Simen Kvaal

Lecture Notes in Quantum Chemistry

COMPILED SEPTEMBER 15, 2026

Publisher of This Book

Copyright © 2026 Simen Kvaal

PUBLISHED BY PUBLISHER OF THIS BOOK

LaTeXtemplate based on TUFTE-LATEX, distributed under the Apache License 2.0

Contents

1 Lecture 1: Linear algebra

Quantum chemistry relies heavily on certain mathematical tools. Chief among these are *linear algebra*. This chapter is an attempt at giving a refresher of the most important concepts.

It is strongly suggested to read Chapter 1 in SO.

To be written.

1.1 Avoided crossing exercise

To be written.

2 Lecture 2: Atomic units and the Born–Oppenheimer approximation

2.1 Atomic units

To be written ...

2.2 The Born–Oppenheimer approximation

The Born–Oppenheimer approximation is a very central construction in quantum chemistry. It is quite necessary for chemical intuition to make sense. We will try to explain this presently.

For this section, we assume familiarity with basic quantum mechanics as taught in, say, KJM2601. You should review the curriculum on basic quantum mechanics.

For us, the most important aspects are the following: In the previous lecture, we recapped linear algebra – the most important tool in quantum chemistry. Linear algebra is important because we formulate our computational methods in terms of it. Our methods are approximations of *The Schrödinger equation*, a *partial differential equation* that relates a *wavefunction* and its energy. For example, for the hydrogen atom, we are looking for a wavefunction $\psi(\mathbf{r})$ which is square integrable,

$$\|\psi\|^2 = \int |\psi(\mathbf{r})|^2 \, d^3r \; = \; 1,$$

meaning that it is an element of the Hilbert space $L^2(\mathbb{R}^3)$. The Schrödinger equation for the hydrogen atom is an eigenvalue equation and reads

$$\left(-\frac{\hbar^2}{2m_e}\nabla^2 + \frac{e^2}{4\pi\epsilon_0}\frac{1}{\|\mathbf{r}\|} \right) \psi(\mathbf{r}) = E\psi(\mathbf{r}).$$

Here the *Hamiltonian operator* is

$$\hat{H} = -\frac{\hbar^2}{2m_e}\nabla^2 + \frac{e^2}{4\pi\epsilon_0}\frac{1}{\|\mathbf{r}\|},$$

and is the observable for total energy. We observe that the Schrödinger equation has the same structure as a matrix eigenvalue equation, but that the unknown is a *function* in Hilbert space.

There will in general be many discrete eigenvalues E_μ, with eigenfunctions (orbitals) $\psi_\mu(\mathbf{r})$. For the hydrogen atom these are well-known. The eigenfunctions are orthonormal,

$$\langle \psi_\mu | \psi_\nu \rangle = \int \psi_\mu(\mathbf{r})^* \psi_\nu(\mathbf{r})\, d^3r = \delta_{\mu\nu}.$$

The clamped-nuclei approximation is extremely important.

The hydrogen atom as presented above is the simplest example of *the clamped-nuclei approximation*, where the nuclei of the molecule are held fixed in space, while electrons move around quantum mechanically. This approximation is *the most important item of this lecture*, and the basis for virtually all subsequent material. However, *the Born–Oppenheimer approximation* is a refinement of the approximation that takes into account nuclei as quantum particles, and of great importance in chemistry, although we will not study it extensively in this course.

So we will have a Hilbert space Schrödinger equation to solve. In subsequent lectures, we will approximate the wavefunction using a finite number of basis functions, transporting us to the realm of linear algebra. This is necessary, because our Hilbert space is infinite dimensional, meaning that it is not possible to find a *finite* basis, in contrast to finite dimensional vector spaces such as $\mathbb{C}^n$. Infinite dimensionality introduces quite a few mathematical subtleties that are beyond the scope of our course, but we will later on briefly mention (but not prove) some aspects of the set of eigenvalues and eigenfunctions of the Hamiltonian.

2.2.1 *From classical to quantum mechanics for the molecule*

Consider a molecule consisting of N_{nuc} *classical* nuclei and N_{el} *classical* electrons. The water molecule is illustrated in the margin. By "clasical" we mean that each massive particle in the theory has a definite position and velocity/momentum, that change in time according to Newton's laws.

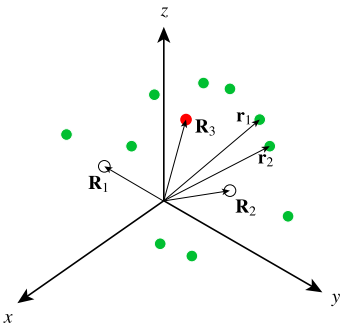


Illustration of the classical configuration of the water molecule.

Nucleus number $a = 1, \ldots, N_{\text{nuc}}$, is located at $\mathbf{R}_a \in \mathbb{R}^3$, has momentum $\mathbf{P}_a$, charge $+eZ_a$, and mass M_a. (Recall that nuclei consist of N_a neutrons and Z_a protons.) On the other hand, electron number $i = 1, \ldots, N_{\text{el}}$ has position $\mathbf{r}_i$, momentum $\mathbf{p}_i$, charge $-e$ and mass m_e.

The total energy of the molecule is, assuming complete isolation from the rest of the universe and no external electromagnetic forces acting,

$$\mathcal{H}_{\text{mol}} = \mathcal{T}_{\text{nuc}} + \mathcal{T}_{\text{el}} + \mathcal{V}_{\text{nuc-nuc}} + \mathcal{V}_{\text{el-el}} + \mathcal{V}_{\text{nuc-el}} \tag{2.1}$$

with kinetic energy contributions from nuclei,

$$\mathcal{T}_{\text{nuc}} = \sum_{a=1}^{N_{\text{nuc}}} \frac{\mathbf{P}_a^2}{2M_a}, \tag{2.2}$$

and electrons,

$$\mathcal{T}_{\text{el}} = \sum_{i=1}^{N_{\text{el}}} \frac{\mathbf{p}_i^2}{2m_e}. \tag{2.3}$$

Potential energy is split according to which particles interact with who: Pairs of nuclei,

$$\mathcal{V}_{\text{nuc-nuc}} = \frac{e^2}{4\pi\epsilon_0} \sum_{a=1}^{N_{\text{nuc}}-1} \sum_{b=a+1}^{N_{\text{nuc}}} \frac{Z_a Z_b}{\|\mathbf{R}_a - \mathbf{R}_b\|}, \tag{2.4}$$

pairs of electrons,

$$\mathcal{V}_{\text{el-el}} = \frac{e^2}{4\pi\epsilon_0} \sum_{i=1}^{N_{\text{el}}-1} \sum_{j=i+1}^{N_{\text{el}}} \frac{1}{\|\mathbf{r}_i - \mathbf{r}_j\|}, \tag{2.5}$$

and finally pairs of one electron and a nucleus,

$$\mathcal{V}_{\text{nuc-el}} = -\frac{e^2}{4\pi\epsilon_0} \sum_{a=1}^{N_{\text{nuc}}} \sum_{i=1}^{N_{\text{el}}} \frac{Z_a}{\|\mathbf{R}_a - \mathbf{r}_i\|}. \tag{2.6}$$

The *state* of the classical system is the position-momentum vector

$$\begin{aligned}\xi = (r, R, p, P) = (&\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{el}}}, \mathbf{R}_1, \dots, \mathbf{R}_{N_{\text{nuc}}}, \\ &\mathbf{p}_1, \dots, \mathbf{p}_{N_{\text{el}}}, \mathbf{P}_1, \dots, \mathbf{P}_{N_{\text{nuc}}}) \in \mathbb{R}^{6(N_{\text{el}}+N_{\text{nuc}})}.\end{aligned} \tag{2.7}$$

Now, turning to the non-relativistic quantum mechanical description, and ignoring *spin* for the time being, the state becomes a complex-valued wavefunction

$$\begin{aligned}\Psi_{\text{mol}}(r, R) = \Psi_{\text{mol}}(&\mathbf{r}_1, \dots, \mathbf{r}_{N_{\text{el}}}, \mathbf{R}_1, \dots, \mathbf{R}_{N_{\text{nuc}}}, \\ &\mathbf{p}_1, \dots, \mathbf{p}_{N_{\text{el}}}, \mathbf{P}_1, \dots, \mathbf{P}_{N_{\text{nuc}}}),\end{aligned} \tag{2.8}$$

while momentum vectors are replaced by differential operators,

$$\mathbf{p}_i \longrightarrow -i\hbar\nabla_{\mathbf{r}_i}, \quad \mathbf{P}_a \longrightarrow -i\hbar\nabla_{\mathbf{R}_a}.$$

The the total energy function becomes the *Hamilton operator*

$$\hat{H}_{\text{mol}} = \hat{T}_{\text{nuc}} + \hat{T}_{\text{el}} + \hat{V}_{\text{nuc-nuc}} + \hat{V}_{\text{el-el}} + \hat{V}_{\text{nuc-el}}, \tag{2.9}$$

where the kinetic energy operators are

$$\hat{T}_{\text{nuc}} = \sum_{a=1}^{N_{\text{nuc}}} \frac{-\hbar^2}{2M_a} \nabla^2_{\mathbf{R}_a}, \quad \hat{T}_{\text{el}} = \sum_{i=1}^{N_{\text{el}}} \frac{-\hbar^2}{2m_e} \nabla^2_{\mathbf{r}_i},$$

and where the potential operators $\hat{V}_{\text{el-el}}$, et.c., are just the operators that multiply with the potential energy functions $\mathcal{V}_{\text{el-el}}$, et.c.

Whereas the classical molecule's state is governed by Newton's laws, the quantum molecule's state is governed by the time-dependent Schrödinger equation,

$$i\hbar \frac{\partial}{\partial t} \Psi_{\text{mol}}(r, R, t) = \hat{H}_{\text{mol}} \Psi_{\text{mol}}(r, R, t).$$

We recall that *stationary states* are solutions of the *time-independent* Schrödinger equation,

$$\hat{H}_{\text{mol}} = \Psi_{\text{mol},k}(r, R) = E_{\text{mol},k} \Psi_{\text{mol},k}(r, R). \tag{2.10}$$

Here, k is an index that counts the different solutions. The solution wavefunctions are called "stationary" since

$$\Psi_{\text{mol}}(r, R, t) = \Psi_{\text{mol},k}(r, R) \exp(-itE_{\text{mol},k}/\hbar)$$

solves the time-dependent Schrödinger equation, but with time-independent probability density.

Solving Eq. (2.10) is an *immense task*. Firstly, the molecular wavefunction, depending on all the massive particles' coordinates at once, is such a monster that it can only be obtained approximately for very modest molecules. Secondly, the molecular wavefunction quite effectively hides molecular structure. The wavefunction does not yield molecules with a clear structure. For example, nuclear positions are only given probabilistically. Of course, the information is there, hidden in all the correlations of the complicated Ψ_{mol}, but it is very hard to obtain.

2.2.2 *The Clamped-nuclei approximation*

Enter the *Clamped-nuclei approximation*: The nuclei are very heavy compared to the electrons. Indeed, the proton mass and the electron mass satisfy

$$\frac{m_p}{m_2} \approx 1836.$$

Our intuition about heavy and light objects tells us that *nuclei will move slowly relative to electrons*. In fact, in the clamped-nuclei approximation one approximates the nuclei as being completely stationary particles, so heavy that they do not even have quantum wavefunctions; just a postition $\mathbf{R}_a$ in space as well as a charge $+eZ_a$. Only the electrons are described by quantum mechanics in this approximation. Thus, we eliminate kinetic energy $\hat{T}_{\text{nuc}}$ of the nuclei, fix $\mathbf{R}_a$ for each nucleus $a = 1, \ldots, N_{\text{nuc}}$, and observe that $\hat{V}_{\text{nuc-nuc}}$ then becomes a constant:

$$\hat{H}_{\text{mol}} = \hat{T}_{\text{nuc}} + \hat{H}_{\text{el}}(R), \tag{2.11}$$

$$\hat{H}_{\text{el}}(R) = \hat{T}_{\text{el}} + \hat{V}_{\text{el-el}} + \hat{V}_{\text{nuc-el}}(R) + \hat{V}_{\text{nuc-nuc}}(R), \tag{2.12}$$

$$\hat{V}_{\text{nuc-el}}(R) = -\frac{e^2}{4\pi\epsilon_0} \sum_{a=1}^{N_{\text{nuc}}} \sum_{i=1}^{N_{\text{el}}} \frac{Z_a}{\|\mathbf{R}_a - \mathbf{r}_i\|}. \tag{2.13}$$

The electrons-only system has a wavefunction

$$\Psi_{\text{el}}(\mathbf{r}_1, \ldots, \mathbf{r}_{N_{\text{el}}}; R),$$

that depends *parametrically* on R, meaning that selecting an R amounts to selecting where the nuclei are – R is not a quantum variable, so choosing R in the wavefunction does not correspond to asking what is the probability amplitude for locating the nuclei at R. (The wavefunction also depends parametrically on the nuclear charges, but as these usually are not changed in a chemical reaction, one omits their specification.) The Hamiltonian of the electrons also depends parametrically on R, since this variable only

enters $\hat{V}_{\text{nuc-el}}$ in terms of where the nuclei are. this explains the appearance of R in parenthesis in $\hat{H}_{\text{el}}$,

$$\hat{H}_{\text{el}}(R) = \hat{T}_{\text{el}} + \hat{V}_{\text{el-el}} + \hat{V}_{\text{nuc-el}}(R) + +\hat{V}_{\text{nuc-nuc}}(R) \tag{2.14}$$

The parentheses "(R)" in $\hat{H}_{\text{el}}(R)$ signifies the parametric dependence. It is not an operator that "acts on" nuclear coordinates – the operator *itself* changes whenever R changes.

We note also that $\hat{V}_{\text{nuc-nuc}}(R)$ is just a constant in the electronic Hamiltonian, since it is a purely R-dependent function. Therefore it is sometimes omitted from the Hamiltonian and added later.

We can now write down the *electronic Schrödinger equation:*

$$\hat{H}_{\text{el}}(R)\Psi_{\text{el}}(r;\, R) = E_{\text{el}}(R)\Psi_{\text{el}}(r;\, R). \tag{2.15}$$

While we have simplified the molecular Schrödinger equation somewhat by removing the nuclei from the picture, the electronic Schrödinger equation is *still very hard to solve*. The main simplification so far is conceptually: Electrons move about fixed nuclei, and we have the nuclear "scaffold" as a well-defined thing.

Note again that we have ignored electron (and nuclear) spin in this presentation. However, adding spin changes nothing in our equations except their complexity and difficulty to read. We will address electron spin properly in the next lecture.

2.2.3 *What is in the molecular geometry?*

For the moment, we now fix a nuclear geometry R. For example, we could think of the simplest diatomic, $H_2{}^+$, consisting of two protons and a single electron. Physical intuition and common sense tell us that *where* the molecule is and *which orientation* it has should not matter for the solutions of the electronic Schrödinger equation. Indeed this is true. Since translation is described by 3 real numbers, and orientation of a rigid body by 3 real numbers (Euler angles), we have in total $3N_{\text{nuc}} - 6$ real numbers, or "degrees of freedom", that describe molecular geometry. Furthermore, for linear molecules, only 2 numbers are needed for orientation due to the symmetric rotation axis, totaling $3N_{\text{nuc}} - 5$ real numbers.

Consider a diatomic like $H_2{}^+$, which is linear. We obtain $3 \cdot 2 - 5 = 1$ degree of freedom. Clearly, the interatomic distance is sufficient, $R = \|\mathbf{R}_1 - \mathbf{R}_2\|$. Any other geometry of the diatomic molecule is obtained by first rotating along two axes, and then translation in space.

2.2.4 *The spectrum of the electronic Hamiltonian*

In this section, we summarize some mathematical facts. Obtaining a complete understanding of these facts is far beyond the scope of this course. We will not have extensive use of these facts later in the course.

As usual with the Schrödinger equation, the eigenfunctions are not unique. Let us write, with some simplification in the notation,

$$\hat{H}_{\text{el}}(R)\Psi_\mu(r;R) = E_\mu(R)\Psi_\mu(r;R) \tag{2.16}$$

The index μ labels the solutions. I mathematics, the set $\sigma(\hat{H}_{\text{el}}(R))$ of possible eigenvalues is called *the spectrum of* $\hat{H}_{el}$:

$$\sigma(\hat{H}_{\text{el}}(R)) = \{\text{all eigenvalues}\} \subset \mathbb{R}.$$

The mathematical term "spectrum" is of course natural in light of its interpretation in physics. The spectrum is now a function of R – in our example the interatomic distance. Thus, we now fix R, and omit its specification for ease of notation.

Mathematicians have proved that the spectrum of the electronic Hamiltonian in general can be divided in two parts: The *discrete spectrum* $\sigma_d(\hat{H}_{\text{el}})$ consisting of isolated points on the real line, and a *continuous spectrum* $\sigma_c(\hat{H}_{\text{el}})$ consisting of a half-interval $[\Sigma, +\infty)$. The number Σ is called *the ionization threshold*, and the distance from the ground state energy is the energy required to remove an electron from the molecule. The spectrum can be visualized as in the margin.

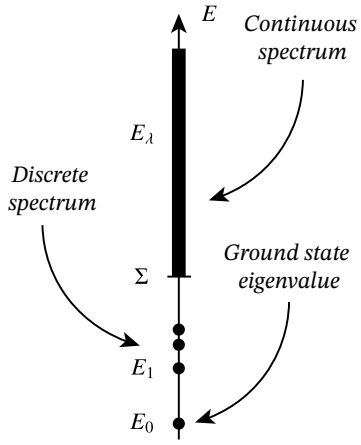


Illustration of the spectrum of the electronic Hamiltonian $\hat{H}_{\text{el}}$.

It is always so that the discrete spectrum is at the bottom, while the continuous spectrum starts at some value Σ and continues to infinity:

$$\sigma(\hat{H}_{\text{el}}) = \{E_\mu \mid \mu = 1, 2, \ldots\} \cup [\Sigma, +\infty).$$

Every neutral molecule has at least one discrete eigenvalue.

The discrete spectrum has eigenfunctions that are square integrable, and are therefore labeled by a discrete index μ:

$$\hat{H}_{\text{el}}\Psi_\mu(r) = E_\mu\Psi_\mu(r) \quad \text{(discrete)}$$

$$\int |\Psi_\mu(r)|^2 \, d^{3N_{\text{el}}}r \; = 1.$$

Indeed, for different eigenvalues, the eigenfunctions are *orthonormal*,

$$\langle\Psi_\mu|\Psi_\nu\rangle = \int \Psi_\mu(r)^*\Psi_\nu(r) \, d^{3N_{\text{el}}}r \; = \delta_{\mu\nu}.$$

On the other hand, eigenfunctions belonging to a point in the continuous spectrum are *not normalizable*, and called a *continuum eigenfunction*. These are therefore labeled by the continuous energy $E \in [\Sigma, +\infty)$:

$$\hat{H}_{\text{el}}\Psi_E(r) = E\Psi_E(r) \quad \text{(contiunuous spectrum)}$$

The eigenfunctions are "δ-normalized", i.e.,

$$\langle\Psi_E|\Psi_{E'}\rangle = \int \Psi_E(r)^*\Psi_{E'}(r) \, d^{3N_{\text{el}}}r \; = \delta(E - E'),$$

where $\delta(x)$ is the Dirac δ-function. In particular, this "function" is infinite at $x = 0$, meaning that the norm of the continuum wavefunction cannot be defined.

The *normalizable* eigenfunctions are called *bound states*. The continuous-spectrum eigenfunctions are called *scattering states*. Since a quantum system must have square integrable states (due to the probability interpretation), the scattering states are not *realizable* in the sense that the electrons cannot live in a single scattering state. These states are idealizations akin to plane waves. On the other hand, the spectrum is *complete* in the sense that we can expand *any square-integrable wavefunction in the eigenfunctions*:

$$\Psi(r) = \sum_{\mu} c_{\mu} \Psi_{\mu}(r) + \int_{[\Sigma,+\infty)} c(E) \Psi_E(r)\, dE. \tag{2.17}$$

Here, we have a sum over discrete eigenvectors, and an integral over the continuum. Mathematicians have proved that when we solve the time-dependent Schrödinger equation with this general initial condition, the part that is expanded in the continuous-spectrum eigenfunctions *escape from the nuclei as time approaches infinity* while the bound-state part does not escape. This explains the term "scattering states".

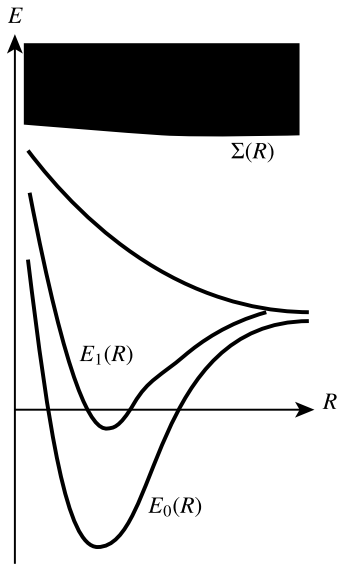


Illustration of the R-dependent spectrum, including the ionization threshold Σ.

2.2.5 *Potential-energy surfaces (PES)*

We now consider what happens with the eigenvalues as function of R. Keep the H_2^+ example in mind. The electronic ground state energy is the minimal energy attainable by the electrons for any given R. For some R, this energy is lower than the energy of isolated atoms, binding the molecule together: nuclei repel each other, and so do electrons – in particular the potential energy is always positive. Kinetic energy is positive, too. So molecules must be bound by electrons and nuclei attracting each other. The result is a curve as shown in the margin.

The excited electronic states have energies above the ground state energy. For H_2^+ there are infinitely many such bound states. They cluster around the value $\Sigma = 0$, where the continuous spectrum starts. In the figure, though, the more general situation is displayed, where the ionizaton threshold can vary with R.

Note that for *every* R we have an electronic energy $E_\mu(R)$ and a corresponding wavefunction $\Psi_\mu(r; R)$. This concept will be attempted illustrated in a future version of the figure.

2.2.6 *Born–Huang expansion*

Let R be a fixed molecular geometry. We now suppose we find a complete set of electronic wavefunctions, and it is customary to ignore the continuous spectrum from now on,

$$\hat{H}_{\text{el}}(R)\Psi_\mu(r;\, R) = E_\mu(R)\Psi_\mu(r;\, R), \tag{2.18}$$

where μ counts the eigenstates, and they are complete. Completeness means that any allowed electronic wavefunction can be expanded according to

$$\Psi(r;\, R) = \sum_{\mu} \chi_\mu(R)\, \Psi_\mu(r;\, R).$$

Here, $\chi_\mu(R)$ is the quantum mechanical amplitude that the electrons are in the eigenstate $\Psi_\mu(r;\, R)$. This is basically the same expansion as we had when we discussed the spectrum of the electronic Hamiltonian.

Again, R is here entering parametrically, and $\chi_\mu(R)$ is just a sequence of complex numbers defined for our fixed R. This sequence satisfies

$$\sum_\mu |c_\mu(R)|^2 = 1.$$

The adiabatic assumption: Recall, that nuclei move very slowly relative to electrons, or, equivalently, that electrons move very fast relative to nuclei. There is a theorem in quantum mechanics called *the adiabatic theorem*: If a quantum system is in its ground state, and if a parameter is changed "infinitely slowly", the system will remain in the ground state for every parameter value. Thus, if the electrons start in the ground state $\Psi_0(r,\ R)$, then they will remain in the ground state as the nuclei move.

We now make the nuclei quantum mechanical. There will be a wavefunction $\chi(R)$ for the nuclei, and the question is: What is the wavefunction of the full molecular system? For the nuclear geometry R, we know with certainty that the electrons are in $\Psi_0(r,\ R)$. There is a way to form the wavefunction of a *combined* quantum system, and that is taking the product of their wavefunctions, i.e.,

$$\Psi_{\text{mol}}(r, R) \approx \chi(R)\Psi_{\text{el},0}(r;\, R).$$

This wavefunction is the *Born–Oppenheimer wavefunction*. Effectively, we choose the first term of the full electronic wavefunction above, and promote the parametric variable R to a quantum variable. Now the function $\chi(R) = \chi_0(R)$ is actually a wavefunction!

We could do a more accurate wavefunction, allowing the electrons to also reside in excited states. This makes sense since the adiabatic approximation is not exact. We then obtain the *Born–Huang wavefunction*

$$\Psi_{\text{mol}}(r,\ R) \approx \sum_\mu \chi_\mu(R)\, \Psi_{\text{el},\mu}(r;\, R),$$

where the sum runs over as many terms as we desire, typically only a handful.

2.2.7 *Born–Oppenheimer approximation*

What is the Schrödinger equation satisfied by the Born–Oppenheimer wavefunction? We start by recalling the molecular Hamiltonian, which acts on the molecular wavefunction:

$$\hat{H}_{\text{mol}} = \hat{T}_{\text{nuc}} + \hat{H}_{\text{el}}(R).$$

Since the derivative terms of $\hat{H}_{\text{el}}$ act only on r, we get

$$\hat{H}_{\text{el}}(R)\Psi_{\text{mol}}(r, R) = \chi(R)[\hat{H}_{\text{el}}(R)\Psi_{\text{el},0}(r, R)] = E_0(R)\Psi_{\text{mol}}(r, R).$$

The nuclear kinetic energy operator is

$$\hat{T}_{\text{nuc}} = \sum_a \frac{-\hbar^2}{2M_a}\nabla_{\mathbf{R}_a} \cdot \nabla_{\mathbf{R}_a}.$$

When applied to the product it will generate several terms:

$$\hat{T}_{\text{nuc}}\,\chi(R)\Psi_{\text{el},0}(r; R) = \sum_a \frac{-\hbar^2}{2M_a}\Big([\nabla^2_{\mathbf{R}_a}\chi(R)]\Psi_0(r; R) + 2\nabla_{\mathbf{R}_a}\chi(R) \cdot \nabla_{\mathbf{R}_a}\Psi_0(r; R) + \chi(R)\nabla^2_{\mathbf{R}_a}\Psi_0(r; R)\Big) \tag{2.19}$$

The two last terms in the big parenthesis are called *nuclear adiabatic coupling terms*, and are assume smalls. This is not always true. The assumption is essentially that the electronic wavefunction changes slowly with R, which is okay when electronic states are well-separated in energy. Ignoring these two terms, therefore, leads to the equation

$$\hat{H}_{\text{mol}}\,\chi(R)\Psi_{\text{el},0}(r; R) = [(\hat{T}_{\text{nuc}} + E_0(R))\chi(R)]\Psi_{\text{el},0}(r; R). \tag{2.20}$$

We multiply from the left with $\Psi_{\text{el},0}(r; R)^*$ and integrate over electronic coordinates. This gives the *nuclear Schroedinger equation in the adiabatic approximation*:

$$[\hat{T}_{\text{nuc}} + \hat{V}_{\text{nuc-nuc}} + E_0(R)]\,\chi_k(R) = E_{\text{mol},k}\,\chi_k(R).$$

Here, $\chi_k(R)$ is the k'th eigenfunction with eigenvalue $E_{\text{mol},k}$. Thus, the nuclei move about in a potential set up by the electrons $E_0(R)$. This potential is unknown, and must be computed for every R we are interested in. Thus, it is quite difficult to obtain the potential energy surface to a sufficient degree of accuracy so we can solve the nuclear Schrödinger equation. Most of our work in subsequent lectures will concern computing $E_0(R)$.

In the margin, we have visualized the electronic PES and the nuclear ground state wavefunction for a Morse potential, a model for diatomic potentials. This visualization ignores rotation of the molecule.

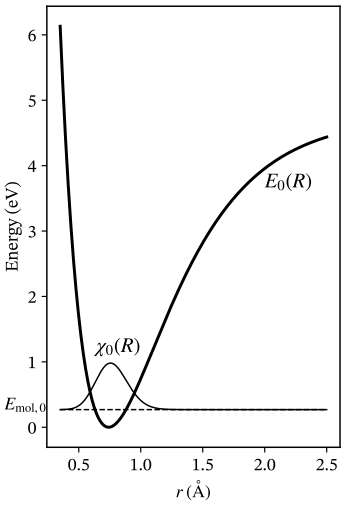


Illustration of the nuclear ground-state wavefunction $\chi_0(R)$ superimposed on the electronic potential-energy surface $E_0(R)$, modeled by a Morse potential. The wavefunction is plotted on a vertical position shifted according to the ground-state eigenvalue, as is common.

2.3 Exercises

2.3.1 Exercise 1: Exact two-body Schrödinger equation, COM/relative coordinates, and clamped nuclei

Consider the problem of solving the Schrödinger equation for the hydrogen atom (an electron and a nucleus) as the simplest "molecular" Schrödinger equation. Let the electron have mass m_e and charge $-e$, and the nucleus have mass m_p and charge $+e$. Let their position operators be $\mathbf{r}_e$ and $\mathbf{r}_p$ in lab coordinates. The exact nonrelativistic Hamiltonian is

$$\hat{H}_{\text{mol}} = -\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}_e}^2 - \frac{\hbar^2}{2m_p}\nabla_{\mathbf{r}_p}^2 - \frac{e^2}{4\pi\varepsilon_0\,|\mathbf{r}_e - \mathbf{r}_p|}. \tag{2.21}$$

(a) Exact problem and translation invariance. Show that the potential depends only on the separation $\mathbf{r}_e - \mathbf{r}_N$. Argue that this implies invariance of solutions of the Schrödinger equation under uniform translations $(\mathbf{r}_e, \mathbf{r}_p) \mapsto (\mathbf{r}_e + \mathbf{a}, \mathbf{r}_p + \mathbf{a})$, and that the total momentum

$$\mathbf{P}_{\text{tot}} = -i\hbar\big(\nabla_{\mathbf{r}_e} + \nabla_{\mathbf{r}_p}\big)$$

commutes with $\hat{H}_{\text{mol}}$.

(b) Center-of-mass and relative coordinates. Introduce

$$M = m_e + m_N, \qquad \mu = \frac{m_e m_p}{M},$$

and define

$$\mathbf{R} = \frac{m_e\mathbf{r}_e + m_p\mathbf{r}_p}{M}, \qquad \mathbf{r} = \mathbf{r}_e - \mathbf{r}_p.$$

Derive the inverse relations

$$\mathbf{r}_e = \mathbf{R} + \frac{m_p}{M}\mathbf{r}, \qquad \mathbf{r}_p = \mathbf{R} - \frac{m_e}{M}\mathbf{r}.$$

(c) Transform the kinetic energy (exact separation). Using the chain rule, show that

$$\nabla_{\mathbf{r}_e} = \frac{m_e}{M}\nabla_{\mathbf{R}} + \nabla_{\mathbf{r}}, \qquad \nabla_{\mathbf{r}_p} = \frac{m_p}{M}\nabla_{\mathbf{R}} - \nabla_{\mathbf{r}}.$$

Define operators ("conjugate momenta")

$$\mathbf{P} = -i\hbar\nabla_{\mathbf{R}}, \qquad \mathbf{p} = -i\hbar\nabla_{\mathbf{r}},$$

and show that

$$-\frac{\hbar^2}{2m_e}\nabla_{\mathbf{r}_e}^2 - \frac{\hbar^2}{2m_p}\nabla_{\mathbf{r}_p}^2 = \frac{\mathbf{P}^2}{2M} + \frac{\mathbf{p}^2}{2\mu}.$$

Conclude that the exact Hamiltonian (2.21) becomes

$$\hat{H}_{\text{mol}} = \underbrace{\frac{\mathbf{P}^2}{2M}}_{H_{\text{cm}}} + \underbrace{\left(\frac{\mathbf{p}^2}{2\mu} - \frac{e^2}{4\pi\varepsilon_0\,|\mathbf{r}|}\right)}_{H_{\text{rel}}}. \tag{2.22}$$

(d) Separation of variables and physical meaning. Let $\Psi(\mathbf{R}, \mathbf{r}) = \Phi(\mathbf{R})\psi(\mathbf{r})$. Show that the Schrödinger equation $\hat{H}_{\text{mol}}\Psi = E_{\text{mol}}\Psi$ with (2.22) separates into

$$\hat{H}_{\text{cm}}\Phi = E_{\text{cm}}\Phi, \qquad \hat{H}_{\text{rel}}\psi = E_{\text{rel}}\psi, \qquad E_{\text{mol}} = E_{\text{cm}} + E_{\text{rel}}.$$

Interpret Φ as the free translational motion of the whole atom, and derive the eigenfunctions of this motion, and interpret ψ as the internal (bound-state) motion.

(e) The clamped-nuclei (infinite-mass) approximation. In the clamped-nuclei approximation one replaces the nuclear kinetic energy term by zero and treats the nucleus as fixed at $\mathbf{r}_p = \mathbf{0}$. The approximate Hamiltonian is then

$$\hat{H}_{\text{el}} = -\frac{\hbar^2}{2m_e}\nabla^2_{\mathbf{r}_e} - \frac{e^2}{4\pi\varepsilon_0\,|\mathbf{r}_e|}. \tag{2.23}$$

1. Compare the clamped-nucleus internal Hamiltonian (2.23) with the exact relative Hamiltonian $\hat{H}_{\text{rel}}$ in (2.22). By identifying $\mathbf{r} = \mathbf{r}_e$ when $\mathbf{r}_p = \mathbf{0}$, show that $\hat{H}_{\text{el}}$ is obtained from $\hat{H}_{\text{rel}}$ by the replacement $\mu \to m_e$ (equivalently, the limit $m_p \to \infty$).

2. Explain (in a sentence or two) which physical effects are removed by clamping the nucleus (e.g. recoil / finite nuclear mass, COM motion, and exact translation invariance of the full two-body problem).

(f) Quantifying the error (spectral shift). Let $E_n^{(\mu)}$ denote the bound-state energies of $\hat{H}_{\text{rel}}$ and $E_n^{(e)}$ those of $\hat{H}_{\text{el}}$. Using the known hydrogenic scaling $E_n \propto -(\text{mass})/n^2$, show that

$$\frac{E_n^{(\mu)}}{E_n^{(e)}} = \frac{\mu}{m_e} = \frac{m_p}{m_e + m_p}.$$

Estimate the relative difference

$$\frac{|E_n^{(e)} - E_n^{(\mu)}|}{|E_n^{(e)}|} = 1 - \frac{\mu}{m_e}.$$

(You may use $m_p/m_e \approx 1836$.)

3 Lecture 3: Electron wavefunctions and Hartree–Fock

3.1 Electron wavefunctions

3.1.1 Spin

Recall that in the clamped-nuclei approximation we are seeking an N-electron wavefunction Ψ_{el} that depend on *all* the electron coordinates. However, before we discuss this wavefunction in detail, we must mention *spin*.

An electron does not only have a position in space $\mathbf{r}$, but also an intrinsic degree of freedom called *spin* . All quantum particles have spin, which takes on half-integral or integral value s. Electrons have spin $s = 1/2$, and are therefore fermions. The spin *coordinate* takes $2s + 1 = 2$ discrete values. For the electron, these values are typically denoted $\omega \in \{\alpha, \beta\}$. The notation varies, sometimes it is $\{\uparrow, \downarrow\}$ or $\{+\frac{1}{2}, -\frac{1}{2}\}$, etc. In the latter case, the spin variable is associated with the *projection of electron spin along a coordinate axis*, typically the z axis.

Particles with integral spin are called *bosons*, while particles with half-integral spin are called *fermions*.

Thus, the electron has a generalized position $x = (\mathbf{r}, \omega)$ in an extended configuration space being equal to two copies of $\mathbb{R}^3$. The wavefunction is $\psi(x)$, and can be written in numerous ways:

$$\psi(x) = \psi(\mathbf{r}, \omega) = \psi_\omega(\mathbf{r}).$$

Picking $\omega = \alpha$, or "$\uparrow$", $\psi_\alpha(\mathbf{r})$ is the probability amplitude for having α-spin at $\mathbf{r}$, while $\psi_\beta(\mathbf{r})$ is the probabiltiy amplitude for having β-spin, or "$\downarrow$", at $\mathbf{r}$.

At a given position $\mathbf{r}$, the electron wavefunction is therefore a two-component vector, a *spinor* $|u\rangle$:

$$\begin{aligned} |u\rangle &= |\alpha\rangle \langle\alpha|u\rangle + |\beta\rangle \langle\beta|u\rangle \\ &= u_\alpha |\alpha\rangle + u_\beta |\beta\rangle \end{aligned} \tag{3.1}$$

where $|\alpha\rangle$ and $|\beta\rangle$ are an *orthonormal basis* for this spinor space. It is customary – but not necessary – to associate each of these basis spinors with the standard basis in $\mathbb{C}^2$, i.e.,

$$|\alpha\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad |\beta\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{3.2}$$

Thus,

$$|u\rangle = \begin{bmatrix} u_\alpha \\ u_\beta \end{bmatrix}. \tag{3.3}$$

The coordinates of the electron is the tuple $x = (\mathbf{r}, \omega)$. Thus, the electron lives in "two copies" of space $\mathbb{R}^3$. The total one-electron wavefunction is a function of x, e.g.,

$$\chi(x) = \chi(\mathbf{r}, \omega) = \begin{bmatrix} \chi(\mathbf{r}, \alpha) \\ \chi(\mathbf{r}, \beta) \end{bmatrix} = \begin{bmatrix} \chi_\alpha(\mathbf{r}) \\ \chi_\beta(\mathbf{r}) \end{bmatrix}. \tag{3.4}$$

Usually, since the spin is a discrete quantity, the spin coordinate is placed as a subscript, as on the right hand side, but mathematically it is of course of no consequence where we put the spin coordinate symbol.

3.1.2 *Orbitals and spin-orbitals*

The function $\chi(x)$ from above is a general single-electron wavefunction. However, in quantum chemistry we usually work with single-electron wavefunctions on a particular form – *spin-orbitals*.

Let $\psi(\mathbf{r})$ be a function of a single spatial coordinate only. Such a function is called an *orbital*. By takin the *direct product* of such an orbital with a basis spinor, we obtain a spin-orbital:

$$\chi(x) = \psi(\mathbf{r})\,|\alpha\rangle = \begin{bmatrix} \psi(\mathbf{r}) \\ 0 \end{bmatrix} \tag{3.5}$$

$$\chi(x) = \psi(\mathbf{r})\,|\beta\rangle = \begin{bmatrix} 0 \\ \psi(\mathbf{r}) \end{bmatrix} \tag{3.6}$$

Thus each orbital gives two linearly independent spin-orbitals.

Notice that if we are given two orbitals ψ_1 and ψ_2 that overlap in space,

$$\langle \psi_1 | \psi_2 \rangle = \int \psi_1(\mathbf{r})^* \psi_2(\mathbf{r})\, d\mathbf{r} \neq 0, \tag{3.7}$$

then the two *spin-orbitals* $\chi_1 = \psi_1\,|\alpha\rangle$ and $\chi_2 = \psi_2\,|\beta\rangle$ are *orthogonal* since

$$\langle \chi_1 | \chi_2 \rangle = \langle \psi_1 | \psi_2 \rangle \langle \alpha | \beta \rangle = 0. \tag{3.8}$$

3.1.3 *N-electron wavefunctions*

The wavefunction of a collection of N electrons depends on *all* the electron coordinates:

$$\Psi_{\text{el}}(x_1, x_2, \cdots, x_N). \tag{3.9}$$

This is a very complicated object! If we were to know the wavefunction at, say, 5 points in space for all electrons, we would have 10^N values ($10 = 5 \times 2$ for space and spin). We would quickly run out of pencil lead, paper, or time if we were to write down the wavefunction of, say, a water molecule with $N = 10$ electrons.

The electrons are *indistinguishable* particles. They are in a fundamental manner *identical*. There is no physical property of two

electrons that allows us to say which is which. Our physical theory must reflect this. If not, then our physical theory allows us to design experiments that could distinguish the electrons! Recall that the physical content of the wavefunction is probabilistic. That is,

$$P(x_1, \cdots, x_N) = |\Psi_{\text{el}}(x_1, \cdots, x_N)|^2 \tag{3.10}$$

is the probability density of locating the ensemble of N electrons in the given configuration. This probability density must reflect the indistinguishability of the particles. For example if we switch the location of two particles i and j, the probability remains the same:

$$P(x_1, \cdots, x_i, \cdots, x_j, \cdots, x_N) = P(x_1, \cdots, x_j, \cdots, x_i, \cdots, x_N). \tag{3.11}$$

Repeating such transpositions of particles lead to the fact that if we *permute* particles, the probability density remains the same:

$$P(x_1, \cdots, x_N) = P(x_{\sigma(1)}, \cdots, x_{\sigma(N)}). \tag{3.12}$$

Turning to the wavefunction, we see that we must have

$$\Psi_{\text{el}}(x_1, \cdots, x_N) = c(\sigma)\Psi_{\text{el}}(x_{\sigma(1)}, \cdots, x_{\sigma(N)}), \tag{3.13}$$

where $c(\sigma)$ is a complex with unit modulus, $|c(\sigma)| = 1$. It turns out, that the only possible values of $c(\sigma)$ are $c(\sigma) = 1$ and $c(\sigma) = (-1)^{|\sigma|}$, the sign of σ. Microscopic particles can only have one or the other. A particle type that has $c(\sigma) = 1$ this have *totally symmetric wavefunctions* and are called *bosons*, while a particle type that has $c(\alpha) = (-1)^{|\sigma|}$ has *totally antisymmetric wavefunctions* and are called *fermions*.[1]

The conclusion is the following postulate, which is called *the Pauli principle*: The N-electron wavefunction must be totally antisymmetric:

$$\psi_{\text{el}}(x_1, \cdots, x_N) = (-1)^{|\sigma|}\psi_{\text{el}}(x_{\sigma(1)}, \cdots, x_{\sigma(N)}). \tag{3.14}$$

Recall that a *permutation* $\sigma \in S_N$ is a rearrangement of N items. Thus σ is a function that takes a number $in\{1, \cdots, N\}$ to a number $\sigma(i) \in \{1, \cdots, N\}$, such that no numbers are repeated or left out.

Any permutation $\sigma \in S_N$ can be written as a sequence of transpositions, the number of which is denoted $|\sigma|$. This is even or odd, regardless of the choice of sequence of transpositions. Thus, any permutation has a well-defined *sign*

$$(-1)^{|\sigma|}.$$

[1] In 1d or 2d, it turns out that there is an intermediate case. Such "particles" are called *anyons*. This was proven by the Norwegian physicists Jon Magne Leinaas and Jan Myrheim in the 1970s. Anyons have, apparently, been experimentally detected.

3.2 *Hartree products*

In quantum chemistry, we build N-electron wavefunctions from single-electron wavefunctions. Let $\chi_i(x)$ be a collection of single-electron wavefunctions. The simplest possible N-electron wavefunction we can concieve of is a Hartree product,

$$\Phi^{\text{HP}}(x_1, \cdots, x_N) = \chi_1(x_1)\chi_2(x_2)\cdots\chi_N(x_N). \tag{3.15}$$

The Hartree product was used by Douglas R. Hartree when he concieved the Hartree method in 1927–1928, the forerunner of the Hartree–Fock method. Of course, the Hartree product is not antisymmetric, so it is not an admissible N-electron wavefunction.[2] However, if we compute the probability density for the Hartree product we obtain (with $N = 2$ for simplicity)

$$P(x_1, x_2) = P_1(x)P_2(x_2), \quad P_i(x) = |\chi_i(x)|^2. \tag{3.16}$$

Douglas R. Hartree (1897–1958)

[2] Electron spin was discovered in 1925, and the Pauli principle was formulated around the same time. However, the Hartree method was the very first quantitative numerical method for electronic systems.

This is the mathematical definition of the probability densities of the random variables x_1 and x_2 being independent: The joint probability distribution is a product of individual distributions. We conclude, that Φ^{HP} describes *independent particles*, even if the wavefuncion is not admissible for electrons.

However, particles with a Hartree product wavefunction are in general distinguishable. The *marginal probabilities* are

$$\begin{aligned} P(x_1 \mid x_2 \text{ anywhere}) &= \int dx_2 P(x_1, x_2) = \int dx_2 P_1(x_1) P_2(x_2) \\ &= P_1(x_1). \end{aligned} \tag{3.17}$$

Similarly,

$$P(x_2 \mid x_1 \text{ anywhere}) = P_2(x_2). \tag{3.18}$$

Since these two functions are not necessarily the same, electrons 1 and 2 have different marginal probability densities, and are hence possible to distinguish. If all the $\chi_i(x)$ are the same functions will the marginal probabilities all be the same. This is possible for bosons, but not for fermions.

We conclude that fermions are not truly independent particles due to antisymmetry.

3.3 Slater determinants

To correct for violation of antisymmetry for the $N = 2$ particle case, we conssider the *Slater determinant*,

$$\Phi^{\mathrm{SD}}(x_1, x_2) = \frac{1}{\sqrt{2}} (\chi_1(x_1)\chi_2(x_2) - \chi_2(x_1)\chi_1(x_2)). \tag{3.19}$$

It is easy to verify that this function is antisymmetric. The general N-electron Slater determinant is defined to be

$$\Phi^{\mathrm{SD}}(x_1, \cdots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \cdots & \chi_1(x_N) \\ \vdots & \ddots & \vdots \\ \chi_N(x_1) & \cdots & \chi_N(x_N) \end{vmatrix}. \tag{3.20}$$

For a given choice of the positions, the Slater determinant is thus the determinant of the matrix X with elements $X_{ij} = \chi_i(x_j)$.

Let us compute the probability density of the two electron determinant:

$$\begin{aligned} P(x_1, x_2) = \frac{1}{2} &(P_1(x_1)P_2(x_2) + P_2(x_1)P_1(x_2) \\ &- Q_{12}(x_1)Q_{12}(x_2)^* - Q_{12}(x_1)^* Q_{12}(x_2)), \end{aligned} \tag{3.21}$$

where $Q_{12}(x) = \chi_1(x)^*\chi_2(x)$. This probability density is more complicated than for the Hartree product. In particular *it does not describe x_1 and x_2 as having independent probability densities!* The electrons are *correlated* due to the antisymmetry of the wavefunction. This is called *exchange correlation*.[3]

It is still customary to say that the Slater determinant is "uncorrelated", and that we have an "independent particle model". This

[3] Not to be confused with "exchange-correlation potential" in density-functional theory!

is because the Slater determinant is the *least correlated* N-electron wavefunction we can think of, and that the electrons are *as independent as possible*.

3.3.1 *Two electrons, opposite spin*

Let us consider two special cases. We begin with two spin-orbitals of opposite spin:

$$\chi_1(x) = \psi_1(\mathbf{r})\,|\alpha\rangle\,, \tag{3.22}$$
$$\chi_2(x) = \psi_2(\mathbf{r})\,|\beta\rangle\,. \tag{3.23}$$

We place these in a 2-electron Slater determinant. Computing marhonal probabilities, This gives:

$$P_1(\mathbf{r},\omega) = |\psi_1(\mathbf{r})|^2\delta_{\omega,0}, \quad P_2(\mathbf{r},\omega) = |\psi_2(\mathbf{r})|^2\delta_{\omega,1}, \quad Q_{12}(\mathbf{r},\omega) = 0. \tag{3.24}$$

We now consider the marginal probability density of locating the electrons in *space*, regardless of *spin* direction:

$$P(\mathbf{r}_1,\mathbf{r}_2) = \sum_{\omega_1,\omega_2} P(\mathbf{r}_1,\omega_1,\mathbf{r}_2,\omega_2) = \frac{1}{2}(P_1(\mathbf{r}_1)P_2(\mathbf{r}_2) + P_2(\mathbf{r}_1)P_1(\mathbf{r}_2)), \tag{3.25}$$

where $P_i(\mathbf{r}) = \sum_\omega P_i(\mathbf{r},\omega) = |\psi_i(\mathbf{r})|^2$.

The probability density for spatial location of the electrons is now properly symmetric in the two coordinates. In particular it is correlated – the particles are not independent. Moreover, the marginal probabilities are seen to be the same for each electron 1 and 2:

$$P(\mathbf{r}_1 \mid \mathbf{r}_2 \text{ anywhere}) = \int P(\mathbf{r}_1,\mathbf{r}_2)\,d\mathbf{r}_2 = \frac{1}{2}P_1(\mathbf{r}_1) + \frac{1}{2}P_2(\mathbf{r}_1). \tag{3.26}$$

3.3.2 *Two electrons, collinear spins*

The next example uses collinear spins:

$$\chi_1(x) = \psi_1(\mathbf{r})\,|\alpha\rangle\,, \tag{3.27}$$
$$\chi_2(x) = \psi_2(\mathbf{r})\,|\alpha\rangle\,. \tag{3.28}$$

$$P_1(\mathbf{r},\omega) = |\psi_1(\mathbf{r})|^2\delta_{\omega,0}, \quad P_2(\mathbf{r},\omega) = |\psi_2(\mathbf{r})|^2\delta_{\omega,0}, \tag{3.29}$$
$$Q_{12}(\mathbf{r},\omega) = \psi_1(\mathbf{r})^*\psi_2(\mathbf{r})\delta_{\omega,0}. \tag{3.30}$$

We now consider the marginal probability density of locating the electrons in *space*, regardless of *spin* direction:

$$\begin{aligned} P(\mathbf{r}_1,\mathbf{r}_2) &= \sum_{\omega_1,\omega_2} P(\mathbf{r}_1,\omega_1,\mathbf{r}_2,\omega_2) \\ &= \frac{1}{2}(P_1(\mathbf{r}_1)P_2(\mathbf{r}_2) + P_2(\mathbf{r}_1)P_1(\mathbf{r}_2) \\ &\quad - Q_{12}(\mathbf{r}_1)^*Q_{12}(\mathbf{r}_2) - Q_{12}(\mathbf{r}_1)Q_{12}^*(\mathbf{r}_2)), \end{aligned} \tag{3.31}$$

where $Q_{12}(\mathbf{r}) = \psi_1(\mathbf{r})^*\psi_2(\mathbf{r})$.

We now observe, that if we set $\mathbf{r}_1 = \mathbf{r}_2$, we obtain

$$P(\mathbf{r}_1, \mathbf{r}_1) = 0, \tag{3.32}$$

so antisymmetry forces the probabilty of two electrons occupying both the same space and spin coordinates to vanish. This phenomenon is called a *Fermi hole*. The Fermi hole did not appear for opposite spins.

3.4 Configuration-interaction wavefunctions

The Slater determinant is certainly the simplest N-electron wavefunction we can imagine that respects the Pauli principle. In order to make a Slater determinant we need N single-electron wavefunctions χ_1 through χ_N. We assume these to be orthonormal, $\langle \chi_p | \chi_q \rangle = \delta_{pq}$. We write the Slater determinant as

$$|\chi_1 \chi_2 \cdots \chi_N\rangle \,. \tag{3.33}$$

Suppose now we have at our disposal in total K single-electron functions χ_p. From this, we can make many Slater determinants:

$$|\chi_{p_1} \chi_{p_2} \cdots \chi_{p_N}\rangle \,, \quad 1 \leq p_i \leq K. \tag{3.34}$$

However, by antisymmetry of the determinant in Eq. (3.20), permuting the χ_{p_i}s only gives a sign prefactor. Thus, the Slater determinant shown is uniquely given by the *set* of indices $\{p_i \mid i = 1, 2, \ldots, N\}$. To see this, considered first the case where the indices p_i are ordered:

For example, $|\chi_1 \chi_4 \chi_5\rangle = -|\chi_4 \chi_1 \chi_5\rangle$.

$$p_1 < p_2 < \cdots < p_N. \tag{3.35}$$

Any permutation leads to an index set that represents the *same* determinant. The ordered index set is clearly a *selection* of N elements from $\{1, 2, \cdots, K\}$.

There are $\binom{K}{N}$ possible selections of N elements out of a set of size K:

$$\binom{K}{N} = \frac{K!}{N!(K - N!)}. \tag{3.36}$$

3.4.1 Taking linear combinations

The determinants $|\chi_{p_1} \cdots \chi_{p_N}\rangle$ are *orthonormal*. This is an exercise to show. Thus, we can use them as *basis for a vector space*, i.e., we can consider an N-electron wavefunction on the form

$$|\Psi^{\mathrm{CI}}\rangle = \sum_{\substack{p_1=1 \\ p_1<p_2<\cdots<p_N}}^{K} \sum_{p_2=1}^{K} \cdots \sum_{p_N-1}^{K} c_{p_1 \cdots p_N} |\chi_{p_1} \cdots \chi_{p_N}\rangle \,. \tag{3.37}$$

This wavefunction is called the *full configuration-interaction (FCI) wavefunction*. It lives in a Hilbert space of dimension $\binom{K}{N}$. We will have more to say about the CI wavefunction later.

3.4.2 *Occupation number representation*

Since the Slater determinant $|\chi_{p_1}\cdots\chi_{p_N}\rangle$ is given by a subset of the K first integers, we can code the Slater determinant as a bit pattern of length K, where bit number p is lit if $p = p_i$ for some i:

$$|\chi_{p_1}\cdots\chi_{p_N}\rangle = |n_1 n_2 \cdots n_K\rangle\,, \quad n_p = \sum_i \delta_{p,p_i}. \tag{3.38}$$

For example, with $K = 8$,

$$|\chi_1\chi_2\rangle = |11000000\rangle\,. \tag{3.39}$$

$$|\chi_1\chi_8\rangle = |10000001\rangle\,. \tag{3.40}$$

$$|\chi_1\chi_5\chi_8\rangle = |10001001\rangle\,. \tag{3.41}$$

The numbers n_p are called *occupation numbers*. Note, however, that permutation of the χs do not make sense using occupation numbers. This is because we are directly encoding the *set* of indices, and not the individual indices with ordering! Thus, for example,

$$|\chi_2\chi_1\rangle = -|\chi_1\chi_2\rangle = -|11000000\rangle\,. \tag{3.42}$$

As a final comment, note that the occupation number formalism does not care about the number of total electrons – the same notation is valid for any number N of electrons. In particular it makes sense to insist that two slater determinants in occupation number formalism are orthogonal if they have differing number of electrons. There is a special determinant called *the vacuum* where there are no electrons:

$$|-\rangle = |00000\ldots\rangle \tag{3.43}$$

This construction, of Slater determinants with arbitrary number of electrons, includig zero, is called *Fock space*.

3.5 *Introduction to the Hartree–Fock method*

3.5.1 *The Hartree–Fock ansatz*

The idea of the Hartree–Fock method is to consider a single Slater determinant with single-electron wavefunctions χ_i, $i = 1, 2, \ldots, N$, and treeat these as unknowns. By minimizing the *Hartree–Fock energy* with respect to all the χ_i, an estimate for the ground-state wavefunction is obtained.

We now describe the General Hartree–Fock method, where the single-electron functions are completely general.

Recall that our goal is to solve the N-electron Schödinger equation in the clamped-nuclei approximation:

$$\hat{H}_{\text{el}}\,|\psi_{\text{el}}\rangle = E_{\text{el}}\,|\psi_{\text{el}}\rangle\,. \tag{3.44}$$

We now take the approximation

$$|\psi_{\text{el}}\rangle \approx |\Psi^{SD}\rangle = |\chi_1\cdots\chi_N\rangle\,, \quad \langle\chi_i|\chi_j\rangle = \delta_{ij}. \tag{3.45}$$

We require the energy expectation to be minimal:

$$E_{\mathrm{HF}} = \frac{\langle \Psi^{\mathrm{SD}} | \hat{H}_{\mathrm{el}} | \Psi^{\mathrm{SD}} \rangle}{\langle \Psi^{\mathrm{SD}} | \Psi^{\mathrm{SD}} \rangle} = \text{min!}. \tag{3.46}$$

Recall that the electronic Hamiltonian is

$$\begin{aligned} \hat{H}_{\mathrm{el}} &= \hat{T}_{\mathrm{el}} + \hat{V}_{\text{nuc-el}} + \hat{V}_{\text{el-el}} \\ &= \sum_{i=1}^{N} \frac{-1}{2} \nabla_i^2 + \sum_{i=1}^{N} V_{\mathrm{nuc}}(\mathbf{r}_i) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ &= \sum_{i=1}^{N} \hat{h}(i) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} \hat{w}(i,j) \end{aligned} \tag{3.47}$$

The operator $\hat{h}$ is a one-electron operator,

$$\hat{h} = \frac{-1}{2} \nabla^2 + V_{\mathrm{nuc}}(\mathbf{r}). \tag{3.48}$$

and $h(i)$ indicates that it is acting on the coordinates of electron i. Similarly, $\hat{w}$ is a two-electron operator.

That the energy is a minimum requires the *Hartree–Fock equations* to be fulfiled:

$$\hat{f} \, |\chi_i\rangle = \epsilon_i \, |\chi_i\rangle \, . \tag{3.49}$$

Here, $\hat{f}$ is the *Fock operator* given by

$$\begin{aligned} \hat{f} &= \hat{h} + \hat{v}^{\mathrm{HF}} \\ &= \frac{-1}{2} \nabla^2 + V_{\mathrm{nuc}}(\mathbf{r}) + V_{\mathrm{H}}(\mathbf{r}) - V_{\text{exchange}}. \end{aligned} \tag{3.50}$$

Here, the *Hartree potential* is

$$V_{\mathrm{H}}(\mathbf{r}) = \int d\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \tag{3.51}$$

where the *electron density* is

$$\rho(\mathbf{r}) = \sum_{j=1}^{N} \sum_{\omega} |\chi_j(\mathbf{r}, \omega)|^2, \tag{3.52}$$

and the *exchange potential* is a *nonlocal potential operator* defined by its action on a spatial orbital $\psi(\mathbf{r})$ as

$$[V_{\text{exchange}} \psi](\mathbf{r}, \omega) = \int d\mathbf{r}' \sum_{j=1}^{N} \sum_{\omega'} \frac{\psi(\mathbf{r}', \omega') \chi_j(\mathbf{r}', \omega')^*}{|\mathbf{r} - \mathbf{r}'|} \chi_j(\mathbf{r}, \omega). \tag{3.53}$$

While complicated-looking, these definitions can be interpreted. The Hartree potential appears already in the Hartree method (where we minimize the energy of the Hartree product as opposed to the Slater determinant)W, and is the electrostatic potential at point $\mathbf{r}$ in space due to the charge density $\rho(\mathbf{r}')$. This is the charge density obtained by assuming that the electrons independently set up a charge density $\sum_\omega |\chi_j(\mathbf{r}, \omega)|^2$. The total density is the sum of the individual electrons' density.[4]

However, we know that the electrons are indistinguishable. They do not have individuality, and therefore we have to antisymmetrize

[4] One can define the *density matrix*

$$\gamma(\mathbf{r}, \mathbf{r}') = \sum_{j=1}^{N} \sum_{\omega} \chi_j(\mathbf{r}, \omega) \chi_j(\mathbf{r}', \omega)^*$$

. The density is then $\rho(\mathbf{r}) = \gamma(\mathbf{r}, \mathbf{r})$, and the exchange potential can also be written in terms of γ.

the Hartree product into a Slater determinant. This leads to the exchange potential, which *does not have a simple classical interpretation*. It is due to bona fide quantum effects in the electron system. From the definition (3.53) we can see that the operator is *non-local*, smearing $\psi(\mathbf{r}, \omega)$ out in space via the integral.

Later in the course, we will derive the Hartree–Fock equations using *calculus of variations*.

3.5.2 *Restricted, urestricted, and general HF*

We have outlined *General Hartree–Fock* (GHF), where the Slater determinant was composed of *general single-electron fucntions* $\chi_i(x)$.

In *Restricted Hartree–Fock* (RHF), we instead do the following: We assume that the number of electrons is $2N$, an even number. We then let $\psi_i(\mathbf{r})$ be N unknown spatial orbitals, and let

$$\chi_{2j-1} = \psi_j |\alpha\rangle, \quad \chi_{2j} = \psi_j |\beta\rangle. \tag{3.54}$$

To compactly denote such a Slater determinant, we write

$$|\chi_1 \cdots \chi_{2N}\rangle = |\psi_1 \overline{\psi_1} \psi_2 \overline{\psi_2} \cdots \psi_N \overline{\psi_N}\rangle. \tag{3.55}$$

Thus, the bars denote β-spin, and no bar denotes α-spin.

In *Unrestricted Hartree–Fock* (UHF) we release the spatial orbitals for α and β-spins. Thus we have $2N$ spatial orbitals ψ_i^α and ψ_i^β:

$$|\chi_1 \cdots \chi_{2N}\rangle = |\psi_1^\alpha \overline{\psi_1^\beta} \psi_2^\alpha \overline{\psi_2^\beta} \cdots \psi_N^\alpha \overline{\psi^\beta{}_N}\rangle. \tag{3.56}$$

What is the difference between UHF and GHF?

Thus, in RHF we *restrict* the spatial orbitals to be identical, while in UHF they can be different.

In RHF, the electron density becomes

$$\rho(\mathbf{r}) = 2 \sum_{j=1}^{N} |\psi_j(\mathbf{r})|^2. \tag{3.57}$$

In UHF, the electron density has two parts, coming from the α and β spins, respectively

$$\begin{aligned} \rho(\mathbf{r}) &= \rho^\alpha(\mathbf{r}) + \rho^\beta(\mathbf{r}) \\ &= \sum_{j=1}^{N} |\psi_j^\alpha(\mathbf{r})|^2 + \sum_{j=1}^{N} |\psi_j^\beta(\mathbf{r})|^2 \end{aligned} \tag{3.58}$$

3.5.3 *H_2 in a minimal basis*

As a very simple example of RHF, we can consider the H_2 molecule in a "minimal basis". Read about this in SO.

The main idea illustrates the LCAO ansatz for Hartree–Fock, where we are given a set of K *atomic orbitals* (AOs), and write the Hartree–Fock solution (molecular orbital, MO) as a linear combination:

$$\psi_i(\mathbf{r}) = \sum_{\mu=1}^{K} \phi_\mu(\mathbf{r}) C_{\mu i}. \tag{3.59}$$

The unknowns become the elements of the *MO coefficient matrix* $C = [C_{\mu i}]$.

Denote the atoms A and B, located at $-\mathbf{R}/2$ and $+\mathbf{R}/2$, respectively, totaling an internuclear distance $R = |\mathbf{R}|$. The orientation of $\mathbf{R}$ does not matter; let $\mathbf{R} = R\mathbf{e}_x = (R, 0, 0)^T$ for simplicity and concreteness.

Each H atom is assigned a single AO, an 1s function ϕ_A and ϕ_B, respectively:

$$\phi_A(\mathbf{r}) = \phi(\mathbf{r} - \mathbf{R}/2), \quad \phi_B(\mathbf{r}) = \phi(\mathbf{r} + \mathbf{R}/2). \tag{3.60}$$

Since we have 2 electrons, we have $N = 1$ in RHF, and seek a *single* MO $\psi(\mathbf{r})$, doubly occupied in the Slater determinant

$$|\Psi^{RHF}\rangle = |\psi\overline{\psi}\rangle\,. \tag{3.61}$$

This MO is

$$\psi(\mathbf{r}) = c_A\phi_A(\mathbf{r}) + c_B\psi(\mathbf{r}). \tag{3.62}$$

Since H_2 has inversion symmetry[5] we know that the MO of this system must be an eigenfunction of inversion. This happens precisely when $c_A = c_B$ or $c_A = -c_B$. The first case has the lowest energy due to not having nodes.

[5] The molecule is inversion symmetric about the origin if the nuclei are permuted (and not changing type) by the map $\mathbf{R}_A \mapsto -\mathbf{R}_A$, for *all* nuclei A.

The size of c_A is given by normalization of ψ:

$$\langle\psi|\psi\rangle = \langle c_A\phi_A + c_B\phi_B|c_A\phi_A + c_B\phi_B\rangle = |c_A|^2(2 + 2\langle\phi_A|\phi_B\rangle) = 1, \tag{3.63}$$

where we assumed that ϕ_A was a real function.

In conclusion, we could solve the RHF problem for the hydrogen molecule in a minimal basis by appealing to symmetry alone.

3.6 Exercises

3.6.1 Exercise 1: Spin operators for a single electron (spin-1/2)

Use the standard spin basis $\{|\alpha\rangle, |\beta\rangle\}$, defined by

$$\hat{S}_z |\alpha\rangle = \frac{\hbar}{2} |\alpha\rangle, \qquad \hat{S}_z |\beta\rangle = -\frac{\hbar}{2} |\beta\rangle.$$

The Pauli matrices are

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and the spin operators are

$$\hat{S}_x = \frac{\hbar}{2}\sigma_x, \qquad \hat{S}_y = \frac{\hbar}{2}\sigma_y, \qquad \hat{S}_z = \frac{\hbar}{2}\sigma_z.$$

Define raising/lowering operators

$$\hat{S}_\pm = \hat{S}_x \pm i\hat{S}_y.$$

A general normalized spinor is written as

$$|\psi\rangle = u_\alpha |\alpha\rangle + u_\beta |\beta\rangle, \qquad |u_\alpha|^2 + |u_\beta|^2 = 1.$$

Problem 1: Operators in the $\{|\alpha\rangle, |\beta\rangle\}$ basis

In this course we treat spin as a quantum degree of freedom: states are vectors and measurable quantities are operators. The Pauli matrices provide a convenient representation of the spin operators for a single electron in the $|\alpha\rangle, |\beta\rangle$ basis.

1. Write the explicit 2×2 matrices for $\hat{S}_x, \hat{S}_y, \hat{S}_z$.
2. Construct $\hat{S}_+$ and $\hat{S}_-$ as matrices, and simplify them.

Problem 2: Action on basis states

Let us see what spin does to basis kets. In particular, $\hat{S}_x$ and $\hat{S}_y$ mix $|\alpha\rangle$ and $|\beta\rangle$, while the ladder operators $\hat{S}_\pm$ move between $\hat{S}_z$ eigenstates.

1. Compute $\hat{S}_x |\alpha\rangle$, $\hat{S}_x |\beta\rangle$, $\hat{S}_y |\alpha\rangle$, $\hat{S}_y |\beta\rangle$.
2. Verify the ladder relations:

$$\hat{S}_+ |\beta\rangle = \hbar |\alpha\rangle, \quad \hat{S}_+ |\alpha\rangle = 0, \quad \hat{S}_- |\alpha\rangle = \hbar |\beta\rangle, \quad \hat{S}_- |\beta\rangle = 0.$$

Problem 3: Commutation relations

Spin is a form of quantum mechanical angular momentum with no classical counterpart. In quantum mechanics, different components of angular momentum operators cannot, in general, be simultaneously determined with certainty. This appears mathematically as nonzero commutators.

1. Using matrix multiplication, evaluate $[\hat{S}_x, \hat{S}_y]$, $[\hat{S}_y, \hat{S}_z]$, and $[\hat{S}_z, \hat{S}_x]$.

2. Show that
$$[\hat{S}_i, \hat{S}_j] = i\hbar\, \varepsilon_{ijk} \hat{S}_k.$$

3. Show that $[\hat{S}_z, \hat{S}_\pm] = \pm\hbar \hat{S}_\pm$ and $[\hat{S}_+, \hat{S}_-] = 2\hbar \hat{S}_z$.

Problem 4: The total spin operator $\hat{S}^2$

The operator $\hat{S}^2$ corresponds to the total spin angular momentum (its magnitude). For an electron the total spin quantum number is fixed to $s = \frac{1}{2}$. This problem shows how that emerges from the operator definitions.

Define $\hat{S}^2 = \hat{S}_x^2 + \hat{S}_y^2 + \hat{S}_z^2$.

1. Compute $\hat{S}^2$ explicitly as a 2×2 matrix.

2. Apply $\hat{S}^2$ to $|\alpha\rangle$ and $|\beta\rangle$, and find the eigenvalue.

3. Compare your result to $s(s+1)\hbar^2$ and identify s.

Problem 5: Expectation values for a general spinor

Even though a measurement of $\hat{S}_z$ yields only $\pm\hbar/2$, a general state $|\psi\rangle$ can have nonzero expectation values $\langle \hat{S}_x \rangle$, $\langle \hat{S}_y \rangle$, and $\langle \hat{S}_z \rangle$. These expectation values behave like components of a 3D vector and are often visualized on something called *the Bloch sphere*.

Let $|\psi\rangle = u_\alpha |\alpha\rangle + u_\beta |\beta\rangle$ with $|u_\alpha|^2 + |u_\beta|^2 = 1$.

1. Derive formulas for $\langle \hat{S}_x \rangle$, $\langle \hat{S}_y \rangle$, $\langle \hat{S}_z \rangle$ in terms of u_α and u_β.

2. Show that
$$\| \langle \hat{\mathbf{S}} \rangle \| \leq \frac{\hbar}{2},$$
and determine the condition(s) on u_α, u_β for equality.

3. Any normalized spinor has *two real degrees of freedom*, because normalization of $|\psi\rangle$ removes one degree of freedom and an overall global phase $e^{i\gamma}$ multiplying $|\psi\rangle$ does not change measurement probabilities. A standard representation of the remaining two degrees of freedom is
$$u_\alpha = \cos(\theta/2), \qquad u_\beta = e^{i\phi} \sin(\theta/2),$$
where $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. Here θ controls the relative weights of $|\alpha\rangle$ and $|\beta\rangle$, while ϕ is their relative phase. Using this parameterization, express $\langle \hat{\mathbf{S}} \rangle$ in terms of θ, ϕ, and interpret (θ, ϕ) as spherical angles for the direction of $\langle \hat{\mathbf{S}} \rangle$. Here you are encouraged to make a sketch. (This exercise point may be a little tough. We will see!)

Problem 6: Eigenstates of $\hat{S}_x$ and $\hat{S}_y$

The basis $|\alpha\rangle, |\beta\rangle$ is special because it diagonalizes $\hat{S}_z$. States with definite spin along x or y are different superpositions of $|\alpha\rangle$ and $|\beta\rangle$. This problem connects "measuring along a different axis" to "changing basis."

1. Find normalized eigenstates of $\hat{S}_x$ and write them as linear combinations of $|\alpha\rangle$ and $|\beta\rangle$.
2. Do the same for $\hat{S}_y$.
3. For the $\hat{S}_x$ eigenstate with eigenvalue $+\hbar/2$, compute the probabilities of measuring $\hat{S}_z = \pm\hbar/2$.
4. Write the original basis of $\hat{S}_z$ eigenstates in terms of the $\hat{S}_x$ eigenstates and the $\hat{S}_y$ eigenstates.

Optional: Rotation about the z-axis

In quantum mechanics, rotations are "generated" by angular momentum operators. For spin-1/2, $\hat{S}_z$ generates rotations about the z-axis. The unitary operator

$$\hat{U}_z(\varphi) = \exp\left(-\frac{i}{\hbar}\varphi\hat{S}_z\right)$$

means "rotate the spin state by angle φ around the z-axis." Because $|\alpha\rangle$ and $|\beta\rangle$ are eigenstates of $\hat{S}_z$, they acquire simple phase factors under this rotation.

1. Apply $\hat{U}_z(\varphi)$ to $|\psi\rangle = u_\alpha|\alpha\rangle + u_\beta|\beta\rangle$. Show explicitly that each component acquires a phase, and identify those phases.
2. Explain why this rotation changes the *relative* phase between the α and β components (even though a global phase would not matter).
3. Connect to Problem 5(3): show that changing the relative phase corresponds to changing the angle ϕ while keeping θ fixed (i.e., a rotation around the z-axis on the Bloch sphere).

3.6.2 *Exercise 2: One-electron spinor wavefunctions and spin expectation values*

In this exercise you will work with a general one-electron (two-component) spinor wavefunction written as

$$\chi(x) = \chi(\mathbf{r},\omega) = \begin{bmatrix}\chi_\alpha(\mathbf{r})\\ \chi_\beta(\mathbf{r})\end{bmatrix}, \qquad \omega \in \{\alpha,\beta\}, \quad x = (\mathbf{r},\omega).$$

Here $\chi_\alpha(\mathbf{r})$ and $\chi_\beta(\mathbf{r})$ are complex-valued spatial functions (spin components).

The inner product between two 1-electrons wavefunctions is

$$\langle \chi | \eta \rangle = \int dx \, \chi(x)^* \eta(x),$$

where the integral is short for

$$\int dx = \sum_{\omega} \int d\mathbf{r},$$

which is mathematically well defined, but its proper definition is beyond the scope of our course.

Spin operators act on the spin index only. See Exercise 1.

Problem 1: From $\chi(x)$ to component form

In quantum chemistry we usually separate the spatial coordinate $\mathbf{r}$ from the spin coordinate ω. A two-component spinor χ is just a compact way to store the two spatial functions associated with $\omega = \alpha$ and $\omega = \beta$.

1. Show that the inner product of two spinors χ and η can be written as
$$\langle \chi | \eta \rangle = \int \left(\chi_\alpha^*(\mathbf{r}) \eta_\alpha(\mathbf{r}) + \chi_\beta^*(\mathbf{r}) \eta_\beta(\mathbf{r}) \right) d\mathbf{r}.$$
2. Write explicitly what $\chi(\mathbf{r}, \alpha)$ and $\chi(\mathbf{r}, \beta)$ mean in terms of $\chi_\alpha(\mathbf{r})$ and $\chi_\beta(\mathbf{r})$.
3. Suppose $\chi_\beta(\mathbf{r}) = 0$ everywhere. What spin state does χ correspond to? What if $\chi_\alpha(\mathbf{r}) = 0$ everywhere?

Problem 2: Probability of measuring $m_s = \pm\hbar/2$

Even when the spinor depends on $\mathbf{r}$, a Stern–Gerlach type experiment/measurement of $\hat{S}_z$ still yields only $\pm\hbar/2$. The spatial dependence affects the *probabilities* through the norms of the components.

1. Show that the probability of measuring $+\hbar/2$ (i.e. finding $\omega = \alpha$) is
$$P_\alpha = \int |\chi_\alpha(\mathbf{r})|^2 \, d\mathbf{r},$$
and similarly
$$P_\beta = \int |\chi_\beta(\mathbf{r})|^2 \, d\mathbf{r}.$$
2. Use normalization to show that $P_\alpha + P_\beta = 1$.
3. Interpret P_α and P_β in the special case where $\chi_\alpha(\mathbf{r}) = u_\alpha \varphi(\mathbf{r})$ and $\chi_\beta(\mathbf{r}) = u_\beta \varphi(\mathbf{r})$ with a normalized spatial orbital φ.

Problem 3: Expectation values of $\hat{S}_x$, $\hat{S}_y$, $\hat{S}_z$

In Exercise 1 you computed $\langle \hat{S}_i \rangle$ for a constant spinor $(u_\alpha, u_\beta)^T$. Here the "coefficients" are functions of $\mathbf{r}$, so expectation values become integrals. The cross terms $\chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r})$ encode the *relative*

phase and overlap of the two components. Recall that the expectation value of an operator $\hat{A}$ is

$$\langle \hat{A} \rangle = \langle \chi | \hat{A} | \chi \rangle .$$

1. Starting from $\langle \hat{S}_z \rangle = \langle \chi | \hat{S}_z | \chi \rangle$, show that

$$\langle \hat{S}_z \rangle = \frac{\hbar}{2} \int \Big(|\chi_\alpha(\mathbf{r})|^2 - |\chi_\beta(\mathbf{r})|^2 \Big)\, d\mathbf{r}.$$

2. Show that

$$\langle \hat{S}_x \rangle = \frac{\hbar}{2} \int \Big(\chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r}) + \chi_\beta^*(\mathbf{r}) \chi_\alpha(\mathbf{r}) \Big)\, d\mathbf{r} = \hbar \, \mathrm{Re} \int \chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r})\, d\mathbf{r}.$$

3. Show that

$$\langle \hat{S}_y \rangle = \frac{\hbar}{2} \int \Big(-i\, \chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r}) + i\, \chi_\beta^*(\mathbf{r}) \chi_\alpha(\mathbf{r}) \Big)\, d\mathbf{r} = \hbar \, \mathrm{Im} \int \chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r})\, d\mathbf{r}.$$

4. Define

$$p(\mathbf{r}) = |\chi_\alpha(\mathbf{r})|^2 + |\chi_\beta(\mathbf{r})|^2.$$

   Explain why $p(\mathbf{r})$ is the probability density for finding the electron at position $\mathbf{r}$ (regardless of spin).

Problem 4: The local spin polarization and magnetization density

In quantum chemistry (and especially density-functional theory) it is common to describe spin information in terms of densities in real space. For a single electron spinor, the difference between α and β densities measures the spin polarization along z at each point.

1. Define the spin-resolved densities

$$\rho_\alpha(\mathbf{r}) = |\chi_\alpha(\mathbf{r})|^2, \qquad \rho_\beta(\mathbf{r}) = |\chi_\beta(\mathbf{r})|^2.$$

   Show that $\int \rho_\alpha(\mathbf{r})\, d\mathbf{r} = P_\alpha$ and $\int \rho_\beta(\mathbf{r})\, d\mathbf{r} = P_\beta$.

2. Define the z-spin density

$$s_z(\mathbf{r}) = \frac{\hbar}{2} \big(\rho_\alpha(\mathbf{r}) - \rho_\beta(\mathbf{r}) \big).$$

   Show that $\int s_z(\mathbf{r})\, d\mathbf{r} = \langle \hat{S}_z \rangle$.

3. The transverse components $s_x(\mathbf{r})$ and $s_y(\mathbf{r})$ involve $\chi_\alpha^*(\mathbf{r}) \chi_\beta(\mathbf{r})$. Explain in words what additional information these components contain that ρ_α and ρ_β alone do not.

Problem 5: $\hat{S}^2$ and when a spinor is a pure α or β state

For one electron, $\hat{S}^2$ always has eigenvalue $\frac{3}{4}\hbar^2$ on any spin state. However, a general spinor may or may not be an eigenfunction of $\hat{S}_z$. This problem clarifies the difference between "having a definite S^2" (which is automatic for one electron) and "having a definite S_z" (which depends on the form of χ).

1. Show that $\langle \hat{S}^2 \rangle = \langle \chi | \hat{S}^2 | \chi \rangle = \frac{3}{4}\hbar^2$ for any normalized one-electron spinor χ.

2. Give a condition on $\chi_\alpha(\mathbf{r})$ and $\chi_\beta(\mathbf{r})$ such that χ is an eigenfunction of $\hat{S}_z$. What are the possible eigenvalues?

3. Suppose $\chi_\alpha(\mathbf{r})$ and $\chi_\beta(\mathbf{r})$ are both nonzero. Explain why measuring $\hat{S}_z$ yields $\pm\hbar/2$ probabilistically, and relate the probabilities to Problem 2.

Problem 6: Spinors built from a spatial orbital and a "global" spinor

A common special case is that both spin components share the same spatial shape, differing only by constant complex coefficients. Then the spin physics reduces to Exercise 1, and the spatial orbital acts only as a spectator (provided it is normalized).

Assume

$$\chi_\alpha(\mathbf{r}) = u_\alpha\, \varphi(\mathbf{r}), \qquad \chi_\beta(\mathbf{r}) = u_\beta\, \varphi(\mathbf{r}),$$

where $\int |\varphi(\mathbf{r})|^2\, d\mathbf{r} = 1$ and $|u_\alpha|^2 + |u_\beta|^2 = 1$.

1. Show that χ is normalized.

2. Show that the expectation values $\langle \hat{S}_x \rangle, \langle \hat{S}_y \rangle, \langle \hat{S}_z \rangle$ reduce to the Exercise 1 formulas with (u_α, u_β).

3. Parameterize $u_\alpha = \cos(\theta/2)$ and $u_\beta = e^{i\phi}\sin(\theta/2)$ and interpret the resulting $\langle \hat{\mathbf{S}} \rangle$ in terms of (θ, ϕ).

Optional: Rotation about the z-axis for spatially dependent spinors

A rotation about z changes the *relative phase* between the α and β components. When $\chi_\alpha(\mathbf{r})$ and $\chi_\beta(\mathbf{r})$ depend on $\mathbf{r}$, the rotation still acts only on the spin index and therefore multiplies each component by a phase factor, independent of $\mathbf{r}$.

1. Apply

$$\hat{U}_z(\varphi) = \exp\left(-\frac{i}{\hbar}\varphi \hat{S}_z\right)$$

to χ and show that

$$\chi_\alpha(\mathbf{r}) \mapsto e^{-i\varphi/2}\chi_\alpha(\mathbf{r}), \qquad \chi_\beta(\mathbf{r}) \mapsto e^{+i\varphi/2}\chi_\beta(\mathbf{r}).$$

2. Show that P_α and P_β (Problem 2) are unchanged by this rotation.

3. Show that $\langle \hat{S}_z \rangle$ is unchanged but that $\langle \hat{S}_x \rangle$ and $\langle \hat{S}_y \rangle$ generally change. (You may use your expressions from Problem 3.)

3.6.3 *Exercise 3: Slater determinants*

In this exercise you will work with antisymmetric N-electron wavefunctions built from one-electron spin-orbitals $\chi_i(x)$, where $x = (\mathbf{r}, \omega)$ and $\omega \in \{\alpha, \beta\}$. Recall that electrons are fermions, so the physical N-electron wavefunction must be antisymmetric under exchange of any two electron coordinates $x_i \leftrightarrow x_j$.

The N-electron Slater determinant is defined as

$$\Phi^{\mathrm{SD}}(x_1, \ldots, x_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \cdots & \chi_1(x_N) \\ \vdots & \ddots & \vdots \\ \chi_N(x_1) & \cdots & \chi_N(x_N) \end{vmatrix},$$

i.e. the determinant of the matrix X with elements $X_{ij} = \chi_i(x_j)$, multiplied by $1/\sqrt{N!}$.

For the case $N = 2$,

$$\Phi^{\mathrm{SD}}(x_1, x_2) = \frac{1}{\sqrt{2}} \big(\chi_1(x_1)\chi_2(x_2) - \chi_2(x_1)\chi_1(x_2)\big).$$

Problem 1: Antisymmetry and basic determinant properties

A determinant changes sign when two columns are exchanged. This is exactly the mathematical mechanism that enforces fermionic antisymmetry.

1. Show explicitly that the two-electron Slater determinant satisfies
$$\Phi^{\mathrm{SD}}(x_2, x_1) = -\Phi^{\mathrm{SD}}(x_1, x_2).$$
2. Consider the general N-electron determinant. Explain (briefly) why swapping two electron coordinates $x_i \leftrightarrow x_j$ changes the sign of Φ^{SD}.
3. Show that if two spin-orbitals are identical (e.g. $\chi_p = \chi_q$ for $p \neq q$), then $\Phi^{\mathrm{SD}}(x_1, \ldots, x_N) = 0$ for all $(x_1, \ldots, x_N)$. Interpret this result in words.

Problem 2: Normalization and orthonormal spin-orbitals

Context. In practical quantum chemistry, Slater determinants are built from an orthonormal set of spin-orbitals,

$$\langle \chi_i | \chi_j \rangle = \int dx\, \chi_i(x)^* \chi_j(x) = \delta_{ij}.$$

With orthonormal spin-orbitals, the prefactor $1/\sqrt{N!}$ ensures that the determinant is normalized.

1. For $N = 2$, assume $\langle \chi_1 | \chi_1 \rangle = \langle \chi_2 | \chi_2 \rangle = 1$ and $\langle \chi_1 | \chi_2 \rangle = 0$. Show that $\langle \Phi^{\mathrm{SD}} | \Phi^{\mathrm{SD}} \rangle = 1$.
2. State (no proof required) what changes if the spin-orbitals are not orthonormal. What object replaces 1 as the normalization factor?

Problem 3: Two-electron probability density and exchange correlation

A Hartree product $\chi_1(x_1)\chi_2(x_2)$ leads to an independent-particle probability density $P(x_1, x_2) = P_1(x_1)P_2(x_2)$. A Slater determinant is still considered the "least correlated" fermionic wavefunction, but antisymmetry introduces *exchange correlation* through interference (cross) terms.

Define, for $N = 2$,

$$P(x_1, x_2) = \left|\Phi^{\mathrm{SD}}(x_1, x_2)\right|^2, \qquad P_i(x) = |\chi_i(x)|^2, \qquad Q_{12}(x) = \chi_1(x)^*\chi_2(x).$$

1. Starting from $\Phi^{\mathrm{SD}}(x_1, x_2)$, show that

$$P(x_1, x_2) = \frac{1}{2}\Big(P_1(x_1)P_2(x_2) + P_2(x_1)P_1(x_2) - Q_{12}(x_1)Q_{12}(x_2)^* - Q_{12}(x_1)^*Q_{12}(x_2)\Big).$$

2. Compare this with the Hartree-product probability density and explain (in 2–4 sentences) why the determinant does not describe independent particles.

Problem 4: Two electrons, opposite spin

Context. A common special case is that the two electrons have opposite spin. Then the spin labels make the cross term $Q_{12}(x)$ vanish, and exchange effects in the *spatial* probability density disappear.

Let

$$\chi_1(x) = \psi_1(\mathbf{r})\,|\alpha\rangle\,, \qquad \chi_2(x) = \psi_2(\mathbf{r})\,|\beta\rangle\,,$$

and let Φ^{SD} be the corresponding two-electron Slater determinant.

1. Show that $Q_{12}(\mathbf{r}, \omega) = 0$ for all $(\mathbf{r}, \omega)$.

2. Define the spatial (spin-summed) probability density

$$P(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\omega_1,\omega_2} P(\mathbf{r}_1, \omega_1; \mathbf{r}_2, \omega_2).$$

Show that

$$P(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2}\Big(|\psi_1(\mathbf{r}_1)|^2|\psi_2(\mathbf{r}_2)|^2 + |\psi_2(\mathbf{r}_1)|^2|\psi_1(\mathbf{r}_2)|^2\Big).$$

3. Compute the one-electron marginal (spatial) probability density

$$P(\mathbf{r}_1 \mid \mathbf{r}_2 \text{ anywhere}) = \int P(\mathbf{r}_1, \mathbf{r}_2)\, d\mathbf{r}_2,$$

and show that it is the same for "electron 1" and "electron 2". Interpret this in words.

Problem 5: Two electrons, collinear spins and the Fermi hole

If both electrons have the same spin (e.g. both α), the cross term is generally nonzero and produces a *Fermi hole*: the probability of finding two equal-spin electrons at the same point in space vanishes.

Let

$$\chi_1(x) = \psi_1(\mathbf{r})\,|\alpha\rangle\,, \qquad \chi_2(x) = \psi_2(\mathbf{r})\,|\alpha\rangle\,.$$

1. Show that $Q_{12}(\mathbf{r}, \omega) = \psi_1(\mathbf{r})^* \psi_2(\mathbf{r})\, \delta_{\omega,\alpha}$.

2. Show that the spatial (spin-summed) probability density can be written as

$$P(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2}\Big(P_1(\mathbf{r}_1)P_2(\mathbf{r}_2) + P_2(\mathbf{r}_1)P_1(\mathbf{r}_2) - Q_{12}(\mathbf{r}_1)^* Q_{12}(\mathbf{r}_2) - Q_{12}(\mathbf{r}_1)Q_{12}(\mathbf{r}_2)^*\Big),$$

   where $P_i(\mathbf{r}) = |\psi_i(\mathbf{r})|^2$ and $Q_{12}(\mathbf{r}) = \psi_1(\mathbf{r})^* \psi_2(\mathbf{r})$.

3. Set $\mathbf{r}_1 = \mathbf{r}_2$ and show that $P(\mathbf{r}_1, \mathbf{r}_1) = 0$. Explain why this is called a *Fermi hole*.

Problem 6: A link to expectation values

In quantum chemistry, expectation values are computed from $\Psi^* \hat{O} \Psi$ integrated over all electron coordinates. For spin-independent operators (depending only on $\mathbf{r}$), antisymmetry still affects the probability density and therefore the expectation value.

1. Let $\hat{A} = \hat{a}(1) + \hat{a}(2)$ where $\hat{a}(i)$ depends only on $\mathbf{r}_i$ (no spin dependence). Write the expectation value

$$\langle \hat{A} \rangle = \int dx_1 \int dx_2\, \Phi^{\mathrm{SD}}(x_1, x_2)^*\, \hat{A}\, \Phi^{\mathrm{SD}}(x_1, x_2).$$

2. (Qualitative.) Using Problems 4 and 5, explain why same-spin and opposite-spin electron pairs can lead to different spatial correlation patterns even when the Hamiltonian contains no explicit spin dependence.

3.6.4 *Exercise 4: Counting Slater determinants from an AO basis*

Assume a set of K spatial atomic orbitals (AOs) $\{\phi_\mu(\mathbf{r})\}_{\mu=1}^{K}$. From each spatial AO we form two spin-orbitals (spin-AOs)

$$\chi_{\mu\alpha}(x) = \phi_\mu(\mathbf{r})\,|\alpha\rangle\,, \qquad \chi_{\mu\beta}(x) = \phi_\mu(\mathbf{r})\,|\beta\rangle\,,$$

where $x = (\mathbf{r}, \omega)$ and $\omega \in \{\alpha, \beta\}$.

Thus the AO basis generates $M = 2K$ spin-orbitals in total. An N-electron Slater determinant is specified by choosing N distinct spin-orbitals from these M and antisymmetrizing.

Problem 1: How many determinants are possible?

Context. In configuration interaction (CI) and many other wave-function methods, the size of the determinant expansion is set by how many distinct Slater determinants can be formed from a given one-electron basis. This is the main source of the "combinatorial explosion".

1. Show that the number of distinct N-electron Slater determinants that can be formed from $M = 2K$ spin-orbitals is

$$\#\mathrm{Det}(N, M) = \binom{M}{N}.$$

2. Evaluate $\binom{2K}{N}$ for the following cases:
   (a) $K = 5$, $N = 2$.
   (b) $K = 5$, $N = 6$.
   (c) $K = 10$, $N = 10$.
3. Explain briefly why $\binom{2K}{N}$ grows very rapidly with K and N.

Problem 2: Counting determinants at fixed N_α and N_β

In many calculations we restrict to a fixed number of α and β electrons:

$$N_\alpha + N_\beta = N.$$

This corresponds to working in a fixed M_S sector, where

$$M_S = \frac{1}{2}(N_\alpha - N_\beta)$$

is the total z projection of the electron spin. Counting determinants in a fixed (N_α, N_β) sector is often much smaller than counting all determinants.

1. Show that the number of determinants with exactly N_α α-spin electrons and N_β β-spin electrons is

$$\#\mathrm{Det}(N_\alpha, N_\beta; K) = \binom{K}{N_\alpha}\binom{K}{N_\beta}.$$

2. By a theorem in combinatorics, the sum over all possible (N_α, N_β) with $N_\alpha + N_\beta = N$ recovers $\binom{2K}{N}$:

$$\sum_{N_\alpha=0}^{N} \binom{K}{N_\alpha}\binom{K}{N-N_\alpha} = \binom{2K}{N}.$$

   Verify this identity for $K = 4$ and $N = 3$, by computing the LHS and RHS independently.

3. For $K = 6$ and $N = 6$, compute the number of determinants in the sectors:

$$(N_\alpha, N_\beta) = (3,3),\ (4,2),\ (5,1).$$

Problem 3: Closed-shell vs. open-shell counting

A "closed-shell" electron count corresponds (in a spin-restricted picture) to $N_\alpha = N_\beta = N/2$ (so N even). An "open-shell" case has $N_\alpha \neq N_\beta$. This problem asks you to compare the sizes of the corresponding determinant spaces.

1. Assume N is even. Find the number of determinants in the closed-shell sector.

2. Consider the neighboring open-shell sector $(N_\alpha, N_\beta) = (N/2+1, N/2-1)$. Write down the the number of determinants in that sector.

3. For $K = 8$ and $N = 8$, compute the actual number of determinants in each of the sectors.

Problem 4: Counting determinants relative to a reference (excitations)

In CI, determinants are often organized by "excitations" relative to a chosen reference determinant (often Hartree–Fock). This provides a structured way to build truncated CI spaces such as CIS (singles) and CISD (singles and doubles).

Assume a reference determinant with N electrons built from M spin-orbitals. Let N spin-orbitals be occupied in the reference and $M - N$ be unoccupied (virtual).

1. Show that the number of single excitations (CIS determinants) relative to the reference is

$$N_{\text{singles}} = N(M-N).$$

2. Show that the number of double excitations (CID determinants) relative to the reference is

$$N_{\text{doubles}} = \binom{N}{2}\binom{M-N}{2}.$$

3. Specialize to an AO-based count by taking $M = 2K$. For $K = 7$ and $N = 6$, compute N_{singles} and N_{doubles}.

Problem 5: Spin-resolved excitation counting (fixed N_α, N_β)

If you work in a fixed (N_α, N_β) sector, then excitations must preserve the number of α and β electrons. This changes the counting relative to Problem 4.

Assume the reference determinant has N_α occupied α spin-orbitals and N_β occupied β spin-orbitals. There are $K - N_\alpha$ virtual α spin-orbitals and $K - N_\beta$ virtual β spin-orbitals.

1. Show that the number of spin-conserving single excitations is

$$N_{\text{singles}}^{(\alpha/\beta)} = N_\alpha(K - N_\alpha) + N_\beta(K - N_\beta).$$

2. Show that the number of spin-conserving double excitations can be split into $\alpha\alpha$, $\beta\beta$, and $\alpha\beta$ types:

$$N_{\text{doubles}}^{(\alpha/\beta)} = \binom{N_\alpha}{2}\binom{K - N_\alpha}{2} + \binom{N_\beta}{2}\binom{K - N_\beta}{2} + \binom{N_\alpha}{1}\binom{K - N_\alpha}{1}\binom{N_\beta}{1}\binom{K - N_\beta}{1}.$$

3. For $K = 6$ and $(N_\alpha, N_\beta) = (3, 3)$, compute $N_{\text{singles}}^{(\alpha/\beta)}$ and $N_{\text{doubles}}^{(\alpha/\beta)}$.

3.6.5 *Exercise 5: Minimal-basis RHF (restricted Hartree–Fock)*

In this exercise you will set up the restricted Hartree–Fock (RHF) model in a *minimal AO basis*. We use $x = (\mathbf{r}, \omega)$ with $\omega \in \{\alpha, \beta\}$. A spatial AO basis is $\{\phi_\mu(\mathbf{r})\}_{\mu=1}^{K}$, and the corresponding spin-orbitals are

$$\chi_{\mu\alpha}(x) = \phi_\mu(\mathbf{r})\,|\alpha\rangle\,, \qquad \chi_{\mu\beta}(x) = \phi_\mu(\mathbf{r})\,|\beta\rangle\,.$$

In RHF for a closed-shell system with N electrons (N even), we use $N/2$ *spatial* molecular orbitals (MOs)

$$\psi_i(\mathbf{r}) = \sum_{\mu=1}^{K} C_{\mu i}\,\phi_\mu(\mathbf{r}),$$

and occupy each ψ_i with two electrons of opposite spin:

$$\chi_{i\alpha}(x) = \psi_i(\mathbf{r})\,|\alpha\rangle\,, \qquad \chi_{i\beta}(x) = \psi_i(\mathbf{r})\,|\beta\rangle\,.$$

The RHF wavefunction is a Slater determinant of the N occupied spin-orbitals.

Define the one-electron integrals (in the AO basis)

$$h_{\mu\nu} = \int \phi_\mu(\mathbf{r})\,\hat{h}(\mathbf{r})\,\phi_\nu(\mathbf{r})\,d\mathbf{r}, \qquad S_{\mu\nu} = \int \phi_\mu(\mathbf{r})\phi_\nu(\mathbf{r})\,d\mathbf{r},$$

and the two-electron (Coulomb) integrals

$$(\mu\nu|\lambda\sigma) = \iint \frac{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1)\phi_\lambda(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)}{r_{12}}\,d\mathbf{r}_1\,d\mathbf{r}_2.$$

(You may assume all basis functions are real.)

Problem 1: Closed-shell Slater determinant in RHF

RHF is the simplest multi-electron model that (i) enforces anti-symmetry via a Slater determinant, and (ii) enforces a closed-shell pairing of spins: each spatial orbital is doubly occupied.

Consider a two-electron closed-shell system ($N = 2$) with one spatial MO $\psi(\mathbf{r})$.

1. Write the two occupied spin-orbitals $\chi_\alpha(x)$ and $\chi_\beta(x)$ in terms of $\psi(\mathbf{r})$ and $|\alpha\rangle\,, |\beta\rangle$.

2. Write the $N = 2$ Slater determinant $\Phi^{\mathrm{RHF}}(x_1, x_2)$ explicitly (as a 2×2 determinant and as an expanded expression).

3. Show that the spin-summed spatial probability density

$$P(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\omega_1,\omega_2} \left|\Phi^{\mathrm{RHF}}(\mathbf{r}_1, \omega_1; \mathbf{r}_2, \omega_2)\right|^2$$

reduces to

$$P(\mathbf{r}_1, \mathbf{r}_2) = |\psi(\mathbf{r}_1)|^2 |\psi(\mathbf{r}_2)|^2.$$

(Interpret this result: what happened to the exchange term and why?)

Problem 2: Minimal basis MO expansion and normalization

In a minimal AO basis, each occupied MO is expanded in a small number of AOs. Because AOs are generally *not* orthonormal, normalization involves the overlap matrix S.

Assume a minimal basis with $K = 2$ AOs, $\{\phi_1, \phi_2\}$, and one spatial MO

$$\psi(\mathbf{r}) = C_1\phi_1(\mathbf{r}) + C_2\phi_2(\mathbf{r}).$$

1. Derive the normalization condition $\int |\psi(\mathbf{r})|^2\, d\mathbf{r} = 1$ in terms of C_1, C_2 and the overlaps S_{11}, S_{22}, S_{12}.
2. In matrix form, show that the normalization is $\mathbf{C}^{\mathrm{T}}\mathbf{S}\,\mathbf{C} = 1$ where $\mathbf{C} = (C_1, C_2)^{\mathrm{T}}$.
3. If the AOs are individually normalized ($S_{11} = S_{22} = 1$), express the normalization condition using only C_1, C_2, S_{12}.

Problem 3: One-electron expectation values in an AO basis

Many important quantities can be written as expectation values of one-electron operators. In an AO basis, these become quadratic forms involving the coefficient vector and the integral matrices.

Let $\hat{a}$ be a one-electron operator acting only on $\mathbf{r}$ (spin-independent). Define $a_{\mu\nu} = \int \phi_\mu(\mathbf{r})\hat{a}\,\phi_\nu(\mathbf{r})\, d\mathbf{r}$.

1. Show that for $\psi(\mathbf{r}) = \sum_\mu C_\mu\phi_\mu(\mathbf{r})$,
$$\langle\psi|\hat{a}|\psi\rangle = \sum_{\mu\nu} C_\mu C_\nu\, a_{\mu\nu}.$$
2. Write this in matrix form as $\langle\psi|\hat{a}|\psi\rangle = \mathbf{C}^{\mathrm{T}}\mathbf{A}\,\mathbf{C}$.
3. Specialize to $\hat{a} = \hat{h}$ and write $\langle\psi|\hat{h}|\psi\rangle$ in terms of $h_{\mu\nu}$.

Problem 4: RHF energy expression (symbolic, minimal basis)

The RHF energy can be written in terms of one-electron and two-electron integrals and the occupied MOs. We focus on setting up the correct expressions, keeping integrals symbolic.

Consider a two-electron closed-shell system ($N = 2$) with one doubly occupied spatial MO ψ.

1. Define the Coulomb and exchange self-interaction quantities (in AO notation) for ψ:
$$J = \iint \frac{|\psi(\mathbf{r}_1)|^2|\psi(\mathbf{r}_2)|^2}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2, \qquad K = \iint \frac{\psi(\mathbf{r}_1)\psi(\mathbf{r}_2)\psi(\mathbf{r}_1)\psi(\mathbf{r}_2)}{r_{12}}\, d\mathbf{r}_1\, d\mathbf{r}_2,$$
and explain briefly why $J = K$ for the same spatial orbital (real ψ).
2. Show that the RHF electronic energy (excluding nuclear repulsion) can be written as
$$E_{\mathrm{RHF}} = 2\,\langle\psi|\hat{h}|\psi\rangle + (2J - K).$$

3. Using $\psi(\mathbf{r}) = \sum_\mu C_\mu \phi_\mu(\mathbf{r})$, express $\langle\psi|\hat{h}|\psi\rangle$ and J in terms of C_μ, $h_{\mu\nu}$, and $(\mu\nu|\lambda\sigma)$. (Leave your final answer as sums over AO indices.)

Problem 5: The RHF density matrix in an AO basis

In practical Hartree–Fock implementations, all occupied orbitals are summarized by the (closed-shell) density matrix. This is the bridge from "orbital coefficients" to compact expressions for energies and Fock matrices.

For a closed-shell system with $n_{\text{occ}} = N/2$ occupied spatial MOs and AO coefficients $C_{\mu i}$, define the AO density matrix

$$P_{\mu\nu} = 2\sum_{i=1}^{n_{\text{occ}}} C_{\mu i} C_{\nu i}.$$

1. For the two-electron case ($n_{\text{occ}} = 1$), show that $P_{\mu\nu} = 2C_\mu C_\nu$.

2. Show that the one-electron energy contribution can be written as

$$E^{(1)} = \sum_{\mu\nu} P_{\mu\nu} h_{\mu\nu}.$$

3. Explain why the factor of 2 appears in $P_{\mu\nu}$ for RHF.

3.7 Exercise 6: Occupation number representation (introducing creation and annihilation operators)

In this exercise you will use the *occupation number representation* as an alternative way of writing Slater determinants built from a fixed list of spin-orbitals $\{\chi_p(x)\}_{p=1}^{M}$, where $x = (\mathbf{r}, \omega)$ and $\omega \in \{\alpha, \beta\}$. An N-electron Slater determinant is uniquely specified by which spin-orbitals are occupied.

We write an occupation number state as

$$|n_1 n_2 \cdots n_M\rangle, \qquad n_p \in \{0, 1\},$$

where $n_p = 1$ means that spin-orbital p is occupied and $n_p = 0$ means it is unoccupied. The electron number is

$$N = \sum_{p=1}^{M} n_p.$$

The vacuum (no-electron) state is

$$|0\rangle \equiv |00\cdots0\rangle.$$

Problem 1: Determinants as occupation number vectors

A Slater determinant is determined by the *set* of occupied spin-orbitals. The occupation number vector is a compact, order-independent label for this set.

1. Consider $M = 8$ spin-orbitals labeled $p = 1, \ldots, 8$. Write the occupation number vector corresponding to the determinant with occupied orbitals $\{1, 3, 4, 8\}$.

2. For the same $M = 8$, interpret the occupation number state $|01011001\rangle$ in terms of which spin-orbitals are occupied and the value of N.

3. Explain briefly why $\{1, 3, 4, 8\}$ and $\{8, 4, 3, 1\}$ represent the same occupation number vector, even though permuting columns of a determinant can change its sign.

Problem 2: Closed-shell and M_S sectors in occupation number language

If the spin-orbitals are ordered so that all α spin-orbitals come first and all β spin-orbitals come last, then an occupation number vector immediately tells you how many α and β electrons you have.

Assume spin-orbitals are ordered as

$$(\phi_1\alpha, \ldots, \phi_K\alpha,\ \phi_1\beta, \ldots, \phi_K\beta),$$

so $M = 2K$.

1. For $K = 4$ (so $M = 8$), write an occupation number vector for a closed-shell determinant with $N = 6$ electrons ($N_\alpha = N_\beta = 3$).

2. For your state in (1), compute N_α, N_β, and
$$M_S = \frac{1}{2}(N_\alpha - N_\beta).$$
3. Give one example of an open-shell determinant with $N = 6$ but $M_S = 1$ in the same ordering.

Problem 3: Introducing creation operators as a way to build *determinants*

The occupation number notation is a static label. Creation operators provide a systematic way to *construct* occupation number states starting from the vacuum. This is the first step toward the operator-based (second-quantized) formalism.

Define the creation operator $\hat{a}_p^\dagger$ by its action on occupation number states:
$$\hat{a}_p^\dagger \left|n_1 \cdots n_p \cdots n_M\right\rangle = \begin{cases} (-1)^{\sum_{q=1}^{p-1} n_q} \left|n_1 \cdots 1 \cdots n_M\right\rangle, & n_p = 0, \\ 0, & n_p = 1. \end{cases}$$
(The sign factor will be motivated in Problem 5.)

1. Starting from $|0\rangle = |000000\rangle$ ($M = 6$), compute
$$\hat{a}_1^\dagger |0\rangle, \qquad \hat{a}_4^\dagger |0\rangle.$$
2. Compute
$$\hat{a}_4^\dagger \hat{a}_1^\dagger |0\rangle \qquad \text{and} \qquad \hat{a}_1^\dagger \hat{a}_4^\dagger |0\rangle.$$
Write each result as an occupation number vector (including any sign).
3. Which occupation number state corresponds to "occupy orbitals 1 and 4"? Explain why both operator orders produce the same occupation pattern but may differ by a sign.

Problem 4: Introducing annihilation operators as a way to remove *electrons*

Annihilation operators remove electrons from occupied spin-orbitals. Together with creation operators, they allow you to move between determinants by "removing" and "adding" electrons in specified orbitals.

Define the annihilation operator $\hat{a}_p$ by its action:
$$\hat{a}_p \left|n_1 \cdots n_p \cdots n_M\right\rangle = \begin{cases} (-1)^{\sum_{q=1}^{p-1} n_q} \left|n_1 \cdots 0 \cdots n_M\right\rangle, & n_p = 1, \\ 0, & n_p = 0. \end{cases}$$

1. Let $|\Phi\rangle = |101100\rangle$ ($M = 6$). Compute $\hat{a}_1 |\Phi\rangle$ and $\hat{a}_2 |\Phi\rangle$.
2. Compute $\hat{a}_4 |\Phi\rangle$ and $\hat{a}_5 |\Phi\rangle$. Interpret the zeros.
3. Starting from $|\Phi\rangle$, apply an annihilation operator and then a creation operator to "move" an electron: compute $\hat{a}_5^\dagger \hat{a}_1 |\Phi\rangle$ and interpret the resulting occupation pattern.

Problem 5: Why the sign factor? (link to antisymmetry of Slater determinants)

Occupation number vectors do not track the sign of a determinant under permutation. The sign factor in $\hat{a}_p^\dagger$ and $\hat{a}_p$ ensures that operator-built states reproduce the correct antisymmetry properties of Slater determinants.

1. Consider the two-electron determinant built from spin-orbitals $i < j$. In occupation-number language it is the state with $n_i = n_j = 1$ and all other $n_p = 0$. Show (using Problem 3.2) that

$$\hat{a}_j^\dagger \hat{a}_i^\dagger \left|0\right\rangle = -\hat{a}_i^\dagger \hat{a}_j^\dagger \left|0\right\rangle .$$

2. Explain (briefly) why this sign change is consistent with swapping two columns in a 2×2 Slater determinant.

3. State (no proof required) the key rule that follows: creation operators anticommute,

$$\hat{a}_i^\dagger \hat{a}_j^\dagger = -\hat{a}_j^\dagger \hat{a}_i^\dagger .$$

Problem 6: Number operator

Once creation and annihilation operators are defined, you can build simple operators that measure occupancy. This provides a quick check that the formalism matches the meaning of n_p.

Define $\hat{n}_p = \hat{a}_p^\dagger \hat{a}_p$.

1. Evaluate $\hat{n}_p \left|n_1 \cdots n_M\right\rangle$ for the cases $n_p = 0$ and $n_p = 1$.

2. Show that $\hat{N} = \sum_{p=1}^{M} \hat{n}_p$ returns the electron number N when acting on $\left|n_1 \cdots n_M\right\rangle$.

4 Lecture 4: Configuration–Interaction method

4.1 The CI method

4.1.1 Recap, CI Hilbert space

Recall the definition of configuration-interaction (CI) wavefunctions, on page 24. We assume that an orthonormal set of single-electron functions χ_p is given, allowing a space of CI functions of dimension $\binom{K}{N}$. The basis functions are the Slater determinants

$$|P\rangle \equiv |\chi_{p_1} \cdots \chi_{p_N}\rangle\,, \quad p_1 < p_2 < \cdots < p_N,$$

where $P = (p_1, \cdots, p_N)$ denotes the *ordered* index set. A general CI wavefunction is of the form

$$|\Psi^{\mathrm{CI}}\rangle = \sum_P c_P\,|P\rangle\,. \tag{4.1}$$

In CI theory this is our Hilbert space. In order to solve the Schrödinger equation, we need to formulate how operatrors act on CI functions.

The ordering is important to have a basis for the CI Hilbert space: If we rearrange the single-particle inddices from $|P\rangle$ to $|P'\rangle$ using a permutation, $|P'\rangle = \pm|P\rangle$ *and hence not linearly independent* and not an independent basis functions. Recall that operators are given by their action on a basis.

However, it is often useful to disregard the ordering of $|P'\rangle$, and instead re-introduce the ordering by a proper permutation sign "a little later". For example, when computing the action of an operator on $|P\rangle$, the result is most easily expressed in terms of unordered determinants.

4.1.2 The electronic Schrödinger equation in CI space

Our goal is to solve the electronic Schrödinger equation,

$$\hat{H}\,|\psi\rangle = E\,|\psi\rangle\,,$$

where we now dropped the subscript "el" for brevity. We will approximate the exact wavefunction with a CI wavefunction,

$$|\psi\rangle = |\psi^{\mathrm{CI}}\rangle + |\mathrm{error}\rangle\,, \tag{4.2}$$

where the error is orthogonal to CI space, i.e., $\langle P|\text{error}\rangle = 0$ for all P. The error is there, since our single-particle basis is not complete. In particular, multiplying the wavefunction by the Coulomb potential results in a function that is likely *not* on CI form, taking us "out of CI space."

We now project our Schrödinger equation onto a CI basis vector. This is the same as appealing to the variational principle. We obtain

$$\sum_Q \langle P|\hat{H}|Q\rangle\, C_Q = EC_P, \tag{4.3}$$

where we used that the error is orthogonal to all Slater determinants in the CI basis. Since the basis was not complete, the eigenvalue will no longer be exact, and we denote it by E^{CI}. This is a matrix eigenvalue problem

$$\mathbf{H}^{\text{CI}}\mathbf{C} = E^{\text{CI}}\mathbf{C}, \tag{4.4}$$

where $\mathbf{C}$ is the coefficient vector in the CI wavefunction, see Eq. (4.1), and where $\mathbf{H}^{\text{CI}}$ is the matrix of $\hat{H}$ in the Slater determinant basis,

$$H^{\text{CI}}_{PQ} = \langle P|\hat{H}|\hat{Q}\rangle\,. \tag{4.5}$$

The matrix eigenvalue problem (4.4) can also be written as

$$\hat{H}^{\text{CI}} = |\psi^{\text{CI}}\rangle = E^{\text{CI}}\,|\psi^{\text{CI}}\rangle\,, \tag{4.6}$$

where

$$\hat{H}^{\text{CI}} = \sum_{PQ} |P\rangle\,\langle P|\hat{H}|Q\rangle\,\langle Q| \tag{4.7}$$

is the operator version of the CI matrix.

Are the operators $\hat{H}^{\text{CI}}$ and $\hat{H}$ the same? Why? Why not?

We stress that in these sums, the sum is over a *basis*, i.e., ordered N-tuples of indices.

A major task in CI theory is therefore to compute the matrix elements $\langle P|\hat{H}|Q\rangle$.

4.1.3 *One- and two-body operators*

There are two main types of operators that we will encounter in CI theory: one-body operators and two-body operators. One-body operators are of the form

$$\hat{A} = \sum_{i=1}^{N} \hat{a}(i), \tag{4.8}$$

One-body operators act only on one particle at a time

where $\hat{a}$ is an operator that acts on single-electron functions. Examples include kinetic energy $-\nabla^2/2$, nuclear potential energy $V_{\text{nuc}}(\mathbf{r})$, and the Fock operator $\hat{f}$.

Two-body operators have the general form

$$\hat{W} = \frac{1}{2}\sum_{\substack{i=1\\ i\neq j}}^{N}\sum_{j=1}^{N} \hat{w}(i,j), \tag{4.9}$$

Two-body operators act on two particles at a time

where $\hat{w}$ is an operator that acts on two-electron functions. It is assumed *symmetric* in the sense that

$$\hat{w}(i,j) = \hat{w}(j,i).$$

Our prime example is the electron-electron repulsion,

$$\hat{w}(i,j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}.$$

The rules for computing the matrix elements of one- and two-body operators are called *Slater–Condon rules*.

The Slater–Condon rules give simple recipes for the action of one- and two-body operators on Slater determinants, in terms of which Slater determinants will appear in the result. The main thing to observe is the following: For one-body operators acting on $|P\rangle$, the result will contain those $|Q\rangle$ where *at most one single-electron function is different*. For two-body operators, $|Q\rangle$ has *at most two single-electron functions different*. This leads to *sparsity* of the representation of the operators in the Slater determinant basis, that is, the Hamiltonian – being a sum of one a and two-body operators – will be a *sparse matrix*.

4.1.4 *Slater determinants and their non-uniqueness*

Let $P = (p_1, \cdots, p_N)$ denote the single-electron function indices of a Slater determinant $|P\rangle$. Due to the antisymmetry of Slater determinants,

$$|\chi_{p_1} \cdots \chi_{p_i} \cdots \chi_{p_j} \cdots \chi_{p_N}\rangle = -|\chi_{p_1} \cdots \underset{\uparrow}{\chi_{p_j}} \cdots \underset{\uparrow}{\chi_{p_i}} \cdots \chi_{p_N}\rangle \qquad (4.10)$$

a permutation of the indices gives the same determinant, up to a $\pm$ sign. In a *basis* for CI space, we need to choose a unique representant for all the $N!$ possible index orderings. It is convenient to chooose $p_1 < p_2 < \cdots < p_N$, but by no means mandatory.

When we below work out the Slater–Condon rules, we will say statments like "Q differ from P in n indices". This must be interpreted as *sets*, since we compare determinants. In mathematical set terminology, $|P \cap Q| = N - n$, they have $N - n$ common indices.

On the other hand, it is very convenient for the compact formulation of the Slater–Condon rules that Q, having n different indices from P, is not arbitrarily ordered – that will make the derivation very complicated. We will instead derive the Slater–Condon rules under the following assumptions on the ordering of Q.

However, if Q differs from P as sets in n elements, we select the ordering of Q that is *as close to the ordering of P as possible*, by lining up equal indices. This is best illustrated by example: Suppose that $P = (p_1, p_2, p_3, p_4)$. If Q differs from P in one element, that must mean that some p_j, say p_2, is not present in Q, but all others are. Therefore, $Q = (p_1, q_2, p_3, p_4)$. note carefully, that Q is now *typically not ordered, even if P is*!

Fixing this ordering means that, if Q is *actually* ordered differently, say $q_1 < q_2 < \cdots < q_N$, we must find the permutation of the q_is that line up the indices according to the above convention.

Let us take a concrete example. Let $P = (1,2,5)$, and let $Q = (1,5,6)$, both being ordered according to the basis set convention $p_1 < p_2 < p_3$. If we now exchange 6 and 5 in the latter determinant, we get $Q = (1,6,5)$. This now lines up with P such that the single different index is in the same location in each determinant, $p_2 = 2$ and $q_2 = 6$.

4.1.5 *Slater–Condon rules for one-body operators*

We will introduce the matrix of the one-body operator $\hat{a}$ in the basis of single-electron functions χ_p:

$$\hat{a}\,|\chi_q\rangle = \sum_{p=1}^{K} |\chi_p\rangle\,\langle p|\hat{a}|q\rangle\,, \tag{4.11}$$

$$\langle p|\hat{a}|q\rangle \equiv \langle \chi_p|\hat{a}|\chi_q\rangle\,. \tag{4.12}$$

Here, we have assumed that $\hat{a}$ acts only within the single-electron basis we are given. This is sufficient, since we are interested in the CI matrix of $\hat{A} = \sum_i \hat{a}(i)$. We have also introduced a notation for a one-body matrix element that is a little more economical. Note that one interprets $\langle p|\hat{a}|q\rangle$ as a *matrix*, and indeed another common notation is

$$a_{pq} \equiv \langle p|\hat{a}|q\rangle\,.$$

Setting $\hat{a} = \hat{1}$, we also get the handy notation

$$\langle p|q\rangle \equiv \langle \chi_p|\chi_q\rangle\,.$$

We start by considering an *arbitrary Slater determinant*. We do *not* have to assume that the indices in $|P\rangle$ are ordered for the following.

We start by writing the Slater determinant using permutations:

$$|P\rangle = \frac{1}{\sqrt{N!}} \sum_{\sigma} (-1)^{|\sigma|} \chi_{p_{\sigma(1)}}(x_1) \cdots \chi_{p_{\sigma(N)}}(x_N). \tag{4.13}$$

It is useful to review permutations for this section.

We act with $\hat{a}(i)$, which latches onto $\chi_{\sigma(i)}(i)$:

$$\hat{a}(i)\,|P\rangle = \frac{1}{\sqrt{N!}} \sum_{\sigma} (-1)^{|\sigma|} \chi_{p_{\sigma(1)}}(x_1) \cdots [\hat{a}\chi_{p_{\sigma(i)}}(x_i)] \cdots \chi_{p_{\sigma(N)}}(x_N). \tag{4.14}$$

All other single-particle functions on the sum are unaffected by $\hat{a}(i)$.

4.1.6 *All single-electron functions identical: diagonal element*

Consider now the case where $Q = P$, i.e., we wish to compute $\langle P|\hat{A}|P\rangle$. For the term $\hat{a}(i)$ we get

$$\langle P|\,\hat{a}(i)\,|P\rangle = \frac{1}{N!} \sum_{\tau} (-1)^{|\tau|} \sum_{\sigma} (-1)^{|\sigma|} \langle \chi_{p_{\tau(1)}}|\chi_{p_{\sigma(1)}}\rangle \cdots \langle \chi_{p_{\tau(i)}}|\hat{a}|\chi_{p_{\sigma(i)}}\rangle \cdots \langle \chi_{p_{\tau(N)}}|\chi_{p_{\sigma(N)}}\rangle\,(x_N). \tag{4.15}$$

By orthonormality, we must have $\tau(j) = \sigma(j)$ for all $j \neq i$. But then there is no other possibility than $\tau(i) = \sigma(i)$ as well. Thus, the sum over σ has only one non-zero term, $\tau = \sigma$.

$$\begin{aligned}\langle P|\,\hat{a}(i)\,|P\rangle &= \frac{1}{N!}\sum_{\tau}\underbrace{(-1)^{2|\tau|}}_{=1}\underbrace{\langle\chi_{p_{\tau(1)}}|\chi_{p_{\tau(1)}}\rangle}_{=1}\cdots\langle\chi_{p_{\tau(i)}}|\hat{a}|\chi_{p_{\tau(i)}}\rangle\cdots\underbrace{\langle\chi_{p_{\tau(N)}}|\chi_{p_{\tau(N)}}\rangle}_{=1}\\ &= \frac{1}{N!}\sum_{\tau}\langle\chi_{p_{\tau(i)}}|\hat{a}|\chi_{p_{\tau(i)}}\rangle\,.\end{aligned} \tag{4.16}$$

We are now summing over all permutations τ. $\tau(i)$ will take on all possible values 1 through N. Given k, There are $(N-1)!$ permutations that have $\tau(i) = k$. Thus,

$$\langle P|\,\hat{a}(i)\,|P\rangle = \frac{(N-1)!}{N!}\sum_{k=1}^{N}\langle\chi_{p_k}|\hat{a}|\chi_{p_k}\rangle \tag{4.17}$$

Notice that the right hand side is independent of i, appearing on the left hand side. Summing over i therefore gives

$$\boxed{\langle P|\hat{A}|P\rangle = \sum_{i=1}^{N}\langle p_i|\hat{a}|p_i\rangle\,.} \tag{4.18}$$

Slater–Condon rule for diagonal element of one-body operator

where we took the liberty to change the name of the dummy summation index.

4.1.7 *All but one single-electron function are identical*

The next case is that $Q = P$ except for one single index, i.e., $q_i = p_i$ *except* for one position, $q_j \neq p_j$. Thus,

$$q_1 = p_1,\ q_2 = p_2,\ \cdots, q_j \neq p_j,\ \cdots q_N = p_N. \tag{4.19}$$

We compute the inner product,

$$\langle Q|\hat{a}(i)|P\rangle = \frac{1}{N!}\sum_{\sigma,\tau}(-1)^{|\sigma|+|\tau|}\,\langle\chi_{q_{\tau(1)}}|\chi_{p_{\sigma(1)}}\rangle\cdots\langle\chi_{q_{\tau(i)}}|\hat{a}|\chi_{p_{\sigma(i)}}\rangle\cdots\langle\chi_{q_{\tau(N)}}|\chi_{p_{\sigma(N)}}\rangle\,(x_N). \tag{4.20}$$

Since Q and P are equal in every place except one $q_j \neq p_j$, the orthonormality of the χ_ps imply that the only nonzero term is obtained when $\sigma(i) = \tau(i) = j$. Otherwise, we are gauaranteed that one factor is of the form $\langle\chi_{q_j}|\chi_{p_k}\rangle = 0$. Thus we may write:

$$\langle Q|\hat{a}(i)|P\rangle = \frac{1}{N!}\sum_{\sigma,\tau}\delta_{\sigma(i),j}\delta_{\tau(i),j}(-1)^{|\sigma|+|\tau|}\,\langle\chi_{q_{\tau(1)}}|\chi_{p_{\sigma(1)}}\rangle\cdots\langle\chi_{q_j}|\hat{a}|\chi_{p_j}\rangle\cdots\langle\chi_{q_{\tau(N)}}|\chi_{p_{\sigma(N)}}\rangle\,(x_N). \tag{4.21}$$

We now observe that $\sigma = \tau$, similar to the diagonal element calculation above: if $\sigma \neq \tau$, then at least *two* $\sigma(k) \neq \tau(k)$:

$$\langle Q|\hat{a}(i)|P\rangle = \frac{1}{N!}\sum_{\sigma}\delta_{\sigma(i),j}\,\langle\chi_{q_j}|\hat{a}|\chi_{p_j}\rangle\,. \tag{4.22}$$

How many permutations σ satisfy $\sigma(i) = j$? That is $(N-1)!$:

$$\langle Q|\hat{a}(i)|P\rangle = \frac{(N-1)!}{N!}\,\langle\chi_{q_j}|\hat{a}|\chi_{p_j}\rangle\,. \tag{4.23}$$

The right hand side is independent of i. Summing the N terms gives the final Slater–Condon rule for Slater determinants that differ in one position:

$$\boxed{\langle Q|\hat{A}|P\rangle = \langle q_j|\hat{a}|p_j\rangle\,,\ Q \text{ and } P \text{ differ only in position } j.} \tag{4.24}$$

Slater–Condon rule for matrix elements of one-body operator between determinants that differ in one index

4.1.8 *More than one single-electron function is different*

It is left as an exercise to show that:

$$\boxed{\langle Q|\hat{A}|P\rangle = 0,\ Q \text{ and } P \text{ have more than one different index.}} \tag{4.25}$$

Slater–Condon rule for matrix elements of one-body operator between determinants that differ in two or more indices

4.2 *Slater–Codon rules for two-body operators*

The case of two-body operators is similar to the case of one-body operators, but a little more involved.

We will introduce the matrix of the two-body operator $\hat{b}$ in the basis of Hartree products of single-electron functions χ_p:

$$\boxed{\begin{aligned}\hat{b}(1,2)\,|\chi_r\otimes\chi_s\rangle &= \sum_{p,q=1}^{K} |\chi_p\otimes\chi_q\rangle\,\langle pq|\hat{b}|rs\rangle\,, &(4.26)\\ \langle pq|\hat{b}|rs\rangle &= \langle\chi_p\otimes\chi_q|\hat{a}|\chi_r\otimes\chi_s\rangle\,. &(4.27)\end{aligned}}$$

Here, we used the *tensor product notation* for Hartree products:

$$(\psi\otimes\phi)(x_1,x_2) \equiv \psi(x_1)\phi(x_2)$$

The tensor product is a way to combine two (or more) basis sets to a higher-dimensional basis set. This construction is very common in quantum mechanics, and in the method of separation of variables.

As for one-particle operators, we have assumed that $\hat{b}$ acts only within the (tensor product) of single-electron basis we are given. This is sufficient, since we are interested in the CI matrix of $\hat{B} = \sum_{i<j}\hat{b}(i,j)$.

We assume that $\hat{b}(1,2) = \hat{b}(2,1)$, i.e., that $\hat{b}$ is *particle-symmetric*. This leads to the relation

$$\boxed{\langle pq|\hat{b}|rs\rangle \equiv \langle qp|\hat{b}|sr\rangle\,.} \tag{4.28}$$

There is also need for a second matrix element of $\hat{b}$, namely the *antisymmetrized* matrix element:

$$\boxed{\langle pq|\hat{b}|rs\rangle_{\text{AS}} \equiv \langle pq|\hat{b}|rs\rangle - \langle pq|\hat{b}|sr\rangle\,.} \tag{4.29}$$

This appears naturally in the Slater–Condon rules below.

These matrix elements are referreed to as "physicists matrix elements". Chemist tend to also like their own notation, using the fact that virtually the only interesting two-electron operator is the Coulomb interaction:

$$\boxed{\begin{aligned}[pq|rs] &\equiv \langle pr|\hat{w}|qs\rangle &(4.30)\\ \langle pq||rs\rangle &\equiv \langle pq|\hat{w}|rs\rangle_{\text{AS}} = [pr|qs] - [ps|qr]. &(4.31)\end{aligned}}$$

Chemist's notation for Coulomb matrix elements using general single-electron functions and spin-orbitals

4.2.1 *All single-electron fuctions identical: Diagonal element*

We apply the operator $\hat{b}(1,2)$ to a Slater determinant, noting that it latches onto $\chi_{p_{\sigma(1)}} \otimes \chi_{p_{\sigma(2)}}$, but leaves all other single-electron functions alone:

$$\hat{b}(1,2)\,|P\rangle = \frac{1}{\sqrt{N!}} \sum_{\sigma} (-1)^{|\sigma|} [\hat{b}\chi_{p_{\sigma(1)}}(x_1)\chi_{p_{\sigma(2)}}(x_2)] \cdots \chi_{p_{\sigma(N)}}(x_N). \tag{4.32}$$

We project onto $\langle P|$:

$$\langle P|\,\hat{b}(1,2)\,|P\rangle = \frac{1}{N!} \sum_{\sigma}\sum_{\tau} (-1)^{|\sigma|+|\tau|} \langle p_{\tau(1)} p_{\tau(2)}|\hat{b}|p_{\sigma(1)} p_{\sigma(2)}\rangle \langle \chi_{p_{\tau(3)}}|\chi_{p_{\sigma(3)}}\rangle \cdots \langle \chi_{p_{\tau(N)}}|\chi_{p_{\sigma(N)}}\rangle . \tag{4.33}$$

For a nonzero contribution in the sum we must have $\sigma(j) = \tau(j)$ for $j > 2$. For a given σ, there are 2 permutations τ that achieve this. Either $\tau(1) = \sigma(1)$ and $\tau(2) = \sigma(2)$, *or*, $\tau(1) = \sigma(2)$ and $\tau(2) = \sigma(1)$. The first case is $\sigma = \tau$, while the other case is $\tau = \sigma \circ T_{12}$, where T_{12} denotes the permutation that only switches 1 and 2, a transposition. Notice that for $\sigma = \tau$, $|\sigma| + |\tau| = 2|\sigma|$ is an even number, but in the second case, $|\sigma| + |\tau| = 2|\sigma| + 1$ is an odd number. This gives:

Insert illustration of permutations here ...

$$\langle P|\,\hat{b}(1,2)\,|P\rangle = \frac{1}{N!} \sum_{\sigma} (\langle p_{\sigma(1)} p_{\sigma(2)}|\hat{b}|p_{\sigma(1)} p_{\sigma(2)}\rangle - \langle p_{\sigma(1)} p_{\sigma(2)}|\hat{b}|p_{\sigma(2)} p_{\sigma(1)}\rangle). \tag{4.34}$$

This expression is independent of $\sigma(3)$ through $\sigma(N)$, in total $(N-2)!$ terms. On the other hand $\sigma(1)$ runs from 1 to N, while $\sigma(2)$ runs from 1 to N as well, jumping over $\sigma(i)$:

$$\langle P|\,\hat{b}(1,2)\,|P\rangle = \frac{1}{N(N-1)} \sum_{i=1}^{N}\sum_{j=1}^{N} (\langle p_i p_j|\hat{b}|p_i p_j\rangle - \langle p_i p_j|\hat{b}|p_j p_i\rangle). \tag{4.35}$$

We note that the last term can be simplified using the antisymmetric matrix element, Eq. (**??**). Consider now that $\hat{B} = \sum_{i<j} \hat{b}(i,j)$. We have only considered $(i,j) = (1,2)$, but the right hand side is independent of which pair we looked at. In fact, the result will be exactly the same. This can most easily seen by permuting i to 1 and j to 2 using $T_{i1}T_{j2}$, and repeating the argument, because

$$T_{i1}T_{j2}\,|P\rangle = |P\rangle\,.$$

Thus, summing up we will get

$$\boxed{\langle P|\,\hat{B}\,|P\rangle = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \langle p_i p_j|\hat{b}|p_i p_i\rangle_{\mathrm{AS}}\,.} \tag{4.36}$$

Diagonal element of two-body operator

4.2.2 *All but one single-electron functions identical*

The next case is that $Q = P$ in all positions except one, as for the one-body matrix element considered earlier.

We start with Eq. (4.32), and project onto $\langle Q|$:

$$\langle Q|\hat{b}(1,2)|P\rangle = \frac{1}{N!} \sum_{\sigma,\tau} (-1)^{|\sigma|+|\tau|} \langle q_{\tau(1)} q_{\tau(2)}|\hat{b}|p_{\sigma(1)} p_{\sigma(2)}\rangle \langle \chi_{q_{\tau(3)}}|\chi_{p_{\sigma(3)}}\rangle \cdots \langle \chi_{q_{\tau(N)}}|\chi_{p_{\sigma(N)}}\rangle . \tag{4.37}$$

We must have that $q_{\tau(i)} = p_{\sigma(i)}$ for $i > 2$, otherwise we get zero factors from orthonormality. Thus $\tau(1)$ or $\tau(2)$ is equal to j, where we have the differing index.

Assuming this condition, there are there are $(N-2)!$ permutations τ that are thus. Moreover, as before, we can have $\sigma(1) = \tau(1)$ and $\sigma(2) = \tau(2)$ or the opposite case $\sigma(2) = \tau(1)$ and $\sigma(1) = \tau(2)$. We get

$$\langle Q|\hat{b}(1,2)|P\rangle = \frac{(N-2)}{N!}\sum_{\sigma}\langle q_{\sigma(1)}q_{\sigma(2)}|\hat{b}|p_{\sigma}(1)p_{\sigma(2)}\rangle_{\mathrm{AS}}, \tag{4.38}$$

We now have that $\sigma(1)$ runs over all $k = 1$ through N, while $\sigma(2) = j$ or the opposite case. We get

$$\langle Q|\hat{b}(1,2)|P\rangle = \frac{(N-2)}{N!}\sum_{k\neq j}(\langle p_k q_j|\hat{b}|p_k p_j\rangle_{\mathrm{AS}} + \langle q_j p_k|\hat{b}|p_j p_k\rangle_{\mathrm{AS}}). \tag{4.39}$$

We now use that $\hat{b}(i,j) = \hat{b}(j,i)$, so that the two matrix elements in the sum are equal. We also observe that the matrix element on the right hand side is independent of which pair we have on the left hand side. Using a permutation argument as we did for the one-body matrix element, we arrive at

$$\boxed{\langle Q|\hat{B}|P\rangle = \sum_{\substack{k=1\\k\neq j}}^{N}\langle p_k q_j|\hat{b}|p_k p_j\rangle_{\mathrm{AS}}.} \tag{4.40}$$

Slater–Condon rule for matrix elements of two-body operators between determinants that differ in one index

4.2.3 *All but two single-electron functions identical*

The final Slater–Condon rule we will need to derive involves P and Q, where Q differ in *two* positions: j and k, where we take $j < k$ for simplicity. Thus, $q_i = p_i$ for all i, except $q_j \neq p_j$ and $q_k \neq p_k$.

By now you might see a pattern. There will only be one term here, and that is

$$\boxed{\langle Q|\hat{B}|P\rangle = \langle q_j q_k|\hat{b}|p_j p_k\rangle_{\mathrm{AS}}.} \tag{4.41}$$

Two single-particle functions differ, two-body matrix element

To prove this relation, we start with Eq. (4.32) again, and observe that $(\tau(1),\tau(2)$ must match either (j,k) or (k,j). The same statement is true for $(\sigma(1),\sigma(2))$. Otherwise, we will get an inner product $\langle q_j|p_k\rangle = 0$ or $\langle q_k|p_k\rangle = 0$ in the term.

Thus we must have $\tau(3) = \sigma(3)$ up to $\tau(N) = \sigma(N)$. Thus, Eq. (4.38) is still true. However, due to the restrictions on $\sigma(1)$ and $\sigma(2)$, and finally summing over $\hat{b}(i,j)$ as before, we get

$$\boxed{\langle Q|\hat{B}|P\rangle = \langle q_j q_k|\hat{b}|p_j p_k\rangle_{\mathrm{AS}}}. \tag{4.42}$$

Slater–Condon rule for two-body operators, with Slater determinants differing in two positions

There are no other non-zero matrix elements for two-body operators in the sense that if more than two single-particle functions differ, the matrix element is zero.

4.3 Summary of Slater–Condon rules

We here summarize the Slater–Condon rules for handy reference. The Slater–Condon rules are derived for pairs of Slater determinants P and Q. It is assumed that Q is related to P as follows:

- $P = (p_1, p_2, \cdots, p_N)$, a vector of indices.
- $Q = (q_1, q_2, \cdots, q_N)$, another vector of indices.
- Q will differ from P as sets, having zero ($Q = P$), one, two, or more, elements that differ.
- One: $q_i = p_i$ for all i except for $q_k \neq p_k$ for a single value k
- Two: $q_i = p_i$ for all i except for $q_k \neq p_k$ and $q_l \neq p_l$ for two distinct values k and l
- More: $q_i = p_i$ for all i except for three or more distinct values

For one-body operators, recall that

$$\langle p|\hat{a}|q\rangle \equiv \langle \chi_p|\hat{a}|\chi_q\rangle \,. \tag{4.43}$$

One-body operators		
$\langle P\|\hat{A}\|P\rangle$	$= \sum_{i=1}^{N} \langle p_i\|\hat{a}\|p_i\rangle$	Diagonal element
$\langle Q\|\hat{A}\|P\rangle$	$= \langle q_k\|\hat{a}\|p_k\rangle$	Q and P differ in exactly one position k
$\langle Q\|\hat{A}\|P\rangle$	$= 0$	Q and P differ in more than one position

For two-body operators, recall the definition

$$\langle pq|\hat{b}|rs\rangle_{\text{AS}} \equiv \langle pq|\hat{b}|rs\rangle - \langle pq|\hat{b}|sr\rangle \,, \tag{4.44}$$

where

$$\langle pq|\hat{b}|rs\rangle \equiv \langle \chi_p \otimes \chi_q|\hat{b}|\chi_r \otimes \chi_s\rangle \,, \tag{4.45}$$

i.e., matrix elements in terms of Hartree products.

Two-body operators		
$\langle P\|\hat{B}\|P\rangle$	$= \frac{1}{2}\sum_{i,j=1}^{N} \langle p_i p_j\|\hat{a}\|p_i p_j\rangle_{\text{AS}}$	Diagonal element
$\langle Q\|\hat{B}\|P\rangle$	$= \sum_{i=1}^{N} \langle p_i q_k\|\hat{B}\|p_i p_k\rangle_{\text{AS}}$	Q and P differ in exactly one position k
$\langle Q\|\hat{B}\|P\rangle$	$= \langle q_k q_l\|\hat{b}\|p_k p_l\rangle_{\text{AS}}$	Q and P differ in exactly two positions k and l
$\langle Q\|\hat{B}\|P\rangle$	$= 0$	Q and P differ in more than two positions

4.4 What if single-particle functions are not ordered asin the derivation?

Suppose $P = (p_1, \cdots, p_N)$ and $Q = (q_1, \cdots, q_N)$ are given vectors of single-particle indices. Suppose that as as sets, they

have n elements not in common. We write $P \setminus Q \equiv (p'_1, \cdots, p'_n)$ and $Q \setminus P \equiv (q'_1 \cdots, q'_n)$, the different indices in each set, and $Q \cap P = (p''_1, \cdots, p''_{N-n})$, the common indices.

Find a permutation σ such that

$$P' = (p_{\sigma(1)}, \cdots, p_{\sigma(N)}) = (p'_1, \cdots, p'_n, p''_1, \cdots, p''_{N-n}).$$

Find another permutation τ such that, similarly,

$$Q' = (q_{\tau(1)}, \cdots, q_{\tau(N)}) = (q'_1, \cdots, q'_n, p''_1, \cdots, p''_{N-n}).$$

Thus, the permutations bring the differing indices to the front. Notably, *these new index sets are ordered according to the convention in the Slater–Condon rules*, but additionally, we have placed the differing indices first.

We now obtain, using permutation antisymmetry of Slater determinants,

$$\langle P|\hat{O}|Q\rangle = (-1)^{|\sigma|+|\tau|} \langle P'|\hat{O}|Q'\rangle . \tag{4.46}$$

We can now apply the Slater–Condon rules to the right-hand side.

4.5 Exercises

4.5.1 Exercise 1: The variational principle and constrained minimization

Let $\hat{H}$ be a Hermitian Hamiltonian operator. A *trial wavefunction* $|\Phi\rangle$ is any (sufficiently well-behaved) state vector that we use to approximate an eigenstate of $\hat{H}$. It is typically dependent on a set of K complex variables:

$$\Phi = \Phi(\mathbf{z}), \quad \mathbf{z} \in \mathbb{C}^K.$$

The *Rayleigh quotient* is

$$E[\Phi] = \frac{\langle\Phi|\hat{H}|\Phi\rangle}{\langle\Phi|\Phi\rangle},$$

and becomes a function of $\mathbf{z}$. In general $\Phi(z)$ can be a *complicated* function. We will, however, study the linear case presently.

In CI theory, the variables $\mathbf{z}$ are the CI expansion coefficients $\mathbf{c} = \{c_P\}$:

$$\Phi = \sum_{k=1}^{K} |P_k\rangle\, c_k.$$

The sum extends over whichever determinants P_k are included in our basis, e.g., singles and doubles for CISD.

Problem 1: The variational principle applied to the CI wavefunction

In this problem, you will prove that the variational principle applied to the CI trial wavefunction gives a linear eigenvalue problem.

1. Show that Φ is a linear function of $\mathbf{c}$.

2. Show that
$$\langle\Phi|\Phi\rangle = \mathbf{c}^\dagger\mathbf{c}.$$

3. Show that
$$\langle\Phi|\hat{H}|\Phi\rangle = \mathbf{c}^\dagger\mathbf{H}\mathbf{c},$$
and identify the matrix $\mathbf{H}$ and its elements.

4. Assume $\mathbf{H}$ to be a real matrix, and $\mathbf{c}$ to be a real vector. (This is no loss of generality in this case.) Compute the partial derivatives
$$\frac{\partial}{\partial c_k}(\mathbf{c}^\dagger\mathbf{c}),$$
and
$$\frac{\partial}{\partial c_k}(\mathbf{c}^\dagger\mathbf{H}\mathbf{c}).$$

5. Use the product rule/quotient rule and the chain rule for differentiation of multivariable functions to show that
$$\frac{\partial}{\partial c_k}E(\mathbf{c}) = \frac{2}{\langle\Phi|\Phi\rangle}\left(\sum_j H_{kj}c_j - E(\mathbf{c})c_k\right).$$

6. Show that if $\nabla_{\mathbf{c}} E(\mathbf{c}) = 0$, then

$$\mathbf{Hc} = E\mathbf{c}.$$

Conversely, show that if $\mathbf{c}$ is an eigenvector of $\mathbf{H}$ with eigenvalue λ, i.e.,

$$\mathbf{Hc} = \lambda\mathbf{c}, \tag{4.47}$$

then $\nabla_{\mathbf{c}} E(\mathbf{c}) = 0$, and $\lambda = E(\mathbf{c})$.

Problem 2: More on the eigenvalue problem

1. Show that $\mathbf{H}^\dagger = \mathbf{H}$
2. Show that any eigenvalue E of $\mathbf{H}$ must be real-valued. (Hint: Take the inner product of Eq. (4.47) with the corresponding eigenvector.
3. Suppose $E^{(i)}$ and $E^{(j)}$ are two *different* eigenvalues, with eigenvectors $\mathbf{c}^{(i)}$ and $\mathbf{c}^{(j)}$, respectively. Show that these eigenvectors are orthogonal. (Hint: Project something onto the eigenvalue equations and compare.)
4. A theorem of linear algebra states that the eigenvectors of a Hermitian matrix $\mathbf{H}$ form an orthonormal basis for the vector space at hand. Why then, cannot the *exact* ground state of $\hat{H}$ be recovered from CI?
5. Interpret the CI eigenvalue equation quantum mechanically: What are the allowed states/Hilbert space? What is the Hamiltonian? What are observables?

Problem 3: Variational bound

Assume $\hat{H}$ to have a complete set of eigenvectors,

$$\hat{H}\,|\Psi^{(i)}_{\text{exact}}\rangle = E^{(i)}_{\text{exact}}\,|\Psi^{(i)}_{\text{exact}}\rangle\,.$$

where $E^{(0)}_{\text{exact}}$ is the ground-state energy, and for all i, $E^{(i)}_{\text{exact}} \leq E^{(i+1)}_{\text{exact}}$.

1. Show the *variational principle*:

$$\min_{\Psi \neq 0} E[\Psi] = E^{(0)}_{\text{exact}}.$$

2. Show the variational upper bound property: the CI ground state eigenvalue $E^{(0)}$ satisfies

$$E^{(0)}_{\text{exact}} \leq E^{(0)}$$

Problem 4: Block structure from conserved quantum numbers

This problem may be considered a little hard, simply because I did not have the time to develop/write it clearly ...

It is suggested that you review what block matrices are, and what it means that a matrix is block diagonal.

Let $\hat{O}$ be an operator that commutes with $\hat{H}$, i.e. $[\hat{H}, \hat{O}] = 0$, and such that the corresponding matrices also satisfy $[\mathbf{H}, \mathbf{O}] = 0$.

1. Explain why the above assumption is significant from a quantum mechanical viewpoint.

2. Explain briefly why we can assume that the CI basis vector $|P_k\rangle$ is an eigenvector of $\mathbf{O}$, the matrix of $\hat{O}$ in the CI basis.

3. Explain why reordering a basis has no impact on what vectors can be expanded. That is, show that if $|P_k\rangle$ is our CI basis, and $|P'_j\rangle$ is a reordering of our basis, then every CI vector $\sum_k |P_k\rangle\, c_k$ can also be written
$$\sum_j |P'_j\rangle\, c'_j,$$
for a different coefficient vector $\mathbf{c}'$.

4. Assume now that the CI basis is reordered according to the $\hat{O}$-eigenvalue. Show that $\mathbf{H}$ is block diagonal with respect to the eigenvalue of $\mathbf{O}$. (Hint: Multiply Eq. (4.47) by $\mathbf{O}$ and use the commutator.)

5. Specialize to $\hat{O} = \hat{S}_z$. Explain why determinants with different (N_α, N_β) have different $\hat{S}_z$.

6. Explain why $\mathbf{H}$ (the CI matrix) is block diagonal in $M_S = \frac{1}{2}(N_\alpha - N_\beta)$.

7. Explain why block diagonality is relevant for computational savings.

Problem 4: Sparsity from one- and two-body operators

Write the electronic Hamiltonian as
$$\hat{H} = \sum_{i=1}^{N} \hat{h}(i) + \sum_{i<j} \hat{g}(i,j).$$

1. Explain why $H_{PQ} = 0$ if $|P\rangle$ and $|Q\rangle$ differ by more than two occupied spin-orbitals.

2. List the only three determinant relationships that can yield nonzero H_{PQ}.

3. Explain briefly how sparsity and block structure change the practical cost of solving the CI eigenvalue problem.

4.5.2 *Programming exercise: Slater determinants and CI matrix structure*

In this exercise, you will develop essential parts of a simple CI program in Python.

Our approach will rely on the Python functionality for *sets*. A set in Python is like a list, only that the order does not matter, and each element can only appear once. Here are some examples:

```
A = {0, 1, 2}  # notation for a set literal
B = {0, 2, 1}  # should be the same set as A
print(A - B) # prints set(), the empty set

C = {0, 1, 3}
D = C.union(A)  # set union
print(D)
```

`itertools.combinations` function, which generates subsets of a set. Here is a minimal example:

```
from itertools import combinations

# Define the set of single-particle indices
sp_indices = set(range(10))  # {0, 1, ..., 9}

# Generate all determinants of 3 electrons
determinants = [set(comb) for comb in combinations(
    sp_indices, 3)]

# Print the subsets
for (k,det) in enumerate(determinants):
    print(f'|Phi_{k}> = |{det}>')
```

Problem 1: Construct all determinants

Let K be a given number of orthonormal single-particle functions χ_p, $p = 0, 2, \ldots, K-1$, and let N be the number of electrons in a molecule.

Set up a list of all Slater determinant indices for K single-particle functions and N electrons.

Problem 2: Compare determinants

For each determinant, create a vector with the list indices of all determinants that have exactly 1 different index. Here, it is useful to use the capabilities of Python's set class. To take the difference between two sets P and Q, use $P - Q$. To count the number of elements, use $\text{len}(P - Q)$.

To find the list index of a set Q, you can use something like:

```
try:
    index = list_of_sets.index(P)
    print(index)
except ValueError:
    print("Set not found")
```

Repeat for all index sets that have exactly 2 different indices.

Thus, you should end up with two lists of lists, where element number k is a list of indices k'.

Problem 3: Sparsity pattern

Set up a matrix $A = A_{kk'}$, where $A_{kk'} = 1$ if determinant numbers k and k' differ by at most 2 indices.

Visualize the matrix A.

5 Lecture 5: Hartree–Fock part 2

5.1 Introduction

In this lecture, we continue with the Hartree–Fock method. We will formulate the *General Hartree–Fock* (GHF) method using, well, general single-electron functions/spin-orbitals.

5.2 The Hartree–Fock energy

The HF method is about minimizing the energy of a single Slater determinant. Therefore, we first need to find an expression for this energy.

Let the HF determinant be

$$|\Phi^{\mathrm{HF}}\rangle = |\chi_1\chi_2\cdots\chi_N\rangle\,. \tag{5.1}$$

Recall that $\chi_i(x)$ is a general function of space $\mathbf{r}$ and spin ω.

Here, we have assumed that the χ_i are orthonormal,

$$\langle\chi_i|\chi_j\rangle = \delta_{ij}.$$

Recall, that for orthonormal single-particle functions, we have $\langle\Phi^{\mathrm{HF}}|\Phi^{\mathrm{HF}}\rangle = 1$. We appeal to the Slater–Condon rules and find

$$\begin{aligned} E[\chi_1,\cdots,\chi_N] &= \frac{\langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle}{\langle\Phi^{\mathrm{HF}}|\Phi^{\mathrm{HF}}\rangle} = \langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle \\ &= \langle\Phi^{\mathrm{HF}}|\underbrace{\hat{T}+\hat{V}_{\mathrm{nuc}}}_{\hat{H}_0}|\Phi^{\mathrm{HF}}\rangle + \langle\Phi^{\mathrm{HF}}|\hat{W}|\hat{\Phi^{\mathrm{HF}}}\rangle \\ &= \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j}\langle ij|\hat{w}|ij\rangle_{\mathrm{AS}} \\ &= \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{N}\left(\langle ij|\hat{w}|ij\rangle - \langle ij|\hat{w}|ji\rangle\right) \end{aligned} \tag{5.2}$$

It is a little more standard to call the electron-electron repulsion operator $\hat{W}$ instead of $\hat{V}_{\text{el-el}}$, so this is what we do presently.

Our task now is to derive the conditions for a critical point of the energy $E[\chi_1,\cdots,\chi_N]$ with respect to variations in the unknowns, which are $\chi_1\cdots,\chi_N$, i.e., equations that describe that the energy has zero derivative.

5.3 Calculus of variations: finite dimensions

We will use *calculus of variations* to derive the HF equations. The calculus of variations generalizes ordinary multivariate calculus to

functions where the variables are *functions*. Thus, we may even have an infinite dimensional space for our unknowns. On one hand, this description is more abstract, but it is also more "geometric", and once one gets used to the nomenclature, it is quite intuitive.

For ordinary multivariate functions we have the following description: Let $E(x_1, \cdots, x_n)$ be a multivariate function, sufficiently well-behaved so we can differentiate with respect to all the x_i and obtain a continuous result at least twice. We say that E is of class C^2. The first derivative is the *gradient* (or Jacobian),

$$\nabla E(\mathbf{x}) = \begin{pmatrix} \frac{\partial E}{\partial x_1} \\ \cdots \\ \frac{\partial E}{\partial x_n} \end{pmatrix},$$

and is a vector of functions, each of which can be differentiated again to obtain the *Hessian*,

$$\nabla\nabla^T E(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 E}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 E}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 E}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2}{\partial x_n \partial x_n} \end{pmatrix}.$$

The elements of the Hessian are all continuous functions, and the Hessian is a symmetric matrix under the conditions stated above. [See, e.g., Marsden and Tromba]

Suppose we have found a *critical point* of $E(\mathbf{x})$, i.e., a point $\mathbf{x^0} = (x_1^0, \cdots, x_n^0)$ where all the partial derivatives of first order vanish, i.e., $\nabla E(\mathbf{x}^0) = 0$:

$$\frac{\partial E(\mathbf{x}^0)}{\partial x_i} = 0 \quad (\mathbf{x}^0 \text{ critical point})$$

The Taylor expansion of $E(\mathbf{x})$ around $\mathbf{x} = \mathbf{x}^0$ now reads

$$\begin{aligned} E(\mathbf{x}^0 + \delta\mathbf{x}) &= E(\mathbf{x}^0) + \delta\mathbf{x}^T \nabla E(\mathbf{x}^0) + \frac{1}{2}\delta\mathbf{x}^T (\nabla\nabla^T E(\mathbf{x}^0)\delta\mathbf{x} + \text{h.o.t.} \\ &= E(\mathbf{x}^0) + \frac{1}{2}\sum_{i,j=1}^{n} \delta x_i \frac{\partial^2 E(\mathbf{x}^0)}{\partial x_i \partial x_j}\delta x_j + \text{h.o.t.} \end{aligned} \tag{5.3}$$

Here, $\delta\mathbf{x}$ is a "small change" of $\mathbf{x}^0$, which can be expanded in the standard basis $\mathbf{e}_i$ of $\mathbb{R}^n$ as

$$\delta\mathbf{x} = \sum_{i=1}^{n} \delta x_i \mathbf{e}_i.$$

How small is "small"? The point is that we can recognize a critical point by the condition that the gradient vanishes. In the Taylor expansion, the gradient couples to the variation $\delta\mathbf{x}$. On the other hand, the second-order term gives a tiny correction, and $\delta\mathbf{x}$ must be a vector of such small magnitude that all higher order terms can be

neglected. If the change is really, really tiny, then we can ignore the second-degree part of the Taylor polynomial as well and evaluate the tiny change in E:

$$\begin{aligned}\delta E(\mathbf{x}^0) &= E(\mathbf{x}^0+\delta\mathbf{x}) - E(\mathbf{x}^0) \approx E(\mathbf{x}^0) + \delta\mathbf{x}^T\nabla E(\mathbf{x}^0) - E(\mathbf{x}^0) \\ &= \delta\mathbf{x}^T\nabla E(\mathbf{x}^0).\end{aligned} \tag{5.4}$$

Thus, a critical point $\mathbf{x}^0$ satisfies the condition that $\delta E(\mathbf{x}^0) = 0$ for all possible "infinitesimal" variations $\delta\mathbf{x}^0$. Note however, *that the tiny change $\delta E(\mathbf{x}^0)$ is the total derivative of E at $\mathbf{x}^0$ in the direction of $\delta\mathbf{x}$.*

> Given a smooth multivariate function $E(\mathbf{x})$, the point $\mathbf{x}^0$ is a critical point if and only if $\delta E = 0$ for all possible infinitesimal variations $\delta\mathbf{x}$.

Additionally, we can recognize whether $\mathbf{x}^0$ is a local minimum, maximum, or saddle point depending on the Hessian matrix $\nabla\nabla^T E(\mathbf{x}^0)$, but this is out of scope for the present discussion.

The point of the vanishing variation δE is that it is often easier to work with in the more abstract setting where the variables are functions.

Let us take a simple one-dimensional function as an example:

$$f(x) = x^2 - x + 1.$$

We now do a small change δx and compute our function value:

$$\begin{aligned}f(x+\delta x) &= (x+\delta x)^2 - (x+\delta x) + 1 = x^2 + 2x\delta x + \delta x^2 - x - \delta x + 1 \\ &= x^2 = x + 1 + (2x-1)\delta x + \text{h.o.t}\end{aligned} \tag{5.5}$$

We compute the variation in the function value:

$$f(x+\delta x) - f(x) = (2x-1)\delta x + \text{h.o.t}$$

and obtain

$$\delta f(x) = (2x-1)\delta x$$

to first order in δx. The condition for a critical point becomes

$$2x - 1 = 0.$$

5.4 Derivation of the Hartree–Fock equations

5.4.1 Variations of single-electron functions

The HF single-particle functions are such that they minimize the energy expectation value. That means that if we make small variations $\chi_i \to \chi_i + \delta\chi_i$, the energy must be stationary, i.e., must not change to first order in $\delta\chi_i$, for *all* such small variations.

What is a legal variation in χ_i? Such a variation must be a legal single-electron wavefunction. Thus, we write $\delta\chi_i \in \mathcal{H}_1$, single-electron space. The set $\{\chi_i \mid i \neq N\}$ are called *occupied spin-orbitals*,

but there are only N of these. From these we can construct an N-dimensional hyperplane of single-electron wavefunctions. These are

$$u(x) = \sum_{i=1}^{N} \chi_i(x) u_i \in \mathcal{H}_{\text{occ}}, \quad u_i = \langle \chi_i | u \rangle . \tag{5.6}$$

Some legal variations in χ_i are in this space, i.e., we may consider

$$\delta\chi_i = \sum_{j=1}^{N} \chi_j \langle \chi_j | \delta\chi_i \rangle \in \mathcal{H}_{\text{occ}}. \tag{5.7}$$

However, there is an infinite number of dimensions in $\mathcal{H}_1$, and our variations may also be along either of these infinitely many dimensions *not* captured by the occupied single-electron functions. Thus, the general form of $\delta\chi_i$ is:

$$\delta\chi_i = \sum_{j=1}^{N} \chi_j \langle \chi_j | \delta\chi_i \rangle + \delta\chi_{i,\perp}, \quad \langle \chi_j | \delta\chi_{i,\perp} \rangle = 0. \tag{5.8}$$

The remaining variation is thus orthogonal to all HF single-particle functions, and thus orthogonal to the space $\mathcal{H}_{\text{occ}}$.

This covers all variations we can do.

5.4.2 *The variation of the Slater determinant*

Given variations in χ_i, what is the variation in Φ^{HF}?

Suppose first that $\delta\chi_i = \epsilon\chi_j$, where ϵ is a number, and where $i \neq j$. Then,

Write out this determinant to see that the variation vanishes.

$$\begin{aligned} \delta |\Phi^{\text{HF}}\rangle &= |\chi_1 \cdots (\chi_i + \epsilon\chi_i) \cdots \chi_N\rangle - |\Phi^{\text{HF}}\rangle \\ &= \epsilon |\chi_1 \cdots \chi_j \cdots \chi_N\rangle \\ &= 0. \end{aligned} \tag{5.9}$$

The variation vanishes because the determinant $|\chi_1, \cdots, \chi_j, \cdots, \chi_N\rangle$ has two repeated single-electron functions. Thus there is no variation in the Slater determinant! Clearly, the energy does not change if we do such variations, either, so it is not necessary consider such variations.

If $i = j$, the variation is

$$\delta |\Phi^{\text{HF}}\rangle = \epsilon |\Phi^{\text{HF}}\rangle .$$

In this case the energy does not change either. In fact, such variations violate the constraint of orthonormality, so they can be ruled out from this consideration as well.

Show this by inserting into the first equality of Eq. (5.2).

On the other hand, if $\delta\chi_i = \delta\chi_{i\perp}$ which is orthogonal to all the χ_j, then the variation in the determinant does not vanish:

Write out this determinant to aid in visualizing

$$\delta |\Phi^{\text{HF}}\rangle = |\chi_1 \cdots \delta\chi_{i,\perp} \cdots \chi_N\rangle .$$

We now consider a *general* variation in the Slater determinant, when *all* the variations in the χ_i are allowed to happen simultaneously:

$$\delta |\Phi^{\text{HF}}\rangle = \sum_i \epsilon_i |\Phi^{\text{HF}}\rangle + |\chi_1 \cdots \delta\chi_{i,\perp} \cdots \chi_N\rangle . \tag{5.10}$$

As in the n-dimensional example above, we have used the chain rule for differentiation to obtain the total change in the Slater determinant, which depends on all the single-particle functions.

> For the variational principle applied to the Hartree–Fock determinant, we need only consider variations that are orthogonal to all single-particle functions

5.4.3 *Applying the variational principle*

Recall that we are going to minimize

$$E[\chi_1, \cdots, \chi_N] \equiv E[\chi_1, \cdots, \chi_N] = \sum_{i=1}^{N} \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{N} (\langle ij|\hat{w}|ij\rangle - \langle ij|\hat{w}|ji\rangle).$$

This function changes with the single-particle functions χ_i. We found that it is sufficient to do variations on the form

$$\delta\chi_i = \delta\chi_{i,\perp}.$$

We then have to find an equation that is equivalent to the condition

$$\delta E = 0 \quad \text{for all variations } \delta\chi_i = \delta\chi_{i,\perp} \text{ for all } i.$$

5.4.4 *Variation of one-body part*

Since the derivative is linear, we can treat each term separately, and then add up later. It turns out that the condition that $\delta\chi_i$ is orthogonal to all the χ_j has no bearing on the following purely algebraic manipulations. The orthogonality comes into play a little later.

We consider first the variation of a one-body matrix element: $\delta\langle\chi_i|\hat{h}|\chi_i\rangle$. The matrix element is an inner product, and these are *additive* in each argument:

$$\begin{aligned}\langle\chi+\chi'|\hat{h}|\phi\rangle &= \int dx\overline{[\chi(x)+\chi(x)]}\hat{h}\phi(x)\\ &= \int dx\overline{\chi(x)}\hat{h}\phi(x) + \int dx\overline{\chi'(x)}\hat{h}\phi(x)\\ &= \langle\chi|\hat{h}|\phi\rangle + \langle\chi'|\hat{h}|\phi\rangle\end{aligned} \tag{5.11}$$

Similarly,

$$\begin{aligned}\langle\phi|\hat{h}|\chi+\chi'\rangle &= \int dx\overline{\phi(x)}\hat{h}[\chi(x)+\chi'(x)]\\ &= \int dx\overline{\phi(x)}\hat{h}\chi(x) + \int dx\overline{\phi(x)}\hat{h}\chi'(x)\\ &= \langle\phi|\hat{h}|\chi\rangle + \langle\phi|\hat{h}|\chi'\rangle\end{aligned} \tag{5.12}$$

This mathematical rule is all we need to compute the variation of the one-body part!

The variation in the one-body part is

$$\begin{aligned}
\delta \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle &= \sum_{i=1}^{N} \langle \chi_i + \delta\chi_i | \hat{h} | \chi_i + \delta\chi_i \rangle - \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle \\
&= \sum_{i=1}^{N} \Big(\langle \chi_i | \hat{h} | \chi_i \rangle + \langle \delta\chi_i | \hat{h} | \chi_i \rangle + \langle \chi_i | \hat{h} | \delta\chi_i \rangle + \langle \delta\chi_i | \hat{h} | \delta\chi_i \rangle \Big) - \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle \\
&\approx \sum_{i=1}^{N} \Big(\langle \delta\chi_i | \hat{h} | \chi_i \rangle + \langle \chi_i | \hat{h} | \delta\chi_i \rangle \Big) \quad \text{(first-order part only)} \\
&= \sum_{i=1}^{N} 2 \,\mathrm{Re}\, \langle \delta\chi_i | \hat{h} | \chi_i \rangle
\end{aligned} \tag{5.13}$$

We now consider a *different* choice of variations, namely $\delta'\chi_i = i\delta\chi_i$. This gives

$$\begin{aligned}
\delta' \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle &= \sum_{i=1}^{N} \langle \chi_i + i\delta\chi_i | \hat{h} | \chi_i + i\delta\chi_i \rangle - \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle \\
&= \sum_{i=1}^{N} \Big(\langle \chi_i | \hat{h} | \chi_i \rangle - i \langle \delta\chi_i | \hat{h} | \chi_i \rangle + i \langle \chi_i | \hat{h} | \delta\chi_i \rangle + \langle \delta\chi_i | \hat{h} | \delta\chi_i \rangle \Big) - \sum_{i=1}^{N} \langle \chi_i | \hat{h} | \chi_i \rangle \\
&\approx i \sum_{i=1}^{N} \Big(- \langle \delta\chi_i | \hat{h} | \chi_i \rangle + \langle \chi_i | \hat{h} | \delta\chi_i \rangle \Big) \\
&= 2 \,\mathrm{Im} \sum_{i=1}^{N} \langle \delta\chi_i | \hat{h} | \chi_i \rangle
\end{aligned} \tag{5.14}$$

5.4.5 *Variation of two-body part*

We compute the variation in the two-body part, starting with a single matrix element, and writing $\chi_i' = \chi_i + \delta\chi_i$, we get:

$$\begin{aligned}
\delta \langle \chi_i \chi_j | \hat{w} | \chi_i \chi_j \rangle &= \langle \chi_i' \chi_j' | \hat{w} | \chi_i' \chi_j' \rangle - \langle \chi_i \chi_j | \hat{w} | \chi_i \chi_j \rangle \\
&= \langle \delta\chi_i \chi_j | \hat{w} | \chi_i \chi_j \rangle + \langle \chi_i \delta\chi_j | \hat{w} | \chi_i \chi_j \rangle + \langle \chi_i \chi_j | \hat{w} | \delta\chi_i \chi_j \rangle + \langle \chi_i \chi_j | \hat{w} | \chi_i \delta\chi_j \rangle \\
&= \langle \delta\chi_i \chi_j | \hat{w} | \chi_i \chi_j \rangle + \langle \delta\chi_j \chi_i | \hat{w} | \chi_j \chi_i \rangle + \langle \chi_i \chi_j | \hat{w} | \delta\chi_i \chi_j \rangle + \langle \chi_j \chi_i | \hat{w} | \delta\chi_j \chi_i \rangle
\end{aligned} \tag{5.15}$$

Here, we used the symmetry of $\hat{w}(1,2) = \hat{w}(2,1)$. We also did not even bother to write down terms that had a double variation or higher, since these are of order more than 1, and irrelevant for the variation in the energy!

We also have a second matrix element to vary, and the calculation is almost identical:

$$\begin{aligned}
\delta \langle \chi_i \chi_j | \hat{w} | \chi_j \chi_i \rangle &= \langle \chi_i' \chi_j' | \hat{w} | \chi_j' \chi_i' \rangle - \langle \chi_i \chi_j | \hat{w} | \chi_j \chi_i \rangle \\
&= \langle \delta\chi_i \chi_j | \hat{w} | \chi_j \chi_i \rangle + \langle \chi_i \delta\chi_j | \hat{w} | \chi_j \chi_i \rangle + \langle \chi_i \chi_j | \hat{w} | \delta\chi_j \chi_i \rangle + \langle \chi_i \chi_j | \hat{w} | \chi_j \delta\chi_i \rangle \\
&= \langle \delta\chi_i \chi_j | \hat{w} | \chi_j \chi_i \rangle + \langle \delta\chi_j \chi_i | \hat{w} | \chi_i \chi_j \rangle + \langle \chi_i \chi_j | \hat{w} | \delta\chi_j \chi_i \rangle + \langle \chi_j \chi_i | \hat{w} | \delta\chi_i \chi_j \rangle
\end{aligned} \tag{5.16}$$

We now add these tidbits together:

$$\begin{aligned}
\delta \sum_{i,j} \langle ij|\hat{w}|ij\rangle_{\mathrm{AS}} &= \sum_{i,j} \langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle + \langle \delta\chi_j\chi_i|\hat{w}|\chi_j\chi_i\rangle + \langle \chi_i\chi_j|\hat{w}|\delta\chi_i\chi_j\rangle + \langle \chi_j\chi_i|\hat{w}|\delta\chi_j\chi_i\rangle \\
&\quad - \langle \delta\chi_i\chi_j|\hat{w}|\chi_j\chi_i\rangle + \langle \delta\chi_j\chi_i|\hat{w}|\chi_i\chi_j\rangle + \langle \chi_i\chi_j|\hat{w}|\delta\chi_j\chi_i\rangle + \langle \chi_j\chi_i|\hat{w}|\delta\chi_i\chi_j\rangle \\
&= \sum_{i,j} \langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle + \langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle + \langle \chi_i\chi_j|\hat{w}|\delta\chi_i\chi_j\rangle + \langle \chi_i\chi_j|\hat{w}|\delta\chi_i\chi_j\rangle \\
&\quad - \langle \delta\chi_i\chi_j|\hat{w}|\chi_j\chi_i\rangle + \langle \delta\chi_i\chi_j|\hat{w}|\chi_j\chi_i\rangle + \langle \chi_j\chi_i|\hat{w}|\delta\chi_i\chi_j\rangle + \langle \chi_j\chi_i|\hat{w}|\delta\chi_i\chi_j\rangle \\
&= \sum_{i,j} 2\langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle_{\mathrm{AS}} + 2\langle \chi_i\chi_j|\hat{w}|\delta\chi_i\chi_j\rangle_{\mathrm{AS}} \\
&= 2\,\mathrm{Re}\sum_{i,j} \langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle_{\mathrm{AS}}
\end{aligned} \tag{5.17}$$

As for the one-body terms, we do the second variation and obtain

$$\delta' \sum_{i,j} \langle ij|\hat{w}|ij\rangle_{\mathrm{AS}} = 2\,\mathrm{Im}\sum_{i,j} \langle \delta\chi_i\chi_j|\hat{w}|\chi_i\chi_j\rangle_{\mathrm{AS}} \tag{5.18}$$

5.4.6 *The non-canonical Hartree–Fock equations*

Adding our components together, we obtain the condition of stationary energy. The variations $\delta\chi_i$ are all independent. We can choose their values at will. We select one particular $i = k$, and $\chi_k = \gamma$ – we don't need the index – and $\delta\chi_i = 0$ for all $i \neq k$. The sum over i in the variations collapse to $i = k$:

$$\delta E[\chi_1, \cdots, \chi_N] = 2\,\mathrm{Re}\left(\langle\gamma|\hat{h}|\chi_k\rangle + \sum_j \langle\gamma\chi_j|\hat{w}|\chi_k\chi_j\rangle_{\mathrm{AS}}\right). \tag{5.19}$$

$$\delta' E[\chi_1, \cdots, \chi_N] = 2\,\mathrm{Im}\left(\langle\gamma|\hat{h}|\chi_k\rangle + \sum_j \langle\gamma\chi_j|\hat{w}|\chi_k\chi_j\rangle_{\mathrm{AS}}\right). \tag{5.20}$$

Adding the two variations together "$\delta + i\delta'$" gives the variational condition:

$$\delta E[\chi_1, \cdots, \chi_N] = 0 \iff \langle\gamma|\hat{h}|\chi_k\rangle + \sum_j \langle\gamma\chi_j|\hat{w}|\chi_k\chi_j\rangle_{\mathrm{AS}} = 0, \tag{5.21}$$

for all single-particle functions $\chi \perp \chi_i$, for all i.

5.4.7 *Summing up the variations*

We are on a journey to turn the condition (5.21) into an equation we can do computations with. We consider the "direct" and "exchange" terms of this condition:

$$\sum_j \langle\gamma\chi_j|\hat{w}|\chi_k\chi_j\rangle \quad \text{Hartree/direct term} \tag{5.22}$$

$$\sum_j \langle\gamma\chi_j|\hat{w}|\chi_j\chi_k\rangle \quad \text{Exchange term} \tag{5.23}$$

The Hartree term is so named because if we do not antisymmetrize the Slater determinant and only have a Hartree–Product, this is exactly what will appear. The exchange term is so named because the two particles are, well, exchanged in the ket, and only appear when the wavefunction is properly antisymmetrized.

5.4.8 *The Hartree term*

Let us write the Hartree term as an integral. We now specialize to the Coulumb interaction,

$$\hat{w}(1,2) = \frac{1}{\|\mathbf{r}_1 - \mathbf{r}_2\|}.$$

We write

$$\begin{aligned}\sum_j \langle \gamma\chi_j|\hat{w}|\chi_k\chi_j\rangle &= \int dx_1 \int dx_2 \overline{\gamma}(x_1) \sum_j \overline{\chi_j(x_2)} \frac{1}{r_{12}} \chi_k(x_1)\chi_j(x_2) \\ &= \int dx_1 \overline{\gamma(x_1)} \left[\int dx_2 \sum_j \frac{|\chi_j(x_2)|^2}{r_{12}} \right] \chi_k(x_1) \\ &= \int dx_1 \overline{\gamma(x_1)} v^{\mathrm{H}}(\mathbf{r}_1)\chi_k(x_1) \\ &= \langle \gamma|\hat{v}^{\mathrm{H}}|\chi_k\rangle\,.\end{aligned} \tag{5.24}$$

Here we defined the Hartree potential

$$\begin{aligned}v^{\mathrm{H}}(\mathbf{r}_1) &= \int dx_2 \sum_j \frac{|\chi_j(x_2)|^2}{r_{12}} = \int d\mathbf{r}_2 \frac{\sum_\omega |\chi(\mathbf{r}_2,\omega)|^2}{\|\mathbf{r}_1 - \mathbf{r}_2\|} \\ &= \int d\mathbf{r}_2 \frac{\rho(\mathbf{r}_2)}{\|\mathbf{r}_1 - \mathbf{r}_2\|},\end{aligned} \tag{5.25}$$

which in particular is independent of the spin coordinate appearing in the inner product.

We have also introduced the *electron density* of the Hartree–Fock wavefunction,

$$\rho(\mathbf{r}) \equiv \sum_j \sum_\omega |\chi_j(\mathbf{r},\omega)|^2. \tag{5.26}$$

Thus, the Hartree potential is simply Coulomb's law for the electrostatic potential generated by the charge density $\rho(\mathbf{r})$.

5.4.9 *Exchange term*

The exchange term is a little more complicated to deal with. We desire a one-body bracket as we obtained for the Hartree term, but to achieve this, we have to introduce a *nonlocal potential operator* $\hat{v}^{\mathrm{ex}}$.

$$\begin{aligned}\sum_j \langle \gamma\chi_j|\hat{w}|\chi_j\chi_k\rangle &= \int dx_1 \int dx_2 \overline{\gamma}(x_1) \sum_j \overline{\chi_j(x_2)} \frac{1}{r_{12}} \chi_j(x_1)\chi_k(x_2) \\ &= \int dx_1 \overline{\gamma(x_1)} \left[\int dx_2 \sum_j \frac{\overline{\chi_j(x_2)}\chi_k(x_2)}{r_{12}} \right] \chi_j(x_1) \\ &= \int dx_1 \overline{\gamma(x_1)} \left[\hat{v}^{\mathrm{ex}}\chi_k\right](x_1) \\ &= \langle \gamma|\hat{v}^{\mathrm{ex}}|\chi_k\rangle\,.\end{aligned} \tag{5.27}$$

Notice here that we were not able to isolate a local potential, but we were still able to isolate an operator! The definition of this operator

is in terms of how it acts on a single-particle function:

$$[\hat{v}^{\text{ex}}u](\mathbf{r}_1,\omega_1) = \int d\mathbf{r}_2 \sum_j \frac{\sum_{\omega_2} \overline{\chi_j(\mathbf{r}_2,\omega_2)} u(\mathbf{r}_2,\omega_2)}{r_{12}} \chi_j(\mathbf{r}_1,\omega_1) \quad (5.28)$$

Note how $\hat{v}^{\text{ex}}$ is a *linear* operator: If we insert $u = u_1 + u_2$ in the integral, we get two terms, thus $\hat{v}^{\text{ex}}(u_1+u_2) = \hat{v}^{\text{ex}}u_1 + \hat{v}^{\text{ex}}u_2$. Howver, the operator is *non-local*, as it smears out u over all space via the integration.

5.4.10 *The Fock operator*

We define the Fock operator as

Szabo and Ostlund write this as $\hat{f} = \hat{h} + \hat{\mathcal{J}} - \hat{\mathcal{K}}$.

$$\hat{f} = \hat{h} + \hat{v}^{\text{H}} - \hat{v}^{\text{ex}}. \quad (5.29)$$

The Fock operator depends on the single-particle functions through the Hartree and exchange potentials. The variational condition in Eq. (5.21) becomes

$$\langle\gamma|\hat{f}|\chi_k\rangle = 0 \quad \text{for all } k. \quad (5.30)$$

5.4.11 *Hartree–Fock equations*

Our variations are now orthogonal to all the χ_i:

$$\langle\gamma|\chi_i\rangle = 0.$$

Since $\hat{f}\,|\chi_k\rangle$ is a single-particle function, we can expand it in two parts: one parallel to the single-particle functions, and one orthogonal to the single-particle functions,

$$\hat{f}\,|\chi_k\rangle = \sum_j |\chi_j\rangle\,\lambda_{jk} + |g_k\rangle\,, \quad |g_k\rangle = (\hat{f}\,|\chi_k\rangle)_\perp.$$

Here, λ_{jk} are some unknown coefficients. We just know they exist. Choose the variation $\gamma = \epsilon g_k$, where ϵ is a tiny number. This is legal, since g_k is orthogonal to all the χ_i. Taking the inner product with γ reveals that

$$\langle\gamma|\,\hat{f}\,|\chi_k\rangle = \epsilon\,\langle g_k|g_k\rangle = \epsilon\|g_k\|,$$

which is only zero if $g_k = 0$.

On the other hand, no variation in χ_k can ever affect the λ-terms. Thus:

> $\delta E = 0$ for all variations orthogonal of the single-particle functions if and only if there are constants λ_{jk} such that
>
> $$\hat{f}\,|\chi_k\rangle = \sum_j |\chi_j\rangle\,\lambda_{jk} \quad (5.31)$$

5.4.12 *The canonical Hartree–Fock equations*

Equation 5.31 is called the *non-canonical Hartree–Fock equation*. It is "almost" an eigenvalue equation, except that the right-hand side allows couplings between the single-particle functions.

We now observe that $\Lambda = [\lambda_{jk}]$ is a Hermitian matrix. To see this, we compute

$$\lambda_{jk} = \langle \chi_j | \hat{f} | \chi_k \rangle = \langle \chi_j | \hat{h} + \hat{v}^{\mathrm{H}} + \hat{v}^{\mathrm{ex}} | \chi_k \rangle \,.$$

The first term is clearly Hermitian, since $\hat{h}$ is Hermitian. It is left as an exercise to show that the Hartree and exchange operators are also Hermitian. Since Λ is Hermitian, we can diagonalize it:

Solve this exercise!

$$\Lambda = UEU^H, \quad E = \mathrm{diag}(\epsilon_1, \cdots, \epsilon_N).$$

In components,

$$\lambda_{jk} = U_{jl} \epsilon_l \overline{U_{kl}}.$$

From $\Lambda = UEU^H$ we also obtain the identity

$$U^H \Lambda = E U^H.$$

We will use this relation shortly.

We next observe, that the Slater determinant Φ^{HF} is left unchanged if we take a unitary matrix $U = [U_{ik}]$ of size $N \times N$, and redefine our single-particle functions according to

Show this by using $\det(AB) = \det(A)\det(B)$ after inserting the expansion into the definition of a Slater determinant.

$$|\chi'_j\rangle = \sum_l |\chi_l\rangle \, U_{lj}. \tag{5.32}$$

Since $U^{-1} = U^H$ by unitarity, the inverse transoformation is

$$|\chi_j\rangle = \sum_l |\chi'_l\rangle \, \overline{U_{jl}}. \tag{5.33}$$

The choice of the unitary matrix U comes from the diagonalization of Λ above. Now something fantastic happens!

Inserting this expression into the non-canonical HF equations (5.31), we obtain

$$\begin{aligned}
\hat{f} \, |\chi_k\rangle &= \sum_l \hat{f} \, |\chi'_l\rangle \, \overline{U_{kl}} = \sum_{j,l} |\chi'_l\rangle \, \overline{U_{jl}} \lambda_{jk} \\
&= \sum_{j,l} |\chi'_l\rangle \, \overline{U_{jl}} \lambda_{jk} = \sum_l |\chi'_l\rangle \, (U^H \Lambda)_{lk} \\
&= \sum_l |\chi'_l\rangle \, (E U^H)_{lk} \\
&= \sum_l |\chi'_l\rangle \, \epsilon_l \overline{U_{kl}}
\end{aligned} \tag{5.34}$$

We now multiply with U_{ki} and sum over k: The left-hand side becomes

$$\sum_{l,k} \hat{f} \, |\chi'_l\rangle \, \overline{U_{kl}} U_{ki} = \sum_l \hat{f} \, |\chi'_l\rangle \, (U^H U)_{li} = \sum_l \hat{f} \, |\chi'_l\rangle \, \delta_{li} = \hat{f} \, |\chi'_i\rangle \,.$$

On the right-hand side, we instead obtain

$$\sum_{k,l} |\chi'_l\rangle \, \epsilon_l \overline{U_{kl}} U_{ki} = \sum_l |\chi'_l\rangle \, \epsilon_l (U^H U)_{li} = \sum_l |\chi'_l\rangle \, \epsilon_l \delta_{li} = |\chi'_i\rangle \, \epsilon_i.$$

We combine these two equations into the ...

Canonical Hartree–Fock equations:

$$(\hat{h} + \hat{v}^{\mathrm{H}} - \hat{v}^{\mathrm{ex}}) \, |\chi_i\rangle = \epsilon_i \, |\chi_i\rangle \, . \tag{5.35}$$

This is a nonlinear eigenvalue problem in the sense that the Hartree and exchange potentials depend on the eigenvectors. Thus, we must solve the equations self-consistently.

5.4.13 *Discussion of the Hartree–Fock equations*

We have derived two sets of Hartree–Fock equations: Non-canonical, where the single-particle functions were mixed in the right-hand side, and the canonical, where the single-particle functions were not mixed. The latter is a generalized nonlinear eigenvalue problem, while the former is not.

The two equations are equivalent: Given a solution of the non-canonical equations we demonstrated explicitly that the χ_i can be transformed into *new* functions that satisfy the canonical equations. Concersely, if we start with a solution of the canonical set of equations, we can do unitary transformations and obtain solutions of the non-canonical equations.

In practice, the non-canonical equations are solved in computer codes. It is important to be aware, though, that the canonical solutions are just a particular choice, a convention, for a Hartree–Fock solution. The Slater determinant, which is the wavefunciton, does not change between canonical and non-canonical solutions.

5.5 Exercises

5.5.1 Exercise 1: Deriving the Hartree–Fock energy using Slater–Condon rules

Let the Hartree–Fock determinant be

$$|\Phi^{\mathrm{HF}}\rangle = |\chi_1\chi_2\cdots\chi_N\rangle\,, \qquad \langle\chi_i|\chi_j\rangle = \delta_{ij}.$$

The electronic Hamiltonian is written as

$$\hat{H} = \hat{H}_0 + \hat{W}, \qquad \hat{H}_0 = \hat{T} + \hat{V}_{\mathrm{nuc}}, \qquad \hat{H}_0 = \sum_{k=1}^{N} \hat{h}(k), \qquad \hat{W} = \sum_{k<l} \hat{w}(k,l),$$

where $\hat{h}$ is a one-electron operator and $\hat{w}(k,l) = r_{kl}^{-1}$ is the electron–electron repulsion operator. In this exercise, you wull use the Slater–Condon rules to derive Eq. (5.2):

$$E[\chi_1,\ldots,\chi_N] = \langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle = \sum_{i=1}^{N}\langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{i,j=1}^{N}\Big(\langle ij|\hat{w}|ij\rangle - \langle ij|\hat{w}|ji\rangle\Big).$$

Problem 1: Normalization and energy decomposition

1. Argue that $\langle\Phi^{\mathrm{HF}}|\Phi^{\mathrm{HF}}\rangle = 1$, and therefore

$$E[\chi_1,\ldots,\chi_N] = \frac{\langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle}{\langle\Phi^{\mathrm{HF}}|\Phi^{\mathrm{HF}}\rangle} = \langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle\,.$$

2. Using $\hat{H} = \hat{H}_0 + \hat{W}$, show that

$$\langle\Phi^{\mathrm{HF}}|\hat{H}|\Phi^{\mathrm{HF}}\rangle = \langle\Phi^{\mathrm{HF}}|\hat{H}_0|\Phi^{\mathrm{HF}}\rangle + \langle\Phi^{\mathrm{HF}}|\hat{W}|\Phi^{\mathrm{HF}}\rangle\,.$$

Problem 2: One-electron part $\langle\Phi^{\mathrm{HF}}|\hat{H}_0|\Phi^{\mathrm{HF}}\rangle$

Start with the one-electron part explicitly as a sum over electrons:

$$\hat{H}_0 = \sum_{k=1}^{N}\hat{h}(k).$$

1. State the Slater–Condon rule for the diagonal matrix element of a one-body operator $\hat{A} = \sum_{k=1}^{N}\hat{a}(k)$ evaluated on a determinant $|\Phi\rangle = |\chi_1\cdots\chi_N\rangle$ built from orthonormal spin-orbitals.

2. Apply that rule to $\hat{H}_0$ to show that

$$\langle\Phi^{\mathrm{HF}}|\hat{H}_0|\Phi^{\mathrm{HF}}\rangle = \sum_{i=1}^{N}\langle\chi_i|\hat{h}|\chi_i\rangle\,.$$

Problem 3: Two-electron part $\langle\Phi^{\mathrm{HF}}|\hat{W}|\Phi^{\mathrm{HF}}\rangle$

1. Write the electron–electron repulsion as

$$\hat{W} = \sum_{k<l}\hat{w}(k,l).$$

Explain briefly why the sum is over $k < l$ rather than all combinations of k and l.

2. State the Slater–Condon rule for the diagonal element of a two-body operator $\hat{G} = \sum_{k<l} \hat{g}(k,l)$ with a determinant built from orthonormal spin-orbitals. Your stated rule should involve Coulomb- and exchange-type two-electron integrals.

3. Apply the rule to $\hat{W}$ and show that

$$\langle \Phi^{\mathrm{HF}} | \hat{W} | \Phi^{\mathrm{HF}} \rangle = \sum_{i<j} \Big(\langle ij|\hat{w}|ij\rangle - \langle ij|\hat{w}|ji\rangle \Big).$$

4. Show that the $i<j$ sum can be rewritten as a double sum with a prefactor:

$$\sum_{i<j} f(i,j) = \frac{1}{2} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} f(i,j).$$

Then argue why, for the antisymmetrized integrand

$$f(i,j) = \langle ij|\hat{w}|ij\rangle - \langle ij|\hat{w}|ji\rangle ,$$

you may extend the sum to all i,j without changing the value:

$$\frac{1}{2} \sum_{\substack{i,j=1 \\ i\neq j}}^{N} f(i,j) = \frac{1}{2} \sum_{i,j=1}^{N} f(i,j).$$

Problem 4: Assemble the Hartree–Fock energy expression

Combine your results from Problems 2 and 3 to obtain Eq. (5.2).

5.6 *Exercise 2: Calculus of variations in finite dimensions*

Let $E(\mathbf{x})$ be a C^2 function of $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. For an infinitesimal variation

$$\delta\mathbf{x} = \sum_{i=1}^{n} \delta x_i \mathbf{e}_i,$$

define the first-order variation

$$\delta E(\mathbf{x}) = E(\mathbf{x} + \delta\mathbf{x}) - E(\mathbf{x})$$

keeping only terms linear in $\delta\mathbf{x}$.

Problem 1: Warm-up with one variable

1. Let $f(x) = x^2 - x + 1$. Compute $f(x+\delta x)$, expand in powers of δx, and identify

$$\delta f(x) = f(x+\delta x) - f(x)$$

to first order in δx.

2. Use the condition "$\delta f(x^0) = 0$ for all δx" to find all critical points x^0.

3. Verify that your critical points agree with the usual derivative condition $f'(x^0) = 0$.

Problem 2: Variation equals directional derivative

1. Let $E(\mathbf{x})$ be differentiable. Show (by Taylor expansion to first order) that
$$\delta E(\mathbf{x}) \approx \delta \mathbf{x}^T \nabla E(\mathbf{x}).$$

2. Let $\mathbf{v}$ be a vector, and let
$$L = \left\{ \mathbf{x}^0 + s\mathbf{v} \mid s \in \mathbb{R} \right\}.$$
Describe L using words.

3. Define the function $f(s) = E(\mathbf{x}^0 + s\mathbf{v})$, where $\mathbf{v}$ is a direction vector. Use the chain rule to show that
$$f'(0) = \mathbf{v}^T \nabla E(\mathbf{x}^0).$$

4. Fix a direction vector $\mathbf{v}$ and set $\delta\mathbf{x} = \epsilon\mathbf{v}$. Show that
$$\delta E(\mathbf{x}) \approx \epsilon\, \mathbf{v}^T \nabla E(\mathbf{x})$$
and interpret the coefficient of ϵ as the directional derivative of E at $\mathbf{x}$ in the direction $\mathbf{v}$.

Problem 3: "$\delta E = 0$ for all $\delta\mathbf{x}$" is the same as $\nabla E = 0$

1. Assume $\delta E(\mathbf{x}^0) = \delta\mathbf{x}^T \nabla E(\mathbf{x}^0) = 0$ for all variations $\delta\mathbf{x}$. Show that this implies $\nabla E(\mathbf{x}^0) = \mathbf{0}$.

2. Conversely, assume $\nabla E(\mathbf{x}^0) = \mathbf{0}$. Show that $\delta E(\mathbf{x}^0) = 0$ for all variations $\delta\mathbf{x}$.

Problem 4: Compute variations of explicit multivariate functions

1. Let $E(x_1, x_2) = x_1^2 + 3x_1x_2 + 2x_2^2 - x_1 - 2x_2$. Compute δE to first order in $(\delta x_1, \delta x_2)$.

2. Read off $\nabla E(\mathbf{x})$ from your expression for δE and verify it by computing partial derivatives.

3. Solve $\nabla E(\mathbf{x}^0) = 0$ for the critical point $\mathbf{x}^0$.

Problem 5: Quadratic form notation and the Hessian

1. Let $E(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\,\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where $\mathbf{A}$ is a constant $n \times n$ matrix, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. Compute δE to first order in $\delta\mathbf{x}$.

2. Show that if $\mathbf{A}$ is symmetric, then
$$\nabla E(\mathbf{x}) = \mathbf{A}\,\mathbf{x} + \mathbf{b}, \qquad \nabla\nabla^T E(\mathbf{x}) = \mathbf{A}.$$

3. Use the condition $\delta E(\mathbf{x}^0) = 0$ for all $\delta\mathbf{x}$ to show that a critical point satisfies
$$\mathbf{A}\,\mathbf{x}^0 + \mathbf{b} = \mathbf{0}.$$

Problem 6: Lagrange multipliers in finite dimensions

The method of Lagrange multipliers is a technique to find critical points of functions when the argument $\mathbf{x} \in \mathbb{R}^n$ is constrained. It turns the constrained optimization, which is in general hard since one needs to find a parameterization of the constrained set of points, into an unconstrained optimization, which may be easier, albeit at the cost of more variables.

Let $E(\mathbf{x})$ be a C^1 function and let $g(\mathbf{x}) = 0$ be a constraint with g differentiable. Define the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda) = E(\mathbf{x}) - \lambda\, g(\mathbf{x}),$$

which is a function of one more variable, $\mathbf{y} = (\mathbf{x}, \lambda) \in \mathbb{R}^{n+1}$.

1. Show that

   $$\delta\mathcal{L} = \delta\mathbf{x}^T \nabla E(\mathbf{x}) - \lambda\, \delta\mathbf{x}^T \nabla g(\mathbf{x}) - (\delta\lambda)\, g(\mathbf{x})$$

   by writing $\delta\mathbf{y} = (\delta\mathbf{x}, \delta\lambda)$; independent infinitesimal variations.

2. Use the requirement $\delta\mathcal{L} = 0$ for all $\delta\mathbf{x}$ and $\delta\lambda$ to show the stationary conditions

   $$\nabla E(\mathbf{x}^0) = \lambda\, \nabla g(\mathbf{x}^0), \qquad g(\mathbf{x}^0) = 0.$$

3. Let $\delta\mathbf{x}$ be a variation which additionally satisfies $\delta g(\mathbf{x}^0) = \delta\mathbf{x}^T \nabla g(\mathbf{x}^0)$. Show that $\mathbf{x}^0 + \delta\mathbf{x}$ satisfies the constraint (to first order).

4. Show that under the conditions of the previous two points,

   $$\delta E(\mathbf{x}) = 0.$$

   Interpret in terms of optimizing E under the constraint.

5. Apply the method of Lagrange multipliers to find the unique critical point of

   $$E(x_1, x_2) = x_1^2 + x_2^2$$

   subject to the constraint $g(x_1, x_2) = x_1 + x_2 - 1$, and find the stationary point(s) (x_1^0, x_2^0).

6. In a 2D axis draw a picture that illustrates the function E and the constraint g. Draw level curves for E, and draw the whole locus of $g = 0$.

7. Compute $\nabla E(x_1^0, x_2^0)$ and $\nabla g(x_1^0, x_2^0)$. Draw them as vectors anchored at the critical point. Interpret in terms of the condition $\nabla E = \lambda \nabla g$.

8. Let E be an n-variable C^2 function, and let g_1 and g_2 be *two* constraints. Define the Lagrangian

   $$\mathcal{L}(\mathbf{x}, \lambda_1, \lambda_2) = E(\mathbf{x}) - \lambda_1 g_1(\mathbf{x}) - \lambda_2 g_2(\mathbf{x}).$$

   Explain why unconstrained optimization $\delta\mathcal{L} = 0$ will give the optimization of E subject to *both* constraints $g_1 = 0$ and $g_2 = 0$.

Problem 7: Connecting to the upcoming Hartree–Fock derivation

1. In Hartree–Fock, the "variables" are functions (the spin-orbitals χ_i) rather than numbers $(x_1, \ldots, x_n)$. In one sentence: what plays the role of $\delta\mathbf{x}$?

2. The HF determinant uses orthonormal spin-orbitals, $\langle\chi_i|\chi_j\rangle = \delta_{ij}$. In the derivation of the HF equations, we did not explicitly enforce this condition. Write a constraint function g_{ij} (analogous to $g(\mathbf{x})$ above) whose zero enforces this condition.

3. Explain that unconstrained optimization of

$$\mathcal{L} = E[\chi_1, \cdots, \chi_N] - \sum_{ij} \lambda_{ij} g_{ij}[\chi_1, \cdots, \chi_N]$$

will yield an optimal Slater determinant with orthonormal single-particle functions.

5.6.1

Problem 8: Alternative derivation of the Hartree–Fock equations (very rough exercise sketch)

The following exercise could really benefit from having more steps. I have not had the time to quality check the points, and write more steps ...

1. Show that $g_{ij} = \langle\chi_i|\chi_j\rangle - \delta_{ij}$ is a proper set of constraints for orthonormality of single-particle functions, where $i, j = 1, \ldots, N$.

2. We now treat χ_i and the complex conjugates as if they were independent variables. This is actually allowed. Show that the variation with respect to χ_k^* alone gives the equation

$$\delta\mathcal{L} = \delta E - \sum_j \lambda_{kj}\chi_j = 0$$

Here, δE is the variation of the energy under this variation of χ_k.

3. (A bit of work.) Retrace the steps of the derivation of the HF equations in the lecture notes, to show that

$$\hat{f}\chi_k = \sum_j \lambda_{kj}\chi_j.$$

4. Compare with the equations derived in the lecture.

6 Lecture 6: Hartree–Fock in a basis

6.1 Restricted and unrestricted Hartree–Fock wavefunctions

We now take our canonical Hartree–Fock equations and turn them into something programmable on a computer. To this end, we assume that our single-particle functions are expanded in a given *atomic orbital functions* $\phi_\mu(\mathbf{r})$. The index is chosen to be distinct from the index i used on χ_i, so that it is easy to recognize them, and also since they cound different things. For example, there are in general only N of the latter, but this is not related to the number of the former, which is typically in the hundreds.

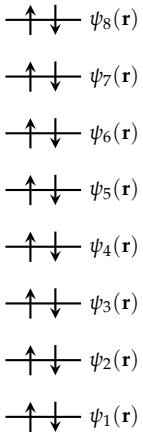


Illustration of a RHF Slater determinant

For *restricted Hartree–Fock* (RHF) we have, for an even number of electrons N, and for $j = 1, 2, \ldots, N/2$,

$$\chi_{2j-1}(\mathbf{r}, \omega) = \psi_j(\mathbf{r})\alpha(\omega), \tag{6.1}$$

$$\chi_{2j}(\mathbf{r}, \omega) = \psi_j(\mathbf{r})\beta(\omega). \tag{6.2}$$

Recall the special notation

$$|\Phi^{\text{RHF}}\rangle = |\psi_1\overline{\psi_1}\cdots\psi_{N/2}\overline{\psi_{N/2}}\rangle\,, \tag{6.3}$$

exhibiting how electrons are paired up in spatial orbitals. The expansion in AOs is given by

$$\psi_i(\mathbf{r}) = \sum_\mu C_{\mu i}\phi_\mu(\mathbf{r}). \tag{6.4}$$

The unknowns in RHF are therefore the matrix elements of $C = C_{\mu i}$, called *MO coefficients*. Thus, we have the following matrix structure:

$$C = \begin{bmatrix} & \overset{C_{:,i}}{\curvearrowright} & \\ \vert & \vert & \\ \vert & \vert & \cdots \\ \vert & \vert & \end{bmatrix}$$

For *unrestricted Hartree–Fock* (UHF), we have

$$\chi_{2j-1}(\mathbf{r}, \omega) = \psi_j^\alpha(\mathbf{r})\alpha(\omega), \tag{6.5}$$

$$\chi_{2j}(\mathbf{r}, \omega) = \psi_j^\beta(\mathbf{r})\beta(\omega). \tag{6.6}$$

The electron spin is still predetermined, but the spatial parts of each spin orbital are "unrestricted":

$$|\Phi^{\text{UHF}}\rangle = |\psi_1^\alpha\overline{\psi_1^\beta}\cdots\psi_{N/2}^\alpha\overline{\psi_{N/2}^\beta}\rangle\,, \tag{6.7}$$

In terms of the AOs,

$$\psi_i^\alpha(\mathbf{r}) = \sum_\mu C_{\mu i}^\alpha \phi_\mu(\mathbf{r}). \tag{6.8}$$

$$\psi_i^\beta(\mathbf{r}) = \sum_\mu C_{\mu i}^\beta \phi_\mu(\mathbf{r}). \tag{6.9}$$

Thus, we now have *two* MO coefficient matrices. Typically, these are stored in a larger matrix on block form, $C = [C^\alpha \mid C^\beta]$. The $N/2$ first columns are the α MO coefficients, while the last $N/2$ columns are the β MO coefficients.

The above Slater determinants are only defined for an even number of electrons. UHF is straightforwardly applied to an odd number of electrons: Simply add one more electron to, say the α part of the wavefunction. To treat odd number of electrons in RHF, the proper approach is *restricted open-shell Hartree–Fock* (ROHF) which we will not treat.

For the *general Hartree–Fock method* (GHF) method we have

$$\chi_i(\mathbf{r},\omega) = \begin{pmatrix} \psi_i^\alpha(\mathbf{r}) \\ \psi_i^\beta(\mathbf{r}) \end{pmatrix}, \quad \psi_i^\alpha = \sum_\mu C_{\mu i}^\alpha \phi_\mu(\mathbf{r}), \quad \psi_i^\beta = \sum_\mu C_{\mu i}^\beta \phi_\mu(\mathbf{r}). \tag{6.10}$$

Relative to UHF, we see that *every* single-particle function has both α and β MO coefficients – the dimension of these matrices are therefore doubled in GHF. The GHF method is quite uncommon, and we will not discuss it further.

6.2 The AO basis set

We now briefly discuss the AO basis sets in use. What are they? The short answer is that these functions $\phi_\mu(\mathbf{r})$ are almost universally chosen as *linear combinations of primitive gaussian functions*:

$$\phi_\mu(\mathbf{r}) = \sum_I d_I g_{\mu,I}(\mathbf{r}), \tag{6.11}$$

where $g_{\mu,I}(\mathbf{r})$ denotes a number of *primitive gaussians* defined for basis function number μ. These take the form

$$g_{\mu,I}(\mathbf{r}) = N_\mu x^a y^b z^c \exp[-\alpha \|\mathbf{r} - \mathbf{c}\|^2], \tag{6.12}$$

where N_μ is a normalization constant, $(x,y,z) = \mathbf{r} - \mathbf{c}$ are cartesian coordinates relative to the center $\mathbf{c}$, a,b,c are non-negative integers, and $\alpha > 0$ controls the overall width of the gaussian.

Thus, each AO is not a single gaussian, but a sum as such. Gaussians are very flexible, and a few primitive gaussians can mimic very general functions.

An AO basis is chosen per molecule: For each atomic nucleus at position $\mathbf{R}_A$ with charge Z_A, a set of N_A functions $\{\phi_\mu^A(\mathbf{r})\}$ is chosen using a predetermined table of gaussian parameters. Then, the final AO basis set is taken as the union of these, giving in total $\sum_A N_A$ basis functions.

Care must be taken so that the basis set is numerically linearly independent. That is, if the overlap matrix $S_{\mu\nu} \equiv \int \overline{\phi_\mu}(\mathbf{r})\phi_\nu(\mathbf{r})\, dr$ is numerically singular, one or more basis functions can be written as linear combinations of the others. These are removed, until S is no longer numerically singular.

The predetermined table of basis functions is typically what is referred to as "the basis set". Examples are 'STO-3G', 'cc-pVDZ', and so on.

Why gaussian functions? This has a pragmatical answer: Gaussian functions are easy to compute with, and integrals involved in electronic structure calculations (see Equations .....) are analytically or semi-analytically available. They can be calculated with extreme efficiency using well-established publicly available software libraries such as `libint`. Without such libraries, quantum chemistry would be very hard business!

On the other hand, we know for a fact that the exact orbitals of a Hartree–Fock calculation behaves in a manner not compatible with gaussians: (a) They have a *cusp* at the nucleus due to the Coulomb singularity there. (b) They behave asymptotically like $\exp(-\beta r)$ for large r, again due to the long range behavior of the nuclear Coulomb potential. However, a linear combination of gaussians will *never* have a cusp, and they fall of much faster than an exponential. This results in certain unphysical behavior. For example, long-range interactions between molecules are hard to correctly model using gaussians.

AO basis sets that use exponentials instead of gaussians are called *Slater-type orbitals* (STOs). These are much harder to utilize in computations, but are better suited from a physics point of view.

A third option is *numerical basis sets*, which resolve AO radial functions on a grid. It is typical to use spherical harmonics $Y_{lm}(\theta, \phi)$ instead of cartesian monomials $x^a y^b z^c$ in this case. Numerical basis sets are best suited for Hartree–Fock calculations on atoms or atomic ions.

6.3 Restricted Hartree–Fock

6.3.1 Energy expression

We now compute the energy of the RHF wavefunction in terms of the MO coefficients. Recall the energy expression for the HF Slater determinant, Eq. (5.2). Since χ_i is a spin-orbital, and since $\hat{h}$ and $\hat{w}$ do not depend on spin, we obtain some simplifications. We first deal with the single-electron part of the energy.

$$\begin{aligned}\langle\psi_i|\hat{h}|\psi_i\rangle &= \int d\mathbf{r} \sum_\omega \overline{\psi_i(\mathbf{r})\alpha(\omega)}\hat{h}\psi_i(\mathbf{r})\alpha(\omega) \\ &= \langle\alpha|\alpha\rangle \int d\mathbf{r}\overline{\psi_i(\mathbf{r})}[-\frac{1}{2}\nabla^2 + V(\mathbf{r})]\psi_i(\mathbf{r}) \\ &= (\psi_i|\hat{h}|\psi_i).\end{aligned} \tag{6.13}$$

For a single-electron operator that does not act on spin, we have defined the *spatial integral*

$$(\psi_i|\hat{h}|\varphi_i) \equiv \int d\mathbf{r}\, \overline{\psi_i(\mathbf{r})}\hat{h}\varphi_i(\mathbf{r}) \tag{6.14}$$

Spatial operator matrix element

We obtain for the single-electron energy:

$$\sum_j \langle\chi_j|\hat{h}|\chi_i\rangle = 2\sum_{i=1}^{N/2}(\psi_i|\hat{h}|\psi_i), \tag{6.15}$$

where the factor 2 comes from the two spin directions.

We now turn to the two-electron energy. We need to evaluate a matrix element of $\hat{w}(1,2)$. To that end, we consider 4 generic spin-orbitals

$$\chi_i = \psi_i\gamma_i, \quad \gamma_i \in \{\alpha,\beta\}. \tag{6.16}$$

Computing a matrix element we find

$$\begin{aligned}\langle\chi_1\chi_2|\hat{w}|\chi_3\chi_4\rangle &= \sum_{\omega_1\omega_2}\int d\mathbf{r}_1 d\mathbf{r}_2 \overline{\psi_1(\mathbf{r}_1)\gamma_1(\omega_1)\psi_2(\mathbf{r}_2)\gamma_2(\omega_2)}\hat{w}(\mathbf{r}_1,\mathbf{r}_2)\psi_3(\mathbf{r}_1)\gamma_3(\omega_1)\psi_4(\mathbf{r}_2)\gamma_4(\omega_2)\\ &= \langle\gamma_1|\gamma_3\rangle\langle\gamma_2|\gamma_4\rangle\int d\mathbf{r}_1 d\mathbf{r}_2\, \overline{\psi_1(\mathbf{r}_1)\psi_2(\mathbf{r}_2)}\hat{w}(\mathbf{r}_1,\mathbf{r}_2)\psi_3(\mathbf{r}_1)\psi_4(\mathbf{r}_2)\\ &= \delta_{\gamma_1\gamma_3}\delta_{\gamma_2\gamma_4}(\psi_1\psi_2|\hat{w}|\psi_3\psi_4),\end{aligned} \tag{6.17}$$

where we have defined a spatial-only matrix element of $\hat{w}$. We are now ready to compute the two-electron energy. We start with the "direct part" of the antisymmetrized sum:

$$\begin{aligned}\frac{1}{2}\sum_{ii'}^{N}\langle\chi_i\chi_{i'}|\hat{w}|\chi_i\chi_{i'}\rangle &= \frac{1}{2}\sum_{jj'}^{N/2}\langle\psi_j\psi_{j'}|\hat{w}|\psi_j\psi_{j'}\rangle + \langle\overline{\psi_j}\psi_{j'}|\hat{w}|\overline{\psi_j}\psi_{j'}\rangle\\ &\quad + \langle\psi_j\overline{\psi_{j'}}|\hat{w}|\psi_j\overline{\psi_{j'}}\rangle + \langle\overline{\psi_j\psi_{j'}}|\hat{w}|\overline{\psi_j\psi_{j'}}\rangle\\ &= 2\sum_{jj'}^{N/2}(\psi_j\psi_{j'}|\hat{w}|\psi_j\psi_{j'}).\end{aligned} \tag{6.18}$$

Now for the exchange part:

$$\begin{aligned}\frac{1}{2}\sum_{ii'}^{N}\langle\chi_i\chi_{i'}|\hat{w}|\chi_{i'}\chi_i\rangle &= \frac{1}{2}\sum_{jj'}^{N/2}\langle\psi_j\psi_{j'}|\hat{w}|\psi_{j'}\psi_j\rangle + \langle\overline{\psi_j}\psi_{j'}|\hat{w}|\psi_{j'}\overline{\psi_j}\rangle\\ &\quad + \langle\psi_j\overline{\psi_{j'}}|\hat{w}|\overline{\psi_{j'}}\psi_j\rangle + \langle\overline{\psi_j\psi_{j'}}|\hat{w}|\overline{\psi_{j'}\psi_j}\rangle\\ &= \sum_{jj'}^{N/2}(\psi_j\psi_{j'}|\hat{w}|\psi_{j'}\psi_j).\end{aligned} \tag{6.19}$$

Here, two of the terms in the last sum vanish since the spins do not align.

We now obtain for the total RHF energy:

$$E_{\mathrm{RHF}} = 2\sum_i(\psi_i|\hat{h}|\psi_i) + 2\sum_{ij}(\psi_i\psi_j|\hat{w}|\psi_i\psi_j) - \sum_{ij}(\psi_i\psi_j|\hat{w}|\psi_j\psi_i). \tag{6.20}$$

RHF energy decomposed into one-body part, Hartree energy, and exchange energy

6.3.2 *Energy in AO basis*

We now express the energy in terms of the AO basis. We then insert the expansion (6.4) into Eq. (6.21). The result is:

$$E_{\text{RHF}} = 2\sum_{\mu\nu}\overline{C_{\mu i}}(\phi_\mu|\hat{h}|\phi_\nu)C_{\nu j} + 2\sum_{\mu\nu\gamma\delta}\overline{C_{\mu i}C_{\nu j}}(\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta)C_{\gamma_i}C_{\delta j} - \sum_{\mu\nu\gamma\delta}\overline{C_{\mu i}C_{\nu j}}(\phi_\mu\phi_\nu|\hat{w}|\phi_\delta\phi_\gamma)C_{\gamma_i}C_{\delta j} \tag{6.21}$$

The matrix elements $(\phi_\mu|\hat{h}|\phi_\nu)$ and $(\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta)$ are called *AO integral tensors*, and are usually computed by library routines. We will rarely have to consider them in detail, as researchers have already spent a lot of effort into handling them efficiently.

The Hartree and exchange terms look quite similar. Indeed, if we define

RHF antisymmetrized electron-repulsion integral tensor

$$[\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta] \equiv (\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta) - \frac{1}{2}(\phi_\mu\phi_\nu|\hat{w}|\phi_\delta\phi_\gamma) \tag{6.22}$$

we can write the energy more compactly as

$$E_{\text{RHF}} = 2\sum_{\mu\nu i}\overline{C_{\mu i}}(\phi_\mu|\hat{h}|\phi_\nu)C_{\nu i} + 2\sum_{\mu\nu\gamma\delta ij}\overline{C_{\mu i}C_{\nu j}}[\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta]C_{\gamma_i}C_{\delta j} \tag{6.23}$$

6.4 *The RHF equations*

We now turn to the HF equation using a RHF ansatz. Recall the HF equations:

$$\hat{f}\chi_j = \epsilon_j\chi_j.$$

We now use the definition of the RHF spin-orbitals to obtain

$$\hat{f}\psi_i\alpha = \epsilon_{2i-1}\psi_i\alpha \tag{6.24}$$

$$\hat{f}\psi_i\beta = \epsilon_{2i}\psi_i\beta. \tag{6.25}$$

We recall the Fock operator definition

The book writes $\mathcal{J} = v^{\text{H}}$ and $\mathcal{K} = v^{\text{ex}}$.

$$\hat{f} = \hat{h} + \hat{v}^{\text{H}} - \hat{v}^{\text{ex}}.$$

The Hartree and exchange operators do not change the spin of a spin-orbital as can be seen from Eqs. (5.28) and (5.25). Thus, if we project (i.e., compute the inner product) of the HF equation with a spin function, we see that we obtain two *identical* spatial eigenvalue problems, except that the eigenvalues are ϵ_{2i-1} and ϵ_{2i}, respectively. But $\psi_i(\mathbf{r})$ cannot have two different eigenvalues! We therefore conclude that

$$\epsilon_{2i-1} = \epsilon_{2i} \equiv \epsilon_i,$$

and that the Hartree–Fock equation for the RHF case reads:

$$(\hat{h} + \hat{v}^{\text{H}} - \hat{v}^{\text{ex}})\psi_i(\mathbf{r}) = \epsilon_i\psi_i(\mathbf{r}), \quad i = 1, \ldots, N/2. \tag{6.26}$$

6.5 The Roothan–Hall equations

Our next task is to express the RHF equation in an AO basis. To this end, we begin by expanding the RHS,

$$\text{RHS} = \epsilon_i \psi_i(\mathbf{r}) = \epsilon_i \sum_\nu C_{\nu i} \phi_\nu(\mathbf{r}). \tag{6.27}$$

RHS in AO basis

We next project onto ϕ_μ:

$$(\phi_\mu|\text{RHS}) = \sum_\nu \epsilon_i S_{\mu\nu} C_{\nu i} \tag{6.28}$$

We have introduced the *overlap matrix*

$$S_{\mu\nu} \equiv (\phi_\mu|\phi_\nu) = \int d\mathbf{r}\, \overline{\phi_\mu(\mathbf{r})} \phi_\nu(\mathbf{r}). \tag{6.29}$$

We next consider the left-hand side, and we begin with the operator $\hat{h}$. We again introduce the AO expansion for ψ_i and project onto ϕ_μ:

First part of LHS in AO basis

$$(\phi_\mu|\hat{h}|\psi_i) = \int d\mathbf{r}\, \overline{\phi_\mu(\mathbf{r})} \hat{h} \sum_\nu C_{\nu i} \phi_\nu(\mathbf{r}) = \sum_\nu (\phi_\mu|\hat{h}|\phi_\nu) C_{\nu i}. \tag{6.30}$$

We next turn to the Hartree potential, which involves the electron density:

$$\rho(\mathbf{r}) = 2\sum_j |\psi_j(\mathbf{r})|^2 = \sum_{j\gamma\delta} \overline{C_{\gamma j}} C_{\delta_j} \overline{\phi_\gamma(\mathbf{r})} \phi_\delta(\mathbf{r}). \tag{6.31}$$

The Hartree potential becomes

$$v^{\text{H}}(\mathbf{r}) = \int d\mathbf{s}\, \frac{\rho(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} = 2\sum_{\gamma\delta j} \overline{C_{\gamma j}} C_{\delta j} \int d\mathbf{s}\, \frac{\overline{\phi_\gamma(\mathbf{s})}\phi_\delta(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \tag{6.32}$$

We now act with this operator on $\psi_i(\mathbf{r})$ and obtain:

$$v^{\text{H}}(\mathbf{r})\psi_i(\mathbf{r}) = 2\sum_{\gamma\delta j} \overline{C_{\gamma j}} C_{\delta j} \int d\mathbf{s}\, \frac{\overline{\phi_\gamma(\mathbf{s})}\phi_\delta(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \sum_\nu C_{\nu i} \phi_\nu(\mathbf{r}). \tag{6.33}$$

Finally, we project onto $\phi_\mu(\mathbf{r})$ and obtain:

Second part of LHS in aO basis

$$\begin{aligned}(\phi_\mu|v^{\text{H}}|\psi_i) &= 2\sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta j} \int d\mathbf{r} \int d\mathbf{s}\, \overline{\phi_\mu(\mathbf{r})} \frac{\overline{\phi_\gamma(\mathbf{s})}\phi_\delta(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \phi_\nu(\mathbf{r}) C_{\nu i} \\ &= 2\sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta j} (\phi_\mu\phi_\gamma|\hat{w}|\phi_\nu\phi_\delta) C_{\nu i}\end{aligned} \tag{6.34}$$

We next turn to the exchange operator:

$$[\hat{v}^{\text{ex}}\psi_i](\mathbf{r}) = \int d\mathbf{s} \sum_j \frac{\overline{\psi_j(\mathbf{s})}\psi_i(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \psi_j(\mathbf{r}). \tag{6.35}$$

Here, there is no factor 2, since in Eq. (5.28) the sum over spin ω_2 only gives a contribution for those spin-orbitals that have the same spin as $\psi(\mathbf{r})\alpha$ (or β). Inserting the AO expansions gives

$$[\hat{v}^{\text{ex}}\psi_i](\mathbf{r}) = \sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta j} \int d\mathbf{s}\, \frac{\overline{\phi_\gamma(\mathbf{s})}\phi_\delta(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \phi_\nu(\mathbf{r}) C_{\nu i}. \tag{6.36}$$

Projection onto ϕ_μ finally gives

$$\begin{aligned}(\phi_\mu|\hat{v}^{\text{ex}}\psi_i) &= \sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta i} \int d\mathbf{r} \int d\mathbf{s}\, \overline{\phi_\mu(\mathbf{r})} \frac{\overline{\phi_\gamma(\mathbf{s})}\phi_\delta(\mathbf{s})}{|\mathbf{r}-\mathbf{s}|} \phi_\nu(\mathbf{r}) C_{\nu j} \\ &= \sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta i} (\phi_\mu\phi_\gamma|\hat{w}|\phi_\nu\phi_\delta) C_{\nu j} \\ &= \sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta j} (\phi_\mu\phi_\gamma|\hat{w}|\phi_\delta\phi_\nu) C_{\nu i}\end{aligned} \tag{6.37}$$

In the last step, we renamed $\delta \leftrightarrow \nu$.

6.6 *Putting all together*

We now combine all the terms of the left-hand side as well as the right-hand side, and obtain:

$$\sum_\nu (\phi_\mu|\hat{h}|\phi_\nu) C_{\nu i} + 2 \sum_{\gamma\delta\nu j} \overline{C_{\gamma j}} C_{\delta j} [\phi_\mu\phi_\gamma|\hat{w}|\phi_\delta\phi_\nu] C_{\nu i} = \epsilon_i \sum_\nu (\phi_\mu|\phi_\nu) C_{\nu i}, \tag{6.38}$$

where we introduced the antisymmetrized electron-repulsion integral (ERI) tensor from Eq. (6.22).

This equation is quite compact! But we can do better: We introduce the *density matrix* as

$$D_{\nu\mu} \equiv (2CC^H)_{\nu\mu} = 2\sum_j C_{\nu j}\overline{C_{\mu j}}. \tag{6.39}$$

Notice that the left-hand side of the RHF equations depend on the orbitals through $D_{\mu\nu}$. We furthermore introduce several other matrices:

$$H_{\mu\nu} \equiv (\phi_\mu|\hat{h}|\phi_\nu) \tag{6.40}$$

$$J_{\mu\nu} \equiv (\phi_\mu|\hat{v}^{\text{H}}|\phi_\nu) = \sum_{\gamma\delta} D_{\delta\gamma} (\phi_\mu\phi_\gamma|\hat{w}|\phi_\nu\phi_\delta) \tag{6.41}$$

$$K_{\mu\nu} \equiv (\phi_\mu|\hat{v}^{\text{ex}}|\phi_\nu) = \frac{1}{2}\sum_{\gamma\delta} D_{\delta\gamma} (\phi_\mu\phi_\gamma|\hat{w}|\phi_\delta\phi_\nu) \tag{6.42}$$

$$V^{\text{HF}}_{\mu\nu} \equiv J_{\mu\nu} - K_{\mu\nu} = \sum_{\delta\gamma} D_{\delta\gamma} [\phi_\mu\phi_\gamma|\hat{w}|\phi_\nu\phi_\delta] \tag{6.43}$$

$$F_{\mu\nu} \equiv H_{\mu\nu} + V^{\text{HF}}_{\mu\nu} \tag{6.44}$$

We now obtain the following form of the Roothan–Hall equation (6.38):

Roothan–Hall equation

$$FC = SC\epsilon, \quad \epsilon = \text{diag}(\epsilon_1, \cdots, \epsilon_{N/2}) \tag{6.45}$$

Since column number i of C forms the MO coefficients of ψ_i, we write $c_i \equiv C_{:,i}$, i.e.,

$$C = \begin{bmatrix} | & | & \cdots & | \\ c_1 & c_2 & \cdots & c_{N/2} \\ | & | & \cdots & | \end{bmatrix}. \tag{6.46}$$

We can therefore write the Roothan–Hall equations as:

$$Fc_i = \epsilon_i S c_i. \tag{6.47}$$

This may more clearly illustrate that we are dealing with a *genarlized eigenvalue problem*: It contains an overlap matrix S instead of the identity on the right-hand side. . Additionally, since F depends on the solution, this is a *nonlinear generalized eigenvalue problem*.

We can choose to split the Fock matrix F into its constituents in different ways:

Fock matrix split into plain one-body operator and nonlinear Hartree and Exchange matrices.

$$F = H + V^{\mathrm{HF}} = H + J - K. \tag{6.48}$$

We observe that $F = F(D) = F(2CC^H)$ depends on the orbitals through the density matrix.

6.7 Occupied and virtual orbitals

The RHF equations result in a number $N/2$ of HF orbitals. However, F is a square matrix of dimension equal to the number of AOs, N_{AO}. This matrix is Hermitian, so we can compute its eigenvectors, resulting in a *square matrix* C_{full}, (full diagonalization):

$$F(D)C_{\mathrm{full}} = SC_{\mathrm{full}}\epsilon_{\mathrm{full}}.$$

Assuming that the eigenvalues are sorted according to increasing eigenvalue (which is typically done by common Hermitian eigenvalue problem solvers), we then have that the first $N/2$ columns are the occupied orbital MO coefficients. The remaining $N_{\mathrm{AO}} - N/2$ columns are *virtual orbitals*:

$$C_{\mathrm{full}} = [C \mid C_{\mathrm{virt}}]. \tag{6.49}$$

$$\epsilon_{\mathrm{full}} = \mathrm{diag}(\underbrace{\epsilon_1, \cdots, \epsilon_{N/2}}_{\epsilon}, \underbrace{\epsilon_{N/2+1}, \cdots, \epsilon_{N_{\mathrm{AO}}}}_{\epsilon_{\mathrm{virt}}}). \tag{6.50}$$

It is customary to use the index i for occupied MOs, and a for virtual MOs. Thus the columns of C are denoted c_i and the columns of C_{virt} are denoted c_a. Similarly, the orbital energies are ϵ_i and ϵ_a for the virtual orbitals. This is very handy for post-Hartree–Fock methods. Thus, noting the naming of the summation index,

$$D = 2CC^H = 2\sum_i c_i c_i^H$$

and, the RHF equations for the occupied MOs alone become

$$F(D)C = SC\epsilon \iff F(D)c_i = \epsilon_i S c_i \quad \text{(occupied)}$$

and for the virtual MOs alone

$$F(D)C_{\mathrm{virt}} = SC_{\mathrm{virt}}\epsilon_{\mathrm{virt}} \iff F(D)c_a = \epsilon_a S c_a \quad \text{(virtual)}$$

and if we write the equations together, we get the full diagonalization of the Fock matrix,

$$F(D)C_{\mathrm{full}} = SC_{\mathrm{full}}\epsilon_{\mathrm{full}} \iff F(D)c_p = \epsilon_p S c_p \quad \text{(general)}.$$

We stress that the columns of C_{virt} are the virtual orbital MO coefficients, and are not part of the RHF wavefunction, but can be used to produce better wavefunctions using *Post-Hartre–Fock methods* like CI or coupled-cluster methods.

6.8 Self-consistent field method

The RHF working equations (Roothan–Hall) are nonlinear equations that must be solved *self-consistently*. To see this, note that the Fock operator $F = F(D)$ is a function of the density matrix $D = 2CC^H$. So, if we compute the eigenvectors of $F(D)$, we will get a new matrix C', and if $C' \neq C$, we have not solved our RHF equations!

The standard approach is the *self-consistent field (SCF) method*, where an initial guess $C^{(0)}$, with density matrix $D^{(0)}$, is used to find a new guess by diagonalizing $F(D^{(0)})$, resulting in $\epsilon^{(1)}$ and $C^{(1)}$. The process is repeated with $C^{(1)}$, to yeld $\epsilon^{(2)}$ and $C^{(2)}$, etc., until some convergence criterion has been reached.

Thus,

$$\begin{aligned} F(D^{(0)})C^{(1)} &= SC^{(1)}\epsilon^{(1)} \\ F(D^{(1)})C^{(2)} &= SC^{(2)}\epsilon^{(2)} \\ &\vdots \\ F(D^{(n)})C^{(n+1)} &= SC^{(n+1)}\epsilon^{(n+1)} \\ &\vdots \end{aligned}$$

Each iteration is costly, requiring the solution of a *generalized eigenvalue problem* $FC_{\text{full}} = SC_{\text{full}}\epsilon_{\text{full}}$, which costs on the order of N_{AO}^3 FLOPs to diagonalize the Fock matrix. Moreover, the iterations may not converge. It is therefore usual to implement *extrapolation/-convergence acceleration methods*. In quantum chemistry, the most common method is *direct inversion in iterative subspace* (DIIS) invented by P. Pulay. We will not discuss this scheme here.

In Section 3.4.5 of Szabo & Ostlund, the generalized eigenvalue problem is converted to a standard eigenvalue problem $F'C' = C'\epsilon$ via *orthogonalization*. We will not cover this procedure here. It is possible to avoid orthogonalization by solving the generalized eigenvalue problem directly.

6.9 Setting up the SCF iterations

We now consider the task of solving the RHF iterations via SCF iterations. A choice has been made regarding

- The molecule: The nuclear positions and the total number of electrons, an even number.
- The AO basis set.

In order to perform the SCF iterations, the following ingredients must be at hand, obtainable via library functions that we will *not* code ourselves. PySCF can readily give us these ingredients:

- The matrix elements $H_{\mu\nu}$ of the one-body Hamiltonian, often called the *core Hamiltonian*
- The overlap matrix elements $S_{\mu\nu}$
- The ERI tensor $(\phi_\mu\phi_\gamma|\hat{w}|\phi_\nu\phi_\delta)$

To bootstrap the SCF iterations, one needs an initial guess for the density matrix $D = CC^H$. A natural choice is obtained by diagonalizing the core Hamiltonian, which amounts to ignoring the electron repulsion:

$$D^{(0)} = C^{(0)}(C^{(0)})^H, \quad HC^{(0)} = SC^{(0)}\epsilon^{(0)}. \tag{6.51}$$

In Python, one would do something like:

```
# Import generalized eigenvector routine
from scipy.linalg import eigh

# Perform complete diagonalization.
# This routine solves H C = S C E,
# and returns eigenvalues sorted according
# to increasing value.
E_occ_and_virt, C_occ_and_virt = eigh(H, S)

# Extract the first N//2 eigenvectors
C = C_occ_and_virt[:,:N//2]

# Compute density matrix
D = 2 * C @ C.T
```

In order to do an SCF iteration starting from a guess $D^{(n)}$ of the density matrix, one now must:

1. Compute the Fock matrix $F(D^{(n)}) = H + J(D^{(n)}) - K(D^{(n)})$. This is typically done using for loops, or contraction routines such as `np.einsum` for a faster alternative.
2. Diagonalize the Fock matrix:
$$FC^{(n+1)}_{\text{full}} = SC^{(n+1)}_{\text{full}}\epsilon^{(n+1)}_{\text{full}}$$
3. Extract the occupied MO coefficients, assuming that $\epsilon_0 \leq \epsilon_1 \leq \cdots$:
$$C^{(n+1)}_{\text{full}} = [C^{(n+1)} \mid C^{(n+1)}_{\text{virt}}].$$
See the above code snippet for the diagonalization of the core Hamiltonian for the general procedure.
4. Compute the new density matrix $D^{(n+1)}$ from the occupied MO coefficients.
5. Determine if the iteration has converged. This can be done in several ways:

- Check if the MO coefficient matrix has converged, i.e., if $C^{(n+1)} - C^{(n)}$ is sufficiently small using, e.g., `np.linalg.norm`
- Check if the HF eigenvalues have converged, i.e., if $\epsilon^{(n+1)} - \epsilon^{(n)}$ is sufficiently small, using, e.g., root mean square (RMS) error, or maximum absolute deviation (MAD).

6. If the iteration has not converged, or if the maximum number of allowed iterations $n_{\max}$ have been reached, set $n \leftarrow n + 1$ and repeat from the top.

6.10 *DIIS acceleration of iterations*

In many cases, the SCF iterations fail to converge, or they converge slowly. It is common to apply the *direct inversion in the iterative subspace* (DIIS) acceleration technique due to P. Pulay to overcome non-convergence. The idea is that after a number of iterations n_0 (typically a handful, 4–7), these iterations are combined to a better guess. The SCF iterations then proceed using this better guess. The DIIS acceleration is then located typically after point 3. in the above iteration outline.

DIIS is defined for general iteration schemes: Suppose we want to solve $f(\mathbf{x}_*) = 0$, and we have guesses $\mathbf{x}_i$, $i = 1, 2, \ldots, n_{\text{DIIS}}$ available, together with *error vectors* $\mathbf{e}_i$, such that $\mathbf{x}_i \approx \mathbf{x}_* + \mathbf{e}_i$. These errors are of course not known exactly. The idea is to find coefficients c_i such that $\sum_i c_i = 1$ and set

$$\mathbf{x}_{\text{DIIS}} = \sum_{i=1}^{n_{\text{DIIS}}} c_i \mathbf{x}_i.$$

The cefficients c_i are chosen so that the resulting error is minimal, i.e.,

$$\| \sum_i c_i \mathbf{e}_i \| = \text{min!}$$

Why do the coefficients have to sum to 1? Consider that for all iterations,

$$\mathbf{x}_i \approx \mathbf{x}_* + \mathbf{e}_i.$$

The DIIS guess satisfies

$$\mathbf{x}_* \approx \mathbf{x}_{\text{DIIS}} - \mathbf{e}_{\text{DIIS}} = \sum_i c_i \mathbf{x}_i - \mathbf{e}_{\text{DIIS}} = \sum_i c_i \mathbf{x}_* + \sum_i c_i \mathbf{e}_i - \mathbf{e}_{\text{DIIS}}.$$

We see that unless $\sum_i c_i = 1$, the error terms will not be small.

The DIIS algorithm is a constrained optimization problem with Lagrangian

$$L(c_1, \cdots, c_{n_{\text{DIIS}}}, \mu) = \| \sum_i c_i \mathbf{e}_i \|^2 - \mu (\sum_i c_i - 1).$$

Here is a generic Python implementation:

```
def diis_extrapolate(x_list, e_list):
    """
```

```
    Perform DIIS extrapolation given a list of
        solution vectors and corresponding error
        vectors.

    Parameters:
        x_list : list of np.ndarray
            List of previous solution vectors.
        e_list : list of np.ndarray
            List of corresponding error/residual
                vectors.

    Returns:
        np.ndarray
            The DIIS-extrapolated solution vector.
    """
    m = len(e_list)

    # Build the B-matrix from error inner products.
    B = np.empty((m, m))
    for i in range(m):
        for j in range(m):
            # Use ravel() to ensure the vectors are
                1-D
            B[i, j] = np.dot(e_list[i].ravel(),
                e_list[j].ravel())

    # Construct the augmented matrix for the DIIS
        linear system:
    #  [ B    -1 ]
    #  [ -1^T  0 ]
    A = np.empty((m + 1, m + 1))
    A[:m, :m] = B
    A[:m, m] = -1.0
    A[m, :m] = -1.0
    A[m, m] = 0.0

    # Right-hand side: zeros except for the last
        element being -1.
    rhs = np.zeros(m + 1)
    rhs[m] = -1.0

    # Solve for the coefficients (the last entry is
        the Lagrange multiplier)
    sol = np.linalg.solve(A, rhs)
    coeffs = sol[:m]

    # Form the extrapolated solution as the linear
        combination of past iterates.
    x_new = np.zeros_like(x_list[0])
    for c, x in zip(coeffs, x_list):
        x_new += c * x
    return x_new
```

For SCF iterations, the guesses $\mathbf{x}_i$ are the iterates $C^{(n+i-1)}$, and the error vectors are the corresponding *residuals* of the Roothan–

Hall equations,

$$R^{(n)} = F(D^{(n)})C^{(n)} - SC^{(n)}\epsilon^{(n)}.$$

Even if these are not exactly "error vectors", since $R^{(n)}$ do not estimate the error in $C^{(n)}$ directly, the residual typically "behaves like" the error, especially close to the solution, in the sense that the residuals and errors are roughly linearly related.

6.11 Unrestricted Hartree–Fock

In unrestricted Hartree–Fock, the α and β parts of the HF equations are no longer identical. Instead they will be coupled through the Hartree potential.

We here summarize the equations, but will not do a detailed derivation.

We consider the general case where there are N_α and N_β spins of each type. We expand the orbitals:

$$\psi_i^\alpha(\mathbf{r}) = \sum_\nu C_{\nu i}^\alpha \phi_\nu(\mathbf{r}), \quad i = 1, 2, \ldots, N_\alpha \tag{6.52}$$

$$\psi_i^\beta(\mathbf{r}) = \sum_\nu C_{\nu i}^\beta \phi_\nu(\mathbf{r}), \quad i = 1, 2, \ldots, N_\beta \tag{6.53}$$

We define α and β density matrices:

$$D^\alpha = C^\alpha C^{\alpha H}, \quad \text{i.e.,} \quad D_{\nu\mu}^\alpha \equiv \sum_{j=1}^{N_\alpha} C_{\nu j}^\alpha \overline{C_{\mu j}^\alpha} \tag{6.54}$$

$$D^\beta = C^\beta C^{\beta H}, \quad \text{i.e.,} \quad D_{\nu\mu}^\beta \equiv \sum_{j=1}^{N_\beta} C_{\nu j}^\beta \overline{C_{\mu j}^\beta} \tag{6.55}$$

$$D = D^\alpha + D^\beta. \tag{6.56}$$

The Roothan–Hall equations become the *Pople–Nesbet equations*

$$(H + J - K^\alpha)C^\alpha = SC^\alpha\epsilon^\alpha \tag{6.57}$$

$$(H + J - K^\beta)C^\beta = SC^\beta\epsilon^\beta. \tag{6.58}$$

Here, the exchange operator depends on spin:

$$K_{\mu\nu}^\alpha \equiv \sum_{\gamma\delta} D_{\delta\gamma}^\alpha (\phi_\mu \phi_\gamma | \hat{w} | \phi_\delta \phi_\nu) \tag{6.59}$$

$$K_{\mu\nu}^\beta \equiv \sum_{\gamma\delta} D_{\delta\gamma}^\beta (\phi_\mu \phi_\gamma | \hat{w} | \phi_\delta \phi_\nu). \tag{6.60}$$

6.12 UHF contains RHF as solutions

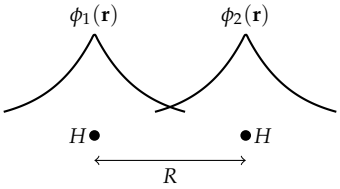
For RHF, we have $N_\alpha = N_\beta = N/2$. Additionally, $D^\alpha = D^\beta = D/2$. We see that the α and β parts of the Pople–Nesbet equations become identical to the Roothan–Hall equations, and thus there is a solution $C^\alpha = C^\beta = C$ (where C is the RHF MO coefficient matrix) to the Pople–Nesbet equations.

There are often other solutions as well. This is investigated for the hydrogen molecule in the next section.

6.13 UHF vs RHF applied to the hydrogen molecule

We will now do a study of the RHF and UHF model in a minimal basis for the hydrogen molecule with H atoms 1 and 2 located a distance R from each other.

In a minimal basis, there is exactly one AO per electron for the H atoms: A single 1s function centered at proton 1, and a single 1s function located on proton 2. Using the STO-3G basis set, each 1s function (a Slater-type orbital) is approximated with 3 primitive gaussians. We obtain 2 AOs $\phi_1(\mathbf{r})$ and $\phi_2(\mathbf{r})$. (Here, the numbering is conveniently associated with a nucleus. This is not in general true, of course.)



6.13.1 Form of the MOs

For RHF, there is a single doubly occupied MO:

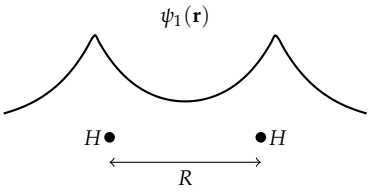
$$\psi_1(\mathbf{r}) = c_{11}\phi_1(\mathbf{r}) + c_{12}\phi_2(\mathbf{r}).$$

$$|\Phi^{\text{RHF}}\rangle = |\psi_1\overline{\psi_1}\rangle\,.$$

Appealing to the symmetry of the problem, the charge density must be symmetric about the inversion point located at $R/2$. This implies that

$$c_{11} = \pm c_{12}.$$

Physical intuition tells us that energy is lowered by having a *symmetric and nodeless MO*, i.e., $c_{11} = c_{12}$. This holds for *all* internuclear distances R. This MO is illustrated in the margin to the right.



Bonding MO, $c_{11} = c_{12}$.

We observe, that at very large separations, say a lightyear, the MO is *still* on this form, indicating that *each electron is forced to be delocalized*, with probability $1/2$ to be at each nucleus. This is absurd, and reflects that RHF is *not size-consistent*, which we discuss below.

We now turn to UHF. For UHF, we have separate α and β MOs, each singly occupied. These are free to vary:

$$\psi_1^\alpha(\mathbf{r}) = c_{11}^\alpha\phi_1(\mathbf{r}) + c_{12}^\alpha\phi_2(\mathbf{r}),$$

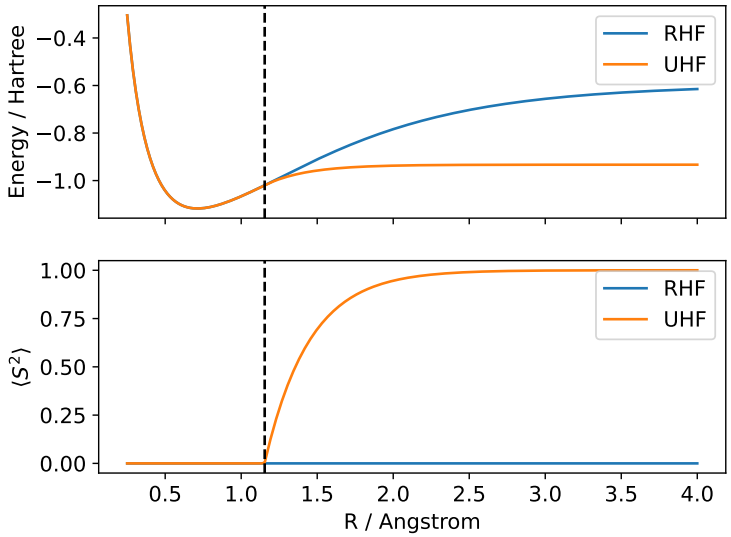
$$\psi_1^\beta(\mathbf{r}) = c_{11}^\beta\phi_1(\mathbf{r}) + c_{12}^\beta\phi_2(\mathbf{r}),$$

$$|\Phi^{\text{UHF}}\rangle = |\psi_1^\alpha\overline{\psi_1^\beta}\rangle\,.$$

As previously discussed, there is a solution to the UHF problem where $\psi_1^\alpha = \psi_1^\beta = \psi^1$ (from the RHF model). However, when the atoms are very far apart, a solution appears where $\psi_1^\alpha \approx \phi_1$ and $\psi_1^\beta \approx \psi_2$. This solution is energetically more favorable. Thus each atom has a single localized electron, which is the physically reasonable situation. Importantly, this type of solution is not possible in RHF, which constrains orbitals to be doubly occupied. This illustrates that for this molecule at least, the UHF model is *size-consistent*, as discussed further below.

6.13.2 *Plot of the RHF and UHF energy curves*

Here is a plot of the UHF and RHF energy curves for the hydrogen molecule. We have also plotted the expectation value of $\hat{S}^2$ – the total electron spin, for each calculation.



We have marked a point, $R_0 \approx 1.15$ Å, where it seems like the UHF solution starts to deviate from the RHF solution. Indeed, for $R < R_0$, the RHF solution is obtained from the UHF calculaton: $\phi_1^\alpha = \psi_1^\beta$. For $R > R_0$, the UHF solution no longer matches the RHF solution: $\psi_1^\alpha \neq \psi_1^\beta$.

Recall that for two electrons, the total spin is either 0 or 1 for an eigenfunction of the Hamiltonian. It is desirable that approximate wavefunctions are eigenfunctions of spin, too. The total spin of RHF is always zero. This is because all orbitals are doubly occupied. We say that the RHF wavefunction is a singlet. But in the plot we see that the UHF wavefunction is in general *not an eigenfunction of spin!* In fact, we see that at R_0, the spin suddenly starts to grow, and gets closer to 1 as $R \to +\infty$. The UHF wavefunction makes a gradual transition from a singlet to a triplet state! Another way of saying this is that UHF becomes *spin contaminated*.

The correct ground-state wavefunction for the hydrogen molecule is the siglet state near equilibrium, correctly predicted by RHF, but a triplet at large separations, which RHF is not able to describe. Note the artificially high energy for RHF in the dissociation limit.

6.13.3 *Size-consistency*

In the above example, RHF was not *size-consistent*: As we pull the molecule apart, the ground-state energy computed is *not* equal to the ground state energy of two isolated atoms. In UHF, on the other hand, this is the case. Each electron is localized on its proton in this limit.

Size-consistency is an important concept in electronic-structure theory. A computational method that is not size-consistent will often yield unphysical results as we move around the potential energy surface.

6.14 Exercises

6.14.1 Exercise 1: Deriving the Roothaan–Hall equations for RHF using Lagrange multipliers

In Lecture 5, the continous Hartree–Fock equations were derived by taking general variations in the single-electron functions that satisfied the constraints. We then projected these equations onto the AO basis in the current lecture, Lecture 6. Notably, we did not *derive* the Roothan–Hall equations. We will now derive the Roothan–Hall equations using the method of Lagrange multipliers. You learned about the method of Lagrange multipliers in the exercises of Lecture 5.

Assume an AO basis $\{\phi_\mu(\mathbf{r})\}_{\mu=1}^K$ and an RHF ansatz with N even and $i = 1, \ldots, N/2$,

$$\chi_{2i-1}(\mathbf{r},\omega) = \psi_i(\mathbf{r})\alpha(\omega), \qquad \chi_{2i}(\mathbf{r},\omega) = \psi_i(\mathbf{r})\beta(\omega),$$

where each spatial MO is expanded as

$$\psi_i(\mathbf{r}) = \sum_\mu C_{\mu i}\phi_\mu(\mathbf{r}).$$

Define AO overlap and one-electron integrals

$$S_{\mu\nu} = \int d\mathbf{r}\,\overline{\phi_\mu(\mathbf{r})}\phi_\nu(\mathbf{r}), \qquad h_{\mu\nu} = (\phi_\mu|\hat{h}|\phi_\nu) \equiv \int d\mathbf{r}\,\overline{\phi_\mu(\mathbf{r})}\,\hat{h}\,\phi_\nu(\mathbf{r}),$$

and two-electron AO integrals

$$(\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta) \equiv \iint d\mathbf{r}_1 d\mathbf{r}_2\,\frac{\overline{\phi_\mu(\mathbf{r}_1)\phi_\nu(\mathbf{r}_2)}\,\phi_\gamma(\mathbf{r}_1)\phi_\delta(\mathbf{r}_2)}{r_{12}}.$$

Use the RHF antisymmetrized tensor (as in the lecture)

$$[\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta] \equiv (\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta) - \frac{1}{2}(\phi_\mu\phi_\nu|\hat{w}|\phi_\delta\phi_\gamma).$$

Assume the RHF energy written in the AO basis is

$$E_{\mathrm{RHF}}[C] = 2\sum_{\mu\nu i}\overline{C_{\mu i}}\,h_{\mu\nu}\,C_{\nu i} + 2\sum_{\mu\nu\gamma\delta\,ij}\overline{C_{\mu i}}\,\overline{C_{\nu j}}\,[\phi_\mu\phi_\nu|\hat{w}|\phi_\gamma\phi_\delta]\,C_{\gamma i}C_{\delta j}.$$

Problem 1: The orthonormality constraint in an AO basis

1. The RHF MOs are required to satisfy

   $$(\psi_i|\psi_j) = \delta_{ij}.$$

   Show that in the AO basis this becomes

   $$\sum_{\mu\nu}\overline{C_{\mu i}}\,S_{\mu\nu}\,C_{\nu j} = \delta_{ij}.$$

2. Write the same condition in matrix form. Let C be the $K \times (N/2)$ matrix with entries $C_{\mu i}$. Show that the orthonormality condition is

   $$C^\dagger S C = I,$$

   where S is the AO overlap matrix and I is the $(N/2) \times (N/2)$ identity matrix.

Problem 2: Set up the variational problem with Lagrange multipliers

1. Let$\Lambda = (\lambda_{ij})$ (dimension $(N/2) \times (N/2)$) be a matrix of Lagrange multipliers. Explain how the following functional is the proper Lagrangian functional to rewrite the constrained optimization of the RHF energy as an unconstrained optimization problem:

$$\mathcal{L}(C, \Lambda) = E_{\mathrm{RHF}}[C] - 2\sum_{i,j} \lambda_{ij} \Big((\psi_i | \psi_j) - \delta_{ij} \Big).$$

2. Rewrite the constraint term entirely in AO quantities. Hence express $\mathcal{L}$ only in terms of $C_{\mu i}$, $\overline{C_{\mu i}}$, λ_{ij}, and the AO tensors.

3. Λ is expected to be Hermitian at a stationary point. Can you explain this?

Problem 3: First variation of the Lagrangian

We now perform the unconstrained variations, treating $C_{\mu i}$ and $\overline{C_{\mu i}}$ as independent variables.

1. Show that

$$\delta(\psi_i | \psi_j) = \sum_{\mu\nu} \delta\overline{C_{\mu i}}\, S_{\mu\nu}\, C_{\nu j} + \sum_{\mu\nu} \overline{C_{\mu i}}\, S_{\mu\nu}\, \delta C_{\nu j}.$$

2. Show that the first-order variation of the one-electron energy term is

$$\delta E^{(1)} = 2\sum_{\mu\nu i} \delta\overline{C_{\mu i}}\, h_{\mu\nu}\, C_{\nu i} + 2\sum_{\mu\nu i} \overline{C_{\mu i}}\, h_{\mu\nu}\, \delta C_{\nu i}.$$

3. Consider the two-electron energy term

$$E^{(2)} = 2 \sum_{\mu\nu\gamma\delta\, ij} \overline{C_{\mu i}}\, \overline{C_{\nu j}}\, [\phi_\mu \phi_\nu | \hat{w} | \phi_\gamma \phi_\delta]\, C_{\gamma i} C_{\delta j}.$$

   Compute the part of $\delta E^{(2)}$ that is linear in $\delta\overline{C_{\mu i}}$ (hold $\delta C = 0$). Write your answer as

$$\delta E^{(2)}\big|_{\delta\overline{C}} = 2\sum_{\mu i} \delta\overline{C_{\mu i}}\, (\cdots)_{\mu i},$$

   where you must identify the expression $(\cdots)_{\mu i}$ explicitly as a sum over AO and MO indices.

Problem 4: Introduce the RHF density matrix and the Fock matrix

1. Define the RHF AO density matrix

$$P_{\gamma\delta} = 2 \sum_{j=1}^{N/2} C_{\gamma j}\, \overline{C_{\delta j}}.$$

   Explain briefly the origin of the factor 2.

2. Using $P_{\gamma\delta}$, show that the object multiplying $\delta\overline{C_{\mu i}}$ from Problem 3.3 can be rewritten in the form

$$\sum_{\nu} G_{\mu\nu}\, C_{\nu i},$$

where

$$G_{\mu\nu} = \sum_{\gamma\delta} P_{\gamma\delta}\, [\phi_\mu \phi_\nu | \hat{w} | \phi_\gamma \phi_\delta].$$

3. Define the RHF Fock matrix in the AO basis as

$$F_{\mu\nu} = h_{\mu\nu} + G_{\mu\nu}.$$

Rewrite $F_{\mu\nu}$ in terms of Hartree and exchange terms, exhibiting $P_{\mu\nu}$ in the resulting expressions.

4. Show that the part of δE_{RHF} linear in $\delta\overline{C_{\mu i}}$ can be written compactly as

$$\delta E_{\text{RHF}}\big|_{\delta\overline{C}} = 2\sum_{\mu\nu i} \delta\overline{C_{\mu i}}\, F_{\mu\nu}\, C_{\nu i}.$$

Problem 5: Stationarity condition and the Roothaan–Hall equations

1. Write the part of $\delta\mathcal{L}$ that is linear in $\delta\overline{C_{\mu i}}$:

$$\delta\mathcal{L}\big|_{\delta\overline{C}} = 2\sum_{\mu\nu i} \delta\overline{C_{\mu i}}\, F_{\mu\nu}\, C_{\nu i} - 2\sum_{i,j} \lambda_{ij} \sum_{\mu\nu} \delta\overline{C_{\mu i}}\, S_{\mu\nu}\, C_{\nu j}.$$

2. Use the fact that $\delta\overline{C_{\mu i}}$ are arbitrary numbers to show that $\delta\mathcal{L} = 0$ if and only if

$$\sum_{\nu} F_{\mu\nu} C_{\nu i} = \sum_{\nu}\sum_{j} S_{\mu\nu} C_{\nu j} \lambda_{ji} \qquad \text{for all } \mu, i.$$

3. Show that this equation has the matrix form

$$FC = SC\Lambda.$$

This is the *Roothaan–Hall* generalized eigenvalue problem for RHF.

4. Show that if you choose a unitary transformation among the occupied orbitals that diagonalizes Λ, then the equations become the *canonical* Roothaan–Hall equations

$$FC = SC\varepsilon,$$

where ε is diagonal with elements ϵ_i (orbital energies).

Problem 6: Self-consistency

1. Explain why the Roothaan–Hall equations are not a single linear eigenvalue problem: which matrix depends on the solution C itself?

2. Using your definition of $P_{\gamma\delta}$, explain briefly how a practical RHF algorithm iterates: $C \to P \to F \to C$.

7 Lecture 7: Computing expectation values in Hartree–Fock & Gaussian basis sets

7.1 Observables and expectation values

What we obtain in a Hartree–Fock calculation is a set of (spin-)orbitals, that together comprise the (approximate) wavefunction of the system. The wavefunction is the *state* of the system, the totality of information we need to make any physical prediction.

Physical quantities that we can measure are called *observables*. These are represented by Hermitian operators. Examples include:

- Kinetic energy, $\hat{T} = \sum_{i=1}^{N} \hat{t}(i)$.
- Nuclear potential energy, $\hat{V}_{\text{nuc}} = \sum_{i=1}^{N} V_{\text{nuc}}(\mathbf{r}_i)$
- Total electronic energy, with operator $\hat{H}_{\text{el}}$
- Electronic repulsion energy, $\hat{W} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} |\mathbf{r}_1 - \mathbf{r}_2|^{-1}$
- Total electronic energy, with operator $\hat{H}_{\text{el}} = \hat{T} + \hat{V}_{\text{nuc}} + \hat{W}$
- Dipole moment, a vector-valued observable, with operators $\hat{\boldsymbol{\mu}} = (\hat{\mu}_x, \hat{\mu}_y, \hat{\mu}_x)$,

$$\hat{\boldsymbol{\mu}} = e \sum_{i=1}^{N} \mathbf{r}_i.$$

  In atomic units the elementary charge is $e = -1$.

Many observables are obtained from classical observables by means of operator substitution. All the examples above are of this type.

The *expectation value* of an observable $\hat{\Omega}$ when the state of the system is $|\Psi\rangle$ is

$$\langle \Omega \rangle = \frac{\langle \Psi | \hat{\Omega} | \Psi \rangle}{\langle \Psi | \Psi \rangle}. \tag{7.1}$$

The expectation value is the mean outcome of a large number of identical experiments on identically prepared systems. However, the individual outcomes of an experiment are eigenvalues of the operator.

It is important to realize, that the wavefunction itself is not an observable. There is no experiment on a quantum system we can do to reveal its wavefunction.

7.2 The HF total energy

7.2.1 Electronic energy

The most immediate observable we have access to is the total electronic energy. Indeed, this is the observable that we minimize, so its expectation value is just the Hartree–Fock energy. Introducing the density matrix into Eq. (6.21), we obtain

$$\langle \hat{H} \rangle = \sum_{\mu\nu} D_{\mu\nu} H_{\nu\mu} + \frac{1}{2} \sum_{\mu\nu} D_{\mu\nu} V^{\mathrm{HF}}_{\nu\mu} = \mathrm{Tr}(DH) + \frac{1}{2}\mathrm{Tr}(DV^{\mathrm{HF}}) \quad (7.2)$$

Recall that the trace of a matrix is the sum of its diagonal elements: If $A = [A_{\mu\nu}]$ is a matrix, then $\mathrm{Tr}(A) = \sum_\mu A_{\mu\mu}$. The trace satisfies $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$.

Make sure you see from Eq. (6.21) where the factor $1/2$ comes from! In general, the density matrix has an α and β component,

$$D_{\mu\nu} = D^{\alpha}_{\mu\nu} + D^{\beta}_{\mu\nu},$$

and in RHF the two components are identical, while in UHF they may differ.

It is in general true that the expectation value of a one-body operator $\hat{A} = \sum_{i=1}^{N} \hat{a}(\mathbf{r}_i)$ is obtained as

$$\langle \hat{A} \rangle = \langle \Phi^{\mathrm{HF}} | \sum_{i=1}^{N} \hat{a}(i) | \Phi^{\mathrm{HF}} \rangle = \mathrm{Tr}(DA), \quad A_{\mu\nu} = \langle \phi_\mu | \hat{a} | \phi_\nu \rangle .$$

The density matrix is a general concept, and can be defined for *any* ab into quantum chemical method, not only the Hartree–Fock method, in such a way that expectation values take this form.

Here, it is assumed that the operator $\hat{A}$ does *not act on spin*. On the other hand, the Roothan–Hall equations state that

$$(H + V^{\mathrm{HF}})C = SC\epsilon.$$

Multiplying with C^H from the left, we see that

$$C^H(H + \mathrm{HF})C = \epsilon.$$

Taking the trace, we obtain:

$$\sum_{i=1}^{N/2} \epsilon_i = \mathrm{Tr}(C^H(H + V^{\mathrm{HF}})C) = \frac{1}{2}\mathrm{Tr}(H + V^{\mathrm{HF}})D).$$

Here, we used that $CC^H = D/2$, and that $\mathrm{Tr}(XYZ) = \mathrm{Tr}(ZXY)$. From this we obtain:

$$E_{\mathrm{RHF}} = 2\sum_{i=1}^{N/2} \epsilon_i - \frac{1}{2}\mathrm{Tr}(V^{\mathrm{HF}}D). \quad (7.3)$$

Thus, the sum of the orbital energies is *not* the total Hartree–Fock energy!

7.2.2 Adding nuclear repulsion energy

The total energy of a molecule in the clamped-nuclei approximation also includes nuclear repulsion energy. Since the nuclei are treated classically, this is just a constant shift of the electronic energy, independent of the electronic wavefunction:

$$E_{\mathrm{tot}}(\{\mathbf{R}_A\}) = E_{\mathrm{RHF}}(\{\mathbf{R}_A\}) + \sum_{A=1}^{N_{\mathrm{atom}}-1} \sum_{B=A+1}^{N_{\mathrm{ATOM}}} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}. \quad (7.4)$$

This energy is usually included in the RHF energy. This is true for PySCF, in particular. Note that for geometry optimization, where we try to determine the geometry of a molecule at the minimum total energy, this is the relevant energy to optimize.

7.3 Total dipole moment

The classical dipole moment of a collection of charged classical particles with coordinates $\mathbf{r}_i$ and charges q_i is given by

$$\boldsymbol{\mu} = \sum_i q_i \mathbf{r}_i. \tag{7.5}$$

In electronic-structure theory, the nuclei are classical, while the electrons are quantum mechanical, leadint to the total dipole moment

$$\boldsymbol{\mu}_{\text{tot}} = \langle \Psi | \hat{\boldsymbol{\mu}} | \Psi \rangle + \sum_A Z_A \mathbf{R}_A. \tag{7.6}$$

Using the above formula for the electronic expectation value in terms of the spatial density matrix, we obtain

$$\langle \Psi | \boldsymbol{\mu} | \Psi \rangle_k = \sum_{\mu\nu} D_{\mu\nu} (\phi_\nu | r_k | \phi_\mu), \quad k \in \{x, y, z\}. \tag{7.7}$$

We see that in addition to the (spin)orbitals of our HF calculation, we also need the integrals of the operators x, y, and z in the AO basis. This is easy to obtain using integration libraries.

The above expression is valid for RHF, UHF, and GHF.

7.3.1 Spin and space density

Consider a Slater determinant built from spin-orbitals. It gives spin-dependent density matrices D^α and D^β, such that the total density matrix is $D = D^\alpha + D^\beta$. Where does the term "density matrix" come from? The answer is that it is the basis expansion coefficients of the *density* $\rho(\mathbf{r})$:

$$\rho(\mathbf{r}) = \sum_{\mu\nu} D_{\nu\mu} \phi_\nu(\mathbf{r}) \overline{\phi_\mu(\mathbf{r})}. \tag{7.8}$$

Why is the density useful? Let us look at the expectation value of a local potential $v(\mathbf{r})$,

$$\hat{V} = \sum_{i=1}^{N} v(\mathbf{r}_i).$$

$$\begin{aligned} \langle V \rangle &= \sum_{j=1}^{N} \int dx_1 \cdots dx_N \overline{\Phi^{\text{HF}}(x_1, \cdots, x_N)} v(\mathbf{r}_j) \Phi^{\text{HF}}(x_1, \cdots, x_N) \\ &= \sum_{j=1}^{N} \langle \chi_j | v | \chi_j \rangle = \sum_{j=1}^{N} \sum_{\omega} \int d\mathbf{r} |\chi_j(\mathbf{r}, \omega)|^2 v(\mathbf{r}), \end{aligned} \tag{7.9}$$

where we used the Slater–Condon rules. For UHF, half of the χ_j have spin $m_{s,j} = \alpha$, and spatial MO coefficients C^α, and the rest have spin $m_{s,j} = \beta$, and spatial MO coefficients C^β:

$$\chi_{2i-1} = \sum_\mu C^\alpha_{\mu i} |\alpha\rangle .$$

$$\chi_{2i} = \sum_{\mu} C^{\beta}_{\mu i} |\beta\rangle .$$

Thus,

$$\begin{aligned}\sum_{j=1}^{N}\sum_{\omega}|\chi_j(\mathbf{r},\omega)|^2 &= \sum_{j}\sum_{\mu\nu}[C^{\alpha}_{\mu j}\overline{C^{\alpha}_{\nu j}} + C^{\beta}_{\mu j}\overline{C^{\beta}_{\nu j}}]\phi_\mu(\mathbf{r})\overline{\phi_\nu(\mathbf{r})} \\ &= \sum_{\mu\nu}[D^{\alpha}_{\mu\nu} + D^{\beta}_{\mu\nu}]\phi_\mu(\mathbf{r})\overline{\phi_\nu(\mathbf{r})} \\ &= \sum_{\mu\nu} D_{\mu\nu}\phi_\mu(\mathbf{r})\overline{\phi_\nu(\mathbf{r})} \equiv \rho(\mathbf{r}).\end{aligned} \tag{7.10}$$

We now compute the expectation value as an integral, and obtain:

$$\langle V\rangle = \int \rho(\mathbf{r})v(\mathbf{r}). \tag{7.11}$$

For RHF, we only get the slight modification that $D^{\alpha} = D^{\beta}$.

For UHF/RHF we observe the interesting fact that

$$D^{\alpha}SD^{\alpha} = D^{\alpha}, \quad \mathrm{Tr}(D^{\alpha}S) = N_{\alpha}, \tag{7.12}$$

$$D^{\beta}SD^{\beta} = D^{\beta}, \quad \mathrm{Tr}(D^{\beta}S) = N_{\beta}. \tag{7.13}$$

More generally, the density can in turn be viewed as the expectation value of the *density operator* $\hat{\rho}(\mathbf{r})$,

$$\hat{\rho}(\mathbf{r}_0) = \sum_{i=1}^{N}\delta^3(\mathbf{r}-\mathbf{r}_0).$$

That is, the expectation value of a "potential" which is zero everywhere, except at the position $\mathbf{r}_0$, where it is "infinite". The density can thus be defined for all quantum mechanical methods, not only Hartree–Fock, and its expression will depend on the method in question.

The delta "potential" can now be used as a basis for expanding a general potential! Write

$$v(\mathbf{r}) = \int d\mathbf{s}\; v(\mathbf{s})\delta^3(\mathbf{r}-\mathbf{s}).$$

Then we compute

$$\langle V\rangle = \int d\mathbf{s}\; v(\mathbf{s})\,\langle\delta^3(\mathbf{r}-\mathbf{s})\rangle = \int d\mathbf{s}\; v(\mathbf{s})\rho(\mathbf{s}).$$

This shows the universality of the expectation value of a one-body potential as an integral over the density. This will become very useful when we learn about density-functional theory!

7.4 *Gaussian AO basis sets*

We have previously briefly discussed gaussian basis sets. We now do a slightly more comprehensive discussion.

7.4.1 *Numerical vs. algebraic approach*

The phrase "numerical" has for chemists the connotation "grid resolved MOs". This is rather uncommon, and proably arose historically in the dichotomy between the analytic simple "algebraic approach" and the use of a grid, where the basis functions were "numerical" rather than "analytical". In science, "numerical methods" comprise all computational methods that use finite precision arithmetic – even the use of AO basis sets!

There are two basic approaches to parameterizing a molecular orbital: The "numerical approach" resolves the MO on a grid. This is feasible only for atoms or linear molecules, where the molecule possesses a high degree of symmetry. In these cases, the numerical approach can yield highly accurate results. Numerical approaches include finite difference methods, finite element methods, pseudospectral methods, and so on. We will not have opportunity to discuss numerical approaches in this course.

For other molecules, the "algebraic approach" remains the only feasible method. Here, the MO is expanded in a a fixed set of *basis functions*,

$$\psi_i = \sum_\mu C^\alpha_{\mu i} \phi_\mu(\mathbf{r}) \left|\sigma\right\rangle, \quad \sigma \in \{\alpha, \beta\}.$$

These functions are called "basis functions" since they are interpreted as the basis for a finite-dimensional space of functions.

What basis functions are "good" basis functions?

Firstly, a good basis set should be systematically extendable towards a *complete* basis set, that can resolve any one-electron functions. It is unavoidable that any complete basis set is infinitely large, since one-electron Hilbert space is infinite dimensional. Mathematically, a complete basis set is such that for *any* one-electron function $\psi(\mathbf{r})$, and for any $\epsilon > 0$, we can find coefficients c_μ such that

$$\|\psi - \sum_{\mu=1}^{\infty} c_\mu \phi_\mu\| < \epsilon.$$

Secondly, the systematic extension of the basis set should lead to *rapid* convergence for MOs, so that only a small number of basis functions are needed for highly accurate calculations.

Thirdly, the basis functions should be easy to manipulate and use in various calculations, such as evaluating integrals needed for quantum chemical calculations.

Orthogonal or near-orthogonal functions are better than highly non-orthogonal functions. It should be possible to invert the overlap matrix without much problems.

In practice it is hard to satisfy all requirements at the same time. There are always compromises in quantum chemistry.

7.4.2 *Atom centered basis functions*

The by far most common approach in the algebraic approach is to use "atomic orbitals" that are centered on the atoms in the molecule. This is natural, as molecules consist of atoms, and since each atom after all remains a largely identifiable unit of a quantum mechanical molecule. Thus, a good description of each atom should give a good description of the overall molecule – at least in principle!

It is important to realize, that even if the basis functions are called "atomic orbitals", they do not have to have any physical sig-

nificance, i.e., they are not exact solutions of some important physical problem. Their name is largely historical, and their sole purpose is to give a fast convergent, practical, and physically reasonable description of molecular wavefunctions.

If we consider a single electron atom (such as a hydrogen atom), the exact eigenfunctions are on the form

$$\psi_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_{lm}(\theta,\phi),$$

where $R_{nl}(r)$ is a radial function, and where Y_{lm} is a *spherical harmonic*. This angular symmetry is also inherited for Hartree–Fock calculations on neutral atoms or ions. It is therefore natural to seek atomic orbitals with spherical symmetry.

If we consider an N-electron atom, and the desire to let ψ_{nlm} be Hartree–Fock eigenstates for the atom, we know that $R_{nl}(r)$ is in general continuously differentiable and exponentially falling off as $r \to +\infty$, i.e., it behaves line $\exp(-\beta r)$ for some $\beta > 0$. For small r, it behaves like r^l due to the boundary conditions for nonzero angular momentum. How does it behave for intermediate r? It turns out that *any* radial function can be expanded in *Laguerre functions*, leading to Slater-type orbitals. We discuss this below, after dealing with the angular part first, since this is the simplest, and really fixed once and for all.

7.4.3 *The angular part*

An alternative exists in using *cartesian coordinates*.

The angular part of the AOs can be taken to be spherical harmonics in some form or another. (There are different ways to write them.) The spherical harmonics $Y_{lm}(\theta,\phi)$ are eigenfunctions of the orbital angular momentum operator $\hat{L}^2$ and its z-projection $\hat{L}_z$:

$$\hat{L}^2 Y_{lm}(\theta,\phi) = l(l+1)Y_{lm}(\theta,\phi), \tag{7.14}$$

$$\hat{L}_z Y_{lm} = mY_{lm}(\theta,\phi). \tag{7.15}$$

We have

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta), e^{im\phi} \tag{7.16}$$

where $P_l^m(x)$ are the *associated Legendre polynomials*. The numbers l and m satisfy

$$l = 0, 1, 2, \cdots, \quad |m| \leq l.$$

There is a very nice generating expression for the associated Legendre polynomials, called a *Rodrigues expression*,

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{m/2}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l. \tag{7.17}$$

7.5 *Laguerre expansion and Slater type orbitals*

The radial part $R_{nl}(r)$ depends only on the principal quantum number n and the total angular momentum l. A fixed l consstitutes

an *angular momentum shell*. In other words, for a fixed l, there are $2l+1$ functions with the same n and l. Thus, R_{nl} gives rise to $2l+1$ AOs with identical radial part.

As previously mentioned, the radial part behaves like $\exp(-\zeta r)$ for large r, and is continuously differentiable. A complete set of functions are the *Laguerre functions*, given by

$$R_{nl}^{\text{Laguerre}}(r) = (2\zeta)^{3/2}\sqrt{\frac{(n-l-1)!}{(n+l+1)!}}(2\zeta r)^l L_{n-l-1}^{2l+2}(2\zeta r)\exp(-\zeta r).$$

Here, L_{n-l-1}^{2l+2} are *associated Laguerre polynomials* of degree $n-l-1$. These functions are also orthonormal, and lead to rapid convergence of the expansion of MOs.

Completeness means that

$$R(r) = \sum_{n=l+1}^{\infty} c_n R_{nl}^{\text{Laguerre}}(r).$$

Looking at the radial Laguerre function $R_{nl}^{\text{Laguerre}}(r)$, we see that the factor $(2\zeta r)^l$ makes the total polynomial prefactor being of the form

$$a_0(2\zeta r)^l + \cdots + a_{n-1}(2\zeta r)^{n-1}$$

leading to a polynomial of degree $n-1$. Thus, we can instead work with AOs that have *monomial* prefactors $(2\zeta r)^k$, with k ranging from l to infinity. The monomials also constitute a complete basis, which is simpler, but no longer orthonormal. This leads to *Slater type orbitals*:

$$R_n^{\text{STO}}(r) = \frac{(2\zeta)^{3/2}}{\sqrt{\Gamma(2n+1)}}(2\zeta r)^{n-1}\exp(-\zeta r). \tag{7.18}$$

The prefactor is chosen so the STO is normalized. The complete AO including angular parts becomes

$$\phi_{nlm}^{\text{STO}}(r,\theta,\phi) = R_n^{\text{STO}}(r)Y_{lm}(\theta,\phi). \tag{7.19}$$

By varying n, the original Laguerre basis functions can be reobtained. The STO orbitals are therefore *complete*.

The complicated-looking normalization factor makes each STO normalized, but they are not orthogonal for different n.

7.6 Gaussian type orbitals

The Slater-type orbitals are ideal for flexibly representing physically correct atomic orbitals. They have the correct physical behavior for both small, intermediate, and large r. However, they are not so easy to compute with. For example, electron-repulsion integrals become quite cumbersome. So cumbersome in fact, that STOs are not that popular.

An alternative exists in *Gaussian type orbitals* (GTOs). These are popular because gaussian functions are able to approximate almost any function by taking linear combinations. Moreover, they lead to

analytical integrals, that are computed extremely fast with modern algorithms, needing only a few CPU cycles per integral!

Chemists are always ready to throw some important physical properties out of the window if it can allow calculations that are faster, on larger molecules, still retaining accuracy!

They do have the drawback that they are not able to describe the *cusp* of STOs and electronic wavefunctions in general at $r = 0$, and that they fall off too fast as $r \to +\infty$.

Let $\alpha > 0$ be an exponent, and define *the spherical primitive gaussian* as

$$g^{\mathrm{GTO}}_{\alpha l m}(r, \theta, \phi) = R^{\mathrm{GTO}}_{\alpha l} Y_{lm}(\theta, \phi). \tag{7.20}$$

The radial part is

$$R^{\mathrm{GTO}}_{\alpha l} = N_{\alpha l} r^l e^{-\alpha r^2}. \tag{7.21}$$

The normalization constant is

$$N_{\alpha l} = \frac{(\sqrt{2\alpha})^{l+3/2}}{\pi^{1/4}} \sqrt{\frac{2^{l+2}}{(2l+1)!!}} \tag{7.22}$$

Let l be given. We wish to define a set of AOs with this angular momentum. We define a vector α^l_k (which has different lengths M_l for each l) of exponents. Then, we define the *radial part* of the AOs for quantum number l (and m) as a linear combination of gaussians with these exponents:

$$R_{nl}(r) = \sum_{k=1}^{M_l} d^l_{kn} R^{\mathrm{GTO}}_{\alpha^l_k l}(r), \quad n = 1, 2, \ldots, N_l. \tag{7.23}$$

This results in a total of $N_l \times (2l+1)$ AOs with angular momentum l:

$$\phi_{nlm}(r) = R_{nl} Y_{lm}(\theta, \phi), \quad |m| \leq l, \; n = 1, 2, \ldots, N_l. \tag{7.24}$$

7.6.1 *Example: STO-3G*

Basis Set Exchange is an online databases of gaussian basis sets for quantum chemistry. Let us look up an example, with the STO-3G basis set for hydrogen and beryllium. This basis set would then be used for a calculation on compounds such as BeH_2 (which is a typical benchmark model system).

https://www.basissetexchange.org/basis/aug-cc-pvtz/format/nwchem/?version=1&elements=1,4

```
1  #----------------------------------------------------------------------
2  # Basis Set Exchange
3  # Version 0.11
4  # https://www.basissetexchange.org
5  #----------------------------------------------------------------------
6  #   Basis set: STO-3G
7  # Description: STO-3G Minimal Basis (3 functions/AO)
8  #        Role: orbital
9  #     Version: 1  (Data from Gaussian09)
10 #----------------------------------------------------------------------
11
12
13 BASIS "ao basis" SPHERICAL PRINT
14 #BASIS SET: (3s) -> [1s]
15 H    S
16       0.3425250914E+01        0.1543289673E+00
17       0.6239137298E+00        0.5353281423E+00
18       0.1688554040E+00        0.4446345422E+00
19 #BASIS SET: (6s,3p) -> [2s,1p]
20 Be   S
21       0.3016787069E+02        0.1543289673E+00
22       0.5495115306E+01        0.5353281423E+00
23       0.1487192653E+01        0.4446345422E+00
24 Be   SP
25       0.1314833110E+01       -0.9996722919E-01        0.1559162750E+00
26       0.3055389383E+00        0.3995128261E+00        0.6076837186E+00
27       0.9937074560E-01        0.7001154689E+00        0.3919573931E+00
28 END
```

Lone 1–10 form a comment block. Line 13 states that an AO basis in spherical coordinates are given. At line 14, a comment states

that we take 3 primitives of angular momentum $l = 0$ (s functions) and build 1 AO with $l = 0$. This is an approximation to an STO, hence the name of the basis set. Line 15 states that we give a basis for hydrogen (H) with $l = 0$ (S). Next follows the specification: The first column is α_k^0, and the second column is d_{kn}^0, the expansion coefficients for the nth function, and there is only one function here. In total we get 1 AO for the H atom, since $l = 0$.

At line 19, another comment states that we take 6 s functions and produce 2 s AOs, and 3 p-functions and produce 1 p function.

At line 20, the beryllium basis starts, and the format is identcal to the previously discussed part for the hydrogen atom.

At line 24 starts a block labeled SP, which means that the exponents are used *both* for an s function and a p function. Thus, the first column contains the exponents, the second column expansion coefficients for an s-function, the third column contains a p function.

The SP block could have been split in two blocks, one S block and a P block, but in the current format, some storage space is saved.

What do the functions look like?

TODO: Insert a visualization

7.6.2 *Example: Dunning basis sets.*

Wheras the STO-3G basis set is "minimal", containing just enough functions for a Hartree–Fock calculation, correlated calculations using DFT, coupled-cluster theory, or CI theory requires larger basis sets. A popular family of basis sets is the Dunning *correlation consistent* basis sets cc-pVXZ:

cc= "correlation consistent"

p= "polarized"

V= "valence"-only correlation

XZ= quality level, starting at DZ (double-zeta), TZ (triple-zeta), etc. Each level includes more and more valence polarization functions, i.e., higher angular momenta.

If the prefix "aug-" is added, then *diffuse* functions are added. These are useful for improving long-range interactions, since the gaussians decay too fast for large r. The key idea is to obtain systematic convergence towards the complete basis set limit (CBS limit). The prefix "d-aug-" gives double augmented basis sets, with very diffuse functions.

Let us look at the aug-cc-pVDZ basis set for Be and H:

```
1 BASIS "ao basis" SPHERICAL PRINT
2 #BASIS SET: (5s,2p) -> [3s,2p]
3 H    S
4       1.301000E+01           1.968500E-02           0.000000E+00
5       1.962000E+00           1.379770E-01           0.000000E+00
6       4.446000E-01           4.781480E-01           0.000000E+00
7       1.220000E-01           5.012400E-01           1.000000E+00
8 H    S
9       0.0297400              1.0000000
```

```
10 H     P
11         7.270000E-01            1.0000000
12 H     P
13         0.1410000               1.0000000
14 #BASIS SET: (10s,5p,2d) -> [4s,3p,2d]
15 Be    S
16         2.940000E+03            6.800000E-04           -1.230000E-04           0.000000E+00
17         4.412000E+02            5.236000E-03           -9.660000E-04           0.000000E+00
18         1.005000E+02            2.660600E-02           -4.831000E-03           0.000000E+00
19         2.843000E+01            9.999300E-02           -1.931400E-02           0.000000E+00
20         9.169000E+00            2.697020E-01           -5.328000E-02           0.000000E+00
21         3.196000E+00            4.514690E-01           -1.207230E-01           0.000000E+00
22         1.159000E+00            2.950740E-01           -1.334350E-01           0.000000E+00
23         1.811000E-01            1.258700E-02            5.307670E-01           0.000000E+00
24         5.890000E-02           -3.756000E-03            5.801170E-01           1.000000E+00
25 Be    S
26         1.790000E-02            1.000000E+00
27 Be    P
28         3.619000E+00            2.911100E-02            0.000000E+00
29         7.110000E-01            1.693650E-01            0.000000E+00
30         1.951000E-01            5.134580E-01            0.000000E+00
31         6.018000E-02            4.793380E-01            1.000000E+00
32 Be    P
33         1.110000E-02            1.000000E+00
34 Be    D
35         2.354000E-01            1.0000000
36 Be    D
37         7.220000E-02            1.000000E+00
38 END
```

We have dropped the header block for simplicity.

Lines 3–13 define the H basis set. There are 2 S blocks, and 2 P blocks. The first S block defines 2 s-functions in terms of 4 primitive gaussians, while the next S block adds a single diffuse gaussian (note the exponent value). The P blocks contain one gaussian each. The last block is the diffuse function.

The two diffuse functions are not present in the un-augmented basis cc-pVDZ, otherwise the AOs are identical.

Lines 15–27 define the Be basis set. The first block defines 3 S functions in terms of 9 primitives, and the next block has one single diffuse S function. It is interesting to note that the second S function has both positive and negative coefficients, signifying at least 1 node in the radial function!

The P blocks are similarly defined, and we also here have 2 D blocks, with $l = 2$.

8 Lecture 8: Configuration interaction

8.1 Recap

Suppose we have done an RHF or UHF calculation, resulting in a set of N occupied spin-orbitals $\chi_i(\mathbf{r})$ and a remaining set of $N_u = 2N_{\text{AO}} - N$ unoccupied/virtual orbitals $\chi_a(\mathbf{r})$. Only the occupied orbitals take part in the Slater determinant that is the Hartree–Fock state:

We set $N_o = N$ and $N_u = 2N_{\text{AO}} - N$.

$$|\Phi_0\rangle \equiv |\Phi^{\text{HF}}\rangle = |\chi_1 \cdots \chi_N\rangle \,.$$

In Chapter **??** we studied the CI method in a general setting where we did not distinguish between occupied and unoccupied orbitals. We derived the *Slater–Condon rules* for the matrix elements of one and two-body operators.

The goal of this section is to introduce CI as a *post-Hartree–Fock model*, where we systematically improve upon Hartree–Fock. We will put the Slater–Condon rules to use.

8.2 Excited determinants and the CI hierarchy

The Hartree–Fock state has none of the unoccupied orbitals. On the other hand, we may be dissatisfied with the quality of the results of the RHF method. In particular, we may find that the energy E_{HF} was not good enough. Since the HF wavefunction is uncorrelated (except for the exchange correlation due to the Pauli principle), we set

$$E_0 = E_{\text{HF}} + E_{\text{corr}},$$

where E_{corr} is the *correlaton energy*, and where E_0 is the exact ground state energy. Similarly, the ground state wavefunction is correlated

$$|\Psi_0\rangle = c_0 \, |\Phi_0\rangle + |\Psi_{\text{corr}}\rangle \,.$$

The number c_0 is required to keep the total wavefunction normalized, when we assume that

$$\langle \Phi_0 | \Psi_{\text{corr}} \rangle = 0.$$

We need to find a recipe for $|\Psi_{\text{corr}}\rangle$. One option is the configuration-interaction hierarchy. Another option is the coupled-cluster hierarchy; see later in the course.

Suppose we improve upon the RHF wavefunction by allowing *single excitations*, yielding the CIS wavefunction:

$$|\Psi^{\mathrm{CIS}}\rangle = c_0 |\Phi_0\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle \,. \tag{8.1}$$

Here, $|\Phi_i^a\rangle$ is obtained from the wavefunction by *substituting* χ_i *by* χ_a:

$$|\Phi_i^a\rangle = |\chi_1 \cdots \underset{\text{was } i}{\chi_a} \cdots \chi_{N_o}\rangle \,.$$

There are clearly $N_o \times N_v$ possible such substitutions, since i takes on N_o values and a takes on N_v such values.

We can note that the placement of indices as sub- and superscript pictorially indicate that particles are "excited" from lower-lying levels to higher-lying levels, energetically speaking.

The next level of improvement would be to include *double exitations*, yielding the CISD wavefunction:

$$|\Psi^{\mathrm{CISD}}\rangle = c_0 |\Phi_0\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle + \sum_{i<j} \sum_{a<b} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \,. \tag{8.2}$$

Here, the restriction to $i < j$ and $a < b$ ensures that we sum only over unique Slater determinants.

We can continue in this fashion, defining CISDT, where triple excitations are included, CISDTQ which includes quarduples, and all the way up to N_o-fold excitations, which is called full CI (FCI).

Clearly the higher we are willing to go, the better the wavefunction will be. On the other hand, the number of excited determinants grows rapidly with the truncation level (S, D, T, etc.). CISD already contains $N_o^2 N_v^2/4$ excited determinants. Full CI at the extreme end contains $\binom{N_o+N_v}{N_o}$ determinants – this renders full CI completely intractable but for the simplest molecules in very small basis sets!

8.3 *The CI eigenvalue problem*

As previously discussed in an earlier lecture, the CI wavefunction is determined by matrix diagonalization, i.e., the CI wavefunction is given by an eigenvector of a matrix, Eq. (4.4). We now have introduced a *block structure* of the CI wavefunction, starting with the HF reference, then singles, doubles, etc:

$$\mathbf{c} = \begin{pmatrix} c_0 \\ \{c_i^a\} \\ \{c_{ij}^{ab}\} \\ \vdots \end{pmatrix} \tag{8.3}$$

The Hamiltonian matrix has a corresponding structure:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{00} & \mathbf{H}_{0,S} & \mathbf{H}_{0,D} & \cdots \\ \mathbf{H}_{S,0} & \mathbf{H}_{S,S} & \mathbf{H}_{S,D} & \cdots \\ \mathbf{H}_{D,0} & \mathbf{H}_{D,S} & \mathbf{H}_{D,D} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{8.4}$$

Here, $\mathbf{H}_{00} = [\langle \Phi_0 | \hat{H} | \Phi_0 \rangle]$ is a simple 1×1 matrix. The matrix $\mathbf{H}_{0,S}$ is an $1 \times (N_o \times N_v)$ mtarix with elements $\langle \Phi_0 | \hat{H} | \Phi_i^a \rangle$. The matrix $\mathbf{H}_{S,S}$ contains matrix elements that couples single excitations with single excitations: $\langle \Phi_{i'}^{a'} | \hat{H} | \Phi_i^a \rangle$. And so on and so forth. As we progress towards higher extitations, the matrix blocks become larger and larger.

Thus, for CISD one would need the matrix blocks written down explicitly in Eq. (8.4).

Note that since $\hat{H} = \hat{H}^\dagger$, the matrix $\mathbf{H} = \mathbf{H}^H$ is Hermitian. Thus, for example, $\mathbf{H}_{0,S} = \mathbf{H}_{S,0}^H$.

8.4 *Brillouin's Theorem*

There is one matrix block of the CI matrix which is identically zero when we start from the Hartree–Fock ground state as reference Φ_0 for excitations:

$$\mathbf{H}_{0,S} = \mathbf{H}_{S,0}^H = 0.$$

This is *Brillouin's Theorem* and follows directly from the Hartree–Fock equations and the Slater–Condon rulles.

The Slater–Condon rules give the following expression:

$$\langle \Phi_i^a | \hat{H} | \Phi_0 \rangle = \langle \chi_a | \hat{h} | \chi_i \rangle + \sum_j \langle \chi_a \chi_j | \hat{w} | \chi_i \chi_j \rangle_{\text{AS}} . \tag{8.5}$$

On the other hand, we have the following expression for the Fock operator:

$$\hat{f} = \hat{h} + \sum_j \langle \cdot \, \chi_j | \hat{w} | \cdot \, \chi_j \rangle_{\text{AS}} . \tag{8.6}$$

Computing the action on χ_i we get

$$\hat{f} \, | \chi_i \rangle = \hat{h} \, | \chi_i \rangle + \sum_j \langle \cdot \, \chi_j | \hat{w} | \chi_i \chi_j \rangle_{\text{AS}} . \tag{8.7}$$

Forming the inner product with χ_a gives

$$\langle \chi_a | \hat{f} \, | \chi_i \rangle = \langle \chi_a | \hat{h} \, | \chi_i \rangle + \sum_j \langle \chi_a \chi_j | \hat{w} | \chi_i \chi_j \rangle_{\text{AS}} . \tag{8.8}$$

But since we have a Hartree–Fock solution in the spin-orbitals χ_i, we have that

$$\hat{f} \, | \chi_i \rangle = \epsilon_i \, | \chi_i \rangle ,$$

which means that $\langle \chi_a | \hat{f} | \chi_i \rangle = 0$. Thus,

$$\langle \Phi_i^a | \hat{H} | \Phi_0 \rangle = \langle \chi_a | \hat{f} | \chi_i \rangle = 0.$$

8.5 *Simplifications from RHF*

[To be written]

8.6 *Dissociation of H_2*

[To be written]

8.7 *Truncated CI is not size-consistent*

[To be written]

9 Lecture 9: Coupled-cluster theory

9.1 Introduction

We now turn to the most popular post-Hartree–Fock wavefunction method in quantum chemistry. (Density–Functional Theory is by far the workhorse in quantum chemistry, but it is not a wavefunction theory!) The method is the *coupled-cluster method* (CC). Like CI, it is divided into a hierarchy of approximate theories: CCS (CC with singles), CCSD (with singles and doubles), etc. The most widely-used approximation is termed CCSD(T) – the CC method with singles, doubles and perturbative triples. The latter means that while triples are included, they are only treated approximately.

Why introduce CC theory? The motivaion stems from the fact that CI theory is not size-extensive. We will see a Jupyter Notebook where an experiment is performed on several water molecules. CISD is not able to give the correct energy as we let the water molecules drift apart, while CCSD does: CCSD is (under appropriate conditions) *size-extensive/consistent*.

9.2 Excitation/substitution operators

In CI theory, we worked with a linear combination of *excited determinants*, where we took occupied spin-orbitals (indexed by i, j, etc.) in the HF determinant and replaced them by virtual spin-orbitals (indexed by a, b, etc.):

$$|\Psi^{CISD}\rangle = c_0 |\Phi_0\rangle + \sum_{ia} c_i^a |\Phi_i^a\rangle + \sum_{i<j} \sum_{a<b} c_{ij}^{ab} |\Phi_{ij}^{ab}\rangle \,. \tag{9.1}$$

Recall that the singly excited determinants were obtained by *substitution of occupied spin-orbitals*. We can formulate this as an operator, which will allow us to define the CC method.

Recall that an operator is determined by its action on a basis. We therefore need to formulate the action of the substitution on a basis. This is pretty straightforward, since a basis is given by the collection of Slater determinants.

9.2.1 Singles

We define an *operator* X_i^a that performs this substitution:

$$|\Phi_i^a\rangle = X_i^a |\Phi_0\rangle \,. \tag{9.2}$$

This defines the operator on $|\Phi_0\rangle$, but we can also define it on any Slater determinant: Let $|\chi_{p_1}\cdots\chi_{p_N}\rangle$ be a Slater determinant. There are two possibilities: Either $p_k = i$ for some k. In that case, we set

$$X_i^a\,|\chi_{p_1}\cdots\chi_{p_N}\rangle \equiv |\chi_{p_1}\cdots \underset{\text{was } i}{\chi_a} \cdots\chi_{p_N}\rangle \quad (\chi_i \text{ present})$$

If $p_k \neq i$ for every k, then we set

$$X_i^a\,|\chi_{p_1}\cdots\chi_{p_N}\rangle \equiv 0 \quad (\chi_i \text{ not present})$$

This defines X_i^a on every Slater determinant.

The operator X_i^a is called an *excitation operator* since it kicks χ_i up to χ_a, or a *substitution operator*, since it substitutes χ_i with χ_a. Both names are common and equally good.

If we now let $|\Psi\rangle$ be an arbitrary vector in FCI space, it is a general linear combination of Slater determinants:

Recall that I is an index that counts the HF function with $I = 0$, singles, doubles, etc.

$$|\Psi\rangle = \sum_I c_I\,|\Phi_I\rangle\,. \tag{9.3}$$

If we now act with X_i^a, defined as a linear operator, we get

$$|\Psi'\rangle = X_i^a\,|\Psi\rangle = \sum_I c_I X_i^a\,|\Phi_I\rangle\,. \tag{9.4}$$

We see that having defined X_i^a on arbitrary Slater determinants is enough to define X_i^a on any CI vector!

9.2.2 *Doubles and higher*

Recall that a doubles excitation from the HF determinant is just substitution of i by a and then j by a. Thus, we have

Double substitution/excitation

$$X_{ij}^{ab} = X_j^b X_i^a,$$

the operator product. Since X_i^a is defined as an operator on FCI space, applying two such operators in succession is also a well-defined operator.

[A couple of examples]

We can define triples similarly:

Triple substitution/excitation

$$X_{ijk}^{abc} = X_k^c X_j^b X_i^a,$$

Quadruples, quintuples, etc., follow in the same manner.

9.2.3 *Excitations commute*

It does not matter if we first substitute i to a, then j to b, or the other way around. Mathematically, the excitation operators satisfy

$[\hat{A},\hat{B}] = 0$ means that $\hat{A}\hat{B}\,|\Psi\rangle = \hat{B}\hat{A}\,|\Psi\rangle$ for every wavefunction $|\Psi\rangle$. Thus the order of operation does not matter.

$$[X_i^a, X_j^b] = X_i^a X_j^b - X_j^b X_i^a = 0.$$

always. They *commute*, and the order of application does not matter.

9.2.4 *CI theory with excitation operators*

The next step is to use the excitation operators to reformulate the CI wavefunction. Since

$$|\Phi_i^a\rangle = X_i^a |\Phi_0\rangle\,, \quad |\Phi_{ij}^{ab}\rangle = X_{ij}^{ab} |\Phi_0\rangle\,,$$

and so on, we can write

$$\begin{aligned} |\Psi^{\text{CISD}}\rangle &= c_0 |\Phi_0\rangle + \sum_{ia} c_i^a X_i^a |\Phi_0\rangle + \sum_{ijab} c_{ij}^{ab} X_{ij}^{ab} |\Phi_0\rangle\,. \\ &= (c_0 \hat{1} + \sum_{ia} c_i^a X_i^a + \sum_{ijab} c_{ij}^{ab} X_{ij}^{ab}) |\Phi_0\rangle\,. \end{aligned} \tag{9.5}$$

We observe that the ting in parenthesis is an operator, being a linear combination of excitation operator:

$$\begin{aligned} \hat{C} &= c_0 \hat{1} + \sum_{ia} c_i^a X_i^a + \sum_{ijab} c_{ij}^{ab} X_{ij}^{ab} \\ &= c_0 \hat{1} + \hat{C}_1 + \hat{C}_2. \end{aligned} \tag{9.6}$$

Here, $\hat{1}$ means the identity operator, $\hat{1}\,|\Psi\rangle = |\Psi\rangle$. The operator $\hat{C}$ is called a *cluster operator*. The operator $\hat{C}_1$ contains the singles part, and $\hat{C}_2$ the doubles part. Higher excitations have similar notation. The history of the name "cluster operator" is muddied by the waters of ancient history, but we will later see how it may have come to pass.

Notice that the product of two cluster operators is again a cluster operator. [Give examples.] Mathematically, cluster operators form an algebra. Since cluster operators commute, we can compute with cluster operators much like with numbers! However, cluster operators are *nilpotent*, since there are only a finite number of electrons we can excite!

Intermediate normalization is common in perturbation theory

We now introduce a concept called *intermediate normalization*: An eigenvector of the Hamiltonian is only determined up to its length (and complex phase). At the same time, if Hartre–Fock is supposed to be an initially good approximation, it makes sense to fix $c_0 = 1$ in the CI wavefunction. This also means that $\langle \Psi^{\text{CISD}} | \Psi^{\text{CISD}} \rangle \geq 1$ in general, and > 1 whenever at least one CI coefficient is nonzero, so we need to divide by this quantity to evaluate the energy (since the norm is not 1). But the length of a Hilbert space vector has no physical meaning, so this does not change physics at all. Intermediate normalizations does require, however, that the wavefunction is never exactly orthogonal to $|\Phi_0\rangle$.

We now have:

$$|\Psi^{\text{CISD}}\rangle = (\hat{1} + \hat{C})\,|\Phi_0\rangle = (\hat{1} + \hat{C}_1 + \hat{C}_2)\,|\Phi_0\rangle :$$

If we add triples and so on, that would simply mean that we extend $\hat{C}$ with the corresponding terms $\hat{C}_3$, etc., defining the CISDT wavefunction, and so on.

The wavefunction is now written as an operator equation: The cluster operator *acts on* the Hartree–Fock wavefunction, building up

the complete wavefunction by means of excitations. This is a very useful concept.

We now formulate CI theory as an operator equation: The Schrödinger equation reads: Find a cluster operator $\hat{C}$ (with the appropriate truncation, say CISD) such that the following equation is satisfied:

$$\hat{H}(\hat{1}+\hat{C})\,|\Phi_0\rangle = E(\hat{1}+\hat{C})\,|\Phi_0\rangle\,.$$

Projecting onto $\langle\Phi_0|$, $\langle\Phi_i^a|$, and $\langle\Phi_{ij}^{ab}|$ gives us the *CI energy equation*, and the CISD *amplitude equations*:

$$\langle\Phi_0|\hat{H}(\hat{1}+\hat{C})|\Phi_0\rangle = E \tag{9.7}$$
$$\langle\Phi_i^a|\hat{H}(\hat{1}+\hat{C})|\Phi_0\rangle = Ec_i^a \tag{9.8}$$
$$\langle\Phi_{ij}^{ab}|\hat{H}(\hat{1}+\hat{C})|\Phi_0\rangle = Ec_{ij}^{ab}. \tag{9.9}$$

This is nothing but the CISD matrix eigenvalue problem but where the wavefunction is expressed using the cluster operator $\hat{C}$.

9.3 *Size-extensivity revisited*

We now formulate the non-size-extensivity of CI theory using cluster operators. This will be most enlightening, and we will see the power of the cluster operators!

Suppose we have two molecular fragments A and B, with N_A and N_B electrons, respectively. These could be, say, hydrogen or water molecules. The Hamiltonian operator for such a system is

$$\hat{H}_{A+B} = \hat{H}_A + \hat{H}_B + \hat{W}_{AB},$$

where $\hat{W}_{AB}$ is the part of the $N_A + N_B$-electron Hamiltonian describing interactions between electrons of A and B. We now assume that the fragments are very far from each other. Then, they do not interact, and $\hat{W}_{AB} = 0$, i.e., $\hat{H}_{A+B} = \hat{H}_A + \hat{H}_B$.

The total Hamiltonian is then solvable by means of *separation of variables*:

$$\hat{H}_A\,|\Psi_A\rangle = E_A\,|\Psi_A\rangle \tag{9.10}$$
$$\hat{H}_B\,|\Psi_B\rangle = E_B\,|\Psi_B\rangle \tag{9.11}$$
$$E_{A+B} = E_A + E_B, \quad |\Psi_{A+B}\rangle = |\Psi_A\Psi_B\rangle\,. \tag{9.12}$$

A few words on exactly what the product of these wavefunctions mean: On the right-hand side of the last equation, we have written a product of two wavefunctions. This must be understood in the sense that the Slater determinants that enter the product contain spin-orbitals from each fragment A and B, with coefficients that are multiplied together. This ensures proper antisymmetry: If $|\Phi_I^A\rangle = |\chi_{p_1}^A \cdots \chi_{p_{N_A}}^B\rangle$ and $|\Phi_J^B\rangle = |\chi_{q_1}^B \cdots \chi_{q_{N_B}}^B\rangle$ are Slater determinants built from MOs at fragment A and B (where $I = (p_1, p_2, \cdots, p_{N_A})$ and $J = (q_1, \cdots, q_{N_B})$ are multiindices), then $IJ = (p_1, \cdots, p_{N_A}, q_1, \cdots, q_{N_B})$ is a multiindex for $A + B$ and

$$|\Phi_{IJ}^{A+B}\rangle = |\chi_{p_1}^A \cdots \chi_{p_{N_A}}^A \chi_{q_1}^B \cdots \chi_{q_{N_B}}^B\rangle$$

is a slater determinant for the complete molecule $A + B$. The fragment wavefunctions are

$$|\Psi_A\rangle = \sum_I c_I^A |\Phi_I^A\rangle, \quad |\Psi_B\rangle = \sum_J c_J^B |\Phi_J^B\rangle,$$

and finally the product wavefunction

$$|\Psi_A \Psi_B\rangle = \sum_{IJ} c_I c_J |\Phi_{IJ}^{A+B}\rangle$$

We observe two fundamental concepts, that embody the concept of size-extensivity

> For non-interactig molecular fragments, the energy is additively separable, and the wavefunction is multiplicatively separable.

If a many-body calculation method does not respect this, the results will have the wrong scaling properties with respect to the size of the system. Quantum chemical methods that are size-extensive *behave much better than non-size-extensive methods*, especially on larger molecules. Moreover, potential energy surfaces will be much better for properly size-extensive methods.

Let us consider the *exact* FCI wavefunctions for these fragments in terms of cluster operators:

$$|\Psi_A\rangle = (\hat{1} + \hat{C}_A) |\Phi_0^A\rangle. \tag{9.13}$$
$$|\Psi_B\rangle = (\hat{1} + \hat{C}_B) |\Phi_0^B\rangle. \tag{9.14}$$

Here, $|\Phi_{0,A}\rangle$ is the Hartree–Fock determinant for A, and similarly for B:

$$|\Phi_0^A\rangle = |\chi_1^A \cdots \chi_{N_A}^A\rangle, \quad |\Phi_0^B\rangle = |\chi_1^B \cdots \chi_{N_B}^B\rangle.$$

Note that since the fragments are very far from each other, the molecular orbitals in the two HF determinants are orthogonal to each other. The combined system therefore has HF determinant

$$\begin{aligned} |\Phi_0^{A+B}\rangle &= |\Phi_0^A \Phi_0^B\rangle = |\chi_1^A \cdots \chi_{N_A}^A \chi_1^B \cdots \chi_{N_B}^B\rangle \\ &= |\chi_1^{A+B} \cdots \chi_{N_A+N_B}^{A+B}\rangle. \end{aligned} \tag{9.15}$$

The last equation indicates the HF spin-orbitals obtained from a HF calculation on both molecules simulaneously. These will match the orbitals from the separately calculated MOs.

(We assume that a HF formulation has been used so this is true. For eample, if each fragment was a H atom, we know that RHF will not produce this form, while UHF will. If each fragment has an even number of electrons, RHF will work out fine.)

A FCI calculation on the combined molecular system AB will now yield the exact result, and the wavefunction

$$|\Psi_{A+B}\rangle = (\hat{1} + \hat{C}_{A+B}) |\Phi_0^{A+B}\rangle.$$

The cluster operator $\hat{C}_{A+B}$ now contains in principle *all possible excitations from every possible occupied i to every possible unoccupied a.* Since the fragments do not interact, excitations that move electrons from A to B or vice versa will have zero amplitudes.

On the other hand, multiplicative separability gives

$$\begin{aligned} |\Psi_{A+B}\rangle = |\Psi_A \Psi_B\rangle &= (\hat{1} + \hat{C}_A)(\hat{1} + \hat{C}_B)\,|\Phi_0^{A+B}\rangle \\ &= (\hat{1} + \hat{C}_A + \hat{C}_B + \hat{C}_A\hat{C}_B)\,|\Phi_0^{A+B}\rangle \end{aligned} \tag{9.16}$$

Notice here, that the product of two cluster operators is again a cluster operator.

We now consider the case of a truncated CI method, such as CISD. We compare the application to each molecular fragment A and B, and the application to $A + B$. Are the results size-extensive, i.e., is the energy additively separable and the wavefunction multiplicatively separable? The answer is *no*, and with the aid of cluster operators, not so hard to see!

The CISD calculation on the fragment gives

$$|\Psi_A^{\mathrm{CISD}}\rangle = (\hat{1} + \hat{C}_A^{\mathrm{CISD}})\,|\Phi_0^A\rangle\,,$$

$$|\Psi_B^{\mathrm{CISD}}\rangle = (\hat{1} + \hat{C}_B^{\mathrm{CISD}})\,|\Phi_0^B\rangle\,.$$

The product of these wavefunctions is

$$|\Psi_A^{\mathrm{CISD}}\Psi_B^{\mathrm{CISD}}\rangle = (\hat{1} + \hat{C}_A^{\mathrm{CISD}} + \hat{C}_B^{\mathrm{CISD}} + \hat{C}_A^{\mathrm{CISD}}\hat{C}_B^{\mathrm{CISD}})\,|\Phi_0^{A+B}\rangle\,.$$

We observe that the final term contains quadruples excitations. Compare now with a CISD calculation on the system $A + B$. The wavefunction can only contain up to double excitations:

$$|\Psi_{A+B}^{\mathrm{CISD}}\rangle = (\hat{1} + \hat{C}_{A+B}^{\mathrm{CISD}})\,|\Phi_0^{A+B}\rangle\,.$$

Here, $\hat{C}_{A,B}^{\mathrm{CISD}}$ describes simultaneous excitations on A and B, as well as excitations on A and B only. But in truncated CI, the level of these excitations would be bounded by the level of the CI theory: for CISD, only singles on A and B simultaneously would be allowed. On the other hand, the product wavefunction has much higher excitations present – CISDTQ for the CISD case!

Thus, truncated CI is not multiplicatively separable in the wavefunction in general, and the energy is then neither additively separable.

The only exception is FCI, where all excitations are taken into account, irrespective of the number of electrons present.

9.4 *Exponential ansatz in CC theory*

In CC theory, the non-size extensivity is corrected by means of the *exponential ansatz*:

$$|\Psi^{\mathrm{CC}}\rangle = e^{\hat{T}}\,|\Phi_0\rangle\,. \tag{9.17}$$

Here, $\hat{T}$ is a cluster operator:

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \dots, \tag{9.18}$$

$$\hat{T}_1 = \sum_{ia} t_i^a X_i^a, \tag{9.19}$$

$$\hat{T}_2 = \sum_{ijab} t_{ij}^{ab} X_{ij}^{ab}, \tag{9.20}$$

$$\vdots \tag{9.21}$$

The $\hat{T}_1$ operator is called the singles operator, $\hat{T}_2$ operator the doubles operator, and so on. The numbers t_i^a and t_{ij}^{ab} are called singles and doubles cluster amplitudes.

The operator has now a new symbol, because it is numerically *not the same* as the CI cluster operator $\hat{C}$.

There are now two important questions: *What* is the exponential of this operator, and *why* do we choose this exponential?

The exponential of an operator is defined in terms of the Taylor series of the exponential function:

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \dots \tag{9.22}$$

So the exponential of $\hat{T}$ is

$$e^{\hat{T}} = \hat{1} + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \dots$$

As for *why*: Does the exponential fix multiplicative separability? Yes! Let is consider a truncated CC ansatz, say CCSD, on fragments A and B, meaning that we let $\hat{T}_A = \hat{T}_{A,1} + \hat{T}_{A,2}$ similarly for $\hat{T}_B$:

$$|\Psi_A^{\mathrm{CC}}\rangle = e^{\hat{T}_A} |\Phi_0^A\rangle \tag{9.23}$$

$$|\Psi_B^{\mathrm{CC}}\rangle = e^{\hat{T}_B} |\Phi_0^B\rangle . \tag{9.24}$$

When we multiply these two wavefunctions together, something magical happens:

$$|\Psi_A^{\mathrm{CC}}\Psi_B^{\mathrm{CC}}\rangle = e^{\hat{T}_A} e^{\hat{T}_B} |\Phi_0^{A+B}\rangle = e^{\hat{T}_A + \hat{T}_B} |\Phi_0^{A+B}\rangle \tag{9.25}$$

In the last step we used the rule

$$e^x e^y = e^{x+y},$$

This can be taken as a defining feature of the exponential function

which is valid for numbers x and y, *and also for commuting operators*. Note now, that $\hat{T}_A + \hat{T}_B$ has the *same total truncation level* as the individual $\hat{T}_A$ an $\hat{T}_B$. If we start with singles and doubles, the sum also has singles and doubles. Thus, unlike CI, multiplying the wavefunctions of the fragments together does *not* increase the need for excitations in the total molecule, when using the exponential ansatz!

What we have not done, is to specify *how* the CC cluster operators are determined for the fragments, and show that the CC cluster operator for $A + B$ will actually turn out to be the sum. For that, we need equations that define the CC method. This is the next step.

10 Lecture 10: Coupled-cluster theory, continued

10.1 The CC equations

We now consider a single molecule, and no longer the splitting into fragments. We first consider the variational principle. The expectation value of the energy is

$$E = \frac{\langle \Phi_0 | e^{\hat{T}^\dagger} \hat{H} e^{\hat{T}} \Phi_0 \rangle}{\langle \Phi_0 | e^{\hat{T}^\dagger} e^{\hat{T}} | \Phi_0 \rangle}. \tag{10.1}$$

Looking hard at this equation shows that evaluating the energy is very costly. The reason is that $e^{\hat{T}} |\Phi_0\rangle$ is a full CI wavefunction, and there seems to be no formal simplifications in sight. Thus, computing the energy is as expensive as in full CI. This will not be tractable!

However, there turns out to be a nice trick to make CC a viable option. Instead of doing something with the variational principle, we look directly at the Schrödinger equation:

$$\hat{H} e^{\hat{T}} |\Phi_0\rangle = E e^{\hat{T}} |\Phi_0\rangle . \tag{10.2}$$

We observe that

$$e^{-\hat{T}} e^{\hat{T}} = e^{-\hat{T}+\hat{T}} = e^0 = \hat{1}.$$

Thus, multiplying from the left with $e^{-\hat{T}}$ we obtain the *similarity transformed SE*:

$$e^{-\hat{T}} \hat{H} e^{\hat{T}} |\Phi_0\rangle = E |\Phi_0\rangle . \tag{10.3}$$

This equation is somewhat curious: On the right hand side, we have an eigenvalue E multiplied by the HF determinant. On the left hand side, we have an operator

$$\bar{H} \equiv e^{-\hat{T}} \hat{H} e^{\hat{T}}$$

acting on the HF state. Thus, Eq. (10.3) can be phrased as a search for $\hat{T}$ so that $\bar{H}$ – an explicit function of $\hat{T}$ – has the HF state as its eigenvector! To be clear, this is a *nonlinear problem*. In quantum chemistry, we abhor the cost of FCI so much that we are willing to introduce nonlinearity upon nonlinearity to produce schemes that in the end are cheaper than FCI ...

$\bar{H}$ is often called *the correlated Hamiltonian*, since correlations are moved away from the wavefunction and put into the Hamiltonian instead.

How do we turn the similarity transformed SE into something ameble for implementation on a computer? We will shortly discuss this, but first we do some projections:

Since the Slater determinants $|\Phi_0\rangle$, $|\Phi_i^a\rangle$, etc., form a basis for FCI space, we can project and obtain the energy equation,

We have that $\hat{T}^\dagger |\Phi_0\rangle = 0$ which makes $\langle\Phi_0| e^{-\hat{T}} = \langle\Phi_0|$

$$\langle\Phi_0|\hat{H}e^{\hat{T}}|\Phi_0\rangle = E, \tag{10.4}$$

and the amplitude equations:

$$\langle\Phi_\mu|e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Phi_0\rangle = 0, \tag{10.5}$$

where μ is a general index that counts excitations:

$$\mu = \begin{pmatrix} a \\ i \end{pmatrix}, \quad \mu = \begin{pmatrix} ab \\ ij \end{pmatrix}, \quad \ldots \tag{10.6}$$

When all the amplitude equations are satisfied, the similarity transformed SE (10.3) is satisfied. We then have the *exact solution* of the CI problem.

The energy equation we also obtained in CI theory, but with the cluster operator $\hat{1} + \hat{C}$ instead of $e^{\hat{T}}$.

10.2 Truncations in the CC wavefunction

The full set of projected CC equations (10.5) is no simpler than full CI, since there are exactly as many amplitudes in $\hat{C}$ as in $\hat{T}$. We therefore introduce *truncated CC*, where we truncate $\hat{T}$ after a fixed number of terms,

$$\hat{T} \approx \hat{T}_1 + \hat{T}_2 + \ldots + \hat{T}_K, \tag{10.7}$$

giving the CCSD$\cdots$K wavefunction. We need exactly one equation per unknown. Consider for example CCSD: Since $\hat{T}_1 = \sum_{ia} t_i^a X_i^a$, projection onto all the $\langle\Phi_\mu| = \langle\Phi_i^a|$ gives equationa for each unknown t_i^a. Similarly, $\hat{T}_2 = \sum_{ijab} t_{ij}^{ab} X_{ij}^{ab}$, and we obtain one equation per unknown if we project onto $\langle\Phi_\mu| = \langle\Phi_{ij}^{ab}|$. This principle gives the following:

> For the CCSD$\cdots$K method, we project the similarity transformed SE onto all excited determinants up to K-fold excitations

10.3 The BCH expansion

Let us look at the "correlated Hamiltonian" $\bar{H}$.. A similarity transformation such as $\bar{H}$ can be expanded using the Baker–Campbell–Hausdorff expansion. This reads

$$\begin{aligned} e^{-\hat{T}}\hat{H}e^{\hat{T}} = \hat{H} + [\hat{H},\hat{T}] + \frac{1}{2!}[[\hat{H},\hat{T}],\hat{T}] \\ + \frac{1}{3!}[[[\hat{H},\hat{T}],\hat{T}],\hat{T}] + \frac{1}{4!}[[[[\hat{H},\hat{T}],\hat{T}],\hat{T}],\hat{T}] + \ldots \end{aligned} \tag{10.8}$$

The BCH expansion is a general expansion, true for all similarity transformations, and not dependent on $\hat{T}$ being a cluster operator.

However, with regards to CC theory, it can now be shown that:

For a Hamiltonian

$$\hat{H} = \sum_i h(i) + \sum_{i<j} w(i,j),$$

containing only up to two-body terms, the BCH expansion *terminates identically after presented 5 terms*. This holds *regardless of the number of electrons* and *regardless of the truncation of* $\hat{T}$.

Proving this almost magical fact is beyond the scope of this course, but can be related to the fact that $\hat{T}$ substitutes orbitals, e.g., X_i^a changes χ_i to χ_a in any Slater determinant, and that $\hat{H}$ also substitutes orbitals, but now any orbital can be substituted by any orbital. For example, the two-electron interaction works at *two spin-orbitals at a time* due to the Slater condon rules. In fact it is possible to write

$$\hat{H} = \sum_{pq} h_{pq} X_q^p + \frac{1}{2} \sum_{pqrs} \langle pq|\hat{w}|rs\rangle \, X_{sr}^{pq},$$

where now X_q^p substitutes χ_q with χ_r, and X_{sr}^{pq} substitutes χ_r by χ_p and χ_s by χ_q. This will be especially transparent in the language of *second quantization* which we so far have avoided.

To prove the result on the termination of the BCH expansion, one can now explicitly compute

$$\exp(-t_\mu X_\mu) X_{sr}^{pq} \exp(t_\mu X_\mu) \tag{10.9}$$

using the BCH formula, and note that it stops after 5 terms. One then notes that

$$\hat{T} = \sum_\mu t_\mu X_\mu; \quad \exp(\hat{T}) = \prod_\mu \exp(t_\mu X_\mu)$$

since all the substitution operators X_μ commute among themselves.

Even if the present outline of a proof is not complete or transparent, one should take note of the fact that the expansion (10.9) does *not* depend on the number of electrons, and *not* on the truncation level of $\hat{T}$, since it is valid for any μ. The universality, so to speak, of the truncation of the BCH expansion for $\bar{H}$ is the important take-home message here.

10.4 *Closed algebraic forms of the CC equations*

In order to implement CC theory on a computer, closed-form algebraic expressions of the amplitude equations are needed. For CCSD, one needs closed-form expressions for the amplitude equations

Recall $\bar{H} = e^{-\hat{T}} \hat{H} e^{\hat{T}}$

$$f_i^a(t) \equiv \langle \Phi_i^a | \bar{H} | \Phi_0 \rangle = 0, \tag{10.10}$$

$$f_{ij}^{ab}(t) \equiv \langle \Phi_{ij}^{ab} | \bar{H} | \Phi_0 \rangle = 0, \tag{10.11}$$

as well as the energy equation,

$$E_{\mathrm{CC}}(t) \equiv \langle \Phi_0 | \hat{H} e^{\hat{T}} | \Phi_0 \rangle \,. \tag{10.12}$$

In these equations $t = (t_i^a, t_{ij}^{ab})$ are the CCSD amplitudes.

Our Hamiltonian is on the form

$$\hat{H} = \hat{H}_0 + \hat{W} = \sum_i \hat{h}(i) + \sum_{i<j} \hat{w}(i,j). \tag{10.13}$$

Here, $\hat{h} = \hat{t} + \hat{v}$ ($\hat{H}_0 = \hat{T} + \hat{V}$) is the total of kinetic energy and the nuclear Coulomb potential. Recall that in Hartree–Fock theory, the Fock operator is given by

$$\hat{f} = \hat{h} + \hat{v}^{\mathrm{HF}},$$

where the HF potential is the sum of the Hartree and exchange potentials. We can add and subtract this potential to $\hat{H}$ and obtain

$$\hat{H} = \sum_i \hat{f}(i) + \left(-\sum_i \hat{v}^{\mathrm{HF}}(i) + \sum_{i<j} \hat{w}(i,j) \right) \equiv \hat{F} + \hat{W}_{\mathrm{fluc}}, \tag{10.14}$$

where $\hat{F}$ is the N-electron Fock operator, and where $\hat{W}_{\mathrm{fluc}} \equiv \hat{W} - \hat{V}^{\mathrm{HF}}$ is called the *fluctuation potential*. Since HF is supposed to be a good approximation, and since $\hat{v}^{\mathrm{HF}}$ is the mean potential seen by any electron due to all the others, the fluctuation potential should be *small*.

The fluctuation potential is expressed in the antisymmetric matrix elements $\langle pq \| rs \rangle \equiv \langle pq | \hat{w} | rs \rangle_{\mathrm{AS}}$. The fluctuation potential and and the Fock operator are the ingredients needed for the CC amplitude and energy equations.

The following is a paraphrasation of the equations listed in T. Daniel Crawford's highly-regarded introduction to CC theory in quantum chemistry. Nowadays, derivation and implementation of CC amplitude equations are usually done *automatically*, with specially designed codes. This minimizes opportunities for errors, and also automate a very time-consuming process.

We begin with the CCSD energy:

$$E_{\mathrm{CCSD}} = E_{\mathrm{HF}} + \sum_{ia} f_{ia}\, t_i^a + \tfrac{1}{4} \sum_{aibj} \langle ij \| ab \rangle\, t_{ij}^{ab} + \tfrac{1}{2} \sum_{aibj} \langle ij \| ab \rangle\, t_i^a\, t_j^b \tag{10.15}$$

Observe how the leading term is the Hartree–Fock energy, so that the remaining terms form the CCSD *correlation energy*. Note that the Fock operator matrix elements are usually $f_{pq} = \epsilon_p \delta_{pq}$, so that, e.g., $f_{ai} = 0$. However, the present algebraic expressions are *also valid* when the basis spin-orbitals are not Hartree–Fock solutions.

The singles amplitude equations are:

$$\begin{aligned} f_i^a(t) = {} & f_{ai} + \sum_c f_{ac}\, t_i^c - \sum_k f_{ki}\, t_k^a + \sum_{kc} \langle ka \| ci \rangle\, t_k^c + \sum_{kc} f_{kc}\, t_{ik}^{ac} + \tfrac{1}{2} \sum_{kcd} \langle ka \| cd \rangle\, t_{ki}^{cd} \\ & - \tfrac{1}{2} \sum_{klc} \langle kl \| ci \rangle\, t_{kl}^{ca} - \sum_{kc} f_{kc}\, t_i^c\, t_k^a - \sum_{klc} \langle kl \| ci \rangle\, t_k^c\, t_l^a - \sum_{kcd} \langle ka \| cd \rangle\, t_k^c\, t_i^d \\ & - \sum_{klcd} \langle kl \| cd \rangle\, t_k^c\, t_i^d\, t_l^a + \sum_{klcd} \langle kl \| cd \rangle\, t_k^c\, t_{li}^{da} - \tfrac{1}{2} \sum_{klcd} \langle kl \| cd \rangle\, t_{ki}^{cd}\, t_l^a - \tfrac{1}{2} \sum_{klcd} \langle kl \| cd \rangle\, t_{kl}^{ca}\, t_i^d \end{aligned} \tag{10.16}$$

The doubles amplitude equations are:

$$
\begin{aligned}
f_{ij}^{ab}(t) = \langle ab\|ij\rangle &+ \sum_c \left(f_{bc}\, t_{ij}^{ac} - f_{ac}\, t_{ij}^{bc} \right) - \sum_k \left(f_{kj}\, t_{ik}^{ab} - f_{ki}\, t_{jk}^{ab} \right) + \tfrac{1}{2} \sum_{kl} \langle kl\|ij\rangle\, t_{kl}^{ab} \\
&+ \tfrac{1}{2} \sum_{cd} \langle ab\|cd\rangle\, t_{ij}^{cd} + P(ij)P(ab) \sum_{kc} \langle kb\|cj\rangle\, t_{ik}^{ac} + P(ij) \sum_c \langle ab\|cj\rangle\, t_i^c \\
&- P(ab) \sum_k \langle kb\|ij\rangle\, t_k^a + \tfrac{1}{2} P(ij)P(ab) \sum_{klcd} \langle kl\|cd\rangle\, t_{ik}^{ac}\, t_{jl}^{db} + \tfrac{1}{4} \sum_{klcd} \langle kl\|cd\rangle\, t_{ij}^{cd}\, t_{kl}^{ab} \\
&- P(ab) \tfrac{1}{2} \sum_{klcd} \langle kl\|cd\rangle\, t_{ij}^{ac}\, t_{kl}^{bd} - P(ij) \tfrac{1}{2} \sum_{klcd} \langle kl\|cd\rangle\, t_{ik}^{ab}\, t_{jl}^{cd} + P(ab) \tfrac{1}{2} \sum_{kl} \langle kl\|ij\rangle\, t_k^a\, t_l^b \\
&+ P(ij) \tfrac{1}{2} \sum_{cd} \langle ab\|cd\rangle\, t_i^c\, t_j^d \\
&- P(ij)P(ab) \sum_{kc} \langle kb\|ic\rangle\, t_k^a\, t_j^c + P(ab) \sum_{kc} f_{kc}\, t_k^a\, t_{ij}^{bc} \\
&+ P(ij) \sum_{kc} f_{kc}\, t_i^c\, t_{jk}^{ab} - P(ij) \sum_{klc} \langle kl\|ci\rangle\, t_k^c\, t_{lj}^{ab} + P(ab) \sum_{klcd} \langle ka\|cd\rangle\, t_{ij}^{db}\, t_k^c \\
&+ P(ij)P(ab) \sum_{kcd} \langle ak\|dc\rangle\, t_i^d\, t_{jk}^{bc} + P(ij)P(ab) \sum_{klc} \langle kl\|ic\rangle\, t_l^a\, t_{jk}^{bc} \\
&+ P(ij) \tfrac{1}{2} \sum_{klc} \langle kl\|cj\rangle\, t_i^c\, t_{kl}^{ab} - P(ab) \tfrac{1}{2} \sum_{kcd} \langle kb\|cd\rangle\, t_k^a\, t_{ij}^{cd} - P(ij)P(ab) \tfrac{1}{2} \sum_{kcd} \langle kb\|cd\rangle\, t_i^c\, t_k^a\, t_j^d \\
&+ P(ij)P(ab) \tfrac{1}{2} \sum_{klc} \langle kl\|cj\rangle\, t_i^c\, t_k^a\, t_l^b - P(ij) \sum_{klcd} \langle kl\|cd\rangle\, t_k^c\, t_i^d\, t_{lj}^{ab} - P(ab) \sum_{klcd} \langle kl\|cd\rangle\, t_k^c\, t_l^a\, t_{ij}^{db} \\
&+ P(ab) \tfrac{1}{4} \sum_{klcd} \langle kl\|cd\rangle\, t_k^a\, t_l^b\, t_{ij}^{cd} + P(ij)P(ab) \sum_{klcd} \langle kl\|cd\rangle\, t_i^c\, t_l^b\, t_{kj}^{ad} \\
&+ P(ij)P(ab) \tfrac{1}{4} \sum_{klcd} \langle kl\|cd\rangle\, t_i^c\, t_k^a\, t_j^d\, t_l^b
\end{aligned}
\tag{10.17}
$$

On this equation, the symbols $P(ij)$ and $P(ab)$ are permutation symbols:

$$P(ij)f(i,j) = f(i,j) - f(j,i),$$

$$P(a,b)f(a,b) = f(a,b) - f(b,a).$$

10.5 *Quasi-Newton iterations of the amplitude equations*

Recall that the Hamiltonian is split as $\hat{H} = \hat{F} + \hat{W}_{\text{fluc}}$, where the fluctutation potential is assumed small, equivalently that the main component of $\hat{H}$ is the Fock operator. Looking at the functions $f_i^a(t)$ and $f_{ij}^{ab}(t)$, we see that the leading terms are grouped according to this "largeness"; first constant terms, then terms proportional to the Fock matrix, and then two-electron interaction terms.

Consider now a Newton–Rhapson approach for finding the root t_* of $f(t_*) = 0$. We start from a guess t_n, and write

$$f(t_*) = 0 = f(t_n + \delta t) \approx f(t_n) + J\delta t + O(\delta t^2), J = \nabla_t f(t_n).$$

Here, J is the Jacobian matrix. Rearranging,

$$t_{n+1} = t_n + \delta t = t_n - J^{-1} f(t_n).$$

This is the basic Newton iteration. However, exact evaluation of the Jacobian is very costly for such a complicated function as $f(t)$. On the other hand, we observe that the Fock matrix – the main part of

the Hamiltonian – only appears linearly in $f(t)$. It should therefore form the main part of the Jacobian! We define therefore a quasi-Newton approach where we approximate the Jacobian by the terms arising from the Fock matrix. The idea is to have a Jacobian which is diagonal, beacuse then the quasi-Newton iterations are very fast.

We now assume that the Fock matrix is diagonal, for simplicity. We set $D_i^a = f_{aa} - f_{ii}$, and $D_{ij}^{ab} = f_{bb} + f_{aa} - f_{ii} - f_{jj}$. We can then rewrite $f_i^a(t)$ as

$$f_i^a(t) = D_i^a t_i^a + g_i^a(t) = 0 \tag{10.18}$$

where $f_i^a(t)$ contains all terms in $f_i^a(t)$, except those that are linear in the Fock matrix and the singles amplitudes. Rearranging, we obtain the quasi-Newton iteration

$$(t_i^a)^{n+1} = -(D_i^a)^{-1} g_i^a(t^n). \tag{10.19}$$

This is the singles quasi-Newton iteration. The doubles iteration is similar: Starting from

$$f_{ij}^{ab}(t) = D_{ij}^{ab} t_{ij}^{ab} + g_{ij}^{ab}(t) = 0, \tag{10.20}$$

rearranging gives the iteration

$$(t_{ij}^{ab})^{n+1} = -(D_{ij}^{ab})^{-1} g_{ij}^{ab}(t^n). \tag{10.21}$$

This quasi-Newton scheme is standard in CC theory, but it may have slow convergence in general. DIIS acceleration fixes this in most cases.

10.6 *Factorization of amplitude equations*

The doubles equations contain a term

$$f_{ij}^{ab}(t) \leftarrow \tfrac{1}{4} \sum_{klcd} \langle kl \| cd \rangle \, t_{ij}^{cd} t_{kl}^{ab}$$

The left-arrow indicates that it is one out of many terms that goes into the amplitude equations.

The brute-force cost of this term is $O(N_o^2 N_v^6)$, since there are N_v^4 summation terms, but the summation has to be done for $N_o^2 N_v^2$ equations. This is quite expensive!

However, we note that if we define the *intermediate tensor*

$$X_{ij}^{kl} = \sum_{cd} \langle kl \| cd \rangle \, t_{ij}^{cd},$$

with cost $O(N_o^2 N_v^4)$, we may now write

$$f_{ij}^{ab}(t) \leftarrow \sum_{kl} X_{ij}^{kl} t_{kl}^{ab}$$

The cost of the original term has been brought down to $O(N_o^2 N_v^4) + O(N_o^4 N_v^2)$.

It turns out that *all* terms in the CCSD amplitude equations can be similarly factorized, bringing the cost of the CCSD amplitude equations down to $O(\mathcal{N}^6)$, where $\mathcal{N}$ measures system size; effectively lying somewhere between N_o and N_v, depending on system size and implementation details.

10.7 CC Lagrangian

The CC method is not variational. We have not devised a scheme for variationally minimizing the expectation value of the energy. We have instead opted for projecting the similarity transformed SE. This is equivalent to *constrained minimization of the CC energy*. For each amplitude t_μ, introduce an auxillary amplitude λ_μ, and define the function

$$L(t,\lambda) = E_{\text{CC}}(t) + \sum_\mu \lambda_\mu f_\mu(t). \tag{10.22}$$

Here, the summation runs over those amplitudes we have present in our approximation, say, singles and doubles.

We interpret λ_μ as Lagranbe multipliers for constrained optimization of $E_{\text{CC}}(t)$, the constraints being that the amplitude equation $f_\mu(t) = 0$ is satisfied. Thus,

$$\frac{\partial L}{\partial \lambda_\mu} = 0 \iff f_\mu(t) = 0.$$

Values for the Lagrange multipliers are obtained from

$$\frac{\partial L}{\partial t_\mu} = 0.$$

If we insert the definitions of the CC energy and of $f_\mu(t)$ we obtain

$$L(t,\lambda) = \langle \Phi_0 | (\hat{1} + \hat{\Lambda}) e^{-\hat{T}} \hat{H} e^{\hat{T}} | \Phi_0 \rangle, \quad \hat{\Lambda} = \sum_\mu \lambda_\mu X_\mu^\dagger. \tag{10.23}$$

The constrained optimization viewpoint is very useful if one wants to compute expectation values of other observables than the energy, say $\hat{A}$.

[Explain Hellman-Feynman.]

10.8 Exercises

10.8.1 Exercise 1: Substitution operators and simple algebra

In these exercises, we study the substitution operators. Let

$$|\Phi_0\rangle = |\chi_1\chi_2\cdots\chi_N\rangle$$

be the Hartree–Fock determinant, where χ_p are spin orbitals. Occupied spin orbitals in $|\Phi_0\rangle$ are indexed by $i, j, \ldots$, and virtual spin orbitals are indexed by $a, b, \ldots$.

The single-substitution operator X_i^a is defined by its action on a Slater determinant:

$$X_i^a\,|\chi_{p_1}\cdots\chi_{p_N}\rangle = \begin{cases} |\chi_{p_1}\cdots\chi_a\cdots\chi_{p_N}\rangle\,, & \text{if } \chi_i \text{ is present,} \\ 0, & \text{if } \chi_i \text{ is not present.} \end{cases}$$

Thus X_i^a replaces the occupied spin orbital χ_i by χ_a whenever this is possible.

We also use the notation

$$|\Phi_i^a\rangle = X_i^a\,|\Phi_0\rangle\,, \qquad |\Phi_{ij}^{ab}\rangle = X_{ij}^{ab}\,|\Phi_0\rangle\,,$$

where X_{ij}^{ab} denotes a double substitution.

Problem 1: Action of a single substitution operator

This problem is about interpreting the definition of X_i^a.

1. Explain in words what the operator X_i^a does when it acts on a Slater determinant.

2. Let $|\chi_{p_1}\cdots\chi_{p_N}\rangle$ be an arbitrary Slater determinant. Explain why

   $$X_i^a\,|\chi_{p_1}\cdots\chi_{p_N}\rangle = 0$$

   whenever χ_i is not present among the occupied spin orbitals.

3. Explainy why

   $$X_i^a\,|\chi_{p_1}\cdots\chi_{p_N}\rangle = 0$$

   whenever χ_a is present among the occupied spin orbitals.

4. Explain why the operator X_i^a is called both a substitution operator and an excitation operator.

Problem 2: Successive substitutions

We now examine what happens when two substitutions are applied.

1. Consider $X_j^b\,|\Phi_0\rangle$. Write down in words what determinant is obtained.

2. Next consider
$$X_i^a X_j^b |\Phi_0\rangle .$$
Explain what this expression means operationally.

3. Under what conditions on the indices i, j, a, b should this double substitution be nonzero?

4. Explain why the cases $i = j$ and/or $a = b$ must be excluded if the result is to represent a double substitution.

5. Assume $i \neq j$ and $a \neq b$. Compare $X_i^a X_j^b |\Phi_0\rangle$ and $X_j^b X_i^a |\Phi_0\rangle$. Do they describe the same orbital replacements? State clearly what is the same, even if sign conventions are not analyzed in detail.

6. Consider the action of X_i^a and X_j^b on an arbitrary determinant $|\chi_{p_1} \cdots \chi_{p_N}\rangle$. Show that
$$[X_i^a, X_j^b] = 0.$$

Problem 3: Products of substitution operators

This problem concerns products of substitution operators.

1. Explain why
$$X_i^a X_i^b |\Phi_0\rangle = 0.$$

2. Explain why
$$X_i^a X_j^a |\Phi_0\rangle = 0.$$

3. Explain why
$$X_i^a X_i^a |\Phi_0\rangle = 0.$$

4. Formulate a general rule for when a product of two substitution operators must vanish.

5. Recall that $X_{ij}^{ab} = X_i^a X_j^b$. How many unique such operators are there, given N_{occ} occupied and N_{virt} virtual orbitals in total?

6. Show that
$$X_{ij}^{ab} = X_{ji}^{ba}, \quad \text{and} \quad X_{ij}^{ab} = -X_{ji}^{ab}, \quad \text{and} \quad X_{ij}^{ab} = -X_{ij}^{ba}.$$

Problem 4: Singles and doubles operators in coupled-cluster theory

We now connect the substitution operators to the cluster operators.

1. Recall the definitions
$$\hat{T}_1 = \sum_{ia} t_i^a X_i^a, \qquad \hat{T}_2 = \sum_{ijab} t_{ij}^{ab} X_{ij}^{ab}.$$
Explain why $\hat{T}_1$ is called the singles operator and $\hat{T}_2$ is called the doubles operator.

2. Show that $\hat{T}_1 |\Phi_0\rangle$ is a linear combination of singly substituted determinants.

3. Show that $\hat{T}_2 |\Phi_0\rangle$ is a linear combination of doubly substituted determinants.
4. Explain the role of the amplitudes t_i^a and t_{ij}^{ab}.
5. Why do you think it is useful to collect many substitutions into the operators $\hat{T}_1$ and $\hat{T}_2$ instead of writing each substituted determinant separately?

10.8.2 *Exercise 2: The exponential ansatz at a formal level*

We now consider only the formal structure of the coupled-cluster ansatz,

$$|\Psi^{\mathrm{CC}}\rangle = e^{\hat{T}} |\Phi_0\rangle , \qquad \hat{T} = \hat{T}_1 + \hat{T}_2 + \cdots .$$

The purpose is not to derive amplitude equations, but only to understand what the exponential means algebraically.

Problem 1: Expanding the exponential

The expomential of an operator is used to define the exponential of an operator. The taylor series is:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!} = 1 + z + \ldots .$$

1. Show that

$$e^{\hat{T}} |\Phi_0\rangle = \left(\hat{1} + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \cdots \right) |\Phi_0\rangle .$$

2. Explain why, for CCSD, the term $(\hat{1} + \hat{T}) |\Phi_0\rangle$ generates a CISD wavefunction, and why it is irrelevant that there is no coefficient c_0 in front of $\hat{1}$.
3. For the term $\hat{T}^2 |\Phi_0\rangle$, for CCSD, show that we have both double, triple, and quadruply excited determinants in the expansion, but no singly excited determinant.
4. What is the highest level of excitation included in $|\Psi^{CC}\rangle$?

Problem 2: Keeping only singles

Assume for this problem that $\hat{T} = \hat{T}_1$.

1. Consider the CC wavefunction:

$$e^{\hat{T}_1} |\Phi_0\rangle = \left(\hat{1} + \hat{T}_1 + \frac{1}{2}\hat{T}_1^2 + \cdots \right) |\Phi_0\rangle .$$

   Insert $\hat{T}_1 = \sum_{ia} t_i^a X_i^a$, and write the expansion up to and including doubles using explicit sums over occupied and virtual indices.
2. Explain why $\hat{T}_1 |\Phi_0\rangle$ gives singly substituted determinants.
3. Explain why $\hat{T}_1^2 |\Phi_0\rangle$ gives doubly substituted determinants.
4. What does this tell you about the difference between the operator $\hat{T}_1$ and the wavefunction $e^{\hat{T}_1} |\Phi_0\rangle$?

Problem 3: Keeping singles and doubles

Assume now that $\hat{T} = \hat{T}_1 + \hat{T}_2$.

1. Write out
$$\hat{T}^2 = (\hat{T}_1 + \hat{T}_2)^2.$$
2. Identify the four unique terms that appear in this product.
3. Explain in words what kinds of substitutions may arise from each of the terms
$$\hat{T}_1^2, \qquad \hat{T}_1\hat{T}_2, \qquad \hat{T}_2\hat{T}_1, \qquad \hat{T}_2^2.$$
4. Why does the exponential ansatz naturally generate determinants with higher substitution level than the truncation level of the operator $\hat{T}$ itself?

Problem 4: Formal separability of the exponential ansatz

This problem revisits the formal observation made in the notes for two non-interacting fragments A and B.

1. Let
$$|\Psi_A^{\mathrm{CC}}\rangle = e^{\hat{T}_A}|\Phi_0^A\rangle, \qquad |\Psi_B^{\mathrm{CC}}\rangle = e^{\hat{T}_B}|\Phi_0^B\rangle.$$
Write down the product wavefunction $|\Psi_A^{\mathrm{CC}}\Psi_B^{\mathrm{CC}}\rangle$.
2. Assuming that $\hat{T}_A$ and $\hat{T}_B$ commute, it is a fact that
$$e^{\hat{T}_A}e^{\hat{T}_B} = e^{\hat{T}_A+\hat{T}_B}.$$
Show that this is *not* in general true if the operators did not commute. Hint: Insert Taylor expansions on both sides, and show that they differ.
3. Explain why $\hat{T}_A + \hat{T}_B$ has the same truncation level as the separate fragment operators.
4. Explain why this is an important difference between coupled-cluster theory and truncated CI.

10.8.3 *Exercise 3: Similarity-transformed coupled-cluster equations*

The coupled-cluster wavefunction is

$$\Psi_{\mathrm{CC}} = e^{\hat{T}}\Phi_0,$$

where Φ_0 is the Hartree–Fock determinant and $\hat{T}$ is the cluster operator. The similarity-transformed Hamiltonian is

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

The excited determinants are denoted Φ_μ, where μ may stand for $\binom{a}{i}$, $\binom{ab}{ij}$, and so on.

Problem 1: From the Schrödinger equation to the similarity-transformed equation

Start from the Schr"odinger equation

$$\hat{H}e^{\hat{T}}\Phi_0 = Ee^{\hat{T}}\Phi_0.$$

1. Multiply the equation from the left by $e^{-\hat{T}}$ and show that

$$\bar{H}\Phi_0 = E\Phi_0.$$

2. Explain why this is not an ordinary linear eigenvalue problem for the unknowns in $\hat{T}$.

3. Explain in words what is meant by the statement that the correlations are moved from the wavefunction into the Hamiltonian.

Problem 2: Projection equations and truncation principle

Assume that the Slater determinants $\Phi_0, \Phi_i^a, \Phi_{ij}^{ab}, \ldots$ form a basis.

1. Project the equation

$$\bar{H}\Phi_0 = E\Phi_0$$

from the left with $\langle\Phi_0|$ and show that

$$\langle\Phi_0|\hat{H}e^{\hat{T}}|\Phi_0\rangle = E.$$

You do not have to show why "$e^{-\hat{T}}$" vanishes from $\bar{H}$ in this case.

2. Project from the left with $\langle\Phi_\mu|$ and show that the amplitude equations are

$$\langle\Phi_\mu|e^{-\hat{T}}\hat{H}e^{\hat{T}}|\Phi_0\rangle = 0.$$

3. For CCSD, write down which determinants Φ_μ must be used in the projection in order to obtain one equation for each unknown amplitude.

4. Generalize this statement to CCSD$\cdots$K.

5. Explain why the projection is truncated at the same excitation level as the cluster operator.

6. Explain the statement "the amplitude equations are nonlinear equations in the cluster amplitudes" using your own words.

Problem 3: Formal properties of the BCH expansion

The Baker–Campbell–Hausdorff expansion gives

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

1. Write out explicitly the first five terms of the BCH expansion for $\bar{H}$.

2. Identify which term contains one commutator, which contains two nested commutators, and so on.

3. For a Hamiltonian with only one-body and two-body terms, the BCH expansion terminates identically after the fifth term. Explain why this means that the amplitude equations are multivariate polynomial equations. What is the degree of the polynomials?

4. The BCH termination does not depend on the number of electrons or on the truncation level of $\hat{T}$. Can you think of why this is considered to be an important structural property of coupled-cluster theory?

Problem 4: Permutation operators in the doubles equations

In the doubles amplitude equations, permutation symbols are used. They are defined by

$$P(ij)f(i,j) = f(i,j) - f(j,i),$$

$$P(ab)f(a,b) = f(a,b) - f(b,a).$$

1. Expand explicitly
$$P(ij)A_{ij}^{ab}.$$

2. Expand explicitly
$$P(ab)A_{ij}^{ab}.$$

3. Expand explicitly
$$P(ij)P(ab)A_{ij}^{ab}.$$

4. Show that the result from the previous item contains four terms.

5. Explain why such permutation operators are useful when writing amplitude equations.

Problem 5: Reading structure from the CCSD energy expression

The CCSD energy expression in the notes is

$$E_{\mathrm{CCSD}} = E_{\mathrm{HF}} + \sum_{ia} f_{ia}\, t_i^a + \tfrac{1}{4}\sum_{aibj}\langle ij\|ab\rangle\, t_{ij}^{ab} + \tfrac{1}{2}\sum_{aibj}\langle ij\|ab\rangle\, t_i^a t_j^b.$$

1. Identify the Hartree–Fock contribution.

2. Identify the terms that make up the correlation energy.

3. State which term is linear in the singles amplitudes t_i^a.

4. State which term is linear in the doubles amplitudes t_{ij}^{ab}.

5. State which term is quadratic in the amplitudes.

6. Assume canonical Hartree–Fock spin orbitals, so that $f_{ai} = 0$. Rewrite the energy expression under this assumption.

Problem 6: Quasi-Newton form of the amplitude equations

Assume that the Fock matrix is diagonal. Define

$$D_i^a = f_{aa} - f_{ii}, \qquad D_{ij}^{ab} = f_{aa} + f_{bb} - f_{ii} - f_{jj}.$$

The notes rewrite the amplitude equations in the form

$$f_i^a(t) = D_i^a t_i^a + g_i^a(t) = 0,$$

$$f_{ij}^{ab}(t) = D_{ij}^{ab} t_{ij}^{ab} + g_{ij}^{ab}(t) = 0.$$

1. Explain how the singles equations are rearranged to obtain the quasi-Newton iteration

$$(t_i^a)^{n+1} = -(D_i^a)^{-1} g_i^a(t^n).$$

2. Rearrange the doubles equation to obtain the corresponding update formula for t_{ij}^{ab}.

3. Explain why the diagonal form of the approximate Jacobian is computationally convenient.

4. Look to Eqs. (**??**) and (**??**). Underline the terms that go into $g_i^a(t)$ and $g_{ij}^{ab}(t)$. What is common to these terms, in terms of which symbols/arrays are required? Which symbols are *not* required?

5. The quasi-Newton iterations are motivated by considering some parts of the Hamiltonian as dominating and some as small perturbation terms. Explain, in words, and using the previous point, which part of the Hamiltonian is treated as the dominant part in this approximation.

Problem 7: Formal comparison of CI and CC

In configuration interaction one writes the wavefunction in the form

$$\Psi_{\mathrm{CI}} = (\hat{1} + \hat{C})\Phi_0,$$

whereas in coupled cluster one writes

$$\Psi_{\mathrm{CC}} = e^{\hat{T}}\Phi_0.$$

1. Write the energy equation obtained in CI theory.

2. Write the corresponding energy equation obtained in CC theory.

3. State clearly which operator replaces $\hat{1} + \hat{C}$ in coupled-cluster theory.

4. Explain one formal advantage of using the similarity-transformed Schrödinger equation rather than a direct variational evaluation of

$$\frac{\langle \Phi_0 | e^{\hat{T}^\dagger} \hat{H} e^{\hat{T}} | \Phi_0 \rangle}{\langle \Phi_0 | e^{\hat{T}^\dagger} e^{\hat{T}} | \Phi_0 \rangle}.$$

10.8.4 *Programming exercise: Potential energy curves, correlation, and size extensivity in LiH*

In this exercise, you will use PySCF to compute simple potential energy curves for LiH with restricted Hartree–Fock (RHF), configuration interaction with single and double substitutions (CISD), and coupled cluster with single and double substitutions (CCSD). The purpose is twofold. First, you will compare the three methods on a bond-stretching problem. Second, you will test numerically whether a method is size extensive by comparing a separated LiH dimer with two isolated LiH monomers.

Use a restricted closed-shell reference throughout. Let R denote the Li–H bond length in Å, and let

$$E_{\mathrm{RHF}}(R), \qquad E_{\mathrm{CISD}}(R), \qquad E_{\mathrm{CCSD}}(R)$$

denote the corresponding total electronic energies. In the size-extensivity test, define

$$\Delta_{\mathrm{method}}(R) = E^{(2)}_{\mathrm{method}}(R) - 2E^{(1)}_{\mathrm{method}}(R),$$

where $E^{(1)}_{\mathrm{method}}(R)$ is the energy of one LiH molecule and $E^{(2)}_{\mathrm{method}}(R)$ is the energy of two LiH molecules placed far apart. For a size-extensive method, $\Delta_{\mathrm{method}}(R)$ should be close to zero.

In all problems below, use a small basis such as `sto-3g` or `6-31g`. The smaller basis is sufficient for a first implementation.

If helpful, begin from the following minimal PySCF structure:

```
import numpy as np
from pyscf import gto, scf, ci, cc

# Suggestion: start with a small grid of bond lengths
R_values = np.linspace(1.2, 4.0, 8)

# Suggestion: store results in Python lists and
    convert to arrays later
E_rhf = []
E_cisd = []
E_ccsd = []
```

A useful way to construct a monomer is:

```
def make_lih(R, basis="sto-3g"):
    mol = gto.Mole()
    mol.atom = f"Li␣0.0␣0.0␣0.0;␣H␣0.0␣0.0␣{R}"
    mol.unit = "Angstrom"
    mol.basis = basis
    mol.spin = 0
    mol.charge = 0
    mol.verbose = 0
    mol.build()
    return mol
```

You will need to consult the PySCF documentation, but a useful first step is to verify which objects store the RHF, CISD, and CCSD total energies.

Problem 1: Compute a potential energy curve for LiH

Compute the RHF, CISD, and CCSD potential energy curves for LiH over a range of bond lengths.

1. Choose a bond-length grid $R_1, R_2, \ldots, R_n$ that covers both the equilibrium region and a moderately stretched bond.

2. For each bond length R, build the LiH molecule and carry out an RHF calculation.

3. Using the RHF solution as reference, perform CISD and CCSD calculations. Record the three total energies

$$E_{\mathrm{RHF}}(R), \qquad E_{\mathrm{CISD}}(R), \qquad E_{\mathrm{CCSD}}(R).$$

4. Plot the three potential energy curves in the same figure.

5. Compute and plot the correlation energies

$$E^{\mathrm{corr}}_{\mathrm{CISD}}(R) = E_{\mathrm{CISD}}(R) - E_{\mathrm{RHF}}(R), \qquad E^{\mathrm{corr}}_{\mathrm{CCSD}}(R) = E_{\mathrm{CCSD}}(R) - E_{\mathrm{RHF}}(R).$$

6. Discuss the following points:

   - Which method gives the lowest energy at a fixed bond length?
   - Is CCSD uniformly lower than CISD on the grid you have chosen?
   - Does the qualitative shape of the RHF curve differ from the correlated curves when the bond is stretched?

The following code skeleton may be useful, but it is intentionally incomplete:

```
for R in R_values:
    mol = make_lih(R, basis="sto-3g")

    # Step 1: RHF calculation
    mf = scf.RHF(mol)
    mf.kernel()

    # Step 2: CISD based on the RHF reference
    myci = ci.CISD(mf)
    # TODO: run the calculation and identify the
        total energy

    # Step 3: CCSD based on the RHF reference
    mycc = cc.CCSD(mf)
    # TODO: run the calculation and identify the
        total energy

    # Step 4: append total energies to E_rhf, E_cisd,
         E_ccsd
```

Problem 2: Test size extensivity with two separated LiH molecules

Now investigate whether RHF, CISD, and CCSD behave correctly when two noninteracting subsystems are combined.

Construct a dimer consisting of two LiH molecules separated by a large distance, for example 20 Å. One convenient geometry is

$$\mathrm{Li}_1 = (0,0,0), \qquad \mathrm{H}_1 = (0,0,R),$$

$$\mathrm{Li}_2 = (20,0,0), \qquad \mathrm{H}_2 = (20,0,R).$$

For sufficiently large separation, the two molecules should interact only very weakly.

1. Write a function that constructs this dimer geometry for a chosen bond length R.
2. For each R on the same grid as in Problem 1, compute the monomer and dimer energies at the RHF, CISD, and CCSD levels.
3. For each method, compute the deviation from exact additivity,

$$\Delta_{\mathrm{method}}(R) = E^{(2)}_{\mathrm{method}}(R) - 2E^{(1)}_{\mathrm{method}}(R).$$

4. Plot $\Delta_{\mathrm{RHF}}(R)$, $\Delta_{\mathrm{CISD}}(R)$, and $\Delta_{\mathrm{CCSD}}(R)$ in the same figure.
5. Interpret the result. Which method appears to be most nearly size extensive? Which method shows the clearest deviation from size extensivity?

A useful starting point is:

```
def make_lih_dimer(R, separation=20.0, basis="sto-3g"
    ):
    mol = gto.Mole()
    mol.atom = f"""
    Li 0.0           0.0 0.0
    H  0.0           0.0 {R}
    Li {separation} 0.0 0.0
    H  {separation} 0.0 {R}
    """
    mol.unit = "Angstrom"
    mol.basis = basis
    mol.spin = 0
    mol.charge = 0
    mol.verbose = 0
    mol.build()
    return mol
```

It is useful to define a helper function that runs one electronic-structure method chain and returns the relevant total energies:

```
def run_all_methods(mol):
    mf = scf.RHF(mol)
    # TODO: run RHF and store the RHF total energy
```

```
    myci = ci.CISD(mf)
    # TODO: run CISD and store the CISD total energy

    mycc = cc.CCSD(mf)
    # TODO: run CCSD and store the CCSD total energy

    # TODO: return the three energies in a convenient
        format
```

Problem 3: Compare CISD and CCSD critically

Use the numerical data from Problems 1 and 2 to assess the statement that CCSD is an improvement over CISD.

1. From the monomer potential energy curves, compare RHF, CISD, and CCSD. In particular, identify the general ordering of the total energies.

2. Explain why the size-extensivity test

$$E^{(2)}(R) - 2E^{(1)}(R)$$

   is more informative than a comparison of monomer energies alone.

3. Based on your plots, explain in what sense CCSD improves on CISD. Your answer should refer both to correlation energy and to size extensivity.

4. State clearly at least one caveat. For example, discuss whether RHF, CISD, and CCSD are all expected to remain reliable at very large bond lengths, where static correlation may become important.

5. Summarize your conclusions in a short paragraph.

11 Lecture 11: Density-functional theory

11.1 Introduction

In this chapter we introduce *density-functional theory* (DFT), which is by far the most widely-used *ab initio* method in quantum chemistry.

In DFT, the N-electron Schrödinger wavefunction $\Psi(1, 2, \ldots, N)$ is replaced by its spatial density $\rho(\mathbf{r})$. Thus, the Schrödinger equation is replaced by a corresponding problem for determining the density. Clearly, this is advantageous, since the density is a much simpler quantity to model than the full wavefunction.

That this is possible is not at all trivial, and its discovery by Hohenberg and Kohn is one of the truly monumental milestones of quantum chemistry. Indeed, Walter Kohn was awarded the Nobel prize in chemistry in 1999.

Of course, there is no free lunch: Even if the transition to a density-based picture is possible in principle, the resulting equations are very complicated. The "universal functional" of DFT is unknown, except for its analytical properties, which are truly ugly.

As I see it, one can identify two schools of DFT:

1. Hohenberg–Kohn school, founding DFT on the famous Hohenberg–Kohn theorems.

2. Levy–Lieb school, founding DFT on the constrained-search formalism.

The latter appraoch is more transparent, so we will largely stay within the Levy–Lieb approach to formulating the basics of DFT.

11.2 From the Schrödinger equation to the Levy–Lieb constrained search formalism

11.2.1 The Schrödinger equation

We recap the Schrödinger equation for N electrons in an "external potential" $v(\mathbf{r})$. This potential, together with N, describes the molecule under consideration. Recall that a molecule is specified by a nuclear arrangement $R = (\mathbf{R}_1, \cdots, \mathbf{R}_{N_{\text{nuc}}})$ and charges $Z = (Z_1, \cdots, Z_{N_{\text{nuc}}})$. Given this data, the molecular potential is

$$v(\mathbf{r}) = \sum_{A=1}^{N_{\text{nuc}}} \frac{-Z_A}{\|\mathbf{R}_A - \mathbf{r}\|}. \tag{11.1}$$

The Hamiltonian is

$$\hat{H}(v) = \hat{T} + \hat{W} + \hat{V}, \tag{11.2}$$

where

$$\hat{V} = \sum_{i=1}^{N} v(\mathbf{r}_i). \tag{11.3}$$

We introduce the v-dependence of the Hamiltonian. Changing v changes the molecule, for example by moving the nuclei around. The part $\hat{T} + \hat{W}$ does not depend on the molecule (apart from N), and are termed *universal*. We are interested in computing the *ground-state energy* of $\hat{H}(v)$:

$$E(v) \equiv \min_{\Psi,\, \|\Psi\|=1} \langle \Psi | \hat{H}(v) | \Psi \rangle = \langle \Psi_0 | \hat{H}(v) | \Psi_0 \rangle, \tag{11.4}$$

where Ψ_0 is the minimizer, the lowest-lying eigenfunction of $\hat{H}(v)$,

$$\hat{H}(v)\Psi_0 = E(v)\Psi_0. \tag{11.5}$$

11.2.2 *The density*

Central to DFT is the electronic density. It is a fact that for a potential $v(\mathbf{r})$, with N-electron operator $\hat{V}$ as above, and given an N-electron wavefunction $\Psi(1, \cdots, N)$, the expectation value is given by

$$\langle \Psi | \hat{V} | \Psi \rangle = \int_{\mathbb{R}^3} \rho(\mathbf{r}) v(\mathbf{r})\, d\mathbf{r}, \tag{11.6}$$

where $\rho(\mathbf{r})$ is the N-electron density coming from Ψ. This density is defined as follows:

Every wavefunction has a unique density ρ, but there could be many wavefunctions that has one particular ρ as density!

$$\rho(\mathbf{r}) = N \sum_{\omega} \int dx_2 \cdots dx_N |\Psi((\mathbf{r}, \omega), x_2, \cdots, x_N))|^2. \tag{11.7}$$

Note here that we are starting with the N-electron space-spin probability density $|\Psi|^2$ and integrate out everything *except* a single electron spatial coordinate. All spin coordinates are integrated (summed).

It is straightforward to see that:

- $\int \rho(\mathbf{r})\, d\mathbf{r} = N$,
- $\rho(\mathbf{r}) \geq 0$.

N-representability conditions of electron densities

Additionally, one can prove that if the wavefunction has finite kinetic energy, then

- $\int |\nabla \sqrt{\rho}|^2 < +\infty$.

It can be shown that conversely, any function ρ satisfying the above three criteria come from *some* electronic wavefunction with finite kinetic energy. The conditions are therefore called "N-representability conditions", and we say that ρ is N-representable if these conditions are met.

The interpretation of the density is that $\rho(\mathbf{r})d\mathbf{r}$ is the *number of electrons in a small box of volume* $d\mathbf{r}$ *at* $\mathbf{r}$.

We now prove Eq. (11.6).

$$\begin{aligned}
\langle\Psi|\hat{V}|\Psi\rangle &= \int dx_1 \cdots dx_N \sum_{i=1}^{N} v(\mathbf{r}_i)|\Psi(x_1,\cdots,x_N)|^2 \\
&= \sum_{i=1}^{N} \int dx_1 \cdots dx_N v(\mathbf{r}_i)|\Psi(x_1,\cdots,x_i,\cdots,x_N)|^2 \\
&= \sum_{i=1}^{N} \int dx_i \cdots dx_1 \cdots dx_N v(\mathbf{r}_1)|\Psi(x_i,\cdots,x_1,\cdots,x_N)|^2 \quad \text{(rename integration vars)} \\
&= \sum_{i=1}^{N} \int dx_1 \cdots dx_N v(\mathbf{r}_1)|\Psi(x_1,\cdots,x_i,\cdots,x_N)|^2 \quad \text{(perm. symm./reorder } \textstyle\int\text{)} \\
&= N \int dx_1 \cdots dx_N v(\mathbf{r}_1)|\Psi(x_1,\cdots,x_N)|^2 \\
&= \int d\mathbf{r} v(\mathbf{r}) \sum_{\omega} \int dx_2 \cdots dx_N |\Psi((\mathbf{r},\omega),x_2,\cdots,x_N)|^2 \quad \text{(shuffle and rename)} \\
&= \int d\mathbf{r} v(\mathbf{r})\rho(\mathbf{r}). \quad \text{Q.E.D.}
\end{aligned} \tag{11.8}$$

A special case of a density is that coming from a Slater determinant: Let

$$|\Phi\rangle = |\psi_1\overline{\psi}_1 \cdots \psi_{N/2}\overline{\psi_{N/2}}\rangle\,, \tag{11.9}$$

i.e., a Slater determinant as encountered in RHF. Then it is straightforward to compute

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2} |\psi_i(\mathbf{r})|^2. \tag{11.10}$$

11.2.3 *Constrained-search theory*

Let Ψ be a normalized N-electron wavefunction, and let ρ_Ψ denote the density of Ψ. We then have

$$\begin{aligned}
\langle\Psi|\hat{H}(v)|\Psi\rangle &= \langle\Psi|\hat{T}+\hat{W}|\Psi\rangle + \langle\Psi|\hat{V}|\Psi\rangle \\
&= \langle\Psi|\hat{T}+\hat{W}|\Psi\rangle + \int v(\mathbf{r})\rho(\mathbf{r})\,d\mathbf{r},
\end{aligned} \tag{11.11}$$

where we used the result of the previous subsection.

Observe that the molecule dependent term of the expectation value only depends on the density of Ψ, only a tiny part of the information of the wavefunction. We therefore split the minimization into two steps: In an outer minimization, we minimize over all possible densities ρ that come from some normalized wavefunction (i.e., they are N-representable), and in an inner minimization, we minimize over all the wavefunctions that actually have this density. The result is the *Levy–Lieb variational principle*

$$\begin{aligned}
E(v) &= \min_{\rho}\left\{\left(\min_{\Psi\mapsto\rho}\langle\Psi|\hat{T}+\hat{W}|\Psi\rangle\right) + \int v\rho\right\} \\
&\equiv \min_{\rho}\left(F_{\mathrm{LL}}(\rho) + \int v\rho\right).
\end{aligned} \tag{11.12}$$

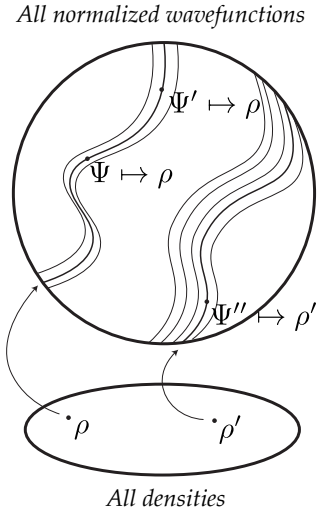
Here, we defined the *Levy–Lieb constrained-search universal functional*

$$F_{\mathrm{LL}}(\rho) \equiv \min_{\Psi \mapsto \rho} \langle \Psi | \hat{T} + \hat{W} | \Psi \rangle \,. \tag{11.13}$$

The functional is termed *universal*, since it does not depend on the molecule in consideration, only the fact that we have N electrons.

We can visualize the constrained search as follows, see Figure 11.2.3.

At the bottom, we have the space of all N-representable densities. At the top, we have the space of all normalized N-electron wavefunctions. Since each wavefunction has a density, the latter space gets divided into strata of wavefuntions that share the same density. Thus, we can nest the minimizaton over all wavefunctions into an outer minimization over densities and an inner minimization of the stratum of all wavefunctions sharing that density.



Visualization of constrained-search division of the unit sphere in Hilbert space.

We have now achieved something remarkable, being the core of DFT: We have replaced the variational principle, which minimized over very complicated N-electron wavefunctions, with the Levy–Lieb variational principle, which minimizes over *densities*. The set of all densities is much more easy to describe and model than the set of all wavefunctions. Of course, there is not really any free lunch here – the universal functional F_{LL} is *extremely* ill-behaved and complicated. Indeed, to evaluate it we must reconstruct all the wavefunctions with a given density. This is a hard task, and basically no easier than solving the original Schrödinger equation.

On the other hand, there might be very useful approximations to F_{LL} that produce relatively easy-to-solve computational problems. This will be discussed in the next lecture, on Kohn–Sham DFT.

11.2.4 *Interpretation of the Levy–Lieb functional*

Let us briefly interpret the Levy–Lieb functional F_{LL}. Suppose v is a potential with a ground-state wavefunction Ψ_0 whose density is ρ_0. Then

$$\begin{aligned} E(v) &= \langle \Psi_0 | \hat{T} + \hat{W} + \hat{V} | \Psi_0 \rangle \\ &= \langle \Psi_0 | \hat{T} + \hat{W} | \Psi_0 \rangle + \int \rho_0 v \\ &= F_{\mathrm{LL}}(\rho_0) + \int \rho_0 v. \end{aligned} \tag{11.14}$$

Thus,

$$F_{\mathrm{LL}}(\rho_0) = \langle \Psi_0 | \hat{T} + \hat{W} | \Psi_0 \rangle \,. \tag{11.15}$$

The Levy–Lieb functional is therefore interpreted as *the intrinsic energy* of a density, since it measures kinetic plus electron repulsion energy wheneger ρ_0 is a ground state density.

11.2.5 *Critical point condition*

Whenver we are given a minimization problem with a differentiable objective function, the minimum must satisfy a critical point

condition. In our case, we are minimizing

$$G(\rho) = F_{\mathrm{LL}}(\rho) + \int v\rho$$

over all N-representable ρ. If F_{LL} were differentiable, we would have

$$\frac{\delta}{\delta\rho} F_{\mathrm{LL}}(\rho) + v = 0, \tag{11.16}$$

where we used the *functional derivative*, which generalizes the partial derivative to "functions of functions". The functional derivative of a functional $A = A(\rho)$, when it exits, is defined as a *function* $\delta A(\rho)/\delta\rho$ that satisfies

$$\lim_{\epsilon\to 0^+} \epsilon^{-1} \left(A(\rho + \epsilon\eta) - A(\rho)\right) = \int \frac{\delta A(\rho)}{\delta\rho} \eta(\mathbf{r})\, d\mathbf{r}, \tag{11.17}$$

for all possible functions $\eta(\mathbf{r})$. We will not have the need for any deeper understanding of the functional derivative than it being a generalization of the partial derivative of a multivariate function. Also, we used the fact that

$$\frac{\delta}{\delta\rho} \int v\rho = v.$$

Now, we have claimed that F_{LL} is highly pathological, and in fact it is not differentiable at all. But virtually all DFT textbooks assumes that it is so, and calculates its derivatives with abandon. We will follow this practice, and merely mention that it can all be made perfectly rigorous. For that, one needs advanced mathematical theory, namely that of subdifferentiation of lower-semicontinuous convex functionals. The interested student is directed towards the seminal research article by Lieb (1983).[1]

[1] E.H. Lieb. "Density Functionals for Coulomb Systems". In: *Int. J. Quant. Chem.* 24.3 (1983), pp. 243–277. ISSN: 0020-7608

It is a fact that at the end of the day, we will do our calculations with *differentiable model functionals* instead of F_{LL}. This further justifies the manipulation of expressions with functional derivatives that don't exist in the usual sense.

Returning to Eq. (11.16), we see that ρ is the ground-state density of the Hamiltonian $\hat{H}(v)$ if

$$-v = \frac{\delta}{\delta\rho} F_{\mathrm{LL}}(\rho).$$

This is, of course, assuming that we are allowed to take derivatives. This relation will be key in the derivation of the Kohn–Sham approach to DFT.

11.2.6 *Fork in the road*

We have come to a proverbial fork in the road. There are two ways forward to produce computable approximations. The conventional aproach is *Kohn–Sham DFT*. It is the undispited workhorse of quantum chemistry, and described in the next lecture. Kohn–Sham DFT became rapidly popular after its invetion, but progress has stalled

somewhat. As a consequence, *Orbital-free DFT* is the less conventional, yet increasingly popular alternative. Its goal is to find useful approximations to the universal functional F_{LL} and directly solve Eq. (11.16).

11.3 Hohenberg–Kohn theory

We now give a presentation of the original "school of thought" in DFT, based on the Hohenberg–Kohn theorems.

11.3.1 The first HK theorem

> **First Hohenberg–Kohn Theorem:** Consider two systems of N electrons in external potentials $v_1(\mathbf{r})$ and $v_2(\mathbf{r})$, respectively. Let the ground state densities be $\rho_1(\mathbf{r})$ and $\rho_2(\mathbf{r})$, respectively, and assume that the ground state wavefunctions are non-degenerate. Then, $\rho_1(\mathbf{r}) = \rho_2(\mathbf{r})$ everywhere implies that $v_1(\mathbf{r}) = v_2(\mathbf{r}) + \mu$, where μ is some constant.

In other words, for a density $\rho(\mathbf{r})$ with N electrons that is the ground-state density of some potential, the potential is uniquely given up to a constant μ.

This striking result has an important consequence: Let $\rho(\mathbf{r})$ be the ground-state density of *some* potential. Since then $v(\mathbf{r})$ is unique (up to μ), the Hamiltonian $\hat{H}(v)$ is also unique (up to a constant potential). The ground-state wavefunction can then in principle be found, and this function does *not* depend on the arbitrary constant. All expectation values of any observable can now be computed. In conclusion:

> **Corollary:** If $\rho(\mathbf{r})$ is a ground-state density of a potential with a non-degenerate ground state, the potential (and this molecule) and all its physical properties are uniquely given by the density.

This observation has led researchers to think of the density as "determining everything", and therefore being a "basic variable" of quantum chemistry.

Is the HK theorem mathematically rigorous? Modulo some formal points omitted in the original paper by Hohenberg and Kohn[2] (and in most DFT texts), yes. However, for many years, the proof of the HK theorem contained a flaw, and mathematicians have spent a lot of effort patching this non-rigorous argument. Only recently has it been established once and for all.[3].

There is a catch, though: That the potential is essentially uniquely given by the density does not mean that the potential is *continously given*. There are examples of densities ρ_1 and ρ_2 that are arbitrarily close together, but whose potentials v_1 and v_2 actually differ by

[2] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas". In: *Phys. Rev.* 136.3B (Nov. 1964). Publisher: American Physical Society, B864–B871. DOI: 10.1103/PhysRev.136.B864. URL: http://link.aps.org/doi/10.1103/PhysRev.136.B864

[3] L. Garrigue. "Unique Continuation for Many-Body Schrödinger Operators and the Hohenberg–Kohn Theorem". In: *Math. Phys. Anal. Geom.* 21 (2018), p. 27

arbitrarly *large* amounts. Thus, very close densities can describe wildly different N-electron systems! This catch makes the intuitive power of the HK theorem less useful in rigorous arguments.

11.3.2 *The HK mapping*

The function $v \mapsto \rho$ that takes potentials into their ground-state densities is well-defined, for all densities that actually have ground states. The *inverse mapping* is provided by the HK theorem: $\rho \mapsto \{v + \mu \mid \mu \in \mathbb{R}\}$, meaning that the inverse mapping maps to a class of potentials.

11.3.3 *Proof of the first HK theorem*

We start with a lemma: If v_1 and v_2 are potentials that have the same non-degenerate ground-state wavefunction Ψ, then $v_1 = v_2 + \mu$ for some constant.

The lemma is proven by contradiction: Let v_1 and v_2 be two potentials that differ by more than a constant, i.e., $v_1 \neq v_2 + \mu$, yet they share the same ground state Ψ. We have two Schrödinger equations

$$\hat{H}(v_1)\Psi = E(v_1)\Psi,$$

and

$$\hat{H}(v_2)\Psi = E(v_2)\Psi.$$

Subtracting these, we get

$$[\hat{V}_1 - \hat{V}_2 - E(v_1) + E(v_2)]\Psi = 0.$$

Dividing by the wavefunction, we see that $\hat{V}_1 = \hat{V}_2 + E(v_1) - E(v_2)$, and therefore $v_1 = v_2 + (E(v_1) - E(v_2))/N$, a contradiction. Therefore, it is not possible that the potentials are different by more than a constant yet the wavefunctions are the same. Thus, the lemma is proven.

We note that dividing by Ψ is only allowable if $\Psi \neq 0$ almost everywhere. This was the non-rigorous part of the original HK theorem and proof. Patching this required the effort of very clever mathematicians over many years.

Armed with the lemma, we prove the main theorem. Again we argue by contradiction. Suppose that $v_1 \neq v_2 + \mu$ and that their ground state densities satisfy $\rho_1 = \rho_2$. We want to show that this leads to a contradiction. We know from the lemma that $\Psi_1 \neq \Psi_2$, the ground states for the two potentials. Since Ψ_2 is *not* the ground state of $\hat{H}(v_1)$ (by non-degeneracy assumption), the variational principle gives

$$\begin{aligned} E(v_1) &< \langle \Psi_2 | \hat{H}(v_1) | \Psi_2 \rangle = \langle \Psi_2 | \hat{H}(v_2) | \Psi_2 \rangle + \langle \Psi_2 | \hat{V}_1 - \hat{V}_2 | \Psi_2 \rangle \\ &= E(v_2) + \int \rho(v_1 - v_2), \end{aligned} \tag{11.18}$$

where we used that the densities are the same by assumption. By symmetry we similarly obtain

$$E(v_2) < E(v_1) + \int \rho(v_2 - v_1). \tag{11.19}$$

Adding the two inequalities,

$$E(v_1) + E(v_2) < E(v_1) + E(v_2),$$

where the density integrals canceled. This is clearly a contradiction, showing that it is not possible to have potentials differing my more than a constant with the same ground-state densities. The converse must then be true, and the proof is complete.

11.4 *Second HK theorem*

We already know of *one* unversal functional, F_{LL} derived using constrained-search methodology. The original foundation for DFT was, however, the following:

> **Second Hohenberg–Kohn Theorem:** There is a unique universal functional $F_{\text{HK}}(\rho)$, defined on the set of ground-state densities, such that
>
> $$E(v) = \min_{\rho} \left(F_{\text{HK}}(\rho) + \int v\rho \right),$$
>
> where the minimum is taken over the set of ground-state densities.

The second HK theorem is easily obtained using our preparations, and the HK mapping. Let ρ be a density, assumed to have be v-representable. Then its potential is v_ρ, up to a constant μ, but the ground-state wavefunction Ψ_ρ is unique. We define

$$F_{\text{HK}}(\rho) \equiv E(v_\rho) - \int \rho v_\rho.$$

We now let v be an arbitrary potential with a ground state Ψ, and ρ be an arbitrary v-representable density. It has associated a unique ground-state wavefunction Ψ_ρ and a potential v_ρ. If ρ is not the density of v, then $v_\rho \neq v + \mu$, for every constant μ. Then, by the lemma leaning up to HK1, the wavefunction $\Psi_\rho \neq \Psi$ as well. Thus,

$$\begin{aligned} E(v) &< \langle \Psi_\rho | \hat{H}(v) | \Psi_\rho \rangle \\ &= \langle \Psi_\rho | \hat{H}(v_\rho) | \Psi_\rho \rangle - \langle \Psi_\rho | \hat{V}_\rho - \hat{V} | \Psi_\rho \rangle \\ &= E(v_\rho) - \int \rho(v_\rho - v) \\ &= F_{\text{HK}}(\rho) + \int \rho v \end{aligned} \tag{11.20}$$

On the other hand, if $v_\rho = v$, then we clearly have

$$E(v) = F_{\text{HK}}(\rho) + \int \rho v,$$

by definition of $F_{\mathrm{HK}}(\rho)$. Thus,

$$E(v) = \min_{\rho} \left(F_{\mathrm{HK}}(\rho) + \int \rho v \right), \tag{11.21}$$

as claimed. Q.E.D.

11.4.1 *Admissible functionals*

Even though we have two very similar variational principles for the ground-state density, there are subtle differences:

$$E(v) = \min_{\rho} \left(F_{\mathrm{LL}}(\rho) + \int v\rho \right), \tag{11.22}$$

and

$$E(v) = \min_{\rho} \left(F_{\mathrm{HK}}(\rho) + \int v\rho \right). \tag{11.23}$$

The LL functional must agree with the HK functional at the densities that happen to be ground-state densities. Otherwise, the ground state energy would change. However, the LL functional is defined on a *larger* set than the HK functional, since we did not require that ρ was a ground-state density for its definition!

There are other universal functionals as well, that all lead to the same $E(v)$ when used in minimization. The precise definition *does not matter* for practical DFT. We shall henceforth use the notation $F(\rho)$, and call it "the" universal functional.

11.5 *Supplemental: Lieb theory (not lectured)*

The material scratched here is detailed in the saminal article by Lieb [4].

[4] E.H. Lieb. "Density Functionals for Coulomb Systems". In: *Int. J. Quant. Chem.* 24.3 (1983), pp. 243–277. ISSN: 0020-7608

We have been a little non-rigorous in our presentation so far. One particular issue is that we have uncritically written "min" when minimizing over densities. We have implicitly assumed that our potential had a ground state Ψ, and then we know that $E(v) = \langle\Psi|\hat{H}(v)|\Psi\rangle$. But there are potentials (i.e., molecules) that do not have ground states! Consider for example trying to place 100 electrons in the hydrogen potential.

In general therefore, the variational principle must be written

$$E(v) = \inf_{\Psi} \langle\Psi|\hat{H}(v)|\Psi\rangle,$$

where the minimization is carried out over all normalized wavefunctions of finite kinetic energy. This in turn leads to the LL variational principle taking the form

$$E(v) = \inf_{\rho\in\mathcal{I}_N} \left(F_{\mathrm{LL}}(\rho) + \int v\rho \right).$$

The infimum is taken over all N-representable densities.

We now need to consider *Lebesgue spaces*. These are vector spaces of functions, with norms that make them complete, i.e., *Banach*

spaces. The interested reader is directed towards the excellent book by Kreyszig[5].

[5] Erwin Kreyszig. *Introductory functional analysis with applications*. en. New York: Wiley, 1978. ISBN: 978-0-471-50731-4

Let $1 \leq p < +\infty$, and $\Omega \subset \mathbb{R}^n$ be an open domain, i.e., a "nice set" without boundary. The Lebesgue space $L^p(\mathbb{R}^3)$ is defined as the set of all functions $f : \Omega \to \mathbb{R}$ such that

$$\|f\|_p := \left(\int_\Omega |f(x)|^p \, d^n x \right)^{1/p} < +\infty. \tag{11.24}$$

This defines a norm, and $L^p(\Omega)$ is complete under this norm. (There is also a subtle point glossed over here, that $f \in L^p(\Omega)$ only defined up to a set of *measure zero*, since such sets do not contribute to the integral. Thus f and the function $\tilde{f}$ obtained by changing f at, say, a countable number of points actually represent the *same* vector in our Banach space.) For $p = +\infty$ one says that $f : \Omega \to \mathbb{R} \in L^\infty(\mathbb{R}^3)$ if there is a constant $C \geq 0$ such that $|f(x)| \leq C$ almost everywhere.

"Almost everywhere", often abbreviated, "a.e.", means that the condition given can be violated violated at most at a set of measure zero.

The set of N-representable densities can be described mathematically as

$$\mathcal{I}_N = \{\rho \in L^1(\mathbb{R}^3) \mid \rho \geq 0 \text{ a.e., } \nabla \rho^{1/2} \in L^2(\mathbb{R}^3)\}. \tag{11.25}$$

Lieb demonstrated that for every $\rho \in \mathcal{I}_N$ we also have $\rho \in L^3(\mathbb{R}^3)$. Thus,

$$\rho \in X \equiv L^1(\mathbb{R}^3) \cap L^3(\mathbb{R}^3).$$

The space X is a Banach space with norm

$$\|u\|_X = \|u\|_1 + \|u\|_3.$$

For the integral $\int v\rho$ to make sense, v must be sufficiently regular. Moreover, we want molecular potentials (Coulomb potentials) to be meaningful. Lieb chose the space X precisely because the dual space X' consists precisely of those functions $v : \Omega \to \mathbb{R}$ for which the integral $\int v\rho$ makes sense for every $\rho \in X$. Furthermore, X' consists of functions on the form

$$v = v_{3/2} + v_\infty,$$

where $v_{3/2} \in L^{3/2}(\mathbb{R}^3)$ is the *short range part* and $v_\infty \in L^\infty(\mathbb{R}^\infty)$ is the *long range part*. Indeed, the hydrogen Coulomb potential can be written

$$\frac{1}{r} = \frac{1}{r}\exp(-r) + \frac{1}{r}(1 - \exp(-r)).$$

It is readily verified that the first term is in $L^{3/2}$ (by direct integration) and that the second term is bounded. Similarly, any molecular nuclear potential is in X'.

Lieb now defined a universal functional $F_\mathrm{L} : X \to \mathbb{R} \cup \{+\infty\}$ (where we allow the functional to take infinite values!) by the expression

$$F_\mathrm{L}(\rho) = \sup_{v \in X'} \left(E(v) - \int \rho v \right). \tag{11.26}$$

This functional can be proven to be such that

$$E(v) = \inf_{\rho \in X} \left(F_{\mathrm{L}}(\rho) + \int \rho v \right). \tag{11.27}$$

The Lieb functiononal is the *smallest* functional that gives the correct ground-state energy $E(v)$ when used in the minimization procedure. Among the admissible funcitonals it is also the only convex and lower-semicontinuous functional. However, it is not differentiable, just like the HK and LL functionals ....

The take-home message is the following:

> In DFT, the ground-state problem for a potential v is obtained as a Legendre transform of a universal functional $F(\rho)$,
>
> $$E(v) = \min_{\rho} \left(F(\rho) + \int v\rho \right).$$
>
> This functional is not unique, and we have seen in particular the Levy–Lieb constrained-search functional, the Hohenberg–Kohn functional, and the Lieb functional. The specific choice does not matter in practice, since we will introduce model functionals at the end of the day.

12 Lecture 12: Density-functional theory, continued

12.1 Kohn-Sham theory

12.1.1 Solving a non-interacting problem

Central to KS-DFT is the introduction of a "fictitious" non-interacting system (abbreviated 's') in an "effective potential" v_s. We will come back to later what exactly this means, but for the moment we ask ourselves: If I have a general external potential $v_\mathrm{s}(\mathbf{r})$, what is the ground-state wavefunction of N non-interacting electrons in this potential?

The solution is obtained via *separation of variables* and making sure that the Pauli principle is satisfied. The N-electron Hamiltonian is

$$\hat{H}_\mathrm{s} = \sum_i \hat{h}_\mathrm{s}(i), \quad \hat{h}_\mathrm{s} = -\frac{1}{2}\nabla^2 + v_\mathrm{s}(\mathbf{r}).$$

Suppose we can diagonalize $\hat{h}_\mathrm{s}$ completely, or at least find the first few eigenfunctions:

$$\hat{h}_\mathrm{s}\varphi_p(\mathbf{r}) = \epsilon_p\varphi_p(\mathbf{r}), \quad p = 1, 2, \ldots$$

$$\epsilon_1 \leq \epsilon_2 \leq \cdots$$

$$\int \overline{\varphi_p(\mathbf{r})}\varphi_q(\mathbf{r})\, d\mathbf{r} = \delta_{pq}$$

Then we can build spin-orbitals $\chi_{2p-1} = \varphi_p\,|\alpha\rangle$ and $\chi_{2p} = \varphi_p\,|\beta\rangle$ in the usual maner. A Slater determinant built with these spin-orbitals will be an eigenfunction of $\hat{H}_\mathrm{s}$:

$$|p_1 p_2 \cdots p_N\rangle = \frac{1}{\sqrt{N!}}\det[\chi_{p_i}(j)].$$

$$\hat{H}_\mathrm{s}\,|p_1 p_2 \cdots p_N\rangle = E_\mathrm{s}\,|p_1 p_2 \cdots p_N\rangle\,, \quad E_\mathrm{s} = \sum_{i=1}^{N}\epsilon_{p_i}$$

The *ground state* of $\hat{H}_\mathrm{eff}$ has minimal energy. This is obtained by using the N lowest-energy spin-orbitals, i.e., by doubly occupying the $N/2$ first φ_p. The *density* of this ground state is.

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2}|\varphi_i(\mathbf{r})|^2$$

12.1.2 *DFT for non-interacting electrons*

We will now use the symbol F for "the" universal functional. It is not so important which one we choose.

Recall that the universal functional $F(\rho)$ minimizes the intrinsic energy of ρ, and was defined for a system of *interacting electrons*. What happens if the starting system was described with $\hat{H}_{\text{eff}}$, which has no interactions between the electrons? We can still use constrained-search as before,

$$E_{\text{s}}(v_s) = \min_{\rho} \min_{\Psi \mapsto \rho} \langle \Psi | \hat{T} | \Psi \rangle + \int \rho v_{\text{s}} = \min_{\rho} F_s(\rho) + \int \rho v_s$$

Here, we introduced the subscript "s" on the energy functional and the intrinsic energy to distinguish notationally from the interacting "true" system.

Now, we observe that we *know* the ground-state wavefunction and its density:

$$\Psi_{\text{s}} = |\varphi_1 \overline{\varphi_1} \cdots \varphi_{N/2} \overline{\varphi_{N/2}}\rangle \,.$$

$$\rho_{\text{s}}(\mathbf{r}) = 2 \sum_{i=1}^{N/2} |\varphi_i(\mathbf{r})|^2$$

This is assuming that we can solve the "trivial" single-electron problem. We then obtain

$$F_{\text{s}}s(\rho_{\text{s}}) = T_{\text{s}}(\rho_{\text{s}}) = \langle \Psi_{\text{s}} | \hat{T} | \Psi_{\text{s}} \rangle = \sum_{i=1}^{N/2} \langle \nabla \varphi_i, \nabla \varphi_i \rangle \,.$$

Thus, when we *know* that ρ_{s} is the ground-state density of *merely some* potential, then $F_{\text{s}}(\rho_{\text{s}})$ has an explicit functional form. The universal functional is known for non-interacting electrons!

12.1.3 *Kohn-Sham splitting of the universal functional*

We return to the interacting system. Kohn and Sham introduced the idea that one could separate

$$F(\rho) = T_s(\rho) + E_{\text{Hxc}}(\rho) = T_s(\rho) + E_{\text{H}}(\rho) + E_{\text{xc}}(\rho).$$

Here, E_{H} is the Hartree energy,

$$E_{\text{H}}(\rho) = \frac{1}{2} \int \frac{\rho(\mathbf{r}_1)\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \, d\mathbf{r}_1 d\mathbf{r}_2.$$

Furthermore, $E_{\text{xc}}(\rho)$ is called the "exchange-correlation energy", and is *defined* by the above formula as $E_{\text{xc}}(\rho) = F(\rho) - T_s(\rho) - E_{\text{H}}(\rho)$. The idea is that the functional $F(\rho)$ should predominantly be given by the non-interacting kinetic energy and the Hartree energy, everything else, being the hard to model part, is $E_{\text{xc}}(\rho)$.

Recall that the Levy-Lieb variational principle states that

$$E(v) = \min_{\rho} \left(F(\rho) + \int v\rho \right).$$

To find the minimum we differentiate and set equal to zero:

$$\frac{\delta}{\delta \rho} F(\rho) + v = 0.$$

This is an equation where ρ is the unknown.

We now note that

$$\frac{\delta}{\delta\rho}E_{\mathrm{H}}(\rho) = V_{\mathrm{H}}(\rho)(\mathbf{r}) = \int \frac{\rho(\mathbf{r_2})}{|\mathbf{r}-\mathbf{r}_2|}\,d\mathbf{r}_2.$$

This is the classical electrostatical potential set up by the electrons. We also note that

$$\frac{\delta}{\delta\rho}E_{\mathrm{xc}}(\rho) \equiv V_{\mathrm{xc}}(\rho) = ???$$

This functional derivative is unknown. As discussed in the previous lecture, it is not even well-defined. But we shall pretend that we are given a *model* for the exchange-correlation potential, and that this model is differentiable. The derivative is called the exchange-correlation potential. So, assuming that we can differentiate everything, we get:

$$\frac{\delta}{\delta\rho}T_s(\rho) + \underbrace{V_H(\rho) + V_{\mathrm{xc}}(\rho) + v}_{\equiv v_{\mathrm{s}}} = 0$$

We now observe that this is the DFT minmization problem for the non-interacting system with potential v_{s}. But this is exactly solvable as above; we simply find the $N/2$ first eigenfunctions of the single-electron operator

$$\hat{h}_{\mathrm{s}} = -\frac{1}{2}\nabla^2 + v_{\mathrm{s}}(\mathbf{r}).$$

$$\hat{h}_{\mathrm{s}}\varphi_i(\mathbf{r}) = \epsilon_i\varphi_i(\mathbf{r}).$$

$$\rho(\mathbf{r}) = 2\sum_{i=1}^{N/2}|\varphi_i(\mathbf{r})|^2.$$

12.1.4 *Self-consistent field iterations*

We note that v_{s} depends on the density ρ. The eigenfunctions coming out out $\hat{h}_{\mathrm{eff}}$ must have the same density. Indeed, we have solved our ground-state problem in the Kohn–Sham formulation *precisely* when we have a self-consistent solution:

1. Set $n = 0$, and choose a density guess $\rho^{(0)}$.
2. Compute $v_{\mathrm{s}}^{(n)} = v_{\mathrm{s}}(\rho^{(n)}) = v + v_{\mathrm{H}}(\rho^{(n)}) + v_{\mathrm{xc}}(\rho^{(n)})$
3. Solve non-interacting Schroedinger equation with potential $v_{\mathrm{s}}^{(n)}$ to obtain orbitals $\varphi_i^{(n+1)}$ and the corresponding density $\rho^{(n+1)}$.
4. Is the new density sufficiently close to the old density? If yes, we have obtained our ground-state density. The energy is computed by
$$E(v) = T_s(\rho) + E_{\mathrm{H}}(\rho) + E_{\mathrm{xc}}(\rho).$$
If no, set $n \leftarrow n+1$ and go to 2.

This procedure is analogous to the SCF iterations of Hartree–Fock. In fact, it is precisely SCF iterations, but now the exchange energy/potential is replaced by the exchange correlation energy/potential!

This explains the term "exchange correlation": It contains not only exchange as in Hartree–Fock, but also correlation, as DFT describes correlations between electrons.

12.1.5 *The wavefunction is not a Slater determinant*

One fact is important to be aware of: Even though we are solving for a Slater determinant in the non-interacting system, this system is not real, i.e. it is fictitious and only a mathematical device to get a computable approximation. The actual ground state wavefunction Ψ for the *interacting system* just happens to have the same density as the fictitious system. The wavefunction is correlated, and is a sum of many Slater determinants using the Kohn–Sham orbitals. But the density is ρ.

Thus, in DFT, *we do not have a wavefunction!*

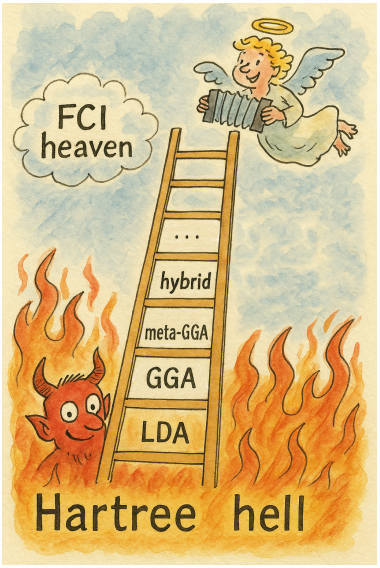
One should therefore be careful with interpreting the obtained MOs physically.

12.2 *Approximate functionals*

Having established the basic SCF procedure for Kohn-Sham DFT, the next issue is the determination of exchange correlation energy models.

John Perdew introduced "Jacob's Ladder" as a pedagogical device to illustrate the levels of approximation in DFT. At the bottom we have "Hartree hell", and at the top we have "FCI heaven". The lowest rung of the ladder is LDA. Then comes GGA, meta-GGA, and finally hybrid functionals.

Using a religious metaphor like this reminds us that the hierarchy of increasingly more complicated functionals in DFT can be viewed as wishful thinking. The functionals are rarely rigorously defined; rather they are semi-empirical in nature. Research can be claimed to have stalled the last decades, and the most widely-used functionals such as B3LYP, a hybrid functional, were developed already in the 1990s!



Jacob's Ladder in DFT, visualized with the aid of ChatGPT and some Photoshop.

12.2.1 *Local density approximation (LDA)*

The exact XC functional is of course extremely complicated, and will depend on ρ in a highly nonlocal manner. On the other hand, the *simplest* models we can make use only the denisty at local points.

The starting point is the assumption that locally the system behaves like a uniform electron gas. That is, $\rho(\mathbf{r})$ varies *slowly* in space. Thus, we seek the exchange correlation energy functional

of a gas of electrons with constant (uniform) density. The general functional form in LDA is

$$E_{\text{xc}}^{\text{LDA}}(\rho) = \int \rho(\mathbf{r})\epsilon_{\text{xc}}(\rho(\mathbf{r}))\, d\mathbf{r}.$$

Here, $\epsilon_{\text{xc}}(\rho)$ is the exchange correlation energy per particle and is an unknown function to be determined. Th The e corresponding exchange-correlation potetial becomes

$$V_{\text{xc}}^{\text{LDA}}(\rho)(\mathbf{r}) = \frac{\delta E_{\text{xc}}^{\text{LDA}}(\rho)}{\delta\rho} = \epsilon_{\text{xc}}(\rho(\mathbf{r})) + \rho(\mathbf{r})\epsilon'_{\text{xc}}(\rho(\mathbf{r})).$$

It is custumary to write $\epsilon_{\text{xc}} = \epsilon_{\text{x}} + \epsilon_{\text{c}}$. The exchange part can be obtained analytically for the uniform electron gas. The result is:

$$\epsilon_{\text{x}}(\rho) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\rho^{1/3}.$$

The total LDA exchange energy becomes

$$E_{\text{x}}^{\text{LDA}} = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int \rho(\mathbf{r})^{4/3}\, d\mathbf{r},$$

which is also called the Dirac exchange energy functional for the uniform electron gas. We have

$$V_{\text{x}}^{\text{LDA}}(\rho)(\mathbf{r}) = -\left(\frac{3}{\pi}\right)^{1/3}\rho(\mathbf{r})^{1/3}.$$

The correlation energy density $\epsilon_{\text{c}}(\rho)$ is not analytically solvable, however. It is most commonly obtained by fitting analytical formulas to advanced quantum Monte Carlo simulations, for example. The interested student can look up Parr and Yang's book "Density-Functional Theory of Atoms and Molecules", Appendix E, for actual expressions

12.2.2 *Applicability of LDA*

The basic assumption of LDA, that the density is slowly varying, is often justified in solids, but not in atoms and molecules. Still, LDA can provide quite good results.

12.2.3 *The Xα approximation*

To illustrate LDA, we consider a very old method by Slater, who introduced a simplification of the Hartree–Fock method called the $X\alpha$ method in 1951. This was done in the 1920s, way before Kohn and Sham. The idea was that the exchange term in Hartree–Fock is the by far most complicated term to compute. Recall

$$\hat{f} = -\frac{1}{2}\nabla^2 + v(\mathbf{r}) + \underbrace{v_{\text{H}}(\mathbf{r}) + \hat{v}_{\text{x}}}_{\hat{v}^{\text{HF}}}.$$

Recall that the exchange operator $\hat{v}_{\text{x}}$ in Hartree–Fock is a non-local operator, necessitating complicated and expensive integrals for evaluating the action of $\hat{f}$ on an orbital.

Slater took results from the uniform electron gas and proposed to replace the exchange operator by the Dirac exchange term, leading to the $X\alpha$ method. He introduced a scaling factor α to the exchange term,

$$E_{\mathrm{x}}^{X\alpha}(\rho) = \alpha E_{\mathrm{x}}^{\mathrm{LDA}}(\rho).$$

$$v_{\mathrm{x}}^{X\alpha} = -\alpha \left(\frac{3}{\pi}\right)^{1/3} \rho(\mathbf{r})^{1/3}.$$

Thus, HF and KS-DFT are both SCF problems. We can turn HF into an approximate LDA version of KS-DFT, where correlation is completely neglected. This is the $X\alpha$ method.

12.2.4 *Generalized-gradient approximation (GGA)*

In LDA, it was assumed that the exchange and correlation energies per particle only depended on the local density $\rho(\mathbf{r})$. However, this is a too simple approximation. The next step would be to introduce the gradient of the density in the energy per particle:

$$\epsilon_{\mathrm{xc}}^{\mathrm{GGA}} = \epsilon_{\mathrm{xc}}^{\mathrm{GGA}}(\rho, \nabla\rho),$$

leading to a total energy

$$E_{xc}^{\mathrm{GGA}}(\rho) = \int \rho(\mathbf{r}) \epsilon_{\mathrm{xc}}^{\mathrm{GGA}}(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))\, d\mathbf{r}.$$

To obtain the exchange-correlation *potential*, we must compute the functional derivative, which becomes more involved than before. We will not detail it here.

Various forms of $\epsilon_{\mathrm{xc}}^{\mathrm{GGA}}$ lead do different GGA functionals. What does the energy density typically look like? We introduce a dimensionless quantity $s(\mathbf{r})$ called *relative inhomogenity*,

$$s(\mathbf{r}) = \frac{|\nabla\rho(\mathbf{r})|}{2k_F(\mathbf{r})\rho(\mathbf{r})},$$

where the *local Fermi wavenumber of the uniform electron gas* is

$$k_F(\mathbf{r}) = (3\pi^2\rho(\mathbf{r}))^{1/3}$$

and measures the density-dependent characteristic momentum scale: For a uniform gas with density $\rho(\mathbf{r})$, all plane waves with momenta up to $k_F(\mathbf{r})$ are occupied. Thus,

$$s(\mathbf{r}) \sim \frac{\text{density gradient}}{\text{momentum scale} \times \text{density}},$$

measures the inhomogenity of the density. For example, if the density is totally uniform, then $s(\mathbf{r}) = 0$. The larger $s(\mathbf{r})$, the faster the density changes locally.

This measure is used to build GGA functionals:

$$\epsilon_{\mathrm{x}}^{\mathrm{GGA}}(\rho, \nabla\rho) = \epsilon_{\mathrm{x}}^{\mathrm{LDA}}(\rho) \cdot F_x(s),$$

where $F_{\mathrm{x}}(0) = 1$, recovering LDA at the uniform limit. An example of a GGA functional on this form, is the Berdew-Becke-Ernzerhof

(PBE) functional, which is perhaps the most popular GGA functional today.

GGA functionals improve LDA in several respects. In particular *atomization energies* and *bond lengths* are greatly improved.

12.2.5 *Meta-GGA functionals*

The next level of approximation involved bringing in the orbitals directly. These are not truly related to the *wavefunction* of the interacting system, but it is reasonable to assume that some information is coded in the orbitals. For example, the kinetic energy density should be useful:

$$\tau(\mathbf{r}) = \sum_{i=1}^{N/2} |\nabla \varphi_i(\mathbf{r})|^2.$$

The kinetic energy of the Slater determinant of the noninteracting system is now $\int \tau(\mathbf{r})\ d\mathbf{r}$. In meta-GGA functionals, we have the exchange-correlation energy density form

$$\epsilon_{\text{xc}}^{\text{meta-GGA}}(\rho, \nabla\rho, \tau),$$

leading to the exchange-correlation energy

$$E_{\text{xc}}^{\text{meta-GGA}}[\rho] = \int \rho(\mathbf{r}) \epsilon_{\text{xc}}^{\text{meta-GGA}}(\rho(\mathbf{r}, \nabla\rho(\mathbf{r}), \tau(\mathbf{r}))) \, d\mathbf{r}.$$

One nice thing with the kinetic energy density $\tau(\mathbf{r})$ is that one can distinguish regions with covalent bonds or metallic regions, as well as the tail region of electron densities. This allows for functionals to distinguish these regions by having different behavior for different kinetic energy densities.

Two examples of standard meta-GGA functionals are TPSS (Tao–Perdew–Staroverov–Scuseria, 2003), and SCAN (Strongly Constrained and Appropriately Normed, 2015). These functionals have no empirical parameters.

12.2.6 *Hybrid functionals*

Hybrid functionals take the viewpoint that the "exact" *nonlocal* exchange of Hartree–Fock will give a useful correction to the exchange energy. In Hybrid functionals one sets

$$E_{\text{xc}}^{\text{hybrid}} = a E_{\text{x}}^{\text{HF}} + (1-a) E_{\text{x}}^{\text{DFT}} + E_{\text{c}}^{\text{DFT}},$$

where "DFT" indicates one of the above described types of semilocal exchange and correlation energy functionals. Hybrid functionals become expensive due to the Hartree–Fock exchange energy, which is fully nonlocal, and are admittedly ad hoc in structure, since the HF wavefunction has nothing to do with the actual interacting wavefunction, except for having the same density.

The most famous hybrid functional is B3LYP. PBE0 is also a widely-used funtional.

12.2.7 *Common DFAs and Their Properties*

This section is generated with ChatGPT, and the information is verified.

Functional	Type	Exchange-Correlation	Key Features	Notes
Slater (Xα)	LDA	Local exchange only (scaled uniform gas)	Simplest local approximation	Historical, used before full DFT
VWN (Vosko-Wilk-Nusair)	LDA	Uniform gas correlation	Local correlation parametrized from QMC	Standard LDA correlation functional
PBE (Perdew-Burke-Ernzerhof)	GGA	Nonempirical exchange and correlation	Satisfies known exact constraints	Very widely used for solids
PW91 (Perdew-Wang 1991)	GGA	Nonempirical exchange and correlation	Early constraint-based GGA	Predecessor to PBE
BLYP (Becke88 + LYP)	GGA	Empirical exchange and correlation	Fit to atomization energies and molecular properties	Popular in quantum chemistry
TPSS (Tao-Perdew-Staroverov-Scuseria)	meta-GGA	Nonempirical exchange and correlation	Adds kinetic energy density $\tau(\mathbf{r})$	First fully nonempirical meta-GGA
SCAN (Strongly Constrained and Appropriately Normed)	meta-GGA	Nonempirical exchange and correlation	Satisfies all 17 exact constraints	Very accurate across diverse systems
M06-L	meta-GGA	Empirically fitted exchange-correlation	Strong fit to molecular data	Useful in chemistry; less physical rigor
B3LYP	Hybrid	20% HF exchange + BLYP	Semiempirical hybrid	Extremely popular in quantum chemistry
PBE0	Hybrid	25% HF exchange + PBE	Nonempirical hybrid	More physically justified than B3LYP
HSE (Heyd-Scuseria-Ernzerhof)	Hybrid (Range-separated)	Short-range HF exchange + PBE	Screened exchange for solids	Good for band gap prediction

12.3 *Quick Trends*

Aspect	**LDA**	**GGA**	**meta-GGA**	**Hybrid**
Locality	Local	Semi-local	Semi-local + orbital density	Nonlocal (exact exchange)
Accuracy for molecules	Poor	Better	Very good	Excellent
Accuracy for solids	Reasonable	Very good (PBE)	Very good (SCAN)	Very good (HSE)
Band gaps	Badly underestimated	Underestimated	Still underestimated	Improved (esp. with range separation)
Cost	Very low	Low	Moderate	High

13 Lecture 13: Computational cost

This lecture also introduced a recap of the methods introduced so far. This recap is skipped in these notes.

13.1 About computational cost

In quantum chemistry we are faced with the problem of solving the N-electron Schrödinger equation. The FCI problem has combinatorial storage requirements due to the length of the coefficient vector. We have to bring the cost down, and this can be viewed as the driving force behind the development of all the various quantum chemical methods. This has also been the main driving force behind choosing GTOs as basis functions in quantum chemistry.

Thus computational cost is a big issue. This is the topic of today's lecture.

Computational cost has several sides, among which we find:

- What are the memory requirements of a calculation?
- What is the cost of performing the calculations?
- To what degree can a calculation be parallelized, i.e., distributed over several CPUs, GPUs, or supercomputer nodes?

13.1.1 Single precision floating point numbers

When calculating numerically, computers use *floating point numbers* with *finite precision*. For the following discussion, recall that 1 byte (B) = 8 bits.

A *single precision* floating point number typically occupies 4 B (32 bits) of memory. It is structured as:

$$(-1)^s \times 1.f \times 2^{e-127}$$

- s: 1 bit, giving the sign,
- e: 8 bit integer, i.e., 1 B, for the exponent (with a bias of 127), which gives exponents in the range -127 to $+128$
- f: 23 bits for the fraction (mantissa), with an implicit leading 1 in normalized numbers.

This gives about *7–8 decimal digits* of precision and a range of approximately 10^{-38} to 10^{38} for the absolute values of numbers.

Note that floating point numbers come with limitations. For example, the number 10^{100} cannot be represented in single precision. Moreover, the result of $1 + 10^{-9} = 1$ in finite precision, since 10^{-9}, while well in the range of what single precision numbers can represent, is too small to be visible in the 23 bits allowed for the fraction part of a number.

13.1.2 *Double Precision Floating Point Numbers*

A `float` in Python is a double precision floating point number, by default.

A *double precision* floating point number uses 8 B (64 bits), and is vastly more expressible than single precision numbers:

$$(-1)^s \times 1.f \times 2^{e-1023}$$

- s: 1 bit, giving the sign,
- e: 11 bit integer for the exponent (with a bias of 1023), which gives exponents in the range -1023 to $+1024$
- f: 52 bits for the fraction (mantissa), with an implicit leading 1 in normalized numbers.

Double precision numbers have about *15–17 decimal digits* of precision and a range of roughly 10^{-308} to 10^{308} for the absolute values of numbers.

Most calculations are today done in double precision, since modern computer architecture is optimized for this. An (important) exception is calculations on graphics processor units (GPUs), typical in machine learning, where single precision is used.

Virtually all calculations that produce a floating point number as result can be decomposed into a number of additions/subtractions and multiplications/divisions. In practice we count *multiplications* since these are by far the most costly on modern computers. This gives a rough measure, but in practice it is sufficient.

13.1.3 *Computer memory*

Computer memory is a complex topic, but let us summarize the most important qualitative features: The computer's *random access memory (RAM)* is where numbers are stored when the computer is working with them. RAM is fast but volatile: when the computer is turned off, RAM is wiped. A typical laptop has 16-32 GB of RAM, which must hold the operating system (OS) and all running software, as well as its data, including numerical data for quantum chemical algorithms.

The IBM 3081 mainframe class supercomputer introduced 1980 had 16 MB of RAM and apparently cost around 3-5 million USD at the time. Yes, this whole setup is the computer, including printer, tape drives, terminals, etc. For an interesting personal viewpoint, see https://www.tomshardware.com/picturestory/508-mainframe-computer-history-2.html (Photo: Norsk Teknisk Museum.)

The *hard drive* provides long-term storage for programs and files, and uses different technology than RAM. Previously, hard drives used magnetic disks and moving parts for reading and writing. They were therefore relatively slow and noisy. Today, more and more hard drives are *solid state drives (SSDs)*, which have no moving parts, physically much smaller, and much faster. However, SSDs are

still *much slower* than RAM. Nowadays, hard drives have capacity in the terabyte (TB) range.

The operating system also provides *virtual memory*, which means that a chunk of the hard drive is used as *swap space*, allowing to emulate extra RAM. When RAM is full – which happens a lot in large simulations – the computer can move data between virtual memory (on the disk) and RAM, referred to as *swapping*. This drastically slows the computer down.

13.1.4 *Storage requirements of vectors, matrices, and tensors*

A vector of double precision floating point numbers (which is default in Python) requires 8 B per element. Thus, for a vector of length N we need a memory of

$$N \times 8 \text{ B}$$

In contrast, an $N \times N$ matrix requires

$$N^2 \times 8 \text{ B}.$$

More generally, a tensor with m dimensions of equal length N requires

$$N^m \times 8 \text{ B}.$$

The maximum N is dramatically reduced when m increases. Let us look at examples.

In the 1980s, the typical high-end computer had 16 MB of RAM! Jan Almlöf, who worked as professor in Oslo, introduced the idea that ERI tensor elements should be computed *on the fly*. Such versions of HF and other algorithms are called "direct".

What is the largest matrix that can fit in 32 GB RAM, assuming that *all* of the laptop's memory was available?

$$32 \times 10^9 = 8N^2 \implies N = 4^{0.5} \times 10^{0.5 \cdot 9} \approx 6.3 \times 10^3$$

This is a pretty large matrix. Consider now the storage of the electron-repulsion integral (ERI) tensor, $(\mu\nu|r_{12}^{-1}|\alpha\beta)$, which is a 4-index symbol. With N atomic orbitals, we need to store N^4 floats. In 32 GB we can fit

$$32 \times 10^9 = 8N^4 \implies N = 8^{0.25} \times 10^{0.25 \cdot 9} \approx 3.0 \times 10^2$$

Thus, with merely $N = 300$ AO basis functions, RAM is full! (Clearly, some tricks are needed.)

13.1.5 *Counting FLOPs*

A *floating point operation* (FLOP) is a single arithmetic calculation involving two double precision floating point numbers done on the computer, such as addition or multiplication/division. Most calculations can be broken down into arithmetic operations as well as look-up tables. The main cost of a calculation consists of the multiplications/divisions. Thus, if we count these, we get a picture of the cost of a calculation.

For example, consider the multiplication of an $N \times N$ matrix with an N-vector, $y = Ax$. A single element y_i is

$$y_i = \sum_{j=1}^{N} A_{ij} x_j.$$

This comprises N multiplications (and N additions). However, since there are N elements of y, the *complete* matrix-vector product costs N^2 multiplications.

Consider next the multiplication of A with another $N \times N$ matrix B, i.e., $C = AB$. A single element of C is

$$C_{ij} = \sum_{k=1}^{N} A_{ik} B_{kj}.$$

This is again N multiplications, but now there are N^2 elements, totalling N^3 multiplications.

(One can also see this by noting that B consists of N column vectors, so we get N times the cost of a single matrix-vector product.)

Other examples include the diagonalization of a matrix A:

```
E, U = np.linalg.eigh(A)
```

The `eigh` function calls a highly optimized library called LAPACK, which implements an algoritm that requires an operation count proportional to N^3. However, wheras the matrix-matrix product was exactly N^3 operations, the matrix diagonalization has a *prefactor* larger than 1. Thus, matrix diagonalization in is *slower* than matrix-matrix multiplication, but *asymptotocically*, i.e., for large matrices, they are proportional in cost.

13.1.6 *Big-O notation*

We are interested in the asymptotic behavior of calculations, as function of the characteristic size parameters. This can be a matrix dimension, number of particles, number of AOs, etc. A mathematical tool is the *big-O notation*.

The matrix-matrix product and matrix diagonalizations are examples of $O(N^3)$ costs. If $f(N)$ counts the number of operations, we say that

$$f(N) = O(N^p)$$

if $f(N)$ is dominated by CN^p for some constant C when N is large. Mathematically,

$$|f(N)| \leq CN^p, \quad \text{large } N.$$

Thus, matrix-vector multiplication is $O(N^2)$.

In general, the cost can be a complicated function, for example the number of multiplications could be found to be

$$f(N) = 1 + 20N + 2N^3.$$

This is $O(N^3)$ since this is the highest power present, and the term $2N^3$ will be *much larger* than the other terms when N grows. For

example,

$$f(10) = 2,002,001, \quad 2 \cdot 10^3 = 2,000,000,$$
$$f(10,000) = 2,000,000,200,001, \quad 2 \cdot 10,000^3 = 2,000,000,000,000.$$

There are also other types of asymptotic behavior of interest. For example, the fast Fourier-transform (FFT) algorithm, which transforms a signal of length N into its frequency components, is $O(N \log N)$, meaning that

$$f_{\text{FFT}}(N) \leq CN \log N, \quad \text{large } N.$$

Note that $N \leq N \log N \leq N^2$ whenever $N > 1$, so this cost is *between* $O(N)$ and $O(N^2)$.

13.1.7 *Scaling is not all*

Suppose we have two algorithms A and B for the calculation of some quantity. We find that A is $O(N^2)$ and B is $O(N^3)$. Clearly for large N, algorighm A should be prefered. However, for *moderate* N the best algorihm depends on the *prefactor* in the $O(N^p)$ definition. For example, A could be

$$f_{\text{A}}(N) = 1,000N^2,$$

while

$$f_{\text{B}}(N) = N^3.$$

Clearly, for small $N \leq 1,000$, algorithm B will be the winner.

Note that for $f(N) \sim CN^p$, where $\sim$ means that we only exhibit the dominating term in $f(N)$, we have

$$\log f(N) = \log C + p \log N.$$

Therefore, plotting $f(N)$, which will be proportional to execution time, in a double logarihmic plot, we expect straight lines. These lines will cross at some N for A and B, indicating where each algorithm will be the fastest.

13.2 *Computational cost of Hartree–Fock*

We now discuss the computational cost of Hartree–Fock using an AO basis with K elements. We focus on RHF and $N = 2N_{\text{o}}$ electrons. Thus,

$$N_{\text{o}} = \frac{N}{2}, \quad N_{\text{v}} = K - N_{\text{o}}.$$

The important ingredients needed for doing a HF calculation were (see page 85):

- Core matrix $H_{\mu\nu}$:
  - K^2 matrix elements computed by library routines, $O(K^2)$ cost, assuming each matrix element requires roughy the same number of FLOPs. Only done once.

- Electron repulsion integral (ERI) tensor $(\mu\alpha|r_{12}^{-1}|\nu\beta)$
  - Four-index symbol, so K^4 integrals. Computed via library routines, $O(K^4)$ cost. Only done once. ← Bottleneck!
- Density matrix $D_{\mu\nu} = 2\sum_{i=1}^{N_o} C_{\mu i}\overline{C_{\nu i}} = 2(CC^H)_{\mu\nu}$
  - K^2 matrix elements, each costing N multiplications, so $O(NK^2)$. Done in each SCF iteration.
- Hartree/direct potential $J_{\mu\nu} = \sum_{\alpha\beta}^{K} D_{\alpha\beta}(\mu\alpha|r_{12}^{-1}|\nu\beta)$
  - K^2 matrix elements, each being a sum over K^2 terms with one multiplication each. So $O(K^4)$. Done in each SCF iteration. ← Bottleneck!
- Exchange potential $K_{\mu\nu} = \frac{1}{2}\sum_{\alpha\beta}^{K} D_{\alpha\beta}(\mu\alpha|r_{12}^{-1}|\beta\nu)$.
  - Same as the Hartree term. ← Bottleneck!
- Diagonalization of Fock matrix in SCF iterations. This has $O(K^3)$ cost.

We see that since each SCF iteration requires an $O(N^4)$ calculation, *the bottleneck is the assembly of the mean-field matrices J and K* as well as the computation of the ERI tensor. For large systems, these steps will completely dominate, and if one can bring them down to $O(K^3)$, Hartree–Fock can treat an order of magnitude larger systems.

13.2.1 *Density fitting*

... (The rest of this lecture not yet typed up.)

14 Lecture 14: Perturbation theory

Perturbation theory is covered well by Szabo and Ostlund. These lecture notes will be updated later ...

14.1 What is perturbation theory?

Suppose you are given a function $f(x)$ of which you want to find a root. For example,

$$f(x) = 2x + \lambda(1 + \cos(x)), \quad \lambda = \text{parameter}, \tag{14.1}$$

and you wish to find x_0 such that $f(x_0) =$. This equation is hard to solve, in the sense that an analytical solution is not available. However, for small λ, the solution should be close to the one obtained for $\lambda = 0$. *That* problem is easy:

$$2x = 0 \iff x = 0. \tag{14.2}$$

The idea of perturbation theory is now to look for solutions under the assumption that λ is small, using a power series expansion. Thus, write

$$\begin{aligned} x_0 &= x_0^{(0)} + \lambda x_0^{(1)} + \lambda^2 x_0^{(2)} + \dots \\ &= x_0^{(0)} + \Delta x_0, \end{aligned} \tag{14.3}$$

and insert into the root equation:

$$\begin{aligned} f(x_0) &= f(x_0^{(0)} + \Delta x_0) = 0 \\ &= \Delta x_0 f'(x_0^{(0)}) + \frac{(\Delta x_0)^2}{2} f''(x_0^{(0)}) + \dots \end{aligned} \tag{14.4}$$

Here, we used that $f(x_0^{(0)}) = 0$. Inserting $\Delta x_0 = \sum_{n=1}^{\infty} \lambda^n x_0^{(n)}$ into the latter equation leads to a series expansion in powers of λ:

$$f(x_0) = \lambda A_1 + \lambda^2 A_2 + \dots. \tag{14.5}$$

Here, we used that there are no zero-order contributions, since Δx_0 has leading term $\lambda x_0^{(1)}$.

This expansion must vanish for all λ. Therefore, every power of λ must vanish individually.

In our example,

$$f'(x) = x - \lambda \sin(x) \tag{14.6}$$

$$f''(x) = -\lambda \cos(x) \tag{14.7}$$

$$\dots \tag{14.8}$$

This leads to

$$(\lambda x_0^{(1)} + \lambda x_0^{(2)} + \cdots)(x_0^{(0)} - \lambda \sin(x_0^{(0)})) + \frac{1}{2}(\lambda x_0^{(1)} + \lambda x_0^{(2)} + \cdots)(-\lambda \cos(x_0^{(0)})) + \cdots = 0. \tag{14.9}$$

To find $x_0^{(1)}$ is is enough to consider the leading terms:

[To be completed]

14.2 *Rayleigh–Schrödinger perturbation theory for matrix eigenvalue problems.*

RSPT

Matrix eigenvalue perturbation theory

Illustrate with two level system. Example, bonding in HF molecule.

Matrix, in electron volts:

$$H = \begin{pmatrix} \epsilon_H & \beta \\ \beta & \epsilon_F \end{pmatrix} \approx \begin{pmatrix} -13.6 & \beta \\ \beta & -18.6 \end{pmatrix} \tag{14.10}$$

Typical values $\beta \sim -1, -2$ electron volts.

Analyze the exact eigenvalues, estimate location of singularities.

14.3 *Orbital perturbation theory*

RSPT for the many-electron problem when the zero-order and perturbation is a one-body operator. We can reuse results from the previous section.

Why relevant: Application of electric field, perturbation theory in semi-empirical methods (e.g., Hückel theory), and more.

14.4 *Møller–Plesset perturbation theory*

15 Lecture 15: Introduction to multireference theory

15.1 When Hartree–Fock is a bad approximation

In many situations, Hartree–Fock is a *good* approximation, and most of the wavefunction is described by a single Slater determinant Φ_{HF}. Recall that the Slater determinant is an uncorrelated wavefunction (except for unavoidable Pauli correlation). Post-Hartree–Fock methods aim to make corrections to the Slater determinant, i.e., build a *correlated wavefunction*:

$$\Psi = \Phi_{\text{HF}} + \Psi_{\text{corr}} \quad \text{(small correction)} \tag{15.1}$$

This in turn leads to a correction to the HF energy:

$$E = E_{\text{HF}} + E_{\text{corr}} \tag{15.2}$$

By *definition*, E_{HF} is an uncorrelated energy. By the varational principle, $E_{\text{corr}} < 0$, since a better wavefunction always leads to a lower energy.

The post-HF methods we have seen so far are:

- Configuration interaction, where the wavefunction is corrected by *excitations* out of the HF *reference*:

$$|\Psi_{\text{CI}}\rangle = |\Phi_{\text{HF}}\rangle + \sum_{ia} c_i^a \, |\Phi_i^a\rangle + \sum_{i<j} \sum_{a<b} c_{ij}^{ab} \, |\Phi_{ij}^{ab}\rangle \quad \text{(CISD wavefunction)}$$

- Coupled-cluster theory, where the wavefunction is correlated with the exponential ansatz:

$$|\Psi_{\text{CC}}\rangle = \exp(\hat{T}) \, |\Phi_{\text{HF}}\rangle \, .$$

- Møller–Plesset perturbation theory, where the wavefunction and energy is expanded in a perturbation series.

When HF is a good approximation, the correlation arises from *description of the movement of electrons due to the repulsion between them*. This type of correlation is termed *dynamic correlation*.

15.1.1 Static correlation

When HF fails to be a good approximation, we have situation where a single determinant is not a good starting point, and more

Slater determinants are essential for even a meaningful starting point. This termed *static correlation*. In this situation, *Hartree–Fock is qualitatively wrong*.

This arises in several situations:

- Bond breaking: Simple example H_2 molecule
- Diradicals: Simple example, O_2 ground state; a triplet.
- Transition metal compounds: Where an unfilled d shell leads to many closely lying electronic configurations.
- Molecules near conical intersections.
- Conjugated systems.
- Et.c.

We will consider the perhaps simplest example of static correlation in the next subsection.

15.1.2 *Dissociation of H_2 in a minimal basis*

The perhaps simplest example is *dissociation of the hydrogen molecule*, where we have merely 2 electrons:

$$|\Phi_{\text{RHF}}\rangle = |\psi\overline{\psi}\rangle\,,$$

a single doubly occupied orbital $\psi(\mathbf{r})$. RHF produces excellent results near equilibrium, but when the atoms are pulled apart, the RHF wavefunction becmomes physically unreasonable, forcing both electrons to be delocalized, with 50 % probability of residing on each atom. Clearly, this is not the case in reality.

Let us introduce a minimal basis: Each H atom comes with a single 1s function ϕ_A and ϕ_B, and the single MO ψ is a linear combination of these:

$$\psi = c_A\phi_A + c_B\phi_B.$$

One can show that by symmetry, the ground state MO will be

$$\psi(\mathbf{r}) = 1\sigma_g(\mathbf{r}) = \mathcal{N}(\phi_A(\mathbf{r}) + \phi_B(\mathbf{r})),$$

where $\mathcal{N}$ is just the normalization constant of the MO. This is a σ orbital due to rotational invariance around the molecular axis, and it is even ("gerade") when inverted around the center of mass.

There is one single unoccupied/virtual MO with the minimal basis, an antibonding orbital

$$\psi^*(\mathbf{r}) = 1\sigma_u^*(\mathbf{r}) = \mathcal{M}(\phi_A(\mathbf{r}) - \phi_B(\mathbf{r})),$$

where the asterisk reminds us about the antibonding nature of the MO. Again a σ orbital, but it is odd ("ungerade") and has a node – hence higher energy.

At infinite separation, ϕ_A and ϕ_B do not overlap, and $\mathcal{N} \approx 1/\sqrt{2}$, leading to the probability $1/2$ of finding an electron at A and B, respectively. The MO is *delocalized*.

Having dealt with RHF, let us look at UHF. Here, we have *two* MOs, one for the α electron and one for the β electron:

$$|\Phi_{\text{UHF}}\rangle = |\psi^{\alpha}\overline{\psi^{\beta}}\rangle\,.$$

Here,

$$\psi^{\alpha}(\mathbf{r}) = c_A^{\alpha}\phi_A(\mathbf{r}) + c_B^{\alpha}\phi_B(\mathbf{r}),$$
$$\psi^{\beta}(\mathbf{r}) = c_A^{\beta}\phi_A(\mathbf{r}) + c_B^{\beta}\phi_B(\mathbf{r}).$$

Near equilibrium, we recall that UHF became identical to RHF, but that as we stretch the bond *a transition occurs where the symmetry is broken*. In particular, at large separations we obtain

$$\psi^{\alpha}(\mathbf{r}) \approx \phi_A(\mathbf{r}),$$
$$\psi^{\beta}(\mathbf{r}) \approx \phi_B(\mathbf{r}).$$

Here, we see that the electrons are correctly localized on A and B. *However, the wavefunction is no longer a singlet wavefunction*, total spin is no longer a good quantum number.

If we turn to a FCI wavefunction starting from the RHF reference, one can show that it will be:

$$|\Psi\rangle = c_{gg}\,|1\sigma_g\overline{1\sigma_g}\rangle + c_{uu}\,|1\sigma_u^*\overline{1\sigma_u^*}\rangle\,. \tag{15.3}$$

Note that the last Slater determinant is a doubly excited determinant, where $i = 1$, the ground state MO, is promoted to $i = 2$, the excited antibonding MO. The singly excited determinants will not contribute due to symmetry, but are in theory present in the FCI wavefunction.

At infinite separation, we have that

$$1\sigma_g = \frac{1}{\sqrt{2}}(\phi_A + \phi_B), \quad 1\sigma_u^* = \frac{1}{\sqrt{2}}(\phi_A - \phi_B).$$

Inserting the expressions for $1\sigma_g$ and $1\sigma_u^*$ in terms of the atomic orbitals, we obtain:

$$\begin{aligned}
|1\sigma_g\overline{1\sigma_g}\rangle &= |\frac{1}{\sqrt{2}}(\phi_A + \phi_B),\ \frac{1}{\sqrt{2}}(\overline{\phi_A} + \overline{\phi_B})\rangle \\
&= \frac{1}{2}\left(|\phi_A\overline{\phi_A}\rangle + |\phi_A\overline{\phi_B}\rangle + |\phi_B\overline{\phi_A}\rangle + |\phi_B\overline{\phi_B}\rangle\right), \\
|1\sigma_u^*\overline{1\sigma_u^*}\rangle &= |\frac{1}{\sqrt{2}}(\phi_A - \phi_B),\ \frac{1}{\sqrt{2}}(\overline{\phi_A} - \overline{\phi_B})\rangle \\
&= \frac{1}{2}\left(|\phi_A\overline{\phi_A}\rangle - |\phi_A\overline{\phi_B}\rangle - |\phi_B\overline{\phi_A}\rangle + |\phi_B\overline{\phi_B}\rangle\right).
\end{aligned}$$

Inserting into the FCI wavefunction, we get

$$|\Psi\rangle = \frac{1}{2}(c_{gg} + c_{uu})(|\phi_A\overline{\phi_A}\rangle + |\phi_B\overline{\phi_B}\rangle) + \frac{1}{2}(c_{gg} - c_{uu})(|\phi_A\overline{\phi_B}\rangle + |\phi_B\overline{\phi_A}\rangle).$$

The first two terms describe electrons either both at A or both at B, i.e., they are delocalized, whereas the final two terms describe electrons distributed one each on each atom. Therefore, we conclude that $c_{gg} + c_{uu} \to 0$ in the dissociation limit. Thus,

$$|\Psi\rangle \to c_{gg}\left(|\phi_A\overline{\phi_A}\rangle + |\phi_B\overline{\phi_B}\rangle\right),$$

where $c_{gg} = 1/\sqrt{2}$ for normalization. In terms of the bonding and antibonding orbitals,

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|1\sigma_g\overline{1\sigma_g}\rangle - \frac{1}{\sqrt{2}}|1\sigma_u^*\overline{1\sigma_u^*}\rangle\,.$$

which is a proper spin singlet wavefunction describing one electron on atom A and one on atom B, in a superposition consistent with indistinguishability and spin symmetry.

This final state is qualitatively correct: it represents two localized electrons on separate atoms, preserving spin symmetry (unlike UHF), and excluding unphysical charge delocalization (unlike RHF).

Thus, we conclude that a qualitatively correct dissociation limit for H_2 requires a linear combination of at least two Slater determinants — the bonding and antibonding configurations — as found in FCI. Neither RHF nor UHF alone provides a physically accurate and symmetry-consistent description at large bond distances. RHF fails by enforcing delocalization, and UHF fails by breaking spin symmetry. Only a linear combination of slater determinants (here FCI in the minimal basis) simultaneously describes localization and spin symmetry.

15.2 Static correlation

The previous example is an example of *symmetry-driven static correlation*.: In order to have correct symmetry (here, a proper spin eigenfunction), we must have several Slater determinants contributing substantially to the total wavefunction.

15.3 Multiconfigurational Self-Consistent Field theory

In situations where HF is not a sufficiently good approximation, the natural generalization is a multiconfiguratopional approach, which is defined as follows: Consider a set of K (unknown) spin-orbitals χ_p, where $1 \leq p \leq K$. These can form $\binom{K}{N}$ Slater determinants Φ_I, which we can combine to a wavefunction as

$$|\Psi_{\text{MCSCF}}\rangle = \sum_I c_I\,|\Phi_I\rangle\,, \tag{15.4}$$

where the sum runs over all the Slater determinants. Each MO is written as

$$\chi_p(\mathbf{r}) = \sum_\mu C_{\mu p}\phi_\mu(\mathbf{r}), \tag{15.5}$$

with an unknown MO coefficient matrix C.

This wavefunction can now be optimized to yield the lowest possible energy: both the coefficients c_I and the spin-orbital coefficients $C_{\mu p}$ are parameters that can be tuned to yield the best possible result. Since for a *given* MO set, the optimal CI coefficients are the lowest-energy eigenvector components of the CI matrix,

$$\sum_{I'} \mathbf{H}_{II'}c_{I'} = E_{\text{MCSCF}}c_I,$$

where the sum runs over all determinants in the MCSCF wavefunction. On the other hand, the MO coefficients are found by solving an equation that is a generalized mean-field-type equation. In particular it is nonlinear.

The difference with Hartree–Fock is now that there are more than one Slater determinant. The difference with a CI expansion is that now the MOs are optimized in tandem.

This *multiconfigurational self-consisten field* approach is very powerful, but in general it requires expertise to apply.

In the next section we discuss the most widely-used variety: *complete-active space self-consistent field* (CASSCF).

15.4 CASSCF

Consider a set of MOs $\psi_p(\mathbf{r})$. We divide these into 3 classes:

- Inactive/frozen core orbitals $\psi_i(\mathbf{r})$: always doubly occupied in Slater determinants
- Active orbitals $\psi_u(\mathbf{r})$: Occupied or unoccupied in Slater determinants
- Virtual orbitals $\psi_a(\mathbf{r})$: Never occupied in Slater determinants

If we have in total N electrons and N_{core} core orbitals, there are $n = N - 2N\text{core}$ electrons to distribute in the active orbitals, the number of which is N_{active}. The result is denoted CAS(n, N_{active}).

The MO coefficient matrix is now block structured:

$$C = [C^{\text{core}} \mid C^{\text{active}} \mid C^{\text{virtual}}].$$

The matrix C is square: there are as many MOs as there are AOs, just as in Hartree–Fock.

A CAS is illustrated in Figure 15.1 in the margin.

For a given molecule, choosing a CAS is not always an easy task. The CAS must be large enough to resolve static correlation, and small enough to be wieldy. Note that inside the CAS, a full CI expansion is performed. Therefore, CASSCF has *exponential scaling with respect to number of orbitals/electrons*, and is in principle equally costly as is FCI as the numbers of electrons and active orbitals grow! For example, CAS(6,6), which is needed for the breaking of the triple bond of N_2, contains 400 determinants. In Figure 15.2, the dimensions of CAS($2k$,$2k$) is listed. Even 24 electrons is almost completely intractable without further approximations!

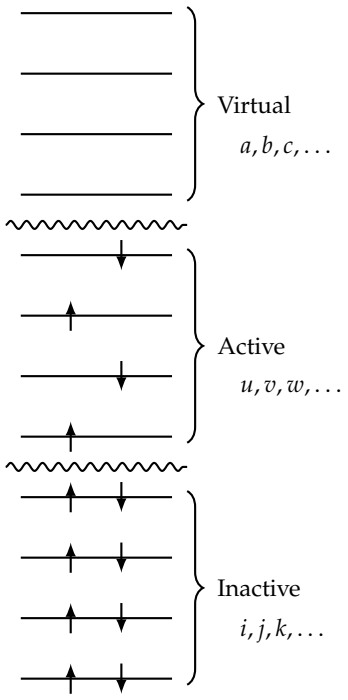


Figure 15.1: A complete-active space CAS(4,4), of 4 electrons in 4 active orbitals. Additionally, there are 8 frozen electrons in 4 inactive orbitals.

$2k$	# SD's
2	4
4	36
6	400
8	4,900
10	63,504
12	853,776
14	11,778,896
16	165,636,896
18	2,363,904,260
20	34,134,777,856
24	7,312,459,672,336

Figure 15.2: Number of Slater determinants for $2k$ electrons in $2k$ spatial orbitals.

15.4.1 Example: Breaking the N_2 triple bond

The nitrogen dimer is a standard poster child for static correlation in molecules, as breaking the triple bond introduces many nearly degenerate Hartree–Fock orbitals. Near degeneracy of the highest occupied MO (HOMO) is a strong indication of static correlation.

In Figure 15.3 you can see the standard MO diagram of the nitrogen dimer at equilibrium. As the interatomic distance is increased, the MOs become closer in energy, until at infinite separation where we end up with 6 degenerate MOs with 6 electrons. This indicates that a CAS(6,6) is suitable, putting the 8 electrons below that into inactive orbitals.

In the cc-pVDZ basis set for the N atom, there are 16 functions: 2 1s functions, 2 2s, 6 2p functions, and 6 3d functions. In total, therefore, there are 32 AOs for the N_2 molecule. Therefore, we can say that for the CAS(6,6) case of the nitrogen dimer we have:

- 4 inactive orbitals (always doubly occupied), $i = 1, \ldots, 4$
- 6 active orbitals with 6 electrons. $u = 5, \ldots, 10$.
- 22 virtual orbitals with no electrons. $a = 11, \ldots, 32$.

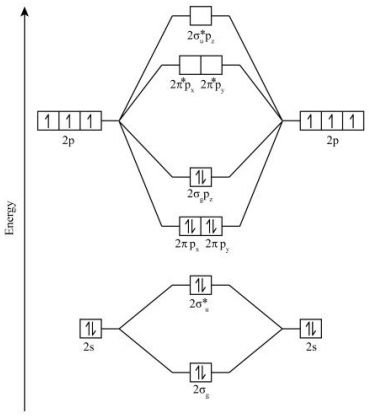


Figure 15.3: The MO diagram of N_2. The 1s orbitals, forming bonding and antibonding $1\sigma_g$ and $1\sigma_u^*$ orbitals are not shown.

15.5 *Multireference configuration-interaction*

While CASSCF seems to be costly, note that one does not try to resolve dynamical correlation with CASSCF, only static correlation. Thus, in "post-MCSCF approaches" one tries to build dynamical correlation *on top of a CAS wavefunction*. Such methods are called *multireference methods* since they do not only have one single reference determinant, but many.

One example of such a method is *multireference configuration-interaction* (MRCI). Here, a wavefunction on the following form is sought:

$$|\Psi_{\text{MRCI}}\rangle = \sum_{I \in \text{CAS}} d_I |\Phi_I\rangle + \sum_{J \in \text{ext}} d_J |\Phi_J\rangle \tag{15.6}$$

Let us explain what this all means: $I \in$ CAS means that we sum over the determinants that are in the CAS, simply according to the rules described above. Thus, the d_I should be close to the c_I from CASSCF, but not equal! We are after all seeking a *correction* on top of CASSCF.

The remaining terms $J \in$ ext, for "external space", sums over a *selection* of determinants that *violate* the rules of CAS. There are two ways to accomplish this:

- One or more electrons are moved *from core orbitals* $i, j, k \ldots$.
- One or more electrons are moved *to virtual orbitals* $a, b, c, \ldots$.

In short, $J \in$ ext means that we have *at least one external label*, where a label is external if it is either inactive/core or virtual.

In FCI, $J \in$ ext takes on *all* possible values so that absolutely all Slater determinants are found. However, truncated MRCI yields vastly fewer determinants. It is typical to truncate at MRCISD, where at most 2 external labels are found in J, as well as the "first-order interaction space" (FOIS), consisting of all those determinants that are coupled to the CASSCF reference via the two-electron

interaction. The FOIS determinants are needed to obtain results that are correct up to second order in perturbation theory (more on this later).

In MRCI we do excitations from the CAS *wavefunction* into excited determinants: In MRCISD we do up to singly and doubly excited detereminants. Note carefully, that we are exciting from the CAS reference wavefunction, and not individual CAS determinants.

In standard single-refrence CI, we obtain a matrix diagonalization problem, and the same is true for MRCI. The MRCISD matrix has the following form:

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}_{\text{CAS}} & \mathbf{V} \\ \mathbf{V}^\dagger & \mathbf{H}_{\text{SD}} \end{pmatrix} \tag{15.7}$$

Here, $\mathbf{H}_{\text{CAS}}$ is the Hamiltonian matrix with respect to the CAS determinants only.

$$\mathbf{H}_{\text{CAS},II'} = \langle \Phi_I | \hat{H} | \Phi_{I'} \rangle \,.$$

It is a fact that the CASSCF wavefunction is an eigenvector of this matrix block:

$$\sum_{I' \in \text{CAS}} \mathbf{H}_{\text{CAS},II'} c_{I'} = E_{\text{CASSCF}} c_I. \tag{15.8}$$

The matrix $\mathbf{V}$ with elements $\mathbf{V}_{IJ}$, where $I \in$ CAS and $J \in$ ext, are defined by

$$\langle \Phi_I | \hat{H} | \Phi_J \rangle \,.$$

Importantly, *only* those J that satisfy

$$\langle \Psi_{\text{CASSCF}} | \hat{H} | \Phi_J \rangle \neq 0 \tag{15.9}$$

are included. This reduces the size of the matrix needed to compute without introducing any additional approximations. *This defines the SD excitation determinant set.* Initially one might think that all excitations would couple to Ψ_{CAS}, but for example symmetries may reduce the set of J that couple quite a bit.

The matrix $\mathbf{H}_{\text{SD}}$ is the matrix of the Hamiltonian $\hat{H}$ between determinants J, J' that are singly or doubly excited from a CAS determinant, and J, J' must satisfy Eq. (15.9).

MRCISD always contains the first-order interaction space, by construction.

15.5.1 *Example: MRCI for H_2*

...

15.5.2 *Example: MRCI for N_2*

We return to the cc-pVDZ example from above. We assume that a CAS(6,6)SCF calculation has been done, leading to a pretty good CAS wavefunction. This wavefunction does *not* describe dynamical correlation; it only resolves the static correlation due to near-degeneracies of the Hartree–Fock MOs.

...

Index