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. This type of weighing might be sensible to make the most important MOs change less, while the less important MOs may change more.

In figure 4.2, we compare different orbital rotation schemes for the dissociation of HF. It shows the change in the coefficient matrix and the change in the CCSD $\hat{T}_2$ amplitudes from the reference, the overlap to the Hartree-Fock wave function, and the difference to the canonical CCSD energy. Procrustes orbitals, natural orbitals and generalized Procrustes orbitals, unlike symmetric orthonormalization, cause small changes in the coefficient matrix and the absolute value of the $\hat{T}_2$ amplitudes. Symmetric orthonormalization leads to a Slater determinant that has much more overlap with the exact HF state than generalized Procrustes orbitals do. Finally, we see that Procrustes orbitals and NOs lead to the same CCSD energies as using canonical orbitals, while the other choices do not, which is the expected behaviour. 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. 13) implies that CCSD is relatively independent of the reference state, too large deviations from the "good" reference nevertheless lead to quantitatively wrong results, as a different problem is being solved.

4.6.1 The effect of avoided crossings for single-reference EVC

In figure 1.1, p. 21, two avoided crossings between two pairs of natural orbitals are marked. The existence of avoided crossings have an important consequence for eigenvector continuation because they make both canonical and natural orbitals unsuitable choices. For natural orbitals, the eigenvalues that almost cross are the natural occupation numbers, while for canonical orbitals, the eigenvalues that almost cross are the molecular energies. Along the trajectory, both the upper lying and the lower lying orbitals are analytic functions of the perturbation, so avoided crossings do not impact the analyticity of the MO coefficient matrix. However, at avoided crossings, the orbitals change qualitatively. To see this, we define the unitary matrix $\tilde{\mathbf{Q}}$

$$\tilde{\mathbf{Q}}(\vec{R}) = \mathbf{S}(\vec{R})^{\frac{1}{2}}\mathbf{C}(\vec{R}) \tag{4.30}$$

where $\mathbf{C}$ is the coefficient matrix representing natural/canonical orbitals and $\mathbf{S}$ the atomic spin orbital overlap matrix. It is hence a geometry-independent representation of the MOs. We assume the columns in $\mathbf{C}$ to be ordered and multiplied with the correct sign in such a way that they are analytic. To see how much MOs change as the nuclei are moved from $\vec{R}_1$ to $\vec{R}_2$, we define a (non-Hermitian) matrix describing how much the MOs have changed

$$\mathbf{O}(\vec{R}_1, \vec{R}_2) = \tilde{\mathbf{Q}}(\vec{R}_1)^\dagger \tilde{\mathbf{Q}}(\vec{R}_2). \tag{4.31}$$

If orbitals change "little" between $\vec{R}_1$ and $\vec{R}_2$, $\mathbf{O}(\vec{R}_1, \vec{R}_2)$ is approximately the identity matrix.

Assume now there is an avoided crossing between natural/canonical orbitals $|\psi_i\rangle$ and $|\psi_j\rangle$ going from $\vec{R}_1$ to $\vec{R}_2$. Then it is possible that O_{ii}, O_{jj} are no longer close to one. Instead, O_{ij}, O_{ji} will be close to 1. But this means that the orbitals have changed qualitatively and $|\psi_i(\vec{R}_1)\rangle$ resembles $|\psi_j(\vec{R}_2)\rangle$ more than $|\psi_i(\vec{R}_2)\rangle$. This is indeed the expected behaviour of eigenvectors of Hermitian matrices at avoided crossings [33]. For that reason, eigenvector continuation will fail, as the qualitative change in the orbitals means that, for example, the excitation coefficients $t_i^a(\vec{R}_1)$ are not even qualitatively close to $t_i^a(\vec{R}_2)$. Thus, we conclude that only Procrustes orbitals are a viable choice out of all presented schemes, as it is the only scheme that allows both to use a HF reference state at each geometry (which is necessary to obtain the same energies as when using canonical orbitals), and ascertain that the MOs do not change qualitatively, which is necessary for EVC to be efficient.

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, the discussion in section 2.2.3, p. 26, is valid, and including extra sampling points will only lead to improvements in the expectation value of the energy. Using qubit representation of states, all geometry independence is in the Hamiltonian $\hat{H}(\vec{R})$, identically to the discussion in sec. 4.2.1. Thus, the qubit-representation of a Slater determinant $|\Phi^{\mathrm{SD}}\rangle$ is geometry independent. For EVC, the principal idea follows the same steps as before:

1. For a set of sample geometries $\vec{R}_m$, $m = 1, \ldots, L$, obtain approximate ground state wave functions $|\Psi(\vec{R}_m)\rangle = \hat{U}(\vec{R}_m)\,|\Phi^{\mathrm{SD}}\rangle$, where $\hat{U}(\vec{R}_m)$ is a unitary operator obtained, for example, using one of the ansätze in sec. 3.5, p. 34. It is important that the reference Slater determinant is chosen in such a way that MO expansion coefficients change analytically, e.g. following any of the descriptions in sec. 4.5. By the discussion of the previous section, Procrustes orbitals are the method of choice.

2. For each $\vec{R}_\odot$, obtain the second-quantized Hamiltonian $\hat{H}(\vec{R}_\odot)$, expressed again in terms of MOs with expansion coefficients that change analytically with $\vec{R}$.

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

$$H_{mn}^{\mathrm{EVC}}(\vec{R}_\odot) = \langle\Psi(\vec{R}_m)|\,\hat{H}(\vec{R}_\odot)\,|\Psi(\vec{R}_n)\rangle \tag{4.32}$$

$$S_{mn}^{\mathrm{EVC}} = \langle\Psi(\vec{R}_m)|\Psi(\vec{R}_n)\rangle\,. \tag{4.33}$$

   Observe that the states $|\Psi(\vec{R}_m)\rangle$ have the same qubit representation (and also the same ON vector representation) independently of the target geometry $\vec{R}_\odot$.

4. Solve the EVC equations.

The only explicit reference to the target geometry $\vec{R}_\odot$ is in the Hamiltonian $H_{ij}^{\mathrm{EVC}}(\vec{R}_\odot) = \langle\Psi(\vec{R}_i)|\,\hat{H}(\vec{R}_\odot)\,|\Psi(\vec{R}_n)\rangle$. Each Hamiltonian is a sum of Pauli strings

$$\hat{H}(\vec{R}) = \sum_{i_1,i_2,\ldots i_Q} c_{i_1,i_2,\ldots i_Q}(\vec{R})\tilde{\sigma}_1^{i_1}\tilde{\sigma}_2^{i_2}\ldots\tilde{\sigma}_Q^{i_Q} \tag{4.34}$$

where the dependence on the geometry is in the coefficients only, corresponding to eq. (4.16). Independently of the geometry, the Hamiltonian $\hat{H}(\vec{R})$ will have the same set of coefficients $c_{i_1,i_2,\ldots i_Q}$ that are nonzero. For that reason, when calculating $\mathbf{H}^{\mathrm{EVC}}(\vec{R})$, the expectation values

$\langle \tilde{\sigma}_1^{i_1} \tilde{\sigma}_2^{i_2} \dots \tilde{\sigma}_Q^{i_Q} \rangle_{ij}$ need only to be sampled at one single geometry, as they are the same for each $\vec{R}$. Similarly, the overlap matrix $\mathbf{S}^{\text{EVC}}$ is geometry-independent. Thus, when calculating a potential energy surface with k points of interest, only $O(L^2M^4)$ expectation values need to be sampled independently of k. In addition, only L unitaries $\hat{U}(\vec{R}_m)$ need to be optimized, as there are L sample wave functions.

4.7.2 Multi-reference approach

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

$$|\Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m)\rangle = \hat{U}_m \hat{K}(\vec{R}_m) |\Phi^{\text{SD}}(\vec{R}_m)\rangle \tag{4.35}$$

where $\hat{U}_m = e^{-\hat{\kappa}_m}$ is a change of basis operator defined from the matrix $\mathbf{U}_m$ as described in section 1.4, p. 12, where $\mathbf{U}_m$ is obtained by solving the generalized Procrustes problem

$$\mathbf{U}_m = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}_{\text{ref}}) - \mathbf{C}(\vec{R}_m)\mathbf{\Omega} \right\| \quad \text{subject to} \quad \mathbf{\Omega}^\dagger \mathbf{\Omega} = \mathbf{1}. \tag{4.36}$$

This is the unitary transformation applied at geometry $\vec{R}_m$ that makes the resulting MO coefficient matrix closest to the one that was used at geometry $\vec{R}_{\text{ref}}$. Unlike the single-reference approach, the resulting wave function $|\Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m)\rangle$ will have a more complicated structure than $|\Psi(\vec{R}_m)\rangle$: If, for example, $|\Psi(\vec{R}_m)\rangle$ is a Slater determinant in its reference basis, then $|\Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m)\rangle$ will generally not be a Slater determinant in the new basis, as mixing occurs between 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 show two paragraphs below. On quantum computers, however, this basis change operation is cheap and readily implemented [84], requiring the Jordan-Wigner mapping. This leads to the following algorithm:

1. For $m = 1, \dots, L$, find approximate ground states $|\Psi(\vec{R}_m)\rangle = K(\vec{R}_m) |\Phi^{\text{SD}}\rangle$. The use of Procrustes orbitals for the reference determinant $|\Phi^{\text{SD}}\rangle$ is possible whenever desired, but using e.g. canonical orbitals works too.
2. Define an (arbitrary) reference geometry $\vec{R}_{\text{ref}}$.
3. For $m = 1, \dots, L$, obtain the states $|\Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m)\rangle$.
4. At the target geometry $\vec{R}_\odot$, express the Hamiltonian $\hat{H}_{\vec{R}_{\text{ref}}}(\vec{R}_\odot)$ in terms of MOs defined by the coefficient matrix $\mathbf{C}(\vec{R}_\odot)\mathbf{U}_\odot$.
5. Find the matrix elements
$$\begin{aligned} H_{mn}^{\text{EVC}}(\vec{R}_\odot) &= \langle \Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m) | \hat{H}_{\vec{R}_{\text{ref}}}(\vec{R}_\odot) | \Psi_{\vec{R}_{\text{ref}}}(\vec{R}_n) \rangle \\ S_{mn}^{\text{EVC}} &= \langle \Psi_{\vec{R}_{\text{ref}}}(\vec{R}_m) | \Psi_{\vec{R}_{\text{ref}}}(\vec{R}_n) \rangle \,. \end{aligned}$$
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 ground state Slater determinant, which is used as the reference state, changes discontinuously. The unitary transform of the wave function into a basis where the Hamiltonian changes analytically is performed *after* the application of a post-HF ansatz.

One might argue why the use of a reference geometry even is necessary. Indeed, not using a reference has a conceptual advantage. For every target geometry, we could instead directly transform to the target geometry, that is, build a state

$$|\Psi_{\vec{R}_\odot}(\vec{R}_m)\rangle = \hat{U}_{m\to\odot}\hat{K}(\vec{R}_m)\,|\Phi^{\mathrm{SD}}(\vec{R}_m)\rangle \tag{4.37}$$

with

$$\mathbf{U}_{m\to\odot} = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{C}(\vec{R}_\odot) - \mathbf{C}(\vec{R}_m)\mathbf{\Omega} \right\| \quad \text{subject to} \quad \mathbf{\Omega}^\dagger\mathbf{\Omega} = \mathbf{1}. \tag{4.38}$$

This approach destroys the advantages of eigenvector continuation on quantum computers. Using $\hat{U}_m$ with a fixed reference, the overlap matrix $\mathbf{S}^{\mathrm{EVC}}$ is independent of $\vec{R}_\odot$, and as previously, sampling of Pauli matrix expectation values only needs to be performed once. Using $|\Psi_{\vec{R}_\odot}(\vec{R}_m)\rangle$, $\mathbf{S}^{\mathrm{EVC}}$ will differ at each each geometry, hence the overlap matrix and Pauli matrix expectation values become geometry-dependent and need to be resampled at each geometry.

We 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 corresponding creation and annihilation operators denoted as a_p, b_p, c_p and $a_p^\dagger, b_p^\dagger, c_p^\dagger$. Let $|\Psi_m^A\rangle$, $|\Psi_n^B\rangle$ be linear combinations 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

$$H_{mn} = \langle\Psi_m^A|\,\hat{H}_C\,|\Psi_n^B\rangle\,, \tag{4.39}$$

we need to express $|\Psi_m^A\rangle\,, |\Psi_n^B\rangle$ in terms of MOs C by performing a unitary transformation on $|\Psi_m^A\rangle$, $|\Psi_m^B\rangle$ as described in section 1.4, p. 12. This leaves the state unchanged, we write $|\Psi_m^A\rangle = \exp\left(-\hat{\kappa}_A\right)|\Psi_m^C\rangle$ and similarly for $|\Psi_m^B\rangle = \exp\left(-\hat{\kappa}_B\right)|\Psi_m^C\rangle$. Now

$$H_{mn} = \langle\Psi_m^C|\exp\left(\hat{\kappa}_B\right)\hat{H}_C\exp\left(-\hat{\kappa}_A\right)|\Psi_n^C\rangle \tag{4.40}$$

applying the operations $\exp\left(-\hat{\kappa}\right)$ 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 be a change of basis operation of the Hamiltonian. However, as this is not the case, there is no simple expression.

4.7.3 Failed ideas to make overlap matrix approximately diagonal

In section 8.1.2. p. 83, it can be seen that very small eigenvalues of 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, $S_{ii}^{\mathrm{EVC}} = 1$, we can write $\mathbf{S}^{\mathrm{EVC}} = \boldsymbol{I} - \mathbf{\Theta}$ and apply the Neumann series where $\mathbf{\Theta}$ fulfills $\|\mathbf{\Theta}\| < 1$ in the operator norm:

$$\left(S^{\mathrm{EVC}}\right)^{-1} = (\boldsymbol{I} - \mathbf{\Theta})^{-1} = \sum_{i=0}^{\infty}\mathbf{\Theta}^i \tag{4.41}$$

from this, we see that by making the off-diagonal elements $\Theta_{ij} = -S_{ij}^{\mathrm{EVC}}$ small, we obtain a matrix inverse with elements that do not "blow up" and stay close to identity. This also

means that **S** will then have no small eigenvalues. 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 new overlap matrix $\mathbf{S}'$ would have the same eigenvalues as **S**. Based on LCU (cf. sec. 3.13, p. 43), we developed a method to calculate matrix elements of arbitrary linear combinations of wave functions. Let $|\Psi_m'\rangle$, $|\Psi_n'\rangle$ be states that can be obtained with LCU. Then, one can use the LCU algorithm to implement the state

$$|\Omega\rangle = \frac{1}{N}\left(\frac{1}{p_m}|0^n\rangle|0\rangle|\Psi_m'\rangle + \frac{1}{p_n}|0^n\rangle|1\rangle|\Psi_m'\rangle\right) \tag{4.42}$$

where n is the number of ancilla qubits necessary for the LCU algorithm, N is a normalization constant and $1/p_m$, $1/p_n$ are success probabilities for the implementation of those states. From eq. (4.42), 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 would be a *Toffoli gate*, which cannot be decomposed into less than 6 CNOT-gates [85], which are considered expensive. Also, the success probability of the individual measurements will be very small, as the (non-normalized) 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 [86], 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. 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 that span the same vector space. In order to do this, we adapt the approach described in section 3.12. 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 $|\Psi(\vec{R}_m)\rangle$ where $m \in \{1, \ldots, k-1\}$. As the states $|\Psi(\vec{R}_m)\rangle$ likely are not ground states at geometry $\vec{R}_k$, we will not obtain excited states, but states with a low energy that are near-orthogonal to $|\Psi(\vec{R}_m)\rangle$. However, this method fails, as is illustrated by the following minimal example.

Assume a symmetric 3×3 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 $\mathbf{v}_m(x)$ 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 eigenvalue $E > E_0$, which makes the method fail. This is because 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 $\text{span}\{\mathbf{v}_0(0), \ldots, \mathbf{v}_0^{(n)}(0)\}$. It can be shown that the first excited state usually has a large contribution to the derivative $\mathbf{v}_0^{(n)}(0)$ [10, ch. 14], but this is not sufficient to do proper EVC.

Chapter 5

Numerical implementation and algorithms

In this chapter, we explain how many of the methods described in the previous chapters were implemented. We 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 discuss our code implementation, which libraries and which functionalities were used, and where to find our code. Finally, we present the chemical systems used to assess how well the EVC methods work.

5.1 Numerical methods

5.1.1 The Moore-Penrose Pseudoinverse

Let $\mathbf{S} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\dagger$ be a matrix with small singular values, that is, some elements of the nonnegative diagonal matrix $\boldsymbol{\Sigma}$ when performing an SVD are small. When inverting $\mathbf{S}$, we see that its inverse

$$\mathbf{S}^{-1} = \mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{U}^\dagger \tag{5.1}$$

will have very large singular values and hence eigenvalues. It will thus change a lot when the smallest eigenvalues of $\mathbf{S}$ are uncertain or numerically unstable. From a computational point of view, one can consider some cutoff parameter ϵ, and remove the smallest singular values

$$\tilde{\Sigma}_{ii} = \begin{cases} \Sigma_{ii} & \text{if } \Sigma_{ii} > \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{5.2}$$

and use as an alternative of the inverse

$$\tilde{\Sigma}^+_{ii} = \begin{cases} 1/\tilde{\Sigma}_{ii} & \text{if } \tilde{\Sigma}_{ii} > 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.3}$$

which leads to a more numerically stable inverse

$$\mathbf{S}^{-1} \approx \mathbf{S}^+ = \mathbf{V}\tilde{\boldsymbol{\Sigma}}^+\mathbf{U}^\dagger. \tag{5.4}$$

$\mathbf{S}^+$ is called *the Moore-Penrose Pseudoinverse* of $\mathbf{S}$ [87]. 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 the sum goes over the R smallest singular values that were removed and is thus small, which justifies this procedure.

5.1.2 The Hermitian generalized eigenvalue problem: Canonical orthogonalization

The Hermitian eigenvalue problem is given by

$$\mathbf{HC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{5.5}$$

where $\mathbf{H}, \mathbf{S}$ are Hermitian matrices. The Hermitian generalized eigenvalue problem can be transformed into a standard Hermitian eigenvector problem by canonical orthogonalization [13], [36], which can be applied when $\mathbf{S}$ is positive semi-definite. As $\mathbf{S}$ is Hermitian, it can be diagonalized as

$$\mathbf{S} = \mathbf{U}\mathbf{s}\mathbf{U}^{\dagger}. \tag{5.6}$$

$\mathbf{s}$ might have some very small diagonal elements, corresponding to small eigenvalues of $\mathbf{S}$ which might lead to round-off errors. Observe that all elements s_{ii} are positive or equal to zero by assumption. We can define a cutoff parameter 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

$$\begin{aligned} \mathbf{HC} &= \mathbf{SC}\boldsymbol{\epsilon} \\ \mathbf{HXC}' &= \mathbf{SXC}'\boldsymbol{\epsilon} \\ \left(\mathbf{X}^{\dagger}\mathbf{HX}\right)\mathbf{C}' &= \mathbf{X}^{\dagger}\mathbf{SXC}'\boldsymbol{\epsilon} \\ \mathbf{H}'\mathbf{C}' &= \mathbf{C}'\boldsymbol{\epsilon} \end{aligned} \tag{5.7}$$

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 [88] 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 even for calculations carried out with data from quantum computers [89].

5.1.3 The non-Hermitian generalized eigenvalue problem

The non-Hermitian generalized eigenvalue problem is given by

$$\mathbf{HC} = \mathbf{SC}\boldsymbol{\epsilon} \tag{5.8}$$

where $\mathbf{H}, \mathbf{S}$ are not Hermitian matrices. This is known to be a difficult problem [37]. We will here present some 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 [90, chap. 7.7], which is less prone to numerical errors, but may fail for ill-conditioned problems. The `GUPTRI` algorithm [91] adapts the generalized Schur decomposition and finds the regular eigenvalues, hence it is applicable to ill-conditioned problems. However, it requires several parameters that require substantial knowledge of the underlying matrix structure. Furthermore, we observed that the smallest eigenvalues were overestimated. A possible alternative is to instead invert $\tilde{\mathbf{S}} = \mathbf{S} + \epsilon \exp\{-\mathbf{S}/\epsilon\}$ for a small constant $\epsilon \in [10^{-10}, 10^{-14}]$, e.g. diagonalizing $\tilde{\mathbf{S}}^{-1}\mathbf{H}$ instead [92]. We have found that WF-CCEVC is 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 best choice of the parameter ϵ, which we considered to be the smallest possible before the eigenvalues became nonsensical or unstable, is problem-dependent, and larger choices for ϵ give higher eigenvalues, but usually, $\epsilon = 10^{-14}$ worked well.

5.2 Calculation of matrix elements between non-orthogonal Slater determinants

We have previously mentioned the Generalized Slater-Condon rules in sec. 1.5, p. 13, which are used to calculate matrix elements between Slater determinants expressed in terms of different sets of MOs. These formulas are correct and fully general, but have an inconvenient scaling. This can be avoided using biorthogonalization [93]. Biorthogonalization can be used to diagonalize the overlap matrix ${}^{wx}\mathbf{S}$, turning the first- and second- order cofactor matrices ${}^{wx}\mathbf{S}^{(i,j)}$ and ${}^{wx}\mathbf{S}^{(ij,kl)}$ into diagonal matrices. Computation of matrix elements without biorthogonalization has been considered previously in Refs. [94] and [95], but scales less conveniently. Consider two non-orthogonal Slater determinants $|{}^{x}\Phi^{\mathrm{SD}}\rangle$ and $|{}^{w}\Phi^{\mathrm{SD}}\rangle$, built from a set of molecular orbitals $\{|{}^{x}\phi_i\rangle\}_{i=1}^{N}$ and $\{|{}^{w}\phi_i\rangle\}_{i=1}^{N}$. Both sets of MOs are constructed from the same set of AOs.

We have already defined the overlap $\langle {}^{w}\Phi^{\mathrm{SD}}|{}^{x}\Phi^{\mathrm{SD}}\rangle = |{}^{wx}\mathbf{S}|$, with ${}^{wx}\mathbf{S}_{ij} = \langle {}^{w}\phi_i|{}^{x}\phi_j\rangle = \left(({}^{w}\mathbf{C}^{o})^{\dagger} \times \mathbf{S}^{\mathrm{AO}} \times ({}^{x}\mathbf{C}^{o})\right)_{ij}$ where $\mathbf{C}^{o}$ is the coefficient matrix of the occupied MOs and $\mathbf{S}^{\mathrm{AO}}$ is the overlap matrix in the atomic spin orbital basis. Using a variant of the SVD, we can diagonalize the overlap matrix as

$$ {}^{wx}\mathbf{S} = \mathbf{U}\mathbf{D}\mathbf{V}^{\dagger} \tag{5.9} $$

where $\mathbf{U}, \mathbf{V}$ are *special unitary matrices* with determinant $+1$ and $\mathbf{D}$ is diagonal, but not necessarily nonnegative.[1] This gives

$$ \left({}^{w}\mathbf{C}^{o}\mathbf{U}\right)^{\dagger} \mathbf{S}^{\mathrm{AO}} \left({}^{x}\mathbf{C}^{o}\mathbf{V}\right) = \mathbf{D}, \tag{5.10} $$

e.g. we apply orbital rotations that leave the states $|{}^{x}\Phi^{\mathrm{SD}}\rangle$ and $|{}^{w}\Phi^{\mathrm{SD}}\rangle$ unchanged and the MOs orthogonal. We refer to the rotated MOs expressed by ${}^{w}\mathbf{C}^{o}\mathbf{U}$ as $\{|{}^{w}\tilde{\phi}_i\rangle\}_{i=1}^{N}$, and to the MOs expressed by ${}^{x}\mathbf{C}^{o}\mathbf{V}$ as $\{|{}^{x}\tilde{\phi}_i\rangle\}_{i=1}^{N}$. Because the overlap matrix $\mathbf{D}$ between the transformed MOs is now diagonal, calculations are simplified. We define the overlap S and the reduced overlap $\tilde{S}$

$$ S = \prod_{i} D_{ii} \tag{5.11} $$

$$ \tilde{S} = \prod_{\substack{i \\ D_{ii}\neq 0}} D_{ii} \tag{5.12} $$

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:

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

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

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

[1] The matrices are easily obtained from the standard SVD by multiplying one diagonal element of $\mathbf{D}$ and one column of $\mathbf{U}$ and/or $\mathbf{V}$ with the necessary phase.

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

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

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. [93], 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 our code. The most expensive calculation is the SVD, scaling as $O(N^3)$.

5.3 Ascertaining the correct eigenvalue trajectories

Let $\mathbf{C}(\vec{R}_1)$, $\mathbf{C}(\vec{R}_2)$ be MO coefficient matrices describing natural orbitals at different geometries. The matrix $\mathbf{O}(\vec{R}_1, \vec{R}_2)$ introduced in section 4.6.1 is a measure for the similarity between the orbitals at the two geometries. If $\vec{R}_1$ is close to $\vec{R}_2$, 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. 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. It is thus 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. This happens especially at avoided crossings. Thus, we also need to construct the NOs at some intermediate geometries

$$\vec{R}_i' = \vec{R}_1 + \frac{i}{K}(\vec{R}_2 - \vec{R}_1) \tag{5.13}$$

for some $K > 1$ and $i = 1, \ldots, K-1$. Finally, we need to consider what happens when some NOs are degenerate at 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. This same discussion is also applicable to canonical orbitals instead of natural orbitals, using orbital energies instead of natural occupation numbers.

5.4 Implementation of AMP-CCEVC

When discussing the AMP-CCEVC method and its parameter-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. (1.71), 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 $\{\tilde{\mathbf{t}}_1, \ldots, \tilde{\mathbf{t}}_L\}$ that span the same space (e.g there are parameters k_{mn} such that $\mathbf{t}_m = \sum_{n=1}^{L} k_{mn}\tilde{\mathbf{t}}_n$), which ascertains that the coefficients c_m, $m = 1, \ldots, L$ do not grow arbitrarily large. Other orthogonalization schemes can of course be used. We also implemented DIIS [96], [97],

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 the CCSD amplitude equations.

5.5 Code implementation

All code is available on Github.[2] 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` [98] and `SciPy` [99]. Methods not existing in those libraries were written from scratch, often relying on `NumPy` data structures.

5.5.1 Quantum chemistry

For quantum chemical calculations, we used both in-house code for coupled cluster calculations and `PySCF` [100], [101] 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` [22] 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 non-orthogonal Slater determinant matrix elements, we need to perform sums over multidimensional arrays/tensors. This is a computational bottleneck, and we used the `opt_einsum` package [102] to speed up those calculations.

5.5.3 Quantum computation

For simulations on a quantum computer, `Qiskit` [103] 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 4.7.2, we obtained the relevant circuit using `OpenFermion` [104] where it is implemented. For optimizing the ansatz $\hat{U}(\vec{\theta})\,|\Phi_0\rangle$, we used the *BFGS* optimization algorithm, which is the most used Quasi-Newton approach and works well for optimization [105]. We avoided the problem of sampling expectation values and gradients by using simulated exact ones, unless we wanted to study the effect of sampling. After obtaining circuits, heavy use of `Qiskit`'s `transpile(optimize=2)` function was made, which collapses several gates into one and uses gate identities to minimize computational cost.

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

5.6 Problems

In this section, we present the chemical systems which are used to assess how well EVC works for chemistry. The systems are small but complicated, and CCSD fails qualitatively or gives unsatisfying results quantitatively. We only consider closed shell singlet systems and use restricted Hartree-Fock Slater determinants as well as symmetry-adapted CCSD to decrease computational cost.

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 only consider the symmetric linear dissociation. This is an interesting problem because of strong static correlation far from the equilibrium geometry [71], and CCSD yields energies that differ from the FCI energy by more than chemical accuracy. For the symmetric linear dissociation, the only free parameter x stands for the internuclear distance.

2. The BeH_2 insertion reaction is a difficult problem [21], [106], [107], 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 as functions of a parameter x 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. The reaction is depicted in figure 5.1.

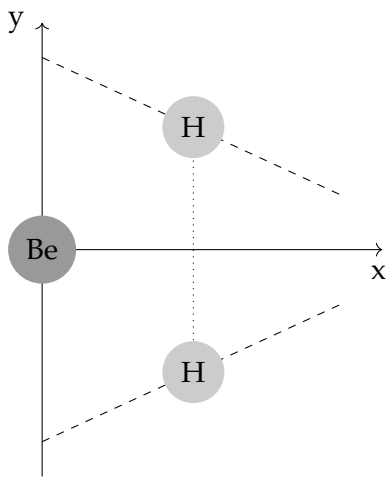


Figure 5.1: The BeH_2 insertion reaction. Beryllium is placed in the origin, the position of the hydrogen atoms as functions of a parameter x is given by $R_x = x$, $R_y = \pm(2.54 - 0.46x)$ (in Bohr).

   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 $|\Phi_2\rangle = |(1a1)^2(2a1)^2(3a1)^2\rangle$. The system is described relatively well 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 $|\Phi_1\rangle \rightarrow |\Phi_2\rangle$ as x is increased at $x \sim 2.85$ (the exact value depends on the basis set). The state $|\Phi_1\rangle$ remains important for $x > 3$. Figure 5.2 shows the FCI energies, the energies of $|\Phi_1\rangle$ and $|\Phi_2\rangle$, as well as the CCSD energies using either $|\Phi_1\rangle$ or $|\Phi_2\rangle$ as the reference determinant.

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

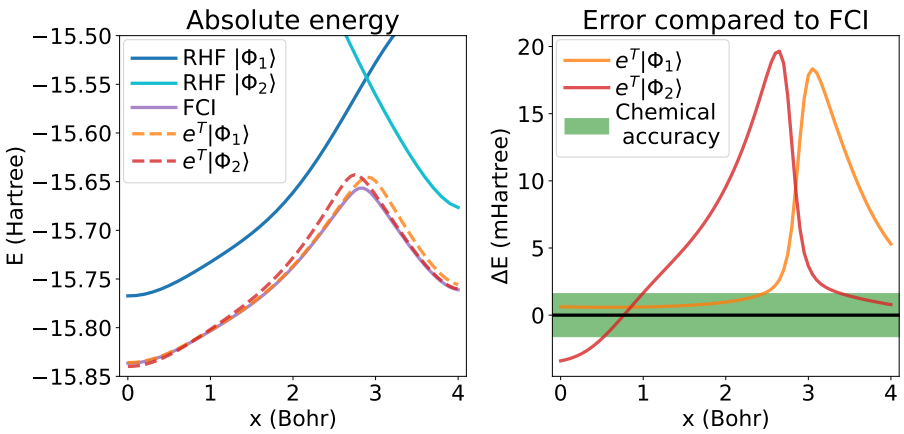


Figure 5.2: Energy of the BeH_2 insertion reaction using the cc-pVDZ basis set. Left: Energy of the two dominant Slater determinants $|\Phi_1\rangle$ or $|\Phi_2\rangle$ and the CCSD energy using either as reference, as well as FCI energy. Right: Deviation from FCI energy for the two CCSD wave functions.

struggle to find a solution. The only free parameter x stands for the internuclear distance.

4. The linear dissociation of HF is an interesting problem, going from a single-reference into a multi-reference state upon dissociation, with the CCSD energy well overshooting the FCI energy far from the equilibrium distance [21]. The only free parameter x stands for the internuclear distance.

Chapter 6

Multi-reference eigenvector continuation

6.1 Sampling along the potential energy surface

6.1.1 Hydrogen fluoride

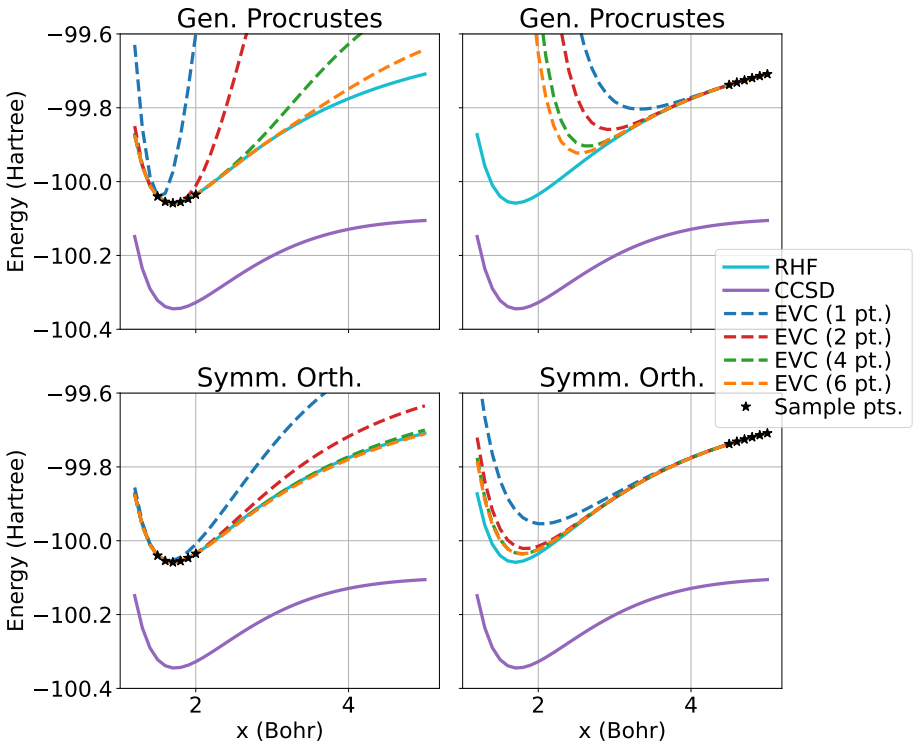


Figure 6.1: EVC potential energy surface using restricted Hartree-Fock wave functions for hydrogen fluoride in the cc-pVTZ basis with sample points close to equilibrium (left) or far from equilibrium (right) using generalized Procrustes orbitals (top) or symmetric orthonormalization (bottom). When not all sample points are included, only the leftmost sample points are included.

We applied the multi-reference EVC procedure described in section 4.4.1, p. 52 with restricted Hartree-Fock wave functions for the hydrogen fluoride molecule. This is shown in figure 6.1. As suggested in the methods section, symmetric orthonormalization of the sample MOs leads to a faster convergence to the RHF energy compared to generalized Procrustes

orbitals because the overlap to the RHF state is larger. Using Slater determinants as sample wave functions and sampling only close to the equilibrium geometry, the RHF potential energy surface is reproduced, but the energy is not improved below the RHF energy either. We calculated the overlap between the RHF-solution at a point x and the EVC solution at the same point $\left|\langle\Phi^{\mathrm{HF}}(x)|\Phi^{\mathrm{EVC}}(x)\rangle\right|^2$. The values for the lower right graph of figure 6.1 are shown in table 6.1, which shows that the EVC states have a very large overlap with the RHF state. The RHF state is essentially reproduced by EVC.

Table 6.1: Overlap between EVC and Hartree-Fock wave functions $\left|\langle\Phi^{HF}(x)|\Phi^{EVC}_{N_{EVC}}(x)\rangle\right|^2$ *for hydrogen fluoride, where* N_{EVC} *stands for the number of sample states. The same sample geometries as in the left lower graph of figure 6.1 were chosen. We do not write the dependence of the wave functions on x for brevity.*

x (Bohr)	$\left\|\langle\Phi^{\mathrm{HF}}\|\Phi^{\mathrm{EVC}}_1\rangle\right\|^2$	$\left\|\langle\Phi^{\mathrm{HF}}\|\Phi^{\mathrm{EVC}}_2\rangle\right\|^2$	$\left\|\langle\Phi^{\mathrm{HF}}\|\Phi^{\mathrm{EVC}}_4\rangle\right\|^2$	$\left\|\langle\Phi^{\mathrm{HF}}\|\Phi^{\mathrm{EVC}}_6\rangle\right\|^2$
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

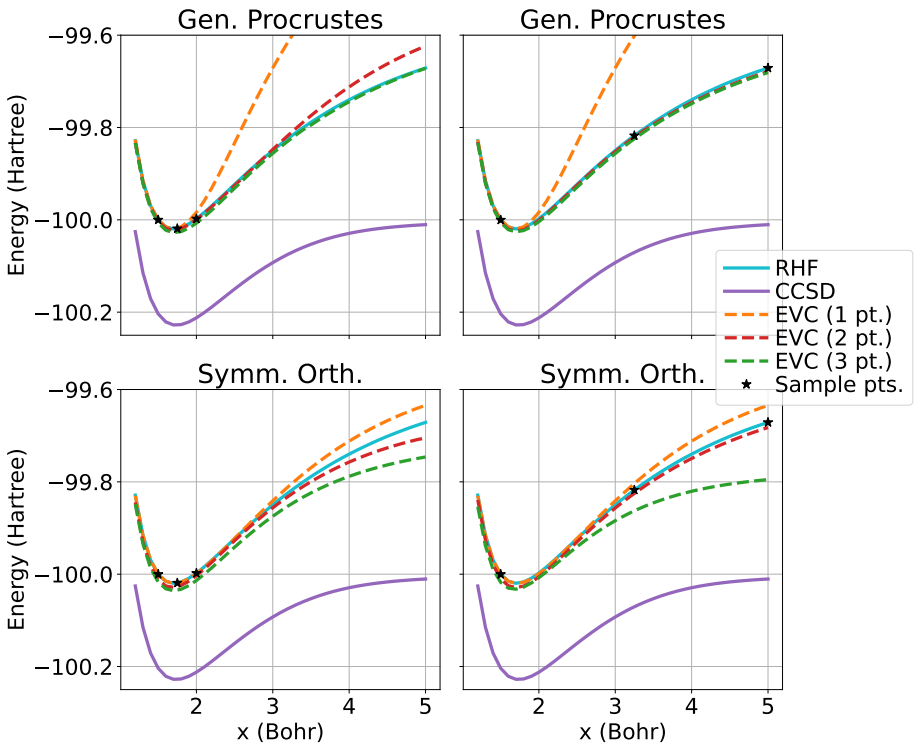


Figure 6.2: EVC energies for hydrogen fluoride in the cc-pVDZ basis with sample points close to equilibrium (left) or spread (right) using generalized Procrustes orbitals (top) or symmetric orthonormalization (bottom), including the HF state and all possible singly excited determinants from every sample geometry. When not all sample points are included, only the leftmost sample points are included.

We studied again hydrogen fluoride with multi-reference EVC, this time including all singly excited determinants for every sample state. This is shown in figure 6.2. Brillouin's theorem does not hold, as the sample states at different geometries no longer are HF states, hence singly excited determinants contribute. When sampling the ground state and the singly excited determinants at just one sample point, it is possible to almost

reproduce the RHF-PES, while the inclusion of additional sample points leads to improved results. However, naive addition of several sample points, as we did here, does not lead to substantial improvements in general.

6.1.2 Beryllium hydride insertion

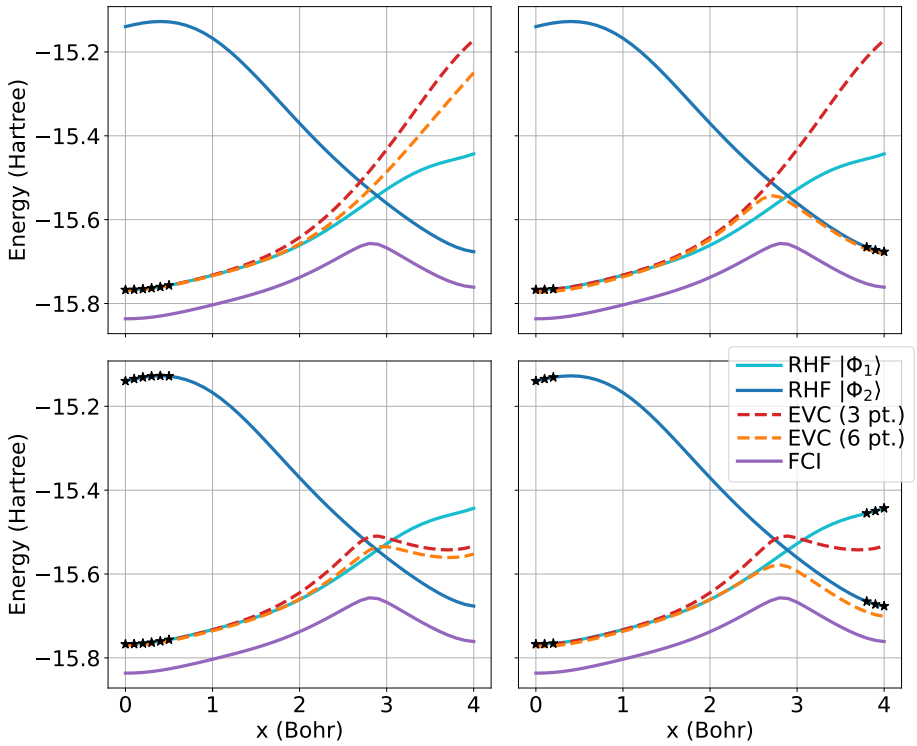


Figure 6.3: EVC energies for the BeH_2 insertion reaction in the cc-pVDZ basis (the largest basis we could do full CI in) with sample points on the left (left), left and right (right), and without (upper) or with (lower) inclusion of both states $|\Phi_1\rangle$ and $|\Phi_2\rangle$. In all cases, symmetric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

Figure 6.3 shows the potential energy surface of the BeH_2 insertion reaction using multi-reference EVC. When only including sample points close to $x = 0$ and only the state $|\Phi_1\rangle$, the energy stays close to that of $|\Phi_1\rangle$ for a part of the trajectory, but the energy is not able to "dip down" to that of state $|\Phi_2\rangle$. This changes when including sample points from both sides, which gives a qualitatively right PES without cusp, even though more sample points would have been necessary to get the energy as good as or better than the RHF energy. To improve upon the HF energy, it is necessary to include both states $|\Phi_1\rangle$ and $|\Phi_2\rangle$ into the sample space at each sample geometry, leading to an improved PES, as is visible in the lower right graph.

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 described in section 4.4.2, p. 52. The results are shown in figure 6.4. The costly part here is to create the sample Slater determinants, not creating and diagonalizing the EVC matrices. For HF, we 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 $|\Phi_1\rangle$

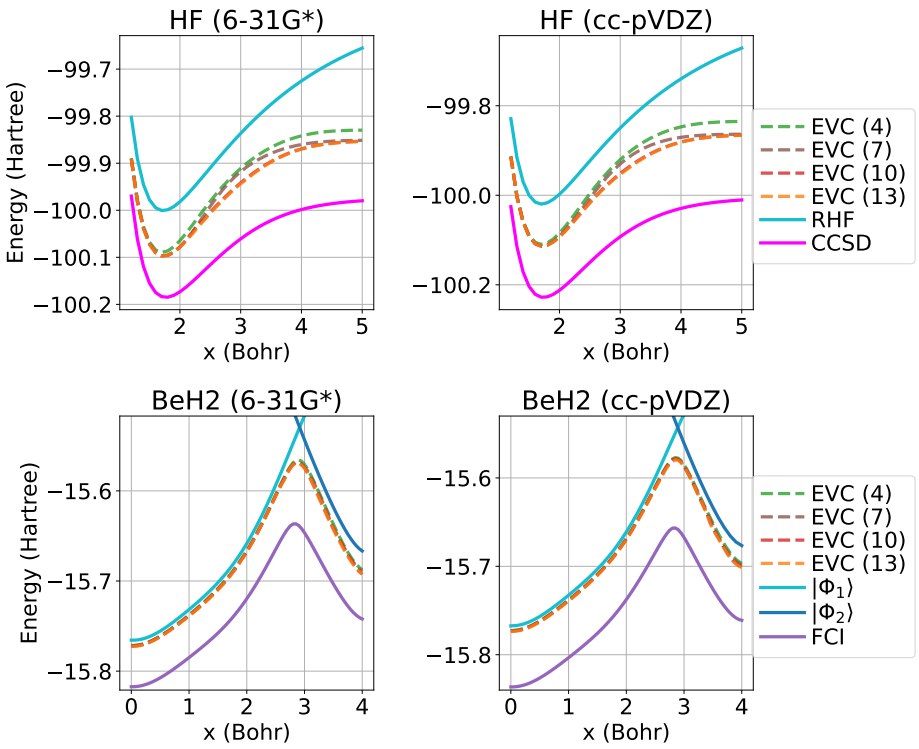


Figure 6.4: EVC energies for the BeH_2 insertion and HF with a tweaked 2-electron repulsion. The number in parentheses EVC(N) refers to the number of included HF states, picking the first N strength factors λ from the list $\lambda \in [1.1, 1.05, \ldots, 0.55, 0.5]$. For BeH_2, we included two determinants with different point group symmetry per λ, corresponding to $|\Phi_1\rangle$ and $|\Phi_2\rangle$.

and $|\Phi_2\rangle$ for each λ, even though it closes less in on the exact FCI solution. For HF, there is basically no difference between using 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 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

We have seen that few sample points chosen along a potential energy surface are sufficient to reproduce the RHF energy with EVC. This might be useful when there are troubles to converge the SCF equations: one can instead find the Slater determinant that maximizes the overlap to the EVC state. 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 a qualitatively correct potential energy surface. Inclusion of determinants from several principal configurations at each sample point leads to improved potential energy surfaces, but requires more knowledge about the system, as simply sampling close to the equilibrium symmetry becomes insufficient. The inclusion of singles, which is an expensive calculation scaling as $O(N^8)$ [31], 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 Hartree-Fock states obtained from different electron interaction strengths as basis at each geometry, the results vary a lot, in the sense that the PES improves greatly for some systems (HF), but less for other systems (BeH_2). The improvement when adding more sample states is seemingly unpredictable, with bigger improvements at some geometries than at others. Furthermore, there is no "natural" way to improve the solution any further without adding singles/doubles, which is prohibitively expensive. While those results are interesting, we do not consider this method to be generally applicable.

Chapter 7

Coupled Cluster eigenvector continuation

In this chapter, WF-CCEVC (cf. sec. 2.3.1, p. 27) and AMP-CCEVC cf. sec. 4.1, p. 45) are always used in conjuncture with CCSD wave functions. We simply write WF-CCEVC and AMP-CCEVC.

7.1 Results

7.1.1 Hydrogen fluoride

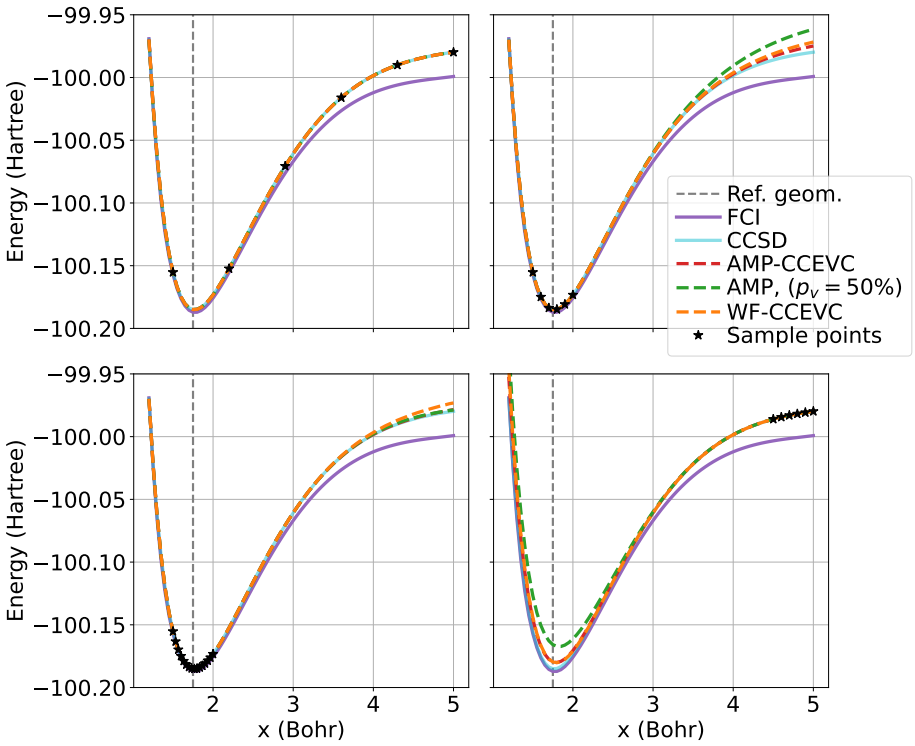


Figure 7.1: Energies of the HF molecule when dissociating using FCI, CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs were used. The 6-31G basis set was used, being the largest basis set for which we could calculate FCI energies. Top left: 6 evenly spread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

Figure 7.1 shows the energy of the HF molecule as a function of the internuclear distance using different methods and Procrustes orbitals in a small basis set (6-31G*). Both WF-CCEVC and AMP-CCEVC interpolate very 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 from 6 to 16 gives almost no improvement for WF-CCEVC, but it does for AMP-CCEVC and parameter-reduced AMP-CCEVC. For interpolation, as few as 6 sample points sampled along the whole PES manage to reproduce the full CCSD potential energy surface for AMP-CCEVC, WF-CCEVC and parameter-reduced AMP-CCEVC. In each case, the FCI energy remains out of reach. Parameter reduction in AMP-CCEVC leads to good interpolation results and also works well when many sample points are used, but it differs from full AMP-CCEVC for extrapolation when few sample points are used. This behaviour is expected and follows our discussion in the methods section.

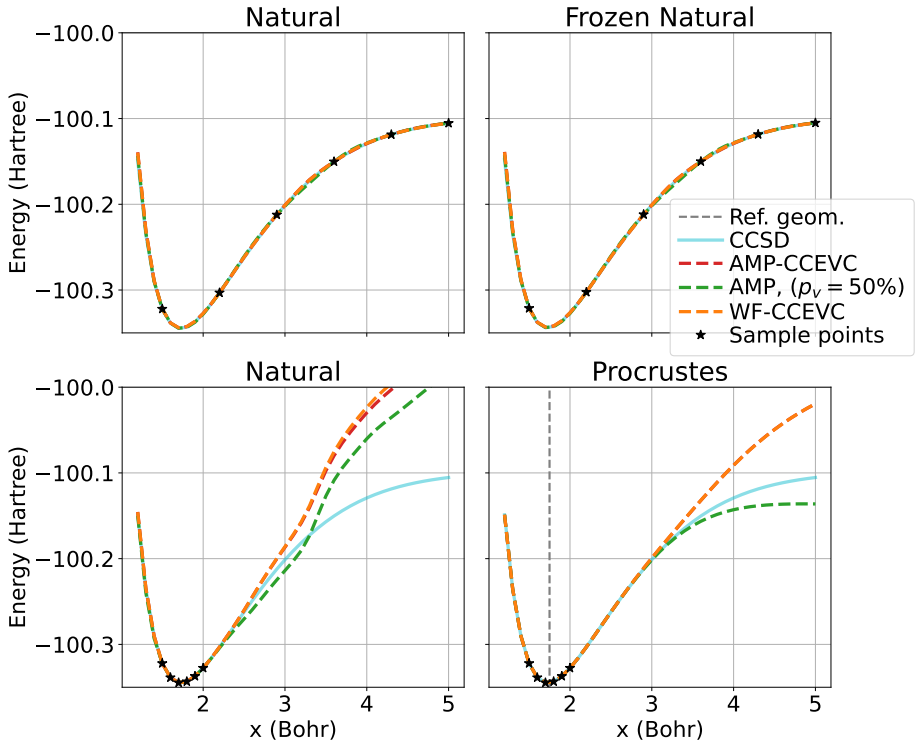


Figure 7.2: Energies of the HF molecule when dissociating using CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. The cc-pVTZ basis set was used. Procrustes MOs were used in the bottom right corner, while the other plots depict natural orbitals. The plot in the top right corner used frozen orbitals, freezing 5 of 44 with an occupation number below 10^{-4} everywhere, 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 hydrogen fluoride, this time using a much larger basis set (cc-pVTZ). In the bottom right corner, we used Procrustes orbitals, which shows that the method's extrapolation ability is not impaired much by larger basis sets, while we used natural orbitals otherwise. With the conservative cutoff of 10^{-4} for the natural occupation number, only 5 of 44 NOs are frozen, and the energy deviations are well below chemical accuracy. Following the discussion in Ref. [27], it is clear that a larger percentage of natural orbitals can be frozen in larger basis sets, and that using 39 NOs obtained from a larger basis set would give improved energies, which we did not consider here in order to be able to compare results to the non-cutoff case. For example, using the cc-pVQZ basis set with the same cutoff of 10^{-4}, only 40 out of 85 NOs would be used (aug-cc-pVTZ: 39 out of 69). Natural orbitals both interpolate and extrapolate worse than Procrustes orbitals, though

the interpolation results remain close to the CCSD energy. This matches with the analysis carried out in section 4.6.1, p. 55, stating that avoided crossings between natural orbitals make them less suited for EVC than Procrustes orbitals.

7.1.2 Nitrogen

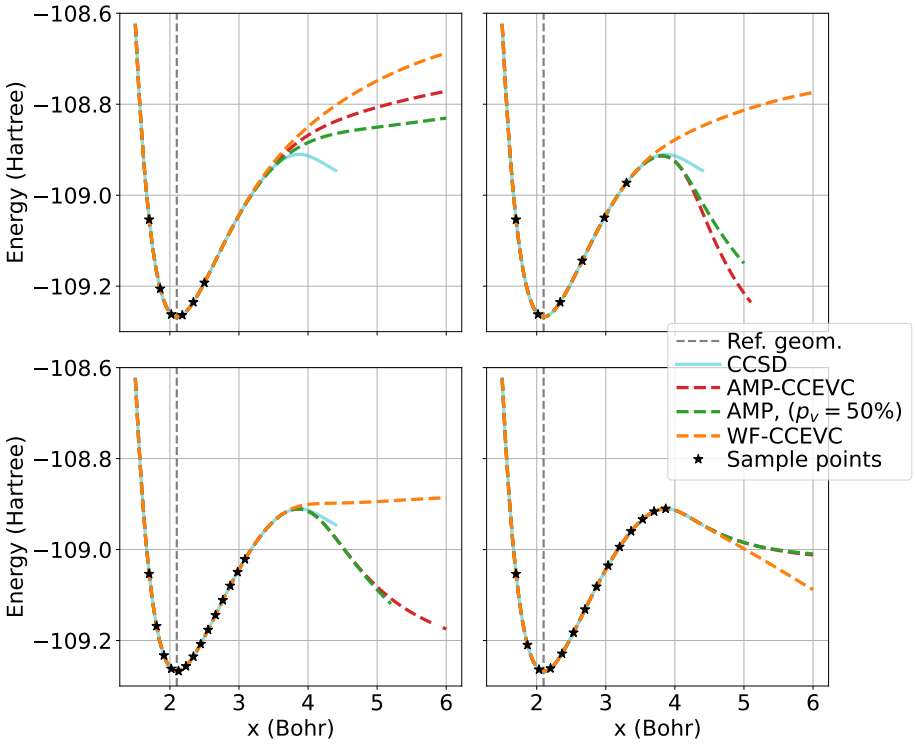


Figure 7.3: Energies of the N_2 molecule when dissociating using CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs and the cc-pVDZ basis set were used. Energies at geometries where no solution to the amplitude equations was found, are not included.

Figure 7.3 shows the energy of the N_2 molecule as a function of internuclear distance using different sets of sample points. The N_2 molecule is an interesting molecule to study, as the CCSD energy *decreases* as the molecule dissociates, which is physically incorrect. Even inclusion of triple excitations in the cluster operator (CCSDT) leads to the same disastrous breakdown [108]. Eigenvector continuation can avoid this breakdown.

AMP-CCEVC extrapolates the CCSD energy well beyond the sample range and has the nonphysical maximum at $x \sim 3.9$ Bohr as well, unless only few sample points that lie far away from the nonphysical region are included. We consider this to be a coincidence. Parameter-reduced AMP-CCEVC gives similar interpolation, but different extrapolation results, as was the case for HF.

The results get more interesting for WF-CCEVC. The upper left and lower right subplots show qualitatively wrong energies, while the other two are qualitatively correct, with the lower left one almost becoming horizontal, which is the correct FCI behaviour [108]. That shows the importance of the sample 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 at at geometries where CCSD is qualitatively wrong makes the WF-CCEVC energy take over the wrong behaviour, as WF-CCEVC is not a variational method. However, a careful choice of the sample points can give proper results. Inclusion of several good sample points improves the WF-CCEVC energy.

7.1.3 Beryllium hydride stretch

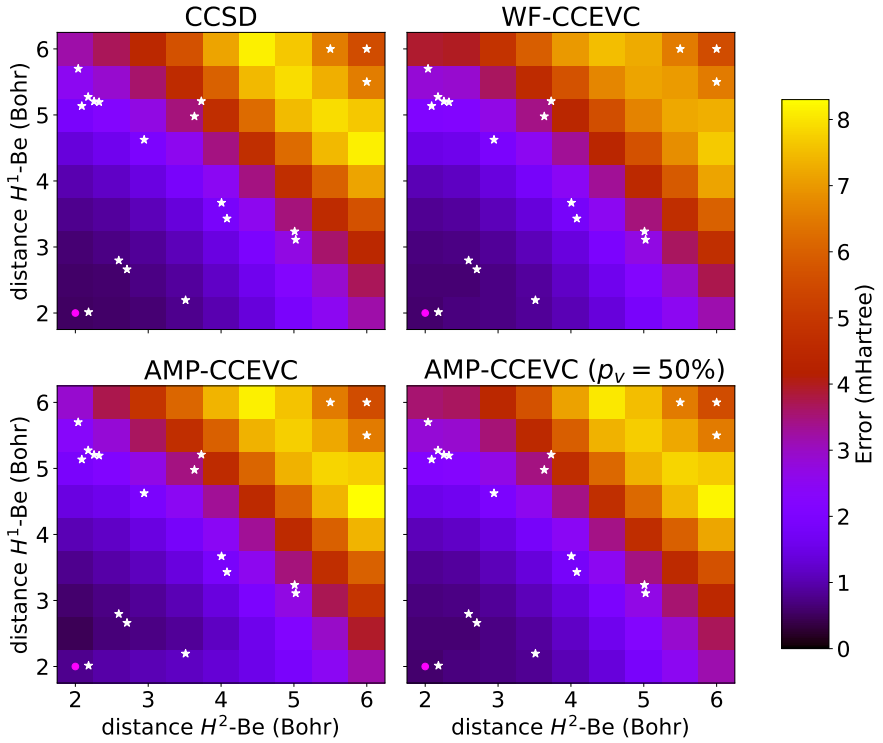


Figure 7.4: Absolute energy difference compared to full CI solution of the asymmetric linear stretch of the BeH_2 molecule using CCSD, WF-CCEVC, AMP-CCEVC and AMP-CCEVC with only 50% of the virtual orbitals. Procrustes MOs were used, with the reference geometry marked in magenta. The cc-pVDZ basis set was used. The 16 sample points are marked with white stars. 13 were chosen randomly in the "simple" region, in addition to 3 that were purposely put in the top right corner.

Figure 7.4 shows the deviation from the FCI energy for the BeH_2 stretch at different geometries, varying 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 on both sides (e.g. the violet to orange region, but not the yellow region), the WF-CCEVC gives decreased errors in the difficult region, such as at $R_1 = 5$, $R_2 = 5.5$ Bohr. This is not the case for the AMP-CCEVC method and the parameter-reduced AMP-CCEVC, which essentially recover the potential energy surface of CCSD.

7.1.4 Beryllium hydride insertion

When considering the beryllium hydride insertion using EVC, special care needs to be taken that the same reference state (either $|\Phi_1\rangle$ or $|\Phi_2\rangle$) is used along the whole PES, as both AMP-CCEVC and WF-CCEVC require the reference determinant to remain the same along the whole region of interest.[1] We only present results for WF-CCEVC.

Figure 7.5 depicts the WF-CCEVC energy using different sample points and either $|\Phi_1\rangle$ or $|\Phi_2\rangle$ as the reference state. In the upper figures, we see that picking sample points along the whole PES essentially recovers the PES. 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

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

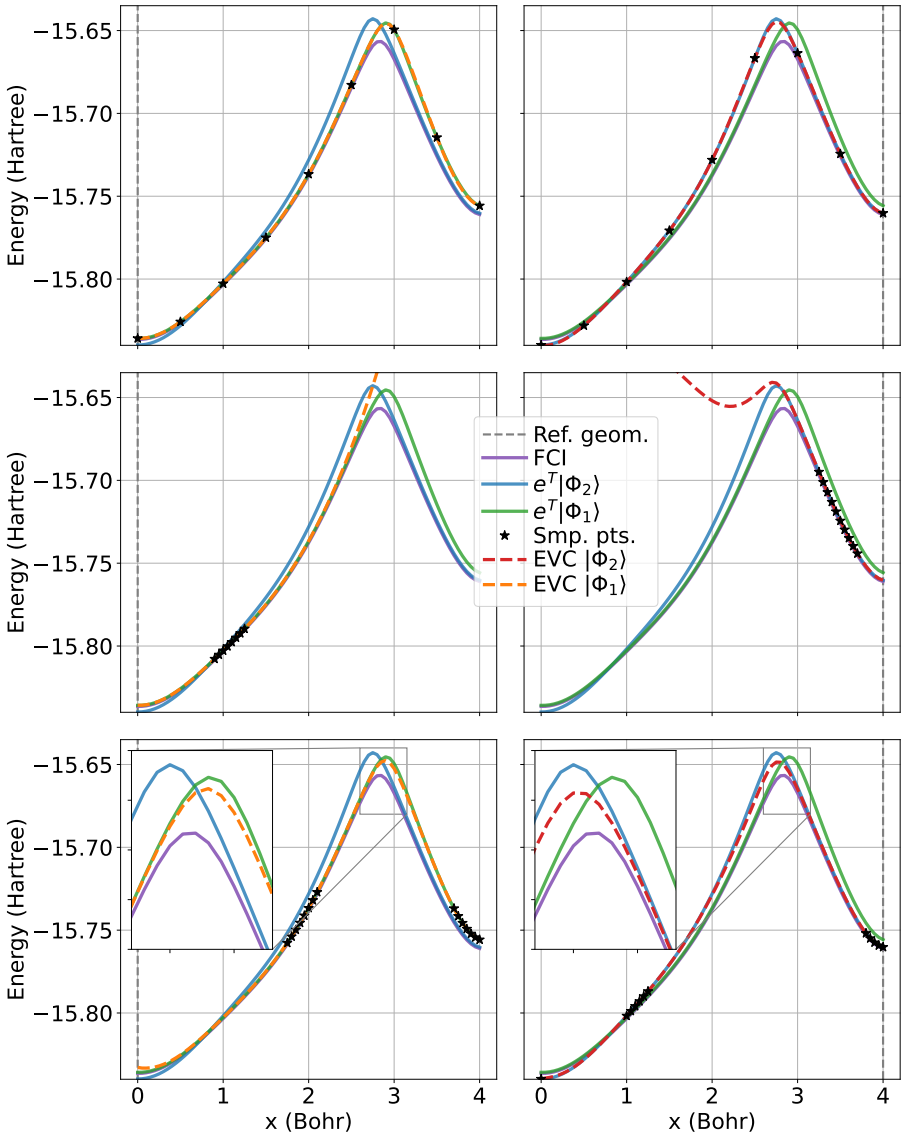


Figure 7.5: WF-CCEVC, CCSD and FCI potential energy curve for the BeH_2 insertion reaction in the cc-pVDZ basis set using Procrustes orbitals. In the legend, EVC stands for WF-CCEVC. Left: Using $|\Phi_1\rangle$ as reference for CCSD and WF-CCEVC. Right: Using $|\Phi_2\rangle$ as reference for CCSD and WF-CCEVC. Different sample geometries were used.

obtain even approximately correct energies. In the bottom, we see that when sampling on both sides, even though $|\Phi_1\rangle$, $|\Phi_2\rangle$ are not good reference states along the whole PES, the PES near the transition point is improved. While chemical accuracy is not obtained, which we believe is because there is no chemical accuracy at all the sample points themselves and the two viable sampling regions are far from another, the WF-CCEVC procedure leads to improved energies near the transition point.

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 sample points from the good region, while the energy is not improved at all for hydrogen fluoride compared to CCSD. WF-CCEVC improves the energy for the BeH_2 insertion when sampling on both "sides". To understand why, it is of value to consider the singular values of the overlap matrix (which is the same at all geometries) 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. Observe that it is not necessary for the sum of singular values to be identical to the number of sample points.

1. For HF in the cc-pVTZ basis set, sampling at 6 evenly spaced points $x \in [1.5, 4]$ (Bohr) gives one singular value above 1 ($\sigma_1 = 5.84, \sigma_2 = 0.17, \sigma_3 = 0.0032$). Adding the points $4.5, 5.0$ does not change that ($\sigma_1 = 7.67, \sigma_2 = 0.41, \sigma_3 = 0.011$).

2. For N_2 in the cc-pVDZ basis set, sampling at 5 evenly distributed points $x \in [1.5, 3.5]$ gives one singular value above 1 ($\sigma_1 = 4.75, \sigma_2 = 0.32$). Adding the points $4.0, 4.5$ increases that number to two ($\sigma_1 = 6.32, \sigma_2 = 2.85, \sigma_3 = 0.068$).

3. For BeH_2 in the cc-pVDZ basis set (cf. fig. 7.5), using configuration $|\Phi_1\rangle$, sampling along the whole PES (top graph) gives two out of 9 singular values above 1 ($\sigma_1 = 8.34, \sigma_2 = 2.75, \sigma_3 = 0.18$). Sampling at geometries on the left (middle graph) gives one out of eight singular values above 1 ($\sigma_1 = 7.9991, \sigma_2 = 0.00093$). Sampling on both sides (bottom graph) gives 2 out of 13 singular values above one ($\sigma_1 = 13.25, \sigma_2 = 4.14, \sigma_3 = 0.033$).

When two singular values are large, there are two important, qualitatively different states. For N_2, there arises a second, incorrect state when including the two extra sample points, leading to wrong results. This state is not included when only sampling in the correct region. When using sample points only from the correct region, Eigenvector continuation hence finds a solution which is a linear combination of configurations from the correct region, even in the wrong region. For HF, there is only one qualitatively important state, and the second most important state is only quantitatively important. For BeH_2, 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, as it is qualitatively correct.

7.1.6 Water around equilibrium geometry

Having established that EVC works well for 1-dimensional and 2-dimensional problems when a sufficient number of sample points is used, we now consider the stochastic convergence of the CCEVC energies compared to the CCSD energy for a three-dimensional problem as a function of the number of sample points. The system considered is water around its equilibrium geometry. Free parameters are the distance between the oxygen and the hydrogen atoms r_1, r_2 as well as the angle α. Both test and sample geometries were sampled 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 deviation from the CCSD energy as a function of the number of sample points for WF-CCEVC and AMP-CCEVC and parameter-reduced AMP-CCEVC. A monotone convergence of the mean and all quartiles for AMP-CCEVC and WF-CCEVC can be observed. In all four scenarios, 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 few as 9 sample geometries are sufficient to have the means converge to chemical accuracy for full AMP-CCEVC and WF-CCEVC. Using parameter-reduced AMP-CCEVC, convergence is slower and noisy, both for the mean and the median. This is not surprising either, as the method is an approximation, and we do not expect that adding further sample geometries necessarily improves energies. We nevertheless observe the same trend and convergence of the mean to just above ($p_v = 20\%$) and below ($p_v = 50\%$) chemical accuracy, justifying the use of the parameter reduction.

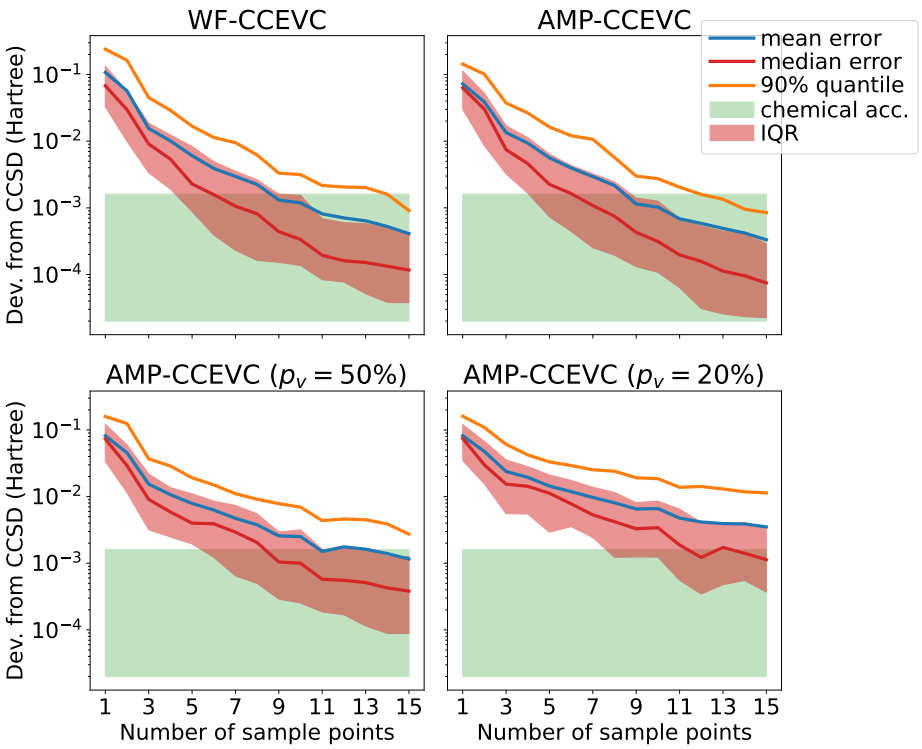


Figure 7.6: Absolute deviation from CCSD energy (mean and median) as well as IQR (range of 2nd and 3rd quartile) and 90% quartile for WF-CCEVC and AMP-CCEVC ($p_v \in \{100\%, 50\%, 20\%\}$) as function of the number of sample points for water. Procrustes orbitals were used, with the reference geometry at equilibrium ($r_1 = r_2 = 1.81$ Bohr, $\alpha = 104.5°$), where r_1, r_2 are the distances between the oxygen and the hydrogen molecules and α the angle between the hydrogen molecules. Both the sample geometries as well as the test geometries were sampled uniformly for $r_1, r_2 \in [1.3, 4.0]$ (Bohr) and $\alpha \in [99.5°, 109.5°]$. The aug-cc-pVDZ basis set was used. Statistics were calculated using 16 random test geometries, and the numerical experiment was repeated 5 times with new sample and test geometries.

7.2 Time usage of AMP-CCEVC

The time-consuming step in AMP-CCEVC is to calculate the error vector $\mathrm{e}(c_i)$ $i = 1, \ldots, L$ at each iteration. This is the value which we want to be equal to zero. In the parameter-reduced version, we approximate $\mathrm{e}(c_i)$, and the cost is reduced. This might impact the accuracy of the amplitude calculations. In table 7.1, we compare 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, where also some occupied orbitals were neglected. We considered a pure interpolation problem, having established that these methods work worse for extrapolation.

Table 7.1: Number of iterations for the quasi-Newton method to reach a projection error $e(c_i)$ below 10^{-5} for each i, total time use per geometry and maximal projection error $\max_\mu \left|\langle \mu | \bar{H} | \Phi^{HF} \rangle\right|$ 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, the amplitudes of the sample points closest to the geometry in question were used. Procrustes orbitals were used with reference geometry $x_{ref.} = 2$ Bohr. As sample geometries, we used 10 evenly spaced samples for $x \in [1.5, 6.0]$ Bohr. Values were obtained for 36 different geometries evenly spaced in the range $x \in [2.5, 5]$ Bohr. `opt_einsum` [102] was used for tensor operations.

Percentage included $M_{\text{virt.}}$	100	80	40	20	10
Num. iterations per geometry	7 ± 4	7 ± 4	7 ± 4	6 ± 3	7 ± 4
Total time per geometry (s)	3.3 ± 1.7	2.8 ± 1.3	1.8 ± 0.8	1.2 ± 0.6	0.7 ± 0.4
Absolute max. projection error	0.0012	0.0012	0.0012	0.0016	0.0016

Table 7.2: Number of iterations for the quasi-Newton method to reach a projection error $e(c_i)$ below 10^{-5} for each i, total time use per geometry and maximal projection error $\max_\mu \left|\langle \mu | \bar{H} | \Phi^{HF} \rangle\right|$ as function of the percentage of included virtual orbitals (rounded) for the stretch of HF in the cc-pVTZ basis set. As initial guess, the amplitudes of the sample points closest to the geometry in question were used. Procrustes orbitals were used with reference geometry $x_{ref.} = 2$ Bohr. As sample geometries, we used 9 evenly spaced samples for $x \in [1.0, 5.0]$ Bohr. Values were obtained for 31 different geometries evenly spaced in the range $x \in [1.5, 4.5]$ Bohr. `opt_einsum` [102] was used for tensor operations.

Percentage included $M_{\text{virt.}}/M_{\text{occ.}}$	100/100	80/80	40/60	20/60	10/60
Num. iterations per geometry	5 ± 2	5 ± 2	5 ± 3	5 ± 3	7 ± 5
Total time per geometry (s)	2.6 ± 1.1	2.2 ± 0.9	1.6 ± 0.8	1.1 ± 0.5	0.74 ± 0.5
Absolute max. projection error	0.0090	0.0090	0.014	0.020	0.019

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 the parameter reduction. The error only increased when using only 20% or less of the virtual orbitals. For HF, there is a slight increase in the number of iterations as more orbitals are neglected, but the maximal projection error does not increase much. In sufficiently large basis sets, inclusion of as little as 10% of the virtual orbitals thus still yields good results and approximate the exact CCSD amplitudes closely. 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 approximating 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 parameter reduction works well and gives almost the same energies as full AMP-CCEVC 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 in a high-performance programming language. 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, as inclusion of extra sample points will likely recover the erroneous behaviour. We found that the energy obtained using full AMP-CCEVC converges monotonously to CCSD, with parameter-reduced AMP-CCEVC getting close to the CCSD energy. Finally, 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 sample points, as shown for N_2 and BeH_2. In particular, it was 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 knowledge of the chemical problem at hand is required and sample points need to be selected carefully whenever an improvement of the CCSD energy is desired. Analyzing the singular values of the overlap matrix tells us when a qualitative change in the wave function occurs and can be used as an indicator whether a sample point is suitable. Furthermore, as the method is not variational, naive inclusion of wrong sample points can increase the error and give rise to nonphysical potential energy surfaces, as we have observed for nitrogen. An increase in the number of sample points gives an increased extrapolation range, but also leads to an increased computational cost. The speedup of WF-CCEVC for calculating potential energy surfaces is orders of magnitudes lower than the one attained in Ref. [4], where WF-CCEVC calculations took one hour, while full CCSD calculations were estimated to have taken 20 years. Because their Hamiltonian was of the form

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

where N_α is a small number, they only had to similarity-transform each Hamiltonian $\hat{H}_i$ once per sample cluster operator. The molecular Hamiltonian needs however to be similarity-transformed at every geometry, thus the $O(M^6)$ scaling is unavoidable. In addition, the molecular Hamiltonian changes in a complicated way when nuclei are moved, and due to this complication, we did not obtain the same quality of extrapolation that can be possible for simpler perturbations.

Just like AMP-CCEVC, WF-CCEVC shows that the CCSD wave function in a large region can be expressed with very few parameters (compared to $O(M^2N^2)$), in the sense that chemical accuracy is obtained with few included sample wave functions. Both WF-CCEVC and the parameter-reduced AMP-CCEVC make use of this and 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 they are relatively simple to implement when the amplitudes and the Λ-amplitudes are available, they are easy-to-implement methods with a cost no higher than CCSD, and both can give substantial improvements when applied correctly.

Chapter 8

Quantum computer eigenvector continuation

8.1 Results

8.1.1 Single-reference EVC with Procrustes orbitals: BeH_2 stretch

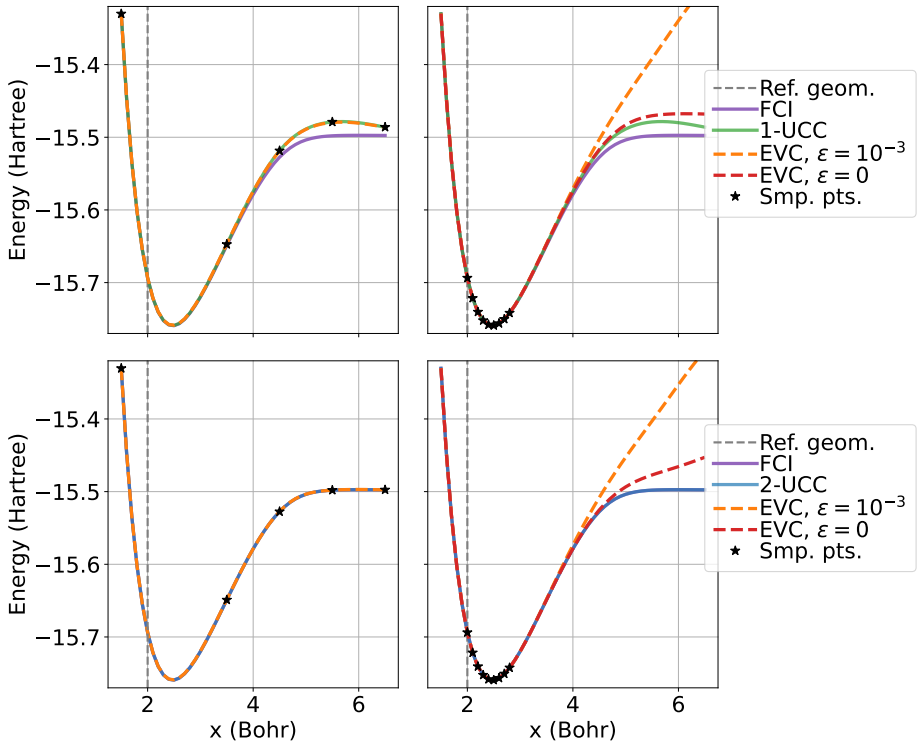


Figure 8.1: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes orbitals on a quantum computer. The STO-6G basis set was used and the two core electrons were frozen. 5 additional qubits were tapered off. We used either a 1-UCCSD (left) or a 2-UCCSD (right) ansatz (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. Two values for the cutoff for canonical orthogonalization ϵ were considered.

The EVC potential energy curve using Procrustes orbitals on a quantum computer (cf. sec. 4.7.1, p. 56) for the symmetric linear dissociation of the BeH_2 molecule is shown in figure 8.1. We used either a 1-UCCSD (left) or a 2-UCCSD (right) ansatz to obtain the sample wave functions. While UCCSD methods are not the most resource- or time-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 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 $\mathbf{H}^{\mathrm{EVC}}$ and $\mathbf{S}^{\mathrm{EVC}}$ were calculated, not sampled, and are hence numerically exact. In order to solve the generalized eigenvalue problem $\mathbf{H}^{\mathrm{EVC}}\mathbf{C} = \mathbf{S}^{\mathrm{EVC}}\mathbf{C}\boldsymbol{\epsilon}$, we used canonical orthogonalization as described in sec. 5.1.2, discarding eigenvectors of $\mathbf{S}^{\mathrm{EVC}}$ with eigenvalues 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 [89].

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 the 1-UCCSD wave function at dissociation). Because we cannot test this for larger basis sets due to the exponential scaling of computational cost as a function of number of simulated qubits, we cannot confirm that this is the case for larger basis sets, too, but since we observed excellent interpolation results for WF-CCEVC with few sample points using large basis sets, we believe this to be the case for EVC on a quantum computer too. We see how a cutoff of $\epsilon = 10^{-3}$ for canonical orthogonalization suffices. Indeed, both for 1-UCCSD and 2-UCCSD, only a single eigenvector of the overlap matrix is discarded. For close-lying sample points, choosing a large cutoff value ϵ 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-UCCSD extrapolation being qualitatively correct. Finally, although 2-UCCSD and 1-UCCSD both are chemically accurate in the interpolation regime, and the wave functions thus very similar, the 2-UCCSD extrapolation follows the 2-UCCSD trajectory, while the 1-UCCSD extrapolation follows the 1-UCCSD trajectory. The 1-UCCSD wave function is qualitatively correct at dissociation, while the 2-UCCSD one is not, which we believe to be coincidental.

8.1.2 Sampling requirements

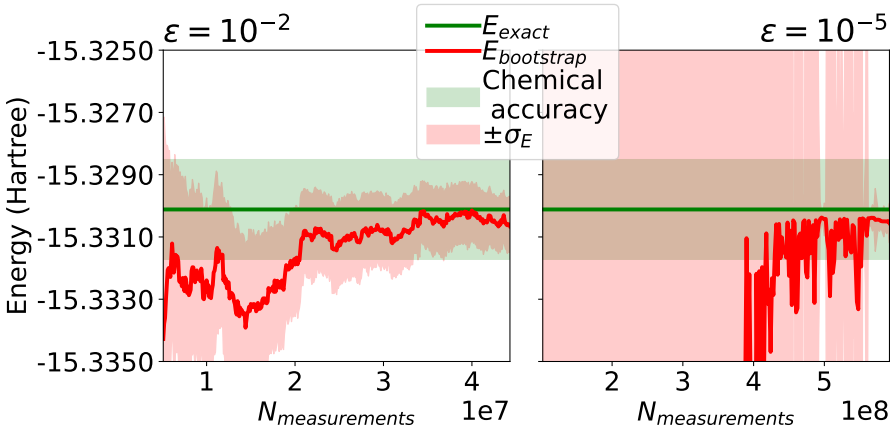


Figure 8.2: Convergence of the sample expectation value $E_{bootstrap}$ at $x = 1.5$ (Bohr) for the symmetric stretch BeH_2 in the STO-6G basis set with the core electrons frozen as the number of measurements is increased for two values of ϵ. Measuring H_{ij}^{EVC} once is considered as one measurement. We used $N_B = 100$ bootstrapping samples. The number of measurements of the matrix element assumed to decrease the standard deviation σ_E the most was increased by 10^5 (left) and $2 \cdot 10^6$ (right) per iteration. Each matrix element was initially measured $2 \cdot 10^5$ (left) and $4 \cdot 10^6$ (right) times. We used a cutoff $\epsilon = 10^{-2}$ (left) and $\epsilon = 10^{-5}$ (right) in the canonical orthogonalization procedure. 2-UCCSD wave functions and Procrustes orbitals with reference geometry $x = 2$ Bohr were used, using sample geometries $x_i \in \{1.5, 3.5, 4.5, 5.5, 6.5\}$(Bohr).

As discussed in the theory and method sections, the matrices $\mathbf{H}^{\mathrm{EVC}}$ and $\mathbf{S}^{\mathrm{EVC}}$ need to be sampled on real quantum computers. We address the sampling requirements to obtain chemical accuracy for different cutoff parameters ϵ for EVC for the BeH_2 stretch using 2-UCCSD sample wave functions and Procrustes orbitals. We used the adaptive scheme discussed in sec. 3.11, p. 42. As a convergence criteria, we demanded that $2\sigma_E$ is less than chemical accuracy upon convergence and that energy fluctuations between consecutive iterations are below chemical accuracy for 20 consecutive increases. This is shown in figure 8.2.

The choice of the cutoff ϵ is paramount. 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 $\mathbf{S}^{\mathrm{EVC}}$ is sampled very accurately: For $\epsilon = 10^{-2}$, the standard deviation σ_E gets gradually smaller as the number of measurements is increased. For $\epsilon = 10^{-5}$, the standard deviation σ_E is extremely large in the beginning, and it is erratic as the number of measurements is increased. Similarly, the expectation energy $E_{bootstrap}$ is initially completely wrong. Figure 8.3 shows that the expectation values of Pauli matrices sampled 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 stated in the methods section. We observe some fluctuations which are of the magnitude of chemical accuracy, which may require further measurements for other systems than the symmetric BeH_2 stretch.

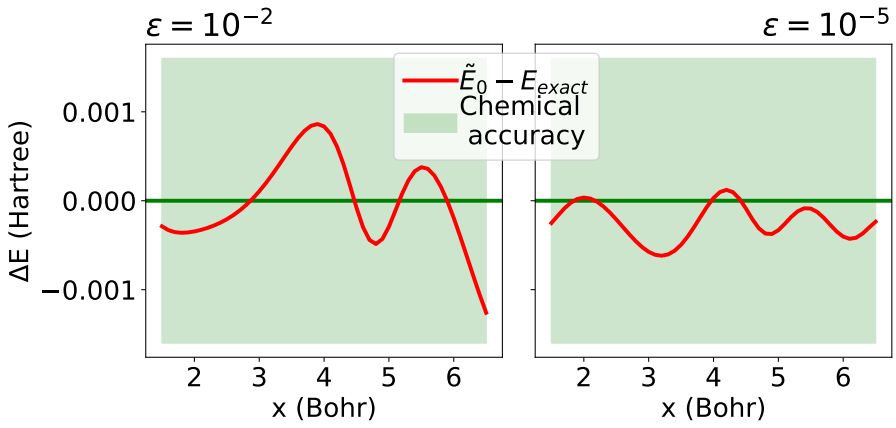


Figure 8.3: Difference between the sampled expectation value $\tilde{E}_0(x)$ (where no bootstrapping was used) and the exact energy $E_{exact}(x) = \langle \hat{H}(x) \rangle$ along the PES for the symmetric stretch of BeH_2. The Pauli matrix expectation values were sampled at $x = 1.5$ Bohr and are identical to the ones in fig. 8.2 after convergence.

8.1.3 Smoothing noisy potential energy surfaces

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. Here, a qualitatively correct PES

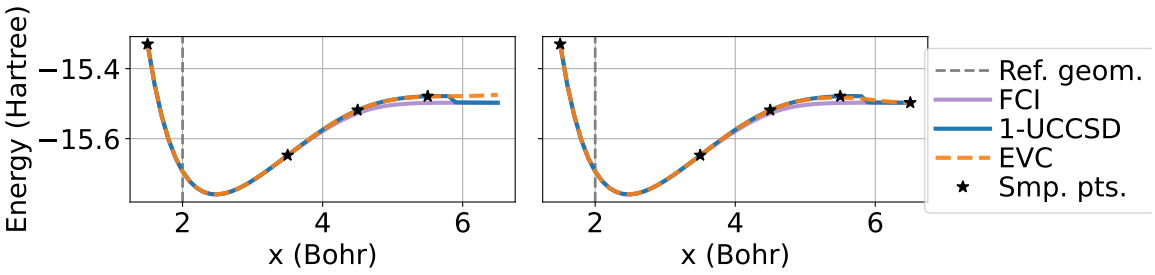


Figure 8.4: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes orbitals. The STO-6G basis set was used and the two core electrons were frozen. The Jordan-Wigner mapping without qubit tapering was used. We used a 1-UCCSD reference. Observe the nonphysical "jump" at $x = 5.8$ Bohr. Left: Including only sample points from one continuous region. Right: Including sample points from both regions. No overlap matrix eigenvectors were discarded.

is obtained by ignoring the region after $x = 5.8$ Bohr, even though adding several sample points $x_i \in [5.8, 6.5]$ (Bohr) might remove the nonphysical "bump" in the right plot.

8.1.4 Comparison between generalized Procrustes and regular Procrustes orbitals

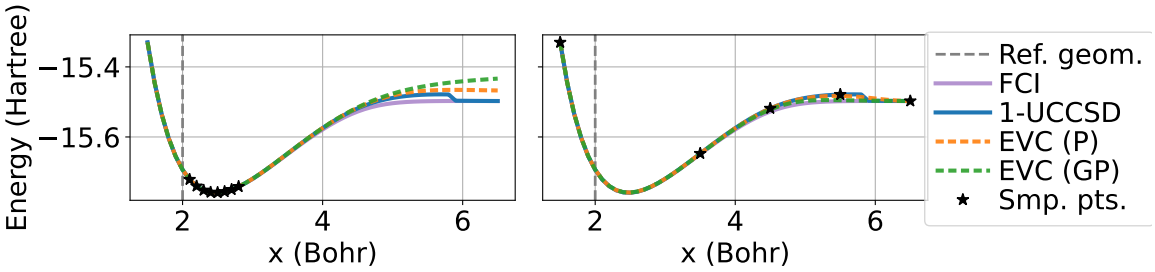


Figure 8.5: Potential energy surface of the symmetric dissociation of BeH_2 using EVC with Procrustes (P) or generalized Procrustes (GP) orbitals. The STO-6G basis set was used and the two core electrons were frozen. The Jordan-Wigner mapping without qubit tapering was used. We used a 1-UCCSD reference. Observe the nonphysical "jump" at $x = 5.8$ Bohr. Left: Dense sampling in equilibrium region. Right: Sampling along the whole PES, including beyond the jump. No overlap matrix eigenvectors were discarded.

We tested whether the use of generalized Procrustes orbitals as described in section 4.7.2, p. 57, can give an advantage over the use of regular Procrustes orbitals. This requires that the Jordan-Wigner mapping is used. In figure 8.5, we compare the extrapolation and interpolation of generalized Procrustes orbitals with regular Procrustes orbitals. Which method works better is inconclusive. Sampling only near the equilibrium region, Procrustes orbitals work better, while along the whole PES, generalized Procrustes orbitals work better. There is an explanation why using Procrustes orbitals works better for extrapolation.

Using Procrustes orbitals, the HF state at geometry $\vec{R}_1$ is associated with the HF state at geometry $\vec{R}_2$. At every geometry $\vec{R}$, the HF state has a qubit representation (possibly with reordered indices) $|\Phi^{\mathrm{HF}}\rangle = |1_1 \dots 1_N 0_{N+1} \dots 0_Q\rangle$. With generalized Procrustes orbitals, this is no longer the case, as occupied and unoccupied orbitals are mixed. When a state is well approximated by the HF state, using generalized Procrustes orbitals will lead to a state that changes more 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 themselves is minimized. Figure 8.6 contains the overlap matrices for the right plot of figure 8.5, which confirms this, as the generalized Procrustes orbital overlap matrix has smaller off-diagonal elements compared to

the Procrustes one. It should also be noted that the overlap matrix for generalized Procrustes orbitals has larger eigenvalues than the one for Procrustes orbitals, which follows from the smaller off-diagonal elements - from that point of view, generalized 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 P stands for Procrustes and GP for generalized Procrustes.

8.1.5 BeH_2 insertion reaction

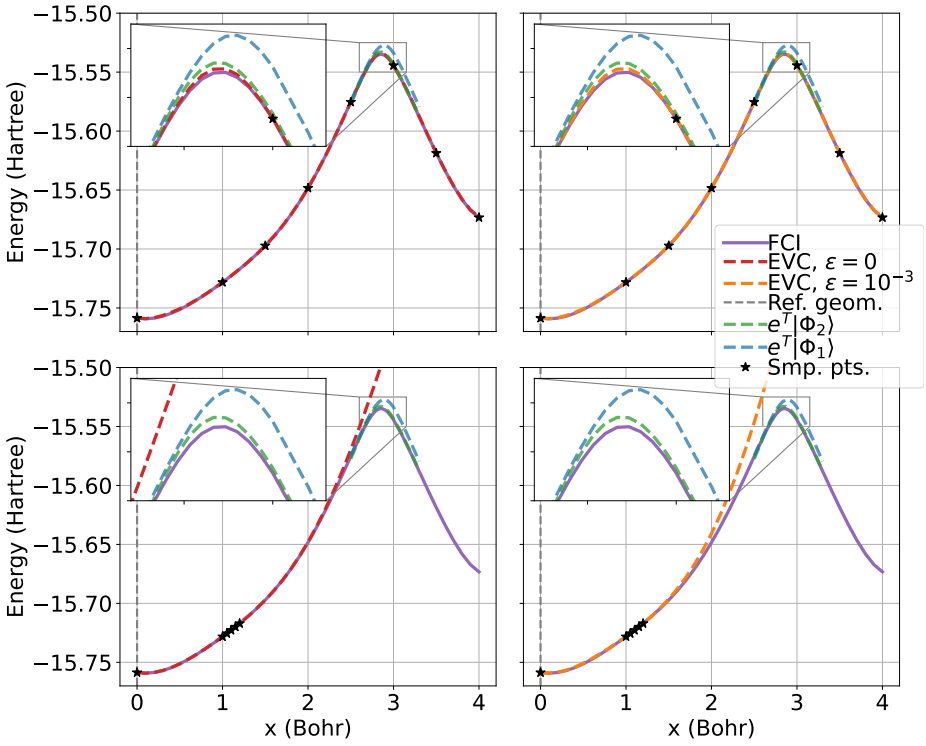


Figure 8.7: Potential energy curve for the BeH_2 insertion reaction using EVC with generalized Procrustes orbitals, with the reference geometry $x = 0$. FCI was used to obtain sample states. The STO-6G basis set was used. Two core electrons were frozen using `Qiskit's ActiveSpaceTransformer` [109]. CCSD solutions in the "difficult" region are shown for comparison. For $x < 3$ (Bohr) use $|\Phi_1\rangle$ as sample reference determinant, for $x \geq 3$ (Bohr), $|\Phi_2\rangle$ is used as sample reference determinant. Left: using no cutoff $\epsilon = 0$. Right: Using a large cutoff $\epsilon = 10^{-3}$.

To see how well EVC works with generalized Procrustes orbitals for systems where the Hartree-Fock ground state changes discontinuously, we considered again the BeH_2 insertion reaction. We used FCI wave functions as sample wave functions, as 1-UCCSD and 2-UCCSD converge to the triplet state at those geometries where the triplet state has a lower energy, and we did not manage to find an ansatz in `qiskit` that did not converge to the triplet state.

This is however not a problem, as existing methods, such as (U)CCSD, give results well below chemical accuracy at the sample geometries considered, and it is hence likely that using any of those ansätze instead of FCI would yield similar results. The resulting PESs are shown in figure 8.7. 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. When the sample points are spread "far" apart, 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 densely. 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 sample points on a quantum computer, as increasing the cutoff ϵ 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 state, 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 Hartree-Fock ground state changes analytically along the PES and is chosen as the reference state, 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 orders magnitude 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. [70]. Careful selection of both sample states and the cutoff ϵ of the eigenvalues of the overlap matrix $\mathbf{S}^{\text{EVC}}$ needs to be performed. This is reflected in the fact that the number of measurements increases with the number of samples L and decrease with the cutoff 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 model eigenvalue problem shown in figure 2.1, p. 25. The overlap matrix for 5 sample points when sampling densely (left plot of aforementioned figure) is shown in figure 8.8 below. It is almost singular, with three eigenvalues below 10^{-5}. Using $\epsilon = 10^{-5}$ as cutoff (and we observed previously that this already corresponds to large sampling requirements), the quality of the EVC eigenvalues deteriorates, which is shown in figure A.1 in the appendix, where adding additional sample points makes no qualitative difference and the correct eigenvalues differ qualitatively and strongly from the EVC eigenvalues. This shows that EVC for extrapolation is not applicable on quantum computers, because sampling the overlap matrix becomes infeasible.

We 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 only need to be sampled at one single geometry. Furthermore, eigenvector continuation can be used with generalized Procrustes orbitals to solve some 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.

$$\mathbf{S}^{\mathrm{EVC}} = \begin{bmatrix} 1 & 0.99998 & 0.99982 & 0.99961 & 0.99893 \\ 0.99998 & 1 & 0.99996 & 0.99978 & 0.99922 \\ 0.99982 & 0.99996 & 1 & 0.99992 & 0.99953 \\ 0.99961 & 0.99978 & 0.99992 & 1 & 0.99983 \\ 0.99893 & 0.99922 & 0.99953 & 0.99983 & 1 \end{bmatrix}$$

Figure 8.8: The overlap matrix for the toy problem shown in the left graphs of figure 2.1 for five sample points, which has eigenvalues $\lambda_0 = 1.1 \cdot 10^{-13}, \lambda_1 = 4.7 \cdot 10^{-10}, \lambda_2 = 7.3 \cdot 10^{-7}, \lambda_3 = 1.3 \cdot 10^{-3}, \lambda_4 = 4.9987$ *and is hence very expensive to sample on a quantum computer.*

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 (HF) states, very few Slater determinants can essentially reproduce the whole potential energy surface (PES) for relatively simple systems. Furthermore, it can "smooth" 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 Hamiltonian with a strengthened/weakened electron-repulsion, qualitative and quantitative improvements are possible. This method is however unreliable and hard to improve further.

The main focus in this thesis was the application of eigenvector continuation to (unitary) coupled cluster singles and doubles ((U)CCSD) wave functions. The CC-reformulation of EVC, subspace-projected CC (in this thesis called WF-CCEVC), can be applied to quantum chemistry and yield promising results. It can be used both as an interpolation tool, where little knowledge about the system is necessary, or in order to improve CCSD energies, in which case one requires knowledge of the system and why CCSD fails. Doing this, we obtained improved potential energy curves for complicated systems such as BeH_2 and N_2. Furthermore, we developed AMP-CCEVC, where the cluster operator is written as a linear combination of sample cluster operators, and found that the true cluster operator is well approximated. In addition, we developed a way to approximate AMP-CCEVC in such a way that it can reduce the cost and time of solving the amplitude equations.

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 can, including systems where the reference determinant changes discontinuously. In particular, the original variational formulation of EVC is restored. Furthermore, expectation values only need to be sampled at one single geometry and can then be reused at all other geometries when using EVC on a quantum computer. At the same time, EVC on a quantum computer cannot be used for extrapolation when sample points are close to one another, as this increases sample need on a quantum computer by several orders of magnitude.

As atomic orbitals and thus the vector space spanned by them change at each nuclear position $\vec{R}$, we considered different ways to make the Hamiltonian matrix change analytically by mapping states from one geometry to another. In order to do so, we compared several schemes to make the molecular orbitals and thus the Hamiltonian matrix change analytically. Using symmetric orthogonalization works well when one wants to have the molecular orbitals change as little as possible in a least square sense when moving them from one geometry to another. When mapping a HF state to a new geometry using symmetric orthogonalization, it has substantial overlap to the exact HF state at the new geometry. Orbitals obtained by solving the orthogonal Procrustes problem are a powerful tool when the reference wave function should be a Hartree-Fock wave function, while minimizing the change in the expansion coefficients of the molecular orbitals in terms of atomic orbitals.

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 [110] leads to extra complications. 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 [111] can be used to effectively reduce the number of amplitudes in AMP-CCEVC and WF-CCEVC. Furthermore, the time improvement using parameter-reduced AMP-CCEVC was not optimized with respect to how data is stored in memory. Another point should be made about the use of Procrustes orbitals. With Procrustes orbitals, the change of all orbitals is changed equivalently, but the weighting might be changed such that the change for some orbitals is penalized more strongly than for others. It is possible that eigenvector continuation and Procrustes orbitals can be used in conjuncture with pair natural orbitals [112] or local natural orbitals [113], [114], which might reduce run time by orders of magnitudes. Furthermore, we have not seen the big computational savings as Ekstrøm and Hagen [4] in the WF-CCEVC approach, because the form of the geometry-dependent Hamiltonian in quantum chemistry still makes it necessary to calculate all projection errors anew at each geometry, which scales as $O(N^2M^4)$. 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 used for similarity transformation at each geometry $\vec{R}$, it is possible that machine learning can be used to get approximate projection errors for each $\hat{T}(\vec{R}_i)$. Furthermore, one 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 viable to develop a variant of WF-CCEVC which can be applied to multi-reference CCSD wave functions, or single-reference CCSD with changing reference determinants. For the Mk-MRCC ansatz, lambda equations are available [115], making it a natural starting point.

Another question to be answered is how the correct sampling geometries are chosen for all EVC methods presented here. 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, but this is not generally the case.

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. [3], 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 have 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 sample parameters. Even though there are no amplitude equations to be solved, the parameter reduction approach can just as well be applied to the stationarity condition $\frac{\partial E}{\partial c_n} = 0$ for $n = 1, \ldots, L$. Thus, one might be able to obtain approximate optimized parameters at reduced cost. The parameter reduction might also be applicable when calculating standard CC amplitudes on top of DIIS, approximating the "best" linear combination of different iterations of cluster operators and thus reducing the total number of iterations faster than DIIS alone.

Appendix A

Figures, Tables and miscellaneous

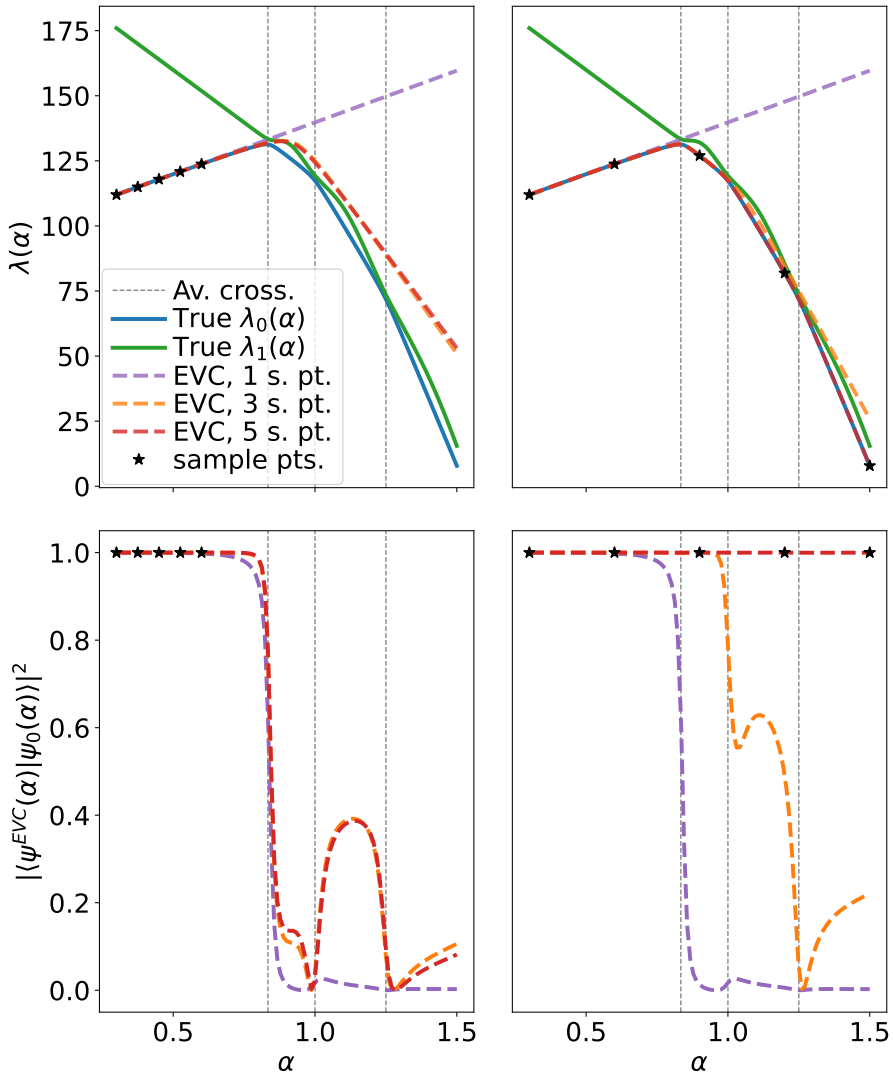


Figure A.1: EVC applied to the toy problem described in sec. A.1 when discarding overlap matrix eigenvectors with value less than $\epsilon = 10^{-5}$, thus discarding the three eigenvectors with the smallest eigenvalue of the overlap matrix when sampling in a small region (left) (but none when sampling along the whole curve (right)).

A.1 Toy model for Eigenvector continuation

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

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

Appendix B

Proofs

B.1 Orthogonal Procrustes problem

Consider the optimization problem

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

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

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

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

$$\mathbf{Q} = \arg\min_{\mathbf{\Omega}} \sum_{i=1}^{m} w_i \|(\mathbf{A\Omega} - \mathbf{B})_i\|^2 \quad \text{subject to} \quad \mathbf{\Omega}^\dagger\mathbf{\Omega} = \mathbf{1} \tag{B.3}$$

where real $w_i \geq 0$ 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. In the proof, we will make use of the Frobenius inner product, defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle_F = \sqrt{\mathrm{Tr}(\mathbf{A}^\dagger\mathbf{B})} \tag{B.4}$$

Theorem B.1.1. *The solution to eq. B.3 is given by* $\mathbf{Q} = \mathbf{U}\mathbf{V}^\dagger$*, where* $\mathbf{U}$ *and* $\mathbf{V}$ *are obtained from the SVD of* $\mathbf{A}^\dagger\mathbf{B}\mathbf{W} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\dagger$*, where* $\mathbf{W} = diag[w_1, \ldots, w_m]$.

Proof.

$$
\begin{aligned}
\mathbf{Q} &= \arg\min_{\mathbf{\Omega}} \sum_{i=1}^{m} w_i \|(\mathbf{A\Omega} - \mathbf{B})_i\|^2 \\
&= \arg\min_{\mathbf{\Omega}} \sum_{i}^{m} w_i \left[\langle (\mathbf{A\Omega})_i, (\mathbf{A\Omega})_i \rangle + \langle \mathbf{B}_i, \mathbf{B}_i \rangle - 2\,\mathrm{Re}(\langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle)\right] \\
&= \arg\min_{\mathbf{\Omega}} \left(-2 \sum_{i}^{m} w_i\, \mathrm{Re}(\langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle) + \text{const.} \right) \\
&= \arg\max_{\mathbf{\Omega}} \sum_{i}^{m} w_i\, \mathrm{Re}(\langle (\mathbf{A\Omega})_i, \mathbf{B}_i \rangle) \\
&= \arg\max_{\mathbf{\Omega}} \sum_{i}^{m} \mathrm{Re}(\langle (\mathbf{A\Omega})_i, w_i \mathbf{B}_i \rangle) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}(\langle \mathbf{A\Omega}, \mathbf{BW} \rangle_F) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}\left(\mathrm{Tr}\left((\mathbf{A\Omega})^\dagger \mathbf{BW}\right)\right) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}\left(\mathrm{Tr}\left(\mathbf{\Omega}^\dagger \mathbf{A}^\dagger \mathbf{BW}\right)\right) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}\left(\langle \mathbf{\Omega}, \mathbf{A}^\dagger \mathbf{BW} \rangle_F\right) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}\left(\langle \mathbf{\Omega}, \mathbf{U\Sigma V}^\dagger \rangle_F\right) \\
&= \arg\max_{\mathbf{\Omega}} \mathrm{Re}\left(\langle \mathbf{U}^\dagger \mathbf{\Omega V}, \mathbf{\Sigma} \rangle_F\right) \\
&\Rightarrow \mathbf{Q} = \mathbf{UV}^\dagger
\end{aligned}
\tag{B.5}
$$

where we made use of the fact that that the identity matrix is the matrix is closest to a diagonal matrix with nonnegative entries [116]. This proof was inspired by a similar proof by deRuiter and Forbes [117]. □

B.2 Analyticity of general Procrustes orbitals

Theorem B.2.1. *General Procrustes orbitals are analytic whenever the overlap matrix* $\mathbf{S}(\vec{R})$ *is analytic and non-singular.*

Proof. Let $\mathbf{B} \in \mathbb{C}^{n\times n}$ be a matrix corresponding to the reference coefficient matrix at $\vec{R}_{\text{ref}}$. In the case of molecular orbitals, where $\mathbf{S}(\vec{R}) \in \mathbb{C}^{n\times n}$ is the Hermitian overlap matrix as a function of a parameter $\vec{R}$, by Lemma 4.3.1, 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

$$
\mathbf{Q}(\vec{R}) = \arg\min_{\mathbf{\Omega}} \left\| \mathbf{S}^{-\frac{1}{2}}(\vec{R})\mathbf{\Omega} - \mathbf{B} \right\|_F \tag{B.6}
$$

with solution, by theorem B.1.1,

$$
\mathbf{Q}(\vec{R}) = \mathbf{U}(\vec{R})\mathbf{V}^\dagger(\vec{R}) \tag{B.7}
$$

where

$$
\mathbf{U}(\vec{R})\mathbf{\Sigma}(\vec{R})\mathbf{V}^\dagger(\vec{R}) = \left(\mathbf{S}^{-\frac{1}{2}}(\vec{R})\right)^\dagger \mathbf{B} = \mathbf{S}^{-\frac{1}{2}}(\vec{R})\mathbf{B} \tag{B.8}
$$

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. □

B.3 Generalized Slater-Condon rules

We derive the General Slater-Condon rules described in sec. 1.5. This is done in a somewhat different way than in Ref. [12], where first quantization was used. To do so, we need the following lemma:

Lemma B.3.1. *Let there be two sets of orthonormal molecular orbitals* $\{|{}^w\phi_i\rangle\}_{i=1}^{M_1}$, $\{|{}^x\phi_i\rangle\}_{i=1}^{M_2}$ *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*

$$[a_i, b_j^\dagger]_+ = \langle|\, a_i b_j^\dagger\,|\rangle = \langle {}^w\phi_i|{}^x\phi_j\rangle \tag{B.9}$$

Proof. Expand $|{}^x\phi_i\rangle$ in a complete basis, such that the first M_1 elements of that basis correspond to the elements $|{}^w\phi_i\rangle$. The basis elements with indices $i = M_1+1,\ldots,\infty$ will also be labelled as $|{}^w\phi_i\rangle = a_i^\dagger\,|\rangle$. Then

$$b_j^\dagger\,|\rangle = |{}^x\phi_j\rangle = \sum_{i=1}^{\infty} \langle {}^w\phi_i|{}^x\phi_j\rangle\, |{}^w\phi_i\rangle = \sum_{i=1}^{\infty} \langle {}^w\phi_i|{}^x\phi_j\rangle\, a_i^\dagger\,|\rangle \tag{B.10}$$

Thus, $b_j^\dagger = \sum_{i=1}^{\infty} \langle {}^w\phi_i|{}^x\phi_j\rangle\, a_i^\dagger$. Then

$$[a_i, b_j^\dagger]_+ = \sum_{k=1}^{\infty} \langle {}^w\phi_k|{}^x\phi_j\rangle\, [a_i, a_k^\dagger]_+ = \sum_{k=1}^{\infty} \langle {}^w\phi_k|{}^x\phi_j\rangle\, \delta_{ik} = \langle {}^w\phi_i|{}^x\phi_j\rangle \tag{B.11}$$

□

B.3.1 The overlap between two Slater determinants in different basis

Lemma B.3.2. *The overlap between non-orthogonal n particle Slater determinants is given by* $\langle {}^w\Psi|{}^x\Psi\rangle = |{}^{wx}\mathbf{S}|$, *with* ${}^{wx}S_{ij} = \langle {}^w\phi_i|{}^x\phi_j\rangle$.

Proof. This proof is by induction on the number of particles. Clearly, in the case $n = 1$, 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 with the $n+1 \times n+1$ overlap matrix ${}^{wx}\mathbf{S}$. Then

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

where, for $i = 1$, $b_1^\dagger\ldots b_{i-1}^\dagger = 1$, and for $i = 2$, $b_1^\dagger\ldots b_{i-1}^\dagger = b_1^\dagger$.

$\langle |\, a_{n+1} \dots a_2 b_1^\dagger \dots b_{i-1}^\dagger a_1 b_i^\dagger b_{i+1}^\dagger b_{n+1} \,| \rangle$ is the overlap between two non-orthogonal Slater determinants with n particles. By the induction hypothesis, it is equal to the determinants of ${}^{wx}\mathbf{S}$ when removing the first row and the ith column of ${}^{wx}\mathbf{S}$

$$\langle |\, a_{n+1} \dots a_2 b_1^\dagger \dots b_{i-1}^\dagger a_1 b_i^\dagger b_{i+1}^\dagger b_{n+1} \,| \rangle = {}^{wx}S_{1,i}^{(1)} \tag{B.13}$$

Recognizing ${}^{wx}S_{1,i}(-1)^{1+i}$ as elements of the cofactor matrix and the last line of eq. (B.12) as the Laplace expansion of $|{}^{wx}\mathbf{S}|$ finishes the proof. □

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

$$\langle {}^w\Psi | h_{\text{nuc}} | {}^x\Psi \rangle = h_{\text{nuc}} \langle {}^w\Psi | {}^x\Psi \rangle \tag{B.14}$$

B.3.2 The one-body matrix element

First, observe that we can rewrite the Hamiltonian $\hat{H}_1$ in second quantization as

$$\hat{H}_1 = \sum_{p,q} h_{pq} b_p^\dagger b_q = \sum_{p,q} h_{pq} \sum_{r=1}^{\infty} S_{rp} a_r^\dagger b_q = \sum_{rq} \left(\sum_p S_{rp} h_{pq} \right) a_r^\dagger b_q = \sum_{rq} {}^{wx}h_{rq} a_r^\dagger b_q \tag{B.15}$$

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

$$\begin{aligned}
\langle {}^w\Psi | \hat{H}_1 | {}^x\Psi \rangle &= \sum_{rq} \langle |\, a_N \dots a_1 \, {}^{wx}h_{rq} a_r^\dagger b_q b_1^\dagger \dots b_N^\dagger \,| \rangle \\
&= \sum_{ij}^{N} \langle |\, a_N \dots a_1 \, {}^{wx}h_{ij} a_i^\dagger b_j b_1^\dagger \dots b_N^\dagger \,| \rangle \\
&= \sum_{ij}^{N} (-1)^{i+j} \, {}^{wx}h_{ij} \langle |\, a_N \dots a_{i+1} a_{i-1} \dots a_1 b_1^\dagger \dots b_{j-1}^\dagger b_{j+1}^\dagger \dots b_N^\dagger \,| \rangle \\
&= \sum_{ij}^{N} (-1)^{i+j} \, {}^{wx}h_{ij} \, {}^{wx}S_{i,j}^{(1)}
\end{aligned} \tag{B.16}$$

B.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

$$\hat{H}_2 = \frac{1}{4} \sum_{pqrs} \langle pq \| rs \rangle \, b_p^\dagger b_q^\dagger b_r b_s = \frac{1}{4} \sum_{pqrs} {}^{wx}\langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger b_r b_s \tag{B.17}$$

where the sums for p, q go to infinity (but that does not matter, hence it is implicit), $\langle pq \| rs \rangle = \langle {}^x\phi_p \, {}^x\phi_q \| {}^x\phi_r \, {}^x\phi_s \rangle$, and ${}^{wx}\langle pq \| rs \rangle = \langle {}^w\phi_p \, {}^w\phi_q \| {}^x\phi_r \, {}^x\phi_s \rangle$. Then

$$\begin{aligned}
&\langle {}^w\Psi | \hat{H}_2 | {}^x\Psi \rangle = \frac{1}{4} \sum_{pqrs} \langle |\, a_N \dots a_1 \, {}^{wx}\langle pq \| rs \rangle \, a_p^\dagger a_q^\dagger b_r b_s b_1^\dagger \dots b_N^\dagger \,| \rangle \\
&= \frac{1}{4} \sum_{ijkl}^{N} {}^{wx}\langle ij \| kl \rangle \, \langle |\, a_N \dots a_1 a_i^\dagger a_j^\dagger b_k b_l b_1^\dagger \dots b_N^\dagger \,| \rangle \\
&= \frac{1}{4} \sum_{ij}^{N} (-1)^{i+j+k+l} \, {}^{wx}\langle ij \| kl \rangle \, \langle |\, a_N \dots a_{i+1} a_{i-1} \dots a_{j+1} a_{j-1} \dots a_1 b_1^\dagger \dots b_{k-1}^\dagger b_{k+1}^\dagger \dots b_{l-1}^\dagger b_{l+1}^\dagger \dots b_N^\dagger \,| \rangle \\
&= \frac{1}{4} \sum_{ijkl}^{N} (-1)^{i+j+k+l} \, {}^{wx}\langle ij \| kl \rangle \, {}^{wx}S_{ij,kl}^{(2)}
\end{aligned} \tag{B.18}$$

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

B.4 Overlap between WF-CCEVC states

We will here prove eq. (2.16). Calculating the matrix elements for the Hamiltonian follows the same pattern. We remind that $|\Phi^{\mathrm{SD}}\rangle$ refers to a Slater determinant and has nothing to do with CC*SD*.

$$\langle\tilde{\Psi}|\Psi\rangle = \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1}+\hat{\Lambda}'\right]e^{\hat{X}}|\Phi^{\mathrm{SD}}\rangle \tag{B.19}$$

$$= \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1}+\hat{\Lambda}'_1+\hat{\Lambda}'_2\right]\left[\mathbb{1}+\hat{X}+\frac{1}{2}\hat{X}^2+\frac{1}{6}\hat{X}^3+\ldots\right]|\Phi^{\mathrm{SD}}\rangle \tag{B.20}$$

$$= \langle\Phi^{\mathrm{SD}}|\left[\mathbb{1}+\hat{\Lambda}'_1+\hat{\Lambda}'_2\right]\left[\mathbb{1}+\hat{X}_1+\hat{X}_2+\frac{1}{2}\hat{X}_1^2\right]|\Phi^{\mathrm{SD}}\rangle \tag{B.21}$$

$$= \langle\Phi^{\mathrm{SD}}|\Phi^{\mathrm{SD}}\rangle + \langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_1\hat{X}_1|\Phi^{\mathrm{SD}}\rangle + \frac{1}{2}\langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_1^2|\Phi^{\mathrm{SD}}\rangle + \langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_2|\Phi^{\mathrm{SD}}\rangle \tag{B.22}$$

$$= 1+\sum_{iajb}\lambda_i^{a\prime}x_j^b\langle\Phi^{\mathrm{SD}}|\hat{a}_i^\dagger\hat{a}_a\hat{a}_b^\dagger\hat{a}_j|\Phi^{\mathrm{SD}}\rangle + \frac{1}{2}\langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_1^2|\Phi^{\mathrm{SD}}\rangle + \langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_2|\Phi^{\mathrm{SD}}\rangle \tag{B.23}$$

$$= 1+\sum_{iajb}\lambda_i^{a\prime}x_j^b\overbrace{\hat{a}_i^\dagger\hat{a}_a\hat{a}_b^\dagger}\hat{a}_j + \frac{1}{2}\langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_1^2|\Phi^{\mathrm{SD}}\rangle + \langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_2|\Phi^{\mathrm{SD}}\rangle \tag{B.24}$$

$$= 1+\sum_{ia}\lambda_i^{a\prime}x_i^a + \frac{1}{2}\langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_1^2|\Phi^{\mathrm{SD}}\rangle + \langle\Phi^{\mathrm{SD}}|\hat{\Lambda}'_2\hat{X}_2|\Phi^{\mathrm{SD}}\rangle \tag{B.25}$$

$$= 1+\sum_{ia}\lambda_i^{a\prime}x_i^a + \sum_{ijklabcd}\lambda_{ij}^{ab\prime}\left(\frac{1}{8}x_k^cx_l^d+\frac{1}{16}x_{kl}^{cd}\right)\langle\Phi^{\mathrm{SD}}|\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k|\Phi^{\mathrm{SD}}\rangle \tag{B.26}$$

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

$$\langle\Phi^{\mathrm{SD}}|\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k|\Phi^{\mathrm{SD}}\rangle \tag{B.27}$$

$$= \overbrace{\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k} + \overbrace{\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k} + \overbrace{\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k} + \overbrace{\hat{a}_i^\dagger\hat{a}_j^\dagger\hat{a}_b\hat{a}_a\hat{a}_c^\dagger\hat{a}_d^\dagger\hat{a}_l\hat{a}_k} \tag{B.28}$$

$$= \delta_{il}\delta_{jk}\delta_{bc}\delta_{ad} - \delta_{ik}\delta_{jl}\delta_{bc}\delta_{ad} - \delta_{il}\delta_{jk}\delta_{bd}\delta_{ac} + \delta_{ik}\delta_{jl}\delta_{bd}\delta_{ac} \tag{B.29}$$

Using the antisymmetry of the λ_{ij}^{ab} and relabeling of dummy variables, we proved eq. (2.16). Going from eq. B.20 to eq. B.21, 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. B.21 to eq. B.22.

Bibliography

[1] P. Sirković and D. Kressner, "Subspace Acceleration for Large-Scale Parameter-Dependent Hermitian Eigenproblems," *SIAM J. Matrix Anal. Appl.*, vol. 37, no. 2, pp. 695–718, May 2016. DOI: 10.1137/15M1017181.

[2] D. Frame, R. He, I. Ipsen, D. Lee, D. Lee, and E. Rrapaj, "Eigenvector Continuation with Subspace Learning," *Phys. Rev. Lett.*, vol. 121, Jul. 2018, Art. no. 032501. DOI: 10.1103/PhysRevLett.121.032501.

[3] A. Sarkar and D. Lee, "Convergence of Eigenvector Continuation," *Phys. Rev. Lett.*, vol. 126, no. 3, Jan. 2021, Art. no. 032501. DOI: 10.1103/PhysRevLett.126.032501.

[4] A. Ekström and G. Hagen, "Global Sensitivity Analysis of Bulk Properties of an Atomic Nucleus," *Phys. Rev. Lett.*, vol. 123, no. 25, Dec. 2019, Art. no. 252501. DOI: 10.1103/PhysRevLett.123.252501.

[5] A. Peruzzo *et al.*, "A variational eigenvalue solver on a photonic quantum processor," *Nat. Commun.*, vol. 5, no. 1, Jul. 2014, Art. no. 4213. DOI: 10.1038/ncomms5213.

[6] P. K. Barkoutsos *et al.*, "Quantum algorithms for electronic structure calculations: Particle-hole Hamiltonian and optimized wave-function expansions," *Phys. Rev. A*, vol. 98, no. 2, Aug. 2018, Art. no. 022322. DOI: 10.1103/PhysRevA.98.022322.

[7] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, NY, US: Dover Publications, Inc., 1982.

[8] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*, 2nd ed. Oxford, U.K.: Oxford University Press, 2017.

[9] E. Gross, E. Runge, and O. Heinonen, *Many-Particle Theory*. Bristol, Philadelphia and New York, US: Adam Higler, 1991.

[10] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic Structure Theory*. Chichester, England: John Wiley & Sons Ltd, 2000.

[11] T. D. Crawford and H. F. Schaefer III, "An Introduction to Coupled Cluster Theory for Computational Chemists," in *Reviews in Computational Chemistry*. New York, NY, US: John Wiley & Sons Ltd, 2000, pp. 33–136. DOI: 10.1002/9780470125915.ch2.

[12] I. Mayer, *Simple Theorems, Proofs, and Derivations in Quantum Chemistry*. New York, NY, USA: Springer Science+Business Media New York, 2003. DOI: 10.1007/978-1-4757-6519-9.

[13] P.-O. Löwdin, "On the Nonorthogonality Problem," in *Advances in Quantum Chemistry*, P.-O. Löwdin, Ed., vol. 5, Cambridge, MA, USA: AP, 1970, pp. 185–199. DOI: 10.1016/S0065-3276(08)60339-1.

[14] P. Echenique and J. L. Alonso, "A mathematical and computational review of Hartree–Fock SCF methods in quantum chemistry," *Mol. Phys.*, vol. 105, no. 23–24, pp. 3057–3098, Dec. 2007. DOI: 10.1080/00268970701757875.

[15] P. Lykos and G. W. Pratt, "Discussion on The Hartree-Fock Approximation," *Rev. Mod. Phys.*, vol. 35, no. 3, pp. 496–501, Jul. 1963. DOI: 10.1103/RevModPhys.35.496.

[16] D. M. Bishop, *Group theory and chemistry*. Mineola, NY, US: Dover Publications, Inc., 1993.

[17] H. G. Hiscock and A. J. W. Thom, "Holomorphic Hartree–Fock Theory and Configuration Interaction," *J. Chem. Theory Comput.*, vol. 10, no. 11, pp. 4795–4800, Nov. 2014. DOI: 10.1021/ct5007696.

[18] H. G. A. Burton, M. Gross, and A. J. W. Thom, "Holomorphic Hartree–Fock Theory: The Nature of Two-Electron Problems," *J. Chem. Theory Comput.*, vol. 14, no. 2, pp. 607–618, Jan. 2018. DOI: 10.1021/acs.jctc.7b00980.

[19] D. Thouless, "Stability conditions and nuclear rotations in the Hartree-Fock theory," *Nucl. Phys.*, vol. 21, pp. 225–232, Nov. 1960. DOI: 10.1016/0029-5582(60)90048-1.

[20] G. Harsha, T. Shiozaki, and G. E. Scuseria, "On the difference between variational and unitary coupled cluster theories," *J. Chem. Phys.*, vol. 148, no. 4, 2018, Art. no. 044107. DOI: 10.1063/1.5011033.

[21] F. A. Evangelista, "Alternative single-reference coupled cluster approaches for multireference problems: The simpler, the better," *J. Chem. Phys.*, vol. 134, no. 22, 2011, Art. no. 224102. DOI: 10.1063/1.3598471.

[22] J. Zhao and G. E. Scuseria, *Drudge/Gristmill*. [Online]. Available: https://tschijnmo.github.io/drudge/ (visited on 05/03/2022).

[23] J. F. Stanton, J. Gauss, J. D. Watts, and R. J. Bartlett, "A direct product decomposition approach for symmetry exploitation in many-body methods. I. Energy calculations," *J. Chem. Phys.*, vol. 94, no. 6, pp. 4334–4345, 1991. DOI: 10.1063/1.460620.

[24] A. G. Taube and R. J. Bartlett, "New perspectives on unitary coupled-cluster theory," *Int. J. Quantum Chem.*, vol. 106, no. 15, pp. 3393–3401, 2006. DOI: 10.1002/qua.21198.

[25] P.-O. Löwdin, "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.*, vol. 97, no. 6, pp. 1474–1489, Mar. 1955. DOI: 10.1103/PhysRev.97.1474.

[26] I. Shavitt, B. J. Rosenberg, and S. Palalikit, "Comparison of configuration interaction expansions based on different orbital transformations," *Int. J. Quantum Chem.*, vol. 10, no. S10, pp. 33–46, 1976. DOI: 10.1002/qua.560100804.

[27] A. G. Taube and R. J. Bartlett, "Frozen natural orbitals: Systematic basis set truncation for coupled-cluster theory," *Collect. Czechoslov. Chem. Commun.*, vol. 70, no. 6, pp. 837–850, 2005. DOI: 10.1135/cccc20050837.

[28] P. J. Hay, "On the calculation of natural orbitals by perturbation theory," *J. Chem. Phys.*, vol. 59, no. 5, pp. 2468–2476, 1973. DOI: 10.1063/1.1680359.

[29] H. Burton, "Holomorphic Hartree-Fock Theory: Moving Beyond the Coulson-Fischer Point," Ph.D. dissertation, Dept. Chem., Univ. of Cambridge, Cambridge, U.K., 2020. DOI: 10.17863/CAM.55079.

[30] J. M. Nite and C. A. Jiménez-Hoyos, "Efficient Multi-Configurational Wavefunction Method with Dynamical Correlation Using Non-Orthogonal Configuration Interaction Singles and Doubles (NOCISD)," Dec. 2019. DOI: 10.26434/chemrxiv.11369646.v1.

[31] H. G. A. Burton, "Generalized nonorthogonal matrix elements: Unifying Wick's theorem and the Slater–Condon rules," *J. Chem. Phys.*, vol. 154, no. 14, 2021, Art. no. 144109. DOI: 10.1063/5.0045442.

[32] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. Berlin, Germany: Springer, 1976.

[33] M. Wilkinson, "Narrowly avoided crossings," *J. Phys. A*, vol. 20, no. 3, pp. 635–645, Feb. 1987. DOI: 10.1088/0305-4470/20/3/026.

[34] P. Demol *et al.*, "Improved many-body expansions from eigenvector continuation," *Phys. Rev. C*, vol. 101, no. 4, Apr. 2020, Art. no. 041302. DOI: 10.1103/PhysRevC.101.041302.

[35] E. A. Hylleraas and B. Undheim, "Numerische Berechnung der 2S-Terme von Ortho- und Par-Helium," *Z. Phys.*, vol. 65, no. 11, pp. 759–772, Nov. 1930. DOI: 10.1007/BF01397263.

[36] P.-O. Löwdin, "Quantum theory of cohesive properties of solids," *Adv. Phys.*, vol. 5, no. 17, pp. 1–171, 1956. DOI: 10.1080/00018735600101155.

[37] M. E. Hochstenbach, C. Mehl, and B. Plestenjak, "Solving Singular Generalized Eigenvalue Problems by a Rank-Completing Perturbation," *SIAM J. Matrix Anal. Appl.*, vol. 40, no. 3, pp. 1022–1046, Jan. 2019. DOI: 10.1137/18m1188628.

[38] R. P. Feynman, "Simulating physics with computers," *Int. J. Theor. Phys.*, vol. 21, no. 6, pp. 467–488, Jun. 1982. DOI: 10.1007/BF02650179.

[39] B. Bauer, S. Bravyi, M. Motta, and G. K.-L. Chan, "Quantum Algorithms for Quantum Chemistry and Quantum Materials Science," *Chem. Rev.*, vol. 120, no. 22, pp. 12 685–12 717, Nov. 2020. DOI: 10.1021/acs.chemrev.9b00829.

[40] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th anniversary ed. Cambridge, U.K.: Cambridge University Press, 2011. DOI: 10.1017/CBO9780511976667.

[41] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge University Press, 1985. DOI: 10.1017/CBO9780511810817.

[42] S. J. Devitt, W. J. Munro, and K. Nemoto, "Quantum error correction for beginners," *Rep. Prog. Phys.*, vol. 76, no. 7, Jun. 2013, Art. no. 076001. DOI: 10.1088/0034-4885/76/7/076001.

[43] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, "Hybrid Quantum-Classical Algorithms and Quantum Error Mitigation," *J. Phys. Soc. Jpn.*, vol. 90, no. 3, 2021, Art. no. 032001. DOI: 10.7566/JPSJ.90.032001.

[44] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca, "Quantum algorithms revisited," *Proc. Math. Phys. Eng. Sci.*, vol. 454, no. 1969, pp. 339–354, Jan. 1998. DOI: 10.1098/rspa.1998.0164.

[45] Y. Cao *et al.*, "Quantum Chemistry in the Age of Quantum Computing," *Chem. Rev.*, vol. 119, no. 19, pp. 10 856–10 915, Aug. 2018. DOI: 10.1021/acs.chemrev.8b00803.

[46] L. Bittel and M. Kliesch, "Training Variational Quantum Algorithms Is NP-Hard," *Phys. Rev. Lett.*, vol. 127, no. 12, Sep. 2021, Art. no. 120502. DOI: 10.1103/physrevlett.127.120502.

[47] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Phys. Rev. A*, vol. 99, no. 3, Mar. 2019, Art. no. 032331. DOI: 10.1103/physreva.99.032331.

[48] H. R. Grimsley, D. Claudino, S. E. Economou, E. Barnes, and N. J. Mayhall, "Is the Trotterized UCCSD Ansatz Chemically Well-Defined?" *J. Chem. Theory Comput.*, vol. 16, no. 1, pp. 1–6, Dec. 2019. DOI: 10.1021/acs.jctc.9b01083.

[49] F. A. Evangelista, G. K.-L. Chan, and G. E. Scuseria, "Exact parameterization of fermionic wave functions via unitary coupled cluster theory," *J. Chem. Phys.*, vol. 151, no. 24, 2019, Art. no. 244112. DOI: 10.1063/1.5133059.

[50] M. Nooijen, "Can the Eigenstates of a Many-Body Hamiltonian Be Represented Exactly Using a General Two-Body Cluster Expansion?" *Phys. Rev. Lett.*, vol. 84, no. 10, pp. 2108–2111, Mar. 2000. DOI: 10.1103/PhysRevLett.84.2108.

[51] D. Mukherjee and W. Kutzelnigg, "Some comments on the coupled cluster with generalized singles and doubles (CCGSD) ansatz," *Chem. Phys. Lett.*, vol. 397, no. 1, pp. 174–179, 2004. DOI: 10.1016/j.cplett.2004.08.100.

[52] J. Lee, W. J. Huggins, M. Head-Gordon, and K. B. Whaley, "Generalized Unitary Coupled Cluster Wave functions for Quantum Computation," *J. Chem. Theory Comput.*, vol. 15, no. 1, pp. 311–324, Jan. 2019. DOI: 10.1021/acs.jctc.8b01004.

[53] J. Tilly *et al.*, "The Variational Quantum Eigensolver: a review of methods and best practices," 2021. arXiv: 2111.05176.

[54] J. Romero, R. Babbush, J. R. McClean, C. Hempel, P. J. Love, and A. Aspuru-Guzik, "Strategies for quantum computing molecular energies using the unitary coupled cluster ansatz," *Quantum Sci. Technol.*, vol. 4, no. 1, Oct. 2018, Art. no. 014008. DOI: 10.1088/2058-9565/aad3e4.

[55] H. R. Grimsley, S. E. Economou, E. Barnes, and N. J. Mayhall, "An adaptive variational algorithm for exact molecular simulations on a quantum computer," *Nat. Commun.*, vol. 10, no. 1, Jul. 2019, Art. no. 3007. DOI: 10.1038/s41467-019-10988-2.

[56] A. Kandala *et al.*, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," *Nature*, vol. 549, no. 7671, pp. 242–246, Sep. 2017. DOI: 10.1038/nature23879.

[57] P. Jordan and E. Wigner, "Über das Paulische Äquivalenzverbot," *Z. Phys.*, vol. 47, no. 9, pp. 631–651, Jul. 1928. DOI: 10.1007/BF01331938.

[58] J. T. Seeley, M. J. Richard, and P. J. Love, "The Bravyi-Kitaev transformation for quantum computation of electronic structure," *J. Chem. Phys.*, vol. 137, no. 22, 2012, Art. no. 224109. DOI: 10.1063/1.4768229.

[59] A. Tranter, P. J. Love, F. Mintert, and P. V. Coveney, "A Comparison of the Bravyi–Kitaev and Jordan–Wigner Transformations for the Quantum Simulation of Quantum Chemistry," *J. Chem. Theory Comput.*, vol. 14, no. 11, pp. 5617–5630, 2018. DOI: 10.1021/acs.jctc.8b00450.

[60] N. Moll, A. Fuhrer, P. Staar, and I. Tavernelli, "Optimizing qubit resources for quantum chemistry simulations in second quantization on a quantum computer," *J. Phys. A*, vol. 49, no. 29, Jun. 2016, Art. no. 295301. DOI: 10.1088/1751-8113/49/29/295301.

[61] S. Bravyi, J. M. Gambetta, A. Mezzacapo, and K. Temme, "Tapering off qubits to simulate fermionic Hamiltonians," 2017. arXiv: 1701.08213.

[62] P. J. J. O'Malley *et al.*, "Scalable Quantum Simulation of Molecular Energies," *Phys. Rev. X*, vol. 6, no. 3, Jul. 2016, Art. no. 031007. DOI: 10.1103/physrevx.6.031007.

[63] K. Setia, R. Chen, J. E. Rice, A. Mezzacapo, M. Pistoia, and J. D. Whitfield, "Reducing Qubit Requirements for Quantum Simulations Using Molecular Point Group Symmetries," *J. Chem. Theory Comput.*, vol. 16, no. 10, pp. 6091–6097, Oct. 2020. DOI: 10.1021/acs.jctc.0c00113.

[64] J. R. McClean, R. Babbush, P. J. Love, and A. Aspuru-Guzik, "Exploiting Locality in Quantum Computation for Quantum Chemistry," *J. Phys. Chem*, vol. 5, no. 24, pp. 4368–4380, Dec. 2014. DOI: 10.1021/jz501649m.

[65] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New J. Phys.*, vol. 18, no. 2, Feb. 2016, Art. no. 023023. DOI: 10.1088/1367-2630/18/2/023023.

[66] M. Motta *et al.*, "Low rank representations for quantum simulation of electronic structure," *npj Quantum Inf.*, vol. 7, no. 1, May 2021, Art. no. 83. DOI: 10.1038/s41534-021-00416-z.

[67] B. Peng and K. Kowalski, "Highly Efficient and Scalable Compound Decomposition of Two-Electron Integral Tensor and Its Application in Coupled Cluster Calculations," *J. Chem. Theory Comput.*, vol. 13, no. 9, pp. 4179–4192, Sep. 2017. DOI: 10.1021/acs.jctc.7b00605.

[68] W. J. Huggins *et al.*, "Efficient and noise resilient measurements for quantum chemistry on near-term quantum computers," *npj Quantum Inf.*, vol. 7, no. 1, Feb. 2021, Art. no. 23. DOI: 10.1038/s41534-020-00341-7.

[69] G. Wang, D. E. Koh, P. D. Johnson, and Y. Cao, "Minimizing Estimation Runtime on Noisy Quantum Computers," *PRX Quantum*, vol. 2, no. 1, Mar. 2021. DOI: 10.1103/prxquantum.2.010346.

[70] W. J. Huggins, J. Lee, U. Baek, B. O'Gorman, and K. B. Whaley, "A non-orthogonal variational quantum eigensolver," *New J. Phys.*, vol. 22, no. 7, Jul. 2020, Art. no. 073009. DOI: 10.1088/1367-2630/ab867b.

[71] N. H. Stair, R. Huang, and F. A. Evangelista, "A Multireference Quantum Krylov Algorithm for Strongly Correlated Electrons," *J. Chem. Theory Comput.*, vol. 16, no. 4, pp. 2236–2245, 2020. DOI: 10.1021/acs.jctc.9b01125.

[72] R. M. Parrish and P. L. McMahon, "Quantum Filter Diagonalization: Quantum Eigendecomposition without Full Quantum Phase Estimation," 2019. arXiv: 1909.08925.

[73] A. C. Davison and D. V. Hinkley, *Bootstrap methods and their application*. Cambridge, U.K.: Cambridge University Press, 1997.

[74] O. Higgott, D. Wang, and S. Brierley, "Variational Quantum Computation of Excited States," *Quantum*, vol. 3, p. 156, Jul. 2019. DOI: 10.22331/q-2019-07-01-156.

[75] A. M. Childs, R. Kothari, and R. D. Somma, "Quantum Algorithm for Systems of Linear Equations with Exponentially Improved Dependence on Precision," *SIAM J. Comput.*, vol. 46, no. 6, pp. 1920–1950, Jan. 2017. DOI: 10.1137/16m1087072.

[76] A. Ralli, P. J. Love, A. Tranter, and P. V. Coveney, "Implementation of measurement reduction for the variational quantum eigensolver," *Phys. Rev. Res.*, vol. 3, no. 3, Aug. 2021, Art. no. 033195. DOI: 10.1103/physrevresearch.3.033195.

[77] R. R. dos Santos, "Introduction to Quantum Monte Carlo simulations for fermionic systems," *Braz. J. Phys.*, no. 1, 2003. DOI: 10.1590/S0103-97332003000100003.

[78] T. U. Helgaker and J. Almlöf, "A second-quantization approach to the analytical evaluation of response properties for perturbation-dependent basis sets," *Int. J. Quantum Chem.*, vol. 26, no. 2, pp. 275–291, 1984. DOI: 10.1002/qua.560260211.

[79] J. Olsen, K. L. Bak, K. Ruud, T. Helgaker, and P. Jørgensen, "Orbital connections for perturbation-dependent basis sets," *Theor. Chem. Acc.*, vol. 90, no. 5, pp. 421–439, 1995.

[80] B. C. Carlson and J. M. Keller, "Orthogonalization Procedures and the Localization of Wannier Functions," *Phys. Rev.*, vol. 105, no. 1, pp. 102–103, Jan. 1957. DOI: 10.1103/PhysRev.105.102.

[81] P. H. Schönemann, "A generalized solution of the orthogonal procrustes problem," *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966. DOI: 10.1007/BF02289451.

[82] S. Banerjee and A. Roy, *Linear Algebra and Matrix Analysis for Statistics*. Chapman and Hall/CRC, Jun. 2014. DOI: 10.1201/b17040.

[83] A. F. Sax, "Localization of molecular orbitals on fragments," *J. Comput. Chem.*, vol. 33, no. 17, pp. 1495–1510, 2012. DOI: 10.1002/jcc.22980.

[84] I. D. Kivlichan *et al.*, "Quantum Simulation of Electronic Structure with Linear Depth and Connectivity," *Phys. Rev. Lett.*, vol. 120, no. 11, Mar. 2018, Art. no. 110501. DOI: 10.1103/physrevlett.120.110501.

[85] V. V. Shende and I. L. Markov, "On the CNOT-cost of TOFFOLI gates," 2008. arXiv: 0803.2316.

[86] G. G. Guerreschi, "Repeat-until-success circuits with fixed-point oblivious amplitude amplification," *Phys. Rev. A*, vol. 99, no. 2, Feb. 2019, Art. no. 022306. DOI: 10.1103/physreva.99.022306.

[87] R. Penrose, "A generalized inverse for matrices," *Math. Proc. Camb. Philos. Soc.*, vol. 51, no. 3, pp. 406–413, 1955. DOI: 10.1017/S0305004100030401.

[88] G. Fix and R. Heiberger, "An Algorithm for the Ill-Conditioned Generalized Eigenvalue Problem," *SIAM J. Numer. Anal.*, vol. 9, no. 1, pp. 78–88, 1972. DOI: 10.1137/0709009.

[89] E. N. Epperly, L. Lin, and Y. Nakatsukasa, "A Theory of Quantum Subspace Diagonalization," 2021. arXiv: 2110.07492.

[90] G. H. Golub and C. F. Van Loan, *Matrix computations*. Baltimore, MD, USA: Johns Hopkins University Press press, 2013.

[91] J. Demmel and B. Kågström, "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.*, vol. 19, no. 2, pp. 175–201, Jun. 1993. DOI: 10.1145/152613.152616.

[92] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, "The Multiconfiguration Time-Dependent Hartree (MCTDH) Method: A Highly Efficient Algorithm for Propagating Wavepackets," *Phys. Rep.*, vol. 324, no. 1, pp. 1–105, 2000. DOI: 10.1016/S0370-1573(99)00047-2.

[93] A. J. W. Thom and M. Head-Gordon, "Hartree–Fock solutions as a quasidiabatic basis for nonorthogonal configuration interaction," *J. Chem. Phys.*, vol. 131, no. 12, 2009, Art. no. 124113. DOI: 10.1063/1.3236841.

[94] F. Prosser and S. Hagstrom, "On the rapid computation of matrix elements," *Int. J. Quantum Chem.*, vol. 2, no. 1, pp. 89–99, 1968. DOI: 10.1002/qua.560020110.

[95] J. Verbeek and J. H. Van Lenthe, "On the evaluation of non-orthogonal matrix elements," *Comput. Theor. Chem.*, vol. 229, pp. 115–137, May 1991. DOI: 10.1016/0166-1280(91)90141-6.

[96] P. Pulay, "Convergence acceleration of iterative sequences. the case of scf iteration," *Chem. Phys. Lett.*, vol. 73, no. 2, pp. 393–398, 1980. DOI: 10.1016/0009-2614(80)80396-4.

[97] G. E. Scuseria, T. J. Lee, and H. F. Schaefer, "Accelerating the convergence of the coupled-cluster approach: The use of the DIIS method," *Chem. Phys. Lett.*, vol. 130, no. 3, pp. 236–239, 1986. DOI: 10.1016/0009-2614(86)80461-4.

[98] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020. DOI: 10.1038/s41586-020-2649-2.

[99] P. Virtanen *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in python," *Nat. Methods*, vol. 17, pp. 261–272, 2020. DOI: 10.1038/s41592-019-0686-2.

[100] Q. Sun *et al.*, "PySCF: the Python-based simulations of chemistry framework," *WIREs Comput. Mol. Sci.*, vol. 8, no. 1, 2018, Art. no. e1340. DOI: 10.1002/wcms.1340.

[101] Q. Sun *et al.*, "Recent developments in the PySCF program package," *J. Chem. Phys.*, vol. 153, no. 2, 2020, Art. no. 024109. DOI: 10.1063/5.0006074.

[102] D. G. a. Smith and J. Gray, "opt_einsum - A Python package for optimizing contraction order for einsum-like expressions," *J. Open Source Softw.*, vol. 3, no. 26, 2018, Art. no. 753. DOI: 10.21105/joss.00753.

[103] M. S. ANIS *et al.*, *Qiskit: An Open-source Framework for Quantum Computing*, 2021. DOI: 10.5281/zenodo.2573505.

[104] J. R. McClean *et al.*, "OpenFermion: The electronic structure package for quantum computers," *Quantum Sci. Technol.*, vol. 5, no. 3, Jun. 2020, Art. no. 034014. DOI: 10.1088/2058-9565/ab8ebc.

[105] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed. New York, NY, USA: Springer Science+Business Media, LLC, 1999.

[106] T. Bodenstein and S. Kvaal, "A state-specific multireference coupled-cluster method based on the bivariational principle," *J. Chem. Phys.*, vol. 153, no. 2, 2020, Art. no. 024106. DOI: 10.1063/5.0009429.

[107] G. D. Purvis III, R. Shepard, F. B. Brown, and R. J. Bartlett, "C2V Insertion pathway for BeH2: A test problem for the coupled-cluster single and double excitation model," *Int. J. Quantum Chem.*, vol. 23, no. 3, pp. 835–845, 1983. DOI: 10.1002/qua.560230307.

[108] I. W. Bulik, T. M. Henderson, and G. E. Scuseria, "Can Single-Reference Coupled Cluster Theory Describe Static Correlation?" *J. Chem. Theory Comput.*, vol. 11, no. 7, pp. 3171–3179, 2015. DOI: 10.1021/acs.jctc.5b00422.

[109] M. Rossmannek, P. K. Barkoutsos, P. J. Ollitrault, and I. Tavernelli, "Quantum HF/DFT-embedding algorithms for electronic structure calculations: Scaling up to complex molecular systems," *J. Chem. Phys.*, vol. 154, no. 11, Mar. 2021, Art. no. 114105. DOI: 10.1063/5.0029536.

[110] T. Helgaker and P. Jørgensen, "An electronic Hamiltonian for origin independent calculations of magnetic properties," *J. Chem. Phys.*, vol. 95, no. 4, pp. 2595–2601, 1991. DOI: 10.1063/1.460912.

[111] H. Koch, A. Sánchez de Merás, and T. B. Pedersen, "Reduced scaling in electronic structure calculations using Cholesky decompositions," *J. Chem. Phys.*, vol. 118, no. 21, pp. 9481–9484, 2003. DOI: 10.1063/1.1578621.

[112] F. Neese, F. Wennmohs, and A. Hansen, "Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method," *J. Chem. Phys.*, vol. 130, no. 11, 2009, Art. no. 114108. DOI: 10.1063/1.3086717.

[113] Z. Rolik and M. Kállay, "A general-order local coupled-cluster method based on the cluster-in-molecule approach," *J. Chem. Phys.*, vol. 135, no. 10, 2011, Art. no. 104111. DOI: 10.1063/1.3632085.

[114] P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the Linear-Scaling Local Natural Orbital CCSD(T) Method: Improved Algorithm and Benchmark Applications," *J. Chem. Theory Comput.*, vol. 14, no. 8, pp. 4193–4215, Aug. 2018. DOI: 10.1021/acs.jctc.8b00442.

[115] E. Prochnow, F. A. Evangelista, H. F. Schaefer, W. D. Allen, and J. Gauss, "Analytic gradients for the state-specific multireference coupled cluster singles and doubles model," *J. Chem. Phys.*, vol. 131, no. 6, 2009, Art. no. 064109. DOI: 10.1063/1.3204017.

[116] K. Fan and A. J. Hoffman, "Some metric inequalities in the space of matrices," *Proc. Amer. Math. Soc.*, vol. 6, no. 1, pp. 111–116, 1955. DOI: 10.1090/S0002-9939-1955-0067841-7.

[117] A. H. J. de Ruiter and J. R. Forbes, "On the Solution of Wahba's Problem on $SO(n)$," *J. Astronaut. Sci.*, vol. 60, no. 1, pp. 1–31, Mar. 2013. DOI: 10.1007/s40295-014-0019-8.